On Assessing the Value of Simulation for Robotics

Liam Paull[1], Anthony Courchesne[1]

***Abstract*— With the emergence of data-driven techniques, two uses of simulators for robotics are often conflated: 1) its ability to be a faithful predictor of agent performance and 2) its ability to be a useful tool for learning. In this abstract, we attempt to formalize the *value* of a simulator for robotics tasks. In particular, we discuss the different sources of discrepancy (reality gap) between simulators and the real world counterparts. Furthermore, we argue that the value of a simulator is condition on the *task* that it is being used to help solve.**

I. Introduction

Historically, the robotics community has had an ambivalent relationship with robot simulators. The sentiment that "simulators are doomed to succeed" has been very prevalent and has consequently labeled simulation-based robotics as "unscientific". The recent advances in data-driven approaches to solve tasks with embodied agents could possibly cause us to re-evaluate this position. Still a proxy for the real world, the simulator is now being used to facilitate agent learning, and hopefully minimize the number of trials that have to be run on the real hardware.

The discrepancy between a simulation and the real environment that it is meant to represent is sometimes referred to as the "reality gap" ([1, 2] among many others). If we can accurately evaluate this gap it is a good proxy for assessing the value of a simulator. The vast majority of works that address the simulation-to-real world transfer learning problem ("sim2real") propose methods for *crossing* this gap without actually explicitly *quantifying* it ([3]–[5] among many others). As a result, algorithms that show good performance in some settings (e.g., where the gap is small) may completely fail in others, and there are no methods of predicting transferability *a priori*. It is also possible that if we can evaluate the reality gap, we can backpropagate this as an error signal into a differentiable simulation (e.g., [6, 7] among many others) in order to minimize it.

Our objective here is to build a framework for *predicting* how useful a specific simulator is to help solve a specific real-world task.

A. Background and Prior Work

A naive definition of the reality gap based on the standard agent/environment abstractions (Fig. 1) would be:

*Def. 1 (**Naive Reality Gap**):* The "difference" in the resulting observations produced by the simulator and the real robot hardware for a predefined sequence of control commands.

There are several issues with Def. 1:

1) It combines in an opaque way the various sources of the discrepancy. Referring to Fig. 1, there could be a "gap" in the dynamics, the environment model (for example how other agents move in the environment), or in the generation of sensor data based on a rendering model. Moreover, errors in upstream models will compound.
2) It presupposes that real-world fidelity is needed. In practice, we only require a form of task-conditioned fidelity: only the things that are important to solve the task at hand must be faithfully reproduced in the simulation.

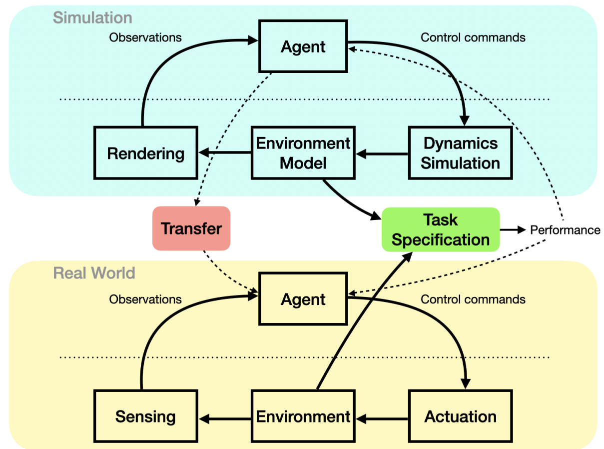

Fig. 1. The agent/environment interface for simulation (top) and the real robot (bottom). An agent receives observations and generates control commands. The **predictive** value of a simulator lies in its ability to faithfully reproduce an estimate of the task performance. A simulator may also be used in a **learning** paradigm. In this case, the value of the simulator lies in how many fewer trials we need to perform on the real robot to achieve equivalent performance.

In [2], the authors propose to estimate a transferability metric based on a simulation-to-reality disparity measure which is defined over robot behaviours. In this case evaluating the metric is likely as arduous as deploying directly onto the real robot in the first place. Another approach is the "Sim2Real Correlation Coefficient" [8], which is defined as the Pearson correlation cofficient over "accuracies" in simulation and reality for n methods. Analyzing the intermediate signals in Fig. 1 is a commonly-used approach to evaluate the discrepancy in one of the components (e.g. to evaluate the dynamics error for manipulation [1]). However, this is myopic and does not consider the ways that different types of errors in the different components interact, nor does it necessarily correspond to a good metric of how useful the simulator is as a tool for learning.

In the remainder of this abstract, we will attempt to formalize the concept of the reality gap and clearly separate the different potential sources of value of a robotics simulator.

II. Preliminaries

We will consider a **simulation** to encapsulate the dynamics simulation, environment modeling, and rendering aspects shown in Fig. 1. As such, a simulation, $S \in \mathcal{S}$ can be considered something that maps control commands $u \in \mathcal{U}$ to observations $z \in \mathcal{Z}$, conditioned on some internal environment model state $x \in \mathcal{X}_{env}$ ($S : \mathcal{U} \times \mathcal{X}_{env} \mapsto \mathcal{Z}$).

A **task**, $T \in \mathcal{T}$, is specified through one or more **evaluation metrics**, M, which map a trajectory of N states (either in the real world or in the simulation) to a real-valued number:

$$T \triangleq \{M_i\}_{i=1}^m, \quad M_i : \mathcal{X}^N \mapsto \mathbb{R} \tag{1}$$

An **agent** contains the algorithm that is used to generate

[1] Montréal Robotics and Embodied AI Lab (REAL) at Université de Montéal, Mila. Corresponding author paulll@iro.umontreal.ca.

control commands from observations and some internal representation of state $x \in \mathcal{X}_{agent}$ ($A : \mathcal{Z} \times \mathcal{X}_{agent} \mapsto \mathcal{U}$). Presumably, the algorithm informing this agent is designed to optimize the specified task evaluation metrics. A **learning agent** is able to adapt its behaviour over time by means of a learning algorithm (A at time k is not necessarily the same as A at time $k+1$). However, we assume here that, for a stationary environment and task, the learning agent will *converge* to a stationary agent for some k large enough.

III. The Simulator as a Predictor

The first "value" of a simulator is as a tool to predict. However, different from Def. 1, we argue that the simulator's ability to predict *task performance* rather than exact observations is what is relevant:

*Def. 2 (**Predictive Reality Gap**):* Given a T defined by evaluation metrics $M_{1:m}$, and an A that generates trajectory $x^{1:N}_{sim}$ in the simulator and $x^{1:N}_{real}$ on the real robot given equivalent starting conditions, then we define the predictive reality gap (PRG) by the discrepancy of the resulting evaluation metrics:

$$\text{PRG} \triangleq \sum_{i=1}^{m} \beta_i |M_i(X^{1:N}_{sim}) - M_i(X^{1:N}_{real})| \tag{2}$$

where the β_i terms are weighting constants that can account for mismatched units or possibly an increased importance of one metric over another.

By Def. 2, a simulator can be considered **perfectly faithful** for a given task if PRG is zero for all possible agents. This definition is roughly equivalent to the definition in [8], except with a particular focus on the task performance as the choice of metric and the 1-norm distance instead of the bivariate correlation.

The need for the β constants in Def. 2 is undesirable since it allows some room for subjectivity. But this can be avoided by considering that in many cases we are interested in *comparing* agents rather than finding exact evaluations of the metrics. As a result we can consider a relaxation of Def. 2 to the relative case. Given that a task may contain several evaluation metrics, agents can be arranged in a partial ordering whose binary relation $\leq$ is defined by dominance along all of the available metrics:

$$A_1 \geq A_2 \rightarrow M_i(X_1) \geq M_i(X_2) \quad \forall i \tag{3}$$

where X_j is shorthand for the trajectory produced by agent A_j (either in the simulator or in the real environment).

*Def. 3 (**Relative PRG**):* Given K agents, the relative predictive ability of a simulator is defined by its ability to accurate predict the binary relations between agents. Let $\mathcal{A}^{sim} = [\alpha_{ij}]_{i,j=1..K}$ be a matrix whose entries are given by:

$$\alpha^{sim}_{ij} = \begin{cases} 1 & A^{sim}_i \geq A^{sim}_j \\ 0.5 & A^{sim}_i \ngeq A^{sim}_j \quad \& \quad A^{sim}_j \ngeq A^{sim}_i \\ 0 & A^{sim}_i \leq A^{sim}_j \end{cases} \tag{4}$$

where A^{sim}_j (A^{sim}_i) is agent j (i) applied to the simulation. We similarly construct $\mathcal{A}^{real}$. Then the relative predictive reality gap (RPRG) is given by the 1-norm between these two matrices that represent the relations in the two partial orders [9]:

$$RPRG(A_{1:K}) = \sum_{i,j=1}^{K} |\alpha^{sim}_{ij} - \alpha^{real}_{ij}| \tag{5}$$

According to Def. 3, a simulator is now perfectly faithful if it produces the identical partial order over agents that would be produced if the agents were run on the real robot. This is closely related to the concept of "rank inversion" [10].

Note that in the case of both PRG and RPRG, the reality gap is *conditioned* on the task and *agnostic* to the agent (only requires some method of generating trajectories). Also note that in practice the performance of the agent in the simulator or (especially) in the real environment will be stochastic and therefore PRG and RPRG should be redifined as metrics over distributions and approximated by sequences of trials.

IV. The Simulator as a Teacher

Independently from the simulator's predictive ability, it may have value as a tool for agents that *learn*. The simulator now becomes a part of the agent generation process since, as shown by the dashed lines in Fig. 1, the task performance may be fed back to the agent. We can assess the value of the simulator by considering the quality of the agents that it is able to produce, compared to agents that learn entirely on the real robot. A simulator is deemed more valuable if it reduces the number of robot trials that are needed. To evaluate a sim2real method, either the number of trials on the real robot is held fixed and the agent performance on a given task with and without the simulator is compared, or the number of trials is compared to achieve equivalent performance with and without simulator pre-training. However, we argue here that choosing an arbitrary performance or an arbitrary number of real robot trials can bias results.

*Def. 4 (**Learning Reality Gap**):* Given that a learning agent, A^{real} trained entirely on hardware is able to achieve a performance of $M_{i..m}$ on task T *at convergence*, then the Learning Reality Gap (LRG) is the number of trials on the real robot needed for an agent trained in simulation and transferred to the real robot, $A^{sim2real}$ to achieve equivalent or better performance ($A^{sim2real} \geq A^{real}$) on the real robot.

In the case of some learning agents or tasks, running trials on the robot is impossible. For example, the case where a reinforcement learning agent is using privileged information in the simulation (such as precise internal state) that is not available on the real robot. In this case, evaluation of the LRG is impossible and we must settle for a metric based on predictive performance such as PRG or RPRG. But this is problematic since it entangles the simulator's learning and predictive values.

Note here that, in contrast to the predictive value, the value of the simulator as a teacher is conditioned on both the task and the learning algorithm itself.

We also note that we are make some simplifying assumptions: 1) we are not doing any form explicit domain transfer, for example domain adversarial transfer [3, 11], since this typically requires some other source of information in the form of a dataset that is not generated by the simulator, 2) we are able to reliably detect when the learning algorithm has reached convergence, both in the simulated case and on the real robot.

References

[1] J. Collins, D. Howard, and J. Leitner, “Quantifying the reality gap in robotic manipulation tasks,” in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6706–6712.

[2] S. Koos, J. Mouret, and S. Doncieux, “The transferability approach: Crossing the reality gap in evolutionary robotics,” *IEEE Transactions on Evolutionary Computation*, vol. 17, no. 1, pp. 122–145, 2013.

[3] H. Bharadhwaj, Z. Wang, Y. Bengio, and L. Paull, “A data-efficient framework for training and sim-to-real transfer of navigation policies,” in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 782–788.

[4] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world,” in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 23–30.

[5] B. Mehta, M. Diaz, F. Golemo, C. J. Pal, and L. Paull, “Active domain randomization,” in *Proceedings of the Conference on Robot Learning*, 2020, pp. 1162–1176.

[6] M. Toussaint, K. Allen, K. Smith, and J. Tenenbaum, “Differentiable physics and stable modes for tool-use and manipulation planning,” in *Proceedings of Robotics: Science and Systems*, Pittsburgh, Pennsylvania, June 2018.

[7] F. de Avila Belbute-Peres, K. Smith, K. Allen, J. Tenenbaum, and J. Z. Kolter, “End-to-end differentiable physics for learning and control,” in *Advances in Neural Information Processing Systems 31*. Curran Associates, Inc., 2018, pp. 7178–7189.

[8] A. Kadian, J. Truong, A. Gokaslan, A. Clegg, E. Wijmans, S. Lee, M. Savva, S. Chernova, and D. Batra, “Are we making real progress in simulated environments? measuring the sim2real gap in embodied visual navigation,” 2019.

[9] W. D. Cook, M. Kress, and L. M. Seiford, “An axiomatic approach to distance on partial orderings,” *RAIRO - Operations Research - Recherche Opérationnelle*, vol. 20, no. 2, pp. 115–122, 1986.

[10] A. Ligot and M. Birattari, *On Mimicking the Effects of the Reality Gap with Simulation-Only Experiments: 11th International Conference, ANTS 2018, Rome, Italy, October 29–31, 2018, Proceedings*, 01 2018, pp. 109–122.

[11] F. Zhang, J. Leitner, Z. Ge, M. Milford, and P. Corke, “Adversarial discriminative sim-to-real transfer of visuo-motor policies,” *The International Journal of Robotics Research*, vol. 38, no. 10-11, pp. 1229–1245, 2019.