Back to Reality for Imitation Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract: Imitation learning, and robot learning in general, emerged due to breakthroughs in machine learning, rather than breakthroughs in robotics. As such, evaluation metrics for robot learning are deeply rooted in those for machine learning. In this paper, we expose a worrying trend that this has led to, with a meta-analysis of imitation learning papers accepted at CoRL 2020. We show that traditional evaluation metrics, which only encourage data efficiency, do not consider how the robot learning technique would actually be implemented in practice, and instead encourage methods that are tailored to simulation environments and benchmarks. To counteract this, we propose that the most appropriate evaluation metric is not data efficiency, but time efficiency, which captures the real-world cost much more truthfully. This is a call to arms for the robot learning community to re-consider our evaluation metrics, and develop our own metrics tailored towards the long-term application of real-world robotics.

Keywords: Imitation Learning, Evaluation, Benchmarking

1 Introduction

How do we define the term *robot learning*? Let us consider two options. Firstly, *The study of machine learning techniques for applications to robotics*. And secondly, *The study of techniques to enable robots to learn*. The choice between these can be distilled down to whether robot learning is considered fundamentally to be a sub-field of machine learning, or a sub-field of robotics, respectively. But whilst this distinction may appear to be subtle, the implication is profound. It determines the evaluation metrics used to judge the performance of new techniques, and therefore, it determines the long-term impact that the field is shaping itself around. In this paper, we argue that in order to deliver real-world impact, these metrics should be driven by the target application of robotics, rather than the underlying machine learning methods. However, this is not what we are observing in robot learning today: evaluation is dominated by traditional machine learning metrics. In this paper, we expose a worrying trend that this has led to, which deserves foresight, attention, and debate.

Robot learning emerged from machine learning, not robotics. Whilst the fields of both machine learning and robotics have existed for decades, the field of robot learning is a more recent development, such as with the arrival of the first CoRL conference in 2017. But its emergence can largely be attributed to breakthroughs in machine learning rather than breakthroughs robotics, and in particular, breakthroughs in deep learning and reinforcement learning. As such, robot learning has adopted the same evaluation metrics as the machine learning community. A typical reinforcement learning paper published at a machine learning conference today, will present graphs of *Success Rate vs Number of Environment Interactions*, and similarly for a typical robot learning paper published at CoRL [1]. For the machine learning community, this is sensible: it encourages data efficiency, which is arguably the most fundamental aim in machine learning. But is data efficiency the most important metric for applications to robotics?

Simulation benchmarks are a double-edged sword. One of the main reasons why robot learning has so readily adopted machine learning metrics, is the dominance of simulation-based benchmarks, such as OpenAI Gym [2] and DeepMind Control Suite [3], particularly for methods based on reinforcement learning. These enable rapid testing of new methods, and quantitative comparison of

Submitted to the 5th Conference on Robot Learning (CoRL 2021). Do not distribute.

different methods. However, these simulations typically focus exclusively on data efficiency, and fail to model two important aspects of real-world robot learning. The first, is that many simulation benchmarks assume ground-truth state representations, such as object poses for a manipulation task. To obtain such states in reality, an image-based pose estimator would need to be pre-trained for each task, which requires data and human supervision. The second, is that most simulation benchmarks assume that resetting the environment after an episode of learning, is achieved at zero cost. In reality, this would require either manual environment resetting and thus constant human supervision, or construction of experimental apparatus [4] for each new task to enable automated resetting.

A choice of two futures. To motivate the remainder of this paper, let us consider the future of robot learning if current evaluation metrics persist. Table 1 shows evaluation results for three different hypothetical methods, each of which involves imitation learning and potentially a period of reinforcement learning. Numbers in the table indicate the point during training at which each method achieves the same test-time performance. According to today's evaluation metrics, Method A is superior to Method B, since it requires the least amount of data. However, Method A also requires continual resetting of the environment, whereas Method B only requires occasional resetting. As such, Method B is in fact a faster overall method, and additionally requires less human effort. Similarly, based on today's evaluation metrics, Method A would typically be considered superior to Method C due to its ability to learn from a single demonstration. But in reality, Method C requires significantly less human effort. The aim of the rest of the paper is to motivate the community towards a future where Methods B and C are both considered superior to Method A, and avoid a future that we are likely to otherwise arrive in, where data efficiency comes at the cost of significant real-world impracticalities.

Method	# Human Demos	# Exploration Episodes	# Environment Resets	Human Supervision Time	Overall Training Time
A	1	10k	10k	1.67k	1.70k
B	1	100k	1k	167	500
C	100	0	0	17	17

Table 1: Hypothetical imitation learning methods. Times are in minutes, based on: exploration time per episode = 5s, human time per demo = 10s, human time per environment reset = 10s.

2 Criteria for real-world imitation learning impact

Imitation learning [5] can be defined as a technique by which a robot can learn a new skill, by observing human demonstrations. This is familiar to many robot learning methods when we consider real-world robot learning. For example, a goal image is often used to define a reward function, and in reality that goal image would be achieved by the human actually performing a demonstration of the task, even if demonstrations are not explicitly required [6]. Therefore, a significantly large number of robot learning methods can be considered as a form of imitation learning, when we move away from simulation benchmarks with pre-defined reward functions, and into the real world. We therefore choose imitation learning as a case study in this paper, and we now define three criteria which we believe a good imitation learning method should offer.

Limited human expertise. A future assistive robot may come out of the factory with a repertoire of basic skills, such as grasping objects and opening doors. However, more complex tasks may need to be learned directly from the user. Therefore, imitation learning methods should be able to operate without requiring the human to have an in-depth understanding of the underlying algorithm. For example, whilst kinesthetic teaching is intuitive and requires little human expertise, setting up apparatus to enable automatic resetting of the environment after each task failure, or curating a dataset for a pre-trained object pose estimator, require a level of technical knowledge from the user.

Clock time efficiency. Traditional machine learning metrics consider only data efficiency, without considering how efficiently that data can be collected in the first place. We believe that a better metric is clock time efficiency, which we define as the total amount of wall clock time a robot takes to learn a new task. For example, a method which requires 30 minutes to learn a new task based on 20k environment interactions, should be considered superior to a method which requires 60 minutes based on 10k environment interactions, when tested on the same hardware.

Human time efficiency. Traditional machine learning metrics do not consider the human effort in supervising the learning process. In typical simulation environments, the costs of setting up experimental apparatus for each new task, and continual resetting of the environment, are entirely ignored. We believe that an important metric is the overall time required by a human during supervision of task learning, which is a good proxy to the overall human effort required. For example, a method which requires 5 minutes of human supervision to learn a new task based on 20k environment interactions, could be considered superior to a method which requires 10 minutes of human supervision based on 10k environment interactions. Now, it is challenging to objectively evaluate our first criteria, that of the amount of human expertise required. However, as a proxy to this, we can simply use the amount of time a non-expert user would need to teach a robot a new task, and this is our human time efficiency metric. Therefore, it is important that this includes not only the time required to provide demonstrations and reset the environment, but any preparation time required for each new task, such as any additional apparatus setup, or training of object pose estimators.

3 Meta-analysis of imitation learning evaluation today

Having established our criteria for good imitation learning methods, we can now analyse the value of existing evaluation metrics used by the robot learning community today. To do this, we performed a meta-analysis of all imitation learning papers published at CoRL 2020. To select these papers, we first searched for all papers which contained either "imitation" or "demonstration" in the paper's abstract. From these, we inspected each paper and retained those which actually involved providing a human demonstration at test time to learn a new task, which resulted in 19 papers. We then analysed each paper to determine whether it evaluated success rate as a function of either clock time or human time. We also recorded whether a paper assumed ground-truth object poses, and whether it required environment resetting. In some cases, the experiments themselves did not require environment resetting but only because the tasks were simple target reaching tasks, and in these cases we made a judgement as to whether environment resetting would be required with more typical, everyday tasks. For each paper, we also recorded whether any real-world experiments were done.

Table 2 shows the results of this meta-analysis. Papers are grouped into 8 groups, where all the papers in one group contained the same yes/no answers across all 5 columns. The first observation we make is that not a single paper evaluated the clock time efficiency or human time efficiency, with most methods instead evaluating data efficiency. Some papers did involve studying success rate as a function of the number of demonstrations, but in those cases, the overall clock time or human time was not evaluated. The second observation we make is that 10 out of the 19 papers assume access to ground-truth object poses, and 13 papers required continual environment resetting. Neither of these properties make for practical, real-world imitation learning. To obtain these poses, some papers required manually pre-training an object pose estimator for each new task [7, 8], but timing information for this was not provided. A final observation we make is that 8 out of the 19 papers only include simulation experiments, an unfortunate trend at all CoRL conferences which results in authors focussing mainly on machine learning metrics, rather than those more appropriate to real-world robotics.

# Papers	Clock Time Efficiency?	Human Time Efficiency?	Assumes Object Poses?	Environment Resetting?	Real-World?
4 [9, 10, 11, 12]	No	No	Yes	Yes	No
3 [13, 14, 8]	No	No	Yes	Yes	Yes
3 [15, 16, 1]	No	No	No	Yes	No
3 [17, 7, 18]	No	No	No	Yes	Yes
3 [19, 20, 21]	No	No	Yes	No	Yes
1 [22]	No	No	No	No	Yes
1 [23]	No	No	No	No	No
1 [24]	No	No	No	No	Yes

Table 2: Meta-analysis of 19 imitation learning papers at CoRL 2020.

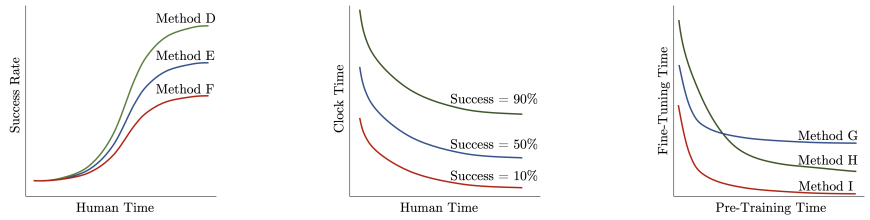
4 A better direction for long-term, real-world impact

Having exposed the flaws in today's evaluation metrics for robot learning, we now propose an alternative direction, with the aim of bringing imitation learning back to reality. In line with our previous arguments, we propose that instead of evaluating data efficiency, imitation learning papers

should evaluate both clock time efficiency, and human time efficiency, when real-world robotics applications are the long-term goal. Time efficiency itself still considers data efficiency, since data collection takes time; but it is a more comprehensive assessment of the true cost to us as humans.

Figure 1 presents various options for visualising these metrics, on hypothetical data. Figure 1a compares three different methods by plotting success rates as a function of human time. This human time should include any time during which the human is physically engaged, such as providing demonstrations, resetting the environment, and setting up apparatus. Whilst Method D has the highest human time efficiency, it is not necessarily the most data efficient, recalling the discussion of Table 1. Therefore, it would not necessarily be considered a desirable method if traditional machine learning metrics were used. Figure 1b then presents a more complete evaluation of one individual method. Here, each line is a Pareto front representing a particular success rate on the task being learned, where every point on one line equally trades off the amount of human time required, and the overall clock time required. This is a useful visualisation for choosing an operating point during deployment, where a user may have a particular preference for minimising overall training time, compared to minimising overall human effort or expertise.

Figure 1c then shows a third way of visualising results, in this case showing the trade-off between pre-training and fine-tuning across three different methods. For example, meta-learning methods typically involve a large amount of pre-training, whereas other methods may learn entirely from scratch. Again, a Pareto front represents all data for a particular success rate. This graph provides an intuitive way to compare methods across entirely different families of imitation learning approaches, which is typically not possible when considering only one of these dimensions. Here, Method I outperforms both Methods G and H, and we can see that Method H is faster than Method G at learning new tasks when there is an abundance of pre-training time, and vice versa.



(a) (b) (c)

Figure 1: Examples of how evaluations could be presented using our clock time efficiency and human time efficiency metrics.

Due to the large number of evaluations required to plot these graphs, particularly those based on Pareto fronts, evaluation would likely need to make use of simulations. These simulation benchmarks for imitation learning should define, for each task, an estimated clock time required for all components of an imitation learning framework, including the time taken for each demonstration and for each environment reset. We believe that by imposing these real-world costs on benchmarks, which are currently ignored in simulation benchmarks, the community will begin to develop methods more in tune with real-world robotics applications.

5 Conclusions

In this paper, we have exposed a worrying trend in the evaluation of imitation learning methods. Due to the historical relationship between robot learning and machine learning, evaluation of imitation learning is typically based on data efficiency. However, we have highlighted that this is not an appropriate metric when the long-term goal is real-world robotics. Instead, two alternative aspects should be considered: the overall time taken for learning a new task, and the amount of time spent by the human supervising the learning. Methods which are the most data efficient may not necessarily be the most time efficient, and may sacrifice consideration of real-world practicalities in an attempt to beat the state-of-the-art on simulation benchmarks. Our intention through this paper, is to stimulate the community to change the current direction of imitation learning evaluation, in an attempt to shape the field, and the methods it develops, towards the real-world scenarios under which imitation learning will actually be deployed in the long-term.

References

[1] Augmenting gail with bc for sample efficient imitation learning. In *Conference on Robot Learning (CoRL)*, 2020.

[2] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. Openai gym. *CoRR*, abs/1606.01540, 2016. URL http://arxiv.org/abs/1606.01540.

[3] Y. Tassa, Y. Doron, A. Muldal, T. Erez, Y. Li, D. de Las Casas, D. Budden, A. Abdolmaleki, J. Merel, A. Lefrancq, T. P. Lillicrap, and M. A. Riedmiller. Deepmind control suite. *CoRR*, abs/1801.00690, 2018. URL http://arxiv.org/abs/1801.00690.

[4] D. Kalashnikov, A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke, and S. Levine. Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. *CoRR*, abs/1806.10293, 2018. URL http://arxiv.org/abs/1806.10293.

[5] T. Osa, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, J. Peters, et al. An algorithmic perspective on imitation learning. *Foundations and Trends in Robotics*, 2018.

[6] G. Schoettler, A. Nair, J. Luo, S. Bahl, J. A. Ojea, E. Solowjow, and S. Levine. Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[7] Model-based inverse reinforcement learning from visual demonstrations. In *Conference on Robot Learning (CoRL)*, 2020.

[8] Acnmp: Skill transfer and task extrapolation through learning from demonstration and reinforcement learning via representation sharing. In *Conference on Robot Learning (CoRL)*, 2020.

[9] f-irl: Inverse reinforcement learning via state marginal matching. In *Conference on Robot Learning (CoRL)*, 2020.

[10] Tolerance-guided policy learning for adaptable and transferrable delicate industrial insertion. In *Conference on Robot Learning (CoRL)*, 2020.

[11] Learning dexterous manipulation from suboptimal experts. In *Conference on Robot Learning (CoRL)*, 2020.

[12] Learning from demonstrations using signal temporal logic. In *Conference on Robot Learning (CoRL)*, 2020.

[13] Generative adversarial training of product of policies for robust and adaptive movement primitives. In *Conference on Robot Learning (CoRL)*, 2020.

[14] Learning from suboptimal demonstration via self-supervised reward regression. In *Conference on Robot Learning (CoRL)*, 2020.

[15] Task-relevant adversarial imitation learning. In *Conference on Robot Learning (CoRL)*, 2020.

[16] Positive-unlabeled reward learning. In *Conference on Robot Learning (CoRL)*, 2020.

[17] Cloud: Contrastive learning of unsupervised dynamics. In *Conference on Robot Learning (CoRL)*, 2020.

[18] Generalization guarantees for imitation learning. In *Conference on Robot Learning (CoRL)*, 2020.

[19] Modeling long-horizon tasks as sequential interaction landscapes. In *Conference on Robot Learning (CoRL)*, 2020.

[20] Interactive imitation learning in state-space. In *Conference on Robot Learning (CoRL)*, 2020.

[21] Learning interactively to resolve ambiguity in reference frame selection. In *Conference on Robot Learning (CoRL)*, 2020.

[22] Learning object manipulation skills via approximate state estimation from real videos. In *Conference on Robot Learning (CoRL)*, 2020.

[23] Transformers for one-shot visual imitation. In *Conference on Robot Learning (CoRL)*, 2020.

[24] Visual imitation made easy. In *Conference on Robot Learning (CoRL)*, 2020.