GTrXL with PPO

NOT-SO-PREDICTING STOCK PREDICTION MODEL USING TRANSFORMERS IN A REINFORCEMENT LEARNING ENVIRONMENT.

Gianluca Monte, Francesco Di Monaco, Elvis Palos

December 2023

Project description

In this project we wanted to figure out the performance of the Transformer Model in Stock Prediction applications. We knew right away that this was not going to work for profit; we just wanted to see how (un)accurate the model would be at predicting one of the most unpredictable things on planet Earth. We decided to use a Transformer for its ability to work exceptionally well with sequences.

The transformer model we used was a modification of the stock transformer model proposed in "Attention is All you Need", which is optimized for Reinforcement Learning applications. This model is better known as the GTrXL (Gated Transformer-XL).

The key differences between GTrXL and normal Transformers are:

- Gated Connections, meaning GTrXL replaces the typical Transformer's residual connections with gated connections. This modification reduces errors that could otherwise flow through the residual connections;
- Re-ordered Layer Normalization: GTrXL applies layer normalization to the Transformer's sub-components but not to the gated connections. This allows the network to preserve information, including information derived directly from the input;

- Longer-term Memory: GTrXL gives sequence-processing neural networks longer-term memory, which can be beneficial in RL applications where the true state of the environment can only be derived from past observations.

As promising as this model might look, unsurprisingly it still wasn't enough to learn any underlying function in the stock market. Another very important challenge in using Reinforcement Learning in predicting a stock market is the rarity of having a positive reward. Being positive rewards so rare compared to negative ones, the model after a few iterations would drop to doing zero market operations to "maximize" the reward (at zero). We tried to counter this problem using various techniques, but even using Proximal Policy Optimization, which includes mechanisms to encourage exploration (such as entropy bonuses) we failed.
In the end, our model is still not powerful or advanced enough to predict the stock market correctly in a consistent way, but with simple functions (simple-dataset is a noisy sin function to simulate opening and closing prices) the model is able to quickly learn and in just 5 to 10 episodes it can generate profits of more than 50.000%.

This proves that the GTrXL with PPO could be suited for the job but is still not powerful enough (we ran it on a desktop computer) to consistently generate profit by making trades in the stock market. Who knows how many parameters would be needed to make this model work... what we are sure about is that for now we can only dream about having such computing power!