LEARNING SUCCESSOR FEATURES WITH DISTRIBUTED HEBBIAN TEMPORAL MEMORY

BACKGROUND: There is little research on temporal memories that can be used in fully online adaptive agents interacting with partially observable changing environments. SOTA methods for sequence modelling, like those of the recurrent neural network family and transformer-based models, are notoriously difficult to train on limited non-stationary data. Drawing inspiration from neurophysiological models of human brain function, we address this problem by proposing DHTM: an ensemble of simple, but easily trainable with local Hebbian-like rules sequence memories based on the Hidden Markov Model.

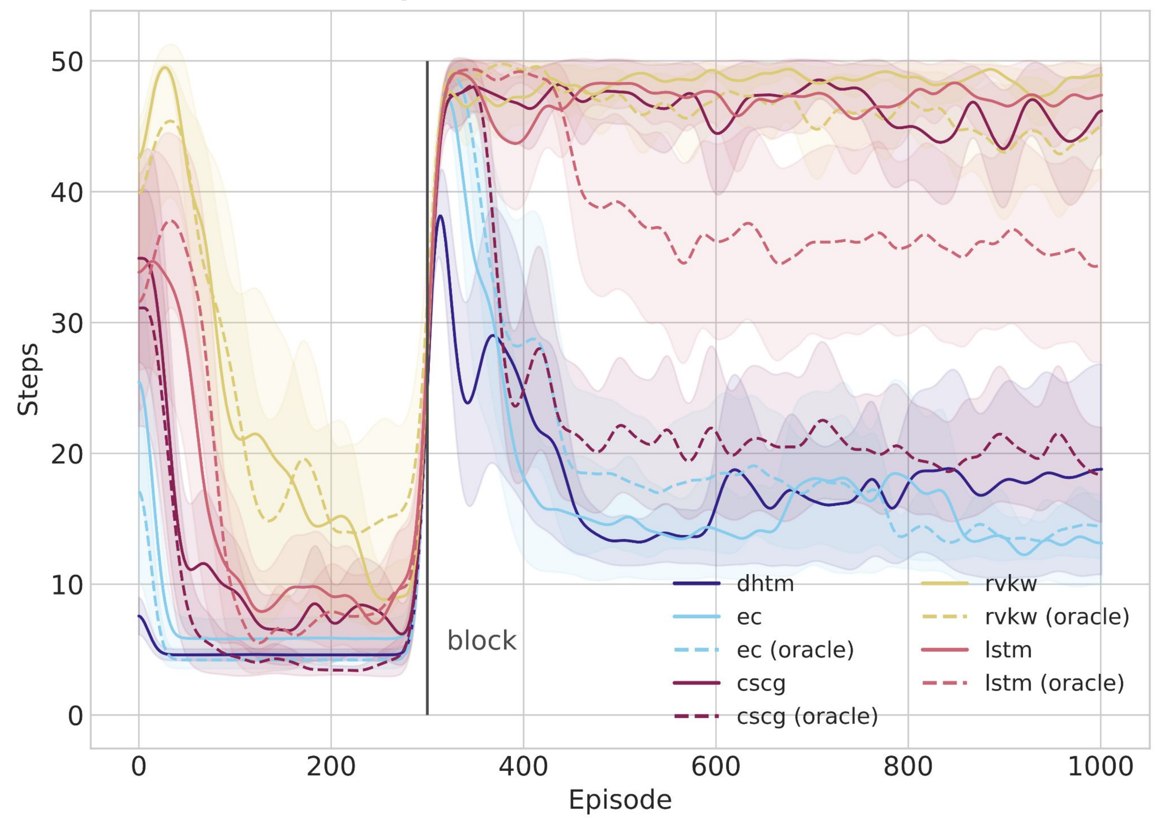
DHTM OVERVIEW

- Inference on Markov chains using a modified version of the sum-product algorithm taking into account chains' interdependencies.
- Monte Carlo sampling on forward pass and Hebbian-like updates to learn transition weights (no backward messages, **no backprop**).
- Sparse neural implementation using notion of dendritic segments.

EXPERIMENTS

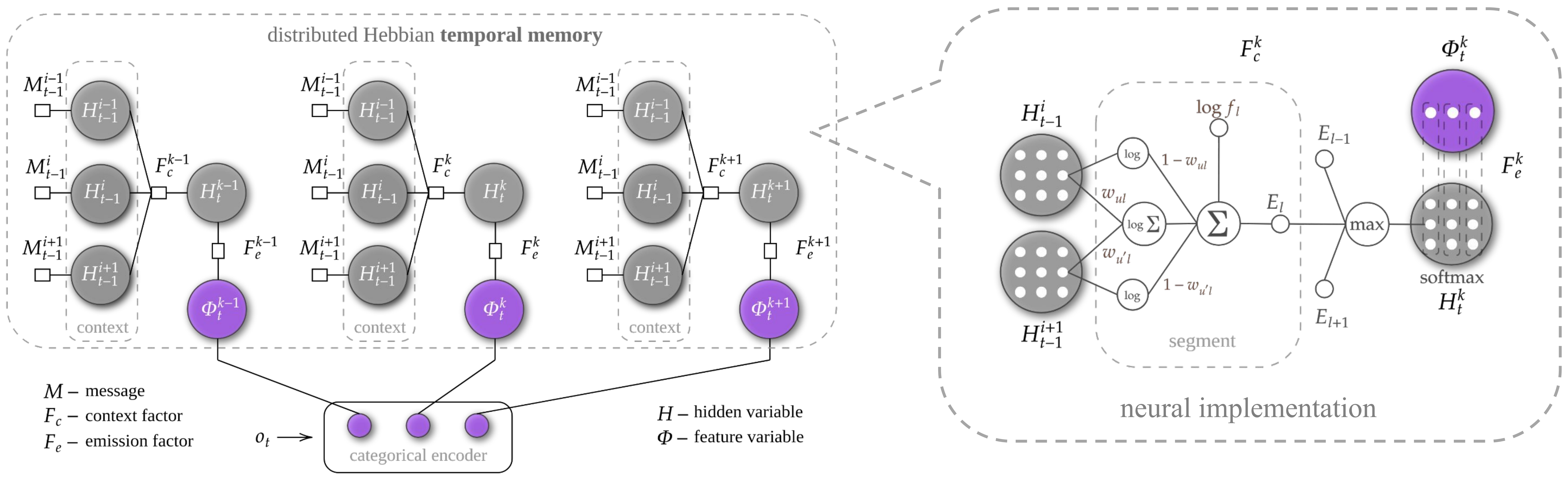
1. Use different temporal memories (DHTM, LSTM, RWKV, CSCG) to form successor features, which are combined with the learnt reward function to form the agent's policy in gridworld with one goal.
2. Block the shortest path to the goal at the 300th episode with and without an oracle that tells when the change happened.

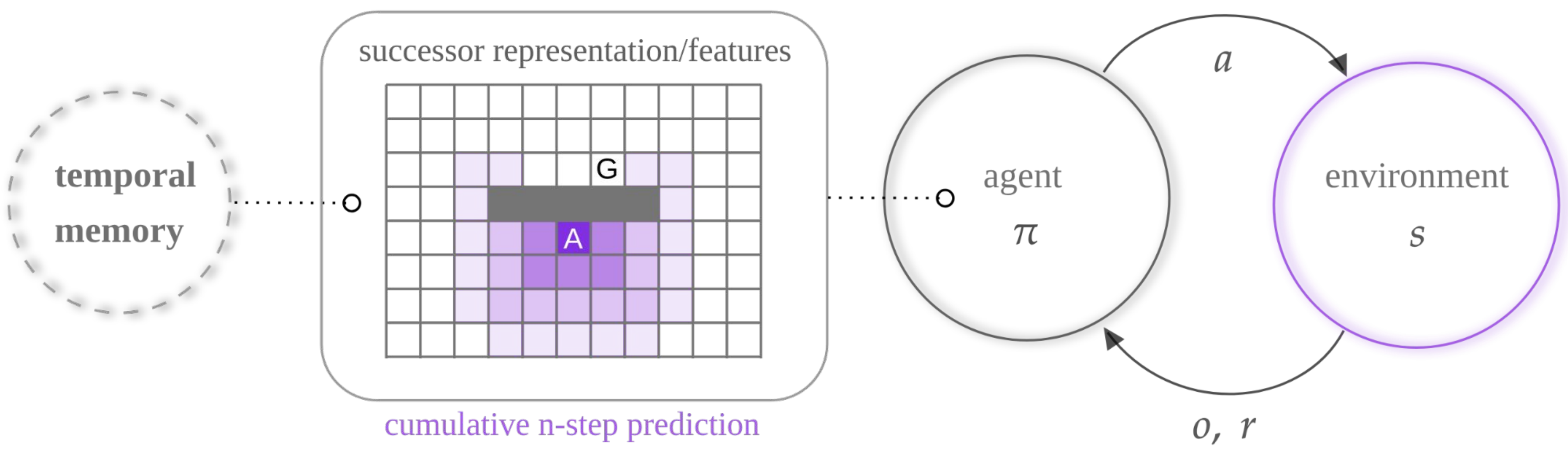
Steps to Goal: 5x5 Gridworld



Ensemble of simple **sequence models** is better than one good model for **online learning** in **changing environments**

full paper link





Evgenii Dzhivelikian,[1,2] Petr Kuderov,[1,2] Aleksandr Panov[1,2]

[1] Moscow Institute of Physics and Technology
Moscow, Russia

[2] AIRI Institute
Moscow, Russia

EPISODIC CONTROL (EC) BASELINE

Algorithm 1 Episodic memory learning

Input: φ_{t+1}, a_t

$h_{t+1} \leftarrow \mathcal{D}_{a_t}(h_t)$

$\varphi^*_{t+1} \leftarrow \underset{\varphi}{\operatorname{argmax}}\, F_e(h_{t+1}, \varphi)$

if h_{t+1} is null **or** φ^*_{t+1} is not φ_{t+1} **then**

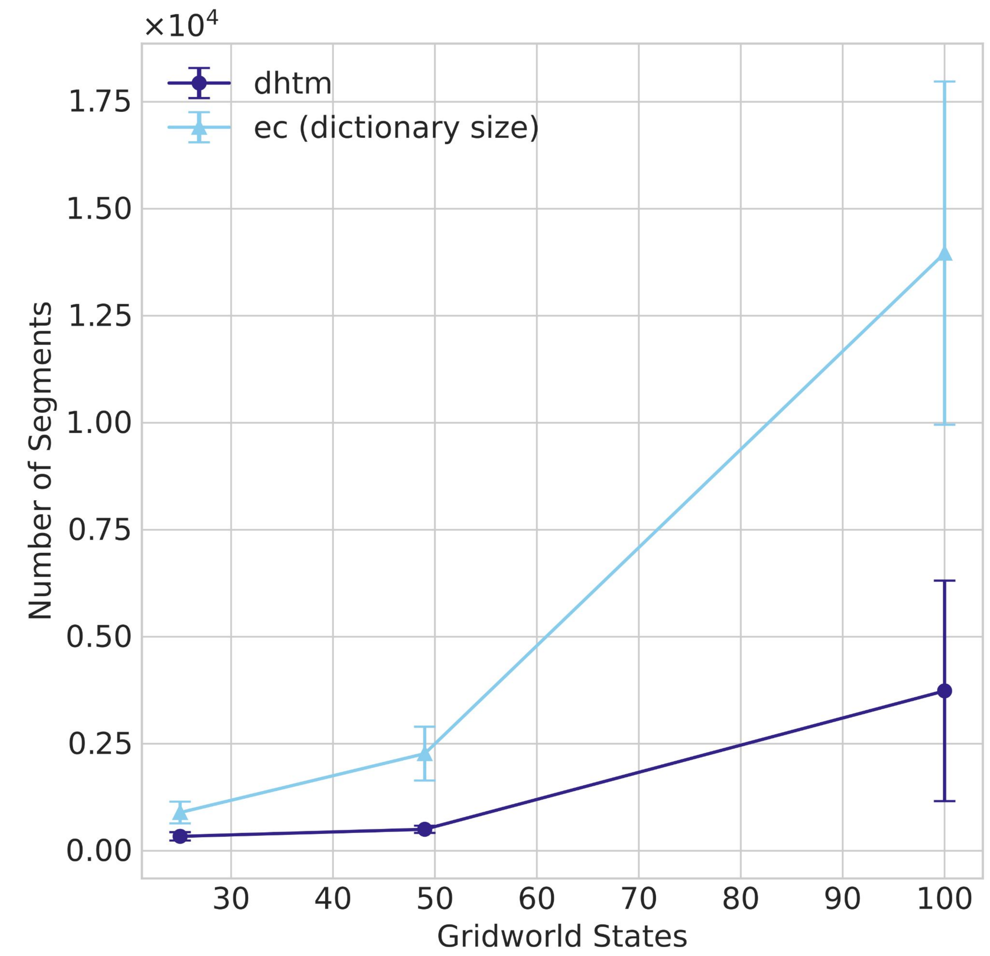
  $h_{t+1} \leftarrow \underset{h \notin K}{\operatorname{argmax}}\, F_e(h, \varphi_{t+1})$ # K is the set of keys for all actions

  $\mathcal{D}_{a_t}(h_t) \leftarrow h_{t+1}$

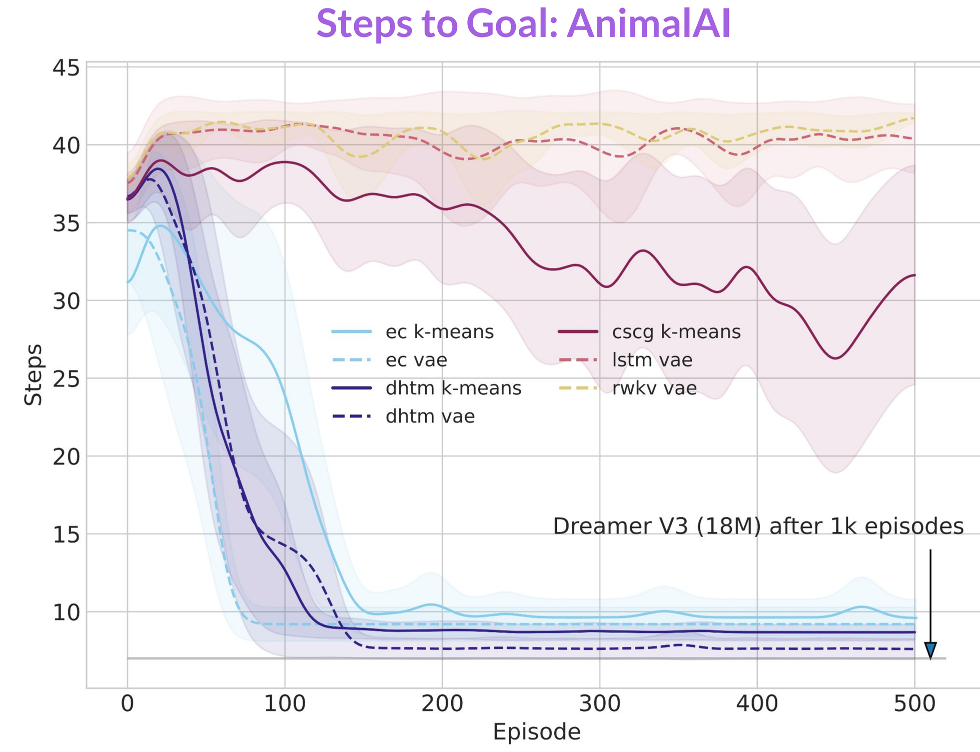
end if

$h_t \leftarrow h_{t+1}$

SCALABILITY AND DISTRIBUTIVITY



Steps to Goal: AnimalAI



Steps to Goal: AnimalAI with Noise