SignalAI: Predicting Financial Sentiment and SPY Returns Using BERT

Abstract

This paper presents a comprehensive study on the application of BERT (Bidirectional Encoder Representations from Transformers) to predict sentiment and SPY returns based on financial texts. The study is divided into three main parts: (I) Training BERT to predict sentiment using a HuggingFace dataset, (II) Training BERT to predict SPY returns based on FOMC meeting minutes, and (III) Improving the results in (II) by experimenting with different holding periods, tuning hyperparameters, and exploring models beyond BERT. My project shows the power of NLP, in analyzing significant financial news and documents, to predict stock

1 Introduction

Financial markets are influenced by a myriad of factors, including investor sentiment and economic policies. Predicting market movements based on textual data, such as news articles or central bank meeting minutes, has become a crucial area of research. This paper explores the use of BERT, a state-of-the-art NLP model, in predicting both sentiment and SPY returns based on financial texts. The research is motivated by the need to understand how central bank communications, particularly the FOMC meeting minutes, impact market behavior and how modern NLP techniques can be utilized to enhance predictive accuracy.

2 Related Work

Sentiment analysis in finance has been extensively studied, with early approaches relying on lexicon-based methods and more recent work utilizing machine learning models. The introduction of transformer-based models like BERT has significantly improved the accuracy of sentiment analysis by capturing contextual nuances in text.

In recent years, there has been significant advancement in leveraging Large Language Models (LLMs) for financial decision-making and trading. One notable work is

FinAgent, a multimodal foundation agent designed specifically for financial trading tasks. This model integrates both textual and visual information to perform comprehensive analyses of the financial markets, leveraging various market data types such as news reports, price trends, and Kline charts.

FinAgent's unique design includes a dual-level reflection module that allows it to adapt to market dynamics quickly, reflecting on both real-time data and historical trends to make more informed trading decisions. It emphasizes tool augmentation, enabling the model to combine insights from expert guidance and traditional financial indicators to ensure that its decisions are robust and data-driven.

In the context of financial markets, several studies have explored the relationship between news sentiment and stock returns. However, predicting market returns directly from central bank communications, such as FOMC minutes, remains an underexplored area. This study builds upon previous work by applying BERT to this specific problem and further enhancing the model's performance through various techniques.

3 Methodology

Part I: Sentiment Prediction Using Twitter Financial Sentiment Dataset

The first part of the study involved training BERT to predict sentiment using a labeled financial sentiment dataset from Twitter. The dataset consists of financial news headlines labeled as positive, negative, or neutral. The following steps were undertaken:

1. **Data Preprocessing**: The text data was cleaned and tokenized using the BERT tokenizer.
2. **Model Training**: A pre-trained BERT model was fine-tuned on the Twitter dataset using the Hugging Face transformers library. The model was trained to classify each headline into one of the three sentiment categories.
3. **Evaluation**: The model's performance was evaluated using metrics such as accuracy, and MSE score..

Part II: Predicting SPY Returns Using FOMC Minutes

In the second part, BERT was trained to predict SPY returns based on FOMC meeting minutes. The holding period is set to one day, with the price gap between open and close price of SPY on the FOMC meeting day as the label of the model. The steps involved are as follows:

1. **Data Collection**: FOMC meeting minutes were downloaded in HTML format in batch using a crawler function and processed to extract the textual content. Corresponding SPY open and close prices were obtained from Yahoo Finance.
2. **Data Alignment**: The FOMC minutes were aligned with the SPY price data by matching the dates. The difference between the SPY open and close prices was calculated for each FOMC meeting date.
3. **Model Training**: BERT was fine-tuned on the aligned dataset to predict the SPY price gap of the day of the FOMC based on the content of the FOMC minutes.
4. **Evaluation**: The model's predictive accuracy was evaluated using a test set with MSE and a comparison to the simple buy and hold strategy.

Part III: Enhancing Model Performance

To improve the results obtained in Part II, several strategies were employed:

1. **Experimenting with Holding Periods**:A longer holding period was tested to see if SPY returns over longer or shorter time frames could be better predicted by the model. Specifically, the label was set to the price gap between the opening price of the day after the FOMC meeting and the closing price a month later.
2. **Exploring Alternative Models**: Other models, such as RoBERTa were tested to see if they could outperform the original BERT model. I wanted to experiment with a larger model to benchmark the performance of my system. I chose the RoBerTa-base model from Huggingface because they are both transformer-based models widely used in natural language processing (NLP) tasks. RoBERTa is essentially an optimized version of BERT, with various modifications aimed at improving its pretraining process and overall performance.

Advanced Techniques

In this project, I utilized web crawling and text extraction techniques to gather the training data for the second task and automate the data collection process of FOMC meeting minutes from the Federal Reserve's website. Web crawling involves systematically navigating through web pages and extracting relevant information. Using

Python libraries such as `requests` and `BeautifulSoup`, my project initiated a web crawler that accessed the FOMC calendar page, identified and followed links to individual meeting minutes, and downloaded these documents in HTML format. The HTML grabbing process involved parsing the downloaded web pages to extract the textual content of the minutes, which was then cleaned and prepared for subsequent analysis. This automated approach allowed for efficient and consistent retrieval of large volumes of data, ensuring that the FOMC minutes were accurately aligned with corresponding financial data, such as SPY returns, for further processing in the machine learning models.

Calculation of Return

Since my model is predicting a price gap, it would not be reasonable to simply sum all the predicted price gaps and compare with the sum of true price gaps. The strategy that I used is to allow the model to open positions in both directions. When the predicted price is positive, it adds the price gap of that label to *cumulative_predicted_price_gap*, and when the prediction is negative, it subtracts the price gap of that label from *cumulative_predicted_price_gap*.

Results

Part I: Sentiment Prediction

The fine tuned BERT model achieved an improvement of over 100% over the base model with a test error of 0.41 and the original model with test error of 1.08. The test error is MSE between the predicted label, either 0, 1 or 2. So this test error value can be interpreted as, on average, the prediction label is 0.4 class away from the true label, whereas the pretrained model's prediction is on average 1 class away from the true label .

Overall, the model performed well in distinguishing between positive and negative sentiments, though it struggled slightly with neutral cases.

Part II: SPY Return Prediction

The initial model trained on FOMC minutes achieved an MSE of 20.92 in predicting SPY returns, indicating some predictive power but leaving room for improvement. The model was able to capture certain market responses to FOMC communications but was less effective in more ambiguous cases. The improvement of the finetuning process is subtle compared to the first task.

Another evaluation metrics that I used is comparing the cumulative profit of my software vs a common trading strategy: Buy and Hold. Our model outperformed the simple strategy by about 5%.

Overall, the BERT model performed as expected since our dataset is extremely limited by the number of FOMC meetings ever happened. However, we still see a positive result compared to naive strategies such as B&H.

Part III: Model Enhancements

- **Holding Periods**: Extending the holding period to one month days improved the model's MSE by 3% , suggesting that the market's reaction to FOMC minutes may unfold over several days.
- **Alternative Models**: Surprisingly, the new model does not perform better than BERT, producing a MSE even higher. After reading and learning about the model, I found out that RoBerta is pretrained on a larger dataset than BERT, thus it is prone to overfitting when a smaller dataset(44 rows of FOMC data in our case) is used.

Discussion

The study demonstrates the potential of BERT in predicting financial sentiment and market returns based on textual data. However, the results also highlight the challenges inherent in predicting market movements, where factors beyond textual content, such as economic indicators and investor sentiment, play a significant role. The improvements achieved through experimenting with holding periods, additional features, and model variants suggest that a multi-faceted approach is necessary for accurate financial forecasting. I would definitely continue my work to further depth after the course has finished.

Challenges and Limitations

- **Limited Data**: The training set is very limited due to the limited number of FOMC meetings available

- **Textual Complexity**: FOMC minutes are complex documents that require nuanced understanding. The model may struggle with ambiguous language or mixed signals in the text.
- **Market Factors**: SPY returns are influenced by a wide range of factors beyond FOMC communications, limiting the model's predictive accuracy when relying solely on text data.

Future Work

Future research could explore the following areas:

1. **Incorporating More Financial Data**: Integrating more granular financial data, such as intraday price movements or sentiment from social media, could improve predictions.
2. **Non-Textual Data**: Including features like the VIX (Volatility Index) and GDP growth improved the model's performance by providing additional context
3. **Advanced NLP Techniques**: Exploring newer transformer architectures or fine-tuning methods specifically designed for financial data could yield better results.
4. **Ensemble Models**: Combining BERT with other models, such as traditional time series models or sentiment analysis tools, could enhance predictive accuracy.

Effort and Learning

This project required a deep understanding of NLP models, financial data processing, and model evaluation techniques. I spent significant time on data preprocessing, particularly aligning FOMC minutes with SPY data. Learning how to fine-tune BERT and optimize its performance was crucial, as was exploring alternative models and features. While the initial approach relied heavily on existing libraries, I found myself spending more time on data gathering, fine-tuning, and model experimentation. Through this project , I also realized that data is of the utmost importance in a good language model. Thus, I will pay extra attention to data gathering and analytics in the future.

Bibliography

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
2. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692.
3. Araci, D. (2019). FinBERT: A Pre Trained Language Model for Financial Communications. arXiv preprint arXiv:1908.10063.
4. Federal Reserve Board. (n.d.). Monetary Policy - FOMC Meeting Minutes. Retrieved from https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm
5. Yahoo Finance. (n.d.). SPY - SPDR S&P 500 ETF Trust. Retrieved from https://finance.yahoo.com/quote/SPY/
6. Li, T., Yang, Y., Yang, L., Sun, Z., Zhou, C., & Wang, C. (2023). FinAgent: A multimodal foundation agent for financial trading with multimodal tools and dual-level reflection. arXiv preprint arXiv:2308.07513. https://arxiv.org/abs/2308.07513