Neural Episodic Control (NEC)

Pritzel et al. (2017) | https://arxiv.org/abs/1703.01988

David Bayha, Dhruv Dhamani, Geet Saurabh Kunumala, Yingjie Ma

2025-02-05

College of Computing and Informatics
University of North Carolina at Charlotte

Outline

1. Introduction 1
   1.1 Can AI learn as fast as Humans? . 2
   1.2 Problem 3
   1.3 Problem Statement 4
   1.4 Motivation 5
2. Background 6
   2.1 DQN and Reinforcement Learning 7
   2.2 Improvements on DQN 8
   2.3 Neural Episodic Control. 9
   2.4 Differentiable Neural Dictionary 10
3. Methods 11
   3.1 Overview of Neural Episodic Control 12
   3.2 Differentiable Neural Dictionary (DND) 13
   3.3 N-step Updates & Training Loop 14
   3.4 Experimental Setup 15
4. Results 16
   4.1 Median Human-Normalized Score... 17

UNIVERSITY OF NORTH CAROLINA CHARLOTTE

1.1 Can AI learn as fast as Humans?

- **Statistic**: In the Atari 2600 set of environments, deep Q-networks required **more than 200 hours** of gameplay in order to achieve scores similar to a human who only played 2 hours. (Bellemare et al., 2013)
- Imagine a human only needing one example shown to learn something, but an AI needs to see that same example millions of times.

DL and Episodic vs. Semantic Memory (1:20-2:40): Watch on YouTube

1.2 Problem

- **What is the Problem?**

Deep Reinforcement Learning (DRL) can exceed human performance, but it requires significantly more interactions to learn, making the process **slow** and **inefficient**.

- **What is the Solution?**

Neural Episodic Control (NEC) is a technique that allows a learning agent to learn faster by remembering past experiences.

- NEC stores and **recalls past successful actions** instead of learning purely through trial and error.
- Inspired by the hypothesized role of the hippocampus in decision-making.

1.3 Problem Statement

- **What is the purpose of the paper?**

Provides an answer to the question:
Why are Deep Reinforcement-Learning agents so slow at learning?

- Seeks to address three major challenges:
  1. **Stochastic Gradient Descent (SGD)** optimization requires using small learning rates – large learning rates cause catastrophic interference (forgetting).
  2. Environments with a **sparse reward signal** make it difficult for a RL agent to learn its environment – there may be very few instances with a non-zero reward.
  3. Many DRL agents using value-based RL methods (ie. Q-learning) learn one step at a time, resulting in **slow reward signal propagation** – *agent may take hundreds of steps before retrieving useful information*

1.4 Motivation

- In order for DRL techniques to be applicable to real-world problems it is essential that faster learning occurs
- **Neural Episodic Control** (NEC) dramatically improves the *efficiency* of RL agents by *storing/recalling successful past experiences* and reducing trial-and-error learning
- Some potential real-world applications include:
  - **Robotics**: Robots require fast learning of their environment
  - **Healthcare**: optimize health care decisions and more efficiently personalize health-care plans
  - **Autonomous vehicles** commonly in high-speed environments

Outline

1. Introduction 1
    1.1 Can AI learn as fast as Humans? . 2
    1.2 Problem 3
    1.3 Problem Statement 4
    1.4 Motivation 5
2. Background 6
    2.1 DQN and Reinforcement Learning 7
    2.2 Improvements on DQN 8
    2.3 Neural Episodic Control. 9
    2.4 Differentiable Neural Dictionary 10
3. Methods 11
    3.1 Overview of Neural Episodic Control 12
    3.2 Differentiable Neural Dictionary (DND) 13
    3.3 N-step Updates & Training Loop 14
    3.4 Experimental Setup 15
4. Results 16
    4.1 Median Human-Normalized Score... 17

UNIVERSITY OF NORTH CAROLINA CHARLOTTE

2.1 DQN and Reinforcement Learning

- Reinforcement Learning is a framework for learning optimal actions through interactions with environment to maximize reward.
- DQN uses Q-learning to learn value function $Q(s_t, a_t)$
- $Q(s_t, a_t)$ takes 2D pixel representation of state st and outputs vector containing value of each action at that state.
- Upon observation, DQN stores $(s_t, a_t, r_t, s_t + 1)$ tuple in replay buffer, which is used for training.

2.2 Improvements on DQN

- Double DQN decouples action selection and action evaluation steps to reduce overestimation bias.
- Prioritized Replay further improves on Double DQN by optimizing replay strategy.
- Many papers have suggested that switching to on-policy learning allows agent to learn faster in Atari environments
- AC3 works on policy gradient, which learns a policy and its associated value function.

2.3 Neural Episodic Control.

- NEC rapidly latches onto successful strategies as soon as they are experienced, instead of waiting many steps.
- The Agent has 3 components:
  1. Convoluted Neural Network - Processes pixel images.
  2. Set of memory modes - One per action.
  3. Final Network - Convert action memories into $Q(s, a)$ values.
- For each action, NEC has a memory module with key-value pairs called differentiable neural dictionary (DND)

2.4 Differentiable Neural Dictionary

- DND has 2 operations, lookup and write
- The output of lookup is a weighted sum of values in memory, whose weights are given by normalized kernels between lookup key and corresponding key in memory.
- After DND is queried, new key-value pair is written into memory. Writes are append-only.

Outline

1. Introduction 1
1.1 Can AI learn as fast as Humans? . 2
1.2 Problem 3
1.3 Problem Statement 4
1.4 Motivation 5
2. Background 6
2.1 DQN and Reinforcement Learning 7
2.2 Improvements on DQN 8
2.3 Neural Episodic Control. 9
2.4 Differentiable Neural Dictionary 10
3. Methods 11
3.1 Overview of Neural Episodic Control 12
3.2 Differentiable Neural Dictionary (DND) 13
3.3 N-step Updates & Training Loop 14
3.4 Experimental Setup 15
4. Results 16
4.1 Median Human-Normalized Score... 17

3.1 Overview of Neural Episodic Control

- *Three Main Components:*
  1. **CNN Embedding Network** for state representation
  2. **Memory Module (DND)**: one per action
  3. **Final Network** to combine memory outputs into $Q(s, a)$
- *Key Idea:* Store (key, value) pairs in a large external memory
  - **Keys** = slow-changing embeddings from CNN
  - **Values** = fast-updated action-value estimates
- *Motivation:* "Episodic memory" allows rapid assimilation of new experiences

3.2 Differentiable Neural Dictionary (DND)

- *Differentiable Neural Dictionary* = Key-Value Store
  - **Lookup**: find nearest keys to current embedding h
  - **Weighted Sum**: output value is a kernel-weighted average of stored values
- **Memory Growth**: append-only writes; update existing entries if the key already exists
- **Efficient Retrieval**: approximate nearest neighbor search (e.g., **kd-trees**) allows large-scale memory

3.3 N-step Updates & Training Loop

- **N-step** Q-learning for **faster reward propagation**:

$$Q^N(s_t, a_t) = \sum_{j=0}^{N-1} \gamma^j r_{t+j} + \gamma^N \max_{a'} Q(s_{t+N}, a')$$

- **Memory Update:**

$$Q_i \leftarrow Q_i + \alpha(Q^N - Q_i)$$

- **Replay Buffer**: small buffer to train the CNN embedding; slow gradient updates to refine representation.

3.4 Experimental Setup

- 57 **Atari 2600** games (Arcade Learning Environment)
- Training from **1M to 40M** frams of gameplay

$$\text{Human-Normalized Score (HNS)} = \frac{\text{score}_{\text{agent}} - \text{score}_{\text{random}}}{\text{score}_{\text{human}} - \text{score}_{\text{random}}}$$

- Recorded performance at specific checkpoints: **1M, 2M, 4M, 10M, 20M**, and **40M frames**
- Compared with: **DQN, Double DQN, Prioritized Replay, A3C, MFEC**
- NEC uses same CNN architecture as DQN for fair comparison

Outline

1. Introduction 1
   1.1 Can AI learn as fast as Humans? . 2
   1.2 Problem 3
   1.3 Problem Statement 4
   1.4 Motivation 5
2. Background 6
   2.1 DQN and Reinforcement Learning 7
   2.2 Improvements on DQN 8
   2.3 Neural Episodic Control. 9
   2.4 Differentiable Neural Dictionary 10
3. Methods 11
   3.1 Overview of Neural Episodic Control 12
   3.2 Differentiable Neural Dictionary (DND) 13
   3.3 N-step Updates & Training Loop 14
   3.4 Experimental Setup 15
4. Results 16
   4.1 Median Human-Normalized Score... 17

4.1 Median Human-Normalized Score

- NEC uses same CNN architecture as DQN for fair comparison