Money Is All You Need Check In #3

Prottoya Chowdhury (pchowdh4), Bailing Hou (bhou14),
Devesh Kumar (dkumar23), Haoyang Xu (hxu126)

25 Apr 2025

1 Introduction

In this project, we will re-implement the AAAI-2024 paper: *MASTER: Market-Guided Stock Transformer for Stock Price Forecasting* by Li et al. 2024 [1] and predict US stock prices by training on the Global Factor Dataset. The paper aims to tackle stock price forecasting, a challenging problem due to the market's high volatility, by solving the limitations of existing works.

Suggested by our team name *Money Is All You Need*, we are interested in the financial market, and stock price prediction is an important topic within this field, with ongoing research. Incorporating a modified Transformer model to tackle this problem elegantly, Li et al. 2024 [1] use deep learning techniques and innovative methods that capture the stock correlation and automatically select relevant features with market information, making it an outstanding application of deep learning to tackle real-world challenges effectively and creatively.

Since the purpose of this model is to use Transformers to train a deep learning model to predict a series of future return data for different stocks, the task of the model would be classified as a structured prediction problem.

2 Challenges

The hardest parts of the project we have encountered so far are as follows, along with our possible solutions:

1. **Problem:** The dimensions of the parameters are often confusing, especially in the case of stocks.
   Solution: We wrote and could write more comments/annotations of sizes besides the parameters to avoid confusion.

2. **Problem:** The codes have strong dependencies across sections, so it would be hard to distribute tasks (i.e., having one person implementing part A while another implementing part B that is strongly connected to part A).
   Solution: As the reading period is beginning soon, we would probably meet more often throughout the next few days to make sure everyone is understanding each part of the code (i.e., to deal with the remaining part of data loading), and we should write better documentation of each function to help readability.

3. **Problem:** We ran into issues during the pre-processing of the global factor data because the file is too big.
   Solution: So, while our teammate focused on unpacking the global factor data, we preprocessed open source data and had '.pkl' files ready to be trained on.

3 Insights

We have currently pre-processed some of the data, but it is too large and complicated to be all processed and is difficult to show the result. Since we have not had a running model, we haven't yet had training and testing results. However, our processed data in '.pkl' file worked in the original implementation, so they are ready to be in use.

4 Plan

Yes, we are currently on track. With that being said, we would still need to dedicate more time to the following:

1. **Training and evaluation.** Once we have implemented a working model, we would need to start training and fine-tuning hyperparameters, as well as conducting evaluations against baseline models.

2. **Fixing Bugs.** It's also important that we go through the code and fix any bugs that may have arisen. This will ensure that the data is as accurate as possible.

In terms of what we would like to change, currently, the code does not have a function to save and load the best-performing model, so we should modify the code to include that function such that we can keep track of the best models.

References

[1] Tong Li, Zhaoyang Liu, Yanyan Shen, Xue Wang, Haokun Chen, and Sen Huang. Master: Market-guided stock transformer for stock price forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1):162–170, Mar. 2024.