Overcoming Partial Observability with RNNs

Guillaume Berger

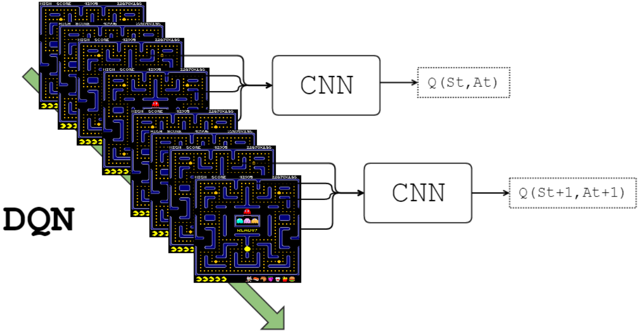
20 avril 2017

The limit of DQNs and POMDPs

Partial Observability :

- in most real-world applications, the entire world is not visible at any moment, but partially observable
- MDP $\Rightarrow$ **POMDP**

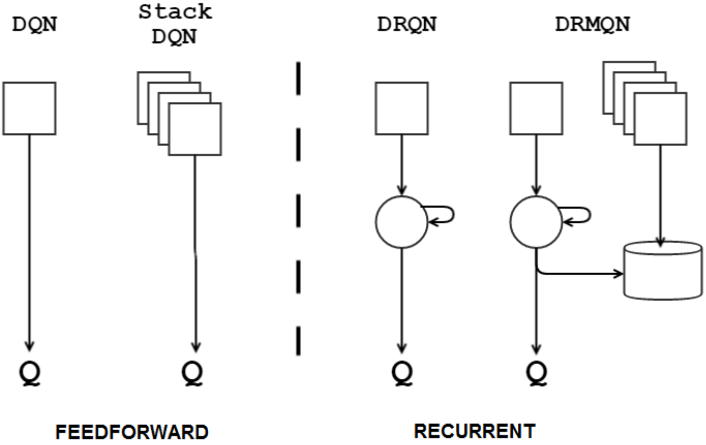
The limit of DQNs : the input receptive field is fixed



RNNs and Memory Networks

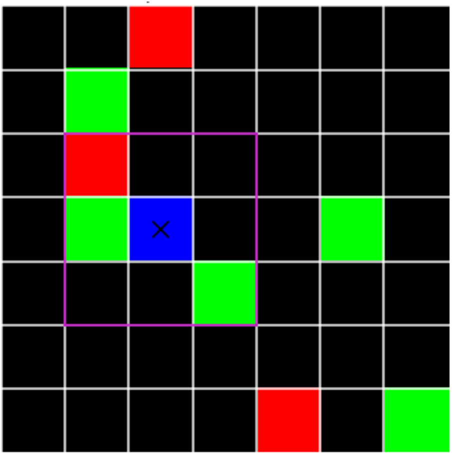
Recurrency might help : $h_{t+1} = \Phi(U\ x_{t+1} + W\ h_t)$

Implement and compare **different neural architectures** on **POMDPs** trained via **deep Q-learning**

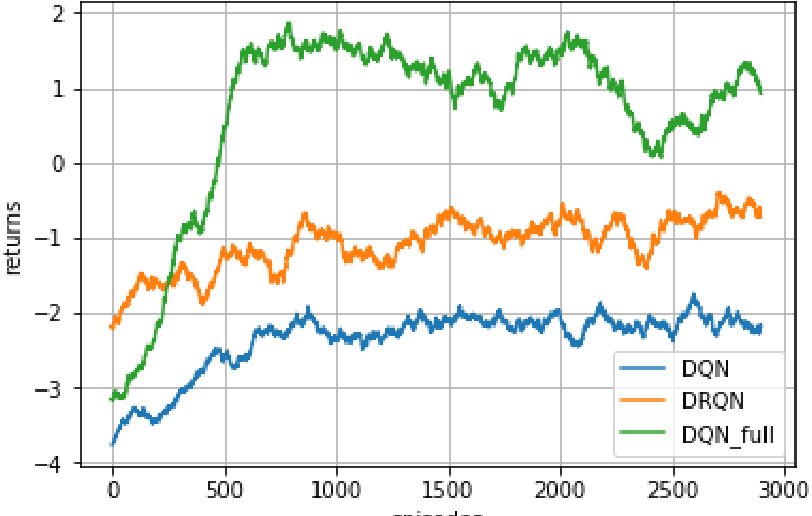


Environments

Experiment A : the myopic agent

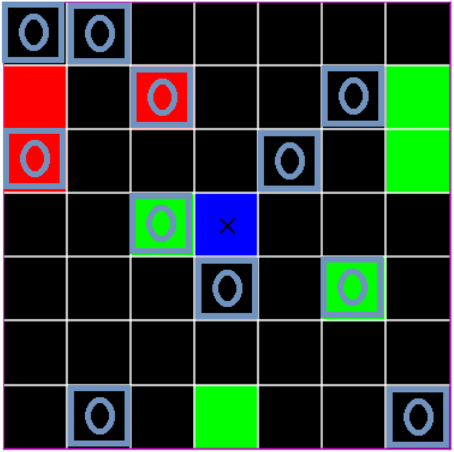


Learning curves



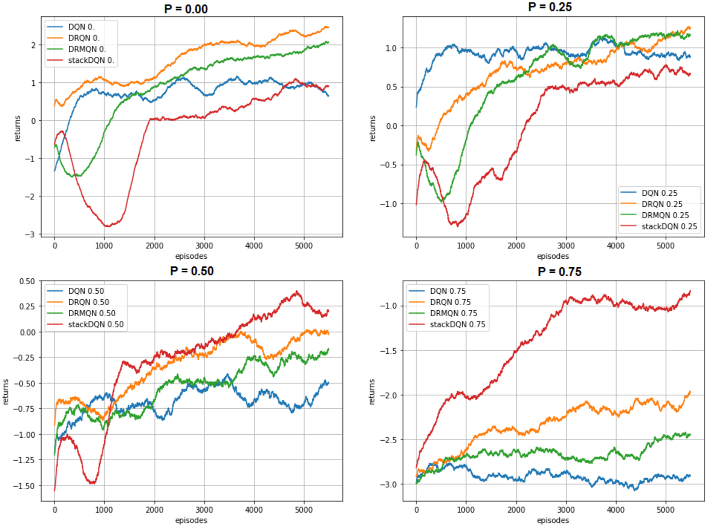
Environments

Experiment B : the masked agent



Dropout rate : P

Learning curves



DQN

Stack DQN

DRQN

DRMQN

Robustness to partial observability

- Training in a fully observable scenario (i.e. no mask, $P = 0$), but **testing with partial observability ($P = 0.5$)**.
- Experiment held on the "Masked" case.

Table: Average return for each architecture

DQN	Stack DQN	DRQN	DRMQN
-2.05	-1.27	-1.31	-1.45

References

[1] Deep Recurrent Q-Learning for Partially Observable MDPs, Matthew Hausknecht and Peter Stone, (2015)

[2] Playing Atari with Deep Reinforcement Learning, Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, Martin Riedmiller (2015)

[3] Control of Memory, Active Perception, and Action in Minecraft, Junhyuk Oh, Valliappa Chockalingam, Satinder Singh, Honglak Lee (2016)

Diferentiable memory module

[4] End-To-End Memory Networks, Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, Rob Fergus (2015)