JOURNAL OF DATA SCIENCE | Vol.2023:21
eISSN:2805-5160

Using Transformer Models for Stock Market Anomaly Detection

Kseniia Biriukova[1], Anol Bhattacherjee[2]

[1,2] University of South Florida, Tampa

Email: [1]xeniatorviana@gmail.com, [2]abhatt@usf.edu

Abstract

Anomaly detection is an important task in financial markets. Detecting anomalies is difficult due to their rarity, multitude of parameters, and lack of labeled data for supervised learning models. Additionally, time series data used in financial models present unique challenges such as irregularity, seasonality, changing trends, and periodicity in data. While prior anomaly detection approaches have used ARIMA and LSTM models, in this paper, we employ a new transformer-based model called TranAD to compare stock market data with its predicted version, measuring deviations from normal price data for anomaly detection. We find that TranAD is an effective approach for financial anomaly detection with a high level of accuracy. We expect that this research will contribute to better detection of financial anomalies and improve market surveillance.

Keywords

Anomaly detection, transformers, financial markets, deep learning

Introduction

An anomaly is defined as "an observation that deviates so significantly from the other observations as to arouse suspicion that it was generated by a different mechanism" (Munir et al., 2019). Anomaly detection or the identification of such observations (outliers) can indicate critical incidents, such as technical malfunctions, a sudden change in consumer preference, and disease diagnosis.

Anomaly detection can sometimes be performed using simple statistical methods. However, this approach is sensitive to irregularity and noise in data and cannot often detect nonlinear patterns (Yu & Yan., 2020). More complex situations, such as when a multitude of parameters should be considered, some of which may be hard to define, or when the anomaly is rare and akin to searching for "a needle in a haystack," may call for complex technological approaches such as machine learning (ML). Two examples of the use of ML in anomaly detection tasks are technical failure identification (Wang et al, 2018) and medical analysis (Han et al, 2021).

In financial markets analysis, identifying observations that differ from earlier ones is called time series anomaly detection. This approach involves setting a baseline for normal behavior, from which deviations can be identified. However, time series data is subject to many challenges such

Submission: 29 November 2023; **Acceptance:** 4 December 2023

as seasonality, changing trends, periodicity, and the presence of random events (Munir et al, 2019), which makes the anomaly detection process particularly difficult (Yu & Yan, 2020). Neural networks, also called deep learning, appear to be better suited for time series data since they are more adaptive and designed to work with errant observations and “polluted” data (White, 1988). One particular deep learning approach designed specifically for sequential data is Long Short-Term Memory (LSTM), which has been shown to work well with time-series data (e.g., Tallboys et al., 2022).

In this paper, we employ a more recent deep learning-based approach based on the transformer architecture for anomaly detection in time-series financial data. This architecture is more computationally efficient and effective than alternate ML approaches because it uses pre-trained models. We examine stock-price data from a basket of mostly technology stocks from Yahoo! Finance to demonstrate the efficacy of our approach for predicting anomalous stock price fluctuations in financial markets.

Related Literature

Research on anomaly detection in financial literature is almost four decades old. In 1988, White (1988) attempted to forecast the price of IBM stock. Subsequent research focused on identifying more effective approaches for stock prediction and comparing these models against real data. For example, Zhang (2003) compared an autoregressive integrated moving average (ARIMA) statistical model with an artificial neural network (ANN) for time series prediction. In this study, ANN outperformed the ARIMA model. Subsequently, ANN and other more advanced deep learning gained prominence in automated trading systems (Vanstone & Finnie, 2009).

One particular deep learning model that has proven effective in time-series anomaly detection is recurrent neural network (RNN) based LSTM approaches. For example, Islam et al. (2018) proposed an LSTM algorithm called ANOMALOUS to detect illegal insider trading from historical stock volume data. In another study, Yu and Yang (2020) showed that an LSTM model outperformed alternative approaches such as SVM, ARIMA, and MLP analysis of stock market data. Subsequently, Generative Adversarial Networks (GANs) have also gained attention in anomaly detection tasks, although GANs have not surpassed LSTM in stock market data analysis (Tallboys et al., 2022).

Previous studies primarily focused on analyzing individual stock prices based on individual trades and/or opening and closing prices. However, this approach is based on historical data and has two limitations. First, it is less useful for detecting anomalies in real time. In contrast, the model we propose in this paper aims to provide real-time market diagnosis and identify divergent assets and corresponding trade hours when anomalies occur. Second, the extensive volume of trading data generated every minute, hour, and day is computationally challenging to analyze. Hence, researchers typically examine shorter periods, typically hours or days, for price prediction. However, this shorter duration may lead to overfitting or may be influenced by seasonal factors and/or unique market conditions. To overcome this limitation, we employ an alternative approach, in which, instead of training the model continuously, we average all trade data for each asset by the hour and train our model on one year of data. We find that this approximation may introduce some inaccuracies but does not significantly impact the anomalies detected.

Methods

In contrast to prior approaches that used ARIMA, LSTM, and other statistical or deep learning approaches, we employ a novel transformer architecture for anomaly detection. Introduced by Vaswani et al. (2017), transformers are attention-based deep learning models that employ an encoder and decoder to convert any input sequence into an output sequence. Transformers are computationally much more efficient than LSTM because they are pretrained on a large corpus of data, and only need to be fine-tuned for specific applications, and their ability to capture global dependencies and contextual information through self-attention mechanisms is well documented in applications such as text summarization (Paulus et al., 2017), image recognition and computer vision (Dosovitskiy et al., 2021), speech recognition (Dong et al., 2018), recommendation systems (Lian et al., 2018), protein folding (Senior et al., 2020), and many more.

One transformer architecture specifically designed for anomaly detection is TranAD (Tuli et al., 2022). This model, illustrated in Figure 1, was employed for anomaly detection in our study. This model was chosen due to its excellent performance in benchmarking anomaly detection tasks. It assumes that the majority of observations would fall within a “normal” range, and thus the few anomalies or outliers that may be present would not cause the model to overfit. To the best of our knowledge, TranAD has previously not been applied to financial market analysis.

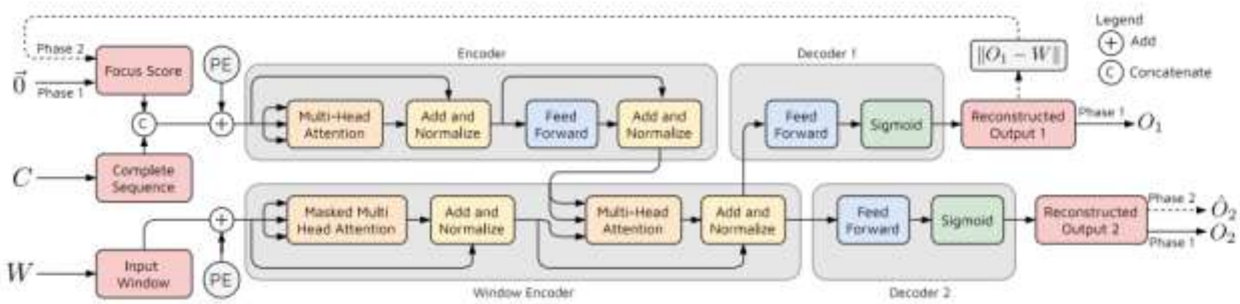

Figure. 1. The TranAD model (Tuli et al., 2022)

The TranAD model builds on transformer architecture by combining two sets of encoders and decoders with multi-head attention. The model learns in two phases. In phase one, an input sequence until the current timestamp is passed to the first encoder, along with a focus score (initially a zero matrix). This is used to generate attention weights, which are then passed along with the input window to the second encoder. This process creates an encoded representation of the input window, which is then used by the first decoder to recreate the input. The deviation from the ground truth is used as a focus score in phase two. In phase two, the same process is used to obtain the output from the second decoder.

It must be noted that while transformers are primarily used for textual data analysis, the TranAD model in question is designed for anomaly detection tasks with numerical data. This is similar to LSTM, which has also been used for both temporal numeric and textual data, Transformers can capture sequential dependencies between observations, which makes this architecture a natural choice for such a task.

We trained the TranAD model using data from Yahoo! Finance, over the period June 2021 to June 2023. Stocks in our sample included Spyder (SPY), Apple (AAPL), Microsoft (MSFT), Tesla (TSLA), Google (GOOG), Nvidia (NVDA), GameStop (GME), Etsy (ETSY), and Disney (DIS). The model incorporated several features of these stocks, including price, volume, and price difference with the last trade.

After the model completed training, each observation was scored based on mean squared error (MSE) between the reconstructed sequence and the true sequence, where "mean" refers to the average hourly price in a rolling 1-hour window. If the MSE of any observation was outside a threshold range, defined by the following equation, it was classified as an "anomaly" (outlier):

$$\text{Threshold} = \text{Mean(MSE)} + 2*\text{Stdev(MSE)}$$

Following Bebee, et al.'s (2021) suggestion that relative strength index (RSI) can also be employed as an additional anomaly indicator, we also used RSI to determine whether an asset is overpriced or underpriced. RSI was computed using the formulae:

$$\text{RS} = \text{Average Gain} / \text{Average Loss}$$
$$\text{RSI} = 100 - 100 / (1 + \text{RS})$$

To facilitate ease of analysis, we grouped the anomalies on a daily basis. For each day, the mean anomaly indicator and the count of anomalies were considered as relevant parameters for further examination and evaluation. The transformer model was implemented in Google Collaboratory, a hosted Jupyter Notebook service, using an Adam optimizer with a learning rate of $3e^{-4}$.

Results

Our transformer model successfully identified anomalous trading data in financial markets. Table 1 shows a sample of anomalies detected for Tesla stock within our examined timeframe (June 2021 to June 2023), along with the respective dates for each anomaly. In addition, we maintained a daily count of anomalies, as shown in Figure 2, which helped us identify days in which we observed a large number of anomalies. Since some of these anomalies could be minor in nature, we also tracked anomaly count by the trading hour in Figure 3, which helped us identify the trading hours that influenced the most anomalies.

Figure 2 suggests several peaks in anomaly count per day. Notably, November 11, 2021, and April 26, 2022, were two of these peaks, which were both connected to Tesla CEO Elon Musk's statements about Tesla's share distribution and Twitter purchase respectively. We also observed that highly anomalous days tend to be surrounded by other anomalous days both before and after those focal dates. This may suggest that significant anomalies tend to occur in groups and are preceded and followed by smaller anomaly occurrences, sometimes resembling a bell shape. This could testify to the general nature of market behavior and anticipation of significant events, or information leaks, as well as aftershocks of an anomaly after its revelation. Although further research is needed, one can assume that smaller anomalies of an increasing frequency may be predictive of larger anomalies.

Table 1. Anomalous events for Tesla stock and their dates

Date	Event
2021-11-09	Tesla down 11.99% after CEO Elon Musk announced over the weekend that he plans to sell 10% of his shares.
2021-11-10	Key inflation report that showed a greater-than-expected jump in consumer prices. Tesla announcement aftershock.
2021-11-22	Elon Musk Tweets About Model S Plaid in China.
2021-11-23	Tesla Model 3 sales announced, LG considers making a battery supply deal with Tesla.
2021-12-01	Stocks rebound after losses due to Omicron worries.
2021-12-06	SEC probes Tesla on whistleblower's claims about solar panel defects.
2022-01-25	Tesla record earnings.
2022-01-26	New Tesla roadmap. No new model announced despite expectations. Investors braced for a FRS rate increase in the coming months. Mixed corporate earnings results.
2022-04-26	Twitter agrees to Elon Musk's purchase deal.
2022-05-05	Elon Musk confirmed plans to serve as CEO of Twitter (TWTR).
2022-05-12	The day before Musk said he would let Trump back on Twitter. Musk announced the deal to acquire Twitter was "temporarily on hold" on May 13, citing pending details to support the microblogging site's claim that spam or fake accounts were less than 5% of its total user bas
2022-07-11	The next day: Twitter sues Elon Musk in the Delaware Court of Chancery, requesting the court to order Musk to proceed with the purchase of Twitter under the previously agreed upon terms.
2022-10-14	Wells Fargo expects Tesla to be the top beneficiary of the electric vehicle incentives in the Inflation Reduction Act.

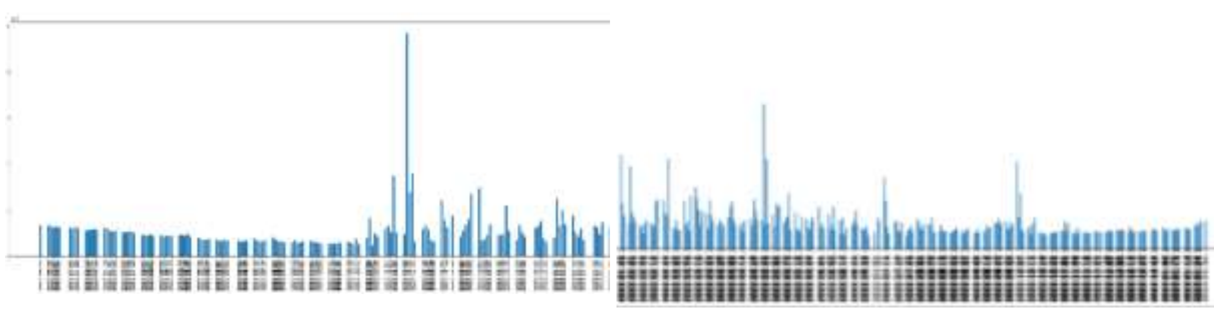

Figure 2. Count of anomalies per day

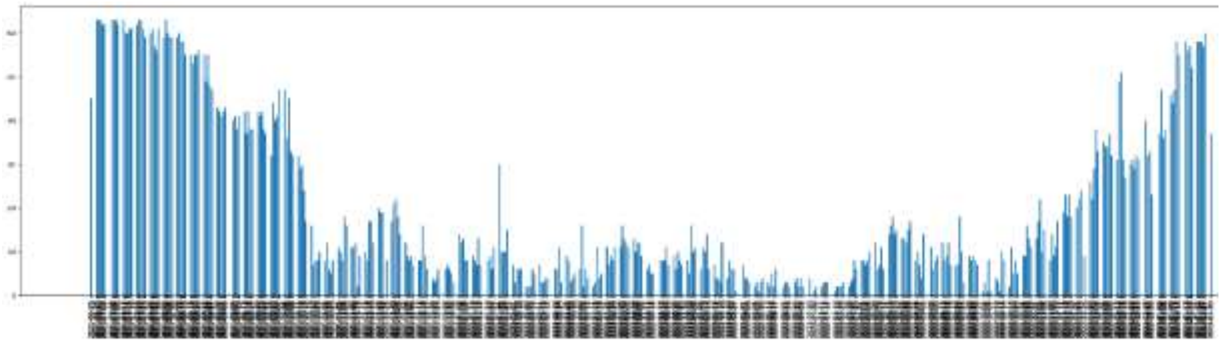

Figure 3. Count of anomalous hours per each day

Model Evaluation

Evaluating TranAD's performance or comparing it with alternative approaches such as ARIMA or LSTM is difficult because of the lack of labeled data. Tuli et al. (2022) demonstrated the model's efficiency in anomaly detection tasks. Our evaluation for anomaly prediction in financial markets was based on an examination of news events that coincided with the anomaly peaks. We illustrated this approach in the previous section where we matched two peaks in Tesla;s stock performance with Tesla's share distribution and Twitter purchase events. However, a more formal evaluation of TranAD's performance, in terms of recall, precision, F1 score, and area under the ROC curve (AUC), and comparing those metrics with that from LSTM or an alternate approach, using labeled data is an opportunity for future research. However, given that the transformer architecture uses pre-trained models, it is likely that this TranAD approach will be computationally more efficient than models that are not pretrained such as LSTM or ARIMA.

Discussion and Conclusions

In this paper, we presented an alternate approach for anomaly detection in financial markets using a transformer-based approach. This model, based on Tuli et al.'s (2022) TranAD architecture, accurately detected market anomalies, in accordance with known events. We analyzed only a small set of mostly technology stocks, but the results look promising. It remains to be seen whether this anomaly prediction approach works equally well for non-technology stocks or a broader basket of stocks or other financial assets such as cryptocurrencies, as well as how this approach compares with alternative approaches such as LSTM or ARIMA.

We anticipate that this work will serve as the initial step in a series of research efforts aimed at improving anomaly detection in financial markets. We expect our proposed approach to be computationally more efficient than alternative approaches. At the very least, this approach represents a practical addition to our existing toolbox of techniques for anomaly detection in financial markets.

References

Bebee, T., & Sabater, D. (2021). Detect anomalies in real time forex data with ML | Google Cloud Blog. Google. https://cloud.google.com/blog/topics/financial-services/detect-anomalies-in-real-time-forex-data-with-ml

L. Dong, S. Xu and B. Xu, "Speech-Transformer: A No-Recurrence Sequence-to-Sequence Model for Speech Recognition," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018, pp. 5884-5888, doi: 10.1109/ICASSP.2018.8462506.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Han, C., Rundo, L., Murao, K. et al. (2021). MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. BMC Bioinformatics 22 (Suppl 2), 31. https://doi.org/10.1186/s12859-020-03936-1

Islam, S.R., Khaled Ghafoor, S., Eberle, W.(2018): Mining Illegal Insider Trading of Stocks: A Proactive Approach. 2018 IEEE International Conference on Big Data (Big Data). pp. 1397{1406

Lian, J., Zhou, X., Zhang, F., Chen, Z., Xie, X., & Sun, G. (2018, July). xdeepfm: Combining explicit and implicit feature interactions for recommender systems. In Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 1754-1763).

Munir, M., Chattha, M.A., Dengel, A., Ahmed, S. (2019). A Comparative Analysis of Traditional and Deep Learning-Based Anomaly Detection Methods for Streaming Data. 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). pp. 561{566

Paulus, R., Xiong, C., & Socher, R. (2017). A deep reinforced model for abstractive summarization. arXiv preprint arXiv:1705.04304.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T. & Kohli, P. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710.

Tallboys, J., Zhu, Y., Rajasegarar, S. (2022). Identification of Stock Market Manipulation with Deep Learning. Advanced Data Mining and Applications. ADMA 2022. Lecture Notes in Computer Science(), vol 13087. Springer, Cham. https://doi.org/10.1007/978-3-030-95405-5_29

Tuli, S., Casale, G., & Jennings, N. R. (2022). Tranad: Deep transformer networks for anomaly detection in multivariate time series data. arXiv preprint arXiv:2201.07284.

Vanstone B., Finnie G. (2009) An empirical methodology for developing stockmarket trading systems using artificial neural networks. Expert Syst Appl 36(3):6668–6680

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Wang, Z., Pi, D., Gao, Y. (2018). A novel unsupervised time series discord detection algorithm in aircraft engine gearbox. Advanced Data Mining and Applications, Springer International Publishing, Cham. pp. 202{210.

White, H, (1988). Economic prediction using neural networks: the case of IBM daily stock returns, Department of Economics University of California, San Diego.

Yu, P., Yan, X. Stock price prediction based on deep neural networks. Neural Comput & Applic 32, 1609–1628 (2020). https://doi.org/10.1007/s00521-019-04212-x
Zhang, G.P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, p. 159–175