Meta-learning for Predictive Knowledge Architectures: A Case Study Using TIDBD on a Sensor-rich Robotic Arm

Extended Abstract

Johannes Günther
University of Alberta
Edmonton, Canada
gunther@ualberta.ca

Alex Kearney
University of Alberta
Edmonton, Canada
kearney@ualberta.ca

Nadia M. Ady
University of Alberta
Edmonton, Canada
nmady@ualberta.ca

Michael R. Dawson
University of Alberta
Edmonton, Canada
mrd1@ualberta.ca

Patrick M. Pilarski
University of Alberta
Edmonton, Canada
pilarski@ualberta.ca

ABSTRACT

Predictive approaches to modelling the environment have seen recent successes in robotics and other long-lived applications. These predictive knowledge architectures are learned incrementally and online, through interaction with the environment. One challenge for applications of predictive knowledge is the necessity of tuning feature representations and parameter values: no single step size will be appropriate for every prediction. Furthermore, as sensor signals might be subject to change in a non-stationary world, predefined step sizes cannot be sufficient for an autonomous agent. In this paper, we explore Temporal-Difference Incremental Delta-Bar-Delta (TIDBD)—a meta-learning method for temporal-difference (TD) learning which adapts a vector of many step sizes, allowing for simultaneous step size tuning and representation learning. We demonstrate that, for a predictive knowledge application, TIDBD is a viable alternative to tuning step-size parameters, by showing that the performance of TIDBD is comparable to that of TD with an exhaustive parameter search. Performance here is measured in terms of root mean squared difference from the true value, calculated offline. Moreover, TIDBD can perform representation learning, potentially supporting robust learning in the face of failing sensors. The ability for an autonomous agent to adapt its own learning and adjust its representation based on interactions with its environment is a key capability. With its potential to fulfill these desiderata, meta-learning is a promising component for future systems.

KEYWORDS

Continual learning; Reinforcement learning; Robot learning; Long-term autonomy

ACM Reference Format:
Johannes Günther, Alex Kearney, Nadia M. Ady, Michael R. Dawson, and Patrick M. Pilarski. 2019. Meta-learning for Predictive Knowledge Architectures:
A Case Study Using TIDBD on a Sensor-rich Robotic Arm. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 3 pages.

Introduction: The real world is non-stationary and complex, so many of the challenges facing an autonomous agent cannot be completely foreseen. An agent should therefore be empowered to adapt to its environment without human assistance. General value functions (GVFs) [16, 18] allow agents to incrementally construct knowledge of the environment purely through interaction [2, 11]. With a GVF architecture, the environment is modelled as a set of forecasts—GVFs—about how signals of interest will behave. An agent's actions affect its world, so these forecasts are made with consideration to a policy of agent behaviour. In this way, these predictions can capture forward-looking aspects of the environment such as, "If I continue moving my arm to the right, how hot do I expect my elbow servo to get?" Many GVFs can be simultaneously made and learned online, in real time [8], using methods such as temporal-difference (TD) learning [15] and other standard learning algorithms from the field of reinforcement learning. GVFs have already shown their potential in multiple real-world domains [e.g., 1, 4, 5, 10, 13]. Many simultaneous GVFs can be learned using a single, shared representation [8], but no single step size will be appropriate for all predictions, and no representation will be equally suitable for all predictions. Therefore, it is desirable to tune both the step size and the representations for each individual prediction. One tuning method is Temporal-Difference Incremental Delta-Bar-Delta (TIDBD) [6], a step-size adaptation method for TD learning.

TIDBD adjusts a vector of many step sizes—one step size for each feature. By adapting step sizes on a per-feature basis, we are able to tune them based on their relevance; features which are highly correlated to the prediction problem should be given large step sizes, while irrelevant features should be given smaller step sizes.

In this work, we use experiments to investigate the effect that TIDBD has on predictions about real sensor signals provided by a sensor-rich robotic arm—the Modular Prosthetic Limb (MPL), shown in Figure 1. The sensor data was recorded during alternating patterns of rest and movement, which can be viewed at `https://blinclab.ca/mpl_teleop_video/`.

Experiment 1: Our primary experiment compared TIDBD, which assigns individual step sizes to features, to classic TD, which uses a fixed step size for all features. For both learning algorithms (TD and TIDBD), we used selective Kanerva coding [17] to choose features to represent the 108-dimensional sensor space for linear function

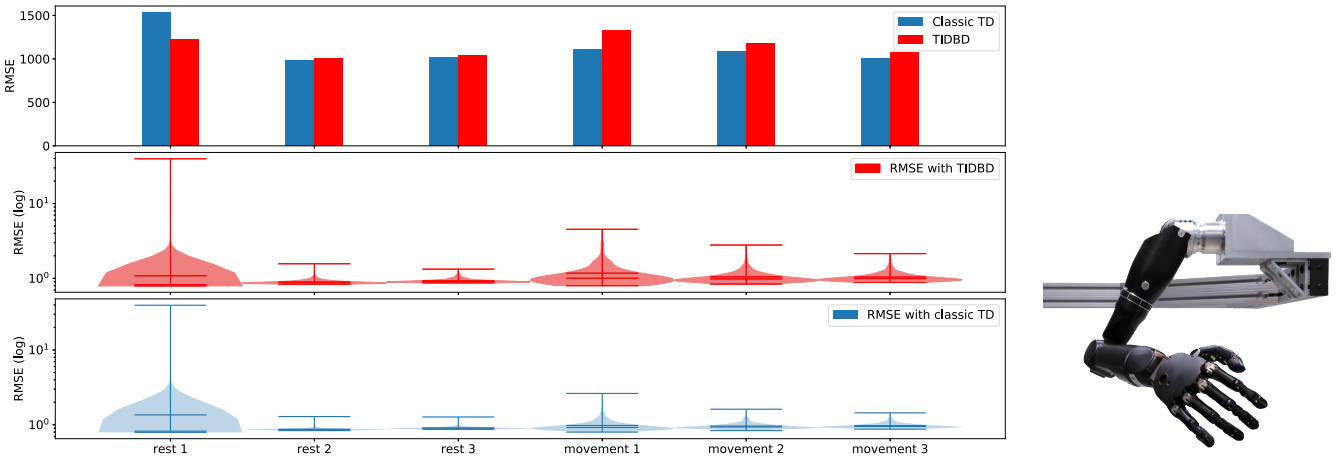

Figure 1: The top pane shows the root mean squared error (RMSE) for classic TD and TIDBD for each time periods. The middle and bottom panes show violin plots for the RMSE, for TIDBD and classic TD. All plots are averages over 30 independent runs.

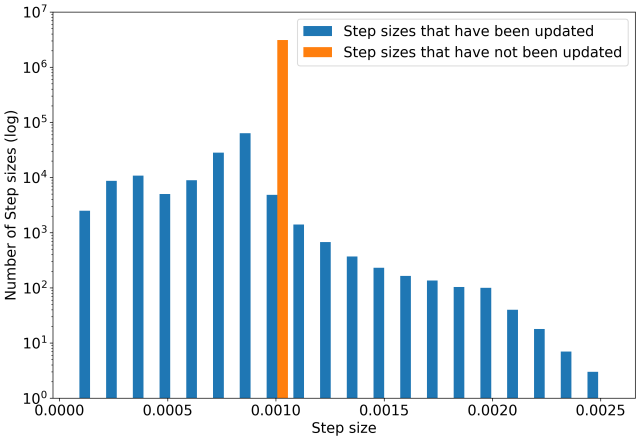

Figure 2: Step-size distribution at termination.

approximation. For each learning algorithm, we performed an extensive (full-factorial) parameter sweep to set the feature representation, and, additionally for TD only, the fixed step size. We did not tune the initial step size for TIDBD, to better demonstrate its performance using a generic value.

We compared the root mean squared error obtained with TIDBD and classic TD. As shown in Figure 1, the two algorithms perform comparably, given careful choice of step size for classic TD. For both TD and TIDBD, the error (shown as violin plots in the lower panes of Figure 1) is primarily due to a small number of GVFs—the majority of the predictions are learned quickly. The most noticeable difference in performance between the two algorithms is that TIDBD shows better early learning (during 'rest 1'). *We therefore suggest that roboticists should consider using a step-size adaptation method like TIDBD for prediction learning* with general value functions, rather than performing an extensive sweep over step sizes for classic TD.

We tuned parameters offline and chose a step size for TD that provided the best performance over the total duration of the data. This advantage is not available for designers of truly online, autonomous agents, so any fixed step size would offer worse performance—which would worsen with wear-and-tear or unexpected changes in the environment. Parameter tuning is computationally expensive and time intensive. It is often skipped and substituted with a general-use parameter value—to the detriment of performance. With this experiment, we demonstrated that a step size adaptation method can replace hand-tuning, and compare well to even the best-case fixed step size. In our experiments, the time to update all step sizes remained within real-time requirements (0.28s).

Experiment 2: As a second series of experiments, we modified the original sensor data to simulate four broken sensors [9], and then to simulate four stuck sensors [7]. We found that TIDBD gradually and automatically excluded the broken sensor signals from the representation by decreasing the associated step sizes. While TIDBD did not automatically resolve the issues created by stuck sensors, the resulting changes in step sizes were clearly distinguishable from those seen during normal functioning of the arm. *These results suggest that it may be possible to automatically identify certain sensor failures by monitoring changes to a group of step sizes.*

Conclusions: These three results—the observation that TIDBD appropriately updates step sizes to accommodate non-stationarity, the distinct reaction of a group of step sizes to stuck sensors, and the automatic feature selection performed by TIDBD for uninformative sensors—are promising key findings for long-term autonomous agents. They empower an agent to not only adapt its learning based on interactions with its environment, but to evaluate and improve its own perception of said environment. As the step sizes provided by TIDBD contain information about the history of each feature, step sizes could also provide an important source of information for the agent itself to learn from. Such introspective signals have already been argued to be a helpful source of information for an agent to better understand its environment and its own functioning within its environment [3, 12, 14]. The insights provided by this work therefore offer a deeper understanding and intuition about TIDBD, aiming to help other designers in creating agents that are capable of autonomous learning and adaptation through interaction with their environment.

REFERENCES

[1] Ashley N. Dalrymple, Dirk G. Everaert, Richard S. Sutton, and Vivian K. Mushahwar. 2018. Online Prediction of Phases of the Gait Cycle for Control of Intraspinal Microstimulation. In *Society for Neuroscience*.

[2] Gary L. Drescher. 1991. *Made-Up Minds: A Constructivist Approach to Artificial Intelligence*. MIT Press.

[3] Johannes Günther, Alex Kearney, Michael R. Dawson, Craig Sherstan, and Patrick M. Pilarski. 2018. Predictions, Surprise, and Predictions of Surprise in General Value Function Architectures. In *AAAI 2018 Fall Symposium on Reasoning and Learning in Real-World Systems for Long-Term Autonomy*, 22–29.

[4] Johannes Günther, Patrick M. Pilarski, Gerhard Helfrich, Hao Shen, and Klaus Diepold. 2016. Intelligent Laser Welding Through Representation, Prediction, and Control Learning: An Architecture with Deep Neural Networks and Reinforcement Learning. *Mechatronics* 34, 1–11.

[5] Gregory Kahn, Adam Villaflor, Bosen Ding, Pieter Abbeel, and Sergey Levine. 2018. Self-Supervised Deep Reinforcement Learning with Generalized Computation Graphs for Robot Navigation. In *Proceedings of the International Conference on Robotics and Automation*, 1–8.

[6] Alex Kearney, Vivek Veeriah, Jaden B. Travnik, Patrick M. Pilarski, and Richard S. Sutton. 2019. Learning Feature Relevance Through Step Size Adaptation in Temporal-Difference Learning. arXiv:1903.03252 [cs.LG]

[7] Xiao-Jian Li and Guang-Hong Yang. 2012. Fault Detection for Linear Stochastic Systems with Sensor Stuck Faults. *Optimal Control Applications and Methods* 33, 1, 61–80.

[8] Joseph Modayil, Adam White, and Richard S. Sutton. 2014. Multi-Timescale Nexting in a Reinforcement Learning Robot. *Adaptive Behavior* 22, 2, 146–160.

[9] Kevin Ni, Nithya Ramanathan, Mohamed Nabil Hajj Chehade, Laura Balzano, Sheela Nair, Sadaf Zahedi, Eddie Kohler, Greg Pottie, Mark Hansen, and Mani Srivastava. 2009. Sensor Network Data Fault Types. *ACM Transactions on Sensor Networks (TOSN)* 5, 3, Article 25 (May 2009), 29 pages.

[10] Patrick M. Pilarski, Michael R. Dawson, Thomas Degris, Jason P. Carey, K. Ming Chan, Jacqueline S. Hebert, and Richard S. Sutton. 2013. Adaptive Artificial Limbs: A Real-Time Approach to Prediction and Anticipation. *IEEE Robotics & Automation Mag.* 20, 1, 53–64.

[11] Mark B. Ring. 1994. *Continual Learning in Reinforcement Environments*. PhD Thesis. University of Texas at Austin, Austin, TX.

[12] Wolfram Schultz and Anthony Dickinson. 2000. Neuronal Coding of Prediction Errors. *Annual Review of Neuroscience* 23, 1, 473–500.

[13] Craig Sherstan, Joseph Modayil, and Patrick M. Pilarski. 2015. A Collaborative Approach to the Simultaneous Multi-Joint Control of a Prosthetic Arm. In *Proceedings of the International Conference on Rehabilitation Robotics*, 13–18.

[14] Craig Sherstan, Adam White, Marlos C. Machado, and Patrick M. Pilarski. 2016. Introspective Agents: Confidence Measures for General Value Functions. In *Proceedings of the International Conference on Artificial General Intelligence*. Springer, 258–261.

[15] Richard S. Sutton. 1988. Learning to Predict by the Methods of Temporal Differences. *Machine Learning* 3, 1 (August 1998), 9–44.

[16] Richard S. Sutton, Joseph Modayil, Michael Delp, Thomas Degris, Patrick M. Pilarski, Adam White, and Doina Precup. 2011. Horde: A Scalable Real-time Architecture for Learning Knowledge from Unsupervised Sensorimotor Interaction. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems, Vol. 2*, 761–768.

[17] Jaden B. Travnik and Patrick M. Pilarski. 2017. Representing High-dimensional Data to Intelligent Prostheses and Other Wearable Assistive Robots: A First Comparison of Tile Coding and Selective Kanerva Coding. In *Proceedings of the International Conference on Rehabilitation Robotics*, 1443–1450.

[18] Adam White. 2015. *Developing a Predictive Approach to Knowledge*. PhD Thesis. University of Alberta, Edmonton, AB, Canada.