Currency Exchange Rate Prediction with Long Short-Term Memory Networks Based on Attention and News Sentiment Analysis

Ching-I Lee
Department of Mathematics
National Central University, Taiwan
Email: joylee9216@gmail.com

Chia-Hui Chang
Department of Computer Science
and Information Engineering
National Central University,Taiwan
Email: chiahui@g.ncu.edu.tw

Feng-Nan Hwang
Department of Mathematics
National Central University, Taiwan
Email: hwangf@math.ncu.edu.tw

***Abstract*—Currency exchange rate prediction is a typical time series prediction problem which has been solved by time-series models, such as Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA) as well as machine learning methods, such as Single Layer Perception (SLP) and Long Short-Term Memory (LSTM). In this work, we aim to predict the future currency exchange prices in collaboration with news sentiment analysis. We use the Australian dollar (AUD) against the US dollar as a case study and study the prediction AUD rate for next day, week, two-week, and month. We conduct a comparative study of the proposed attention-based LSTM with typical models, including ARIMA, SARIMA, SLP, and classical LSTM. The numerical results showed that adding sentiment score of the news articles and matching keywords of "up/increase" can reduce prediction error by at least 15%. For more extended future prediction, the newly trained model is better than the strategy that reuses the next-day model.**

I. Introduction

In the global financial economy, the foreign currency exchange market is one of the largest world's markets, with a trading volume of more than 1.4 trillion US dollars daily in an average [1]. Foreign exchange market refers to the market formed by market participants engaged in foreign exchange transactions. The purpose of trading includes investment and trade, speculation, and hedging. In international trade activities, exchange rate fluctuations directly reflect the economic relationship between countries. The price of import and export commodities fluctuates due to exchange rates. Forex traders can speculatively trade according to exchange rate changes and profit from them. Multinational companies can also take hedging measures to offset the losses caused by exchange rate changes.

Generally speaking, there are two major classes of the future exchange rate predictive tools based on the historical market price changes available in the literature, including the time series models and machine learning methods [2] [3]. Among all time-series models, the autoregressive integrated moving average (ARIMA) and the seasonal ARIMA (SARIMA) is one of the popular methods [4]. However, some comparative research studies showed that the prediction performance of using machine learning methods is better than traditional time series models. For example, Kamrwzaman and Sarke [5] compared three different artificial neural networks (ANN) with ARIMA to predict the exchange rate of the Australian dollar against the other six currencies. The results showed that all ANN models were superior to the traditional time series ARIMA model.

Learned from [6][7][8], we have found some successful applications for sentiment analysis social networks and forum speeches for stock price prediction. In the foreign exchange rate market, market sentiment could also affect the trend of the exchange rate. Thus, emotional analysis of news media becomes more and more important. In this paper, we propose to apply a variant of LSTM, called LSTM-attention, in conjunction with news sentiment analysis, to predict the currency exchange rate. We use the Australian dollar against the US dollar as a case study. We consider various length (7, 30, and 60 days) of historical data as input features for LSTM-attention models, and investigate whether the performance of LSTM-attention with the features extracted from news articles by sentiment analysis will be improved or not. A comparative study of our proposed method with other alternatives, including ARIMA, SARIMA, single Layer Perception (SLP), and LSTM, will be presented.

II. Background Technologies

In this section, we introduce the models that are used in this paper.

A. Long Short-Term Memory (LSTM)

LSTM is a special type of neural networks to extend the short-term memory of recurrent neural networks. It has three mechanisms for controlling memory: input gates, output gates, and forget gates [9]. Consider an LSTM memory cell as shown in Fig. 1 at time k, Here, $k \in$

978-1-7281-4666-9/19$31.00 © 2019 IEEE

$\{1, 2, 3, \ldots, t-1\}$. In the cell, x_k and h_k are denoted by the input and the output of the memory cell, respectively. Other notations include: the forget gate F_k that determines how much of the memory in the previous cell c_{k-1} is put into the memory of the cell c_k, the input gate I_k that determines how much new information will enter the memory and the output gate O_k that determines how much updated information is to output for the next LSTM unit in the network. Then the LSTM unit cell function is defined as follows.

$$h_k = O_k \cdot \tanh(c_k),$$

where

$$\begin{aligned} c_k &= F_k \cdot c_{k-1} + I_k \cdot \tanh(W_x^c x_k + W_h^c h_{k-1} + b_c) \\ F_k &= \sigma(W_x^F x_k + W_h^F h_{k-1} + b_F) \\ I_k &= \sigma(W_x^I x_k + W_h^I h_{k-1} + b_I) \\ O_k &= \sigma(W_x^O x_k + W_h^O h_{k-1} + b_O) \end{aligned}$$

Here, W_x^F, W_h^F, W_x^I, W_h^I, W_x^O, W_h^O, W_x^c, W_h^c represent the weights. b_F, b_I, b_O, b_c represent the bias. Fig. 2 shows

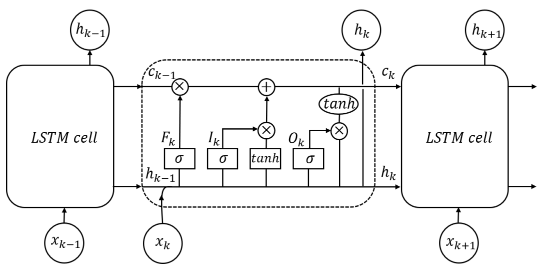

Fig. 1: The structure of an LSTM cell

the network architecture for LSTM used for numerical experiments, which is composed of one input layer, two LSTM layers, and one output layer. Both of the first and second layers consist of 50 LSTM cells and the output layer has 1 neuron. The exchange rate prediction, x_t, is determined by h_{t-1} via

$$x_t = W_s h_{t-1} + b_s,$$

where W_s is the weight, and b_s is the bias. Here, h_{t-1} depends on x_1, x_2, ..., x_{t-1} in the input layer and the output of each LSTM cell, h_1, h_2, ..., h_{t-2}.

B. Long Short-Term Memory Networks based on attention (LSTM-attention)

The main difference between the LSTM with or without attention is whether the outputs of the LSTM cells, h_1, h_2, ..., h_{t-1} at intermediate steps are reused for obtaining the output of this layer, x_t through $x_t = f(r, h_{t-1})$ with $r = \sum_{i=1}^{t-1} \alpha_i h_i$, where r is the output form attention mechanism, and f is the function that generates the final output. As shown in Fig. 3, the basic idea of the attention mechanism is to remove the limitation of the traditional encoder-decoder structure that relies on a fixed-length

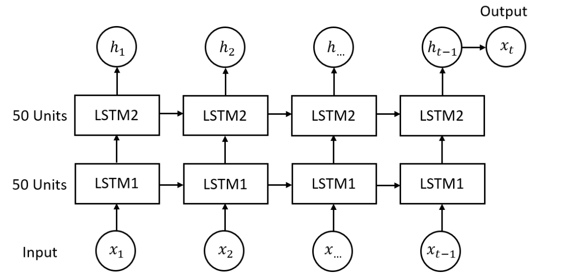

Fig. 2: An architecture of the LSTM network model

vector internally during encoding and decoding, and use a weighted sum of the output at each time step to encode the whole input sequence.

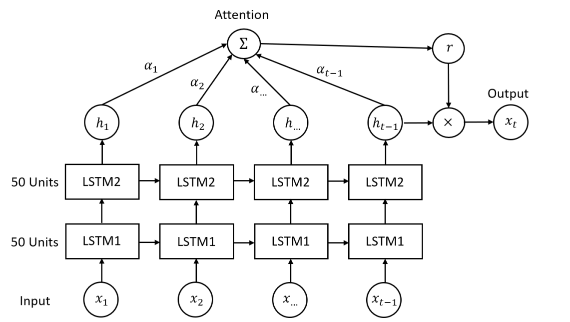

Fig. 3: An architecture of the LSTM-attention network

III. Problem Definition and Proposed Methods

In this paper, we study the problem of next day exchange rate prediction as well as the cases for the next week, two-week, and month prediction. For input features, we compare the use of historical exchange rate data from the last 7, 30, or 60 days and show that the last 7-day historical data with ratio data outperforms historical only input. In addition, simple sentiment analysis with SnowNLP and keyword matching with “increase/up” in the news article is implemented to see if information from news could help solve the task. We compare the performance of LSTM and attention-LSTM models with three baseline models, including ARIMA, SARIMA, and SLP in terms of RMSE and MAPE.

A. The next day prediction task with sentiment analysis

In addition to the historical data, other factors, such as microeconomics, politics, and humans, affect the currency exchange rate. Through sentiment analysis, we hope to extract some useful features from news articles that can reflect the market sentiment to improve the accuracy of the LSTM based predictive tool for predicting the currency exchange rate in the future. By doing so, we collected all news articles that mention the Australian dollar from the newspaper, Liberty Times in Taiwan within a selected period and then produced two types of features. The first one is generated by sentiment analysis through SnowNLP [10], a natural language processing tool, and the second one is obtained by keyword matching. SnowNLP calculates the emotional score for each article. If the score is higher than 0.5, we label it by 1; if not, it is -1. For keyword matching, we searched the keyword “升” (up/increase) and the words with similar meaning. If either one of these two keywords is found, the score is set to be 1. If not, the score is -1.

B. The extended day prediction

We consider two strategies for extended date exchange rate prediction: model reusing and remodeling. The first one is to reuse the next day model until the predicted value on the target date is obtained (as shown in Fig. 4). The intermediate outputs obtained from the previous run are used for input features for the next run. The second one is to build a new model for the target task, such as the currency rate prediction for next week or next month (as shown in Fig. 5). Figs. 4 and 5 show the example of the next-week currency exchange rate prediction by LSTM-attention using the first and second strategies, respectively. For both models, we use seven day historical data as input and the prediction on the next week as output.

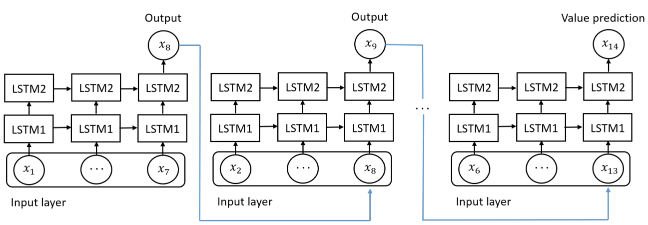

Fig. 4: The first strategy that reuses the next-day prediction model repeatedly

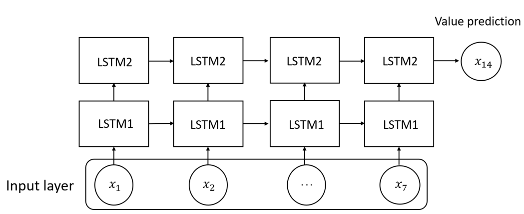

Fig. 5: The second strategy that builds a new model for target date prediction using the seven-day historical data as input and the next seven-day prediction as output.

IV. Numerical results and discussions

A. Data description and performance metrics

A total of three years and three months of historical exchange rate data of Australian Dollar to US Dollar was collected from a website of `www.investing.com` and related news articles on the newspaper, Liberty Times in Taiwan, from January 1, 2016, to March 31, 2019. We use the data from January 1, 2016, to December 31, 2018, as the training set and the data from January 1, 2019, to March 31, 2019, as the testing set. To compare each technique qualitatively, we use two performance measurements, the root mean square error (RMSE) and the mean absolute percentage error (MAPE), which are given by RMSE $= [\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2]^{\frac{1}{2}}$ and MAPE $= \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i-\hat{y}_i}{y_i}|$, where y_i and $\hat{y}_i$ are the true and predicted values of the data, respectively. In implementation, the LSTM and SeqSelfAttention class from Keras [11] were used. The information of LSTM parameters is summarized as follows. The activation function is ReLU, the optimizer is adam, the loss function type is the mean squared error, the number of epochs is 1000, the batch size is 200, and the validation split is 0.2.

B. The next day currency exchange rate prediction

First, we explore how the performance of our proposed method varies with the input preparation for the next day exchange rate prediction. These factors include

- the different length of historical exchange rate prices: 7/30/60 days,
- manipulation of historical data based on difference and ratio, and
- adding news feature extracted from sentiment analysis.

We consider two kinds of historical data manipulations which compute either the difference or ratios of the exchange rates between two consecutive days, i.e., Difference $= x_t - x_{t-1}$ and Ratio $= \frac{x_t}{x_{t-1}}$. Table I summarizes the performance of LSTM and LSTM-attention using different sizes of input features that correspond to the data for 7, 30, and 60 days, respectively. From the table, we found that generally speaking, LSTM-attention performs better than the classical LSTM for all cases. From the same table, we observe that 7-day inputs reduced the test errors for 30-day and 60-day inputs by 15% to 20% and 55% to 60%, respectively. Then we compare the performance of two variants of LSTMs. The performance of LSTM-attention is better than the classical LSTM model and reduces the RMSE of LSTM by 5% to 17%.

Method		LSTM-attention	Classical LSTM
	Input	historical data	
RMSE	7 days	0.0052	0.0055
	30 days	0.0063	0.0067
	60 days	0.0117	0.0141
MAPE	7 days	0.0057	0.0060
	30 days	0.0072	0.0075
	60 days	0.0145	0.0185

TABLE I: The test errors of two variants of LSTMs with different lengths of input features (7/30/60 days) for the next-day currency exchange rate prediction.

Second, we use seven-day data as the input features and compare the effect of adding the exchange rate difference and ratio of two consecutive days. From Table II, we observe that adding the ratio as a feature can reduce the RMSE by 12%. However, adding the difference does not help LSTM with only historical data to reduce the test error.

TABLE II: The test error of LSTM-attention adding difference (D) and ratio (R) of two consecutive for the next-day currency exchange rate prediction. "H" is the historical data.

Method		LSTM-attention		
	Input	H	H+D	H+R
RMSE	7 days	0.0052	0.0055	0.0046
MAPE		0.0057	0.0063	0.0052

Finally, we compare the performance of adding news sentiment. We use historical data and ratios and whether to add news as our input. A shown in Table III, adding news features can reduce the test error of LSTM without news by 15%. In summary, LSTM-attention with 7-day historical exchange rate data and its ratio and news performs the best. We obtain the same conclusion using MAPE as a performance metric.

TABLE III: The test error is reduced when adding news sentiment analysis for the next-day currency exchange rate prediction.

Method		LSTM attention	
	Input	with news	without news
RMSE	7 days H+R	0.0039	0.0046
MAPE		0.0041	0.0052

C. Comparison with different methods

In this section, we present a comparative study of our proposed method with two time-series models, ARIMA and SARIMA, and one machine learning method, SLP. After parameter tuning done in the previous section, the LSTM-attention results are obtained by using the following features, the historical data from previous seven days, the ratio of exchange rates on two consecutive days, the features extracted from news articles through the sentiment analysis and keywords matching. For ARMIA and SARAMA, the optimal parameters, (p, d, q) of ARIMA are $(1, 1, 0)$; the non-seasonal part, (p, d, q), and the seasonal part, (P, D, Q) of the SARIMA method are (1, 0, 6) and (2, 0, 1), respectively. Here, (p, d, q) and (P, D, Q) represent the order of autoregressive model, the degree of differencing, and the order of the moving average model for the nonseasonal and seasonal parts of the ARIMA model, respectively.

For the basic artificial neural network, we use single-layer perceptron, which consists of three parts: the input layer, the hidden layer, and the output layer. Each layer has a large number of artificial neurons. $x_1, x_2, \ldots, x_{t-1}$ are our inputs, $h_1, h_2, \ldots, h_{t-1}$ are the hidden units, and x_t is the exchange rate prediction.

The results for SLP are obtained by using the following settings. A rectified linear unit is used as the activation function, the optimizer is selected to be Adams, the loss function is mean squared error, the number of epochs is 1000, the batch size is 200, and the validation split is 0.2. Figures 6 and 7 show the temporal evolution of predicted values obtained by LSTM-attention, SLP, ARMA, and SARIMA compared to the ground truth data. Obviously, LSTM-attention with news sentiment analysis can capture both of the fluctuation behavior and the trend of the ground truth data while other methods can not. Note that RMSEs of ARIMA, SARIMA, and SLP are 0.0101, 0.0124, 0.0082, which are 2.6, 3.2, and 1.2 times larger than LSTM-attention.

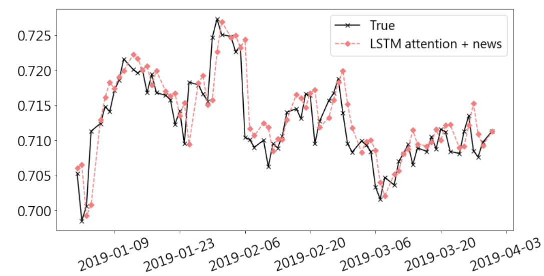

Fig. 6: A comparison of the ground truth and predicted next day currency exchange rates obtained by using LSTM-attention with news.

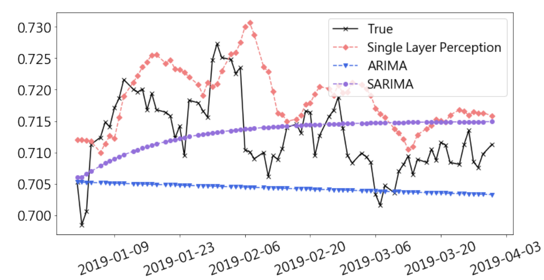

Fig. 7: A comparison of the ground truth and predicted next day currency exchange rates obtained by using the ARIMA and SARIMA models and SLP.

D. Learning curves

Fig. 8 shows three learning curves obtained by using different training set sizes, from one year, two years, to three years. As shown in this figure, the convergence of the validation loss on training data for three years is the fastest. Meanwhile, as shown in Fig. 9, the value of the performance loss of training data for three years is smaller than two other cases. Therefore, all the above results were obtained by using three years of the training data.

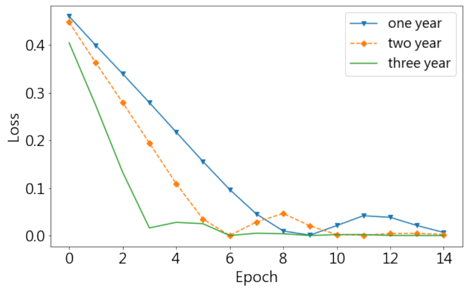

Fig. 8: Our training data validation

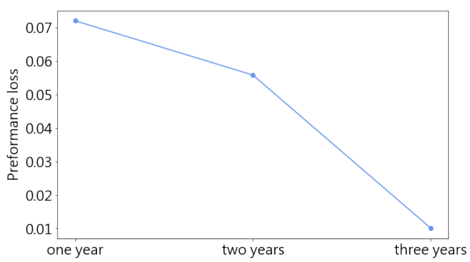

Fig. 9: A comparison of the performance losses for three input data sets.

E. The currency exchange prediction for next week or longer

Next, we study the problem of the currency exchange prediction for next week or longer. Table IV summarizes the test error measured by RMSE and MAPE by using two strategies for the next week, two-week, and month prediction. The input features for LSTM-attention used are the seven-day historical data plus ratio as well as news. The data of this table shows that the second strategy is a better choice than the first one. Instead of reusing the next-day predictive model, we would suggest building a newly trained model for the target task. Figs. 10, 11, and 12 show a comparison of time-series of the predicted values by two strategies against the ground truth data for next week, next two-week, and next month, respectively.

TABLE IV: The test error of two strategies for longer day value prediction.

		First strategy	Second strategy
	Output/Input	H+R+news	H+R+news
RMSE	next week	0.0077	0.0071
	next two-week	0.0147	0.0126
	next month	0.0322	0.0280
MAPE	next week	0.0090	0.0082
	next two-week	0.0184	0.0141
	next month	0.0442	0.0329

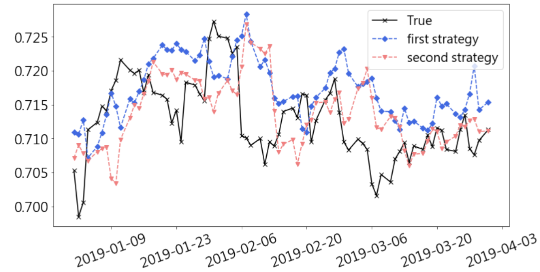

Fig. 10: The performance of two strategy for the next week prediction.

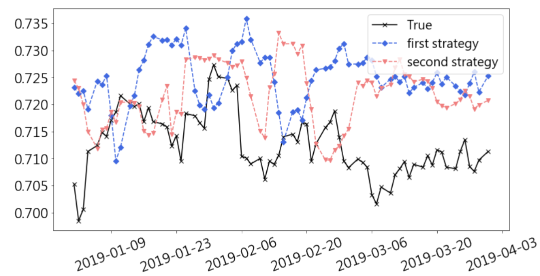

Fig. 11: The performance of two strategies for the next biweek prediction.

V. Conclusions

In this study, we discussed the problem of next day exchange rate prediction by using LSTM-attention and news sentiment analysis as well as the cases for the next week, two-week, and month prediction. For input features, we compared the use of historical exchange rate data from the last 7, 30, or 60 days and showed that the last seven-day historical data with ratio data outperforms historical only input. Besides, simple sentiment analysis with SnowNLP and keyword matching with “up/increase” in the news article can improve performance by 15%. We compared the performance with three baseline models, including ARIMA, SARIMA, SLP, and showed that LSTM-attention model achieved the best performance in terms of RMSE and MAPE. To predict the more extended future

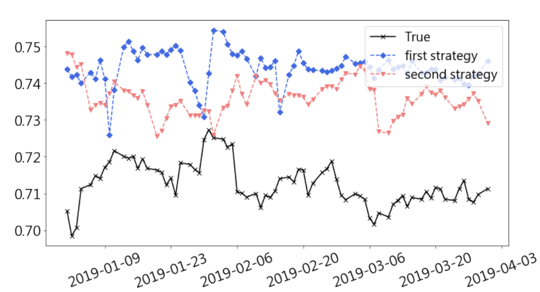

Fig. 12: The performance of two strategies for the next month prediction.

exchange rate, we retrained three new models for each case and compare the performance with the result based on the next day model. The result showed the newly trained models outperformed the next-day model because of the accumulated errors based on a single model.

References

[1] J. Walmsley, *International Money and Foreign Exchange Markets: an Introduction.* Wiley, 1996.

[2] O. V. Kryuchin, A. A. Arzamastsev, and K. G. Troitzsch, "The prediction of currency exchange rates using Artificial Neural Network," *Fachbereich Informatik, Nr. 4*, 2011.

[3] S. Galeshchuk, "Neural networks performance in exchange rate prediction," *Neurocomputing*, vol. 172, pp. 446–452, Jan. 2016.

[4] F.-M. Tseng, G.-H. Tzeng, H.-C. Yu, and B. J. Yuan, "Fuzzy ARIMA model for forecasting the foreign exchange market," *Fuzzy Sets. Syst.*, vol. 118, pp. 9–19, Feb. 2001.

[5] J. Kamruzzaman and R. A. Sarker, "Forecasting of currency exchange rates using ANN: A case study," in *Proceedings of the 2003 International Conference on Neural Networks and Signal Processing*, vol. 1, no. 793–797. IEEE, 2003.

[6] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, "Combining technical analysis with sentiment analysis for stock price prediction," in *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing.* IEEE, 2011, pp. 800–807.

[7] Y. Liu, Z. Qin, P. Li, and T. Wan, "Stock volatility prediction using recurrent neural networks with sentiment analysis," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems.* Springer, 2017, pp. 192–201.

[8] Q. Zhuge, L. Xu, and G. Z., "LSTM neural network with emotional analysis for prediction of stock price." *Eng. Lett.*, vol. 25, pp. 167–175, May 2017.

[9] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, July 2017, e0180944.

[10] "Snownlp," 2015, https://pypi.org/project/snownlp/.

[11] "Keras," 2015, https://keras.io.