2019 IEEE 5th International Conference on Computer and Communications

Application of Regularized GRU-LSTM Model in Stock Price Prediction

Yiwei Liu
Dalian Maritime University
College of Science
Dalian, China
e-mail: 15871180890@163.com

Zhiping Wang
Dalian Maritime University
College of Science
Dalian, China
e-mail: wzp@dlmu.edu.cn

Baoyou Zheng
Dalian Maritime University
College of Science
Dalian, China
e-mail: 17824833955@163.com

***Abstract*—The stock market is a highly complex nonlinear movement system, and its fluctuation law is affected by many factors, so the prediction of the stock price index is a very challenging task. There are many examples showing that Neural Network algorithms are well suited for such time series predictions and often achieve satisfactory results. In this paper, based on the existing models, we proposed a Regularized GRU-LSTM neural network model and applied it to the short-term forecast of closing price of the two stocks. The experimental results show that our proposed model is superior to the existing GRU and LSTM network models in stock time series prediction.**

Keywords-LSTM; GRU; Time series prediction

I. INTRODUCTION

The stock market is more popular in recent years due to its high return rates. In spite of high risk, some investors and institutions still choose the stock market to invest. Therefore, the stock price index prediction has attracted the attention of both private and institution investors. In addition to its inherent complexity, there has been an unchanging argument on the predictability of stock returns and various of methods for predicting and modeling stock price index have been object of study of many different subjects, such as physics, economics, computer science and statistics. In 1970, Fama introduced the Efficient-Market hypothesis[1], which defines that the current price of an asset always reflects all of the previous available information. It is worth mentioning that in 2012, it was estimated that about 85% of the transactions in the US stock market could be carried out by algorithms.

Many methods have been used to forecast the stock price index, including traditional models and the recently popular neural network models. The traditional models include Autoregressive Integrated Moving Average(ARIMA)[2], and Autoregressive Conditional Heteroskedasticity(GARCH) volatility[3]. These models are based on the assumption that a linear correlation structure exists among time series values. Therefore, non-linear patterns cannot be captured by these models. To overcome this limitation, neural network models have been widely used in the prediction of nonlinear time series[4] such as stock price index.

Recurrent Neural Network(RNN) have been proved to be one of the most forceful models for processing sequential data[5], it can recognize complex nonlinear relationships which are difficult to capture using traditional forecasting models[6]. Long Short-Term Memory(LSTM)[7] and Gated Recurrent Unit(GRU)[8] are the two most satisfactory RNN structures. LSTM adopts the memory cell, a unit of calculation, which displaces traditional artificial neuron in the hidden layers of the network. With these memory cells, networks are able to effectively link the memories and the new input, and seize the architecture of data dynamically, which make the prediction more accurate. GRU is very similar to LSTM, the main difference between them is that GRU does not have the output gate as in LSTM. On the basis of the two RNN structures, there are many improved models in recent years, such as the Bi-directional LSTM[9] structure, which has also been used widely. For example, in 2018, Gu et al.[10] used Bi-directional LSTM for short text classification.

Note that, LSTM network has a longer memory capacity for preserving and processing the previous information, then, for large data, the LSTM network may derive better results. However, GRU is much faster than LSTM since it has fewer parameters. In this paper, we combined LSTM and GRU, and proposed a new Regularized GRU-LSTM network model with better performance. With this model, we predicted the closing prices of two stocks.

This paper is organized as follows. In section 2, we discuss relevant past work and we describe some details of our model in section 3. Then we demonstrate our experimental results in section 4. Finally, we conclude the paper with an outline and some future research directions.

II. RELATED WORK

In the past twenty years, many researchers have concentrated on predicting stock price using neural network and have made great improvement. Such as Aladag et al.[11] proposed a new artificial neural network-based predictive combination method, which was applied to the Istanbul Stock Exchange(IMKB) time series index prediction and found to have significant improvement. As early as 2001,

978-1-7281-4743-7/19/$31.00

Ines et al.[12] proposed a neural network model based on partial recurrent neural network, which has a long-term predictive ability in learning phase, the partial recurrent neural network can help improve the accuracy of prediction. Since then, RNN has been widely used in time series forecasting, especially for the prediction of stock price index. Such as in 2011, Hsieh et al.[13] proposed a recurrent neural network based on artificial bee colony algorithm, and used it in the prediction of the price index of multiple stocks including the Dow Jones Industrial Average. However, RNN can't solve the problem of long-term dependence, which becomes more and more serious along with the data increasing. Therefore, other alternative neural network models are used in predicting stock price. The LSTM and GRU neural network models can solve the problem of gradient disappearance in RNN. And these two models have applications in many fields such as computer vision. For example, in 2018, Li et al.[14] proposed an LSTM network model based on residual attention and applied the model successfully to video captioning. In terms of stock price index forecasting, these two models have their own unique advantages. In 2016, Nelson et al.[15] demonstrated the effectiveness of LSTM on stock price forecasts by training a series of different architecture LSTM networks. In[16], the author proposed a model based on LSTM to predict price movements using an input that is not based on text. In 2018, Ma et al.[17] proposed a method of combining k-means clustering algorithm with multi-branch LSTM, the core of the method is to construct the same number of LSTM neural network models according to the number of clusters, and finally train the corresponding LSTM models with clustering results, and this method is used to effectively predict the closing price of Ping An Bank. The GRU network model is a variant of LSTM, which has a faster speed than LSTM. In 2018, DANG et al.[18] proposed a new method called TGRU, which is a method of using deep learning by analyzing financial news articles and historical stock prices. And the method was successfully applied to the prediction of the S&P500 stock index.

In this paper, we use LSTM, GRU and our proposed network model to make short-term forecasts of the closing prices of the two stocks. We find that our proposed method outperforms the other two types of neural network models in the prediction of the stock price index.

III. METHODS

A. The LSTM Network Model

LSTM add a "gate" structure compared to a regularly RNN, each memory cell usually has three gates: the input gate, the forgetting gate and the output gate. The major function of the input gates is to receive a series of information and determine which information can be retained in the cell state, and the forgetting gate decides what kind of information to be discarded from the cell state, this step can be achieved by the sigmoid activation function. The results of the calculations and the output values can be obtained through the output gate.

The cell state is updated based on the outputs form the gates. We can described them using the following equations[19] and the detailed is shown in the Figure 1(a).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

where x_t represents the input vector, f_t is the forget gate vector, and i_t is the input gate vector.

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

In this process, $\tilde{C}_t$ is the cell state vector, and C_t is the new cell state vector.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

where o_t represents the output gate vector, $\sigma(x)$ represents the sigmoid activation function and W, b are the weight and bias parameters, respectively.

The output of each hidden state h_t is passed to the next memory location, and the entire internal process eventually scales the cell state to (0,1). This process is iteration for each memory cell of LSTM.

B. The GRU Network Model

The GRU network was proposed by Cho et al[8] and was a slightly dramatic variant of LSTM. Compared with the LSTM network, the GRU network has one less gate. There are only two gates, namely reset gate and update gate in the GRU network. The reset gate combines the new input with the previous memory and controls the extent to which the previous memory is ignored. The smaller the value of the reset gate, the more memory is ignored. The update gate is used to determine how much memory is reserved before. The greater the value of the update gate, the more information it will generate. If we set all the reset gates values to 1 and the update gate values to 0, we get the general neural network model. The mathematical descriptions of the GRU cell functions are as follows and the detailed process is shown in the GRU layer part of Figure 1(b).

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}_t} \cdot [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

$$y_t = \sigma(W_o \cdot h_t)$$

Just as the operation of the forgetting gate and the input gate in the LSTM, the update gate z_t and the forgetting gate r_t in GRU are derived from the input x_t and the hidden layer h_{t-1} at the previous moment are multipling by the weight matrices W_z and W_r respectively. The final output is compressed to 0-1 by using the sigmoid activation function $\sigma(x)$ on $h(t)$.

C. Regularized GRU-LSTM Model

Regularization is a form of regression which can reduce the complexity and instability of the model during the learning process, thereby reducing the risk of model over-fitting. The general form of regularization[20] can be described as follows.

$$\min_{\omega}\{\sum_{t=1}^{T} l(y_t, f(x_t,\omega)) + \sum_{i=1}^{m}\lambda_i\rho_i(\omega)\} \tag{1}$$

In which, $l(\cdot,\cdot)$ represents the loss function in the model, T represents the number of samples, $\lambda\rho(\omega)$ is the regularized item, λ is to control the equilibrium relationship between the regular term and the loss function, thus called the regularization tunable parameter. $\rho(\omega)$ has many different expression forms, among which the L_1 norm and the L_2 norm are commonly used. In general, different norm penalties will cause different model generalization directions. In our model, we selected L_2 norm as the regularized item and Mean Absolute Error(mae) as the loss function, the objective function (1) is as follows.

$$\min_{\omega}\{\frac{1}{T}\sum_{i=1}^{T}| f(x_i,\omega) - y_i | + \lambda_2 \| \omega \|_2^2\} \tag{2}$$

The Regularized GRU-LSTM model consists of two GRU layers, one LSTM layer, and finally a dense output layer as shown in Figure 1(c). In each layer, we select the default activation function. Since the LSTM layers are non-linear, only one dense layer is needed to accumulate their outputs.

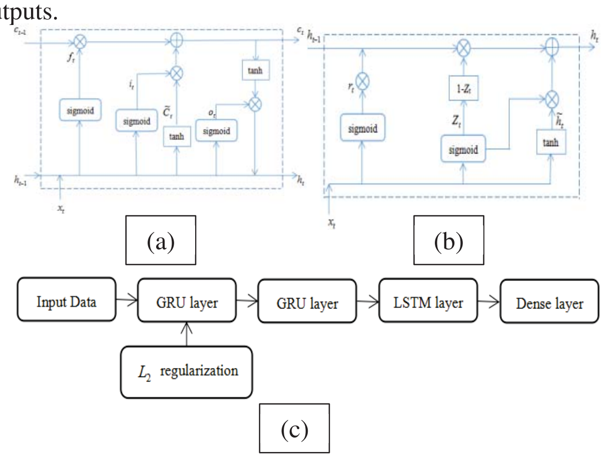

Figure 1 Model internal and external structures. (a) Internal structure of LSTM model. (b) Internal structure of GRU model. (c) The external structure of the proposed model

D. Evaluation Standard

To evaluate the prediction performances of LSTM, GRU and our proposed model , we choose Root Mean-Square Error(RMSE) and Directional Accuracy(DA)[21] as the evaluation criterias for the model. The formulas for these measures are defined as follows:

$$RMSE = \left(T^{-1}\sum_{t=1}^{T}\left(Y_{(t)} - Y_{(t)}^{*}\right)^2 \right)^{1/2}$$

$$DA = \frac{100}{T}\sum_{t=1}^{T} d_t$$

where $d_t = \begin{cases} 1 & (Y_{(t)} - Y_{t-1})(Y^{*}_{(t)} - Y^{*}_{t-1}) \geq 0 \\ 0 & \text{otherwise} \end{cases}$, $Y_{(t)}$ and $Y^{*}_{(t)}$ are the anti-normalized actual and predicted values at time t, respectively, and T is the number of samples. Here RMSE represents the deviation between the actual and predicted values. Therefore, the forecasting performance is better when RMSE is smaller. DA indicates the correctness of the prediction direction and can also be used to assess the accuracy of the prediction, the higher the DA value is the better the prediction is.

IV. EXPERIMENTATION LAYOUT AND RESULTS

A. Raw Data

In this paper, we collected the daily stock prices of Dalian Thermal Power Co.Ltd(600719) and Dalian Friendship Co.Ltd(000679) from March 15, 2008 to March 15, 2018. The data was obtained from Choice Financial Terminal. We only used the open, high, low, close and volume as features for our inputs (The other indicators used in the technical analysis can be calculated based on these five indicators) and the closing prices of the two stocks are shown in Figure 2. We obtained 2217 valid samples after removing the redundant ones for 600719. Among which, 1995(about 90% samples) are used as training samples and the remaining 222 closing prices are testing samples. For 000679, we finally get 2264 valid samples, including 2037 training samples(about 90%) and 227 testing ones. We then take 10% from the training sets as the verification set respectively.

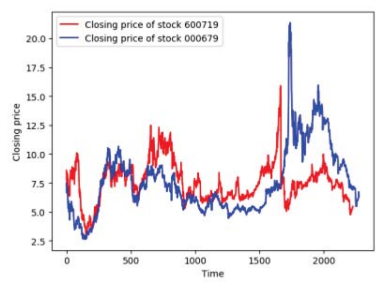

Figure 2. The daily stock closing prices of 600719 and 000679 from March 15, 2008 to March 15, 2018.

B. Data Preparation

In order to reduce the adverse effects caused by the unit disunity, we need to normalize the data before training. First, each day's stock information can be expressed as a vector $x_t = (c_t, h_t, l_t, o_t, v_t)$, where c_t, h_t, l_t, o_t, v_t indicate the close, high, low, open and volume respectively. The formula used for data normalization is as follows.

$$\overline{x}_t = \frac{x_t - \min x_t}{\max x_t - \min x_t}$$

In our experiment, since we chose a step size of 10, so we use $X_t = (\overline{x}_t, \overline{x}_{t+1}, \cdots, \overline{x}_{t+9})$ as input to each model. The input data is a three-dimensional tensor, three dimensions represent the number of samples, time steps and the number of features, respectively. Therefore, the input shape of the first LSTM layer is the last two dimensions of input data.

C. Parameter Determination

Theoretically, the three layers neural network model (including the input layer and the output layer) can fit any nonlinear problem. In our experiment, in order to highlight the performance of the model, each model adopts the five-layer network structure that contains three Layers of hidden layer except the input layer and the output layer. The performance of the model is not only affected by the number of layers of the network, but also related to the parameters. The parameters of model are shown in Table I

TABLE I THE PARAMETERS OF MODEL

Parameter	Value	Parameter	Value
Learn rate	0.001	Optimizer	Adam
Batch_size	32	Epochs	500
Loss function	mae	Dropout	0.1

D. Result Analysis

We used Keras neural network framework to build the prediction model in Python 3.6.0 environment. In order to compare the prediction performance of LSTM, GRU and our model network models, for each model, we will train and predict the closing prices for stock 600719 and 000679 respectively. We repeated the experiment 20 times on these three network models in order to eliminate the contingency.

The effect of the three predictions for stock 600719 and 000679 are shown in Table II.

TABLE II. 600719 AND 000719 PREDICTION ERROR STATISTICS

Evaluation		RMSE		DA Score	
Stock	model	**mean**	**std**	**mean**	**std**
600719	GRU	0.355	**0.060**	0.505	**0.008**
	LSTM	0.370	0.067	0.504	0.008
	Our	**0.287**	0.068	**0.514**	0.010
000679	GRU	0.685	0.059	0.472	**0.012**
	LSTM	0.722	0.078	**0.480**	0.014
	Our	**0.252**	**0.029**	0.479	0.015

It can be seen from Table II in the prediction of stock 600719 closing prices, the average RMSE obtained by our model is lower and the average DA score is higher, compared to the other two models. In the forecast of the stock price of 000679, although there is not much difference among the DA scores of the three models, but the average RMSE of our model is significantly lower than the other two models. Which illustrate that the performance of the our model is better than the GRU and LSTM model when predicting the closing price of the two stocks.

Figure 3 shows the prediction curve of these three models for the two stocks, in which, the blue curves are the actual prices, and the red curves are the predictions. The closer to the blue curve (actual price trend), the better the predictive performance of the model is.

t

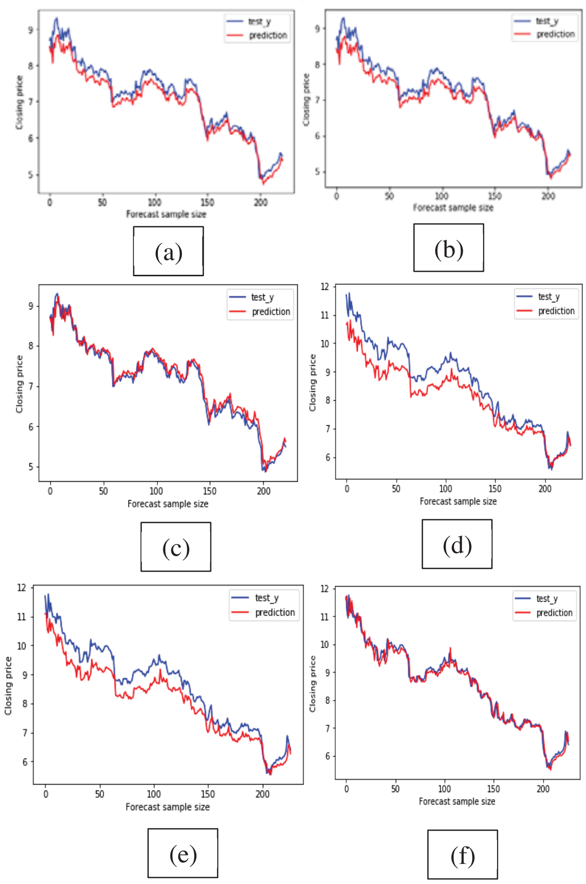

Figure 3. The prediction curve of three models for the two stocks. (a) 600719 prediction results based on GRU model. (b) 600719 prediction results based on LSTM model. (c) 600719 prediction results based on Our model. (d) 000679 prediction results based on GRU model. (e) 000679 prediction results based on LSTM. (f) 000679 prediction results based on Our model

It can be seen that when predicting the closing price of the two stocks, the prediction results of our model(the red curve part in Figure 3(c) and (f)) is closer to the actual price curves than the prediction results of the other two models.

A good neural network model is not only reflected in the prediction accuracy, but also in the number of its parameters. For the same data set, models with fewer parameters can speed up training, and reduce computer memory consumption. Therefore, for a more comprehensive comparison of the three models, we present the numbers of parameters of the three models. As shown in Table III.

TABLE III NUMBER OF PARAMETERS OF EACH NETWORK MODEL

Model	GRU	LSTM	Our
Parameters	**25121**	33489	25905

As it is shown in Table III, the GRU model contains the fewest parameters, followed by our proposed model, and finally the LSTM model. Therefore, the GRU model has the highest training speed, although its prediction effect is not good. Our model has only 784 more parameters than the GRU model, which indicates that our model has little difference in training speed with the GRU model. But the prediction effect of our method is significantly better than the GRU model. Therefore, from a comprehensive perspective, our proposed network model is a good choice for short-term sequence prediction.

V. CONCLUSION

In this work, we proposed a new regularized GRU-LSTM model and compared it with the GRU and LSTM network models. The experimental results showed that our model had a better performance in the short-term forecast of the stock closing price. It has not only high prediction accuracy, but also high convergence speed.

However, the stock market is a frequently changing dynamic system and the prediction of the stock market is a very complex issue, we should not only take into account the problem of the prediction method itself, but also recognize some characteristics of the stocks, and the network system must be trained many times according to the new data produced by the stock market to adapt to the complicated circumstances. Although our regularized GRU-LSTM model has higher prediction accuracy and convergence speed, its stability is sometimes relatively poor. Therefore, our next step is to find an optimization method to improve the stability of the model.

ACKNOWLEDGMENT

The work was supported by the Fundamental Research Funds for the Central Universities Under Contract No.3132019323.

REFERENCES

[1] E. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," The Journal of Finance. vol. 25, pp. 383-417, February 1970.

[2] D. Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction," Engineering Applications of Artificial Intelligence. Vol. 23, pp. 586-594, April 2010.

[3] P. Franses and H. Ghijsels, "Additive outliers, GARCH and forecasting volatility," Internatioal Journal of Forecasting. vol. 15, pp. 1-9, January 1999.

[4] Ferreira, A. E. Tiago, G. C. Vasconcelos, and P. J. L. Adeodato, "A New Intelligent System Methodology for Time Series Forecasting with Artificial Neural Networks," Neural Processing Letters. vol. 28, pp. 113-129, February 2008.

[5] D. T. Pham, "Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems," Monthly Weather Review. vol. 129, pp. 1194, May 2001.

[6] Zhang. G, Hu. M. Y, and Lev. B, "Neural network forecasting of the British Pound/US Dollar exchange rate," Omega. vol. 26, pp. 495-506, April 1998.

[7] Hochreiter, Sepp, Schmidhuber, and Jürgen, "Long Short-Term Memory," Neural Computation. vol. 9, pp. 1735-1780, August 1997.

[8] K. Cho, Van. Merrienboer, and C. Gulcehre, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," arXiv:1502.08029, February 2015.

[9] A. Graves, Jürgen and Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," Neural Networks. vol. 18, pp. 602-610, May 2005.

[10] Gu. Y, Gu. M, and Long. Y, "An enhanced short text categorization model with deep abundant representation," World Wide Web. vol. 21, pp. 1705-1719, March 2018.

[11] Aladag, C. H, E. Egrioglu, and U. Yolcu, "Forecast Combination by Using Artificial Neural Networks," Neural Processing Letters. vol. 32, pp. 269-276, March 2005.

[12] M. Inés, Galván, and P. Isasi, "Multi-step Learning Rule for Recurrent Neural Models: An Application to Time Series Forecasting," Neural Processing Letters. vol. 13, pp. 115-133, February 2001.

[13] T. Hsieh, H. Hsiao, and W. Yeh, "Forecasting Stock Markets Using Wavelet Transforms And Recurrent Neural Networks: An Integrated System Based On Artificial Bee Colony Algorithm," Applied Soft Computing Journal. vol. 11, pp. 2510-2525, February 2011.

[14] Li. X, Zhou. Z, and Chen. L, "Residual attention-based LSTM for video captioning," World Wide Web. vol. 22, pp. 621-636, January 2018.

[15] D. M. Q. Nelson, A. C. M. Pereira, and R. A. D. Oliveira, "Stock market's price movement prediction with LSTM neural networks," IJCNN. Anchorage, Alaska. vol. 6, pp. 1419–1426, May 2017.

[16] Jia. H, "Investigation Into The Effectiveness Of Long Short Term Memory Networks For Stock Price Prediction," arXiv: 1603.07893, August 2016.

[17] Shao. X. L, Ma. D, and Liu. Y, "Short-term forecast of stock price of multi-branch LSTM based on K-mean," ICSAI. Hangzhou, China. vol. 4, pp. 1546–1551, November 2017.

[18] Dang. L and Minh, "Deep Learning Approach for Short-Term Stock Trends Prediction based on Two-stream Gated Recurrent Unit Network," IEEE Access, vol. 6, Sep. 2018, pp. 55392-55404, doi:10.1109/ACCESS.2018.2868970.

[19] S. Selvin, R. Vinayakumar, and E. A. Gopalakrishnan, "Stock price prediction using LSTM, RNN and CNN-sliding window model," ICACCI. Manipal, India. vol. 3, pp. 1643–1647, September 2017.

[20] Ren. J, and Wang. J. H, "Stock Index Prediction Based on Regularized LSTM Model," Computer Applications and Software, vol. 35, Apr. 2018, pp. 45-48, doi:10.3969 /j.issn.1000-386x. 2018.04.008.

[21] Wang. J. J, Wang. J. Z, Zhang. Z. G, and Guo. S. P, "Stock index forecasting based on a hybrid model," Omega, vol. 40, Jun 2012, pp. 758-766, doi:10.1016/j.omega.2011.07.008.