

Research on investment returns prediction based on machine learning

Tingyu yan

Financial Engineering in UCLA, los angles, 90028, USA

18147017910@163.com

Abstract. With the development of the Internet and the continuous improvement of the information disclosure mechanism of the securities market, individual investors have more and more ways to obtain investment information, including company research reports from analysts. However, analysts have mixed qualifications and publish a huge number of reports. For retail investors, how to analyze and distinguish these reports and make informed investment decisions based on them has become the key to profiting in a securities market where information is asymmetrical and risks and opportunities coexist. This is a priority that every retail investor should prioritize. In recent years, machine learning has become increasingly widely used in the financial sector, especially in the prediction of return on investment. In this work, a most advanced machine learning model Transformer is used, which is widely used in a variety of scenarios due to its advantages in processing sequence data. The Transformer model can effectively process and analyze large-scale financial time series data through the self-attention mechanism to capture subtle fluctuations and potential correlations in the market. We designed a Transformer-based architecture to combine multi-source data such as market macroeconomic indicators, stock trading data, and social sentiment analysis for comprehensive learning. Experimental results show that the model significantly improves the accuracy of investment return prediction and provides investors with efficient analysis tools.

Keywords: Investment Returns Prediction, Financial Data Analysis, Machine Learning, Transformer Model

1. Introduction

The development of the investment market has not only contributed to economic growth, but also greatly contributed to the prosperity of the investment analyst profession. This profession first emerged in the American market at the beginning of the last century [1]. Analysts who are certified by the American Investment Management and Research Association typically hold key roles in financial institutions such as securities firms, investment banks, etc. Their work mainly involves collecting, collating and analyzing information on various relevant resources, and making investment recommendations based on these analyses. Analysts use their information processing capabilities and professional knowledge to guide investors to make informed investment decisions, effectively reduce the asymmetry of market information, and thus improve investors' investment returns.

Investment analysis has gradually become a key factor for brokerages to attract customers, expand their business and enhance their core competitiveness, and is an indispensable part of the investment

market. With the development of the Internet and the improvement of the information disclosure system of the securities market, there are more and more channels for individual investors to obtain stock information, including various research reports on listed companies issued by securities companies [2]. These research reports occupy an important position among all market information intermediaries. The company's research reports, written by qualified professional analysts, not only cover the macroeconomic environment, industry analysis and the company's own situation, but also provide predictions about the company's future operations and stock price. These reports can convey in-depth analysis of the company's value and development potential to investors, inspire investment decisions, improve the efficiency of information transmission, reduce information asymmetry, and enhance the efficiency of market resource allocation [3]. In view of the key role of analysts, in recent years, various brokerage institutions have increased their efforts to train analysts, resulting in a significant increase in the size of the analyst team.

Analysts use publicly available and private sources of information in the market to collect data on the macro market, industry, and company level. After careful collation and in-depth analysis, they form and publish research reports that aim to provide decision-making support to investors. These reports help investors better understand market dynamics and company conditions to make more informed investment choices [4].

What is the quality of the company's research report, which is an important reference material for investors to make decisions? Considering the limited attention and cognitive ability of individual investors, can the research reports published by professional securities analysts really bring stock investment returns when they are fully trusted by individual investors? These issues are related to the practical utility of the research report and the trust of the market, and deserve in-depth discussion [5]. Analysts play a crucial role as information intermediaries in the financial markets and are an important part of the market structure.

Their core mission is to conduct in-depth analysis of the fundamentals of listed companies and provide important reference for investors' decision-making. In a complex environment involving conflicts of interest between investors and listed companies, it is a matter of concern whether analysts' research reports can provide valuable content and create excess returns for investors. In addition, for individual investors with weak professional skills, whether they should rely solely on analysts' research reports is also an important research topic [6]. Addressing these issues can help improve market transparency and investment efficiency.

2. Related Work

Initially, Moshirian et al. [7] were the first to empirically analyze the abnormal returns of analysts recommending the purchase and holding of equities in emerging markets (Regression). By looking at data from 13 emerging countries from 1996 to 2005, they found that stock prices reacted significantly to the recommendations and adjustments of securities analysts, that the recommendation to buy yielded significant abnormal returns in emerging markets, and that analysts in emerging markets showed a more positive bias than those in developed markets. Park and Park's [8] study shows that investors who bought "strong-buy" stocks and short-sold "strong-sell" stocks as recommended by U.S. stock market analysts achieved abnormal returns of 4.7%-5.8% per year between 2001 and 2016, even after accounting for transaction costs.

The analysis by Barber et al. [9] shows that the most recommended stocks generate significant market-adjusted average annual returns relative to the least popular stocks, reflecting the long-term value of analyst recommendations. Cai and Cen [10] examined securities analysts' trading recommendations for specific sectors of China's securities market and found that those that received positive recommendations generally outperformed those that received negative recommendations, and that this difference in performance increased over time. If investors follow analysts' buying and selling recommendations, they may make a profit in the medium to long term, and these stocks and sectors tend to outperform the benchmark index in the medium to long term. Conversely, those that receive negative recommendations often underperform the benchmark index.

Additionally, Hobbs et al. [11] found that analysts who frequently modified their investment recommendations achieved better return performance than those who did not. These analysts provide more accurate investment guidance by frequently adjusting their investment recommendations and incorporating incremental market information into their recommendations. Analysts who revise their investment recommendations use their skills and ability to analyze market activity to generate private information that results in better short-term returns.

From the perspective of target price in securities analyst research reports, Thacker et al. [12] found a significant correlation between target price and earnings forecasts and equity returns by analyzing the recommendation data of 657 Malaysian sell-side analysts from January 2010 to December 2015, confirming that the information in analysts' reports can explain 66.67% of equity investment returns (Correlation analysis). Further, researchers Brav and Lehavy [13] studied the impact of the target price on the stock market in the short term and the long-term synergistic change between the stock price and the target price from the perspective of the expected price given by the analyst, and found that the market reacted significantly to the information in the analyst's target price. These studies show that analysts' price targets are important for the market.

3. Methodologies

The Transformer model is based on the self-attention mechanism and can effectively process sequence data. The core advantage of this model is that it can process all data points for the entire sequence at the same time, rather than step-by-step like traditional recurrent neural networks. Using Transformer models for investment returns forecasting can effectively leverage its ability to process large-scale datasets while capturing long-term dependencies and complex patterns in time series data to provide investors with accurate forecasts.

3.1. Notions

Initially, we summarize the primary parameters in following Table 1.

Table 1. Notions.

Notion Symbols	Explanations
Q	Query matrices
K	Key matrices
V	Value matrices
d_k	Dimension of the key vector
h_N	Output of the last Transformer encoder
W_o	Weight matrix
b_o	Bias term
n	Number of samples
\hat{y}_i	Prediction result or the i observation

3.2. Transformer Model

The basic unit of a Transformer is made up of a self-attention layer and a feedforward neural network layer, which are wrapped in multiple encoder and decoder layers. For stock market forecasting, we focus primarily on encoder structures. A standard Transformer encoder layer consists of a self-attention layer and a feedforward neural network. The self-attention layer calculates the weight of the influence of each element in the input sequence on the other elements.

The feedforward neural network further processes the output of the self-attention layer. The core to the self-attention mechanism is to calculate the effect of each element in the input sequence on the other elements, a process that can be illustrated by the following Equation 1.

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (1)$$

Where Q , K and V are Query, Key, and Value matrices, respectively, which are obtained by transforming the input data through different weight matrices. d_k is the dimension of the key vector, and this division operation is to scale the size of the dot product to prevent the gradient disappearing problem.

In investment returns forecasting, we usually only use the encoder part. The input series are stock prices and other financial indicators over the past period. The encoder processes these inputs through a multi-layer self-attention and feedforward network to extract complex patterns in the time series. The output part of the model can be a layer of linear transformations that convert the output of the encoder into a prediction of future value, which is shown in Equation 2.

$$Output = W_o h_N + b_o \quad (2)$$

Note that parameter h_N is the output of the last Transformer encoder layer, W_o is the weight matrix of the output layer, and b_o is the bias term.

3.3. Optimization Function

When using machine learning models for investment returns forecasting, choosing an appropriate loss function is crucial. The role of the loss function is to evaluate the difference between the predicted value and the true value of the model, so as to guide the optimization of the model during the training process. Mean Square Error (MSE) calculates the mean of the squares of the difference between the predicted value and the true value, which is mathematically expressed as Equation 3.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3)$$

Where n is the number of samples, which is the total number of observations in the dataset. The \hat{y}_i means model's prediction for the i observation. Parameter y_i represents the actual value of the i observation of y .

The MSE provides a quantitative metric to measure the predictive power of a model. By minimizing the MSE, the model learns to adjust its parameters to predict an outcome that is close to true as accurately as possible. The MSE sums the squares of the prediction errors, which means that large errors are penalized more heavily. This helps the model avoid large prediction errors, as large errors contribute much more to the MSE than small ones.

4. Experiments

4.1. Experimental Setups

When conducting machine learning-based ROI forecasting research, RESSET provides a rich source of data, including company financial statements, stock trading data, and macroeconomic indicators. Models can be effectively trained and evaluated through well-designed experimental setups, such as data preprocessing, feature engineering, model selection, and training validation processes. The dataset is divided using time series segmentation, and the robustness of the model is ensured through cross-validation. At the same time, the simulated trading test can further verify the practical application effect of the model prediction. This integrated approach not only reveals the predictive power of the model, but also provides practical guidance for investment decisions, thereby improving investment efficiency and returns.

4.2. Experimental Analysis

Absolute return is a basic and straightforward indicator of the performance of an investment strategy or trading model, and it reflects the overall profit or loss of a portfolio over a specific period of time. This indicator is especially important because it provides a clear view of how the model performs in the actual

market environment. Following Figure 1 shows the absolute returns prediction results with existing methods.

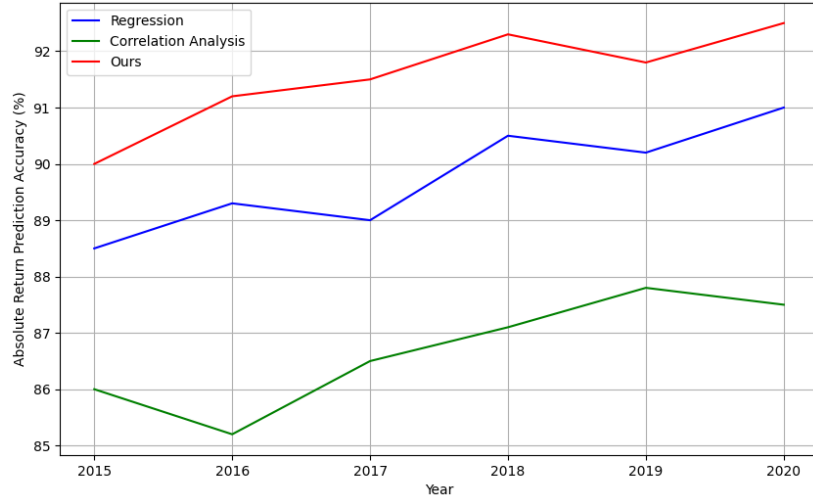


Figure 1. Comparison of Prediction Accuracy Over Years.

The maximum drawdown represents the maximum decrease in the value of the portfolio in simulated trading and is an important indicator of potential risk. A lower maximum drawdown means better capital protection in adverse market conditions. Figure 2 shows the maximum drawdown comparison results.

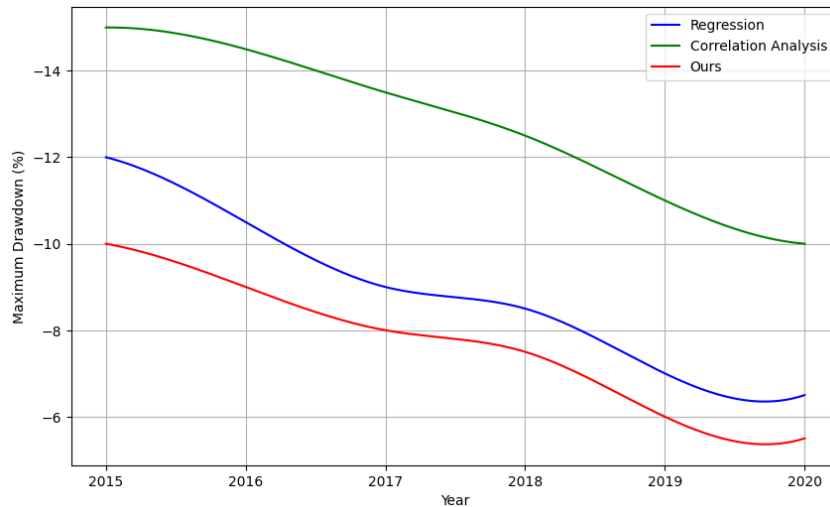


Figure 2. Comparison of Maximum Drawdown Prediction Over Years.

Figure 2 shows the maximum drawdown predicted by the three methods from 2015 to 2020. In the graph, we can see that the maximum drawdown of all methods shows an improvement trend over time. Among them, our method shows the lowest maximum drawdown value, indicating that the method is better able to protect investments from severe losses in the most adverse market conditions. In contrast, the regression analysis and correlation analysis also showed improvements, but with larger drawdowns over the entire period.

5. Conclusion

In conclusion, our proposed model has shown that advanced algorithms such as Transformer can effectively leverage large-scale financial and macroeconomic data to improve the accuracy of

forecasting. Experimental results show that these methods not only provide high forecast accuracy, but also effectively manage risk through simulated trading tests, as shown by the reduction through maximum drawdown. Despite some volatility, overall trends suggest that machine learning methods have significant advantages in identifying market trends, optimizing portfolios, and improving the quality of decision-making. The application of this technology provides investors and financial analysts with powerful tools to achieve better capital protection and return growth.

References

- [1] Thakkar, Ankit, and Kinjal Chaudhari. "Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions." *Information Fusion* 65 (2021): 95-107.
- [2] Lu, Ruochen, and Muchao Lu. "Stock trend prediction algorithm based on deep recurrent neural network." *Wireless Communications and Mobile Computing* 2021.1 (2021): 5694975.
- [3] Sen, Jaydip, Abhishek Dutta, and Sidra Mehtab. "Profitability analysis in stock investment using an LSTM-based deep learning model." 2021 2nd International Conference for Emerging Technology (INCET). IEEE, 2021.
- [4] Kanwal, Anika, et al. "BiCuDNNLSTM-1dCNN—A hybrid deep learning-based predictive model for stock price prediction." *Expert Systems with Applications* 202 (2022): 117123.
- [5] Wu, Wenbo, et al. "A cross-sectional machine learning approach for hedge fund return prediction and selection." *Management Science* 67.7 (2021): 4577-4601.
- [6] Blohm, Ivo, et al. "It's a people's game, isn't it?! A comparison between the investment returns of business angels and machine learning algorithms." *Entrepreneurship Theory and Practice* 46.4 (2022): 1054-1091.
- [7] Moshirian, Fariborz, David Ng, and Eliza Wu. "The value of stock analysts' recommendations: Evidence from emerging markets." *International Review of Financial Analysis* 18.1-2 (2009): 74-83.
- [8] Park, Sung Jun, and Ki Young Park. "Can investors profit from security analyst recommendations?: New evidence on the value of consensus recommendations." *Finance Research Letters* 30 (2019): 403-413.
- [9] Barber, Brad, et al. "Reassessing the returns to analysts' stock recommendations." *Financial Analysts Journal* 59.2 (2003): 88-96.
- [10] Cai, Rongchang, and Zilan Cen. "CAN INVESTORS PROFIT BY FOLLOWING ANALYSTS'RECOMMENDATIONS? AN INVESTIGATION OF CHINESE ANALYSTS'TRADING RECOMMENDATIONS ON INDUSTRY." *Economics, Management and Financial Markets* 10.3 (2015): 11.
- [11] Hobbs, Jeffrey, Tunde Kovacs, and Vivek Sharma. "The investment value of the frequency of analyst recommendation changes for the ordinary investor." *Journal of Empirical Finance* 19.1 (2012): 94-108.
- [12] Thaker, Hassanudin Mohd Thas, et al. "How useful are the equity analysts' report? Evidence from Malaysia." *Reports on Economics and Finance* 4.4 (2018): 221-246.
- [13] Brav, Alon, and Reuven Lehavy. "An empirical analysis of analysts' target prices: Short - term informativeness and long - term dynamics." *The Journal of Finance* 58.5 (2003): 1933-1967.