Neural Network Modeling of Learning to Actively Learn

Lie Yu[1], **Ardavan S. Nobandegani**[2,3], **& Thomas R. Shultz**[1,3]
{lie.yu, ardavan.salehinobandegani}@mail.mcgill.ca
thomas.shultz@mcgill.ca

[1]School of Computer Science, McGill University
[2]Department of Electrical & Computer Engineering, McGill University
[3]Department of Psychology, McGill University

Abstract

Humans are not mere observers, passively receiving the information provided by their environment; they deliberately engage with their environment, actively participating in the information acquisition stage to improve their learning performance. Despite being a hallmark of human cognition, the computational underpinnings of this active (or self-directed) mode of learning have remained largely unexplored. Drawing on recent advances in machine learning, we present a neural-network model simulating the process of learning how to actively learn. To our knowledge, our work is the first neural-network model of learning to actively learn. Extensive simulations demonstrate the efficacy of our model, particularly in handling high dimensional domains. Notably, our work serves as the first computational account of the recent experimental finding by MacDonald and Frank (2016) showing that prior passive learning improves subsequent active learning. Our work exemplifies how a synergistic interaction between machine learning and cognitive science helps develop effective, human-like artificial intelligence.

Keywords: Active learning; deep neural networks; deep reinforcement learning; example generation

1 Introduction

Humans are not mere passive observers of their environment, but actively search for information which helps to improve their learning performance (Gureckis & Markant, 2012). For example, we purposefully search for information online to learn about a topic of interest, decide how to interact with an unfamiliar device to learn its functionality, or ask questions from people around us to learn more about them, helping us to interact with them more effectively in the future. Relatedly, past educational research shows that people learn better if the flow of experience is under their control (e.g., Cherney, 2008; Michael, 2006).

Although active (aka self-directed) information acquisition is a fundamental and extensively studied topic in the educational sciences (e.g., Bruner, Jolly, & Sylva, 1976; National Research Council, 1999), it has been comparably understudied in the psychological literature (Gureckis & Markant, 2012; Markant & Gureckis, 2014), with the psychological processes underpinning this mode of learning remaining largely unexplored (Gureckis & Markant, 2012). Experimental studies of human learning are predominantly passive in that the experimenter tightly controls what information is presented to the learner on every trial.

A growing, but highly theoretical, research area in computer science, called active learning, aims to formally characterize the extent to which self-directed information acquisition can speed up learning (see Hanneke, 2014, for a survey). Despite notable theoretical successes (e.g., Hanneke, 2016), this research area has made little contact with the psychological literature, primarily focused on highly abstract learning problems amenable to theoretical investigations, and predominantly investigated mathematically the performance gain obtained by following specific active learning strategies, paying no attention to the key problem of how learners learn their active learning strategies in the first place.

Drawing on recent advances in machine learning (particularly deep reinforcement learning), we present a novel neural-network model of active learning aiming to simulate the process of learning how to actively learn. By conceptualizing the problem as a reinforcement learning task, our neural-network model learns, during the passive phase of learning (wherein the learner passively receives information from their environment) an effective active learning strategy allowing for faster learning. As an instantiation of our active learning model, in this work we focus on the task of category learning (aka classification).

Our model has several notable features elevating its cognitive plausibility. First, our model uses Markov-adjusted Langevin (MAL) (Savin & Deneve, 2014; Moreno-Bote, Knill, & Pouget, 2011; Nobandegani & Shultz, 2017, 2018), a well-known gradient-based Markov chain Monte Carlo (MCMC) method, allowing active search for maximally informative examples in a computationally-efficient manner. Notably, recent work in theoretical neuroscience has shown that MAL can be implemented in a neurally-plausible manner (Savin & Deneve, 2014; Moreno-Bote et al., 2011). MCMC methods are a family of algorithms for sampling from a desired probability distribution, and have been successful in simulating important aspects of a wide range of cognitive phenomena, e.g., temporal dynamics of multistable perception (Gershman, Vul, & Tenenbaum, 2012; Moreno-Bote et al., 2011), developmental changes in cognition (Bonawitz, Denison, Griffiths, & Gopnik, 2014), category learning (Sanborn, Griffiths, & Navarro, 2010), causal reasoning in children (Bonawitz, Denison, Gopnik, & Griffiths, 2014), and cognitive biases (Dasgupta, Schulz, & Gershman, 2016).

Second, to improve its active learning strategy, our model uses *memory replay*: the idea of accessing memories of multiple past events and integrating them to make useful predictions about an action's consequences (e.g., Káli & Dayan, 2004; Lengyel & Dayan, 2008; Momennejad, Otto, Daw, & Norman, 2018). Mounting evidence shows that memory

replay supports reinforcement learning and planning (e.g., Ólafsdóttir, Bush, & Barry, 2017; Momennejad et al., 2018).

Finally, our model effectively adapts its learned active-learning strategy as it gradually acquires more knowledge about a learning task. This feature of our model is supported by mounting evidence suggesting that people adapt their strategies according to their knowledge and environmental conditions (e.g., Rieskamp & Otto, 2006; Hoffart, Rieskamp, & Dutilh, 2018; Payne, Bettman, & Johnson, 1988; Bröder, 2003; Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011; Lieder & Griffiths, 2017).

Our paper is organized as follows. We begin by introducing our neural-network model, and proceed to show the efficacy of our model with extensive simulations. We conclude by discussing the implications of work for active learning research and point out several fruitful lines of future work.

2 Neural Network Model

Our model consists of three neural network modules:

- **Encoder Network (E-Net)**: This neural network module takes a raw input x_i and ouputs a corresponding state representation s_i. As such, this module simulates perception systems, mapping a stimulus to its representation in psychological space.

- **Classification Network (C-Net)**: This neural network module takes state representation s_i and outputs a class label y_i. As such, this module simulates information processing cortices in the brain supporting concept categorization.

- **Action-Value Network (Q-Net)**: For each representation state s_i (corresponding to raw input x_i), this neural network module, parameterized by a set of weights θ, outputs an *affinity score* $Q(x_i,\theta)$ modeling the learner's confidence in choosing x_i to boost learning. That is, a higher $Q(x_i,\theta)$ corresponds to a higher confidence level. Crucially, the network's output, i.e., affinity scores, encodes information enabling our MCMC method, MAL, to actively search for exemplars most helpful for improving the classification performance of the C-Net.

When searching actively for an informative example x which is likely to maximally improve learning accuracy, our model samples from a target distribution $\pi(x)$ given by:

$$\pi(x) \propto \exp(\beta Q(x,\theta)) \tag{1}$$

where θ denotes the parameters of the Q-network (i.e., the set of network weights), and $\beta \in \mathbb{R}^{>0}$ is a damping factor.

By assigning higher probabilities to those examples x the Q-network believes to maximally improve learning accuracy (i.e., the classification accuracy of the C-Net), Eq. (1) ensures that sampling from $\pi(x)$ yields effective active learning.

To jointly train the E-Net, C-Net, and Q-Net modules of our neural networks model, we use a novel variant of the well-known Deep Q-learning Algorithm (Mnih et al., 2015); see Algorithm 1. Our novel variant of the Deep Q-learning Algorithm has the added advantage of incorporating MCMC in its functionality (Algorithm 1, Line 8), ensuring that sampling from the target distribution $\pi(x)$ would likely yield informative examples x whose knowledge maximally improves the learner's classification accuracy, thus yielding effective active learning.

Algorithm 1 MCMC-Enhanced Deep Q-Learning Algorithm

1: Initialize replay memory D to capacity N
2: Initialize action-value function Q with random weights θ
3: Initialize target action-value function $\hat{Q}$ with weights $\theta^{-} = \theta$
4: Initialize classifier C and encoder E with random weights w_c and w_e, respectively
5: **for** episode = 1 to M **do**
6: Randomly pick an input x_0 and encoded state representation s_0
7: **for** t=1 to T **do**
8: With probability ε sample a random data point x_t
9: Sample a new data point x_t via MCMC with the affinity function:

$$\pi(x_t) \propto \exp(\beta Q(x_i,\theta))$$

10: Compute $q_0 = Q(s_t,a_0;\theta)$ and $q_1 = Q(s_t,a_1;\theta)$
11: If $q_0 > q_1$, discard these data and go to step $T+1$. Otherwise, feed s_t into C and update its parameters w_c.
12: Do evaluation on C and obtain reward r_t
13: Set $s_{t+1} = s_t$, store transition pair (s_t,a_t,r_t,s_{t+1}) in memory D.
14: Sample minibatch of transitions (s_j,a_j,r_j,s_{j+1}) from D
15: Set $y_j = r_j + \gamma max_{a'} \hat{Q}(s_{j+1},a',;\theta^{-})$
16: Perform a gradient descent step on $(y_j - Q(s_j,a_j;\theta))^2$ with respect to θ
17: For every N_Q steps reset $\hat{Q} = Q$

The rationale behind Algorithm 1 is as follows. Line 1 initializes the memory replay capacity of our model. Lines 2-4 randomly initialize the weights of E-Net, C-Net, and Q-Net modules. Crucially, by so doing, we assume no prior knowledge on the part of the learner at the onset of learning. Lines 5-9 (except Line 8) use MCMC to effectively guide the active search toward informative samples, the knowledge of which likely maximally improves learning performance. Line 8, for only a small fraction of times, performs random exploration of the input space during the active learning phase. Being a standard approach in machine learning, Line 8 aims to achieve an effective exploration-exploitation trade-off. Lines 10-12 compute the reward associated with each active learning episode by evaluating learning accuracy on a held-out evaluation set: A higher reward implies that the learning performance of our model has considerably improved by using the samples recommended by the Q-Net module. Line 15 updates the model parameters according to the reward obtained in Line 12. Finally, Lines 12-17 (except Line 15) implement the well-known Q-learning process widely used in modeling model-free reinforcement learning in the machine learning, psychology, and neuroscience literatures (Watkins & Dayan, 1992).

3 Simulations

In this section, we demonstrate with simulations the efficacy of our neural network model in learning how to actively learn. We tackle several learning tasks, ranging from simple (the continuous-XOR Problem) to moderate (the Two-Spirals

Problem) to quite demanding (recognizing high-dimensional images of hand-written digits).

To experimentally investigate optimal scheduling for the active learning phase (i.e., the phase in which the learner begins actively looking for informative examples to improve learning performance), we simulate three types of active learners: Early-Starter, Intermediate-Starter, and Late-Starter. As a learner, by definition, has no control over the information provided passively by the environment, and this passive flow of information can continue indefinitely, we assume that these three types of active learners are constantly engaged in passive learning; that is, they are constantly engaged in improving their learning performance using the information that is passively, yet constantly, provided by the environment. The Early-Starter begins the active learning phase right at the start, together with the passive learning phase. The Intermediate-Starter begins the active learning phase with some delay, at an intermediate stage of passive learning (i.e., when the learner has already acquired some knowledge of the learning task of interest). Finally, the Late-Starter does not begin the active learning phase until a very late stage of passive learning (i.e., when the learner has nearly mastered the learning task at hand). As such, the Early-, Intermediate-, and Late-Starters are constantly engaged in passive learning (using the information passively provided by the environment) even *during* their active learning phase—they only differ in terms of when their active learning phase begins.

Although being simultaneously engaged in both passive and active learning (as our three Early-Starter, Intermediate-Starter, and Late-Starter learners are) is a more psychologically plausible assumption—compared to having learners who either only perform pure active learning or pure passive learning—the foregoing three learners, due to benefiting from different amounts of information, do not provide a fair characterization of the potential boost in learning accuracy afforded by active vs. passive learning.

To provide a completely fair comparison between active and passive modes of learning, and, furthermore, to theoretically corroborate several experimental findings on the efficacy of active learning, in Sec. 3.3 we simulate two new learners (the Active-Passive (AP) learner and Passive-Active (PA) learner), allowing us to directly investigate how active learning fares against passive learning.

3.1 Continuous-XOR Problem

As our first learning task, in this subsection we consider the continuous-XOR classification problem (see Fig. 1(a)). For the passive learning phase, the training set consists of 1000 samples, generated uniformly at random, in the input square $[0,1]^2$, paired with their corresponding labels. The learner receives these training samples in the form of batches of size 32. We implement the C-Net module by a 3-layer perceptron neural network (Rumelhart, Hinton, & Williams, 1985).

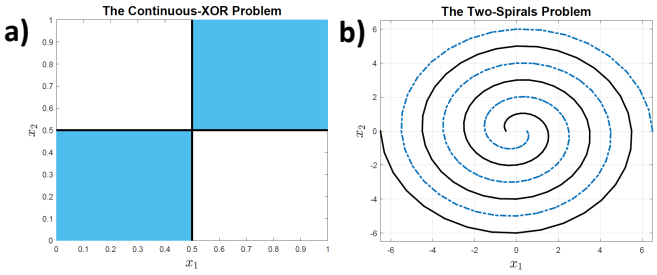

Figure 1: **(a)** The continuous-XOR learning task. the two blue quadrants correspond to the positive category and the two white quadrants correspond to the negative category, with the two solid black lines indicating the boundaries of the two categories. **(b)** The two-spirals learning task. The solid black spiral corresponds to the negative category and the dashed blue spiral corresponds to the positive category.

To quantitatively evaluate the efficacy of our model in learning to actively learn, we simulate the Early-,

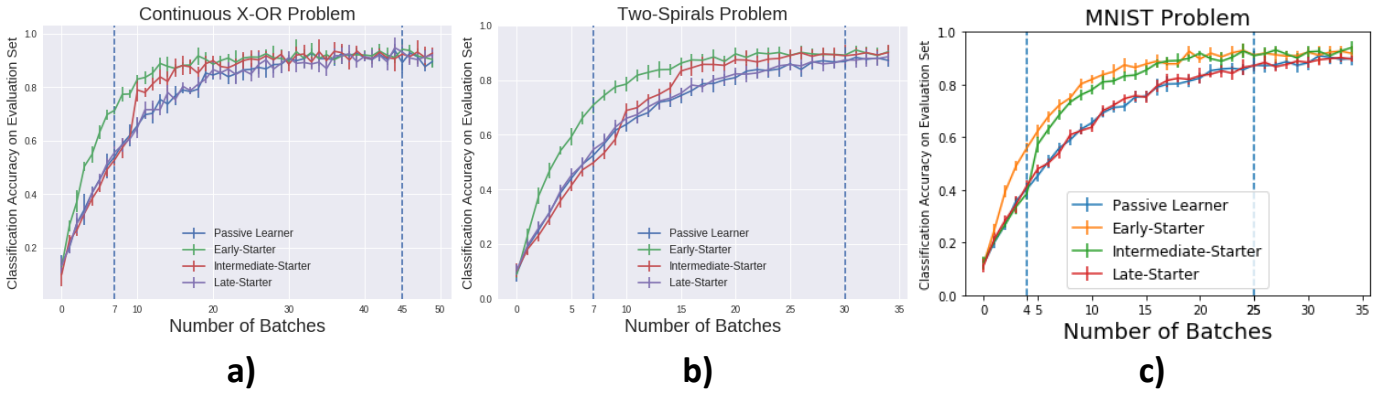

Figure 2: Classification accuracy on a held-out evaluation set by the Early-Starter, Intermediate-Starter, Late-Starter, and a purely passive learner. In each subfigure, the leftmost and the rightmost vertical dashed lines indicate the onset of the active learning phase for the Intermediate-Starter and Later-Starter, respectively. Error bars indicate $\pm$ 1 SEM. **(a)** The continuous-XOR problem. **(b)** The two-spirals problem. **(c)** The MNIST hand-written digits recognition task.

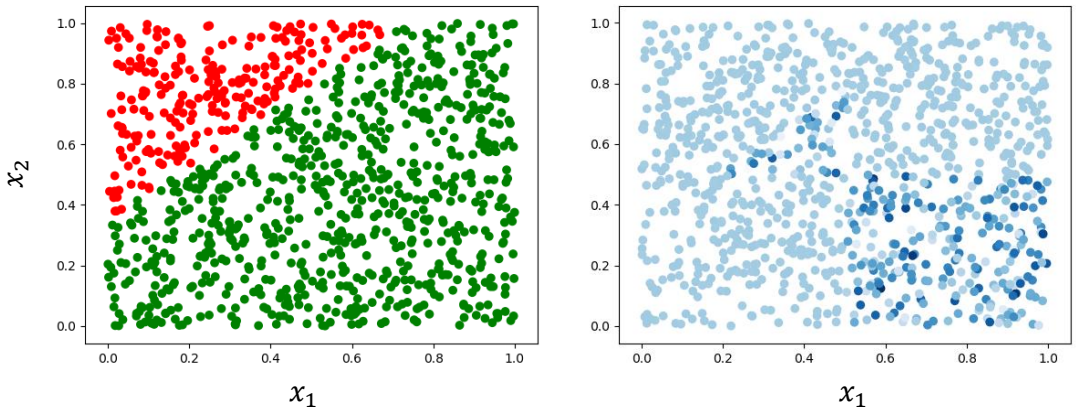

Figure 3: Left: An intermediate learning stage of the Intermediate-Starter learner in the continuous-XOR task. Red and green dots indicate examples that the learner classifies as negative and positive patterns, respectively. Right: The guidance provided by the Q-Net module at the stage of learning indicated in the Left subfigure. By assigning higher affinity scores (indicated by darker blue dots) to those regions of the input space about which the knowledge of the C-Net is lacking/incorrect, the Q-Net ensures that, by actively selecting those darker blue dots, the learning performance of the C-Net module likely improves.

Intermediate-, and Late-Starter learners, and compare their learning accuracy against a purely passive learner (as a baseline condition); see Fig. 2(a). As a measure of learning accuracy, we report percent of correct classification on a held-out evaluation set of size 100. The evaluation set comprises 100 samples, selected uniformly at random from the input square $[0, 1]^2$. Note that the training and evaluation sets do not overlap—their intersection is an empty set.

As Fig.2(a) shows, the Early-Starter predominantly obtains the highest learning accuracy; this performance is later matched by the Intermediate-Starter when it begins its active learning phase. Fig. 2(a) also suggests that any form of active learner (Early-, Intermediate, or Late-Starter) generally outperforms, in learning accuracy, a purely passive learner.

Next, we provide intuition into how the Q-Net module helps the C-Net improves its classification accuracy, by actively guiding the C-Net module toward those input regions the knowledge of which likely maximally improves the learner's classification accuracy. Fig. 3(left) depicts an intermediate learning stage of the Intermediate-Starter learner. As Fig. 3(left) shows, our classifier, i.e., the C-Net module, has already learned some knowledge about the task (that, the top-left quadrant likely corresponds to the negative patterns), but its knowledge about the decision boundaries is still lacking. Fig.3(right) shows the guidance provided by the Q-Net module at this stage of learning: By assigning higher affinity scores (indicated by darker blue dots) to those regions of the input space about which the knowledge of the C-Net is lacking/incorrect, the Q-Net ensures that, by actively selecting those darker blue dots, the learning performance of the C-Net module improves.

3.2 Two-Spirals Problem

As our second learning task, in this subsection we consider the famously difficult Two-Spirals classification problem (see Fig. 1(b)). For the passive learning phase, the training set consists of 2000 samples (1000 samples per spiral), selected uniformly at random, on the two input spirals. The learner receives these training samples in the form of batches of size 32. As was the case in the previous subsection, we implement the C-Net module by a 3-layer perceptron neural network (Rumelhart, Hinton, & Williams, 1985).

To quantitatively evaluate the efficacy of our model in learning to actively learn, we simulate the Early-, Intermediate-, and Late-Starter learners, and compare their learning accuracy against a purely passive learner (as a baseline condition); see Fig. 2(b). As a measure of learning accuracy, we report percent of correct classification on a held-out evaluation set of size 100. The evaluation set comprises 100 samples, selected uniformly at random on the two input spirals. Note that the training and evaluation sets do not overlap—their intersection is an empty set.

As Fig.2(b) shows, the Early-Starter predominantly obtains the highest learning accuracy; this performance is later matched by the Intermediate-Starter when it begins its active learning phase. Fig. 2(b) also suggests that any form of active learner (Early-, Intermediate, or Late-Starter) generally outperforms in learning accuracy a purely passive learner.

3.3 Hand-written Digits Recognition Task

As our last (and hardest) learning task, in this subsection we consider the problem of recognizing high-dimensional images of hand-written digits, using the MNIST dataset, a popular dataset in the deep learning community (Fig. 4). For the passive learning phase, the training set consists of $60,000$ examples of 28×28-pixel hand-written digits. The learner re-

Figure 4: Hand-written digit examples from the widely used MNIST dataset.

ceives these training samples in the form of batches of size 32. We implement the C-Net module by a 6-layer convolutional neural network (LeCun & Bengio, 1995).

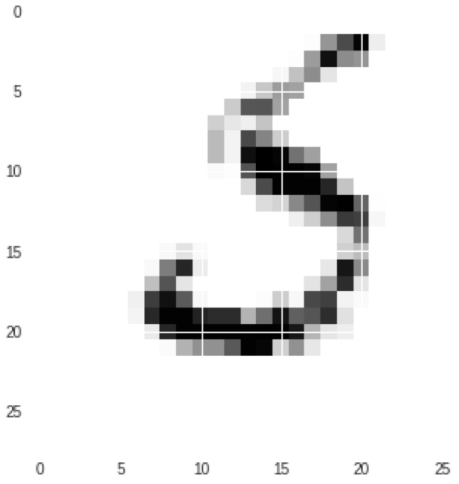

Figure 5: A 28×28-pixel digit actively selected by our model to improve learning performance. More precisely, the Q-Net believes that the classification accuracy of the C-Net can be improved by informing the C-Net that the shown 28×28-pixel image (as a whole) is a 5. Numbers on the vertical and horizontal axes indicate pixel number.

Fig. 5 shows an example produced in the active learning phase of our model; our model believes that, at this stage of learning, informing the C-Net about this example (i.e., that this 28×28-pixel image, as a whole, belongs to the class of Digit 5) significantly boosts the classification accuracy of the C-Net module. To visualize the example depicted in Fig. 5, we used a decoder neural-network module, allowing us to map the corresponding representation from the psychological space into the original 28×28-dimensional space of hand-written digits.

To quantitatively evaluate the efficacy of our model in learning to actively learn, we simulate the Early-, Intermediate-, and Late-Starter learners, and compare their learning accuracy against a purely passive learner (as a baseline condition); see Fig. 2(c). As a measure of learning accuracy, we report percent of correct classification on a held-out evaluation set of size 1000. The evaluation set comprises 1000 samples, selected uniformly at random from the original MNIST test set of size $10,000$. Note that the training and evaluation sets do not overlap.

As Fig.2(c) shows, the Early-Starter predominantly obtains the highest learning accuracy; this performance is later matched by the Intermediate-Starter when it begins its active learning phase. Fig. 2(c) also suggests that any form of active learner (Early-, Intermediate, or Late-Starter) generally outperforms a purely passive learner in learning accuracy.

Recently, MacDonald and Frank (2016) showed that passive-first learning yields better learning performance compared to active-first learning. More specifically, they showed that a passive learning phased followed by an active learning phase yields better ultimate learning performance, compared to the reversed order. As our three Early-, Intermediate-, and Late-Starter learners are constantly engaged in passive learning, even *during* their active learning phase, we cannot directly investigate the key question of which sequence of passive/active learning would ultimately yield better learning performance.

Next, we directly test the effect of passive/active learning sequence on learning. To this end, a la MacDonald and Frank (2016), we simulate two new types of learners: Passive-Active (PA) and Active-Passive (AP). PA performs passive learning during the first stage of his learning and then switches into a purely active learning phase (wherein PA only considers the samples recommended by the Q-Net module). Conversely, AP performs purely active learning during the first stage of his learning and then switches into a passive learning phase.

Fig. 6 clearly shows the superiority of PA, in learning accuracy, over AP. This finding theoretically corroborates, and serves as the first computational account of, the experimental finding by MacDonald and Frank (2016) showing that prior passive learning improves subsequent active learning.

Additionally, our finding that, during the first block of learning (Fig. 6, on the left-hand side of the vertical dashed line), AP performs worse, in learning accuracy compared to PA, is supported by the recent experimental study by Markant and Gureckis (2014) revealing that the quality of active learning is sub-optimal early in learning.

4 General Discussion

Humans are not mere passive observers of their environment, but actively search for information which helps to improve their learning performance. Despite being a hallmark of human cognition, the computational underpinnings of this active (or self-directed) mode of learning have remained largely unexplored (Gureckis & Markant, 2012).

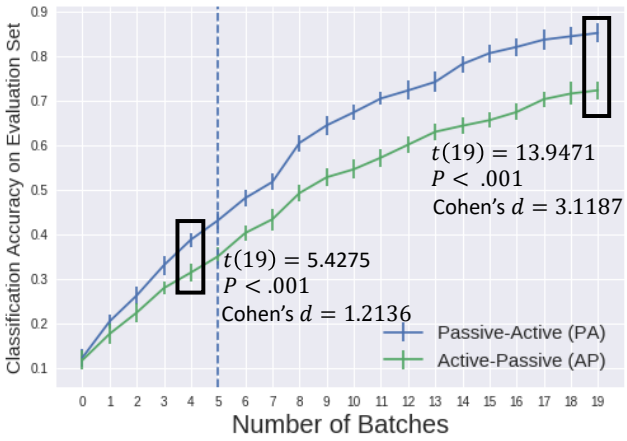

Figure 6: Investigating the effect of passive/active learning sequence on learning. Passive-Active (PA) performs passive learning first and then switches to active learning. Conversely, Active-Passive (AP) performs active learning first and then switches to passive learning. The vertical dashed line indicates the onset of the transition from one mode of learning to the other. Error bars indicate $\pm$ 1 SEM.

Building on recent advances in machine learning, particularly deep reinforcement learning, we present a novel neural-network model simulating the process of learning how to actively learn. Importantly, our neural-network model starts from scratch, having no a priori knowledge of the learning task, nor having any preset active learning heuristic(s) to choose from or to follow. To the contrary, by conceptualizing the problem as a reinforcement learning task, our neural-network model learns, during the passive phase of learning, an effective active learning strategy allowing for faster learning. Extensive simulations demonstrate the efficacy of our model, particularly in handling the high-dimensional learning task of MNIST hand-written digits.

Additionally, our model serves as the first computational account of the recent experimental finding by MacDonald and Frank (2016) showing that prior passive learning improves subsequent active learning, and provides a mechanistic explanation of why the quality of active learning is sub-optimal early in learning, as experimentally demonstrated by Markant and Gureckis (2014).

Markant and Gureckis (2014) also showed that passive learners did not benefit from being "yoked" to active learners' data. Future work should investigate whether our model can also account for this finding.

There is a growing consensus in the artificial intelligence and cognitive science communities that the two fields should establish stronger ties, much like at the dawn of the two fields. Several articles have recently called for bringing the fields of artificial intelligence, cognitive science, and neuroscience closer together (Hassabis et al., 2017, Gershman et al., 2015). Pursuing this approach, our work, like the work of many before us, attests to the effectiveness of this idea by exemplifying how a synergistic interaction between machine learning and cognitive science helps develop effective, human-like artificial intelligence.

Acknowledgments: This works was supported by an operating grant to TRS from Natural Sciences and Engineering Research Council of Canada (NSERC).

References

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014a). Win-stay, lose-sample: A simple sequential algorithm for approximating bayesian inference. *Cognitive Psychology*, *74*, 35–65.

Bonawitz, E., Denison, S., Griffiths, T. L., & Gopnik, A. (2014b). Probabilistic models, learning algorithms, and response variability: sampling in cognitive development. *Trends in Cognitive Sciences*, *18*(10), 497–500.

Bröder, A. (2003). Decision making with the" adaptive toolbox": influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 611.

Bruner, J. S., Jolly, A., & Sylva, K. (1976). Play: Its role in development and evolution.

Cherney, I. D. (2008). The effects of active learning on students' memories for course content. *Active Learning In Higher Education*, *9*(2), 152–171.

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gershman, S. J., Vul, E., & Tenenbaum, J. B. (2012). Multistability and perceptual inference. *Neural Computation*, *24*(1), 1–24.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Hanneke, S. (2014). *Theory of Active Learning* (Tech. Rep.). Available: http://www.stevehanneke.com/.

Hanneke, S. (2016). The optimal sample complexity of pac learning. *The Journal of Machine Learning Research*, *17*(1), 1319–1333.

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, *95*(2), 245–258.

Hoffart, J. C., Rieskamp, J., & Dutilh, G. (2018). How environmental regularities affect people's information search in probability judgments from experience. *Journal of Experimental Psychology. Learning, Memory, and Cognition*.

Káli, S., & Dayan, P. (2004). Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nature Neuroscience*, *7*(3), 286.

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, *3361*(10), 1995.

Lengyel, M., & Dayan, P. (2008). Hippocampal contributions to control: the third way. In *Advances in neural information processing systems* (pp. 889–896).

Lieder, F., & Griffiths, T. L. (2017). Strategy selection as rational metareasoning. *Psychological review*, *124*(6), 762.

MacDonald, K., & Frank, M. (2016). When does passive learning improve the effectiveness of active learning? In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society.*

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94.

Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education*, *30*(4), 159–167.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... others (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529.

Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *eLife*, *7*, e32548.

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences*, *108*(30), 12491–12496.

Nobandegani, A. S., & Shultz, T. R. (2017). Converting cascade-correlation neural nets into probabilistic generative models. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society.*

Nobandegani, A. S., & Shultz, T. R. (2018). Example generation under constraints using cascade correlation neural nets. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society.*

Ólafsdóttir, H. F., Bush, D., & Barry, C. (2018). The role of hippocampal replay in memory and planning. *Current Biology*, *28*(1), R37–R50.

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L., & Goldstein, D. G. (2011). The recognition heuristic: A review of theory and tests. *Frontiers in Psychology*, *2*, 147.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534.

Rieskamp, J., & Otto, P. E. (2006). Ssl: a theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*(2), 207.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation* (Tech. Rep.). California Univ San Diego La Jolla Inst for Cognitive Science.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893.

Savin, C., & Deneve, S. (2014). Spatio-temporal representations of uncertainty in spiking neural networks. In *Advances in Neural Information Processing Systems*.

Watkins, C., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3-4), 279–292.