
Modelling Working Memory using Deep Recurrent Reinforcement Learning

Pravish Sainath^{1,2,3}

Pierre Bellec^{1,3}

Guillaume Lajoie^{2,3}

¹Centre de Recherche, Institut Universitaire de Gériatrie de Montréal

²Mila - Quebec AI Institute

³Université de Montréal

Abstract

1 In cognitive systems, the role of a working memory is crucial for visual reasoning
2 and decision making. Tremendous progress has been made in understanding the
3 mechanisms of the human/animal working memory, as well as in formulating
4 different frameworks of artificial neural networks. In the case of humans, the
5 visual working memory (VWM) task ^[1] is a standard one in which the subjects are
6 presented with a sequence of images, each of which needs to be identified as to
7 whether it was already seen or not. Our work is a study of multiple ways to learn a
8 working memory model using recurrent neural networks that learn to remember
9 input images across timesteps in supervised and reinforcement learning settings.
10 The reinforcement learning setting is inspired by the popular view in Neuroscience
11 that the working memory in the prefrontal cortex is modulated by a dopaminergic
12 mechanism. We consider the VWM task as an environment that rewards the
13 agent when it remembers past information and penalizes it for forgetting. We
14 quantitatively estimate the performance of these models on sequences of images
15 from a standard image dataset (CIFAR-100 ^[2]) and their ability to remember
16 and recall. Based on our analysis, we establish that a gated recurrent neural
17 network model with long short-term memory units trained using reinforcement
18 learning is powerful and more efficient in temporally consolidating the input spatial
19 information. This work is an initial analysis as a part of our ultimate goal to model
20 the behavior and information processing of the working memory of the brain and to
21 use brain imaging data captured from human subjects during the VWM cognitive
22 task to understand various memory mechanisms of the brain.

23 1 Introduction

24 Memory is an essential component for solving many tasks intelligently. Most sequential tasks
25 involve the need for a mechanism to maintain a *context*. In the brain, working memory serves as
26 a work space to encode and maintain the most relevant information over a short period of time, in
27 order to use it to guide behavior for cognitive tasks. Several cognitive tasks have been proposed
28 in the Neuropsychology literature to study and understand the working memory in animals. The
29 Visual Working Memory Task (VWM task) [1] or the classic N-back task is one of the most simple
30 and popular ones. It involves showing sequences of images to subjects and record their responses
31 indicating whether they have seen the image already.

32 On the other hand, with artificial intelligence systems, there has been very good progress in models
33 that learn from sequences of inputs using artificial neural networks as memory for all types of learning

34 (supervised, unsupervised and reinforcement). We intend to use these developments as an ideal
 35 opportunity for synergy to computationally model the working memory system of the brain.
 36 As memory is an important aspect of both artificial intelligence and neuroscience, there are some
 37 good studies that helped choose our models as discussed in Section 2.

38 2 Models

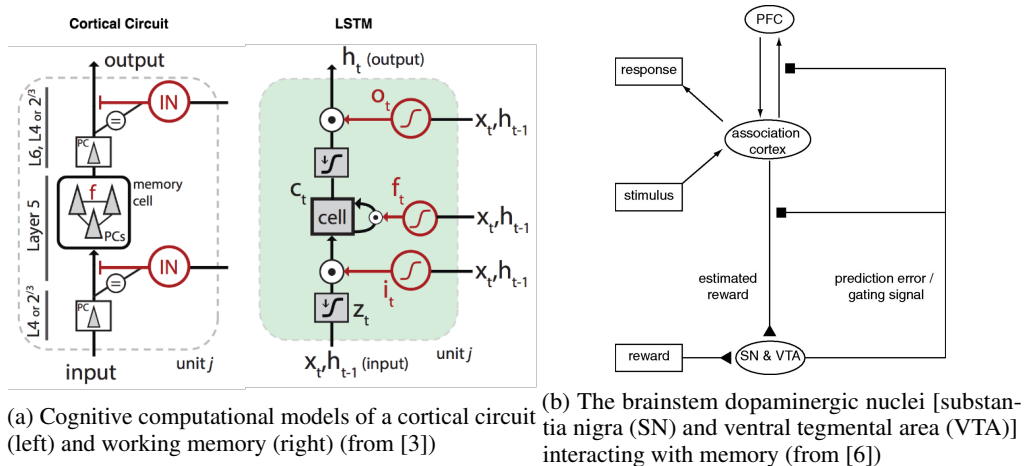


Figure 1: Inspirations for our models and training settings

39 Some ideas that motivated our choice of models :

40 **Gated recurrent-units as cortical circuits** [3] highlights that Long short-term memories (LSTMs
 41 [4]), the most commonly used gated recurrent units with memory cells controlled by input, output
 42 and forget gates and these can be considered (at a high-level) as an abstraction of a cortical circuit
 43 unit composed of layers of pyramidal cells with gating as shown in Figure 1a.

44 **Dopaminergic modulation of working memory** As detailed in [5] and [6] and shown in Figure 1b,
 45 the stimulus receive a reward from the dopaminergic nuclei based on which it is stored in the prefrontal
 46 cortex (PFC) as persistent activations. Thus, biologically there is a reward-processing mechanism
 47 (temporal difference, TD) that is operating outside the cortical memory.

48 **Deep Recurrent Reinforcement Learning** As an extension of Deep Q-Networks (DQN)[7] with
 49 recurrent units were proposed as DRQNs by [8] for partially observable environments. This model
 50 naturally suits the non-Markov decision process involved in learning. ([9; 10; 11])

51 The models are categorized under two different settings :

52 **Supervised Learning models** We consider 2 different models for training recurrent networks to
 53 output the probability of past occurrence of each image in the sequence using supervised learning:
 54 LSTM and LSTM-Attention (LSTM-A). The attention augmented version ([12]) helps learn the
 55 context better. These are used in conjunction with a CNN and trained using backpropagation through-
 56 time (without truncation), as shown in Figure 2a.

57 **Reinforcement Learning models** We consider an environment where each image in the sequence
 58 is a state that the DRQN uses as input to approximate the Q-values of the two different action : 0
 59 for *not seen* and 1 for *seen*. An ϵ -greedy policy is learned using these Q-values to select the action
 60 and receive a reward from the environment (positive value for correct action and negative value for
 61 wrong action) as shown in Figure 2b. These are trained using common TD control algorithms such as
 62 Q-learning and SARSA, indicated as Q+DQRN and SARSA+DQRN respectively.

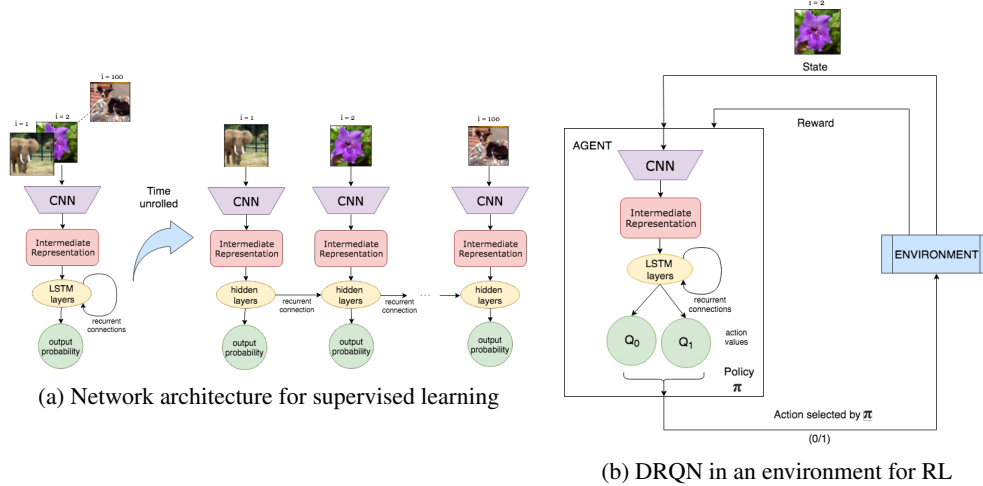
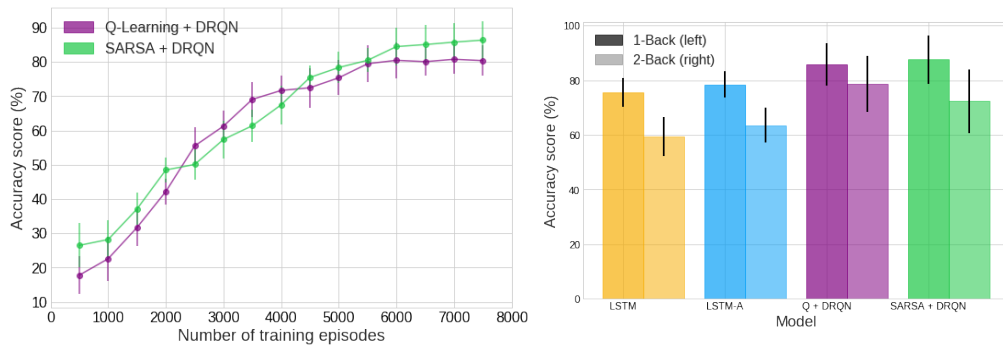


Figure 2: Model architectures

63 3 Experiments and Results

64 For all experiments in both supervised and reinforcement learning settings, 100 images were drawn
 65 from CIFAR-100 ([2]) dataset for each sequence. A CNN that was pre-trained on the same dataset
 66 for the classification task was used for intermediate representation before LSTM layers (as shown in
 67 Figure 2).



(a) Final test performance of DRQN models after every 500 training episodes (sequences)
 (b) Final test performance of the 4 trained models for repeat counts $N = 1$ (left) and $N = 2$ (right)

Figure 3: Final performance of the models (the lines indicate std. dev.)

68 The problem solved by all the models is a binary classification problem, predicting $unseen(0)/seen(1)$.
 69 The performance of all the models in the experiments were measured using the accuracy metric
 70 calculated based on the number of correct predictions for the 100 images in a sequence (as a %). This
 71 evaluation was repeated for 10 independent trials as a part of ablation studies.

72 Figure 3a shows the test accuracy of both the reinforcement learning models after fixed training
 73 episodes to track the progress. Figure 3b depicts the performance of the different models using mean
 74 final accuracy scores for two different conditions of repetitions of the stimuli images: $N = 1$ and
 75 $N = 2$. Figure 4 indicates the variation of test accuracy of each of the trained models over the course
 76 of a testing sequence. It can be seen that the models trained in the reinforcement learning setting
 77 outperform both the LSTM and LSTM-A in the supervised setting. Both on-policy and off-policy
 78 methods help in solving the task and learning good parametric functions in the LSTMs.

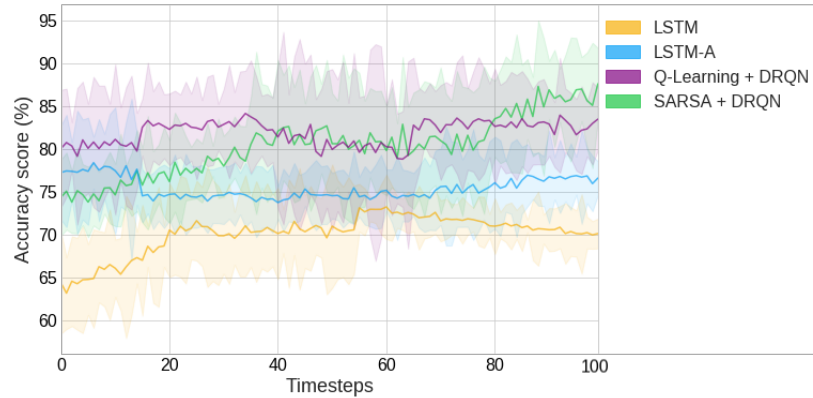


Figure 4: Performance of the trained models measured during test runs within a sequence

79 4 Conclusions and Future work

80 Popular literature in deep reinforcement learning ([8], [11]) outline many architectures and methods
 81 to solve specific (mostly motor) decision making tasks in an environment. In our work, we have taken
 82 an alternative view where we try to solve a strongly memory-oriented task using similar models used
 83 in reinforcement learning to emulate the reward processing happening to store memory.

84 A good working memory model understands what to remember and what to forget. From our
 85 study, we conclude that modelling a working memory using gated recurrent neural networks (such
 86 as LSTMs) to train using a reward-based learning approach offers significant advantages giving a
 87 reasonably superior performance. From the models studied, it can be seen that in addition to the
 88 capacity of the recurrent network, a training in the reinforcement learning setting offers a better
 89 generalization with its power for acting as a good computational model for the biological working
 90 memory for explaining the VWM task. This seems to indicate that the dopaminergic control of
 91 memory in the PFC is a high-level principle that is common to both artificial and biological neural
 92 systems. Also, this observation could be attributed to the fact that deep reinforcement learning is a
 93 better framework for learning in this case, given the non-stationary and non-Markov nature of the
 94 task involved.

95 As a next step, we plan to model other cognitive tasks for memory as a reinforcement learning
 96 problem and compare the performance of different algorithms and gating mechanisms in the networks.
 97 Further, we would like to extend our work by using the brain responses of humans solving the task
 98 from their fMRI data and identify the neural correlates of the visuo-temporal streams of information.
 99 This would throw more light on the functional similarities of the biological and artificial neural
 100 models explaining how the memory system functions computationally.

101 References

- 102 [1] Luck, S. J., Vogel, E. K. (2013) Visual working memory capacity: from psychophysics
 103 and neurobiology to individual differences. *Trends in cognitive sciences*, 17(8), 391–400.
 104 doi:10.1016/j.tics.2013.06.006.
- 105 [2] Krizhevsky, Alex (2009) Learning Multiple Layers of Features from Tiny Images
- 106 [3] Costa, Rui Ponte., Assael, Yannis M., Shillingford, Brendan., de Freitas, Nando., and Vogels,
 107 Tim P. (2017) Cortical microcircuits as gated-recurrent neural networks. In *Advances in Neural
 108 Information Processing Systems 31 (NIPS-17)*, pp. 271–282.
- 109 [4] Hochreiter, Sepp and Schmidhuber, Jurgen. (1997) Long short-term memory. *Neural Computa-
 110 tion*, 9(8): 1735–1780.
- 111 [5] Braver, Todd.S., D.Cohen, Jonathon. (2000) On the control of control: The role of dopamine in
 112 regulating prefrontal function and working memory. *Attention Performance XVIII*:713–737.
- 113 [6] D’Ardenne, Kimberlee., Eshel, Neir., Luka, Joseph., Lenartowicz, Agatha., Nystrom, Leigh E.,
 114 Cohen, Jonathan D. (2012) Role of PFC and DA system in context updating. *Proceedings of the
 115 National Academy of Sciences* 109 (49) 19900–19909; 10.1073/pnas.1116727109

- 116 [7] Mnih, Volodymyr. Kavukcuoglu, Koray, Silver, David, Rusu, Andrei A., Veness, Joel, Bellemare,
117 Marc G., Graves, Alex, Riedmiller, Martin, Fidjeland, Andreas K., Ostrovski, Georg, Petersen,
118 Stig, Beattie, Charles, Sadik, Amir, Antonoglou, Ioannis, King, Helen, Kumaran, Dharshan,
119 Wierstra, Daan, Legg, Shane, and Hassabis, Demis. (2015) Human-level control through deep
120 reinforcement learning. *Nature*, 518(7540):529–533.
- 121 [8] Hausknecht, Matthew and Stone, Peter. (2015) Deep recurrent Q-learning for partially observ-
122 able MDPs, *arXiv preprint*, arXiv:1507.06527.
- 123 [9] Bakker, Bram. (2002) Reinforcement learning with long short-term memory. In *Advances in*
124 *Neural Information Processing Systems 14 (NIPS-01)*, pp. 1475–1482.
- 125 [10] Whitehead, Steven D., Lin, Long-Ji. (1995) Reinforcement learning of non-Markov decision
126 processes. In *Artificial Intelligence*. (73) pp. 271–306
- 127 [11] Li, Lihong Li, Jianfeng Gao, Xiaodong He, Jianshu Chen, Li Deng, and Ji He. (2015) Recurrent
128 reinforcement learning: A hybrid approach. arXiv:1509.03044.
- 129 [12] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov,
130 Rich Zemel, and Yoshua Bengio. (2015) Show, attend and tell: Neural image caption generation
131 with visual attention. In *International conference on machine learning*. pp 2048–2057.