Capstone Project Report

The Practice of QCF - MGT6785

Submitted by:

Hales, Morgan

Sarkari, Zenobia

Tuahivaatetonohiti, Guillaume

Date:

05/02/2024

Table of Contents

Introduction

In the dynamic world of finance, the ability to predict market trends with accuracy and efficiency is invaluable. The S&P 500 index, a barometer for the U.S. economy, is particularly significant due to its broad representation of the market's condition. This project focuses on the application of TimeGPT, a state-of-the-art generative model, to forecast the time series data of the S&P 500 index. The study also involves a comparative analysis with established benchmark models, specifically Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), to evaluate the effectiveness of TimeGPT in this domain. The motivation behind this study stems from the growing complexity and volatility in the financial markets, which demand more sophisticated tools for prediction and analysis.

Traditional models for forecasting financial time series, such as RNNs and CNNs, often struggle with the inherent noise and non-linearity of market data. This project addresses the challenge of predicting the future values of the S&P 500 index, which is crucial for a range of stakeholders, including investors, economists, and policy makers. The objective is to leverage and evaluate TimeGPT's capabilities to enhance prediction accuracy and reliability, overcoming the limitations of older statistical methods and simpler machine learning models.

The motivation for this project is twofold. First, it seeks to demonstrate the practical application of advanced machine learning techniques in real-world financial scenarios. Second, by improving forecast accuracy and comparing the performance of TimeGPT against RNNs and CNNs, this project aims to provide actionable insights that can lead to better investment strategies and more informed economic decisions. Furthermore, the use of TimeGPT could potentially open new avenues in financial modeling, making it a pioneering study in applying generative models to time series forecasting.

This capstone project not only explores an innovative approach to financial forecasting but also contributes to the broader field of financial analysis by integrating cutting-edge AI technologies with economic forecasting.

Data

The data for this project was sourced from Alpha Vantage, an online service providing real-time and historical financial data. The dataset consists of daily trading information for the SPY ETF, an exchange-traded fund that closely tracks the performance of the S&P 500 Index. Data was retrieved using the Alpha Vantage API, with the 'TIME_SERIES_DAILY' function specified to obtain daily resolution data. This dataset is ideal for the project due to its relevance in representing the broader U.S. stock market and its accessibility in a structured format.

The dataset includes the following features for each trading day:

- Open: The price at which the stock first traded upon the opening of the exchange.
- High: The highest price at which the stock traded during the trading day.
- Low: The lowest price at which the stock traded during the trading day.
- Close: The price at which the stock last traded upon the closing of the exchange.
- Volume: The number of shares or contracts traded in a security or an entire market during a given period.

The 'outputsize=full' parameter was used in the API request to ensure the retrieval of the full historical data available for SPY, spanning from its inception to the present. This comprehensive dataset provides a robust foundation for conducting time series analysis and model training.

Upon retrieval, the JSON response from the API was parsed into a Python dictionary, and relevant trading data was extracted and stored in a structured format suitable for analysis. Preliminary data processing included cleaning steps such as handling missing values and outliers, ensuring data integrity, and normalizing the data where necessary to facilitate model training and comparison.

The final dataset comprises over two decades of daily trading data, encompassing various market conditions including bull markets, bear markets, and financial crises. This diversity in data helps in training more robust and generalizable forecasting models.

Methodology

Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs) are a class of neural networks suited for sequential data analysis. Unlike standard feedforward neural networks, RNNs have the unique feature of maintaining a memory (state) that captures information about what has been calculated so far.

In the context of time series forecasting, this allows RNNs to retain state across time steps, making them ideal for predicting future values based on previous sequences of data.

RNNs process time series step by step, maintaining an internal state from time $t-1$ to time t. This is particularly advantageous for financial time series, where the temporal sequence and past market conditions could inform the future market behavior.

For the project, a simple RNN model was constructed using several layers of RNN cells, which were then followed by a dense layer to output a prediction for the next day's closing price. The model was trained on the daily percent change of prices, using a sliding window approach to generate input and label pairs.

Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs), traditionally used for image processing, have been effectively adapted for time series forecasting. CNNs utilize convolutional layers, which apply a convolution operation to the input, passing the result to the next layer. This operation allows CNNs to detect patterns or features in input data, making it useful for identifying trends and anomalies in time series.

For time series, CNNs can treat temporal data as one-dimensional spatial data, extracting features from sequences to forecast future values. This makes them particularly useful for analyzing market data, where the pattern recognition capabilities of CNNs can detect underlying trends in price movements.

The CNN model designed for this project includes several one-dimensional convolutional layers, each followed by a pooling layer.

TimeGPT

TimeGPT, developed by Nixtla, is an innovative generative pre-trained model designed specifically for forecasting time series data. This model adapts the transformer architecture, typically used in natural language processing, to handle sequential data in time series forecasting. TimeGPT can generate accurate forecasts by utilizing historical data alone, without the need for retraining for each new dataset. Its versatility allows it to be applied in various domains, including demand forecasting, anomaly detection, and financial forecasting, which is the focus of this project.

The TimeGPT model operates by "reading" time series data in a sequential manner, analogous to how a human reads text from left to right. It processes windows of past data, referred to as "tokens," and uses these to predict future data points. This approach allows TimeGPT to capture and utilize the temporal dependencies and patterns inherent in time series data effectively.

In our implementation, TimeGPT is employed to forecast future values of the S&P 500 index as represented by the SPY ETF. The model ingests sequences of historical price changes, learning from these "tokens" to predict subsequent movements in the index. This predictive capability is derived from the model's ability to discern patterns in the past data and extrapolate these into future predictions. TimeGPT's API provides a streamlined interface for interacting with the model, enabling easy integration into our forecasting framework.

For this project, TimeGPT was configured to evaluate its performance in predicting stock market trends, benchmarked against traditional models like RNN and CNN. This comparative analysis not only highlights TimeGPT's forecasting prowess but also explores its utility in handling complex financial datasets.

Evaluation Metrics

To assess the accuracy and effectiveness of the TimeGPT, RNN, and CNN models in forecasting the S&P 500 index, two primary statistical metrics were employed: Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics are particularly useful in quantifying the difference between the predicted values of closing prices by the models and the actual values of the dataset.

Mean Square Error (MSE)

Mean Squared Error (MSE) is a widely used metric for measuring the quality of a model in regression tasks. It calculates the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value. Mathematically, it is expressed as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

Where Y_i represents the actual values, $\tilde{Y}_i$ represents the predicted values, and n is the number of observations. MSE is sensitive to outliers as it squares the differences; large errors have a disproportionately large effect on MSE. Thus, it is helpful when large errors are particularly undesirable.

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) measures the average magnitude of the errors in a set of predictions, without considering their direction (i.e., it takes the average over the absolute values of the errors). The MAE is formally defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

where $| Y_i - \hat{Y}_i |$ is the absolute error between the actual and the predicted values. MAE provides a straightforward interpretation of overall error magnitude and is not as sensitive to outliers as MSE. This makes MAE a robust measure of model performance, especially in the presence of anomalies.

Results

	RNN	CNN	TimeGPT	CNN (60 days)	TimeGPT (30 days)
MAE	0.809298	0.0176	2.02037679	0.0177	9.333771
MSE	1.056366	0.00046459	0.000013551	0.00047631	162.4185
RMSE	1.027797		0.003681		12.74435

The evaluation of the forecasting models employed in this study—TimeGPT, RNN, and CNN—was conducted using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) for quantitative accuracy assessment, along with the coefficient of determination (R^2) for the RNN model. The performance metrics indicate varied results across different models and forecasting horizons.

In the context of short-term forecasting (3 days), the TimeGPT model demonstrated superior performance with significantly lower error rates, recording an MAE of 2.020 and an MSE of 0.000013551. This performance suggests an effective capability of TimeGPT to capture and predict near-term market movements accurately. Comparatively, the RNN model exhibited poor performance across all metrics, notably a negative R^2 value of -1.436, indicating that the model failed to capture the variance of the dataset meaningfully.

The CNN models, evaluated over unspecified short-term periods, also performed well with very low MAE (0.0176 and 0.0177) and MSE (0.00046459 and 0.00047631) values, indicating strong predictive accuracy. However, TimeGPT outperformed these benchmarks in terms of MSE, highlighting its efficiency in capturing finer details and dynamics of the market over a few days.

For long-term forecasts (30 days), the results for TimeGPT showed a marked increase in error metrics with an MAE of 9.333 and an MSE of 162.4185, suggesting a decrease in predictive accuracy over extended horizons. This outcome highlights a potential limitation of TimeGPT in maintaining forecast accuracy as the prediction horizon extends, a challenge not uncommon in time series forecasting of volatile domains like stock markets.

The CNN and TimeGPT models generally outperformed the RNN model, which struggled with higher error values and a significantly poor fit as indicated by its negative R^2 value. However, while TimeGPT excelled in the realm of short-term forecasting, its long-term predictive capabilities were less reliable, demonstrating a notable decline in performance as compared to its short-term results.

These findings suggest that while TimeGPT is highly effective for short-term financial forecasting, its application in long-term predictions may require additional adjustments or enhancements to improve its efficacy. Conversely, CNN models displayed consistent performance across their evaluations, albeit without the dramatic fluctuations observed with TimeGPT between different forecasting horizons.

Challenges Faced

During this project, several challenges were encountered that impacted the implementation of our study. These challenges not only influenced the project timelines and outcomes but also provided crucial learning opportunities.

One significant challenge was the ongoing updates to the documentation of the TimeGPT model by its developers, Nixtla. Throughout the project, the documentation for TimeGPT was updated multiple times, which occasionally led to disruptions in our workflow. These updates often included changes in the model's API usage, parameter adjustments, and alterations in the recommended best practices for model deployment.

The frequent updates required us to continually adapt our implementation strategies and sometimes revisit completed sections of our project to align with the new guidelines. This not only affected our project timeline but also necessitated additional time for retraining and re-evaluating our models to ensure consistency with the latest recommendations.

Adaptation Strategies:

Regular Monitoring: We established a routine to check for updates on the TimeGPT documentation weekly, allowing us to quickly adjust our plans and code.

Version Control: Maintaining different versions of our models and scripts to ensure that we could revert to previous setups if newer updates conflicted with our project goals.

Communication with Developers: Engaging with the TimeGPT development community through forums and direct communications to clarify uncertainties and receive updates on upcoming changes.

Conclusion

This project embarked on an exploration of the TimeGPT model's capability in forecasting the S&P 500 index, comparing its performance against traditional models such as RNN and CNN. Our findings reveal that TimeGPT demonstrates remarkable accuracy in short-term forecasting, outperforming the benchmark models in this regard. However, its efficacy diminishes over longer prediction horizons, highlighting a potential area for improvement in applications requiring long-term forecasting accuracy.

Key Findings

Short-Term Accuracy: TimeGPT excelled in predicting the S&P 500 index over short spans (3 days), showing lower error metrics compared to both RNN and CNN models. This underscores its potential utility in applications where short-term predictions are critical, such as trading and short-term investment strategies.

Challenges in Long-Term Forecasting: For longer durations (30 days), TimeGPT's performance was less reliable, with significantly higher error rates. This suggests that while TimeGPT is adept at capturing immediate trends, its long-range forecasting abilities are limited by current model configurations or training approaches.

Consistency of CNNs: The CNN models demonstrated consistent performance across different tests, providing a reliable baseline for neural network-based time series forecasting.

This project contributes to the ongoing discourse on the applicability of generative models in financial forecasting. By providing a direct comparison with more traditional approaches, this study highlights both the strengths and limitations of using advanced AI models like TimeGPT in a highly volatile field such as finance.

Future Directions:

Model Improvement: Future research could explore modifications to TimeGPT's architecture or training process to enhance its long-term forecasting capabilities. Investigating hybrid models or ensemble techniques could also provide pathways to leverage the strengths of multiple forecasting approaches.

Expanding Dataset and Features: Incorporating additional data types, such as macroeconomic indicators or alternative data sources, could improve the model's understanding and forecasting of complex market dynamics.

Adaptive Learning: Implementing techniques for adaptive learning, where the model updates its parameters in response to new data without complete retraining, could make TimeGPT more responsive to market changes.

The ability of generative models like TimeGPT to transform financial forecasting is evident from their performance in short-term predictions. As financial markets continue to evolve, the models we rely on must evolve as well to predict their movements. The challenges encountered and knowledge gained from this project will serve as a foundation for further exploration and innovation in financial time series forecasting.