Paper Review - *"Deep Recurrent Q-Learning for Partially Observable MDPs"*

The authors of the paper build upon the existing model of Deep Q-Networks (DQNs) and present a novel approach to using recurrent neural networks for solving POMDPs (partially observable Markov decision processes) They do so by replacing a small part of DQN which is replacing the last fully connected layer in the network with a recurrent LSTM so that they can focus on the effect of added recurrences.

We understand the need of the model to be able to work on POMDPs in real-world applications where we do not necessarily have complete features and state information. By replacing the fully connected layer with LSTM we see that they are able to handle partial observability which DQN failed to address i.e. DQN needed the last 4 sequences of frame to estimate the next best state whereas DQRN was able to achieve similar results by having just a single frame as the input.

The main contribution of the paper is that by replacing the fully connected layer with the recurrent LSTM, the model is to decipher information from observing just a single frame and yields results at par with the DQN model which observes multiple frames.

The advantage of this method is that it saves on the huge memory utilization which has been a problem in the domain when using DQN. For that also, we could only observe the last 4 frames due to memory constraints. Essentially, using less memory, we are able to achieve similar results. The place where the paper lack is that the model doesn't necessarily beat the results of DQN all the time. It is good if we don't have enough computational resources but if the aim is the get better accuracy, it may not always work.

The authors did many experiments in the domain of atari games which were natively MDPs. It would be interesting to see the results for the games which are POMDPs such as Black Jack and see how the model performs on them as DQN is not the best model for them.

For extending the work, one of the places as suggested above is we can see how the model is able to perform on the games which are undoubtedly POMDPs. Also, we can leverage the latest models like transformers which have a proven record for performing well on sequential data.

It may lead to many interesting results as a much more complex model may try to capture the games better.