J. Chosun Natural Sci.
Vol. 11, No. 4 (2018) pp. 204 – 208
https://doi.org/10.13160/ricns.2018.11.4.204

Comparison of Different Deep Learning Optimizers for Modeling Photovoltaic Power

Prasis Poudel[1], Sang Hyun Bae[2], and Bongseog Jang[1†]

Abstract

Comparison of different optimizer performance in photovoltaic power modeling using artificial neural deep learning techniques is described in this paper. Six different deep learning optimizers are tested for Long-Short-Term Memory networks in this study. The optimizers are namely Adam, Stochastic Gradient Descent, Root Mean Square Propagation, Adaptive Gradient, and some variants such as Adamax and Nadam. For comparing the optimization techniques, high and low fluctuated photovoltaic power output are examined and the power output is real data obtained from the site at Mokpo university. Using Python Keras version, we have developed the prediction program for the performance evaluation of the optimizations. The prediction error results of each optimizer in both high and low power cases shows that the Adam has better performance compared to the other optimizers.

Keywords: Deep Learning, Deep Learning Optimizer, Prediction, Photovoltaic Power Output, Long Short-Term Memory, Neural Networks

1. Introduction

Currently, energy is becoming an emergency problem for economic development in every country in the world. Consumption of traditional fossil fuels such as gasoline and coal caused massive negative effects on the environment, which includes air pollution as well as a greenhouse effect. In order to solve this problem, an enormous application of renewable energy is crucial. Among the various renewable energy technologies, solar energy is most promising, almost limitless, non-pollutant and available free of cost. Massive amount electrical energy could harvest from solar energy using photovoltaic (PV) systems. However, the power output of the PV system is intermittent and non-stationary random process because of the variability of solar irradiation and weather characteristics.

Solar PV power system is an important energy source in Microgrid (MG) system and controlling PV system is needed for optimal performance achievement for MG. Therefore, the modeling and forecasting of a power output of the PV system have been considered as an interesting research topic for MG.

Recently, artificial intelligence based forecasting techniques have been used successfully in many areas such as finance and banking-insurance[1] for analysis of exchange rate evaluation, stock price forecasting, industrial and agricultural production, and medical sectors [2,3]. Numerous forecasting modeling techniques are also applied in PV power output: artificial neural network (ANN) based model[4], time series model[5], and time trend extrapolation model[6]. Among these models, the ANN based modeling technique has more accurate prediction results compared to other modeling methods. However, the ANN based modeling technique is complex and requires huge training data samples.

In this paper, long short-term memory (LSTM) networks[7] have been tested with different optimizers to model PV power output. The purpose of this work is to examine the numerous optimization in deep learning for PV modeling technique and find the best model, which can be deployed in the campus MG management system.

This paper is organized as follows, PV modeling approach using different optimizers in LSTM networks

[1]Department of Multimedia Engineering, Mokpo Nat'l University
[2]Department of Computer Science & Statistics, Chosun University, Gwangju

[†]Corresponding author : jang@mokpo.ac.kr
(Received : December 10, 2018, Revised : December 14, 2018, Accepted : December 15, 2018)

is described in section 2 and prediction results obtained from these optimizers are presented in section 3. Finally, a conclusion is stated in section 4.

2. PV Modeling Using Different Optimizers of LSTM Deep Learning Network

2.1. Long Short-Term Memory Network

LSTM network is based on recurrent neural networks (RNN) and proven to be one of the most powerful RNN models for time series forecasting and other related applications[8,9]. The LSTM networks can be constructed in such a way that they are able to remember long-term relationships in the data. The LSTM networks have been shown to model temporal sequences and their long-range dependencies more accurately than the original RNN model[8]. LSTM networks are applied on sequential data as input, which without loss of generality means data samples that change over time. Input into LSTM networks involves a so-called sequence length parameter (i.e., the number of time steps) that is defined by the sample values over a finite time window[10]. Thus, the sequence length is how we represent the change in the input vector over time; this is the time series aspect to the input data. The architecture of LSTM networks is generally composed of units called memory blocks. The memory block contains memory cells and three controlling gates, i.e., the input gate, forget gate, and output gate.

The memory cell is designed for preserving the knowledge of the previous step with self-connections that can store (remember) the temporal state of the network while the gates control the update and flow direction of information. The structure of LSTM used is shown in Fig. 1.

The LSTM shown in above fig has inputted PV power data Xt, the LSTM updates the memory cell with the help of three gates it, ft and ot, and give an output Yt where t represents the time period[11].

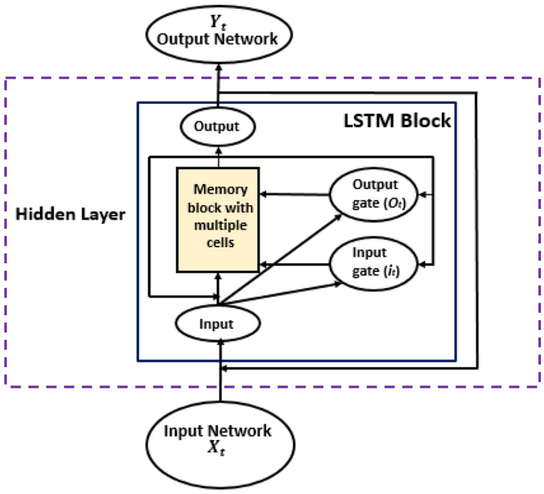

Fig. 1. LSTM learning structure.

2.2. Optimization Algorithms

The internal parameters of a deep learning model play a vital role in efficiently and effectively training a model and produce accurate results. This is why we use various optimization strategies and algorithms to update and calculate appropriate and optimum values of such model's parameters which influence our model's learning process and the output of the PV model.

Types of Optimizations Algorithms

1) Stochastic Gradient Descent (SGD)

SGD updates deep learning model parameters (*theta*(θ)) in the negative direction of the gradient (g) by taking a subset or a mini-batch of data size (n)

$$gradient\ (g) = \frac{1}{n}\nabla_{\theta}\sum_i L\,(f\,(x(i);\ \theta),\ y(i)) \tag{1}$$

$$\theta = \theta - \varepsilon_k \times g \tag{2}$$

In the equations 1 and 2, the deep learning model is represented by ($x(i)$; θ), where $x(i)$ are the training data and $y(i)$ are the training labels, the gradient of the loss L is computed with respect to the model parameters *theta.* Learning rate ε_k determines the size of the step that the SGD algorithms takes along the gradient.

2) Adaptive Gradient Algorithm (AdaGrad)

AdaGrad is an adaptive method for setting the learning rate i.e. it maintains a per-parameter learning rate that improves performance on problems with sparse gradients.

3) Root Mean Square Propagation (RMSProp)

RMSProp maintains per-parameters learning rate that is adapted based on the averages of recent magnitudes of the gradients for the weights values like how quickly it is changing.

4) Adam and variants (Adamax & Nadam)

Adam is a combination of RMSProp and momentum, the update looks like RMSProp expect that a smooth version of the gradient is used instead of the raw stochastic gradient, the Adam update also includes a bias correction mechanism[12].

3. Prediction Results

The input data for the LSTM model is PV power output from the PV system installed at Mokpo National University rooftop in Korea. Initially, we examine the whole March month PV system generated power data in order to find the similar power generation. From this study, we found march 1st and 6th have similar power generation which is highly fluctuating in nature and march 11th and 24th have a similar smooth power generation, as shown in Fig. 2 and 3 respectively.

The LSTM network having a structure of three hidden layers, first hidden layer nodes 200, second hidden layer with 100 nodes and third 50 nodes was implemented using Python Keras Library and was trained with the different optimizers. We set all the parameters for these optimizers as a default value.

Firstly, we used high fluctuating PV power data of March 1st as a training dataset, and the March 6th PV data as a test dataset in LSTM network with different optimizers. Similarly, low fluctuating PV power data of March 11th as a training dataset, and the March 24th PV data as a test dataset. The Fig below 4 and 5 shows the value of the training loss vs iterations of high fluctuating PV data and low fluctuating PV data respectively with different optimizers.

From the training loss versus epochs plot, we can see that in both cases Adam optimizer produce the lowest training loss with highly converged then the other optimizers.

In our experimental study, we used the root mean square error (RMSE) measures technique to evaluate

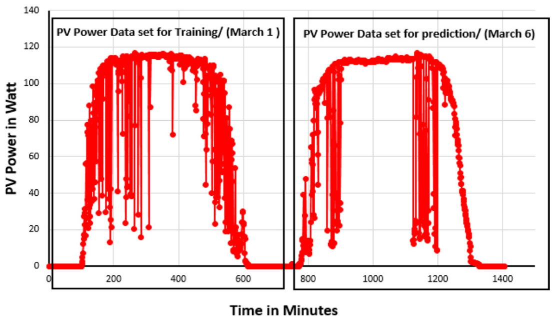

Fig. 2. High fluctuation PV output cases.

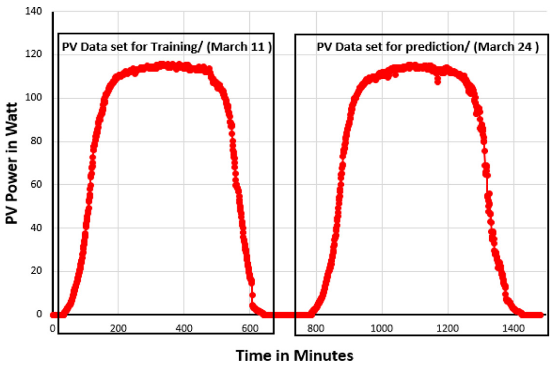

Fig. 3. Low fluctuation PV output cases.

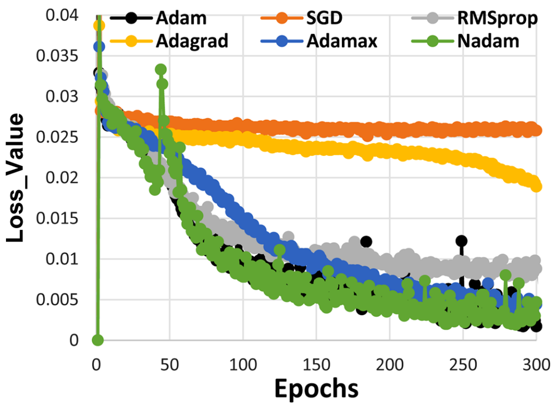

Fig. 4. Training losses for the high fluc. case

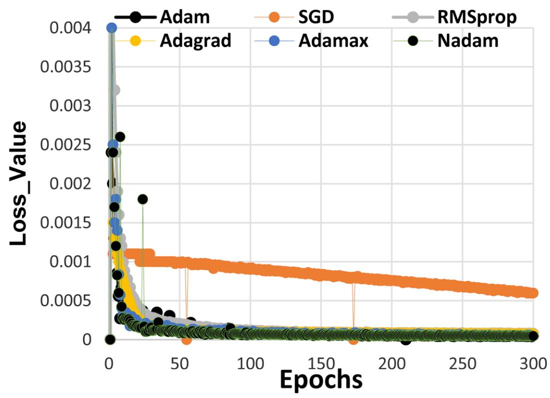

Fig. 5. Training losses for the low fluc. case

the forecasting accuracy, as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y^{(i)} - \hat{y}^{(i)})^2}{N}} \tag{3}$$

Where, $y^{(i)}$ and $\hat{y}^{(i)}$ are the target and forecast (outcomes) values, respectively, of PV power data with a total N observations.

To evaluate the forecasting accuracy of different optimizers used in LSTM network, RMSE of all the optimizers have been evaluated in both cases. At 300 epochs for the training process, the training loss values are converged and the training is stopped and trained model is used for the prediction. The RMSE values obtained from different optimizers during the training and prediction of both fluctuating and low fluctuating cases are shown in Fig. 6 and 7 respectively.

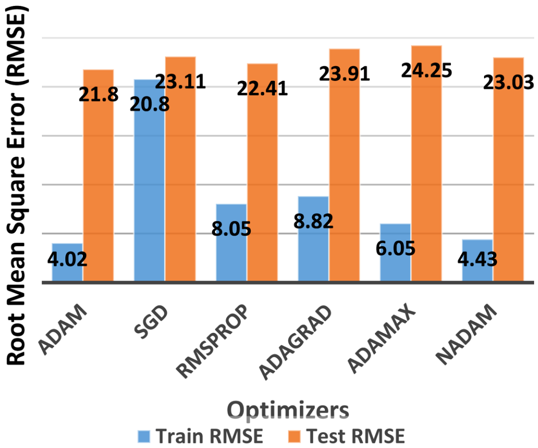

Fig. 6. RMSEs for the high fluctuation case.

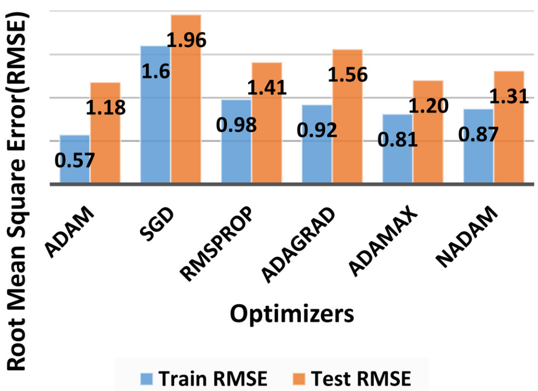

Fig. 7. RMSEs for the low fluctuation case.

The results in Fig. 6 and 7 demonstrate the effectiveness of Adam optimizers in improving forecasting performance in terms of RMSE against the other optimizers in both high and low fluctuation cases of PV power Data. In high fluctuating case of PV data, Adam optimizer has minimum error values of 4.02 RMSE for the training and 21.8 RMSE for the prediction. Similarly, in low fluctuating PV data, Adam optimizer has the lowest error values 0.57 in training RMSE and 1.18 prediction RMSE.

In the high fluctuation PV output case, all optimizer has similar performance results. Due to the sudden PV output changes, the LSTM structure is even hard to find the good results.

4. Conclusion

In this paper, different optimizers namely, Adam, SGD, RMSProp, AdaGrad, Admax, and Nadam have been used to model PV power output in LSTM network. First, we trained each optimizer in both high fluctuating and low fluctuating PV power data to observe the training loss value and used in prediction phase. In both cases, Adam optimizer produces the lowest training loss with highly converged rates. Training and prediction results in both cases also show that Adam optimizers have the better performance compared to the other optimizers in terms of trained and predicted loss values via RMSE.

Also these results imply that the high fluctuation PV output due to the cloud movement with some higher speed is eventually hard to model accurately using any time-series or deep neural network models. The cloud movement is very natural event happenings in some local region and area only. Thus it is hard to imply the movement in real-time modeling methods. To model this cloud movement, we suggest that using multiple irradiation sensors position within the PV plants at some distances. This is the cheapest modeling method to predict very sudden and random the nature events.

Acknowledgements

This research work was supported by research funds of Mokpo National University in 2017.

References

[1] R. Adhikari, and R. K. Agrawal, “A combination of artificial neural network and random walk models for financial time series forecasting,” Neural Computing and Application., Vol. 24, pp. 1441-1449, May. 2014.

[2] P. Das, and S. Chaudhury, “Prediction of retail sales of footwear using feed-forward and recurrent neural networks,” Neural Computing and Application., Vol. 16, pp. 491-502, May. 2007.

[3] G. Biricik, O. O. Bozkurt, and Z. C. Taysi, “Analysis of features used in short-term electricity prices forecasting for deregulated markets,” IEEE Trans. Signal Processing and Communications applications conference (SIU)., pp. 600-603, May. 2015.

[4] A. Shah, S. C. Kaushik, and S. N. Garg, “Assessment of diffuse solar energy under general sky condition using artificial neural network,” Applied Energy, Vol. 86, No. 4, pp. 554-564, 2009.

[5] Y. Cui, Y. C. Sun, and Z. L. Chang, “A review of short-term solar photovoltaic power generation prediction methods,” Resources Science, Vol. 35, No. 7, pp. 1474-1481, 2013.

[6] Y. Li, L. He, and J. Niu, “Forecasting power generation of grid-connected solar PV system based on Markov chain,” Acta Energiae Solaris Sinica, Vol. 35, No. 4, pp. 611-616, 2014.

[7] H. Sak, Andrew Senior, and F. Baufays, “Long Short- Term Memory Recurrent Neural Network Architecture for Large Scale Acoustic Modeling,” Proceedings of the Annual Conference of the International Speech Communication Association, 2014.

[8] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, “An overview and comparative analysis of recurrent neural networks for short term load forecasting,” 2017.

[9] V. P. Utgikar and J. P. Scott, “Energy forecasting: Predictions, reality and analysis of causes of error,” Energy Policy, Vol. 34, No. 17, pp. 3087-3092, 2006.

[10] Z. C. Lipton, J. Berkowitz, and C. Elkan, “A critical review of recurrent neural networks for sequence learning,” 2015.

[11] P. Poudel, and B. Jang, “Solar Power Prediction Using Deep Learning Technique”, Advanced Science and Technology Letters, Vol.146 (FGCN 2017), pp.148-151. http://onlinepresent.org/proceedings/vol146_2017/26.pdf), Nov. 2017.

[12] D. Kingma, and J. Ba, “Adam: A Method for Stochastic Optimization”, ICLR, 2015.