Remaining Useful Life (RUL) Estimation of Lead Acid Battery using Bayesian Approach

Bhalaji N.,
Department of Information Technology,
SSN College of Engineering, Chennai, India.
bhalajin@ssn.edu.in

Rimi Chowdhury
Department of Information Technology,
SSN College of Engineering, Chennai, India.

Abstract. This article deals with Remaining Useful Life (RUL) estimation of Lead Acid Battery using a probabilistic approach which is Bayesian inference of Linear Regression. RUL estimation of lead acid battery plays a very crucial role as it can prevent the catastrophic failure for the system in which it is used to serve as a power supply mainly in automobiles. Although there are various methods for age estimation of lead acid battery, machine learning algorithms always played a major role in the same. In this paper we have implemented one such algorithm for the RUL estimation. Bayesian approach is a probabilistic method which can be used for predicting the RUL of the battery. Firstly, we present a framework for feature extraction and then the RUL estimation model is trained on Bayesian inference of Linear Regression. The proposed approach is then applied to the collected dataset from five differently aged batteries which have undergone some charging/discharging and load cycle test. The experiment result shows that the proposed approach can improve the accuracy of RUL estimation than the regular methods.

1. Introduction

Lead Acid Battery is one of the most popularly used and economical storage device for electrical energy which is widely used in most of the car to power electrical circuits within. Typically for a fully charged lead acid battery the voltage is around 12.6 volts. Over the period of usage which undergoes several charging/discharging, the electrodes will deteriorate below the threshold where a useful current is no longer produced. Usually battery replacement takes place here. Although an average battery life is within five years, but it is highly dependent on the type of usage and also on the operating environment. It has been found that even in a case of well-maintained battery the life time didn't reach to the expected threshold. Hence accurate estimation of RUL based on

the current profile of a battery's parameter becomes extremely important to consider, which can predict the battery replacement time. The above experiment can be done in two ways (1) By understanding the physical circuit model (2) By data driven approach [1]. This project aims to analyze the RUL estimation by data driven approach with machine learning algorithm.

We have started our work by collecting data as there is no such open data-source available for lead acid battery, following which we have analyzed the battery internal characteristics and finally have

Electrical
Engineering and Automation

implemented Bayesian inference of Linear Regression and compare the same with some standard machine learning algorithm.

2. Lead Acid Battery and RUL

The battery has been an essential element of the automobile electrical system for a long time and the principal lead-acid innovation has not changed. It comprises of basically two electrodes submerged in sulfuric acid electrolyte in six cells. A chemical reaction between the plates produces a voltage of approximately 2.1volts per cell, so a total of 12.6 volts. At 12.4 volts it's considered only 50% charged, and at this point the battery wouldn't be able to power systems that require a lot of current, such as the starter motor of the headlights. At 12.0 volts the battery is only 25% charged, and at 10.5% the unit is considered fully discharged. By evaluating element condition and information from past operations the remaining useful life (RUL) of the element are often estimated. The RUL is the length from the actual time to the end of the useful lifetime of an element or system. The RUL may be a stochastic variable and it depends on the present age of the component, the surroundings and therefore the ascertained data of the health of the element. When unceasingly estimating RUL of the element, unsteady usage is taken in account, which makes the estimate a lot of correct.

If X_t is defined as the random variable of the remaining useful life at time t and Y_t as the operating history up to t, the probability density function (PDF) of X_t conditional on Y_t will be denoted as f($x_t|Y_t$). The estimation of RUL can be formulated as estimating f($x_t|Y_t$) or E($X_t|Y_t$). If there is no information about Y_t , then the estimation of f($x_t|Y_t$) becomes:

$$f(x_t|Y_t) = f(x_t) = \frac{f(t + x_t)}{R(t)}$$

Where f(t + x_t) is the value of the PDF at time t + x_t and $R(t)$ is the reliability function at t. If Y_t is available, this will provide more information to make the estimation of RUL more accurate [2]

3. Estimation of RUL with SOC

State Of Charge (SOC) is defined as how much charge is left in the battery which is used to power other electrical circuits. In a new Lead acid battery the open circuit voltage (OCV) is higher which implies higher SOC compare to the aged one. Below table shows a relationship between these two.

Relationship can be built between SOC and RUL to achieve our objective. According to the authors of [2][6][7] generally, RUL is estimated using sensors that give health related information of the component, meaning, there is a possibility to track the state of health related parameters during the

lifetime of the component. Examples of model-based prognostics are given in [3]. Authors in [8] developed a detailed physics-based model of a pneumatic valve in a cryogenic refueling system and predict the RUL of the component based on a discrete sequence of observations and a particle filter as a predictive technique.

Authors of [5] proposed a Deep Learning Framework for another battery type which is a Lithium-ion battery for RUL prediction. In his proposed algorithm he explained the method of feature fusion which is used after feature extraction process. The outcome of the previous process was feed to the final estimation of RUL using Deep Learning. Data-driven models use machine learning methods to either estimate RUL, the health of the component, or prognostic related information. These data-driven methods can be categorized into parametric and non-parametric methods [2]. A parametric approach assumes a parametric form of the distribution of the RUL and the parameters of interest are estimated based on observations which are in turn dependent on the battery SOC. Bahn, H et al [14] presented the "Efficient Management of Probe-Based Nano Storage Devices" Raj, Jennifer S et al [15] put forth an "Power Efficient Topology Control in Wireless Ad Hoc Networks." Smys, S et al [16] devised the "Virtual Reality Simulation as Therapy for Posttraumatic Stress Disorder (PTSD)" Manikandan, R et al [17] conducted a "Controller based performance measures of speed control of electrical motor for industrial applications." Kumar, N. M et al [18] performed the "Energy and Power Efficient System on Chip with Nanosheet FET."

4. Proposed System

For the implementation the project proposes to construct a probabilistic framework using naïve Bayesian theorem along with Linear Regression. In cases where we have a limited dataset, Bayesian model is a great choice for showing our uncertainty in the model. Conventional methods which follow modeling the physical circuits have not proved to be efficient in terms of time and accuracy. Whereas data driven methods with probability framework proved to be better in terms of understanding the participating features.

5. Implementation Work-flow

Our work has been divided into four main sections. We started our work by collecting the data from five differently aged 12V Lead Acid batteries. Load Testing was carried out for all these five batteries. Observations are then recorded and have been used for the dataset of this project. After the data collection, exploratory data analysis (EDA) has been carried out on the collected data. In statistics, exploratory data analysis (EDA) is an approach analyzing data sets to summarize their main characteristics, often with visual methods. After EDA we have implemented our proposed algorithm and compared the result with other standard machine learning algorithms. At the final stage of this workflow we have used our model for making the prediction on the RUL of battery.

5.1. Building the model

A general form of linear regression follows the below model:

$$y = \alpha^{T}X + \in$$

Where the outcome, y, is generated from the model parameters, α times the input matrix, X which is then added to the error due to random sampling noise. In the general linear regression, the model parameters α are calculated by finding the parameters which minimize the sum of squared errors, which are generally MAE or RMSE on the training set of data. In contrast, Bayesian Inference of Linear Regression assumes the responses are sampled from a probability distribution such as the Gaussian distribution:

$$y \sim N(\alpha^{T}X, \sigma^{2})$$

The mean of the Gaussian is the product of the parameters, α and the inputs, X and the standard deviation is σ. In Bayesian Models, both the response and the model parameters are assumed to be sampled from the chosen distribution.

The posterior probability distributions for the model parameters given the inputs, X, and outputs, y is determined by the below formula:

$$P(\alpha|y, X) = \frac{P(y|\alpha, X) * P(\alpha|X)}{P(y|X)}$$

In this project we have used Markov Chain Monte Carlo (MCMC) to draw samples from the posterior in order to approximate the posterior, as calculating the exact posterior distribution is computationally intractable for continuous values. As we draw more samples the posterior probability approximation will converge on the actual posterior distribution for the model parameters. The final result of the proposed inference is not a single estimation but an entire distribution that we can use to make interpretation about new observations. This circulation enables us to show our vulnerability in the model and is one of the advantages of Bayesian Modeling techniques. As the quantity of information focuses expands, the vulnerability should diminish, appearing more elevated amount of conviction in our assessments.

Electrical
Engineering and Automation

5.2. Evaluation of results using RMSE and MAE

The result of the prediction is evaluated using two standard metrics which are Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

```
  Function evaluate_predictions(predictions, true):
 Pass in: actual and predicted value
    mae = np.mean(abs(predictions - true))
    rmse = np.sqrt(np.mean((predictions - true) ** 2))
    return mae, rmse
 EndFunction
```

5.3. Establishing the baseline

```
median_pred = X_train['RUL'].median()
median_preds = [median_pred for _ in range(len(X_test))]
true = X_test['RUL']
mb_mae, mb_rmse = evaluate_predictions(median_preds, true)
print('Median Baseline  MAE: {:.4f}'.format(mb_mae))
print('Median Baseline RMSE: {:.4f}'.format(mb_rmse))
```

5.4. Implementation of Bayesian Inference of Linear Regression

```
with pm.Model() as normal_model:
    family = pm.glm.families.Normal()
    pm.GLM.from_formula(formula, data = X_train, family = family)
    normal_trace = pm.sample(draws=2000, chains = 2, tune = 500, njobs=-1)
```

5.5 Evaluation of Bayesian Model Using Mean of Model Parameters

```
  Function evaluate_trace(trace, X_train, X_test, y_train, y_test, model_results):
  Pass in: sampling trace, training set, testing set, standard model results
    var_dict = {}
    for variable in trace.varnames:
       var_dict[variable] = trace[variable]
    var_weights = pd.DataFrame(var_dict)
    results = pd.DataFrame(index = X_test.index, columns = ['estimate'])
  # Metrics
    actual = np.array(y_test)
    errors = results['estimate'] - actual
    mae = np.mean(abs(errors))
    rmse = np.sqrt(np.mean(errors ** 2))
  EndFunction
```

5.6. Making predictions from model

```
Function test_model(trace, test_observation):
Pass in:  model trace, test observations
   var_dict = {}
  for variable in trace.varnames:
     var_dict[variable] = trace[variable]
  var_weights = pd.DataFrame(var_dict)
  sd_value = var_weights['sd'].mean()
  actual = test_observation['RUL']
  test_observation = test_observation.drop('RUL')
  var_means = var_weights.mean(axis=0)
  mean_loc = np.dot(var_means, test_observation)
  # Estimations of RUL
  estimates = np.random.normal(loc = mean_loc, scale = sd_value,
                  size = 1000)
 EndFunction
```

6 .Results and Discussion

The entire implementation has been carried out in Python3 in Jupyter platform. Implementation includes EDA, Data manipulation, Feature selection, comparison with standard machine learning algorithms, model building, and evaluation of model and making prediction using the proposed model. Predictions have been carried out for both known and unknown data.

6.1 Comparison result of standard machine learning algorithms

A good naive baseline for regression is to guess the median value of the target for every observation in the test data. In our problem, the median is 1.5. In this project we have compared our proposed model with some standard machine learning algorithms such as SVM, Gradient Boosted and other tree based methods. Fortunately, we have seen that all models best the baseline indicating that machine learning will work for this problem. On the comparison, the tree based algorithm and gradient boosted regression method performs the best although Linear Regression (the frequentist approach to linear modeling) also does better than the baseline. Fig 1 shows the comparison graphs.
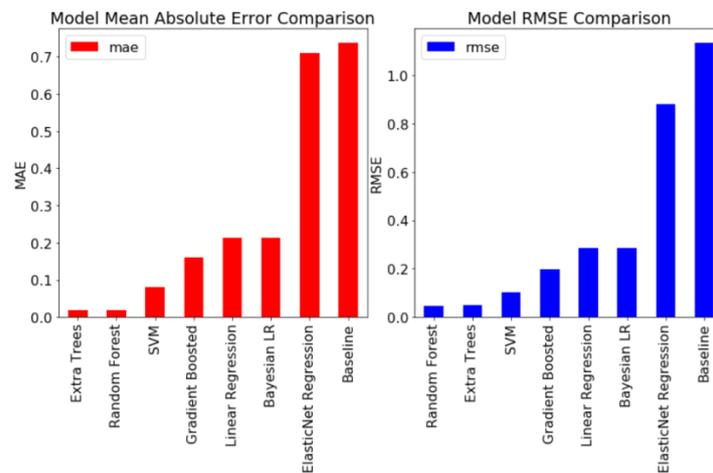
Fig 1 Comparison of RMSE for standard Machine learning algorithms

The main benefits of Bayesian Linear Regression Modeling are not in the accuracy, than the linear regression but in the interpretability and the quantification of the uncertainty in our data.

6.2 Making Predictions

Here we have taken the mean of each model parameter from the trace to serve as the best estimate of the parameter. If we want to make a prediction for a new data point, we can find a normal distribution of estimated outputs by multiplying the model parameters by our data point to find the mean and using the standard deviation from the model parameters. For a new data point, we substitute in the value of the variables and construct the probability density function for the RUL.

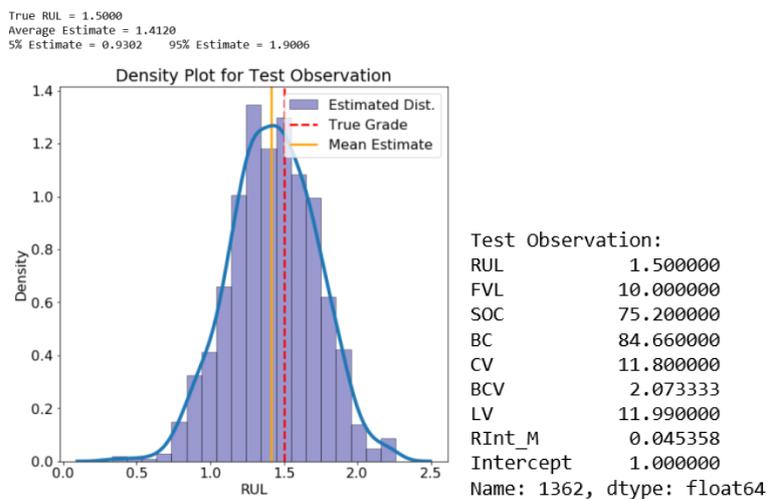### 6.2.1 Testing from test set

Fig 2 Density plot points for test observation for data points from the test set

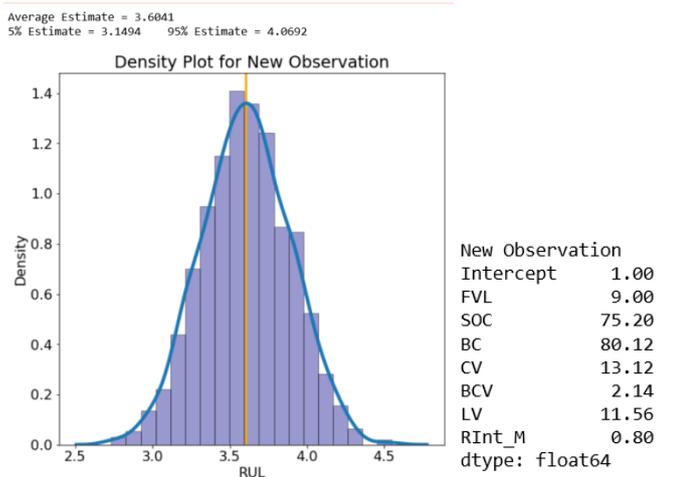### 6.2.2 Testing for a new data point outside test set

Fig 3 Density plot points for test observation for data points outside the test set

32

Fig 2 and Fig 3 shows that the Bayesian inference of Linear Regression is slightly better than the General form of Linear Regression as it can model the **uncertainty of the data.**

7. Conclusion

For storing electrical energy, Lead acid battery has been used for many decades and thus RUL prediction is of great importance to the state estimation and health management of these batteries. Although the behavior and characteristics of battery is highly stochastic, but a proper estimation technique which can capture the past performances and the present situation can always lead to a correct outcome. The developments of machine learning areas provide a new promising method for lead-acid battery RUL prediction. The main contributions can be summarized as follows: (1) exploratory data analysis (EDA) which has helped us to understand the trend of data; and (2) a Bayesian inference of Linear Regression using MCMC sampling method, is proposed for five lead acid battery RUL predictions. We applied the proposed approach to a dataset of 12V lead-acid battery collected from five different profiles, which have undergone charging/discharging and load testing cycle in order to get the battery parameters. The experimental results show the effectiveness and better accuracy of the proposed approach successfully captured the uncertainty of the data and compared superior than traditional regression model.

In future work we plan to improve the proposed approach by applying more advanced machine algorithms even neural network which can help us to get the more accurate result. The main purpose of this implementation is to observe if we can establish a relationship between the input and the output parameter with comparatively less complicated algorithm which takes lesser training time.

References

[1]   Saha, B., et al. 2009. Comparison of prognostic algorithms for estimating remaining useful life of batteries. Transactions of the Institute of Measurement and Control, 31(3-4):293-308.

[2]   Si, X., et al. 2011. Remaining useful life estimation - A review on the statistical data driven approaches. European Journal of Operational Research 213:1-14.

[3]   Saha, B., et al. 2007. An integrated approach to battery health monitoring using Bayesian regression and state estimation. Autotestcon, 2007 IEEE, 646-653.

[4]   Sergii Voronov. "Data Driven Lead Acid Battery Lifetime Prognostics." ICWSM 2016 (2016): 519-522.

[5]   Lei Ren, Li Zhao, Sheng Hong, Shiqiang Zhao, Hao Wang, "Remaining Useful Life Prediction for Lithium-ion Battery: A Deep Learning Approach." IEEE Transactions On Energy Conversion, Vol. 12, March 2018.

[6]   Xiyun Yang , Feifei Jiang, Xiaoning Wu. "Prediction of Lead-acid Storage Battery's Remaining Capacity Based on LM-BP Neural Network" 978-1-4673-5534-6/13/$31.00©2013 IEEE

[7]   Selina S.Y. Ng , Yinjiao Xing, Kwok L. Tsui. "A naive Bayes model for robust remaining useful life prediction of lithium-ion battery." S.S.Y. Ng et al. / Applied Energy 118 (2014) 114–123.

[8]   Claudio Burgos, Doris S_aez, Marcos E. Orchard, Roberto C_ardenas. " Fuzzy modelling for the state-of-charge estimation of lead-acid batteries". C. Burgos et al. / Journal of Power Sources 274 (2015) 355e366

[9]   Wang Y, Pan R, Yang D, et al. "Remaining Useful Life Prediction of Lithium-ion Battery Based on Discrete Wavelet Transform." Energy Procedia, vol.105,pp.2053-2058,2017.

[10]   Hu, Xiaosong, et al. "Technological Developments in Batteries: A Survey of Principal Roles, Types, and Management Needs." IEEE Power & Energy Magazine 15.5(2017):20-31.

[11]   Hu, Xiaosong, D. Cao, and E. Bo. "Condition Monitoring in Advanced Battery Management Systems: Moving Horizon Estimation Using a Reduced Electrochemical Model." IEEE/ASME Transactions on Mechatronics 23.1(2018):167-178.

[12]   Bhalaji N., Kanakeri A.V., Chaitanya K.P., Shanmugam A.: Trust based strategy to resist collaborative blackhole attack in MANET. Int. J. Inf. Process. Manag. 70, 465–474 (2010)

[13]   Hazrat Ali, Swedish Institute of Computer science, "A Performance Evaluation of RPL in Contiki", 2012.

[14]   Bahn, H. (2019). Efficient Management of Probe-Based Nano Storage Devices. Journal of Electronics, 1(01), 1-11.

[15]   Raj, Jennifer S., R. Harikumar, and Joby Philip. "Power Efficient Topology Control in Wireless Ad Hoc Networks." In International Conference on Advanced Computer Theory and Engineering, 4th (ICACTE 2011). ASME Press, 2011.

[16]   Smys, S., & Raj, J. S. (2019). Virtual Reality Simulation as Therapy for Posttraumatic Stress Disorder (PTSD). Journal of Electronics, 1(01), 24-34.

[17]   Manikandan, R., and S. Smys. "Controller based performance measures of speed control of electrical motor for industrial applications." In 2017 International Conference on Inventive Systems and Control (ICISC), pp. 1-5. IEEE, 2017.

[18]   Kumar, N. M. (2019). Energy and Power Efficient System on Chip with Nanosheet Fet. Journal of Electronics, 1(01), 52-59.