Comparison of diagnostic machine learning models with physics based models for predicting the lithium-ion battery degradation behavior

Akshata Kishore Moharir
Analytics and information management services
The Boeing Company
Bangalore, India
akshatakishore.moharir@boeing.com

Srishti Gautam
Boeing Research and technology
The Boeing Company
Bangalore,India
srishti.gautam@boeing.com

Dr. Seema Chopra
Boeing Research and Technology
The Boeing Company
Bangalore,India
seema.chopra@boeing.com

Naveen Kumar Megharaj
Analytics and information management services
The Boeing Company
Bangalore, India
naveen.k.megharaj@boeing.com

***Abstract* - A novel data driven approach is developed for predicting the battery degradation behavior for lithium-ion batteries using Regularized Greedy Forests (RGF). The aim of this novel approach is to improve and compare the prediction accuracy of the lithium ion (Li-ion) batteries using physics based models and surrogate models which are machine learning (ML) models. First, the approach takes in state of charge, state of health, and power of the battery as inputs and predicts the solid electrolyte interface of the Li-ion batteries which lead to the capacity fade which can be the additional cost of batteries. Second, the RGF is proposed as the new ML regressor for failure diagnostics to determine the battery degradation behavior. The prediction is based on the assumption of the availability of real-time observation and historical data. The feasibility of the regressors is validated using Li-ion battery diagnostic data. The experimental results show the following: (1) Fewer data dimensions for the input data are required compared to traditional state-of-art ML models (2) Physics based models describe the battery behavior accurately but they are computationally costly, therefore we have developed surrogate ML models for physic based models which are cost effective (3) The proposed ML regressor provides an effective way of diagnosing battery degradation behavior of the Li-ion battery compared to state-of art Surrogate ML algorithms with small error and has high prediction accuracy.**

Keywords: Battery, State of charge, State of health, Random Forest, Gradient Boosting Trees, Regularized Greedy Forest, Normalization, Extreme Gradient Boosting, Cross Validation.

I. INTRODUCTION

Complex engineering systems, such as aircrafts, industrial processes, power plants where lithium-ion batteries are utilized to perform specific functions in term of reliability, productivity, safety and availability. However, no matter how well a complex system is designed, the system will deteriorate. Thus, maintenance is introduced as an effective routine to sustain the reliability of the system. Maintenance has been traditionally employed as one of two maintenance philosophies: preventative or corrective maintenance. The common characteristic of preventative and corrective maintenance is that neither strategy takes the actual condition of the system into consideration before decided-upon maintenance activates. Therefore, preventive and corrective maintenance become a major expense in different industries; what's more, they lack the accuracy of failure rate at fixed intervals to avoid the catastrophic failures, are labor intensive, and minimize the system's availability. In fact, one third of the cost of maintenance is incurred unnecessarily due to bad planning, improper or misused preventive maintenance, and unavailable equipment that lead to decreased availability and increased maintenance cost in terms of labor and spare parts [3].Li-ion battery was chosen as an example for a complex system because the internal state variables cannot be accessed by sensors or are hard to measure under operational conditions. [1, 2, 4, 5, 6, 7, 8]. Li-ion batteries exhibit high energy densities, long life time, and environmental friendliness. For these reasons, Li-ion batteries contribute to the advancement of technology and are widely used in many applications: from portable electronics and hybrid electric vehicles, to space and aircraft power systems. Failure of a lithium ion battery could lead to irreversible conditions, reduced performance, operational impairment, and under extreme conditions cause catastrophic failure. In order to maximize output from a Li-ion battery, and avoid catastrophic conditions, it is essential that any fault occurring in the battery be quickly detected and accurately diagnosed, and predicting the state of the battery under all operation conditions is necessary in order to prevent fatal failures [5, 6, 7]. Recently, capacity fade of Li-ion batteries has seen a growing interest among industry and the academic community as a hotspot and challenging problem in the fields of reliability.

Physics-based models are most popularly used for identifying the battery behavior including the degradation. There are several kinds of physics based models available in literature [8] [19] but in this work we mainly focus on the P2D models.

Presented at WinTechCon-2018, organized by IEEE CAS Bangalore Chapter, IEEE Bangalore Section, and IEEE WiE Council

The pseudo-two-dimensional is the most popular kind of physics based model. P2D model is by far the most used by battery researchers, and solves for the electrolyte concentration, electrolyte potential, solid-state potential, and solid-state concentration within the porous electrodes and the electrolyte concentration and electrolyte potential within the separator [8]. This model is represented by coupled nonlinear partial differential equations (PDEs) which vary with respect to electrode thickness x, particle radius r, and time t [19].It can take several minutes to simulate and solve them as they are made up of complex formulas and equations [8] [9] [10] hence they are computationally costlier. Attempt to make P2D models easier to implement and faster to solve, surrogate machine learning models are created. Various Machine learning models have been implemented in predicting the battery capacity fade, available in literature [9,10,11,12,13,14,15,16,17,18,19]. our work focuses on extension of the efforts in past with application of most recent and popular ML algorithm such as RGF. We have used regularized greedy forest [21, 22], Random Forest [23], GBT [24], XGBoost [25], algorithms as surrogate models and their prediction accuracies as well as their execution times are compared for predicting the battery degradation behavior. Their ability to predict the dynamic behavior of the physics-based model are examined and the corresponding execution times are extremely encouraging for use in time-critical applications while still maintaining very high (~99%) accuracy. Although surrogate ML models based on the P2D physics model are demonstrated in this paper, the proposed approach in this paper is applicable for other physics based models [7][19].

The rest of the paper is divided into several sections. Section 1 presents a brief background of various efforts related to prediction of battery degradation behaviors and introduction of surrogate ML models and P2D models. Application domain is described in Section 2, which explains the nature of experiments conducted, parameter description. Section 4 starts by describing the overall approach taken and presents details of feature extraction, learning procedure, and prediction algorithms. Section 3 concludes with a brief discussion of underlying learning algorithms that are used in our prediction framework. Section 4 presents the results and discussions, followed by conclusions in Section 5.

A. Abbreviations and Acronyms

Abbreviation	***Definition***
GBT	Gradient Boosting Trees
ML	Machine Learning
MSE	Mean Squared Error
P2D	Pseudo Two Dimensional
RGF	Regularized Greedy Forest
PDE	Partial differential equations
RF	Random Forest
SoC	State of Charge
SoH	State of Health
SEI	Solid Electrolyte Interface
XGBoost	Extreme Gradient Boosting

Table 1 Abbreviations and Acronyms

II. EXPERIMENT DATASET

The proposed ML models are validated using Li-ion battery degradation data. The required Li-ion battery data set was generated by P2D model for Li-ion battery.Several battery performance parameters were measured (Table 1).The objective of these experiments were to be able to compare the results of battery degradation behavior from Surrogate ML models. Since battery degradation behavior cannot be measured directly, indirect estimates of capacity fade based on the measured performance parameters were used in prediction [1].The degradation behavior of the lithium-ion battery is significantly impacted by many factors such as SEI power, SoC, SoH and temperature due to the nature of the lithium-ion mechanism. The assumptions of this current model are: (a) The battery operates in 23 degree Celsius (b) battery fade only happens during charging (c) Initial SoC is 0.2 (d) Initial SoH is 1e-8 [26].

Table 2 Li-ion Battery Parameters used in the Prediction

Li-Ion Battery parameters	**Parameters used for prediction**	
	Parameters	***Definition***
	SoC	State of charge of battery
	SoH	State of health of battery(which is determined by the thickness of the SEI layer)
	Power	Positive sign represents charging; negative sign represents discharging
	SEI	Solid-Electrolyte Interface

III. DATA DRIVEN METHOD

In our work SEI layer growth is used for prediction and it is considered as a regression problem. In the rest of this section the theoretical background of Regularized Greedy Forest is presented along with the data preprocessing and model building approach.

(a) The capacity fade prediction in our case is done using the SEI variable, as battery capacity fade is directly impacted by the SEI layer growth which makes it a natural choice to track the variable that defines the end of life. (b) Various experiments were conducted by building Random Forest, Regularized Greedy Forests, GBT and XGBoost regressors with 10-fold cross validation by changing the value of k ranging from 3 to 10 to obtain the best models for prediction of failures. We have used python 3.6 as development environment. Various python libraries including sklearn, pandas, scipy, numpy, RGF, matplotlib and xgboost were used for data preprocessing, model building model, evaluation and Visualization.

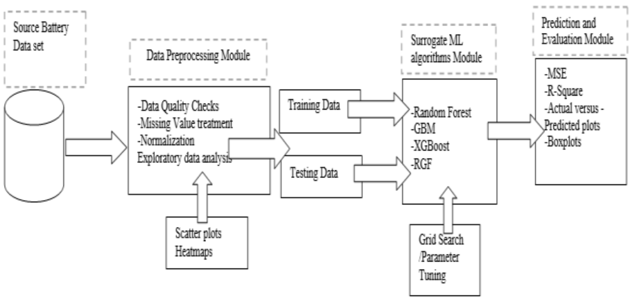

Figure 1 Approach for battery degradation behavior.

A. *Regularized Greedy Forests versus Random Forest, GBT, XGBoost*

In boosting algorithms, each classifier/regressor is trained on data taking into account the previous classifier or regressors success. After each training step, the weights are redistributed. Miss-classified data increases its weights to emphasize the most difficult cases. In this way, subsequent learners will focus on them during their training [21, 22]

However, the boosting methods simply treat the decision tree base learner as a black box and it does not take advantage of the tree structure itself. In a sense, boosting does a partial corrective step to the model at each iteration.

In contrast, RGF performs two steps: (a) RGF looks for the one step structural change to the current forest to obtain the new forest that minimizes the loss function. (b) Adjusts the leaf weights for the entire forest to minimize the loss function.

Search for the optimum structure change:

(a) For computational efficiency, only 2 types of operations are performed in the search strategy: to split an existing leaf node, and to start a new tree.(b) Search is done with the weights of all the existing leaf nodes fixed, by repeatedly evaluating the maximum loss reduction of all the possible structure changes. (c) It is expensive to search the entire forest (and that is often the case with practical applications). Hence, the search is limited to the most recently-created't' trees with the default choice of t =1.

Weight Optimization:

Weights for each of the nodes are also optimized in order to minimize the loss function.

Tree Size:

RGF does not require the tree size parameter (e.g., number of trees, max depth) needed in gradient boosted decision trees. With RGF, the size of each tree is automatically determined as a result of minimizing the regularized loss.

Model Size:

Since RGF performs fully corrective steps on the model/forest, it can train a simpler model as compared to boosting algorithms which require a small learning rate/shrinkage and large number of estimators to produce good results.

B. *Exploratory data analysis*

Data preprocessing is performed to check for the missing values and "NA"s in the data set, which are treated appropriately using advanced data preprocessing techniques like last observation carry forward. (d) Scatter plots are plotted to visually explore the effect of SEI layer growth on SoH of the battery.

The scatterplot show the effect of SEI layer growth on the state of health of battery. In scatterplot the SEI values are on y-axes and power values on the x-axes respectively.

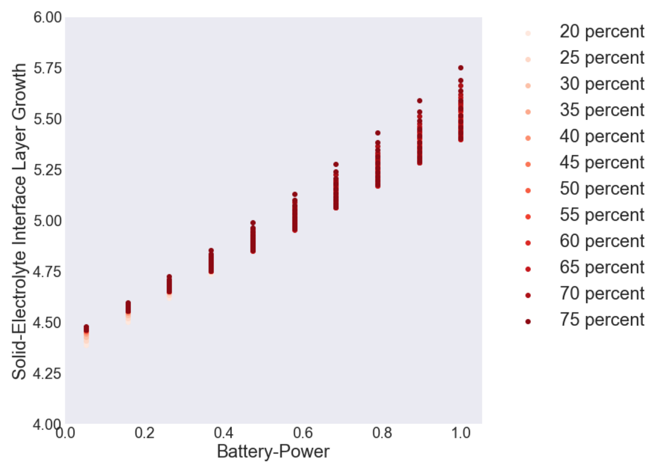

Figure 2 Power values are plotted against the SEI layer growth showing degradation with respect to state of health of battery where the colors indicate the correlation with state of charge.

C. Data Preprocessing and transformation

After thorough exploration of the SEI parameter with power and state of charge, the data set is normalized so that all the variable are in the same scale of unit. Also minimum and maximum values of the power are captured.

D. Model Building and Refinement

Various regression based surrogate models are developed and parameter tuning [20] is performed using the grid search and K-fold cross validation methods for all the nonlinear regressors for comparing their predictive capability. Actual versus predicted values are plotted for surrogate ML models. Boxplots are drawn to visually compare the top performing Surrogate models by plotting the models on x-axes and their R Squared values on the Y axis.

IV. RESULTS

Results of comparing ML based surrogate models for predicting the battery fade were noted down for different nonlinear regressors with respect to their MSE and R-square values on the test data set. Regularized greedy forest was a clear winner compared to rest of the state-of-art regressors.

The actual versus fitted plots for various surrogate ML models were plotted, RGF and XGBoost performed extremely well and random forests performed poorly.

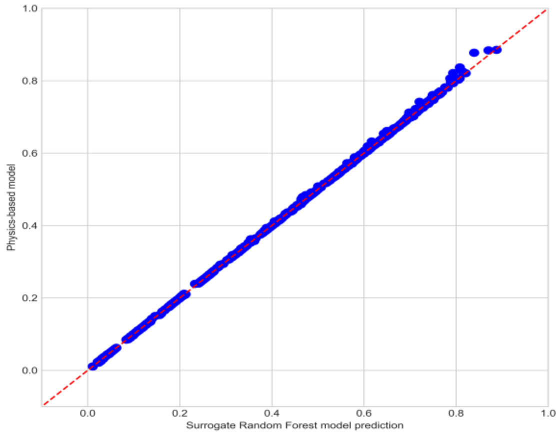

Figure 3 Actual versus predicted plot of SEI using the surrogate Random Forest model

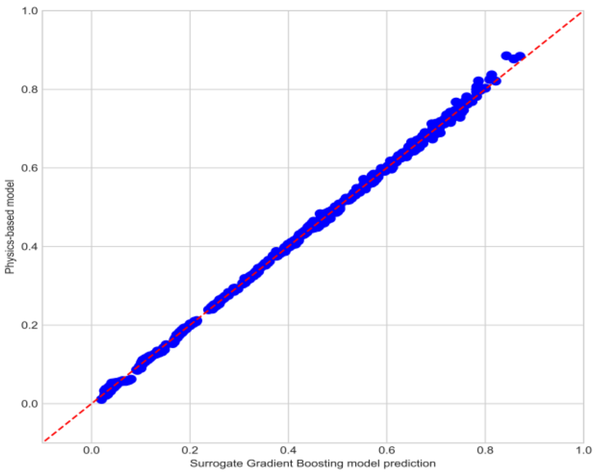

Figure 4 Actual versus predicted plot of SEI using the surrogate Gradient Boosting model

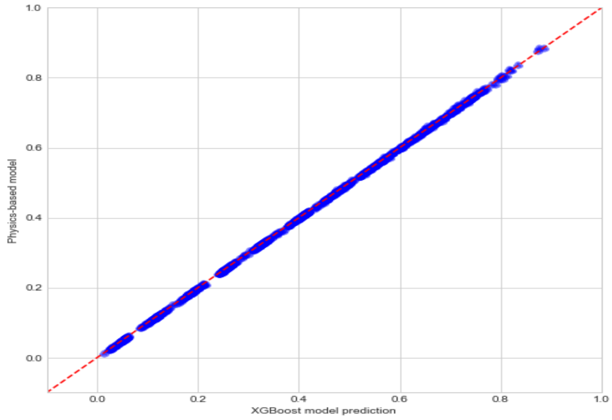

Figure 5 Actual versus predicted plot of SEI using the surrogate XGBoost model.

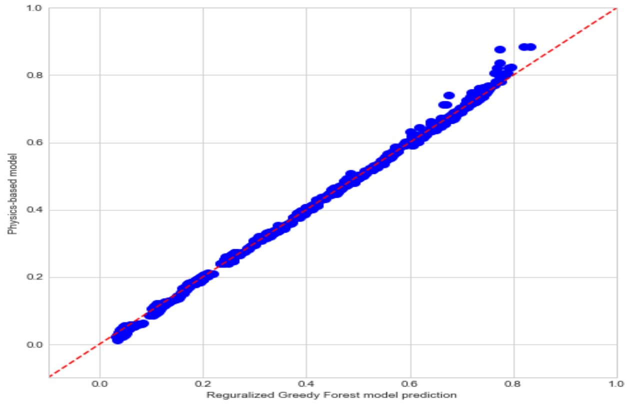

Figure 6 Actual versus predicted plot of SEI using the surrogate RGF

Boxplots were plotted to evaluate the performance of various regressors by plotting the model name on the x-axes and R-sqaure values on the y-axes and from the visual exploration, reguralized greedy forest performed extremely well compared to rest of the models .

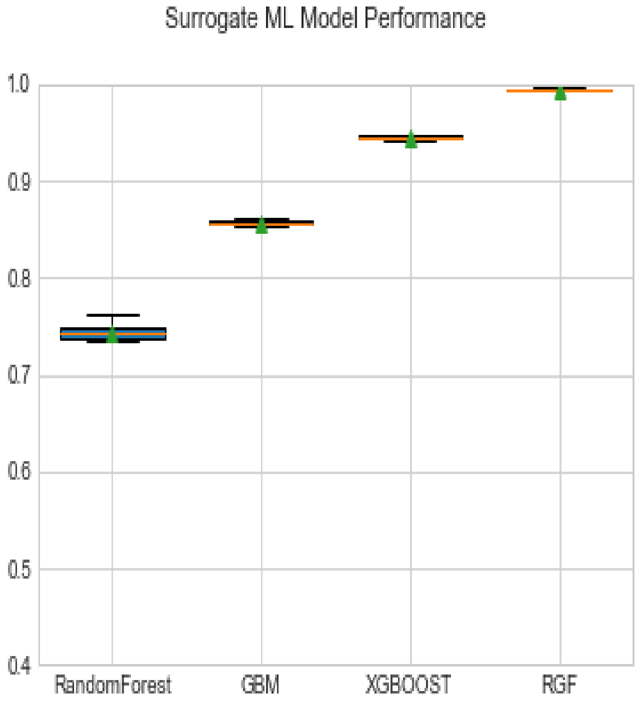

Figure 7 Boxplots showing the comparision of the Surrogate ML models plotted against their R-square values.

Various evaluation metrics are available for accessing the preformance of the regression models which are Root mean square error ,mean square error,R-Square,Mean absolute precentage error. In our analysis we have used MSE and R-Square which are calculated for each of the ML method and the method with minimum error is consided to a robust model.In our analysis RGF has the least error and RF has the maximum error which are shown in the Table 3 of the performance and evaluation.

Table 3. MSE and R-Square of Regressors

Surrogate ML Methods	***MSE***	***R-square***	***Execution time (seconds)***
RF	0.009	77.4	0.23
GBT	0.003	85.7	0.25
XGBoost	0.002	94.5	0.27
RGF	0.001	98.9	0.19

.

V. CONCLUSIONS AND FUTURE WORK

The most important conclusion of this work is application of ensemble based technique Regularized Greedy forest which is still in research and compare its accuracy and performance with state-of-art algorithms on Li-ion battery data set.

Second important conclusion is RGF outperformed in terms of predicting the battery degradation behavior with least minimum error as compared to rest of the regression algorithms. Current regression accuracy is great compared to other machine learning algorithms discussed in the paper, but we can further increase it as we get access to more data and as we replace generic algorithms with more specialized ones like regularized greedy forests which are still under research.

REFERENCES

[1] https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8186223

[2] https://www.researchgate.net/publication/228889501_Comparison_of_prognostic_algorithms_forestimating_remaining_useful_life_of_batteries

[3] G. CHAUHANet al., “Change Processes towards Flexible Lean Manufacturing: A Framework,” International Journal of Performability Engineering, Vol.6, No. 4, pp.363-372, RAMS Consultants, July 2010.

[4] Battery Based on Gaussian Process Regression,” 2012 Prognostics & System Health Management Conference, PHM Beijing 2012.

[5] Amardeep Sidhu et al, “Adaptive Nonlinear Model-Based Fault Diagnosis of Li-Ion Batteries,” IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, VOL. 62, NO.2, FEBRUARY, 2015.

[6] Adnan” Nuhic et al. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods,” Journal of Power Sources 239,pp 680 688,2013.

[7] Juan Carlos et al. “Support Vector Machines Used to Estimate the Battery State of Charge,” IEEE TRANSACTIONS ON POWER ELECTRONICS, VOL. 28, NO. 12, DECEMBER 2013.

[8] http://web.mit.edu/braatzgroup/Modeling_and_simulation_of_lithium_ion_batteries_from_a_systems_engineering_perspective.pdf

[9] https://depts.washington.edu/maple/pubs/66_DataScience_Neal_2018.pdf

[10] https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1164&context=eng_etds

[11] https://depts.washington.edu/maple/pubs/53_JES_SEI_layer_growth_and_capacityfade.pdf

[12] https://www.sciencedirect.com/science/article/pii/S0378775318302350

[13] https://arxiv.org/ftp/arxiv/papers/1210/1210.3672.pdf

[14] https://www.researchgate.net/publication/315745973_Comparison_of_LiIon_Battery_Degradation_Models_for_System_Design_and_Control_Algorithm_Development

[15] https://www.nrel.gov/docs/fy17osti/67102.pdf

[16] https://www.sciencedirect.com/science/article/pii/S0378775313019411

[17] https://www.sciencedirect.com/science/article/pii/S0378775305005082

[18] http://www.ijesrt.com/issues%20pdf%20file/Archive-2017/July-2017/5.pdf

[19] http://jes.ecsdl.org/content/161/14/A2099.full

[20] Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182.

[21] Johnson, Rie and Tong Zhang. 2014. "Learning Nonlinear Functions Using Regularized Greedy Forest." IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, No. 5

[22] https://www.analyticsvidhya.com/blog/2018/02/introductory-guide-regularized-greedy- forests-rgf-python/

[23] https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf.

[24] https://statweb.stanford.edu/~jhf/ftp/trebst.pdf

[25] http://learningsys.org/papers/LearningSys_2015_paper_32.pdf

[26] https://github.com/jerrychens/LiBattDeg