Overview of t-DGR: A Trajectory-Based Deep Generative Replay Method for Continual Learning in Decision Making

Extended Abstract

William Yue
The University of Texas at Austin
Austin, TX, United States
william.yue@utexas.edu

Bo Liu
The University of Texas at Austin
Austin, TX, United States
bliu@cs.utexas.edu

Peter Stone
The University of Texas at Austin,
Sony AI
Austin, TX, United States
pstone@cs.utexas.edu

ABSTRACT

Deep generative replay has emerged as a promising approach for continual learning in decision-making tasks. This approach addresses the problem of catastrophic forgetting by leveraging the generation of trajectories from previously encountered tasks to augment the current dataset. However, existing deep generative replay methods for continual learning rely on autoregressive models, which suffer from compounding errors in the generated trajectories. In this extended abstract, we summarize a simple, scalable, and non-autoregressive method for continual learning in decision-making tasks using a generative model that generates task samples conditioned on the trajectory timestep. We evaluate our method on Continual World benchmarks and find that our approach achieves state-of-the-art performance on the average success rate metric among continual learning methods. Code and a preprint of a complete paper with full details are available at https://github.com/WilliamYue37/t-DGR.

KEYWORDS

Lifelong Learning; Continual Learning; Decision-Making; Imitation Learning; Machine Learning

ACM Reference Format:
William Yue, Bo Liu, and Peter Stone. 2024. Overview of t-DGR: A Trajectory-Based Deep Generative Replay Method for Continual Learning in Decision Making: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Continual learning, also known as lifelong learning, is a critical challenge in the advancement of general artificial intelligence, as it enables models to learn from a continuous stream of data encompassing various tasks, rather than having access to all data at once [9]. However, a major challenge in continual learning is the phenomenon of catastrophic forgetting, where previously learned skills are lost when attempting to learn new tasks [7].

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

To mitigate catastrophic forgetting, replay methods have been proposed, which involve saving data from previous tasks and replaying it to the learner during the learning of future tasks. This approach mimics how humans actively prevent forgetting by reviewing material for tests and replaying memories in dreams. However, storing data from previous tasks requires significant storage space and becomes computationally infeasible as the number of tasks increases.

In the field of cognitive neuroscience, the Complementary Learning Systems theory offers insights into how the human brain manages memory. This theory suggests that the brain employs two complementary learning systems: a fast-learning episodic system and a slow-learning semantic system [3, 5, 6]. The hippocampus serves as the episodic system, responsible for storing specific memories of unique events, while the neocortex functions as the semantic system, extracting general knowledge from episodic memories and organizing it into abstract representations [8].

Drawing inspiration from the human brain, deep generative replay (DGR) addresses the catastrophic forgetting issue in decision-making tasks by using a generative model as the hippocampus to generate trajectories from past tasks and replay them to the learner which acts as the neocortex [12]. The time-series nature of trajectories in decision-making tasks sets it apart from continual supervised learning, as each timestep of the trajectory requires sufficient replay. In supervised learning, the learner's performance is not significantly affected if it performs poorly on a small subset of the data. However, in decision-making tasks, poor performance on any part of the trajectory can severely impact the overall performance. Therefore, it is crucial to generate state-action pairs that accurately represent the distribution found in trajectories. Furthermore, the high-dimensional distribution space of trajectories makes it computationally infeasible to generate complete trajectories all at once.

Existing DGR methods adopt either the generation of individual state observations i.i.d. without considering the temporal nature of trajectories or autoregressive trajectory generation. Autoregressive approaches generate the next state(s) in a trajectory by modeling the conditional probability of the next state(s) given the previously generated state(s). However, autoregressive methods suffer from compounding errors in the generated trajectories. On the other hand, generating individual state observations i.i.d. leads to a higher sample complexity compared to generating entire trajectories, which becomes significant when replay time is limited.

To address the issues in current DGR methods, we propose a simple, scalable, and non-autoregressive trajectory-based DGR method.

We define a generated trajectory as temporally coherent if the transitions from one state to the next appear realistic. Given that current decision-making methods are trained on state-action pairs, we do not require trajectories to exhibit temporal coherence. Instead, our focus is on ensuring an equal number of samples generated at each timestep of the trajectory to accurately represent the distribution found in trajectories. To achieve equal sample coverage at each timestep, we train our generator to produce state observations conditioned on the trajectory timestep, and then sample from the generator conditioned on each timestep of the trajectory.

To evaluate the effectiveness of our proposed method, t-DGR, we conducted experiments on the Continual World benchmarks CW10 and CW20 [13] using imitation learning. Our results indicate that t-DGR achieves state-of-the-art performance in terms of average success rate when compared to other top continual learning methods.

2 METHOD OVERVIEW

Our proposed method, t-DGR, tackles the challenge of generating long trajectories by training a generator which is conditioned on the trajectory timestep to generate state observations. At each timestep, t-DGR generates the j-th state observation of the trajectory using the previous generator conditioned on timestep j, and then labels it with an action using the previous policy. After generating all timesteps in a trajectory, t-DGR adds all labeled state-action pairs in the trajectory to the existing training dataset. Unlike generating state observations i.i.d., our method ensures equal coverage of every timestep during the generative process, significantly reducing sample complexity. Once t-DGR has augmented the training samples from the environment with our generated training samples, t-DGR uses backpropagation to update both the generator and learner using the augmented dataset. The t-DGR algorithm continues this process of generative replay throughout the agent's lifetime, which can be infinite.

We employ a U-net [10] trained with the diffusion loss to implement the generative diffusion model. Given our use of proprioceptive observations in the experiments, we implement the policy using a multi-layer perceptron trained with the mean squared error loss. Full details of t-DGR, including code, are available at https://github.com/WilliamYue37/t-DGR.

3 EXPERIMENTS

We evaluate our method on the Continual World benchmarks CW10 and CW20 [13], along with our own "General Continual Learning" variant of CW10 called GCL10. We compare our method against several baselines: Finetune, Multitask, oEWC [2, 11], PackNet [4], DGR [12], and CRIL [1]. All baselines are evaluated on three metrics: average success rate, average forward transfer, and average forgetting.

In our findings, t-DGR emerges as the leading method, demonstrating the highest success rate on CW10, CW20 (Table 1), and GCL10. Notably, PackNet's performance on the second iteration of tasks in CW20 diminishes, highlighting its limited capacity for continually accommodating new tasks. This limitation underscores the fact that PackNet falls short of being a true lifelong learner, as it necessitates prior knowledge of the task count for appropriate parameter capacity allocation. On the contrary, pseudo-rehearsal methods, such as t-DGR, exhibit improved performance with the second iteration of tasks in CW20 due to an increased replay time. These findings emphasize the ability of DGR methods to effectively leverage past knowledge, as evidenced by their superior forward transfer in both CW10 and CW20.

Table 1: Continual World 20 Benchmark

Method	Success Rate ↑	FT↑	Forgetting↓
Finetune	14.2 ±4.0	-0.5 ±3.0	82.2 ±5.6
Multitask	97.0 ±1.0	N/A	N/A
oEWC	19.4 ±5.3	-2.8 ±4.1	75.2 ±7.5
PackNet	74.1 ±4.1	-20.4 ±3.4	**-0.2** ±0.9
DGR	74.1 ±4.1	18.9 ±2.9	23.3 ±3.3
CRIL	50.8 ±4.4	4.4 ±4.9	46.1 ±5.4
t-DGR	**83.9** ±3.0	**30.6** ±4.5	14.6 ±2.9

GCL10 demonstrates that pseudo-rehearsal methods are mostly unaffected by blurry task boundaries, whereas PackNet's success rate experiences a significant drop-off. This discrepancy arises from the fact that PackNet's regularization technique does not work effectively with less clearly defined task boundaries.

Additionally, the diminishing performance gap between DGR and t-DGR as the replay ratio increases in our experiments indicates that a higher replay ratio reduces the likelihood of any portion of the trajectory being insufficiently covered when sampling individual state observations i.i.d., thereby contributing to improved performance. This trend supports the theoretical sample complexity of DGR, as $\Theta(n \log n + mn \log \log n)$ closely approximates the sample complexity of t-DGR, $\Theta(mn)$, when the replay amount $m \to \infty$. Here, n denotes the trajectory length. However, while DGR can achieve comparable performance to t-DGR with a high replay ratio, the availability of extensive replay time is often limited in many real-world applications.

Overall, t-DGR exhibits promising results, outperforming other methods in terms of success rate in all evaluations. Notably, t-DGR achieves a significant improvement over existing pseudo-rehearsal methods on CW20 using a Welch t-test with a significance level of $\alpha = 0.005$. Its ability to handle blurry task boundaries, leverage past knowledge, and make the most of replay opportunities position it as a state-of-the-art method for continual lifelong learning in decision-making.

ACKNOWLEDGMENTS

This work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (FAIN-2019844, NRT-2125858), ONR (N00014-18-2243), ARO (E2061621), Bosch, Lockheed Martin, and UT Austin's Good Systems grand challenge. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

REFERENCES

[1] Chongkai Gao, Haichuan Gao, Shangqi Guo, Tianren Zhang, and Feng Chen. 2021. CRIL: Continual Robot Imitation Learning via Generative and Prediction Model. arXiv:2106.09422 [cs.RO]

[2] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences* 114, 13 (mar 2017), 3521–3526. https://doi.org/10.1073/pnas.1611835114

[3] Dharshan Kumaran, Demis Hassabis, and James L. McClelland. 2016. What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences* 20, 7 (2016), 512–534. https://doi.org/10.1016/j.tics.2016.05.004

[4] Arun Mallya and Svetlana Lazebnik. 2018. PackNet: Adding Multiple Tasks to a Single Network by Iterative Pruning. arXiv:1711.05769 [cs.CV]

[5] James Mcclelland, Bruce Mcnaughton, and Randall O'Reilly. 1995. Why There are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological review* 102 (08 1995), 419–57. https://doi.org/10.1037/0033-295X.102.3.419

[6] James L. McClelland, Bruce L. McNaughton, and Randall C. O'Reilly. 1995. Complementary Learning Systems within the Hippocampus: A Neural Network Modeling Approach to Understanding Episodic Memory Consolidation. *Psychological Review* 102, 3 (1995), 419–457.

[7] Michael McCloskey and Neal J. Cohen. 1989. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. Psychology of Learning and Motivation, Vol. 24. Academic Press, 109–165. https://doi.org/10.1016/S0079-7421(08)60536-8

[8] Randall C. O'Reilly and Kenneth A. Norman. 2002. Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences* 6, 12 (1 Dec. 2002), 505–510. https://doi.org/10.1016/S1364-6613(02)02005-3

[9] Mark Ring. 1994. *Continual Learning in Reinforcement Environments.* Ph.D. Dissertation. University of Texas at Austin. https://www.cs.utexas.edu/~ring/Ring-dissertation.pdf

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597 [cs.CV]

[11] Jonathan Schwarz, Jelena Luketina, Wojciech M. Czarnecki, Agnieszka Grabska-Barwinska, Yee Whye Teh, Razvan Pascanu, and Raia Hadsell. 2018. Progress Compress: A scalable framework for continual learning. arXiv:1805.06370 [stat.ML]

[12] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. 2017. Continual Learning with Deep Generative Replay. arXiv:1705.08690 [cs.AI]

[13] Maciej Wołczyk, Michał Zając, Razvan Pascanu, Łukasz Kuciński, and Piotr Miłoś. 2021. Continual World: A Robotic Benchmark For Continual Reinforcement Learning. arXiv:2105.10919 [cs.LG]