Not-so-Catastrophic Forgetting in Deep Reinforcement Learning on Sequential Atari

Giacomo Spigler
The BioRobotics Institute
Scuola Superiore Sant'Anna, Pisa (IT)
giacomo.spigler@santannapisa.it

Abstract

In this work we analyze the amount of forgetting on sequential Atari games learning using Deep Reinforcement Learning. Specifically, we show that while catastrophic forgetting is found to be highly disruptive in terms of the behavior of an agent, the changes to the parameters of its neural network are not as drastic as it could seem from the large drop in performance. Indeed, it is found that relatively short periods of retraining on previously learnt tasks are often sufficient to quickly recover and improve the lost performance.

1 Introduction

Catastrophic forgetting is a problem that affects a variety of biological and artificial learning systems, whereby sequential training on two different tasks results in a degraded performance on the first task [1, 2].

A wide variety of solutions have been proposed, from rehearsing previous samples or pseudo-samples [3], to employing sparse representations with reduced overlap between tasks [4], to solutions like Progressive Neural Networks [5], PathNet [6], Incremental Moment Matching [7], Memory Aware Synapses [8], and regularization methods aimed at protecting weights that are important for the previous tasks, like Elastic Weight Consolidation [9] and Synaptic Intelligence [10]. More detailed reviews are available in the literature [11, 12].

Of all the proposed solutions, Elastic Weight Consolidation [9] is the only one that has been benchmarked on a sequential Atari games learning benchmark, achieving a high performance on sets of 10 games repeatedly trained on in random alternation.

Here we use a simplified version of the protocol to analyze the amount of forgetting in Deep Reinforcement Learning on the sequential Atari benchmark. Specifically, we investigate whether a drop in performance on previous games down to random-play levels reflects major changes in the parameters of the neural network of the agent, or whether it is due to more subtle changes. In particular, *we measure the actual degree of catastrophic forgetting due to training on intervening games as the amount of retraining required for the agent to recover the performance lost due to interference*.

2 Methods

We tested a simplified sequential reinforcement learning protocol using pairs of Atari games from the OpenAI Gym library [13] and the Arcade Learning Environment [14]. The protocol consisted in training an agent on a first game for 10M frames, followed by 10M frames of a second game and another 10M frames on the first game. 5 game pairs were tested, `Pong+Tennis`, `FishingDerby+NameThisGame`, `RoadRunner+Pong`, `BeamRider+Assault` and `Atlantis+RoadRunner`. A Deep Reinforcement Learning agent was built using the GA3C [15] variant of the A3C algorithm [16], where all experiences from the worker threads were pooled together and processed by a single training thread. The input frames were converted to grayscale, cropped and re-scaled to 84×84 pixels, and input states consisted of a history of 4 consecutive frames, as done in [17]. The network used was a CNN shaped as $\{Conv2D(32, 8s4)$, $(Conv2D(64, 4s2)$, $\{Conv2D(64, 3s1)$, $FC(512)\}$, followed by two output layers, one of size 1 (critic) and one of size 18 (actor π), with one output for each possible action. The entropy of the policy π was also added to the training loss as done in [16], with coefficient $\beta = 0.01$. The Exponential Linear Unit (ELU) activation [18] was used for every layer except for the output ones, that used a linear activation (critic) and a softmax activation

1st Transylvanian Machine Learning Summer School, 16-22 July 2018, Cluj-Napoca, Romania

(actor). Training was performed using the RMSProp optimizer [19] with $\eta = 0.0003$, decay 0.95, momentum 0.0, and $\epsilon = 0.01$. All gradients were clipped individually to $[-1, 1]$ and then globally (as suggested in [20]) with a clip norm of 50. Instantaneous rewards were also clipped to $[-1, 1]$. A discount factor $\lambda = 0.99$ was used.

3 Results

The results of the simulations are shown in Fig. 1. The games scores were converted to human-normalized scores as in [17]. Catastrophic forgetting was observed in all the simulations for both games on subsequent training of the other game. However, retraining on the first game was found to recover and improve the performance achieved after first training on it. This can be seen in the bottom row, where the learning progress on the first game is compared between its first and second training phases (first and last blocks of 10M frames). The results were averaged over 5 repetitions of each simulation. The performance of the agent was tested every 50000 frames as an average of the total score achieved on each game over 5 episodes of at most 10000 frames, without updating the weights of the network.

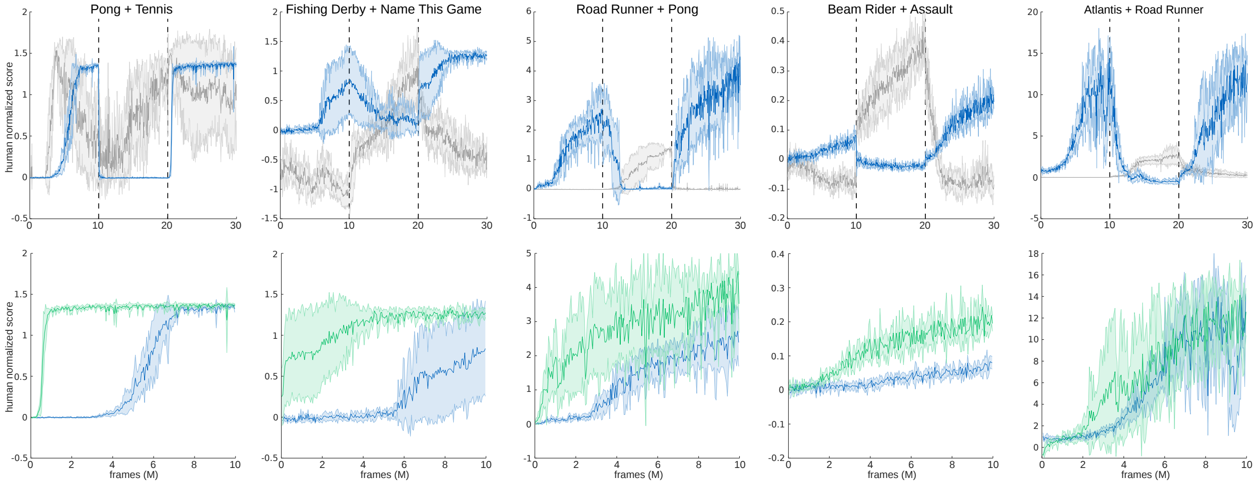

Figure 1: Plot of the human-normalized scores of a GA3C agent trained on pairs of games in alternation (10M frames on the first game, 10M frames on the second game, and 10M frames of retraining on the first game). Catastrophic forgetting is observed for both games on subsequent training of the other game. However, retraining on the first game is found to quickly recover the lost performance. The top row shows the score achieved by the agent on each game during training, tested every 50000 frames. The bottom row overlays the scores achieved on the first game at its first training phase (first 10M frames) and second retraining phase (last 10M frames).

4 Discussion & Conclusion

Catastrophic forgetting was found in all the game pairs tested, with performance on the first game being quickly disrupted by the successive training on the second game, on which performance was in turn disrupted by retraining on the first game. However, it was found that the actual amount of forgetting in the model was milder than that reflected on the decrease in performance. Indeed, retraining the agent on the first game could recover and improve the previous performance with a significantly smaller number of weight updates as originally necessary. Forgetting was thus highly disruptive in terms of the behavior of the agent, but the actual changes to its network's parameters were probably limited. This is line with the intuition that in very high-dimensional weight spaces it may not be difficult to find parameters optimal for a task that are close to the parameters that are optimal for other tasks, as suggested in [9].

It is further important to observe that the amount of interference was specific to the game pairs used. The difference may be explained by the known influence of the degree of overlap between the representations learnt by neural networks trained on different tasks, for which high overlap leads to more disruptive interference [21, 22, 11]. It is also interesting that training on `Pong` (first 10M frames, first column of Fig. 1) resulted in significant performance improvements on `Tennis`, even though that game had never been seen by the agent before.

Finally, the results we presented suggest that systems explicitly designed to mitigate the problem of catastrophic forgetting may still benefit significantly from some alternation in the exposure to the training tasks, as for example was recently done in [9], even with long switching times that would not otherwise work in traditional multi-task learning. These results further suggest that it may even be the case that simpler, approximate solutions to the problem may be able to achieve non-trivial multi-task performance on the Atari games benchmark.

Acknowledgments

The author sincerely thanks Microsoft for providing cloud computing resources via a Microsoft Azure for Research award (2017/2018).

References

[1] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," in *Psychology of learning and motivation*, vol. 24, pp. 109–165, Elsevier, 1989.

[2] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio, "An empirical investigation of catastrophic forgetting in gradient-based neural networks," *arXiv preprint arXiv:1312.6211*, 2013.

[3] A. Robins, "Catastrophic forgetting, rehearsal and pseudorehearsal," *Connection Science*, vol. 7, no. 2, pp. 123–146, 1995.

[4] R. M. French, "Catastrophic forgetting in connectionist networks," *Encyclopedia of cognitive science*, 1991.

[5] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell, "Progressive neural networks," *arXiv preprint arXiv:1606.04671*, 2016.

[6] C. Fernando, D. Banarse, C. Blundell, Y. Zwols, D. Ha, A. A. Rusu, A. Pritzel, and D. Wierstra, "Pathnet: Evolution channels gradient descent in super neural networks," *arXiv preprint arXiv:1701.08734*, 2017.

[7] S.-W. Lee, J.-H. Kim, J. Jun, J.-W. Ha, and B.-T. Zhang, "Overcoming catastrophic forgetting by incremental moment matching," in *Advances in Neural Information Processing Systems*, pp. 4655–4665, 2017.

[8] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, and T. Tuytelaars, "Memory aware synapses: Learning what (not) to forget," *arXiv preprint arXiv:1711.09601*, 2017.

[9] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[10] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *International Conference on Machine Learning*, pp. 3987–3995, 2017.

[11] R. M. French, "Catastrophic forgetting in connectionist networks," *Trends in cognitive sciences*, vol. 3, no. 4, pp. 128–135, 1999.

[12] R. Kemker, A. Abitino, M. McClure, and C. Kanan, "Measuring catastrophic forgetting in neural networks," *arXiv preprint arXiv:1708.02072*, 2017.

[13] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.

[14] M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, "The arcade learning environment: An evaluation platform for general agents.," *J. Artif. Intell. Res.(JAIR)*, vol. 47, pp. 253–279, 2013.

[15] M. Babaeizadeh, I. Frosio, S. Tyree, J. Clemons, and J. Kautz, "Reinforcement learning thorugh asynchronous advantage actor-critic on a gpu," in *ICLR*, 2017.

[16] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International Conference on Machine Learning*, pp. 1928–1937, 2016.

[17] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

[18] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[19] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[20] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International Conference on Machine Learning*, pp. 1310–1318, 2013.

[21] R. M. French, "Semi-distributed representations and catastrophic forgetting in connectionist networks," *Connection Science*, vol. 4, no. 3-4, pp. 365–377, 1992.

[22] M. K. Hetherington, "Catastrophic interference is eliminated in pretrained networks," in *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, pp. 723–728, 1993.