Towards time-scale invariant reinforcement learning

Md Rysul Kabir James Mochizuki-Freeman Zoran Tiganj
Department of Computer Science, Indiana University Bloomington

Abstract

The ability to estimate temporal relationships is critical for both animals and artificial agents. Cognitive science and neuroscience studies provide remarkable insights into behavioral and neural aspects of interval timing. In particular, scale-invariance observed in behavior and supported by neural data is one of the key principles that goes beyond interval timing and governs animal perception. Furthermore, once they learn a task, humans and other animals can rapidly adapt to temporally rescaled versions at a wide range of scales. We show that using a scale-invariant cognitive model of working memory combined with convolutional and max-pool layers gives rise to reinforcement learning (RL) agents that are invariant to temporal rescaling in the environment. We illustrate this using a simple interval bisection task and show that this property is specific to scale-invariant memory and not observed in commonly used recurrent networks.

Keywords: Scale-invariance; Deep RL; Interval timing

Introduction

Learning temporal relationships between cause and effect is critical for successfully obtaining rewards and avoiding punishments in a natural environment. Humans and many other animals can estimate the temporal duration of events and use that estimate as an integral component of decision-making. Furthermore, the ability to do this rapidly and flexibly across a wide range of temporal scales has been critical for survival. While machine learning systems also possess the capacity to represent the elapsed time, typically via recurrent connections, they often struggle with learning temporal relationships and generalizing across multiple scales.

Critically, the mammalian ability to estimate time is known to be scale-invariant across a wide range of temporal scales from seconds to minutes and hours (Buhusi & Meck, 2005; Gibbon, 1977; Buhusi et al., 2009; Balci & Freestone, 2020). A scale-invariant system has a linear relationship between the mean estimated time and the actual time, with a constant coefficient of variation, known as Weber's law (Portugal & Svaiter, 2011). This law is foundational for understanding mammalian perceptions and spans virtually all perceptual domains except for angles (Gibbon, 1977; Wilkes, 2015). Machine learning systems are typically not scale-invariant and they tend to perform well only at a limited set of scales and require adjustments of hyperparameters to learn problems at different scales. Such hyperparameters include learning rate, temporal resolution and temporal discounting (Tiganj, Gershman, Sederberg, & Howard, 2019).

Following previous work on scale-invariant memory (Shankar & Howard, 2012) and its applications in building artificial systems invariant to temporal rescaling (Jacques, Tiganj, Sarkar, Howard, & Sederberg, 2022), we show that knowledge learned at a single temporal scale can be generalized to a wide range of temporal scales in an RL setting. We demonstrate this on a temporal bisection task where agents learned to differentiate three short intervals from three long intervals. Once they learned the task using REINFORCE (Sutton, McAllester, Singh, & Mansour, 1999), the intervals were rescaled by a factor of 2 and a factor of 4. Without additional training, the agents were able to generalize and maintain perfect performance. This is because scale-invariant memory constructs a log-compressed internal timeline. Applying the logarithm to the temporally rescaled environment turns rescaling into translation ($log(ax) = log(a) + log(x)$). We then used CNN combined with max-pool to achieve invariance to rescaling.

Methods

Building on models from computational and cognitive neuroscience (Shankar & Howard, 2012; Howard et al., 2014), we designed a neural network architecture in which the impulse response gives rise to scale-invariant sequentially activated set of neurons (analogous to time cells, reported in mammalian brains (Pastalkova, Itskov, Amarasingham, & Buzsaki, 2008; MacDonald, Lepage, Eden, & Eichenbaum, 2011; Salz et al., 2016; Tiganj, Jung, Kim, & Howard, 2017)). Specifically, this network first constructs an approximation of a real-domain Laplace transform of the temporal history of the input signal $f(t)$:

$$F(s;t) = \int_0^t e^{-s(t-t')} f(t')dt'. \tag{1}$$

The impulse response (response to input $f(t) = \delta(0)$) of $F(s;t)$ decays exponentially as a function of time t with decay rate s: e^{-st} (Fig. 1). The above equation can be discretized and implemented as a recurrent neural network with a fixed (not trainable) diagonal connectivity matrix with $e^{-s\Delta t}$ values along the diagonal. The output of the recurrent layer is mapped through a linear layer with analytically computed weights (Post, 1930) implementing a discrete approximation of the inverse Laplace transform $\tilde{f}$ giving rise to sequentially activated cells (Fig. 1) that together constitute a scale-invariant internal timeline of the past (Fig. 2):

$$\tilde{f}(\overset{*}{\tau};t) = \mathbf{L}_k^{-1} F(s;t) = -\frac{1}{\overset{*}{\tau}} \frac{k^{k+1}}{k!} \left(\frac{-t}{\overset{*}{\tau}}\right)^k e^{k\frac{t}{\overset{*}{\tau}}}, \tag{2}$$

where $\overset{*}{\tau} := k/s$ and k is a parameter that affects the width of the sequentially activated cells (larger k results in smaller width). The above equation is scale-invariant in a sense that rescaling $\tilde{f}(\overset{*}{\tau};t) \longrightarrow \tilde{f}(\overset{*}{\tau};\alpha t)$ can be undone by setting $\overset{*}{\tau}_i \longrightarrow \overset{*}{\tau}_i/\alpha$. Choosing $\overset{*}{\tau}$ to be log-spaced ($\overset{*}{\tau}_i = (1+c)^{i-1}\overset{*}{\tau}_{min}$,

with $c > 0$) makes the rescaling of taustar equivalent to translation: $\overset{*}{\tau}_i = \overset{*}{\tau}_{i+\Delta}$ where $\Delta = log_{1+c}a$. This implies that temporal rescaling will cause a translation of the sequentially activated units (Fig. 3A,B). Our network architecture had 50 log-spaced units with $\overset{*}{\tau}_{min} = 1$, $\overset{*}{\tau}_{max} = 700$, and $k = 8$. Once temporal rescaling is converted into translation, we apply convolution and maxpool, which are translation invariant. Therefore, the network becomes invariant to temporal rescaling (Fig. 3C). The output of the max pool is fed into two dense layers followed by a layer with two neurons and softmax activation.

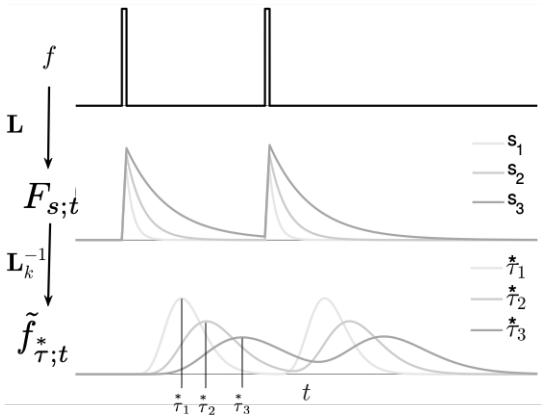

Figure 1: Response of the CogRNN to δ pulses. Neurons in $F_{s;t}$ decay exponentially at a spectrum of time constants $\mathbf{s}$. Neurons in $\tilde{f}_{\overset{*}{\tau};s}$ activate sequentially resembling time cells.

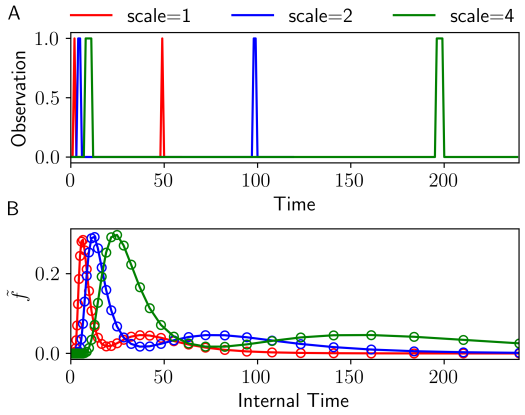

Figure 2: **A.** Observations in the largest interval (48 time steps) at different scales: the two pulses represent the start and stop of the interval. **B.** $\tilde{f}$ activity after the second pulse was presented. Internal time values are equal to values of $\overset{*}{\tau}$: activity of each neuron in $\tilde{f}$ is shown at corresponding $\overset{*}{\tau}$.

Results and Discussion

We evaluated RL agents on a temporal bisection task where they were trained to differentiate three short intervals (30, 33, and 36 time steps) from three long intervals (40, 44 and 48),

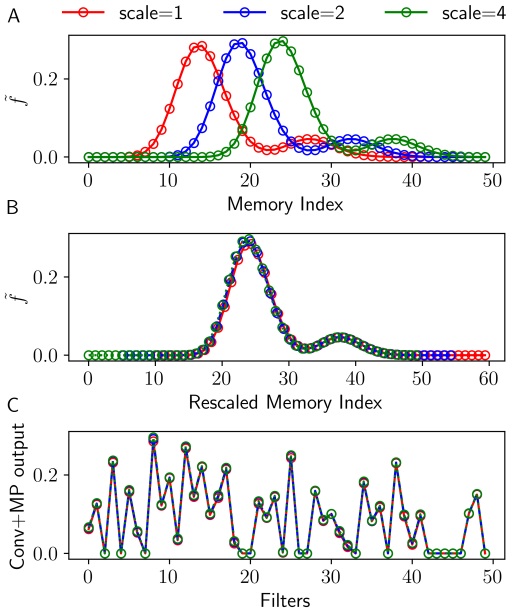

Figure 3: **A.** $\tilde{f}$ activity after the second pulse as a function of $\overset{*}{\tau}$ index (rather than $\overset{*}{\tau}$ value shown in Fig. 2B). Due to the log-compression temporal rescaling results in translation. **B.** Overlapping $\tilde{f}$ activity after $\log_{1+c}(\alpha)$ translation. **C.** Applying convolution and max-pool to $\tilde{f}$ from panel A results in the output invariant to temporal rescaling.

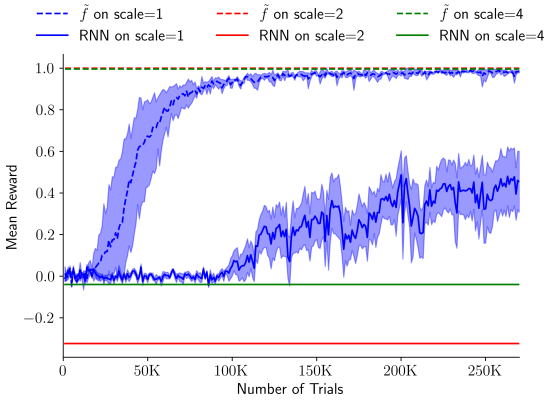

Figure 4: Performance as a function of trial numbers for training at scale=1 and evaluating at scale=2 and 4.

see Fig. 2A for an illustration of the observation space (the durations were based on a neuroscience study by Kim, Ghim, Lee, and Jung (2013)). The proposed agents were able to learn the task after about 200k trials (Fig. 4, dashed blue line). Then, we rescaled the time in the observation space by a factor of 2 and factor 4. Without additional training, the agents reached perfect performance (Fig. 4, dashed red and green lines). This was not the case for agents trained with RNNs, which struggled to learn the task and failed to generalize to different temporal scales. This result illustrates that log-compressed memory can enable generalization to temporal rescaling in the RL setting.

Acknowledgement

We gratefully acknowledge support from the National Institutes of Health's National Institute on Aging, grant 5R01AG076198-02.

References

Balci, F., & Freestone, D. (2020). The peak interval procedure in rodents: a tool for studying the neurobiological basis of interval timing and its alterations in models of human disease. *Bio-protocol*, *10*(17), e3735–e3735.

Buhusi, C. V., Aziz, D., Winslow, D., Carter, R. E., Swearingen, J. E., & Buhusi, M. C. (2009). Interval timing accuracy and scalar timing in c57bl/6 mice. *Behavioral neuroscience*, *123*(5), 1102.

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? functional and neural mechanisms of interval timing. *Nature reviews neuroscience*, *6*(10), 755–765.

Gibbon, J. (1977). Scalar expectancy theory and weber's law in animal timing. *Psychological review*, *84*(3), 279.

Howard, M. W., MacDonald, C. J., Tiganj, Z., Shankar, K. H., Du, Q., Hasselmo, M. E., & Eichenbaum, H. (2014). A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *Journal of Neuroscience*, *34*(13), 4692–4707.

Jacques, B. G., Tiganj, Z., Sarkar, A., Howard, M., & Sederberg, P. (2022). A deep convolutional neural network that is invariant to time rescaling. In *International conference on machine learning* (pp. 9729–9738).

Kim, J., Ghim, J.-W., Lee, J. H., & Jung, M. W. (2013). Neural correlates of interval timing in rodent prefrontal cortex. *Journal of Neuroscience*, *33*(34), 13834–13847.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737–749.

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*(5894), 1322–1327.

Portugal, R., & Svaiter, B. F. (2011). Weber-Fechner law and the optimality of the logarithmic scale. *Minds and Machines*, *21*(1), 73–81.

Post, E. (1930). Generalized differentiation. *Transactions of the American Mathematical Society*, *32*, 723-781.

Salz, D. M., Tiganj, Z., Khasnabish, S., Kohley, A., Sheehan, D., Howard, M. W., & Eichenbaum, H. (2016). Time cells in hippocampal area ca3. *Journal of Neuroscience*, *36*(28), 7476–7484.

Shankar, K. H., & Howard, M. W. (2012). A scale-invariant internal representation of time. *Neural Computation*, *24*(1), 134–193.

Sutton, R. S., McAllester, D., Singh, S., & Mansour, Y. (1999). Policy gradient methods for reinforcement learning with function approximation. In S. Solla, T. Leen, & K. Müller (Eds.), *Advances in neural information processing systems* (Vol. 12). MIT Press.

Tiganj, Z., Gershman, S. J., Sederberg, P. B., & Howard, M. W. (2019). Estimating scale-invariant future in continuous time. *Neural computation*, *31*(4), 681–709.

Tiganj, Z., Jung, M. W., Kim, J., & Howard, M. W. (2017). Sequential firing codes for time in rodent medial prefrontal cortex. *Cerebral Cortex*, *27*(12), 5663–5671.

Wilkes, J. T. (2015). *Reverse first principles: Weber's law and optimality in different senses*. Unpublished doctoral dissertation, UNIVERSITY OF CALIFORNIA, SANTA BARBARA.