Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) Network Models for Forecasting Energy Consumptions

Samir M. Shariff

Abstract — Unlike other sources of energy, electricity can’t be stored. Therefore, an estimation of Energy Consumption (EC) with good accuracy is required to manage demand and supply in the smart grid. Not only good accuracy, but reliability is also on-demand in the prediction model to optimize resource allocation. Therefore, in this study we have implemented and examine two different models: a machine learning model, Autoregressive Integrated Moving Average (ARIMA), and a deep learning-based model Long Short-Term Memory (LSTM). Although ARIMA showed powerful statistical analysis and less robustness, LSTM demonstrated highly accurate results which may stop us to lead false alarming of over-demand and low consumption of energy. In last, we have concluded our result by presenting significant improvement in forecasting energy by LSTM using various evaluation criteria e.g., Mean Square Error (MSE), Root Mean Square Error (RMSE), and other normalized matrices.

Keywords — ARIMA; Demand-Supply; Energy Consumptions; Forecasting; LSTM.

I. INTRODUCTION

Electric vehicles are the revolution in terms of environmental pollution [1]. Stringent rules and guidelines are set by various countries from all over the world and formulated CO2 emission control. This increased demand for electric vehicles. These initiatives made by government’s has produced major demand to elevate the electric vehicle range for each charging cycle [2]. Furthermore, EV manufacturing companies have provided option to home charging, ultra-fast charging stations as well as on road charging capabilities. This will add another new feature to the convenience of consumers [3]. Even, various trending technology like cloud computing optimizes the charging infrastructure [4], but to achieve their demands of charging, we need a tremendous amount of power, and we can do that either by forming new power plants or by proper utilization of existing power plants by forecasting energy requirements for different areas.

We can use different tools stacks such as manual estimation or statistical analysis to forecast the energy demand. But, for the last few years, AI is playing an important role to forecast energy [5], [6]. And most of the work has already been done on this kind of project by using such stacks for example from machine learning [7], [8], they had used support vector machine (SVM) [9], different unsupervised learning models e.g., PCA [10] and through different deep learning models also and no doubt that they got the good results as well. But no one has a proper dataset, all they have hypothetical data, and there is no comparison between the different learning models which gives us better and more efficient results.

We used the dataset of Columbia and took from PJM Interconnection LLC which will eventually improve the prediction demand. Analysis and management of several activities such as market acquisition, day-end planning, unit commitments, and economic expenditure require the realization of future demand and load requirements [11]. Therefore, we worked on two different deep learning models i.e., LSTM [12] and ARIMA [13] models.

LSTM is a Memory Short Term network that is a special type of Recurrent Neural Network (RNN) and the latest version of Artificial Neural Network (ANN) that can learn long-term dependence. They work very well on many different problems and are now widely used. LSTMs are designed to avoid the problem of long-term reliance. Long-term memory is their automatic behavior, not something they strive to learn. LSTM can delete or add information in the next layer, carefully controlled by past results. Whereas ARIMA is known for representing Auto-Regressive Integrated Moving Averages. It is a model segment that captures a set of standard temporary frames in time series data. This Forecasting can also be used in the determination of coming sales in different businesses.

The manuscript is organized as follows: Section II, “Modeling”, describe the detail of the dataset used in the proposed model, models formulation, and evaluation criteria. Section III, “Implementation”, deals with the implementation of both the models. Furthermore, the results and discussion are given in Section IV. Section V gives the concluding remarks about the proposed approach and its performance.

II. MODELING

In this study, we have merged these models to forecast the consumption of energy; the ARIMA and LSTM. Modeling for both systems is divided into the following three steps: A) Data description and preprocessing, B) Model formulation, and C) Evaluation criteria that are described hereunder:
A. Data Description and Preprocessing

The data used in this study and analyzing the models were accessed from a regional transmission organization of Columbia named PJM Interconnection LLC. Data has hourly consumption of energy in Megawatts (MW). To formulate and implementation of the models, we have split the data into two parts: first, data of one year for training, and a month data for testing.

III. MODEL FORMULATION

A. ARIMA

ARIMA is not a novel approach as it is classical statistical analysis. While its model is based on a time series it combines both Auto Regression (AR) along with the concept of Moving Average (MA). We initially apply a test on the provided sample data. This will help us in providing the eligibility of these data before relating it to our ARIMA model. Here we choose to conduct our eligibility using the Augmented Dickey-Fuller (ADF) test.

1) ADF Test

This testing approach is done to check the stationarity of the provided sample data that is dependent to the unit root. This test is conducted using the following hypothesis, 

Null Hypothesis (H₀): A unit root exists.

Alternative Hypothesis (H₁): Unit root does not exist.

where, our statistical hypothesis root test is defined as:

\[ DF_t = \frac{\hat{\gamma}}{\sigma_\gamma} \]  

(1)

\( \hat{\gamma} \) can be obtained from the following equation:

\[ \Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \delta_1 \Delta Y_{t-1} + \cdots + \delta_p \Delta Y_{t-p+1} + \epsilon_t \]  

(2)

However, \( \alpha \) is a constant value, \( \beta \) is defined as the coefficient of time-trend, \( p \) is the number of autoregressive terms, and \( \sigma_\gamma \) is the standard deviation of \( \hat{\gamma} \). In our case, the value of H₀ is rejected while H₁ is accepted is demonstrated by the tabulated values of the sample data being higher than the tabulated values. In addition, the p-value is given as 0.05 that has an interval based incremental values. This shows an acceptable statistical rate of the values of hypothesis H₁.

Furthermore, our data is processed with no more limitation as to adopt the ARIMA model. We adjust our data to fit into an ionospheric containment. Hence, we will introduce the following parameter. Auto-Regressive Terms (p), Integrated-Process (d), and Moving-Average Terms (q).

2) Auto-Regressive Terms (p)

This parameter defines the linearity of time (t) dependence to the variable. This is represented by the following (3):

\[ Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} \cdots \alpha_p Y_{t-p} + \epsilon_t \]  

(3)

It is apparent in (3), that \( Y_t \) has a linear combination of the past time-series terms with the coefficient’s terms of \( \alpha \) as being \( \alpha_1, \alpha_2, \alpha_3 \ldots \alpha_p \) and \( \epsilon_t \) is a random shock at time \( t \).

3) Integrated-Process (d)

Most time-series are observed to have a trend (vertical, and horizontal) approach; this causes instability in the data. Furthermore, this instability may be reduced by differentiating the cumulative terms. Considering a high irregularity in the data, then difference of two successive terms can be a constant and is given in (4):

\[ \Delta Y_t = \epsilon_t \]  

(4)

4) Moving Average Terms (q)

This is a linear containment of various errors reported successively. Where \( q \) defines the number of possible terms that will reduce this error of each time interval \( t \) and is defined in (5):

\[ Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \theta_3 \epsilon_{t-3} \ldots - \theta_q \epsilon_{t-q} \]  

(5)

Considering (3), (4), and (5), will express the ARIMA model (p, d, q) as follows:

\[ Y_t = \delta + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} \ldots \alpha_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \theta_3 \epsilon_{t-3} \ldots - \theta_q \epsilon_{t-q} \]  

(6)

where \( \delta \) is a constant value.

B. LSTM

For the Long Short Term Memory has a distinctive approach as its time series data have both dependent and independent variables. These variables even though are dependent they depend upon an independent variable. That is the target attribute depends on it. An observer is to be installed in order to control the feedback. Similarly, the patterns in each time series of the provided data must follow a certain data of the EC. On regular basis, EC have to maintain certain levels (min/max) values of minor erratum. This erratum is very high particularly during peak events. For example, it is affected by regular and irregular usage such as weekends events, weekend activities, weather, etc.) ANN is an advanced process in such that it is able to recognize and self-learn the non-regular activities that are recorded and analyzed by the given data as well as the upcoming events. In this case the EC time series data will correlate with the past terms. For the next event we will introduce an RNN. The RNN is utilized to pass self-learned nodes that are nearby. Even though RNN gives good data for our analysis, but we introduce TEC. TEC will provide series for a long-term dependent value. These values will be stored in its memory. Our observer will correlate our patterns and provide a long distance from the current predictions. This will give us a better prediction of the utilization of the forecast. While RNN can’t keep any memory of the nodes in its processor, we work out a solution using the LSTM. This will overcome the vanishing gradient situation and will have what is known to be the LSTM-RNN network. This will make learning and forecasting better.
Let input be $X_t$, weight vector $W_t = [w_f \ w_i \ w_c \ w_o]^T$, bias $b_t = [b_f \ b_i \ b_c \ b_o]^T$, and output $O_t$ at time $t$. The forget gate is used to throw the information from the memory which is not necessary for forecasting and this task process by a sigmoid layer.

$$f_t = \sigma(w_i^T[O_{t-1}, X_t] + b_i)$$

(7)

Now, the second important task is to consider the information which is necessary for the forecasting, and this responsibility is fulfilled by the input gate layer.

$$i_t = \sigma(w_i^T[O_{t-1}, X_t] + b_i)$$

(8)

$$C_t = \text{tanh}(w_i^T[O_{t-1}, X_t] + b_i)$$

(9)

Now, there is a need to do updates in cell states ($C_{t-1}$ to $C_t$) forgetting the things not wanted to remain in our observations. After this, to scale values of the state, we add $l_t = C_t^{-}$.

$$C_t = f_t^*C_{t-1} + i_t^*C_t$$

(10)

We decide our output values by using sigmoid and tanh function (that gives value between -1 and 1) of cell state as follows:

$$h_t = \sigma(w_o^T[O_{t-1}, X_t] + b_o)$$

(11)

$$O_t = h_t^*\text{tanh}(C_t)$$

(12)

where, $O_t$, is our final output from the LSTM network at time $t$.

C. Model Evaluation Criteria

Various metrics employed to evaluate the performance of the proposed models; MSE, RMSE, Normalized Mean Square Error (NMSE), Normalized Root Mean Square Error (NRMSE), and Standard Deviation (SD). The calculation methods for each are the following:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\text{EC}_{\text{Predicted}} - \text{EC}_{\text{Actual}})^2$$

(13)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{EC}_{\text{Predicted}} - \text{EC}_{\text{Actual}})^2}$$

(14)

$$\text{NMSE} = \frac{1}{n} \sum_{i=1}^{n} (\text{EC}_{\text{Predicted}} - \text{EC}_{\text{Actual}})^2 / [\max(\text{EC}_{\text{Actual}}) - \min(\text{EC}_{\text{Actual}})]$$

(15)

$$\text{NRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{EC}_{\text{Predicted}} - \text{EC}_{\text{Actual}})^2} / [\max(\text{EC}_{\text{Actual}}) - \min(\text{EC}_{\text{Actual}})]$$

(16)

where $n$ is the number of the observer, EC stands for Energy Consumption.

IV. IMPLEMENTATION

A. ARIMA Implementation

As we have discussed in the previous section, applying ARIMA we need to consider providing stationarity-tests. We have tested the ADF Test on the dataset. The test results from sample data are (-3.431041, -2.861845, -2.566932). These results are clearly higher than the provided tabulated Statistical values (-9.514361). This indicate that $H_0$ is rejected and $H_1$ is accepted. Furthermore, $p$-value is less compared to %5 indicated time interval. This indicate that the acceptance rate is higher as an alternative hypothesis ($H_1$) showed.

As a result, a stable model has been satisfied. Thus, plotting ACF vs. PACF graph respectively.

B. LSTM Implementation

In this experiment, we have examined LSTM. It is noted that the input layer size 12 along with the hidden layer of the same size is optimal to forecast the EC. Furthermore, EC dipped down by a 20 percent rate for each hidden layer that was introduced for the same size of 12 for a single node output layer. This has given an optimal prediction and forecast of EC. In addition, the overfitting problem has had an acceptable rate of difference in ANN structure. Finally, we have fitted LSTM with a 30 iterations and batch size of 12 in order to accommodate an optimal weight for each edge.

V. RESULTS AND DISCUSSIONS

We have made a correlation to accommodate both models in order to achieve an optimal forecast value for the models of ARIMA along with the LSTM models. Based on the
evaluation criteria, we have seen significant improvement in the LSTM model. We have recorded NRMSE of 0.22 MW, 0.03 MW in ARIMA and LSTM, respectively. Fig. 3 represents the forecast values of EC from both the models.

![Energy Forecasting](image)

**Fig. 3. Forecast EC.**

| TABLE I: EVALUATION CRITERIA |
|-------------------------------|
| Model | MSE | RMSE | NMSE | NRMSE |
| ARIMA | 61838.01 | 248.67 | 56.26 | 0.22 |
| LSTM | 1448.44 | 38.05 | 1.31 | 0.03 |

**VI. CONCLUSION**

This work has shown that it is evident that, the EVs are growing rapidly the energy demands. In addition, it is increasing yet manageable for the demand and supply of energy appropriately, we need to know an estimated value of EC. Therefore, in this study, we have employed machine learning based model of ARIMA, including the approach of deep learning-based model LSTM. In summary, we list the results obtained that for the EC forecast we noticed that both ARIMA and LSTM showed optimal solutions. Even though LSTM is accurate compared to ARIMA but still LSTM reduce the number of false over demands. Furthermore, LSTM has a robust and optimal solution once activated alongside the ARIMA model. Further analysis will focus on improving the time complexity of the programable code that is based on the outcomes of the deep learning methodologies.

**CONFLICT OF INTEREST**

Authors declare that they do not have any conflict of interest.

**REFERENCES**

[1] International Energy Agency. Energy and Air Pollution. *World Energy Outlook – Spec. Rep.*, p. 266, 2016.

[2] Mouli G. R. C., Venugopal P. and Bauer P. Future of electric vehicle charging. 2017 *International Symposium on Power Electronics (EE)*, 2017, pp. 1-7, doi: 10.1109/PPEE.2017.8171657.

[3] Saqib M., Hussain M., Alam M. S., Beg M. M. S., and Sawant A. Public Opinion on Viability of xEVs in India. in *ISGW 2017: Compendium of Technical Papers*, 2018.

[4] Saqib, M., Hussain, M.M., Alam, M.S. et al. Smart electric vehicle charging through cloud monitoring and management. *Technol Econ (Smart Grids Sustain Energy)*, 2017:17(2):218. https://doi.org/10.1007/s40866-017-0035-4.

[5] Kim T. Y. and Cho S. B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, 2019:182:72–81.

[6] Xu Y., Ahokangas P., and Louis J. Electricity Market Empowered by Artificial Intelligence: A Platform Approach. 2019.

[7] Seyedzadeh, S., Rahmanian, F., Glesk, I. et al. Machine learning for estimation of building energy consumption and performance: a review. *Vis. in Eng*. 2018:6(5). https://doi.org/10.1186/s40327-018-0064-7

[8] Strategic I., Conference M., Demirel O. F., and Zaim S. Forecasting Electricity Consumption with Neural Networks and Support Vector Regression. *8th Int. Strateg. Manag. Conf.*, 2012:58:1576–1585.

[9] Puspita V. and Ermatta. Time Series Forecasting for Electricity Consumption using Kernel Principal Component Analysis ( kPCA ) and Support Vector Machine Time Series Forecasting for Electricity Consumption using Kernel Principal Component Analysis ( kPCA ) and Support Vector Mac. *IOP Conf. Ser. J. Phys. Conf. Ser.*, 2019.

[10] Berrel R. F., Lopes A. T., Rodrigues A., Varejas F. M., and Oliveira-Santos T. Monthly energy consumption forecast: A deep learning approach. *Proc. Int. Jt. Conf. Neural Networks*, 2017:2017–May:4283–4290.

[11] Muzaffar S., Aftshari A., and Muzaffar S. ScienceDirect ScienceDirect ScienceDirect Direct ScienceDirect Short-Term Load Forecasts Using LSTM Networks on District and Cooling Short-Term Load Forecasts LSTM feasibility of using the heat demand-outdoor Masdar for. *Energy Procedia*, 2019:158:2922–2927.

[12] Suganthi L. and Samuel A. A. Energy models for demand forecasting — A review. *Renew. Sustain. Energy Rev.*, 2012;16(2):1223–1240.

S. M. Shariff Associate Professor at Taibah University Electrical Engineering Dept. Has joined Taibah University in 2005 as an assistant professor. Before that he worked at Saudi Aramco for three years at the PMT department. Before that he worked as projects manager at Saudi Binladin Group, Makkah Harm Projects. Dr. Shariff has received his PhD EE from Wichita State University, Wichita, KS in 2000 and both his MS EE and BS EE from Washington University-St. Louis, MO in 1996 and 1994 respectively.