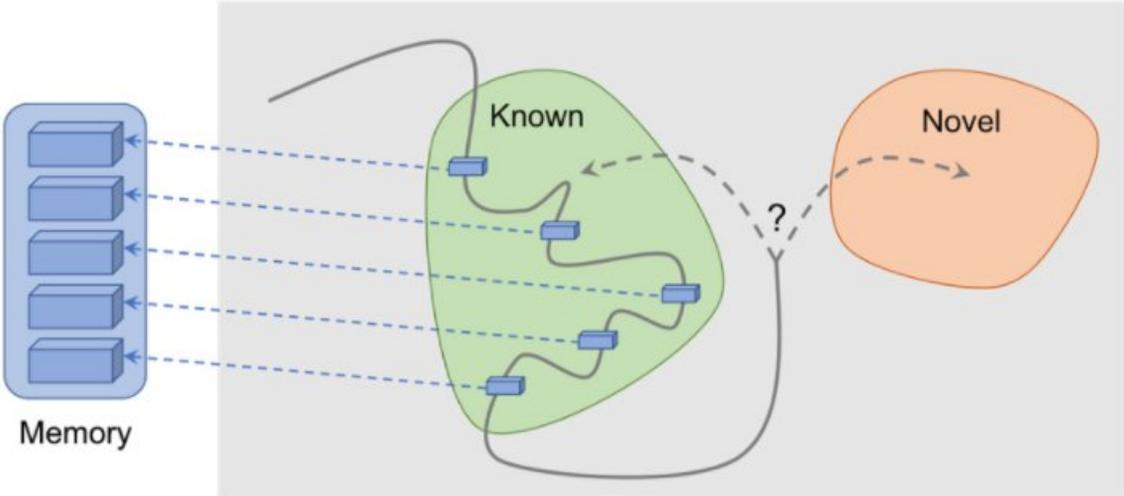


# A new method to instill curiosity in reinforcement learning agents

October 17 2018, by Ingrid Fadelli



How the method works: observations are added to memory, reward is computed based on how far our current observation is from the most similar observation in memory. The agent receives more reward for seeing observations which are not represented in memory yet. Credit: Savinov et al.

Several real-world tasks have sparse rewards and this poses challenges for the development of reinforcement learning (RL) algorithms. A solution to this problem is to allow an agent to autonomously create a reward for itself, making rewards denser and more suitable for learning.

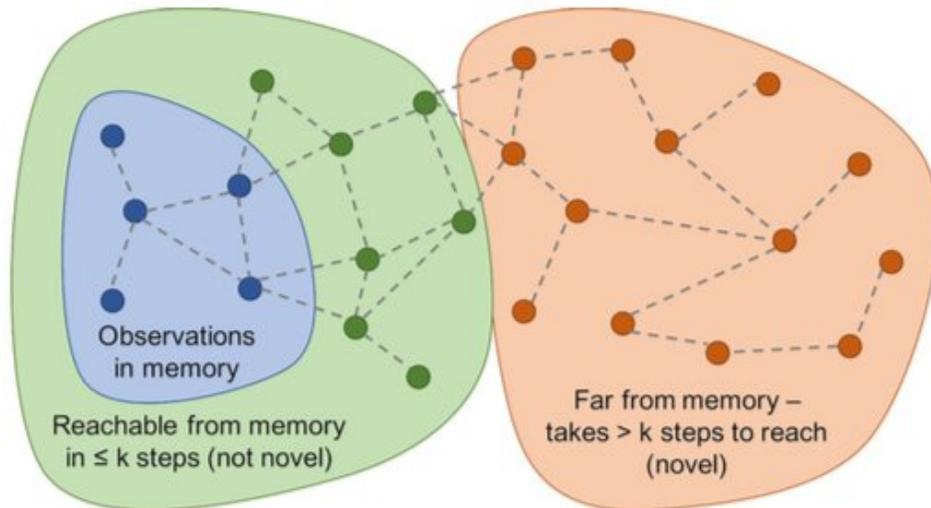
For instance, inspired by the curious behaviour with which animals

explore their environment, an RL algorithm's observation of something new could be rewarded with a bonus. This bonus, summed up with the real task reward, would then allow RL algorithms to learn from a combined reward.

Researchers at DeepMind, Google Brain and ETH Zurich have recently devised a new curiosity method that uses episodic memory to form this novelty bonus. This bonus is determined by comparing current observations and observations stored in memory.

"The main objective of our work was to investigate new memory-based ways of imbuing reinforcement learning (RL) agents with 'curiosity,' by which we mean a drive to explore the environment even in the complete absence of rewards," Tim Lillicrap at DeepMind and Nikolay Savinov at Google Brain told TechXplore in an e-mail. "Curiosity has been approached in various ways by the research community, but we felt that several ideas could benefit from further exploration."

The key ideas explored in this recent paper are based on a previous [study carried out by Savinov](#), which proposed a new memory architecture inspired by mammalian navigation. This architecture allows agents to repeat a route through an environment using only a visual walkthrough. The new method developed by the researchers takes this one step further, trying to achieve good exploration driven by curiosity.



Graph of reachabilities would determine novelty. In practice, this graph is not available — so a neural network approximator is trained to estimate a number of steps between observations. Credit: Savinov et al.

"While acting, the agent stores instances of observation representations in its episodic memory," Lillicrap and Savinov said. "To determine if the current observation is novel or not, it is compared to those in memory. If nothing similar is found, the current observation is deemed novel and the agent is rewarded, otherwise it gets a negative reward. This encourages the agent to explore unfamiliar territory, akin to being curious."

The researchers found that comparing pairs of observations could be tricky, as checking for an exact match is ultimately meaningless in realistic environments. This is because in real-world situations, an agent rarely observes the same thing twice.

"Instead, we trained a neural network to predict if the agent can reach the current observation from those in memory by taking fewer actions

than a fixed threshold; say, five actions," Lillicrap and Savinov explained. "Observations within those five actions are considered similar, while those requiring more actions to make a transition are considered dissimilar."

Lillicrap, Savinov and their colleagues tested their approach in VizDoom and DMLab, two visually rich 3D environments. In VizDoom, the agent learned to successfully navigate to a distant goal at least two times faster than state-of-the-art curiosity method ICM. In DMLab, the algorithm generalized well to new, procedurally generated levels of the game, reaching its desired goal at least two times more frequently than ICM on test mazes with very sparse rewards.



Surprise-based method (ICM) is persistently tagging walls with a laser-like science fiction gadget instead of exploring the maze. This behaviour is similar to the channel switching described before: even though the result of tagging is theoretically predictable, it is not easy and apparently requires a deep knowledge of physics which is not straightforward to acquire for a general agent. Credit: Savinov et al.

"We noticed an interesting drawback in one of the most popular methods to imbue agents with curiosity," Lillicrap and Savinov said. "We found that this method, based on the surprise that is computed by a slowly changing model that tries to predict what will happen next, can result in

an instant gratification response from the agent: instead of solving the task at hand, it will exploit actions which lead to unpredictable consequences in order to get immediate [reward](#)."

This peculiar occurrence, also known as "couch-potato" issues, entails an agent finding ways to instantly gratify itself by exploiting actions that lead to unpredictable consequences. For instance, when given a TV remote, the agent might do nothing other than change channels, even if its original task was entirely different, such as searching for a goal in a maze.

"This shortcoming can be alleviated using episodic memory together with a reasonable measure of observation similarity, which is our contribution," Lillicrap and Savinov said. "This opens up a way to more intelligent exploration."



Our method shows a reasonable exploration. Credit: Savinov et al.

The new curiosity method devised by Lillicrap, Savinov, and their colleagues could help to replicate curiosity-like skills in RL algorithms, allowing them to autonomously create rewards for themselves. In the future, the researchers would like to use [episodic memory](#) not only for granting rewards, but also for planning actions.

"For example, can content retrieved from [memory](#) be used to think about where to go next?" Lillicrap and Savinov said. "This is currently a big scientific challenge: if solved, agents would be able to quickly adapt exploration strategies to new environments, allowing learning to happen at a much faster rate."

**More information:** Episodic curiosity through reachability.  
arXiv:1810.02274v1 [cs.LG]. [arxiv.org/abs/1810.02274](https://arxiv.org/abs/1810.02274)

[sites.google.com/view/episodic-curiosity](https://sites.google.com/view/episodic-curiosity)

© 2018 Tech Xplore

Citation: A new method to instill curiosity in reinforcement learning agents (2018, October 17)  
retrieved 26 April 2024 from  
<https://techxplore.com/news/2018-10-method-instill-curiosity-agents.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.