Applied attention-based LSTM neural networks in stock prediction

Li-Chen Cheng, Yu-Hsiang Huang
Department of Computer Science and Information Management
Soochow University
Taipei, Taiwan
lijen.cheng@gmail.com

Mu-En Wu
Department of Information and Finance Management
National Taipei University of Technology
Taipei, Taiwan
mnasia1@gmail.com

***Abstract*—Prediction of stocks is complicated by the dynamic, complex, and chaotic environment of the stock market. Many studies predict stock price movements using deep learning models. Although the attention mechanism has gained popularity recently in neural machine translation, little focus has been devoted to attention-based deep learning models for stock prediction. This paper proposes an attention-based long short-term memory model to predict stock price movement and make trading strategies**

Keywords—deep learning; stock prediction; attention mechanism;

I. INTRODUCTION

Stock predictions have been the object of study for decades. A common approach is to use machine learning algorithms to predict future prices by learning from historical data. Here we proceed in that direction but study a specific method using attention-based recurrent neural networks (RNNs).

RNN have proven successful in sequential data. Traditional RNNs, however, suffer from the problem of exploding and vanishing gradients and thus do not effectively capture long-term dependencies. Recently, the long short-term memory architecture has overcome this limitation. The attention mechanism, another powerful neural networking technique to solve this problem, is widely used in various works. Yang et al. use attention networks to locate objects and concepts referred to in visual question answering [1]. Luong et al. extract informative nodes to form sentence vectors by using a generalization of the sequential model [2].

This study focuses on two issues. We use historical stock data and technical indicators to predict future stock price movement by using an attention-based long short-term memory model, after which we use these predictions to develop trading strategies..

II. LITERATURE REVIEWS

Neural networks have drawn significant interest from the machine learning community, especially given their recent empirical success. The success of neural networks again brings up the theoretical question: why are multi-layer networks better than one-hidden-layer networks?

Ronen and Ohad prove that a function expressible by a 3-layer feedforward network cannot be approximated by a 2-layer network [3]. Matus also proves that a function expressible by a deep feedforward network cannot be approximated by a shallow network [4]. Deep networks can exploit in their architecture the special structure of compositional functions, whereas shallow networks are blind to this [5]. State-of-the-art deep learning techniques include deep neural networks (DNNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs).

Stock market prediction is traditionally one of the most challenging time series prediction tasks, as stock data is noisy and non-stationary. Stock return predictions are broadly classified into linear and non-linear models: linear autoregressive integrated moving average models, exponential smoothing models, and the generalized autoregressive conditional heteroscedasticity model. Non-linear models include SVMs, genetic algorithms, and state-of-the-art deep learning. Akita build an LSTM model using textual and numerical information to predict ten company's closing stock prices [6]. Ding propose a deep convolutional neural network using event embeddings which combines the influence of long-term events and short-term events to predict stock prices [7]. Nelson propose a model based on LSTM using five historic price measures (open, close, low, high, volume) and 175 technical indicators to predict stock price movement [8]. Fischer apply an LSTM model to a large-scale financial market prediction task on S&P 500 data from December 1992 until October 2015. They show that the LSTM model outperforms the standard deep net and traditional machine learning methods [9].

The concept of attention has gained popularity recently in neural networks, as it allows models to learn algorithms between different modalities, e.g., between image objects and agent actions in dynamic control problems [10], between speech frames and text in speech recognition [11], or between visual features of a picture and its text description in image caption generation [12]. Liu predict stock price movements using a novel end-to-end attention-based event model. They propose the ATT-ERNN model to exploit implicit correlations between world events, including the effect of event counts and short-term, medium-term, and long-term influence as well as the movement of stock prices [13]. Zhao capture market dynamics from multimodal information (fundamental indicators, technical indicators, and market structure) for stock return prediction by using an end-to-end market-aware system. Their market awareness system leads to reduced error and temporal awareness across stacks of market images leads to further error reductions [14]. Song forecast time series using a novel dual-stage attention-based recurrent neural network (DA-RNN) which consists of an encoder with an input attention mechanism and a decoder with a temporal attention mechanism [15].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE
978-1-5386-5035-6/18/$31.00 ©2018 IEEE

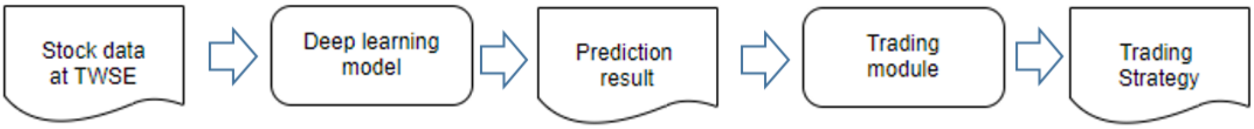

Fig 1 Proposed method

III. PROPOSED FRAMEWORK

This study proposes an attention-based long short-term memory model to predict stock price movement. The proposed framework contains several modules and is illustrated in Figure 1.

A. Dataset

We collect stock price data (Open, Close, Low, High, Volume) from the Taiwan Stock Exchange Corporation (TWSE), and calculate technical indicators (KD, MA, RSV, etc.)

B. Deep learning module

we relay our data to an attention-based long short-term memory model. Long short-term memory networks enable RNNs to capture very long dependencies among input signals [16]. For this purpose, LSTMs still process information sequentially, but introduce a cell state C_i which remembers and forgets information, similar to a memory [17]. This cell state is passed onward, similar to the state of an RNN. However, the information in the cell state is manipulated by additional structures called gates. The LSTM has three gates: an input gate, an output gate, and a forget gate. We add an attention mechanism to the LSTM neural network model as an effective way to enhance the neural network's capability and interpretability. The input to our model is a sequence of stock data including price data and technical indicators. The output of our model is multi-class, where Class 0 represents a stock price increase of more than 3%, Class 1 an increase of 2% to 3%, Class 2 an increase of 1% to 2%, Class 3 an increase of 0% to 1%, Class 4 a flat stock price, Class 5 a stock price decrease of 0% to 1%, Class 6 a decrease of 1% to 2%, Class 7 a decrease of 2% to 3%, and Class 8 a decrease of more than 3%

C. The trading module

We use the prediction result from the deep learning module. When the predicted class belongs to an increasing class – that is, our model predicts that the stock price will go up – the strategy is to buy stock. When a decreasing class is predicted – the stock price is predicted to go down – the strategy is to sell stock. The results are calculated based on the returtn.

IV. EVALUATION

The model was designed to predict stock price movement by using attention-base LSTM model. In this section, we introduce the experimental setting in detail, including the dataset and evaluation.

We build the model by using Google's Tensorflow , which consists of a LSTM input layer that will take pricing data and technical indicators as input and will feed an output layer using softmax activation (1)

$$S_i = \frac{e^{V_i}}{\sum_j e^{V_j}} \quad (1)$$

The input layer has a dimensionality features, that consists of the set of technical indicators (KD, MA, RSV, etc.) and price data (Open, Close, Low, High, Volume). It will have an output later using the tanh function and that will be connected to an attention layer, then it will be connected to the network's output layer.

We use a metrics to evaluate our networks performance. The metrics were accuracy (2), precision (3), recall (4) and F-measure (5), a harmonic mean between precision and recall. Those metrics are calculated based on positive classes (true positives – tp), negative classes (true negatives- tn) and inaccurately for both class (false positive- fp, false negative- fn).

$$\mathrm{A} = \frac{tp + tn}{tp + fp + tn + fn} \quad (2)$$

$$\mathrm{P} = \frac{tp}{tp + fp} \quad (3)$$

$$\mathrm{R} = \frac{tp}{tp + fn} \quad (4)$$

$$\mathrm{F1} = 2\frac{P * R}{P + R} \quad (5)$$

V. CONCLUSION

To produce trading strategies, we apply deep learning to predict stock price movements. We train models with historical price data and technical indicators, after which we use the prediction result to decide on a trading strategy.

ACKNOWLEDGMENT

This study was supported in part by the Ministry of Science and Technology of Taiwan under grant number：MOST 105-2410-H-031 -035 -MY3 and MOST 107-2218-E-007-045. MOST 107-2221-E-027 -104 -MY2. MOST 107-2218-E-001 -009 -

VI. REFERENCE

[1]Yang, Z., et al. Stacked attention networks for image question answering. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[2] Luong, M.-T., H. Pham, and C.D.J.a.p.a. Manning, Effective approaches to attention-based neural machine translation. 2015.
[3] Eldan, R. and O. Shamir. The power of depth for feedforward neural networks. in Conference on Learning Theory. 2016.
[4] Telgarsky, M.J.a.p.a., Benefits of depth in neural networks. 2016.
[5] Mhaskar, H., Q. Liao, and T.A. Poggio. When and why are deep networks better than shallow ones? in AAAI. 2017.
[6] Akita, R., et al. Deep learning for stock prediction using numerical and textual information. in Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on. 2016. IEEE.
[7] Ding, X., et al. Deep learning for event-driven stock prediction. in Ijcai. 2015.
[8] Nelson, D.M., A.C. Pereira, and R.A. de Oliveira. Stock market's price movement prediction with LSTM neural networks. in Neural Networks (IJCNN), 2017 International Joint Conference on. 2017. IEEE.
[9] Fischer, T. and C.J.E.J.o.O.R. Krauss, Deep learning with long short-term memory networks for financial market predictions. 2018. 270(2): p. 654-669.
[10] Luong, M.-T., et al., Addressing the rare word problem in neural machine translation. 2014.
[11] Chorowski, J., et al., End-to-end continuous speech recognition using attention-based recurrent NN: first results. 2014.
[12] Xu, K., et al. Show, attend and tell: Neural image caption generation with visual attention. in International conference on machine learning. 2015.
[13] Liu, J., et al. Attention-Based Event Relevance Model for Stock Price Movement Prediction. in China Conference on Knowledge Graph and Semantic Computing. 2017. Springer.
[14] Zhao, R., et al., Visual Attention Model for Cross-sectional Stock Return Prediction and End-to-End Multimodal Market Representation Learning. 2018.
[15] Qin, Y., et al., A dual-stage attention-based recurrent neural network for time series prediction. 2017.
[16] Goodfellow, I., et al., Deep learning. Vol. 1. 2016: MIT press Cambridge.
[17] Hochreiter, S. and J.J.N.c. Schmidhuber, Long short-term memory. 1997. 9(8): p. 1735-1780.