Predicting the Robot Learning Curve based on Properties of Human Interaction

Sekou Remy and Ayanna M. Howard
Human-Automation System Lab
Georgia Institute of Technology
225 North Avenue,
Atlanta, Georgia 30332

Abstract

In this work we present research on three topics which have implications for future robotic applications. Couched in learning from human provided examples, we study how robots can demonstrate learning curves akin to those observed for human students. Specifically we show how the parameters of robot learning curves relate to those parameters from learning curves generated by human students. Next we show how these parameters and learning process they represent are affected by the quality of instruction provided. Finally, we present a method to generate an estimate of the robot learning curve. This method is of merit since it is based on properties of interaction that can be extracted as learning occurs.

Introduction

For many researchers in the robotics community it is becoming clearer that learning, or the process of skill acquisition and generalization, is an integral feature for future robotic systems. Robotics researchers are not alone in this revelation. Those studying factory automation and autonomous systems design (Schaal, Atkeson, and Vijayakumar 2000) have also realized that the complexity of implementing and programming adequate control strategies poses a challenge that learning can help surmount. Not only an issue of initial implementation and design, there are some (e.g. (Schaal 1999)) who assert that learning is the only viable research approach to create flexible autonomous robots that can perform multiple tasks.

Before continuing further on the topic of robotic learning, attention must first be paid to some relevant issues of the learning process for humans. Once these matters have been addressed we will then proceed to provide motivation and hypotheses for our work and then the body of the work itself.

Background

Several methods have been applied to evaluate the process of human learning. In spite of this, as early as the turn of the century (Mayer-Kress, Newell, and Liu 1998) it was noted that there are strong regularities in the results obtained from these methods. One area where these regularities are evident is the *learning curve*. The learning curve is a representation of the learning assessment (performance) as the learning process occurs.

As noted by (Mayer-Kress, Newell, and Liu 1998; Ritter and Schooler 2002) and several others, learning curves for human subjects are generally described by two families of functions: exponential functions and power-law functions. The exponential family of functions is defined by (1). This equation shows that the performance (P) is dependent on N, the number of practice instances or trials. To accommodate for cases where the learning process is applied and prior learning exists, N_0 is the number of initial instances or trials. $N + N_0$ thus represents the total number of instances or trials. P is also defined by constants A, B, and β. These constants respectively capture the steady state performance level, the range of learning, and the learning rate of the student.

$$P(N) = A + Be^{\beta(N+N_0)} \tag{1}$$

The power-law family of functions is defined in (2) and all the parameters and variables have analogous meanings as those presented in (1).

$$P(N) = A + B(N + N_0)^{\beta} \tag{2}$$

In general learning curves capture the performance of a task over the course of the learning process. It is true that for many human subjects, past experience will influence the learning process, but these parametric forms of the learning curve were developed based on the number of times direct training occurred. The value N_0 refers to prior training, but since it is difficult to effectively capture all types of prior training for human students, these curves are derived from known instances of training. For this reason, common practice sets $N_0 = 0$, then applies a regression method to derive the values B and β. Although this is the most accurate form of (1) and (2), it is commonly seen in the literature that the equations are further simplified by setting $A = 0$.

In terms of identifying which of these two families of functions should be appropriate, (Howard and Paul 2005) indicates that (1) should be applied when fitting performance of a single user exploring a single strategy. This will not be the case in the situation when teachers are interactively teaching students. The relevance of such a scenario will become more obvious. Briefly, in this scenario, teachers typi-

cally adjust their instruction based on observed or apparent learning progress thus creating a system in which the learning attained is directly related to the proportion of what remains to be learned. Such functionality lends itself to the theory of learning curves which follow the power law (Ritter and Schooler 2002).

Usefulness of Learning Curves

When available, learning curves can help to provide a wealth of information about the learning process. These curves can be used to estimate how long or how many instances it will take to display a desired performance level. They can also be used to estimate the performance level after a given amount of time (instances) passes. Such information can be combined to identify at what point the learning process should be terminated since the increases in performance level no longer attain significant changes. The merits of having this information have been presented in (Remy and Howard 2008b). Of specific note is that each learning curve also provides the learning rate, β, which in itself gives great indication of the quality of the learning process.

Challenges with Learning Curves

It must be noted that while heralded as one of the successes of cognitive modeling (Ritter and Schooler 2002), there are some areas of concern when considering applications of the power-law of learning. In work presented in (Roessingh and Hilburn 1999), researchers have identified that the log-log linear model (see Equation 3) will not necessarily fit the power-law model of (2) depending on the nature of the error in the collected data.

$$log(P(N)) = log(B) + \beta log(N + N_0) \quad (3)$$

As such, estimates of B and β can be improperly inferred from the log-log linear transformation of the power-law so while convenient, applying this transformation can introduce errors in the derived learning curves. The researchers suggested that non linear regression models must be used and the presented work follows this suggestion.

Another area of concern identified in (Roessingh and Hilburn 1999) is related to the values that β can attain. For $\beta > 0.5$ these researchers indicate that there is a super linear characteristic when the learning curve is plotted as a function of time rather than number of trials. It was our observation that the shorter trial times were indeed recorded as learning progressed. Further work is needed to investigate their claim that such action is "opposite to the basic assumptions of learning theory". We believe that the performance value will approach A asymptotically, so improvements in performance (and time) are expected with continued practice.

Motivation

With this understanding of learning and learning curves in human study we now consider the situation where learning is applied to robotic students. Of specific interest is learning that occurs through human-robot interaction. In such a scenario, whether through observation or through interactive learning, the robot is learning the task directly from one or more human teachers. As such, we hypothesize that a robot learning curve can be extracted that follows similar characteristics to a human learning curve. We also hypothesize that this learning curve is related to the quality of the instruction provided. Finally, we believe that during the process of learning from the human interactively that the robot can extract enough information for it to generate a proxy of a learning curve.

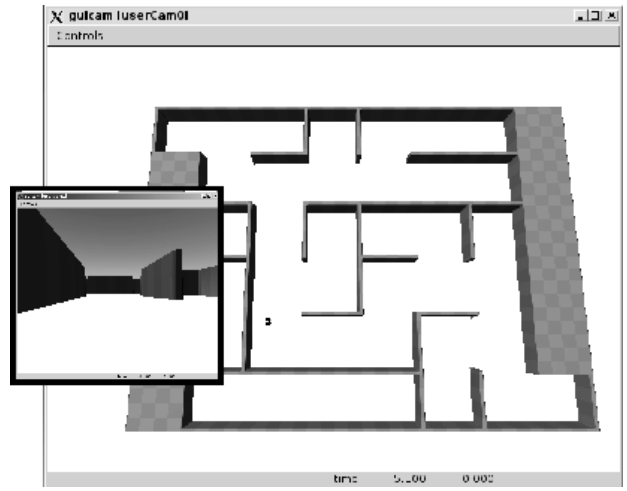

Figure 1: Overhead view of simulation environment with robot view inset.

Implementation and Analysis

The remainder of this paper will be separated into two sections. In the first section is the presentation of the learning curves generated for interaction between an expert user and a learning mobile robot. Interactive learning (Remy and Howard 2008a) and learning from teleoperation were performed and the curves extracted for both types of learning are presented. The second section presents a proxy of the learning curve developed by the robotic agent as the interactive learning process unfolded.

Capturing the Learning Curve

Learning from teleoperation is a process through which a user demonstrates a task and their action is used to provide examples of the task. In a manner similar to Dogged Learning (Grollman and Jenkins 2007) and Interactive Q-Learning (Thomaz and Breazeal 2006), Interactive Learning from teleoperation (abbreviated to Interactive Learning) varies the process of learning from teleoperation by providing means for the robot to learn incrementally and demonstrate learning in a real-time manner. The human operator providing control then provides instruction when needed to correct the robot's function. In essence, Interactive Learning enables just-in-time instruction while learning from teleoperation requires the entire behavior to be demonstrated prior to the beginning of the learning process.

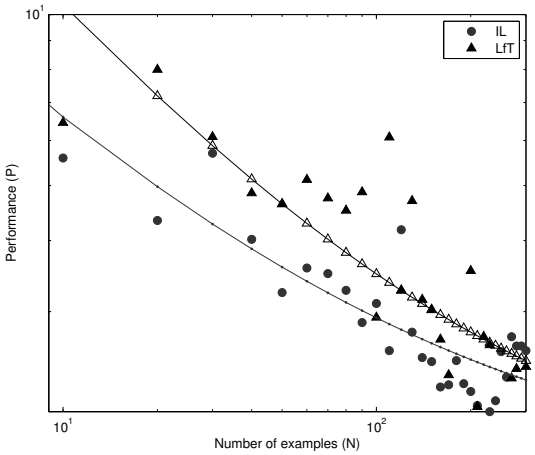

Figure 2: Learning curves for interactive learning and learning from teleoperation.

Teleoperation and Evaluation For this work, teleoperation was accomplished by providing the operator with an overhead view of the robot. The simulated robot was placed in an environment similar to that shown in Fig. 1. Based on visual cues, the operator demonstrated the target behavior and their actions were captured through the use of a joystick and passed on to the robot as they became available (near real-time). Data from the sonar sensors were used to represent the sensory state of the world. An off-board robot controller relayed all information between robot and operator, so the provided instruction is coupled with the appropriate visual cue.

Such a scenario should be increasingly common and is likely to feature in remote care giving. Our intent for such scenarios is to endow the robot with some learning so that some of the repetitive tasks can be automated. Some of these tasks may involve recording the location of landmarks for navigation purposes, traveling between prespecified locations and picking up or dropping off items. For this work we focused on an exploration task, one needed for landmark discovery.

The results presented in this paper will be for a robot that is learning a wall following task in a maze environment. To measure the performance of the robot, a performance metric was devised. The metric, which was evaluated as a post-test measure, incorporated distance (to evaluate task quality) and time of completion to generate the performance value. When teleoperating the robot (no learning), on average, expert human operators produced a performance value of 1.8 with the same metric. As the performance improves, this metric decreases in value.

Curves Generated Using Teleoperation in Different Ways The graph in Fig. 2 presents curves that follow the characteristic shape of the "power-law of performance". As listed in Table 1, the magnitude of the learning rate observed for learning from teleoperation is larger than that of Interactive Learning but it is important to note that Interactive Learning out performs it for much of the learning process (see Fig. 2). This surprising result occurs because more information is provided about the behavior earlier in the learning process. This means that with Interactive Learning, for the same number of examples, better quality instruction is presented earlier in the learning process. The larger magnitude of learning rate for learning from teleoperation and its larger learning range indicate that its performance will eventually outperform Interactive Learning, as would be expected since the entire behavior is being presented. This expectation is confirmed by the smaller value of performance with "infinte instruction" which is captured in value A in Table 1.

Table 1: Learning curves for learning from teleoperation, with and without interaction.

	IL	LfT
β	-0.55	-0.64
B	18.20	39.06
A	1.490	1.440

This section confirms the hypothesis that the robot's learning curve fits the characteristics of human learning curves. While the asymptotic values are both smaller than average value for the human operator (1.8), these curves cross this value with more than 1500 examples and require an equivalent order of magnitude examples to improve performance by ten percent.

Coherence It can be expected that cost will be one of the limiting design factors for service robots used for personal care in the home. As such, it is likely that the sensors and the actuators used will be noisy. It is also likely that the remote caregivers, for a variety of reasons, will introduce additional uncertainty in their actions as well as demonstrate variation in how sensor data is interpreted. To study the effect of these contributions on learning the following three cases are considered. In the first, each channel of the already noisy sensor data presented for learning was augmented with an additional uniform noise profile. The additional noise was uniformly distributed over the range of valid sensor values (Case 1). In the second, each channel of the actuator data was augmented with an additional Gaussian noise profile. The intent of adding this noise was to simulate slippage as well as operator inconsistencies. This noise was normally distributed with mean a_i and variance $\gamma/2$. where a_i was the operator's desired value for actuator i and γ is the upper limit of valid actuator values (Case 2). Finally in the third, changes from both Case 1 and Case 2 were applied (Case 3). Each of these changes to the sensor and actuator data utilized for learning compromised the examples provided by the human operator and the curves shown in Fig. 3 highlight these effects.

The learning rate for Case 3 is smaller than either Case 1 or Case 2 but more importantly, the curve generated for this case approaches a constant which is larger than any other curve. This indicates that corrupting both sensor and actua-

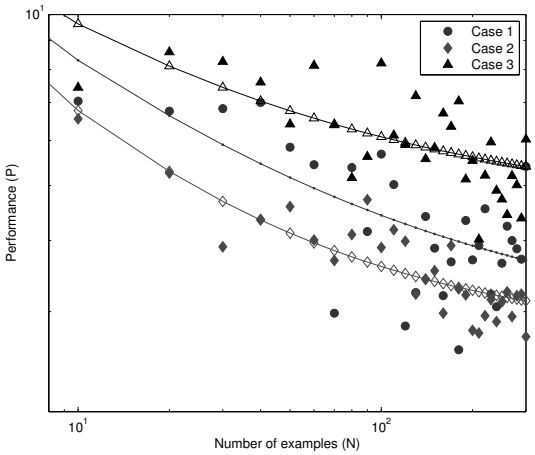

Figure 3: Learning curves in three cases where coherence affects the learning process.

Table 2: Learning curve parameters for Cases 1-3.

	Case 1	Case 2	Case 3
β	-0.51	-0.65	-0.50
B	18.20	18.28	16.34
A	2.715	2.695	4.456

tor data sources had the most negative impact on the learning process. This should not have been surprising since the behaviors learned were based on compromised instruction. It is also valuable to note that in all three cases these curves approached a steady state value larger than the performance of the human operator as well as larger than the performance predicted by each curve in Fig. 2.

These results confirm the second hypothesis of this work since the properties of the learning curves generated vary based on the degree of coherence in instruction, which is a measure of the instruction quality.

Proxy of the Learning Curve

It is only possible to generate learning curves after learning has been completed. It is also only possible to generate learning curves if a closed form evaluation metric exists for the task being learned. In the cases where such a metric does not exist or in the case where an assessment of learning is needed or useful *as learning is occurring*, developing a proxy of the learning curve in situ would be of interest. Such a proxy could enable an agent to autonomously regulate facets of the learning process or even to provide feedback with the intent of soliciting higher quality instruction from the teacher.

Overview In this work we present a proxy that is produced as a by product of the learning method applied. Interactive Learning, a method briefly described earlier and presented with more detail in (Remy and Howard 2008a), is a process that leverages the expertise of the human teacher while minimizing the knowledge required apriori. The essence of the method is that the teacher provides the needed instruction only when they believe it is necessary. This quality of the Interactive Learning process provides a mechanism to estimate the teacher's level of satisfaction with the student's performance. A large degree of the instruction implies that the robot is performing poorly while the opposite can imply that the robot is performing well.

The application of interactive learning can also provide useful information since the trends can be extracted from the classifiers. If the robot is learning the task successfully, the classification errors produced should decrease or at worst not increase after the major components of the task are learned. Large error changes should only occur during the act of incorporating chunks of new knowledge, and this should be a transient process after which formerly new knowledge transitions to old knowledge.

To summarize, in the process of using Interactive Learning to learn the target behavior(s), it is possible to tap into two fundamental qualities: interaction rate and error levels. These qualities can be used to extract a proxy of performance autonomously by the robot as instruction is being provided.

Approach Implemented The quantization error plot shown in Fig. 4 includes both error and interaction information. The point density of the graph when projected onto the x-axis indicates the interaction level, while the slope of the graph indicates how the error changes as additional examples are presented. This graph contains two curves. The first (in circles), is the plot of the error for the case where learning was successful over the interaction time and this can be observed by the decrease in interaction level as well as the plateau in the curve.

The second curve in Fig. 4 (in dots) is the error plot for Case 1 which was defined in the previous section. In this case learning was impeded and this can be observed by the relatively consistent point density as well as by the gradual increase in error over time.

The magnitude of the error in Case 1 is also of relevance, but this property is neglected since it would be difficult to assess a baseline value for error without a great deal of information about the task, the environment, the robot and the instruction that will be provided by the operator. None the less, the error plot for Case 1 shows how information about the success of the learning process can be extracted.

When considering these three cases, the following curves are generated (see Fig. 5). The interpretation permitted by these proxies is consistent with those generated by formal treatment of performance. Case 3 has a smaller learning rate than either Case 1 or Case 2 and all of the cases learn slower than the case where learning was most successful.

This shows that it was possible to generate a proxy of the learning curve by extracting data from the interactive learning process. It should be noted that this proxy, while a crude first attempt that will be refined in the future, is not a learning/performance factor which indicates how performance changes with examples. This proxy captures how the

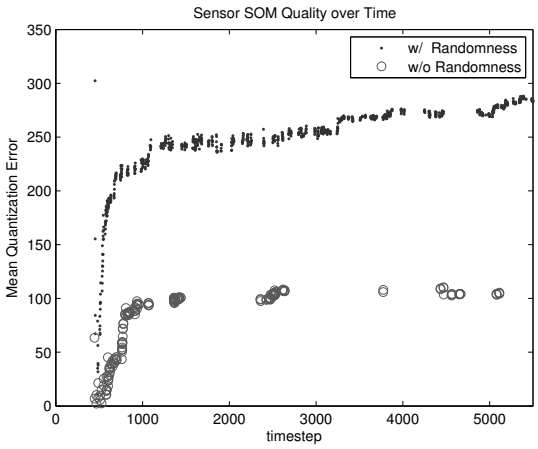

Figure 4: The mean quantization error graph where learning was successful.

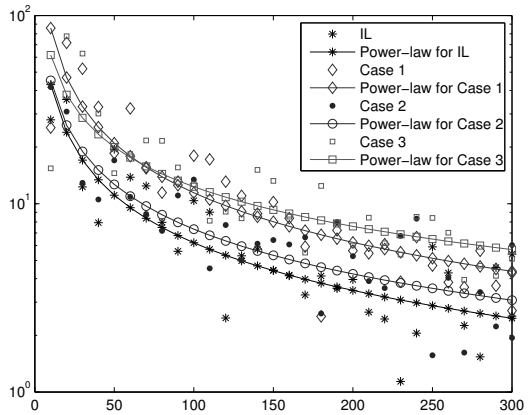

Figure 5: Estimations of the learning curves for interactive learning under four conditions vs. time.

robot is learning over **time**.

Developing Estimates as Data is Provided The values of β are presented as they change with additional examples. It has been mentioned that the estimate of the learning curve can be used to assist the student (robot) to augment the learning process. For such a notion to become feasible, it is assumed that the estimate of performance at any time is credible. Fig. 6 shows the plot of $\beta(N)$. This plot shows that for the task under consideration the value of β changes very little for $N > 200$. The implication of this observation is that after 200 examples, a stable estimate of the performance can be extracted.

Summary and Future Work

In this work we have discussed three topics which have implications for future robotic applications. The first identifies that when robots learn from humans, they generate learning curves that follow similar characteristics as their human instructors (in this case the power-law of learning). A closely related second point is that the parameters of these curves are dependent on the quality of the instruction provided. The third combines the results of the first two and uses them to produce a proxy of the learning curve. This proxy while crude is unobtrusively generated by the robot, as learning is occurring.

Having an estimate of the performance (and learning) would not only provide the robot with the ability to regulate the learning process without increasing the training prerequisites for the human instructors, but it can also be used to enable informative feedback to those instructors. Since feedback and dialog are key components of human communication such an estimate is invaluable. Incorporation of this feedback is thus a logical next step for this research.

Finally, while the goals were attained and the crude proxy generated contains many of the expected properties, further work is required to determine the mapping between the estimated learning curve (the proxy) and the formally derived curve. Such a mapping would enable the proxy to directly provide more information about the actual learning curve and increase the quality of information that autonomous agents can incorporate into their operation.

Table 3: Parameters of estimations of learning curves for Cases 1-3.

	IL	Case 1	Case 2	Case 3
β	-0.8402	-0.8751	-0.7900	-.6993
B	297.24	643.85	278.6	308.4

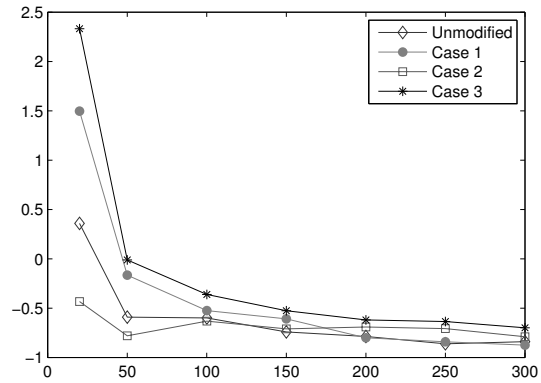

Figure 6: Plot of learning rate (β) vs. trial number N.

References

Grollman, D. H., and Jenkins, O. C. 2007. Dogged learning for robots. In *IEEE International Conference on Robotics and Automation*, 2483–2488.

Howard, A. M., and Paul, W. 2005. A 3d virtual environment for exploratory learning in mobile robot control. In *Proc. IEEE Conference on Systems, Man, and Cybernetics*, 306–310.

Mayer-Kress, G. J.; Newell, K. M.; and Liu, Y.-T. 1998. What can we learn from learning curves? In *International Conference on Complex Systems*.

Remy, S., and Howard, A. M. 2008a. Learning approaches applied to human-robot interaction for space missions. *Intelligent Automation and Soft Computing*.

Remy, S., and Howard, A. M. 2008b. The use of coherence in learning behaviors via teleoperation. In *Proceedings of the International Symposium on Robot and Human Interactive Communication (RO-MAN 08)*.

Ritter, F. E., and Schooler, L. J. 2002. The learning curve. In *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam: Pergamon, Elsevier Science.

Roessingh, J. J. M., and Hilburn, B. G. 1999. The power law of practice in adaptive training applications. In *Proceedings of the Annual Meeting of the Human Factors Society - Europe Chapter,*. NLR-TP-2000-308.

Schaal, S.; Atkeson, C. G.; and Vijayakumar, S. 2000. Real-time robot learning with locally weighted statistical learning. In *ICRA*, 288–293.

Schaal, S. 1999. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences* 3:233–242(10).

Thomaz, A. L., and Breazeal, C. 2006. Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance. In *AAAI*.