2019 IEEE International Conference on Big Data (Big Data)

Stock Prediction using Deep Learning and Sentiment Analysis

Yichuan Xu
Computer Science
Dalhousie University
Halifax, Canada
y.xu@dal.ca

Vlado Keselj
Computer Science
Dalhousie University
Halifax, Canada
vlado@dnlp.ca

***Abstract*—Stock prediction has been a popular research topic and researchers have done a lot of work in this field. Due to its stochastic nature, predicting the future stock market remains a very difficult problem. This paper studies the application of attention-based LSTM deep neural network in future stock market movement prediction. We also build stock aggregate dataset and individual dataset including stock history data, financial tweets sentiment and technical indicators in the US stock market. The experiment studies the time sensitivity of finance tweet sentiment and methods of collective sentiment calculation. This paper also experiments on conventional LSTM and attention-based LSTM for performance comparison. We find the finance tweets that are posted from market closure to market open in the next day has more predictive power on next day stock movement. The weighted sentiment on max follower on StockTwits also outperforms other methods. In our experiment, the result on our individual stock dataset shows a similar pattern like normal distribution.**

***Index Terms*—Deep Learning, Sentiment Analysis, Stock prediction, LSTM, attention-based LSTM, Technical Indicators**

I. Introduction

Being able to predict the future stock market movement is a dream of every investor. In stock market prediction, fundamental analysis and technical analysis are the two major techniques. Researchers put much effort in this field, but it remains a difficult problem. Coming from countless sources and news, the amount of information is huge and very difficult to process. In the past decades, there has been a lot of research in a variety of fields trying to tackle this problem and proposed various approaches.

The recent progress made in the Machine Learning area motivated many researchers and brought the focus back onto this field. In the past three decades, there has been a thriving evolution of artificial intelligence from multilayer perceptron [1] to deep neural networks (DNNs), like convolutional neural network (CNN) [2] and long-short term memory (LSTM). The financial sector is without doubt one of the most popular fields where people see a potential to successfully apply DNNs. Bollen *et al.* [3] used twitter sentiment for stock movement prediction; Chen *et al.* [4] used LSTM and basic stock market data including Open, Close, High, Low (OCHL) prices in the stock market prediction and found improvement over tradition methods. Nelson *et al.* [5] experimented on LSTM and stock technical indicators derived from OCHL, which outperforms tradition methods with few exceptions.

However, the result has shown that the new techniques are still not capable of achieving satisfying accuracy. The stock market itself is adapting new technologies. To explore this problem further, we built a dataset that includes the OCHL prices, finance tweets with sentiment and technical indicators, and experiment on attention-based LSTM model to predict future stock price movement.

We collect historic stock data from Yahoo Finance of 80 stocks in the US market, all of which are chosen from top of the market cap in different sectors. The sentiment is based on finance tweets from StockTwits, one of the most popular social platforms for finance and investment. For hundreds of finance tweets per day, we take a different approach to calculate collective sentiment. Since StockTwits started collecting sentiment score since 2017, in our experiment we adopt the sentiment score from them.

As many papers have discussed, the advantage of LSTM over traditional approach, the objective of this paper is to study an attention-based LSTM variant, and whether the combination of finance tweets sentiment and stock technical indicator gains better performance from

978-1-7281-0858-2/19/$31.00 ©2019 IEEE

this modified LSTM. By predicting stock rise and fall in the next trading day, we also experiment whether the finance tweets posted during intraday, after hours or full day data can have different impact on the prediction accuracy. The result from the previous experiment lead us to explore whether the trained model works for all our picked stocks, or each stock has different price distribution which requires us to train the model separately.

The main contribution of this paper are as follows: (1) a new stock trend prediction model on attention-based LSTM trained on a dataset including stock history, finance tweets sentiment and technical indicators; (2) a comparison of finance tweets posted during different periods of time in a trading day; (3) evaluation of the model under different test cases: after-hour finance tweets and weighted on maximum followers giving best predictive power; and (4) test result that resembles Gaussian Distribution in individual stock dataset brings new interesting topics.

The result shows great potential using financial tweet sentiment and technical indicators compared with non-sentiment and non-technical dataset.

The remainder of this paper is organized as follows: Section II presents an overview of previous work in stock market and deep learning, followed by the problem formulation in Section III. Methods and development are demonstrated in Section IV. The evaluation is made in Section V and empirical results in Section VI. Summary is in Section VII.

II. Background and Related work

Stock price prediction methods generally fall under two categories: Fundamental analysis (FA) and technical analysis (TA). FA is focused on the intrinsic value of a stock by looking at the economic factors, such as revenues, debt, growth rate etc.. FA takes a broader view of a company and considers long term perspective. They believe the return takes time to realize its intrinsic value.

Technical Analysis, on the other hand, concentrates on stock price and tools that were derived from stock price. Due to the sensitivity over the history data, TA is usually considered as an approach for short-term to mid-term investment. Investors use technical indicators to help predicting trend of a stock or index. Common technical indicators include moving average (MA), relative strength index (RSI) and moving average convergence/divergence (MACD). Zhai *at al.* [6] show that the combination of News and Technical Indicators has positive influence on profit than using news or price alone.

In Figure 1, the purple line and green line are MA65 and MA200 respectively. To explain this trend, the stock price is moving higher until it reaches top. Then it starts to pull back and when it reaches the support of MA65 line it bounced back a little bit, but eventually falls below MA65. Floor is also called *support point,* and ceiling called *resistance point.* When the prices drop below floor which is the MA65 line, the floor becomes ceiling. Then the price fluctuates between the MA65, which is now the resistance line and MA200, which is the support line. When the price finally jumps above MA65, it becomes the support line again and the stock starts rallying. Even when few corrections happen, MA65 shows great support for the price rally.

The constant debate on the predictive power of stock returns has been driving research in the past decades. According to Efficient Market Hypothesis (EMH) [8], stock market prices are mostly affected by new information rather than present and history prices. But early research [3] showed using twitter mood to predict the stock market did provide enhancement compared with non-sentiment methods. The classic methods are mostly based on feature engineering [9]. With DNNS drawing much more attention in past years, CNN based method [10] and LSTM [11, 12] based models were able to use larger datasets from text (e.g., news and twitter) and history stock price to produce better results.

Recurrent neural networks (RNNS) were applied in many time-series data problems like speech recognition, Natural Language Processing (NLP). However, vanishing gradient [13] and exploding gradient [14] problems makes it very difficult to train on RNNs especially on long time steps. The appearance of LSTM, which is used in this project, solved many tasks that were not solvable by previous learning algorithms for RNNS [15] by introducing a 'memory cell' that can memorize information in its cells for a long periods of time.

Instead of using text as input for this model, we use OCHL data, collective sentiment and technical indicators to feed the neural network. Furthermore, we want to answer the questions that are mentioned in our experiment.

The application of Attention [16] in NLP is one of the most exciting breakthroughs in past few years. In NLP especially Neural Machine Translation (NMT), the performance of conventional RNN tends to decrease when the input sequence increases, but Attention model could still maintain a good performance. The attention layer does a 're-scan' of the input and extract useful information that has more connection to the target.

Fig. 1. AMD Yearly Chart [7] MA65 denoted by the purple line and MA200 denoted by the green line

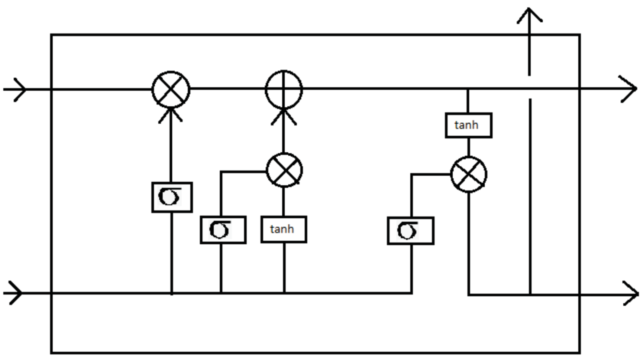

Fig. 2. LSTM cell [11]

III. Problem Formulation

In this paper we are looking at the following questions:

- Is there time-sensitivity in finance tweets with respect to stock market prediction and is it measurable?
- If finance tweets are time-sensitive, in which time period of the day the finance tweet sentiment has more predictive power?
- How does attention-based LSTM perform in comparison with conventional LSTM?
- Does a common model work for all stocks or we have to train a model for each individual stock?

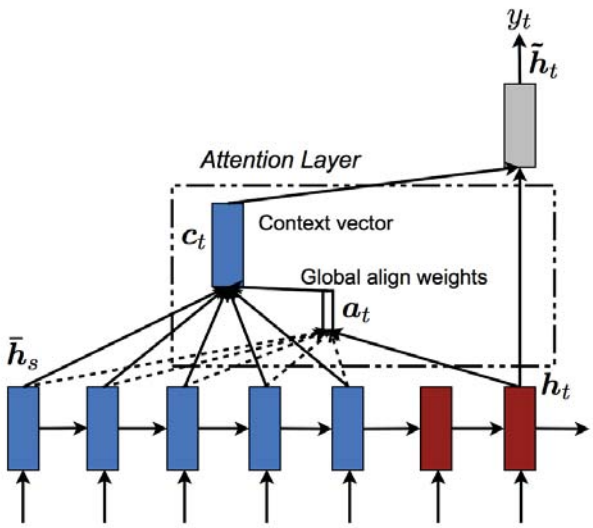

Fig. 3. Attention Model [17]

IV. Methods and Development

In stock market, typical trading day data usually include Open, Close, High, Low (OCHL), and volume. On top of that Bollen *et al.* [3] performed experiments on Twitter sentiment and the result showed that the public sentiment can significantly improve the accuracy of the most basic models in prediction of DJIA closing values.

The classification model that we apply is based on the LSTM and attention mechanism, and the goal is to predict the direction of stock price movement for the next trading day. In other words, it attempts to predict if the target stock will rise or fall on the next trading day. The model is trained on the history data ranging from 11/5/2017 to 31/12/2018.

A. Data Collection

Based on the findings in the previous work, there are 37 features in the dataset including OCHL, volume, technical indicators and finance tweet sentiment. We use ta-lib to generate technical indicators and some indicators are listed in Table I. Ta-lib is a wildly used tool in trading software development. It integrates methods to calculate over 150 technical indicators and run candlestick pattern recognition. We use the python wrapper of ta-lib to process our technical indicator dataset.

Daily stock price data are collected from Yahoo Finance from 11 industries: Basic Materials, Communication Services, Consumer Cyclical, Consumer Defensive, Energy, Financial Services, Health care, Industries, Real Estate, Technology and Utilities[1]. We collect 80 stocks in total from each sector. All stocks are top-market-cap companies within its own sector, including Apple (AAPL), Amazon (AMZN), T (AT&T) and so on. We collect history data of these stocks from 01/01/2016 to 31/12/2018 and the daily price includes Open, Close, High, Low and Volume.

For finance tweets we acquire data from a social media company StockTwits. We collect the bulk data from StockTwits from 01/01/2016 to 31/12/2018. Each bulk data file is a monthly backup in the raw file format. Each row of data is a JSON object for one finance tweet. We parse all data into objects and stored them in a relational database[2]. Instead of our target stocks, the bulk data contain all finance tweets posted on StockTwits, which is much more than what we need. The data query for target stocks is very time consuming on a 150GB database, so we remove all other irrelevant stock data and only kept the data of target stocks. One issue with the StockTwits data that stock tweets were not collected consistently in a daily fashion until 11/05/2017.

B. Technical Indicators

Moving Average (MA) is a widely used indicator in technical analysis that helps smooth out price action by filtering out the “noise” from random short-term price fluctuations [18]. While there are Simple Moving Average and Exponential Moving Average, MA usually

[1]https://finance.yahoo.com/industries

[2]Microsoft SQL Serer database

TABLE I
EXAMPLES OF TECHNICAL INDICATORS

AD	Chaikin A/D Line
ADX	Average Directional Movement Index
EMA	Exponential Moving Average
KAMA	Kaufman Adaptive Moving Average
MA	Moving Average
MACD	Moving Average Convergence/Divergence
RSI	Relative Strength Index
SAR	Parabolic SAR
SMA	Simple Moving Average

refers to Simple Moving Average. In this paper, we will use MA to refer to Simple Moving Average.

Given a trading day d, MA_{50}, MA_{65} and MA_{200} are calculated as follows:

$$MA_k = \frac{\sum_{d-k}^{d} p_i}{k} \tag{1}$$

where k is the number of days in a time window ending with the day d, and p_i is the closing price after each day.

Examples of many other technical indicators are in Table I.

C. Collective Sentiment

The US stock market usually opens at 9:30 a.m. EST and closes at 4:00 p.m. EST for transactions. However, the pre-market trading and after-market trading also affect the movement of stock price. The pre-market [19] trading usually occurs from 8:00 a.m. to 9:30 a.m. EST and after-market from 8:00 a.m. to 9:30 a.m. each trading day. Activities during those periods may affect the stock price dramatically. For example, some companies would release fiscal report or make big announcement after market closes, which sometime results in huge price hike or price plunge.

Bolen *et al.* [3] state that tweets do help predicting the stock price future movement. We want to see if finance tweets in certain time period may have better predictive power; e.g., intraday tweets, after-market tweets and full-day tweets.

Intraday tweets refer to the tweets that are posted during the trading hours; after-market tweets refer to the tweets that are posted from market closes till before market opens in the next trading day; full-day tweets are tweets that are posted in the past 24 hours before the market closes on a target trading day. For the time period of finance tweets, we define three categories: full day, intraday and after hours.

TABLE II
TEST CASES

Tweets Time Period	Sentiment Score
Full Day	Simple Sum
Full Day	Weighted on Max Followers
Full Day	Weighted on Total number of Followers
Intraday	Simple Sum
Intraday	Weighted on Max Followers
Intraday	Weighted on Total number of Followers
After hours	Simple Sum
After hours	Weighted on Max Followers
After hours	Weighted on Total number of Followers

TABLE III
RESULT FOR MSFT

Test Case	Accuracy
Full day SimpleSum	53.33%
Full day Max Followers	52.00%
Full Day Total Followers	58.76%
Intraday Simple Sum	54.67%
Intraday Max Followers	55.44%
Intraday Total Followers	61.33%
After hours Simple Sum	57.44%
After hours Max Followers	**63.78%**
After hours Total Followers	57.33%

However, each tweet may have different influence and predictive power. For instance, my tweet about the market may go unnoticed while Donald Trump's tweet about tariff may cause the market fluctuate dramatically [20]. To calculate the collective sentiment, we compared three different approach to calculate daily collective sentiment C for finance tweets T in target time period:

- Simple summation of tweets sentiment:

$$C = \sum_{i=0}^{n} T_i \tag{2}$$

- Weighted sentiment on tweets followers F for each tweet T:

$$C = \sum_{i=0}^{n} \frac{T_i \cdot F_i}{F_{max}} \tag{3}$$

- Weighted sentiment on total number of followers:

$$C = \sum_{i=0}^{n} \frac{T_i \cdot F_i}{\sum_{j=0}^{n} F_j} \tag{4}$$

In our test cases (Table II), we control variants on tweets time period and sentiment score methods(Table). The first test group uses full day finance tweets dataset and we compare the performance among three collective sentiment approach; the second test group uses intraday finance tweets and the third uses after-hours finance tweets.

To experiment on previous test cases, we train our model on three stocks: Microsoft (MSFT), XPO logistics (XPO) and AMD (AMD). The dataset is collected ranging from 05/10/2017 to 12/31/2018 when StockTwits started collected sentiment score. Then we use the configuration with best result to train on the aggregate dataset of 80 stocks with LSTM and attention-based LSTM model. Eventually we use the superior DNN model and train on each stock separately to compare the difference with overall accuracy and individual stock accuracy.

V. EVALUATION

To evaluate the model performance, we adopt the standard measure of accuracy and Matthews Correlation Coefficient (MCC), following previous work by Xu *et al.* [12]. MCC [21] is used to measure the quality of binary classifications. In a confusion matrix $\begin{pmatrix} Tp & Tn \\ Fp & Fn \end{pmatrix}$, including the number of true positives, true negatives, false positives and false negatives, MCC is calculated as follows:

$$MCC = \frac{Tp \cdot Tn - Fp \cdot Fn}{\sqrt{(Tp + Ff)(Tp + Fn)(Tn + Fp)(Tn + Fn)}} \tag{5}$$

The value of MCC is ranging from -1 to $+1$ where $+1$ represents a perfect prediction, 0 no better than random prediction and -1 shows total disagreement between prediction and observation [22].

VI. EMPIRICAL RESULTS

The results from Table III, Table IV and Table V show that the after-hours and weighted-on-max-followers configuration has overall the best predictive power in our test cases. In other words, the finance tweets posted from market closes till market opens next day has more predictive power in predicting the next-day market movement.

Following the previous result, we use after-hours and weighted-on-max-followers configuration and test on aggregate dataset that contains all 80 stocks. The model we use is the conventional LSTM and time-step is set at 40.

TABLE IV
RESULT FOR XPO

Test Case	Accuracy
Full day SimpleSum	45.33%
Full day Max Followers	54.56%
Full Day Total Followers	52.00%
Intraday Simple Sum	52.00%
Intraday Max Followers	54.67%
Intraday Total Followers	57.33%
After hours Simple Sum	53.33%
After hours Max Followers	**58.67%**
After hours Total Followers	54.67%

TABLE V
RESULT FOR AMD

Test Case	Accuracy
Full day Simple Sum	54.67%
Full day Max Followers	52.33%
Full Day Total Followers	53.47%
Intraday Simple Sum	49.33%
Intraday Max Followers	48.00%
Intraday Total Followers	48.00%
After hours Simple Sum	52.33%
After hours Max Followers	**56.00%**
After hours Total Followers	50.67%

The result of conventional LSTM model trained on aggregate dataset (Table VI) does not perform as good as individual stock dataset. This result raises a question for us whether a model can learn latent knowledge from an aggregate dataset.

To explore further, first we add attention block to the LSTM model and retrain the aggregate dataset. The result from our attention-based LSTM model shows a moderate improvement over traditional LSTM model.

TABLE VI
RESULT FOR AGGREGATE DATASET

Test Case	Accuracy	MCC
After hours Max Followers	52.27%	0.04092

TABLE VII
ATTENTION-BASED LSTM COMPARISON

Model	Accuracy	MCC
LSTM	52.27%	0.04092
attention-based LSTM	**54.58%**	**0.04780**

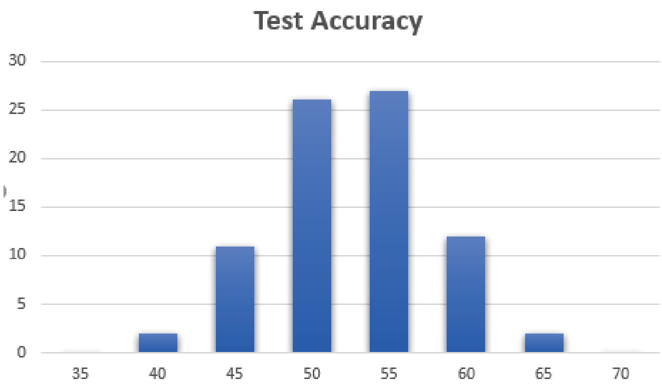

Fig. 4. The distribution of prediction accuracy where X-axis denotes accuracy and Y-axis denotes frequency

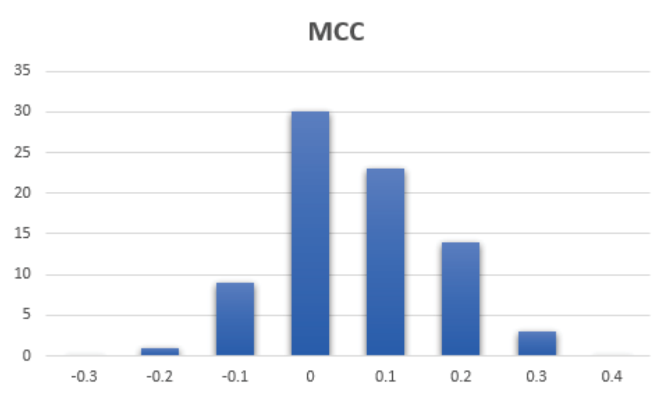

Fig. 5. The distribution of MCC where X-axis denotes accuracy and Y-axis denotes frequency

Based on the result from Table VII, we use the superior attention based LSTM model and train models on 80 stocks separately. To reduce experimental errors, we train five times for each dataset and take the average accuracy and MCC.

From the result in Figure 4, we find an interesting observation: In the histogram, the distribution of the accuracy results almost looks like a Gaussian Distribution. In Figure 5, the distribution of MCC result looks similar, but it leans towards positive side. The results are unexpected but very interesting.

VII. CONCLUSIONS

In this paper we experiment on the performance of DNN with dataset that is combined with finance tweets sentiment, stock history price and stock price technical indicator. We find the finance tweets that are posted from market closure till market open in the next day has more predictive power on next day stock movement.

We also notice that the outcomes of attention-based LSTM model have improvement over conventional LSTM on aggregate dataset. In individual stock dataset, the results are very interesting. The best result we get is near 65%, which is a decent result. However, the distribution of accuracy in Figure 4 is very similar to Gaussian Distribution and it raises a lot more interesting questions to be answered.

A. Forthcoming Research

Given the difference of results in aggregate stock dataset and individual stock dataset, we intend to further investigate the reason behind the scene. Since the attention-based LSTM model shows better performance in some stocks, whether there are some similarities among those stocks, such market cap, industry and their products.

In addition to that, the shape of the results that resembles Gaussian Distribution is worth further research. As central limit theorem defines, in some situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution [23]. Although there has been much progress made in this paper about the application of DNNs in investment, we can not ignore the possibility that the DNN model in stock market prediction learns nothing but makes random guess. Or the stock dataset we chose happen to show this result. Thus, we intend to collect more stocks and technical indicators for further investigation.

VIII. Acknowledgement

This research was supported by an NSERC grant at the Dalhousie University. We also appreciate the help from StockTwits for the access to finance tweets on their platform.

References

[1] Matt W Gardner and SR Dorling. Artificial neural networks (the multilayer perceptron) — A review of applications in the atmospheric sciences. *Atmospheric environment*, 32(14-15):2627–2636, 1998.

[2] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[3] Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[4] Kai Chen, Yi Zhou, and Fangyan Dai. A LSTM-based method for stock returns prediction: A case study of China stock market. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 2823–2824. IEEE, 2015.

[5] David MQ Nelson, Adriano CM Pereira, and Renato A de Oliveira. Stock market's price movement prediction with lstm neural networks. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1419–1426. IEEE, 2017.

[6] Yuzheng Zhai, Arthur Hsu, and Saman K Halgamuge. Combining news and technical indicators in daily stock price trends prediction. In *International symposium on neural networks*, pages 1087–1096. Springer, 2007.

[7] Yahoo. Advanced micro devices, inc. (AMD). https://finance.yahoo.com/chart/AMD. Accessed: 2019-05-30.

[8] Burton G Malkiel and Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

[9] Robert P Schumaker and Hsinchun Chen. Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2):12, 2009.

[10] Ehsan Hoseinzade and Saman Haratizadeh. CNNPred: CNN-based stock market prediction using several data sources. *arXiv preprint arXiv:1810.08923*, 2018.

[11] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[12] Yumo Xu and Shay B Cohen. Stock movement prediction from tweets and historical prices. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1970–1979, 2018.

[13] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02):107–116, 1998.

[14] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. Understanding the exploding gradient problem. *CoRR, abs/1211.5063*, 2, 2012.

[15] Felix A Gers, Douglas Eck, and Jürgen Schmidhuber. Applying lstm to time series predictable through time-window approaches. In *Neural Nets*

WIRN Vietri-01, pages 193–200. 2002.

[16] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[17] Lilian Weng. Attention? attention! https://lilianweng.github.io/lil-log/2018/06/24/attention-attention.html. Accessed: 2019-08-01.

[18] Moving average (ma). https://www.investopedia.com/terms/m/movingaverage.asp. Accessed: 2019-05-30.

[19] Pre-market. https://www.investopedia.com/terms/p/premarket.asp. Accessed: 2019-05-30.

[20] Elena Popina and Vildana Hajric. Stock traders have theories about the timing of Trump's tariff tweet. *Bloomberg*.

[21] Brian W Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451, 1975.

[22] Wikipedia. Matthews correlation coefficient. https://en.wikipedia.org/wiki/Matthews_correlation_coefficient. Accessed: 2019-05-30.

[23] Murray Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43, 1956.