Journal of Institute of Science and Technology, 24(2), 17-21 (2019)
ISSN: 2469-9062 (print), 2467-9240 (e)
Doi: http://doi.org/10.3126/jist.v24i2.27247

Research Article

ANALYSIS OF GRADIENT DESCENT OPTIMIZATION TECHNIQUES WITH GATED RECURRENT UNIT FOR STOCK PRICE PREDICTION: A CASE STUDY ON BANKING SECTOR OF NEPAL STOCK EXCHANGE

Arjun Singh Saud[1*], Subarna Shakya[2]

[1]*Central Department of Computer Science & Information Technology, Tribhuvan University, Kirtipur, Nepal*
[2]*Institute of Engineering, Pulchowk Campus, Tribhuvan University, Pulchowk, Lalitpur, Nepal*
[*]*Corresponding author: arjunsaud@cdcsit.edu.np*

*(**Received:** March 14, 2019; **Revised**: December 10, 2019; **Accepted:** December 20, 2019)*

ABSTRACT

The stock price is the cost of purchasing a security or stock in a stock exchange. The stock price prediction has been the aim of investors since the beginning of the stock market. It is the act of forecasting the future price of a company's stock. Nowadays, deep learning techniques are widely used for identifying the stock trends from large amounts of past data. This research has experimented two big and robust commercial banks listed in the Nepal Stock Exchange (NEPSE) and compared stock price prediction performance of GRU with three widely used gradient descent optimization techniques: Momentum, RMSProp, and Adam. GRU with Adam is more accurate and consistent approach for predicting stock prices from the present study.

Keywords: Stock prediction, GRU, Momentum, RMSProp, Adam

INTRODUCTION

The stock price is the cost of purchasing a security or stock in a stock exchange. A company's stock price often boils down to the matter of supply and demand. The stock price prediction is the act of forecasting the future price of a company's stock. Early models suggested that the stock price cannot be predicted with more than 50 % accuracy (Fama, 1970). Several studies provided pieces of evidence contrary to EMH and random walk hypotheses (Prechter & Parker, 2007). These studies showed that we can predict the stock market to some degree. Technical analysis is the method of forecasting the future stock price movements based on examination of past price movements. A similar type of approach was also used by machine learning approaches for predicting stock price. It is difficult for investors to analyze and forecast the market from a massive volume of past data. Therefore, many Artificial Intelligence (AI) techniques have been investigated to predict stock trends automatically. Some of the first research on the prediction of stock prices dates back to 1994, where a comparative study of machine learning regression models were performed (Refenes *et al.*, 1994).

Machine learning models are capable of learning a function by looking at the data without explicit programming, and performance of these algorithms depends heavily upon the representation of the data they are given (Goodfellow *et al.*, 2017). Unfortunately, the time-series data of the stock market cannot be mapped easily. It is best described as a random walk, which makes the feature engineering and prediction much harder. Therefore deep learning models are the best available tool for computing such nonlinear and complex systems (Heaton *et al.*, 2017).

There are many gradient descent optimization approaches (Ruder, 2017). This research paper has compared stock price prediction performance of three widely used gradient descent optimization techniques: Momentum, RMSProp, and Adam. All of these optimization techniques are used with gated recurrent unit (GRU) networks. Two banking stocks, listed on Nepal Stock Exchange (NEPSE), studied in this research work are: Everest Bank Limited (EBL) and Himalayan Bank Limited (HBL). Main objectives of this research work are to predict stock prices of EBL and HBL using GRU with Momentum, RMSProp & Adam, and to analyze stock price prediction accuracy of the methods as mentioned above and suggest best gradient descent optimization technique that can be used with GRU.

MATERIALS AND METHODS

This research work was carried out by adopting an empirical research method and experimental research design. Data from August 8, 2007 to November 5, 2018 was used in this research work. This data was pre-processed before using it for training and testing the GRU network. First, the raw data were converted into coma separated value (CSV) file, then the required stock data were extracted, and attributes date, and stock symbol was dropped. Next, missing values of low, high, open, and close prices were replaced by average of previous and next day's price, other data like High-Low and Close-Open was computed. Technical indicators like 3day MA, 10day MA, 30day MA, Standard Deviation, Relative Strength Index, and William % R was also calculated, and prediction attribute "Next day's Close Price" was added, which was merely Close price shifted back by one position. Then, data were scaled by using z-score normalization and train/test split was done at 9:1 ratio. Finally, the stock price was predicted, which was in the scaled form and hence difficult to understand. The predicted result was converted back to its original form using inverse transformation, and was plotted. Configuration of GRU network used in this research work was 14×50×50×50×50×1.

Vargas *et al.* (2017) proposed an RCNN model. The results showed that sentence embedding is better than word embedding, RCNN is better than CNN, and the influence of technical indicators leads to better performance. Minh *et al.* (2018) proposed TGRU network. The experiments showed that TGRU could outperform GRU and LSTM. Huynh *et al.* (2017) proposed BGRU for predicting stocks and assessed the impact of the financial news on the price of the stock. The experiment showed that the highest accuracy was obtained in the first 24 hours and accuracy achieved with BGRU was the highest. These researches clearly showed the use of GRU networks and technical indicators in predicting stock prices. Therefore, this research work had incorporated technical indicators in the dataset and used the GRU network for predicting stock prices.

Gated recurrent unit networks

Gated recurrent unit is a type of RNN cell and aims to solve the vanishing gradient problem which comes with a standard recurrent neural network (Haykin, 2008). GRU uses two gates: Update gate and reset gate (Cho *et al.*, 2014). There are fewer parameters used by GRU than LSTM, and hence it can be trained faster and need less data for generalization. But, with a large volume of data, LSTM with higher expressiveness may lead to better results (Weiss *et al.*, 2018). Formulation and architecture of GRU (Fig. 1) are given below.

$$Z_t = \sigma(W_z x_t + U_z H_{t-1}) \qquad R_t = \sigma(W_r x_t + U_r H_{t-1})$$

$$H_t^{'} = \tanh(W_h x_t + (R_t \times H_{t-1})U_h) \quad H_t = (Z_t \times H_t^{'}) + ((1 - Z_t) \times H_{t-1})$$

where, Z and R are update and reset gates and $H_t^{'}$ *and* H_t *are candiadte hidden state and hidden states*

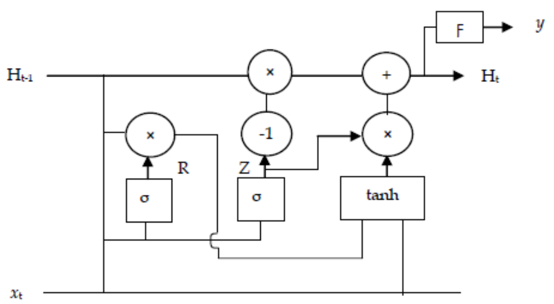

Fig. 1. Architecture of GRU

Concept of gradient descent

Mini-batch gradient descent updates parameters with just a subset of examples. The direction of updates has some variance (Fig. 2). Therefore the path taken by mini-batch gradient descent oscillates toward convergence (Goodfellow *et al.*, 2016).

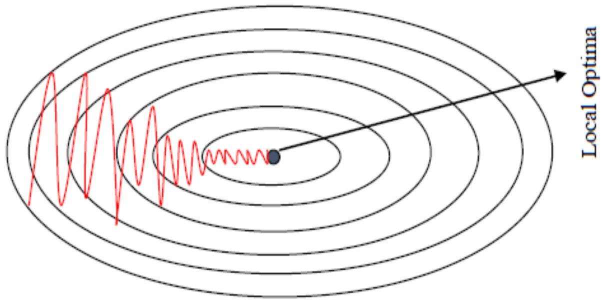

Fig. 2. The convergence of gradient descent

Momentum

Gradient Descent with Momentum uses the past gradients to smooth out the update (Fig. 3). It computes an exponentially weighted average of gradients and then uses this gradient to update the weights. It updates values of parameters w and b during the back propagation phase using dw and db, as given below (Goodfellow *et al.*, 2017).

$$w = w \quad \alpha V_{dw} \qquad b = b \quad \alpha V_{db}$$

$$V_{dw} = \beta V_{dw} + (1 \quad \beta) dw \quad V_{db} = \beta V_{db} + (1 \quad \beta) db$$

Whereβ is momentumwhose vaalue ranges from 0 to 1 and dw and db is gradient

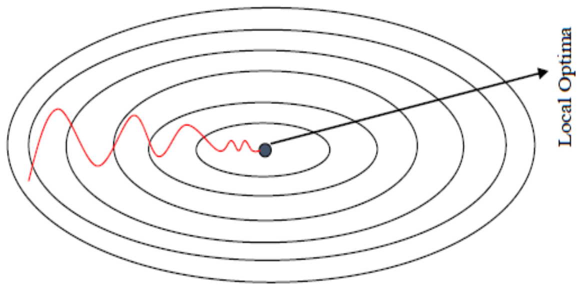

Fig. 3. The convergence of GD with Momentum

RMSProp

RMSProp stands for root mean squared propagation. It was first described in a Coursera class on neural networks taught by Geoffrey Hinton (Hinton et al., 2012). The central idea of RMSProp is to keep the moving average of the squared gradients for each weight. And then divide the gradient by root mean square of the squared gradients. It allows us to use a high learning rate (α) to speed up learning. RMSProp updates coefficients using equations given below.

$$\mathrm{V}_{dw} = \beta \mathrm{V}_{dw} + (1-\beta)dw^2 \qquad \mathrm{V}_{db} = \beta \mathrm{V}_{db} + (1-\beta)db^2$$

$$\mathrm{w} = \mathrm{w} - \alpha \frac{dw}{\sqrt{V_{dw} + \in}} \qquad \mathrm{b} = \mathrm{b} - \alpha \frac{db}{\sqrt{V_{db} + \in}}$$

Adam

Adam stands for Adaptive Momentum estimation. It is a combination of RMSProp and momentum optimization techniques. This technique accurately calculates an exponential moving average of the gradient and the squared gradient, and the parameters β_1 and β_2 control the decay rates of these moving averages (Goodfellow *et al.*, 2017).

$$\mathrm{V}_{dw} = \beta_1 \mathrm{V}_{dw} + (1-\beta_1)dw \qquad \mathrm{V}_{db} = \beta_1 \mathrm{V}_{db} + (1-\beta_1)\, db$$

$$\mathrm{S}_{dw} = \beta_2 V_{dw} + (1-\beta_2)dw^2 \qquad \mathrm{S}_{db} = \beta_2 \mathrm{S}_{db} + (1-\beta_2)db^2$$

If we choose large values of β_1 and β_2, the contribution of initial terms will be very less in moving averages. This results in biased estimates of moving averages. To overcome such situation, Adam adds a bias correction term. Finally, the parameters w and b was updated as below.

$$\mathrm{V}_{dw} = \frac{\mathrm{V}_{dw}}{(1-\beta_1^t)} \quad \mathrm{V}_{db} = \frac{\mathrm{V}_{db}}{(1-\beta_1^t)} \quad \mathrm{S}_{dw} = \frac{\mathrm{S}_{dw}}{(1-\beta_2^t)} \quad \mathrm{S}_{db} = \frac{\mathrm{S}_{db}}{(1-\beta_2^t)}$$

$$\mathrm{w} = \mathrm{w} - \alpha \frac{V_{dw}}{\sqrt{S_{dw} + \in}} \qquad \mathrm{b} = \mathrm{b} - \alpha \frac{V_{db}}{\sqrt{S_{db} + \in}}$$

RESULTS AND DISCUSSION

Stock Price Prediction

In this research work, stock prices were predicted using GRU with three variants of gradient descent optimization: Momentum, RMSProp, and Adam. For each case, ten experiments were performed, which resulted in 60 prediction curves. Therefore, only best-fitted curves are presented here for each gradient descent optimization technique. From Figs. 4 to 9, it can be seen that GRU with Adam can predict stock prices of both stocks (EBL and HBL) more accurately than GRU with RMSProp and GRU with momentum.

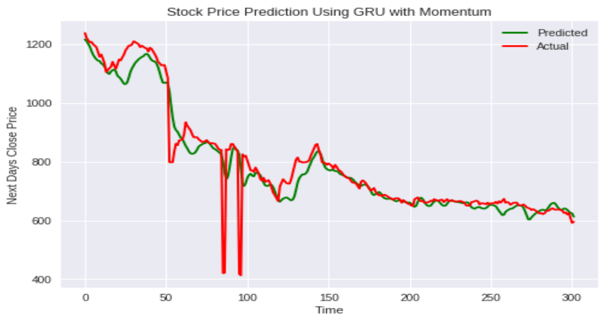

Fig. 4. Best fitted EBL price prediction curve using momentum

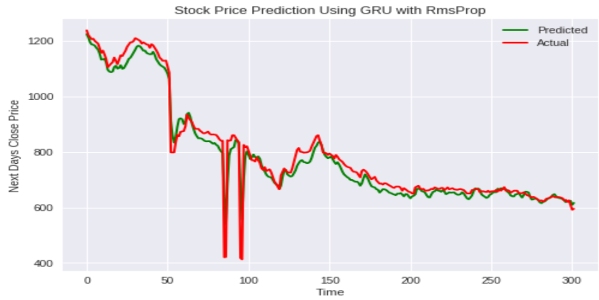

Fig. 5. Best fitted EBL price prediction curve using RMSProp

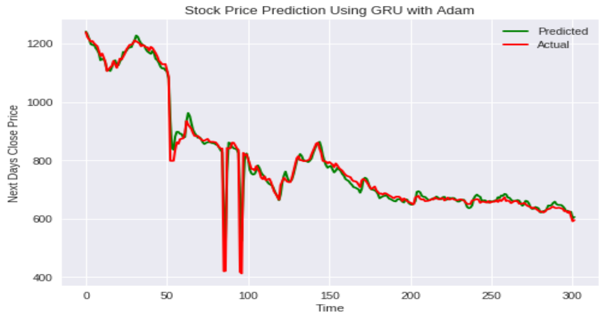

Fig. 6. Best fitted EBL price prediction curve using Adam

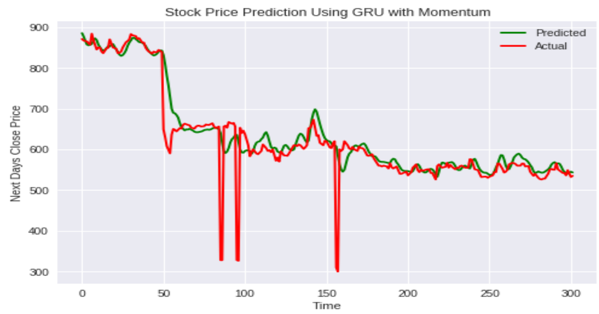

Fig. 7. Best fitted HBL price prediction curve using momentum

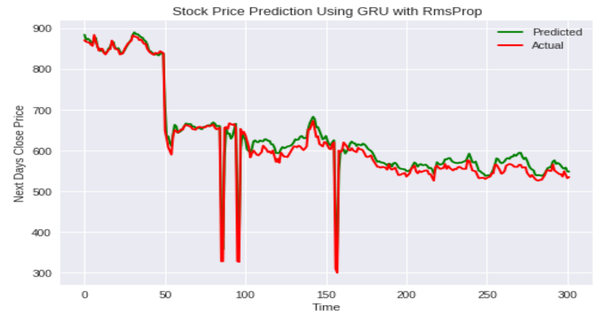

Fig. 8. Best fitted HBL price prediction curve using RMSProp

Best MAPE obtained from GRU with Adam was 1.15 and 2.13 for EBL and HBL, respectively. On the other hand, GRU with RMSProp predicted stock prices of EBL and HBL with MAPE 2.27 and 3.08, respectively, and GRU with momentum was able to predict stock prices of EBL and HBL with MAPE 2.90 and 4.59, respectively.

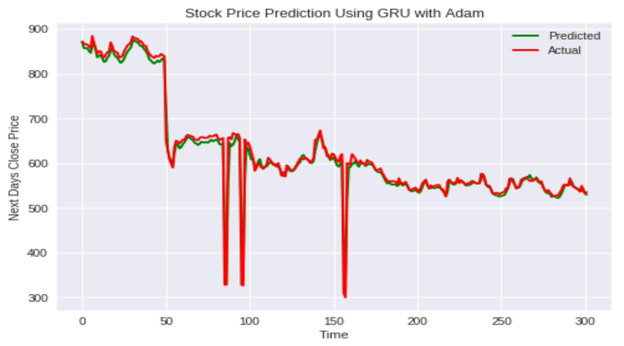

Fig. 9. Best fitted HBL price prediction curve using Adam

Analysis of prediction MAPEs

It can be seen that minimum value of MAPE for stock price prediction of EBL using GRU with Adam was lower than the minimum amount of MAPE for the EBL price prediction using GRU with RMSProp and GRU with momentum, as shown in Fig. 10(a). A similar type of pattern was also observed for HBL. GRU with momentum is worst in terms of minimum MAPE's obtained during stock price prediction. Again, it can see from Fig. 10(b) that maximum values of MAPEs for stock price prediction of EBL and HBL using GRU with Adam was lower than maximum values of MAPEs obtained from GRU with RMSProp and GRU with momentum. GRU with RMSProp is worst in terms of maximum MAPE's obtained during stock price prediction. Finally, from Fig. 10(c), we can see that average values of MAPEs for stock price prediction of EBL and HBL using GRU with Adam is lower than average amounts of MAPEs obtained with GRU with RMSProp and GRU with momentum. GRU with RMSProp is worst in terms of average MAPE's obtained during stock price prediction.

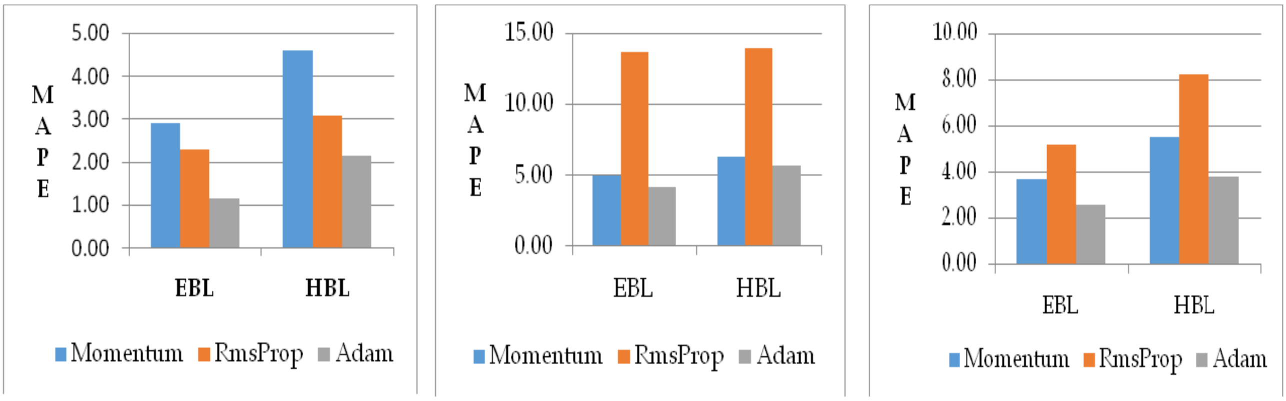

Fig. 10. MAPE's obtained for EBL and HBL: (a) minimum MAPE's (b) maximum MAPE's (c) average MAPE's

CONCLUSION

Prediction of stock market is the field in which investors and academicians have a keen interest. Recently, stock price prediction using deep learning has attracted many researchers and investors because it can give higher prediction accuracy. This research work has compared stock price prediction accuracy obtained from GRU with momentum, RMSProp, and Adam.

The research work showed that GRU with Adam could outperform GRU with Momentum and GRU with RMSProp in each aspect. Another exciting conclusion that can be made from the analysis of prediction MAPE's is that RMSProp is the most inconsistent gradient descent optimization technique while predicting stock prices. Because RMSProp has shown the worst prediction performance in terms of average MAPE's and maximum MAPE's even though minimum MAPE's obtained with RMSProp is better than minimum MAPE's got with Momentum.

REFERENCES

Cho, K., Bart, V., Caglar, G., Dzmitry, B., Fethi, B., & Holger, S. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724-1734), Association for Computational Linguistics, Doha.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383-417.

Goodfellow, I., Bengio, Y., Courville, A., & Farancis, B. (2017). *Deep learning- adaptive computation and machine learning series.* Cambridge, USA: MIT Press.

Haykin, S. O. (2008). *Neural network and learning machines* (3rd ed.). London, United Kingdom: Pearson Education.

Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry, 33*(1), 3-12.

Hinton, G., Srivastava, N., & Swersky, K. (2012). *Rmsprop: normalize the gradient.* Retrieved from Neural Networks for Machine Learning. https://www.youtube.com/watch?v=defQQqkXEfE

Huynh, H. D., Dang, L., & Duong, D. (2017). A new model for stock price movements prediction using deep neural network. *8th International Symposium on Information* (pp. 57-62), ACM, Nha Trang.

Minh, D. L., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep learning approach for short-term stock trends prediction based on the two-stream gated recurrent unit network. *IEEE Access*, *6*, 55392-55404.

Prechter, J., & Parker, W. (2007). The financial/economic dichotomy in social behavioural dynamics: the socioeconomic perspective. *The Journal of Behavioral Finance*, *8*(2), 84-108.

Refenes, A. N., Zapranis, A., & Francis, G. (1994). Stock performance modelling using neural networks: A comparative study with regression models. *Neural Networks*, *7*(2), 375-388.

Ruder, S. (2017). Machine learning: An overview of gradient descent optimization algorithms. Cornell University. Retrieve from https://arxiv.org/pdf/1609.04747.pdf

Vargas, M., Lima, B., & Evsukoff, A. (2017). Deep learning for stock market prediction. *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* (pp. 60-65), IEEE, Annecy.

Weiss, G., Goldberg, Y., & Yahav, E. (2018). On the practical computational power of finite precision RNNs for language recognition. *56th Annual Meeting of the Association for Computational Linguistics* (pp. 740-745), Association for Computational Linguistics, Melbourne, Australia.