Using Intra-Stock, Inter-Stock, and Temporal Aggregation to Predict Stock Market Returns and Stock Selections

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

16 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 high volatility of the market, by solving 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] uses 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 a classified as a structured prediction problem.

2 Related Works

1. Yoo et al. 2021 [2] uses DTML (Data-axis Transformer with Multi-Level contexts) framework to learn stock correlations end-to-end and predicts stock movement by effectively correlating multiple stocks. The DTML framework consists of three main modules:

   (a) attentive context generation that uses attention LSTM to summarize multivariate historical prices;

   (b) multi-level context aggregation that combines local and global market contexts;

   (c) data-axis self-attention that uses a Transformer encoder to calculate asymmetric attention scores.

   Learning from various stock markets of US, China, Japan, and UK, DTML improves on accuracy and the Matthews corelation coefficients compared to previous approaches.

2. List of public implementations of Li et al. 2024 [1]:

   (a) Official implementation

   (b) A Qlib implementation

3 Data

For our initial analysis, we will be using the Global Factor Dataset that contains information about the price of stocks along with over 150 variables pertaining to the stock for stocks across 93 countries.

The overall dataset is very large, about 75GBs, and contains data on stocks that differ fundamentally across countries, economic structures etc. As such, we believe that training our model on the entire dataset is likely to encounter two issues:

1. High Computation Costs: Eveh though we are going to be training our model through utilizing Brown's computer cluster (OSCAR), we could encounter issues where we can't access enough compute power to train our model on the entire dataset.

2. Different Time Horizons: The dataset contains stock related information for companies ranging from about the 1960s to 2020. Economies and stock markets throughout these times were quite different and so stocks and their characteristics from earlier time periods in the data are unlikely to resemble stocks in present day. As such, we limit our time horizon for stock data from 1990 to about 2020. This gives our model the ability to learn different market environments and types of stocks while also making sure that our stock data is more consistent across time.

3. Markets across different geographies and economies can differ: Economies across countries and continents can differ in the amount of investment activity, the level of savings contributed to the stock market and the overall perception of investing wealth into stocks. Moreover, certain parts of the world have less developed financial markets, which means the model would have less data to infer results from those regions. For this reasons, and to improve accuracy of our model, we chose to limit our data to US stocks only.

   In addition to our initial analysis, we hope to further enrich our dataset by including text data. To do so, we will be using the FNSPD dataset that contains 15.7 million articles on US stock markets and companies from 1993-2023. Our goal would be construct embedding vectors for each of these articles and merge them with the factor data. This would incorporate information about overall market conditions and sentiments during the time as captrued by news coverage during the period.

4 Methodology

In terms of architecture, this model has five elements: market-guided gating, intra-stock aggregation, inter-stock aggregation, temporal aggregation, and prediction layers. The market-guided gating element uses a market status vector to scale different features based upon that market status vector. Intra-stock aggregation means creating an embedding that takes into account the temporal context around a specific point in time for a stock. Inter-stock aggregation involves creating an embedding of how stocks compare to each other. This step uses attention. Temporal aggregation means putting temporal embeddings together to form an aggregated stock embedding. The final prediction layers help determine what the returns for a stock will be. [1]

An intuitive understanding of the architecture is as follows. The market-guided gating using the market's status to help determine which features are most important. The inter-stock aggregation helps use overall market dynamics to determine returns. Intra-stock aggregation uses an individual stock's data to determine its returns. The prediction layer is simply just dense layers to determine the final output.

We will train the model using OSCAR. The hardest part about implementing the model will be managing the huge corpus of data, and also using that data to create a model that is able to effectively predict returns.

5 Metrics

We plan on measuring the performance of our model in the following two ways:

1. **Measuring Returns**: We would like to predict the returns of different stock prices, which will be measured in terms of percent changes of a stock's price. The performance of the model, in this part, can be measured by the accuracy between the actual percent change and the predicted percent change using the standard mean-squared error (MSE). The goal of the model would be to minimize the MSE such

that our model's performance can approach the real-world data as closely as possible. Furthermore, even though we will be using this accuracy measurement to select stocks and measure the model's selection skills (as mentioned below), the MSE loss will be the key measurement with which we will train the model to minimize.

2. **Measuring Stock Selection Skills**: While our model's core task is to predict percentage returns of a stock between two time periods, the core use of this model would be to inform investors about how to allocate their capital into certain stocks that are likely to outperform the broader market. As such, to analyze the effectiveness of the model at predicting returns, we need to quantify how much 'investing skill' the model has that is independent of the market performance.

   To infer the 'investing skill' of the model, we will be using the model to construct a portfolio of the top 100 stocks with the highest predicted returns. Then, we will evaluate the returns of these constructed portfolios using two metrics that are:

   (a) Sharpe Ratio: This metric allows us to identify how much return the portfolio generates for each percentage of standard deviation in the returns of the portfolio. Higher sharpes tend to indicate that the portfolio is able to generate more return while incurring lower levels of risk. The formula for sharpe is the following:

   $$\text{Sharpe Ratio} = \frac{E[R_a - R_f]}{\sigma_a}$$

   In the formula above, R_a is the return of the stock, R_f is the risk-free rate, and σ_a is the standard deviation of returns of stock a.

   (b) Information Ratio: This metric allows us to identify whether using the model to generate a portfolio is able to produce higher returns as opposed to the overall return of the market. The formula is :

$$\text{Information Ratio} = \frac{\text{Portfolio Return - Market Return}}{\text{Tracking Error}}$$

Where Tracking Error is the standard deviation of the difference between Portfolio return and Market Return.

3. **Goals and Comparison With Paper**: In this section, we will first discuss the metrics used by the paper we drew inspiration and then discuss our base, target and stretch goals.

The paper our project is based on quantifies the accuracy of the stock returns predicted by the model using variants of information coefficient, which tracks the correlation between predicted and true returns for a stock at a future time period. Furthermore, it generates portfolios of 30 stocks with the highest predicted returns and then observes their returns over different time horizons. As me mentioned prior, the simulation of portfolio returns allows the paper to test the investing skill that the model has accumulated over its training period.

Now, we will discuss our goals for the project.

In terms of our base goal, we aim to recreate the transformer model that was developed in the paper in the paper to predict stock returns.

In terms of target goals, we hope to predict returns and then form portfolios and observe the returns of the portfolio created by using the model we build. The aim here is to observe the portfolio returns over shorter and longer time horizons to see if the model is better at predicting long term returns or short term returns.

Finally, our stretch goal is to incorporate text data using our text database into our transformer model. This will incorporate a lot of latent data on sentiment surrounding financial markets that is likely to influence the trajectory of stock returns.

6 Ethics

- *What is your dataset? Are there any concerns about how it was collected or labeled? Is it representative? What kind of underlying historical or societal biases might it contain?*
  Our dataset comes from the Global Factor Data website. One potential bias with this dataset is that the time span of data for each company is across many decades (dating back to as early as the 1940s), which already implies that those are companies that survive for longer–meaning that the companies represented in the dataset would mostly be more "successful" companies. As such, this dataset may contain biases by only depicting the stock market trends of more successful companies, thus making the model potentially biased in that the trends of less successful companies are not studied. Furthermore, those companies are all U.S.-based companies, so even if the model can be trained to predict U.S. stock market changes well, it would be difficult to expand and generalize this model to the world as certain trends that happen in the U.S. stock market may not happen in other international stock markets.

- *Who are the major "stakeholders" in this problem, and what are the consequences of mistakes made by your algorithm?*
  One major group of stakeholders includes the individual investors, where our deep learning model's results may influence their decisions on when to buy a stock, how much to buy, and when to sell or liquidate. Of course, the major consequences of

mistakes or lack of accuracy in our algorithm are that investors may receive incorrect information about what our algorithm predicts to happen and what actually happens in the stock market. This may lead to consequences such as missing the best time to buy a stock or, even worse, losing money by buying and selling at the wrong times. Another important group of stakeholders includes regulatory agencies such as the U.S. Securities and Exchange Commission (SEC). People tend to be more stressed and thus make more irrational decisions in times when stock prices are unstable, and the role of the SEC is to protect investors from misconduct. Regulatory agencies may be influenced by our algorithm's results in ways such as predicting when future times of stock market instabilities will come. Therefore, any mistakes or lack of accuracy in our algorithm may lead to the consequence that regulatory agencies may receive incorrect predictions of the time and scale of future instabilities, and thus take incorrect actions at the wrong time, leading to potential issues such as unnecessary chaos.

7 Division of Labor

1. All of us will work on reimplementing the model together. We will do so through collaborative coding sessions.

2. Prottoya will take point on working on procuring data, cleaning it, and preparing it for the model and training.

3. Devesh will take point on understanding how to use OSCAR and integrate our model with OSCAR. He'll also use OSCAR to run the model

4. Bailing will take point on writing the paper, and assigning portions to be written by others accordingly.

5. Sunny will take point on creating a presentation for DL day.

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.

[2] Jaemin Yoo, Yejun Soun, Yong-chan Park, and U Kang. Accurate multivariate stock movement prediction via data-axis transformer with multi-level contexts. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 2037–2045, New York, NY, USA, 2021. Association for Computing Machinery.