Paper Review - *"A Distributional Perspective on Reinforcement Learning"*

Summary: The author of the paper presents evidence that having a distributional approach can have better results than the traditional way of reinforcement learning that involves estimating the expected value of future rewards. They introduce a new algorithm called categorical DQN which learns the future reward learn directly from experience and demonstrates that it performs better in complex environments like atari games.

Main idea: Building on the author's vision of the paper, they remove the expectation away from the Bellman equation and instead consider the full distribution of the random variable which they call value distribution. The algorithm Categorical DQN introduced here learns the distribution of the future rewards in addition to estimating the value. It learns this using a variation of Bellman's equation that uses a probability mass function to represent the distribution of rewards. The network is trained using a variant of stochastic gradient descent, and the algorithm uses a replay memory to store and sample past experiences for training.

Strengths: They build the paper by showing a theoretical analysis of why the model works and makes sense. This is a different approach, different than the usual of first proving the hypothesis theoretically and then seeing the applied model on the dataset to confirm that it works in practice as well. It is a novel approach at that time and presented strong empirical results which outperform the SOTA.

Weakness: The experiments were restricted to Atari games only. It does not compare to other datasets and domains which could mimic more real-world scenarios and we could actually see the robustness of the algorithm. The distributional model requires large memory requirements and is computationally expensive. The paper explores only discrete action spaces and it is kinda unclear how to extend it to continuous action spaces.

Experiments: They evaluated the model on atari games, by training on 5 games and testing it out on 52 games. They did see that the model picks up on stochasticity. The model not only outperformed the different models like DQN and Double DQN but the SOTA (on most of the 57 games). The C51 model outperformed DQN only with 1/4th of the training frames (50 million frames). If it's let to have all the 200 million training frame, it can get significantly better results.

Extensions: Since the model was tested only on Atari games, we do not know if it generalizes well to different applications which cater more to the real-world scenario. It will be interesting to see the results evaluated on different datasets like TORCS and Labyrinth. We could also possibly extend the work to continuous action spaces available in atari. Also, could we explore different distributions other than categorical, like gaussian or something?