LEARNING FINITE STATE REPRESENTATIONS OF RECURRENT POLICY NETWORKS

Anurag Koul*, Sam Greydanus†, Alan Fern*

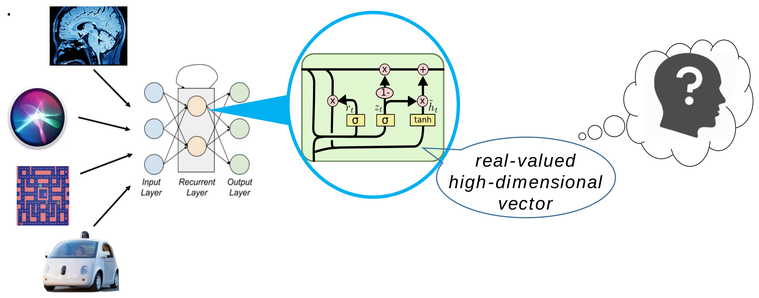
*Oregon State University, †Google Brain

GitHub.com/koulanurag/mmn

Introduction

Recurrent neural networks (RNNs) are an effective representation of control policies for a wide range of reinforcement and imitation learning problems. RNN policies, however, are particularly difficult to explain, understand, and analyze due to their use of continuous-valued memory vectors and observation features. This limits their usage in high-stake applications. In this work, we take a step towards comprehending and explaining RNN policies by learning more compact memory representations.

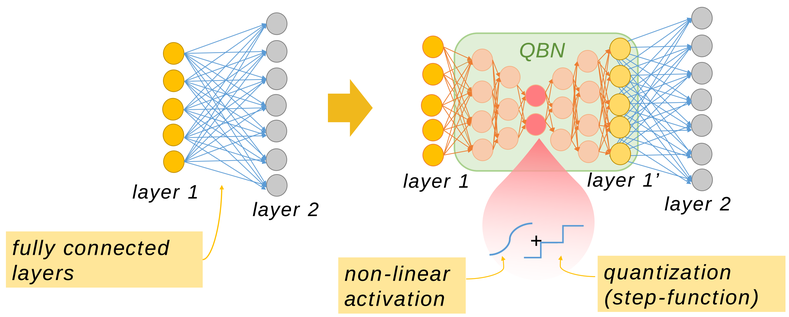
.



Proposal: Moore Machine Network (MMN)

An **RNN** policy is transformed to have finite observation and hidden-state space. This transformed network is referred as **MMN**. It's finite-space nature helps in understanding the behavior of the agent.

We introduce, Quantized bottleneck Networks (**QBN**) Insertion to get finite representations. A QBN is simply an auto-encoder where the latent representation at the bottleneck is constrained to be composed of k-level activation units.



Problem: quantize is mostly non-differentiable

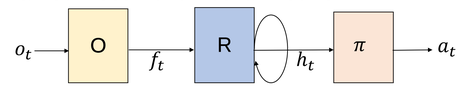
$$quantize'(y) = \begin{cases} 0\,; for\ y \neq \{-1,0,1\} \\ ?\,; for\ y = \{-1,0,1\} \end{cases}$$

Solution: straight-through estimator

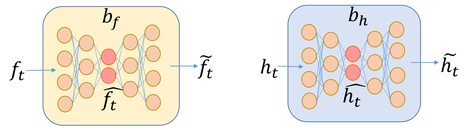
$$quantize'(y) = 1$$

Approach

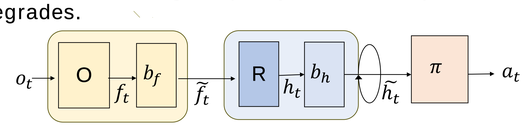
1. Pre-Trained Recurrent Neural Network



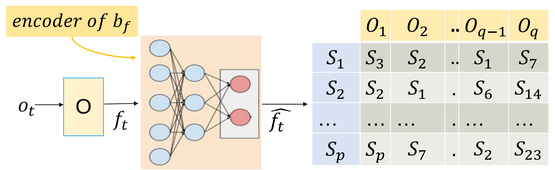
2. QBN Training
   - Perform policy rollouts to create a dataset of $< f_t, h_t >$
   - Train QBN b_f, b_h for finite state transformation of $f_t,\ h_t$



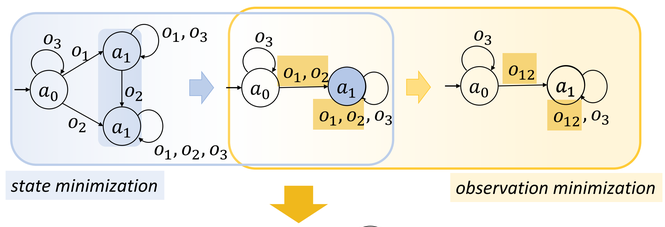
3. Transformation to Moore Machine Network:
   - Insert trained QBN into the network.
   - Fine-tune over original policy behavior if performance degrades.

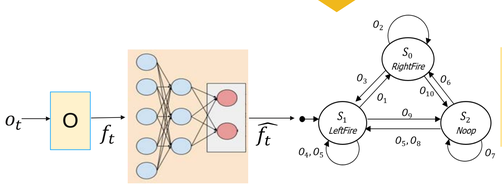


4. Moore Machine (MM)
   - Collect $< \hat{h}_{t-1}, \hat{f}_t, \hat{h}_t, a_t >$ from MMN policy rollouts to create a Moore Machine



- Minimize Moore Machine

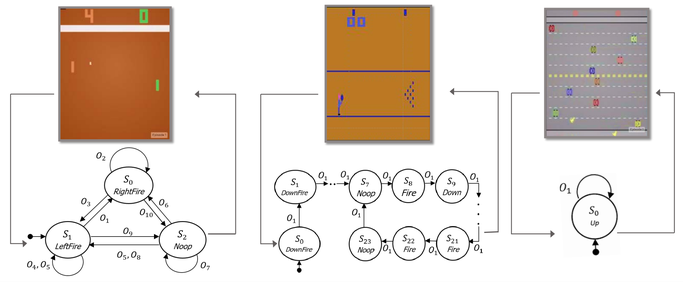




- $(m * n)$ *transaction*
- $m =$ *#min. equiv. states*
- $n =$ *#min. equiv. obs.*
- $m \leq p\,; n \leq q$

Experimental Results

- We evaluated our method over RNNs trained for Mode Counter Environment (Synthetic) , 7 Tomita Grammars and 6 Atari games. After minimization, there was a considerable state-space reduction for most policies.
- **Tomita Grammars**: no fine-tuning was required upon insertion of QBNs and minimal machines resulted into ***exact*** solutions.
- **Deterministic Atari**: We could retain approximate optimal policy on fine-tuning after QBN insertion. In the case of Pong, we had a perfect policy with 3 discrete memory states and 10 observations.
- The state machines provided insight into memory usage, exposing time-steps where **no-memory/only-memory** is used.



Game (# of actions)	RNN (Score)	(B_h, B_f)	Fine-Tuning Score		Before Minimization			After Minimization		
			Before	After	$\lvert\hat{H}\rvert$	$\lvert\hat{O}\rvert$	Score	$\lvert\hat{H}\rvert$	$\lvert\hat{O}\rvert$	Score
Pong (3)	**21**	64,100	20	21	380	374	21	4	12	**21**
		64,400	20	21	373	372	21	3	10	**21**
Freeway (3)	**21**	64,100	21	-	1	1	21	1	1	**21**
		64,400	21	-	1	1	21	1	1	**21**
Breakout (4)	773	64,100	32	423	1898	1874	423	8	30	423
		64,400	25	415	1888	1871	415	8	30	415
Space Invaders (4)	1820	64,100	520	1335	1495	1502	1335	8	29	1335
		64,400	365	1235	1625	1620	1235	12	29	1235
Bowling (6)	**60**	64,100	60	-	49	1	60	33	1	**60**
		64,400	60	-	49	1	60	33	1	**60**
Boxing (18)	**100**	64,100	94	100	1173	1167	100	13	79	**100**
		64,400	98	100	2621	2605	100	14	119	**100**

Summary & Future Work:

Deep neural networks with continuous high dimensional memory (RNNs) can be transformed into finite state machines while maintaining similar performance. This improves interpretability and acceptance in critical applications. A key direction for future work is to develop tools and visualizations for attaching meaning to the discrete observations and in turn states, which will allow for an additional level of insight into the policies.

Oregon State University