Machine Learning Algorithm comparison between Transformer and LSTM for stock market prediction

STEFANO TONINI JACOPO DALLAFIOR
stonini | jacopoda @kth.se

November 18, 2024

Abstract

The prediction in the stock market is daunting because financial markets are of a volatile and nonlinear nature. The study compares the performance of the Long Short-Term Memory network with the Transformer based models in stock price forecasting. Historical data for Apple (AAPL) and Microsoft (MSFT) were collected, then further preprocessing was made to ensure quality for any missing data. Both models were implemented with great care in tuning hyperparameters and tested on MAPE and RMSE. We further hypothesize that the performance of the LSTM outperforms the Transformer model since this model is much more effective in dealing with long-term dependencies. The result of this will be used to identify which of these two models best suits the prediction of stock prices, which might be helpful to investors, and at the same time plays an important role in the development of the state-of-the-art in financial machine learning.

Contents

List of Acronyms and Abbreviations

AAPL Apple

ARIMA Autoregressive integrated moving average

CNN Convolutional Neural Network

GARCH Generalized AutoRegressive Conditional Heteroskedasticity

GPU Graphics processing unit

LSTM Long Short-Term Memory

MAPE Mean Absolute Percentage Error

ML Machine Learning

MSFT Microsoft

NLP Natural language processing

RMSE Root Mean Square Error

RNN Recurrent Neural Network

TPU Tensor Processing Unit

1 Introduction

It was conjectured in [?] that multicasting could provide gains by . . . .

See also [?], the paper [?], and the book [?].

2 Theoretical framework/literature study

2.1 Introduction

The stock market is characterized by volatility and uncertainty. In recent years, techniques to provide better and more accurate predictions have significantly advanced. Machine learning has gained traction in financial markets due to its ability to predict stock prices more accurately than traditional methods [1].

It is common to define machine learning as a broad category since numerous algorithms are used. According to [2], the most widely used algorithms for stock prediction are Artificial Neural Networks (ANN) and Support Vector Machines (SVM) [3]. Financial institutions prefer neural network-based algorithms due to the high complexity of markets. These techniques are data-driven, self-adaptive, and capable of capturing nonlinear behaviors in time series without relying on statistical assumptions [4].

2.2 LSTM

Long Short-Term Memory (LSTM) networks are designed to overcome the limitations of traditional RNNs, such as the vanishing and exploding gradient problem, by introducing a memory cell structure and gating mechanisms [5]. In general, an Artificial Neural Network (ANN) consists of three layers: the input layer, hidden layers, and the output layer. Connections between layers, called synapses, have associated weights that are iteratively optimized during training to minimize prediction errors. The hidden layers apply activation functions, such as sigmoid or hyperbolic tangent, to transform weighted inputs, while backpropagation adjusts the weights to achieve convergence with a minimized error rate [6].
RNNs extend this concept by incorporating feedback loops, allowing them to utilize earlier sequence data for forecasting future trends. However, they are limited in their ability to store long-term dependencies. LSTM networks address this limitation by using memory cells and gates specifically forget, input, and output gates that regulate the flow of information. The forget gate determines which information is retained or discarded, the memory gate selects new data to be stored, and the output gate decides the final output from the cell. These gates collectively enable LSTM to remember and utilize long-term dependencies effectively, which is crucial for sequential data tasks such as time-series forecasting [7][6].
The internal architecture (see Fig. 1) of an LSTM node incorporates a memory line that maintains past data streams. This architecture allows independent cell states to control the flow of information by either disposing of or retaining values based on sigmoid activation outputs. For instance, the forget gate outputs a value between 0 and 1 to either completely ignore or retain specific information, while the input gate adds candidate data to the cell state using a tanh function. The output gate combines the cell state with updated inputs to produce the final output. This ability to regulate and utilize sequence memory distinguishes LSTMs as a powerful tool for tasks involving complex temporal patterns [6].

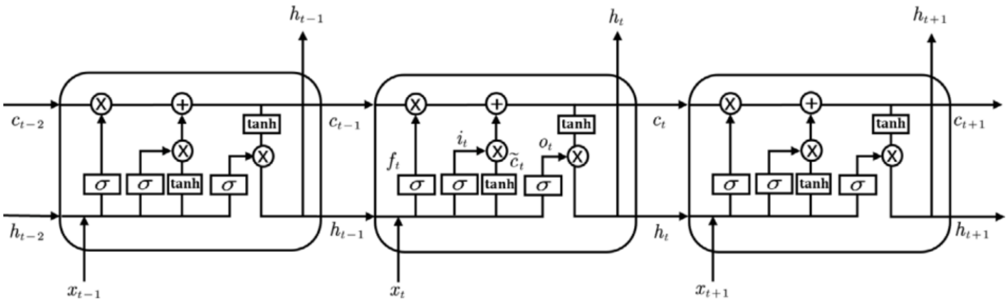

Figure 1: The architecture of an LSTM network [3].

2.3 Transformer Model

The Transformer model represents a significant advancement in deep learning by adopting an attention-based mechanism, replacing the sequential nature of RNNs and LSTMs. This architecture relies on self-attention mechanisms to capture both local and global dependencies within sequences efficiently. Unlike RNNs, Transformers process inputs in parallel, allowing faster training and better performance on long sequences[3].

The Transformer architecture consists of an encoder-decoder structure (see Fig. 2), where each module is composed of multiple layers integrating Multi-Head Attention (MHA) and feed-forward networks. MHA facilitates the model's ability to focus on different parts of the input sequence by applying the scaled dot-product attention mechanism in parallel across multiple "heads," each with its own learnable weights. This allows the network to extract diverse characteristics of the data and combine them into richer representations[8]. The encoder maps input sequences into a continuous representation through positional embeddings and attention layers, while the decoder uses this representation to generate outputs, employing additional masked attention layers to preserve the autoregressive nature of tasks such as translation [8].

A pivotal feature of the Transformer is its exclusive reliance on self-attention mechanisms, which eliminate the need for recurrence or convolutions, addressing challenges like vanishing gradients in sequential models. The self-attention mechanism, implemented via matrices, computes attention scores by scaling the dot product, normalized by the square root of their dimension, followed by a softmax operation. These scores determine the weights for combining values, allowing the model to selectively focus on relevant parts of the input sequence [8].

This architecture enables the Transformer to learn complex dependencies within data and in this way outperforms traditional RNN-based models due to its higher accuracy and enhanced prediction capabilities [3].

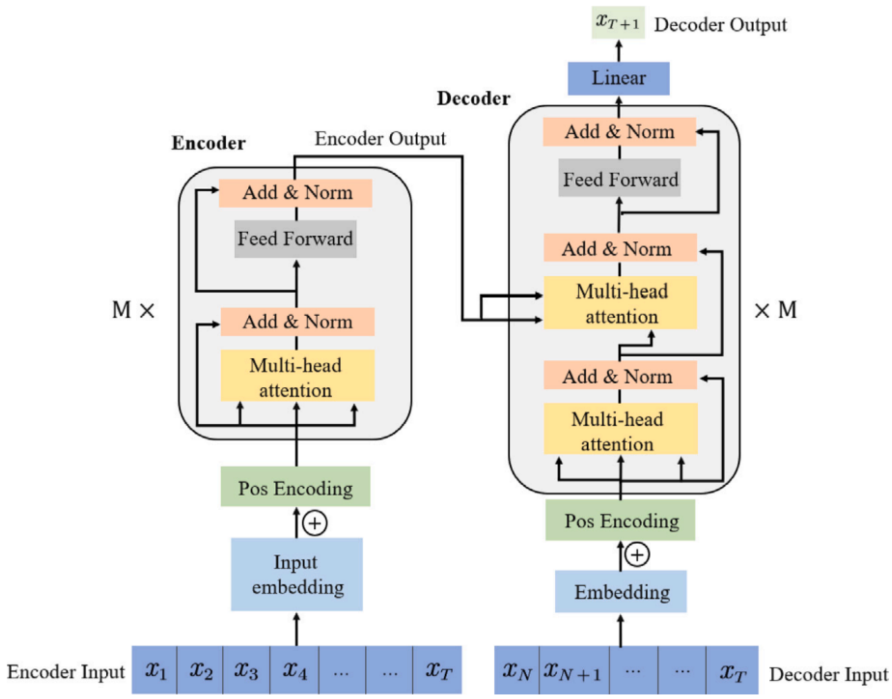

Figure 2: The architecture of an Transformer network [3].

2.4 LSTM in Stock Market Prediction

Long Short-Term Memory (LSTM) networks have become pivotal in stock market prediction due to their capability to model temporal dependencies in financial time series data. Their architecture effectively mitigates the vanishing gradient problem, allowing for the capture of long-term dependencies essential for accurate forecasting. For example, [9] proposed an improved LSTM model based on a Conditional Restricted Boltzmann Machine for stock prediction, demonstrating enhanced performance in capturing temporal features of stock data. Additionally, [10] utilized LSTM models incorporating technical analysis indicators for stock price prediction, highlighting the importance of integrating domain-specific features. These studies collectively emphasize the effectiveness of LSTM networks in handling the complexities of stock market prediction, making them valuable tools for investors and analysts.

2.5 Transformers in Stock Market Prediction

Transformers have gained prominence in the financial sector due to their superior ability to model complex temporal dependencies and integrate diverse data sources. In contrast to traditional time-series models like LSTMs, which process data sequentially, Transformers leverage self-attention mechanisms to analyze entire datasets simultaneously[11]. This capability allows them to identify intricate patterns and long-term relationships in financial data, such as stock prices and market sentiment, which are crucial for accurate forecasting in the volatile and interconnected financial markets [12].

2.6 Final Comparison

3 Research questions, hypotheses

xxxxx xxxx xxxx

4 Method(s)

xxxxx xxxx xxxx

5 Results and Analysis

xxxxx xxxx xxxx

6 Discussion

xxxxx xxxx xxxx

References

[1] "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data: A comparative analysis."

[2] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Computer Science & Engineering Department, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India*.

[3] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, "Stock market index prediction using deep transformer model," *School of Mathematical Science, Jiangsu University, Zhenjiang, China*.

[4] R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, J. P. Nobrega, and A. L. Oliveira, "Computational intelligence and financial markets: A survey and future directions."

[5] Y. Baek and H. Y. Kim, "Modaugnet: A new forecasting framework for stock market index value with an overfitting prevention lstm module and a prediction lstm module," *Department of Financial Engineering, Ajou University, Republic of Korea*.

[6] A. Moghar and M. Hamiche, "Stock market prediction using lstm recurrent neural network," in *International Workshop on Statistical Methods and Artificial Intelligence (IWSMAI 2020)*, IWSMAI. Warsaw, Poland: University Abdelmalek Essaadi, Morocco, April 6-9 2020.

[7] M. Göçken, M. Özçalıcı, A. Boru, and A. T. Dosdoğru, "Integrating metaheuristics and artificial neural networks for improved stock price prediction."

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[9] H. Li, Y. Chen, H. Ma, and F. Liu, "A lstm model based on crbm for stock prediction," *Soft Computing*, 2024. doi: 10.1007/s00500-024-09836-3

[10] K. Rahul and R. Venkatesan, "Lstm-based stock price prediction using technical analysis indicators," in *2023 International Conference on Artificial Intelligence and Machine Learning (ICAIML)*. IEEE, 2023. doi: 10.1109/ICAIML57064.2023.10284713 pp. 1–6.

[11] D. Boyle and J. Kalita, "Spatiotemporal transformer for stock movement prediction," *Proceedings of the Association for the Advancement of Artificial Intelligence*, 2022.

[12] C. Yañez, W. Kristjanpoller, and M. C. Minutolo, "Stock market index prediction using transformer neural network models and frequency decomposition," *Neural Computing and Applications*, vol. 36, pp. 15 777–15 797, 2024. doi: 10.1007/s00521-024-09931-4

A Insensible Approximation

Note that the Appendix or Appendices are Optional.