|  |  |
| --- | --- |
| Image result for university of victoria letterheadSunil **Kumar** | [*sunilkumarsheoran@gmail.com*](mailto:sunilkumarsheoran@gmail.com)  *+1-6043674432*  [linkedin.com/in/sunilkumarsheoran](http://www.linkedin.com/in/sunilkumarsheoran)  *ECS 412, University of Victoria,*  *Victoria, BC, Canada. V8P 5C2* |

In the 1950s, behavioral psychology gave birth to the main idea behind Reinforcement Learning (RL), that all animals learn to perform better in specific tasks if they are given positive rewards after the completion of such tasks. Keeping this idea in mind, RL has become a popular branch of machine learning. RL is different then Supervised Learning because the feedback in RL is not complete, and most rewards are delayed, i.e., a reward is given after doing multiple steps instead of each step. It is also a lot different than Unsupervised Learning because the goal of RL is not to find a hidden pattern in unlabeled random data but to maximize the reward solely. In recent years there have been many developments in the field which give rise to better RL algorithms. This includes some hybrid algorithms which are a mixture of tree search and RL algorithms such as AlphaZero (Google’s DeepMind 2017) and MuZero (Google’s DeepMind 2019).

A recent (2018) comparison of reinforcement learning algorithms was done using a cart-pole problem which is a classical benchmark problem for control purposes but is a very simple problem <https://arxiv.org/ftp/arxiv/papers/1810/1810.01940.pdf>. This does not include any of the hybrid algorithms. As my Thesis, I want to do a comparison between state-of-the-art reinforcement learning algorithms when applied to a much complex game such as chess. The goal is not to create a state-of-the-art engine/AI for chess but to study and list the different advantages and disadvantages of deep reinforcement learning methods by applying the methods to a game of chess. These methods will include state-of-the-art RL methods and AlphaGo zero which is a mix of tree search with RL.

I will start my experiment by first implementing a tree search-based model/bot for chess. Such models can find a perfect solution for a problem with small search space such as tic-tac-toe but for chess, it would need some modification such as reducing the search tree’s depth with position evaluation and reducing the search tree’s width with alpha-beta pruning. After tree-search based engine I would start working on deep reinforcement learning-based models starting from basic policy gradient method then value-based method and then actor-critic method. My initial experiment will end with a model based on AlphaZero.

My thesis will be focusing on a comparison of the Value-based RL method i.e. Q-Learning, Policy gradient method, and AlphaGo when applied at a game of Chess. After having all the models my second step would be getting an Elo rating for all the models. Either by deploying them on an online matchmaking server or playing games against known chess engines. Then I will work on finding more comparison parameters such as time to predict, model size etc.