C9: Playing Doom with a Deep Recurrent Q Network (DRQN)

▼ Table of Content

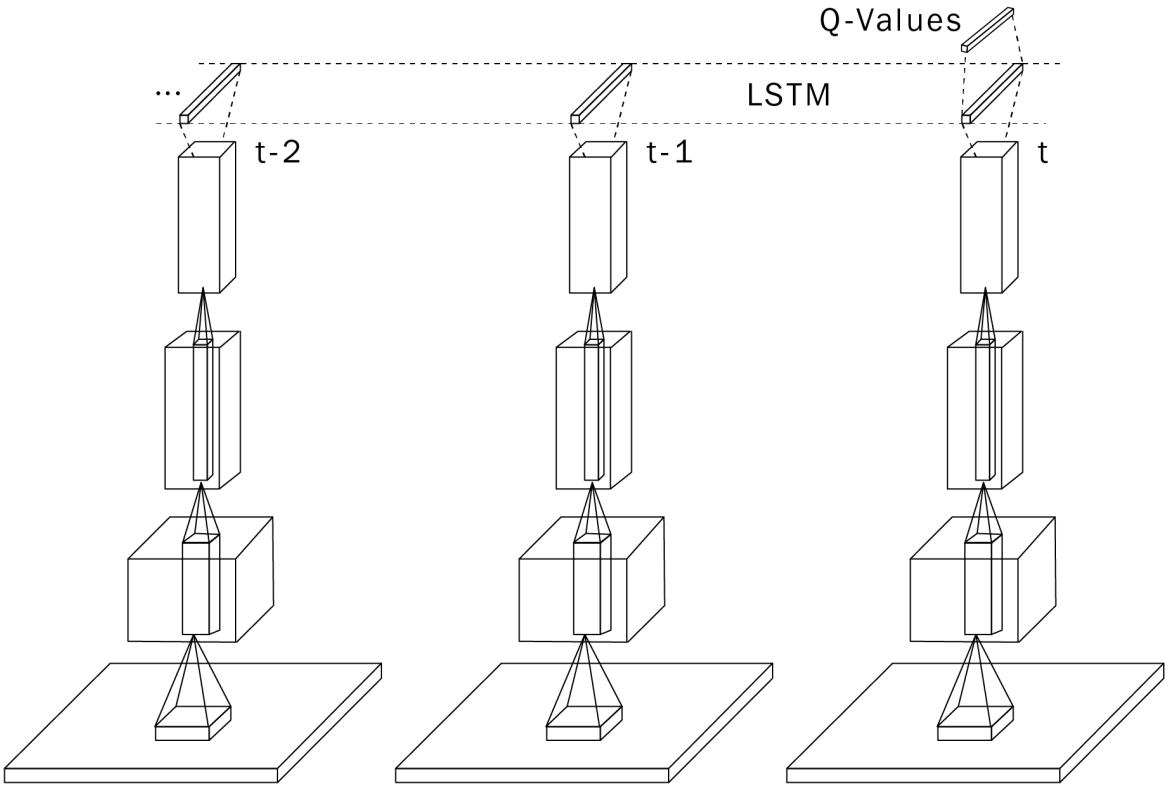
Partially Observable Markov Decision Process (POMDP)
Deep Recurrent Q-Network (DRQN)

Partially Observable Markov Decision Process (POMDP)

- An environment when we have limited set of information available about the environment.
- Thus, information of the past needs to be kept in the memory as it might help the agent better understand the nature of the environment and improve the policy.
- An example would be an agent learning to walk in the real-world environment the agent will not have complete knowledge about the terrain and information outside of its view (Or Otherwise here comes the Laplace Demon)
- In the game of Pong and DQN, past 4 frames is taken into consideration to determine the velocity of the ball but that might not be enough for memory intensive game. An intuitive solution would be to increase the number of frame from 4 to higher number but that comes at a high computing and memory resource (Imagine Replay Memory for past 100 frames).
- Thus, we can modify DQN architecture by augmenting LSTM layer to understand, retaining, forgetting and updating the information required to approximate the Q-function.

Deep Recurrent Q-Network (DRQN)

- The follow picture explains the architecture of DRQN where an LSTM replaces the first post convolutional fully connected layer with LSTM RNN.



- When a game screen is pass as an input, the Convolutional layer convolves the image and produces feature maps. The resultant feature map is passed to LSTM layer that has the memory retaining the information about important previous game states and updates its memory over time as required. It outputs Q-values after passing through a fully connected layer.
- Therefore, unlike DQN, $Q(s_t, a_t)$ is not estimated directly but we estimate $Q(h_t, a_t)$ where h_t is the input returned by the network at previous time step, $h_t = LSTM(h_{t-1}, o_t)$
- For Experience Buffer in DRQN, the entire episode is stored and we randomly sample n consecutive steps from a random batch of episode.
    - This accommodate both randomization and also an experience that actually follows another.