

Forgetting in neural networks just got less catastrophic

March 16 2017, by Nancy Owano

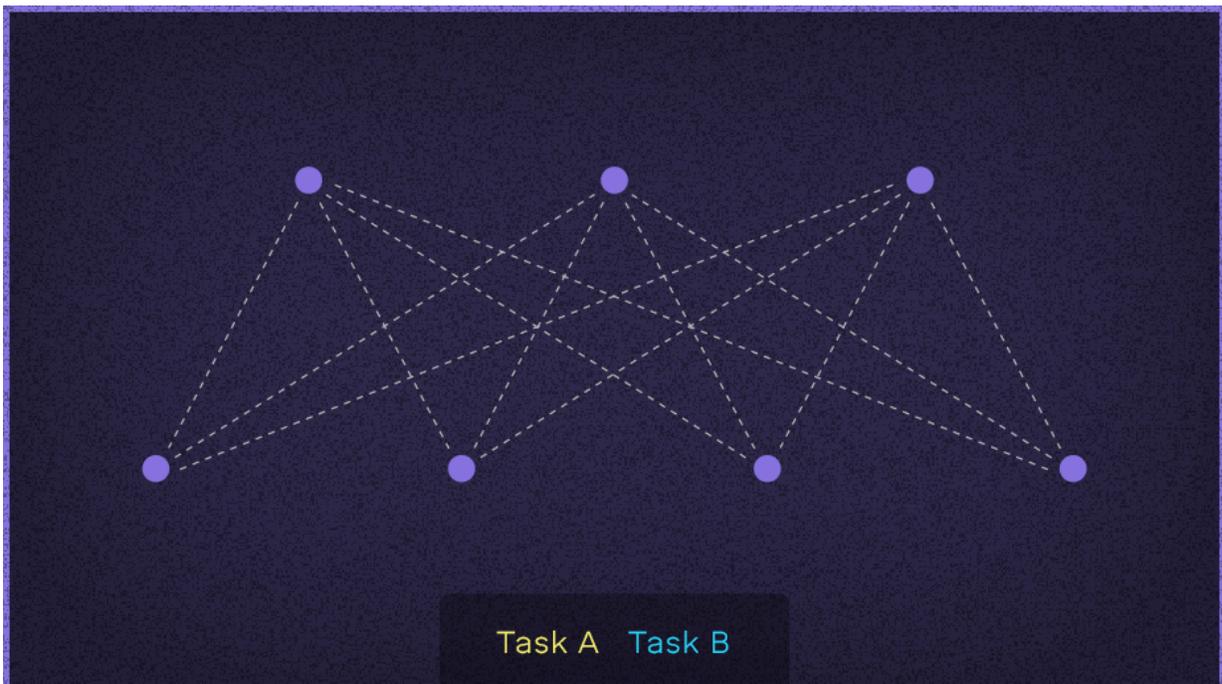


Illustration of the learning process for two tasks using EWC. Credit: DeepMind blog

(Tech Xplore)—How to add memory to AI: Follow the trail of DeepMind researchers, where reports say the AI system can learn to play one Atari game and then use the knowledge to learn another.

Matt Burgess in *Wired* and writers for other news sites sat up and took

notice of the DeepMind work, discussed in their paper and in a blog post on Monday from DeepMind.

Wait, DeepMind? Atari? We have heard something like this before. The computer program had already been reported to have played Atari games. But just don't expect it to remember how well it did.

The blog at DeepMind commented. "Computer programs that learn to perform tasks also typically forget them very quickly."

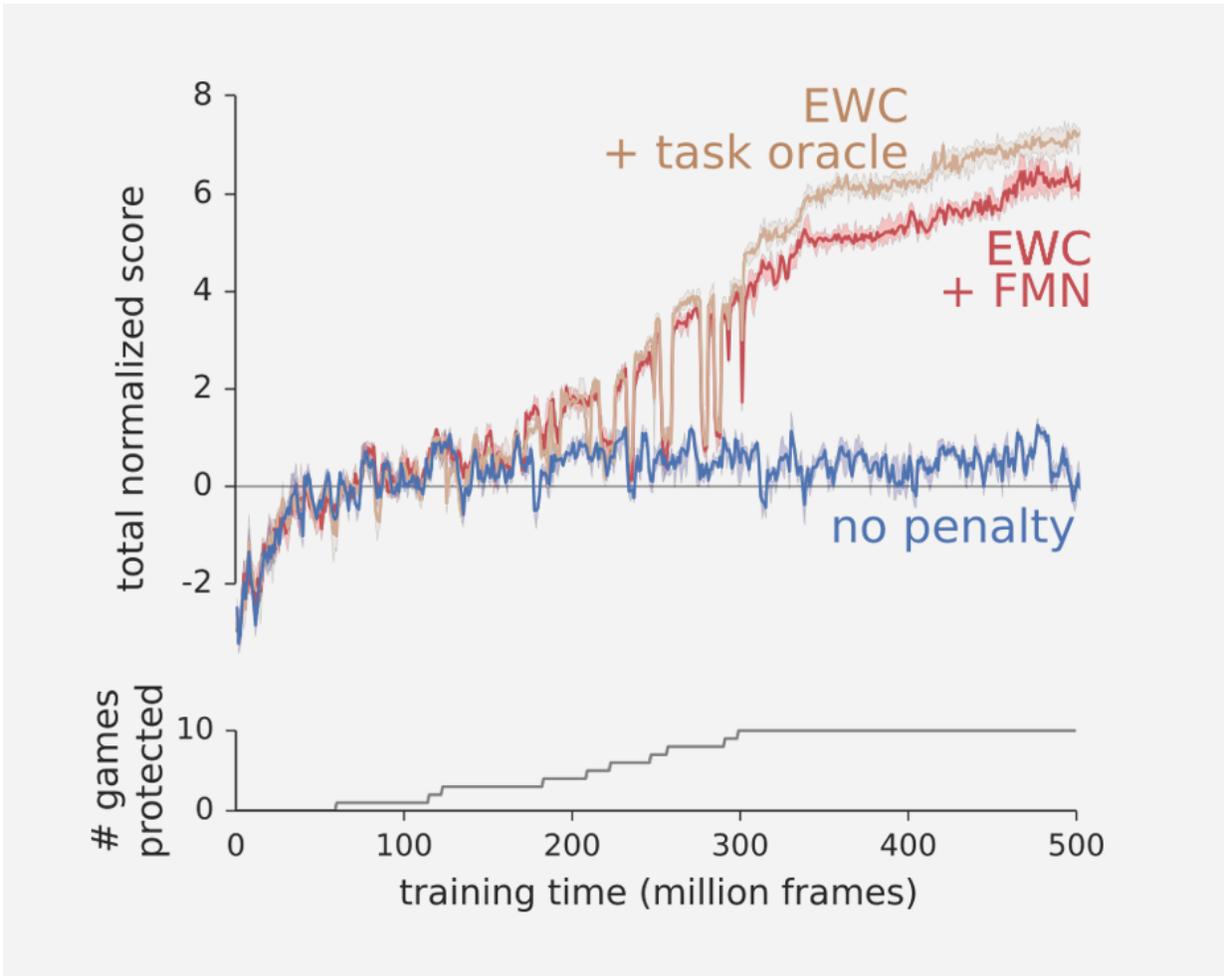
Newer work tells a different story. "Now," said Burgess, "a team of DeepMind and Imperial College London researchers have created an algorithm that allows its neural networks to learn, retain the information, and use it again."

Actually, as the DeepMind bloggers said, we are dealing with what has been a drawback presented by that wonder of wonders, the [neural network](#).

"As a network trains on a particular task its parameters are adapted to solve the task. When a new [task](#) is introduced, new adaptations overwrite the knowledge that the neural network had previously acquired. This phenomenon is known in cognitive science as 'catastrophic forgetting', and is considered one of the fundamental limitations of neural networks."

They showed in their work that a program can remember old tasks when learning a new one.

The work is presented in "Overcoming catastrophic forgetting in neural networks." It is in the *Proceedings of the National Academy of Sciences (PNAS)*. Author affiliations are DeepMind, London, and the Bioengineering Department, Imperial College London.



Credit: DeepMind blog

DeepMind's James Kirkpatrick, the lead author of its research paper, said in *Wired*: "Previously, we had a system that could learn to play any game, but it could only learn to play one game. Here we are demonstrating a system that can learn to play several games one after the other."

How does it work? As *Wired* explained, the deep neural networks, called

Deep Q-Network (DQN), previously used to conquer the Atari games, are this time "enhanced" with the EWC algorithm.

What does EWC stand for? Wait for this. Elastic Weight Consolidation.

Burgess said that, "Essentially, the [deep neural network](#) using the EWC algorithm was able to learn to play one game and then transfer what it had learnt to play a brand new game." The approach remembers old tasks by slowing down learning on certain weights based on how important they are to previously seen tasks.

However, the system is by no means perfect, said Burgess, in that "it isn't able to perform as well as a neural network that completes just one [game](#)."

Kirkpatrick said in *Wired*: "At the moment, we have demonstrated sequential learning but we haven't proved it is an improvement on the efficiency of learning. Our next steps are going to try and leverage sequential learning to try and improve on real-world learning."

Beyond Atari finesse, what is the bigger picture as regards their work? Consider this research effort as an important step towards more intelligent programs that are able to learn progressively and adaptively. They have shown that, as they blogged, catastrophic forgetting is not an insurmountable challenge for neural networks.

They are talking about enabling continual learning in neural networks, which is the title of the blog post.

What is more, the work might shed more light on the brain. In their paper, the authors wrote, "Whereas this paper has primarily focused on building an algorithm inspired by neurobiological observations and theories, it is also instructive to consider whether the algorithm's

successes can feed back into our understanding of the brain."

They said they saw considerable parallels between EWC and two computational theories of synaptic plasticity.

More information: — DeepMind blog:
deepmind.com/blog/enabling-con...-in-neural-networks/

— James Kirkpatrick et al. Overcoming catastrophic forgetting in neural networks, *Proceedings of the National Academy of Sciences* (2017). [DOI: 10.1073/pnas.1611835114](https://doi.org/10.1073/pnas.1611835114)

Abstract

The ability to learn tasks in a sequential fashion is crucial to the development of artificial intelligence. Until now neural networks have not been capable of this and it has been widely thought that catastrophic forgetting is an inevitable feature of connectionist models. We show that it is possible to overcome this limitation and train networks that can maintain expertise on tasks that they have not experienced for a long time. Our approach remembers old tasks by selectively slowing down learning on the weights important for those tasks. We demonstrate our approach is scalable and effective by solving a set of classification tasks based on a hand-written digit dataset and by learning several Atari 2600 games sequentially.

© 2017 Tech Xplore

Citation: Forgetting in neural networks just got less catastrophic (2017, March 16) retrieved 16 April 2024 from <https://techxplore.com/news/2017-03-neural-networks-catastrophic.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is

provided for information purposes only.