Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction

Adithyan b asok
MSc Computer Science
Department of Computer Science
University of Calicut,India
adithyanbasok@gmail.com

Abstract—Stock market plays an important role in the economic development. Due to the complex volatility of the stock market, the research and prediction on the change of the stock price, can avoid the risk for the investors. The traditional time series model ARIMA can not describe the nonlinearity, and can not achieve satisfactory results in the stock prediction. As neural networks are with strong nonlinear generalization ability, this paper proposes an attention-based CNN-LSTM and XGBoost hybrid model to predict the stock price. The model constructed in this paper integrates the time series model, the Convolutional Neural Networks with Attention mechanism, the Long Short-Term Memory network, and XGBoost regressor in a non-linear relationship, and improves the prediction accuracy. The model can fully mine the historical information of the stock market in multiple periods. The stock data is first pre processed through ARIMA. Then, the deep learning architecture formed in pretraining-finetuning framework is adopted. The pre-training model is the Attention-based CNN-LSTM model based on sequence-to-sequence framework. The model first uses convolution to extract the deep features of the original stock data, and then uses the Long Short-Term Memory networks to mine the long-term time series features. Finally, the XGBoost model is adopted for fine-tuning. The results show that the hybrid model is more effective and the prediction accuracy is relatively high, which can help investors or institutions to make decisions and achieve the purpose of expanding return and avoiding risk. Source code is available at https://github.com/ zshicode/Attention-CLX-stock-prediction.

Index Terms—Attention mechanism, Convolutional Neural Networks, Long Short-Term Memory, XGBoost, stock prediction

I. INTRODUCTION

STOCK market plays an important role in the economic development. Due to the high return characteristics of stocks, the stock market has attracted more and more attention from institutions and investors. However, due to the complex volatility of the stock market, sometimes it will bring huge loss to institutions or investors. Considering the risk of the stock market, the research and prediction on the change of the stock price can avoid the risk for the investors.

The traditional time series model ARIMA can not describe the nonlinear time series, and needs to satisfy many pre conditions before modeling, and can not achieve remarkable results in the stock forecasting. In recent years, with the rapid development of artificial intelligence theory and technology, more and more researchers apply artificial intelligence method to the financial market. On the other hand, the sequence mod eling problem, focusing on natural language sequences, protein sequences, stock price sequences, and so on, is important in the field of artificial intelligence research [8], [13]. The most representative artificial intelligence method is neural networks, which are with strong nonlinear generalization ability.

Recurrent Neural Network (RNN) was adopted for analyz ing sequential data via neural network architecture, and Long Short-Term Memory (LSTM) model is the most commonly used RNN. LSTM introduced gate mechanism in RNN, which can be seen as simulation for human memory, that human can remember useful information and forget useless information [6]. Attention Mechanism [7], [16] can be seen as simulation for human attention, that human can pay attention to useful information and ignore useless information. Attention-based Convolutional Neural Networks (ACNN) are widely used for sequence modeling [4], [10]. Combining Attention-based Con volutional Neural Networks and Long Short-Term Memory, is a self-attention based sequence-to-sequence (seq2seq) [15] model to encode and decode sequential data. This model can solve long-term dependency problem in LSTM, hence, it can better model long sequences. LSTM can capture particular long-distance correspondence that fits the sturcture of LSTM itself, while ACNN can capture both local and global corre spondence. Therefore, this architecture is more flexible and robust.

Transformer [16] is the most successful sequential learning self-attention based model. Experiments on natural language processing demonstrates that Transformer can better model long sequences. Bidirectional Encoder Representation Trans former (BERT) with pretraining [2] can perform better than the basic Transformer. Pretraining is a method to significantly improve the performance of Transformer (BERT).

This paper proposes a hybrid deep learning model to predict the stock price. Different from the traditional hybrid prediction model, the proposed model integrates the time series model ARIMA and the neural networks in a non-linear relationship, which combines the advantages of the two vanilla models, and improves the prediction accuracy. The stock data is first preprocessed through ARIMA. The stock sequence is put into neural networks (NN) or XGBoost after preprocessing via ARIMA(p=2,q=0,d=1). Then, the deep learning architecture formed in pretraining-finetuning framework [2], [5] is adopted. The pre-training model is the Attention-based CNN-LSTM model based on sequence-to-sequence framework, where the Attention-based CNN is encoder, and the Bidirectional LSTM is decoder. The model first uses convolution to extract the deep features of the original stock data, and then uses the Long Short-Term Memory networks to mine the long-term time series features. Finally, the XGBoost model is adopted for fine-tuning, which can fully mine the information of the stock market in multiple periods. Our proposed Attention based CNN-LSTM and XGBoost hybrid model is so called AttCLX.

The results show that the model is more effective and the prediction accuracy is relatively high, which can help investors or institutions to make decisions and achieve the purpose of expanding returns and avoiding risks. The source code of this paper is available at https://github.com/zshicode/ Attention-CLX-stock-prediction. We conduct empirical study on the stock price of Back of China (601988.SH) in Chinese stock market. The data is downloaded from Tushare(www. tushare.pro). The stock price data on Tushare is with public availability.

II. MATERIALS AND METHODS

A. ARIMA

Classical stock prediction methods are based on ARMA (Auto Regressive Moving Average) model and ARIMA (Auto Regressive Integrated Moving Average) model. An ARMA(p,q) model

$$s_t = a_0 + \sum_{i=1}^{p} a_i\, s_{t-i} + w_t + \sum_{i=1}^{q} bi\, w_{t-1} \qquad (1)$$

where a,b are parameters, w is noise. ARMA model can be used when sequence s1:N is stationary, which means

E[st] = Constant.

Cov(St, St−k) = Constant.

where t = 1, 2, ..., N and k = 1, 2, ..., t.

When sequence is non-stationary, ARIMA(p,q,d) adopts d order difference to the sequence. In stock prediction, the first order difference x1:N = ˙s1:N (i.e. xk = sk − sk−1) is usually considered as stationary sequence.

We conduct empirical study on the stock price of Back of China (601988.SH) in Chinese stock market. The data is downloaded from Tushare(www.tushare.pro). The stock price data on Tushare is with public availability. The data is selected from the data from January 1, 2007 to March 31, 2022, the data in one day denotes a point of the sequence.

ADF test is adopted for testing the stationary condition of time series. Adopting ADF test for the original sequence and first-order difference sequence. The results are in Table I and II. When p-value is more than 0.562 or Critical Value (1%) is more than -3.44, the sequence is non-stationary. The ADF test shows that the original sequence is non-stationary and first-order difference sequence is stationary. The first-order difference sequence and second-order difference sequence are shown on Fig.1 and 2.

TABLE1

ADF TEST FOR ORGINAL SEQUENCE

Metric	Value
Test Statistic Value	-2.35539
p-value	0.154726
Lags Used	16
Number of Observations Used	3484
Critical Value(1%)	-3.43223
Critical Value(5%)	-2.86237
Critical Value(10%)	-2.56721

TABLE2

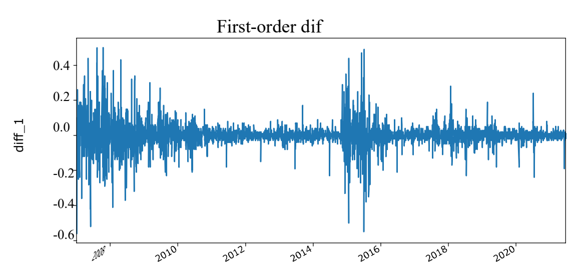
ADF TEST FOR FIRST ORDER DIFF SEQUENCE

Matric	Value
Test Statistic Value	14.7498
p-value	2.4956e-27
Lags Used	15
Number of Observations Used	3484
Critical Value(1%)	-3.43223
Critical Value(5%)	-2.86237
Critical Value(10%)	-2.56721

B.Deep learning on sequential data

In basic feed-forward neural network (FFNN), output of current moment ot is only determined by input of current moment it, which suppress the ability of FFNN to model time series data. In recurrent neural network (RNN), a delay is used to save the latent state of latest moment ht−1, then, latent state of current moment ht is determined by both ht−1 and it. See Fig. 3. [3] suggested that RNN may vanish the gradient as error propagates through time dimension, which leads to long term dependency problem. Human can selectively remember information. Through gated activation function, LSTM (long short-term memory) model can selectively remember updated information and forget accumulated information.

Sequence-to-sequence (seq2seq) [15] model adopted au toencoder (i.e. encoder-decoder architecture) for analyzing sequential data. Sequence-to-sequence model (seq2seq) [15] is constructed through an encoder-decoder architecture, which enhances the ability of LSTM to learn hidden information through data with noise. In seq2seq, the encoder is an LSTM that encodes the inputs into the context (commonly the hidden state at last hN), then decode the context in the decoder. In the decoder, output at previous moment is input at next moment.



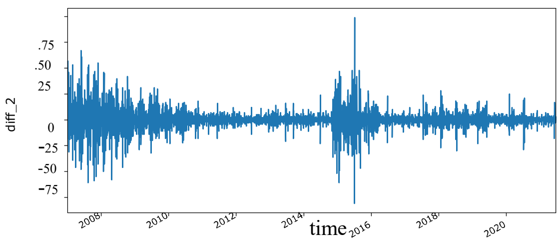


Fig. 2. The second-order difference.

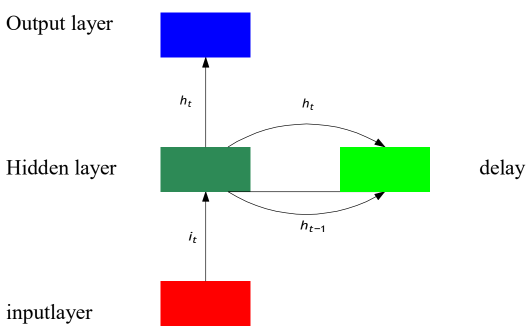


Fig.3RNN unit

To optimize the decoded sequence, beam search [17] was adopted in seq2seq model. Both beam search and Viterbi algo rithm in hidden Markov model (HMM) are based on dynamic programming. Solving optimal estimation of current state according to observation and previous state, is called decoding or inference in HMM. Solving p(xk|y1:k) and p(xk|y1:N) respectively, is equivalent to forward-backward algorithm in HMM. The optimal bidirectional estimation can be obtained through the distribution of xk. That is the probabilistic per spective that the bidirectional LSTM which combines forward and backward, is proposed [6]

.

B. Attention mechanism

Human usually pays attention to salient information. At tention mechanism is a technique in deep learning based on human cognitive system. For input X = (x1, x2, ..., xN), give query vector q, depict the index of selected information by attention z = 1, 2, ..., N, then the distribution of attention

$$\alpha_i = p(z = i|X, q) = \frac{\exp\big(s(x_i, q)\big)}{\sum_{j=1}^{N} \exp\,(s(X_{j,q}))}$$

i.e,

$$\alpha_i = softmax\big(s(Xi, q)\big) \quad (5)$$

Here

$$s(Xi, q) = \frac{X_i^T q}{\sqrt{d}}$$

is attention score through scaled dot product, d is the dimen sion of input. Suppose the input key-value pairs (K, V) = [(k1, v1), · · · ,(kN , vN)], for given q, attention function

$$att\big((K, V), q\big) = \sum_{i=1}^{N} \alpha_i\, v_i = \sum_{i=1}^{N} \frac{\exp\big(s(k_i, q)\big)}{\sum_j \exp\big(s(k_j, q)\big)} vi$$

Multi-head mechanism is usually adopted through multi query Q = [q1, · · · , qM] for attention function computation.

att((K, V), Q) = (att ((K, V), q1)‖ · · · ‖att ((K, V), qM)). (8)

Here, ‖ denotes Concatenate operation. This is so called multi head attention (MHA).

Attention mechanism can be adopted to generate data-driven different weights. Here, Q, K, V are all obtained through linear transform of X, and WQ, WK, WV can be adjusted dynamicly.

$$Q = W_Q X, K = W_K X, V = W_V X \quad (9)$$

This is so called self-attention. Similarly, output

hi = att ((K, V), qi) . (10)

Hence

$$h_{i=\sum_{j=1}^{N} \alpha_{ij} vi = \sum_{j=1}^{N} softmax\left(s(k_j, q_i)\right) vj} \quad (11)$$

Adopting scaled dot product score, the output

$$H = V\, softmax\left(\frac{K^T Q}{\sqrt{d}}\right) \quad (12)$$

D. Method

1) Preprocessing: This paper proposes a hybrid deep learn ing model to predict the stock price. Different from the tradi tional hybrid prediction model, the proposed model integrates the time series model ARIMA and the neural networks in a non-linear relationship, which combines the advantages of the two vanilla models, and improves the prediction accuracy.

The stock data is first preprocessed through ARIMA. The stock sequence is put into neural networks (NN) or XGBoost after preprocessing via ARIMA(p=2,q=0,d=1). Putting original stock market data into ARIMA, can output a new series that depict state more effectively.

The ADF results in Table I and II shows that the original sequence is non-stationary and first-order difference sequence is stationary. After determining d=1, we need to determine AR(p) and MA(q) in ARIMA. We use the Autocorrelation Figure (ACF) and Partial Autocorrelation Figure (PACF). The ACF and PACF of the original sequence and first-order difference sequence are shown on Fig. 4 and 5. Fig. 4 shows that PACF truncates when order=2, meaning that we should adopt AR(2). The ACF is with long tail for any order, meaning that we should adopt MA(0). Hence p=2, q=0.

2) Pretraining: Then, the deep learning architecture formed in pretraining-finetuning framework is adopted. The pre-training model is the Attention-based CNN-LSTM model based on sequence-to-sequence framework, where the Attention-based CNN is encoder, and the Bidirectional LSTM is decoder. The model first uses convolution to extract the deep features of the original stock data, and then uses the Long Short-Term Memory networks to mine the long-term time series features. Finally, the XGBoost model is adopted for fine-tuning, which can fully mine the information of the stock market in multiple periods