LITHIUM-ION BATTERIES SOH ESTIMATION, BASED ON SUPPORT-VECTOR REGRESSION AND A FEATURE-BASED APPROACH

Iacopo Marri [a], *Emil Petkovski* [a],*Loredana Cristaldi*, Marco Faifer*

[a] Politecnico di Milano, DEIB, Milan, Italy, iacopomarri@gmail.com, emil.petkovski@polimi.it, loredana.cristaldi@polimi.it, marco.faifer@polimi.it,
* Corresponding author E-mail address: loredana.cristaldi@polimi.it

Abstract − Lithium-Ion batteries, have become enormously used in many systems and applications, and are the most widespread energy storage system. Optimizing the usage of batterie is therefore very important to increase the safety of systems like electric vehicles or portable devices, to reduce economic loss in industrial environments, and to increase their availability. An accurate State of Health (SoH) estimation is important since it allows us to know battery conditions and make an appropriate use of it, and it improves the accuracy of other diagnostic measures, like State of Charge (SoC). In this paper, an approach for SoH estimation is proposed, based on Support Vector Regression machine learning algorithm and a smart feature extraction process, finding a good trade-off between applicability, light computation effort, and accuracy of results. Features selection and parameters tuning are discussed, and performances are measured on a dataset from the Prognostics Center of Excellence at NASA, considering 3 batteries of the dataset.

Keywords: SoH, machine learning, lithium-ion batteries, degradation diagnostic.

1. INTRODUCTION

One of humanity's primary concerns today is global warming due to greenhouse gas emissions from the combustion of fossil fuels. Along with the depletion of non-renewable resources, the use of renewable energy has become the government's highest priority. In general, in the generation of renewable energy through various means of production such as PVs, and wind, that are not available continuously during the day and night, it is essential to have battery systems and energy storage systems.

In this scenario, electric vehicles (EVs), which are powered by rechargeable batteries, also provide excellent performance in terms of efficiency and reduction of CO2 emissions and are widely used in the automotive industry [1].

There are many energy storage technologies, ranging from lead acid to NiMH to lithium-ion batteries, which are being employed. Lithium-ion (Li-ion) batteries are widely used in energy storage systems because of their high power, energy density and acceptable service life. These advantages lead to more focus and investment on this technology to increase its robustness and stability.

Because of their high power and energy density, lithium-ion batteries are also an excellent choice for electric cars in the transportation system business because of their minimal weight.

In order to protect battery systems and provide a reliable control and supply system, a battery management system (BMS) that protects the battery from being over charged or over-discharged and ensures cell balancing has to be implemented. When BMS is used, the major purpose is to keep the batteries in a safe, dependable, and efficient state while also preventing the battery pack from being used for a shorter amount of time than expected. To achieve this the BMS must accurately predict the different characteristics of the battery [2]. State of charge (SoC) and state of health (SoH) are two of the most important characteristics to consider. SoC and SoH can be derived by measuring related measurable factors such as battery voltage and current and battery temperature.

SoC is related to accessible capacity of the batteries. By knowing this factor, BMS retrains from overcharging or discharging of batteries. SoC can be estimated by using a battery model system and algorithm, or statistical approaches: in [3], SoC is estimated by means of the Kalman filtering method.

SoH provides information about aging status of the battery pack. Due characteristics and working condition of the battery pack, internal resistance grows gradually over time, as well as the capacity decreases.

In literature, many algorithms and model have been studied and used to perform SoH estimation: in [4], SoH is estimated by using a model-based approach to model the aging behaviour. In [5] and [6], artificial neural networks have been successfully applied to estimate the SoH of li-ion batteries, while in [7], random forest algorithm is used for the same task.

2. NASA DATASET

NASA Ames Prognostics Center of Excellence (PCoE) published a data repository composed of 6 datasets of aged Li-Ion batteries [8]. According to their guidelines, only the first of these datasets is suitable for prognostic degradation prediction.

The tests have been carried out in a climatic chamber to keep the environment temperature under control. The dataset is composed of 4 batteries run to failure, which is fixed at 70% of the remaining capacity (from 2 Ah to 1.4 Ah). The charging phase is divided into a constant current mode, with 1.5 A until the voltage gets to 4.2 V, and after that with a constant voltage phase, until charging current lowers to 20 mA. Discharge phase is carried out at 2.7 V, 2.5 V, and 2.2 V depending on the battery.

Cycles are divided into charging, discharging or impedance cycle. For each cycle, the sequent values are measured: time, voltage, current, temperature, voltage charge, current charge, capacity (only discharge).

3. SUPPORT VECTOR REGRESSION

Support Vector Machine (SVM) is one of the most classic and used ML supervised learning models, which performs binary classification, and it's been widely used for prognostic and diagnostic tasks: in [9], least-square SVM is used along with a partial discharge curve approach to estimate the SoH of li-ion batteries. In [10], SVM is applied on a dataset generated by simulations of electric vehicles (EV) battery usage profile, to estimate the SoH.

Given a set of samples, SVM finds the separating hyperplane with the biggest distance from each training point, so not only tries to classify the points but also to find the best and most robust possible hyperplane to do so [11].

In many cases, the points are not linearly separable in the input space, and a higher dimensional feature space is needed to find a linear hyperplane which can divide them. The Kernel Trick allows to move the data from the input space to a higher-dimensional space, keeping feasible the computational burden of the dot products between arrays of points.

Support Vector Regression (SVR) is a version of SVM adapted to perform regression tasks [12]. Compared to normal linear regression which finds the minimum of the squared error for the target values, SVR just fits the error of its predictions within a limit ϵ, while minimizing the loss function in (1), which is called the L2 loss:

$$
\begin{cases}
\min \dfrac{1}{2} \beta' \beta \\
|Y_n - (X'_n \beta + b)| \leq \epsilon \ \forall n
\end{cases} \tag{1}
$$

where:
β', β - weights arrays, normal and transposed
Y_n - target values
X'_n - transposed descriptor array
b - bias
ϵ - maximum allowed error

The ϵ constraint is then relaxed introducing the slack variables and applying what is called the soft margin approach.

$$
\begin{cases}
\min \dfrac{1}{2} \beta' \beta + C \sum_{n=1}^{N} (\xi_n + \xi_n^*) \\
Y_n - (X'_n \beta + b) \leq \epsilon + \xi_n \ \forall n \\
(X'_n \beta + b) - Y_n \leq \epsilon + \xi_n^* \ \forall n
\end{cases} \tag{2}
$$

where:
ξ_n, ξ_n^* - slack variables for positive and negative error
C - weight associated to slack variables

If we write the optimization problem, the loss function and the constraints, in its dual Lagrange form, the problem is more easily solvable and the parameters β of the model, as well as the predicted value for new observations, become a function of the training samples, according to this formulation:

$$
\begin{cases}
\beta = \sum_{n=1}^{N} (\alpha_n - \alpha_n^*) X_n \\
f(x) = \sum_{n=1}^{N} (\alpha_n - \alpha_n^*)(X'_n X) + b
\end{cases} \tag{3}
$$

where α_n and α_n^* are the Lagrangian multipliers. The prediction is therefore expressed as a function of the data and not of the weights, in particular of those data with $\alpha \neq 0$, called support vectors.

4. FEATURE BASED APPROACH

In most applications the charging phase of the batteries is guaranteed to be carried out under constant current, since it is given by the device, while discharge is likely not constant. Charging phase is anyway not guaranteed to be fully completed, for this reason it was decided to use a small portion of the charging curve (PCC), to extract features.

Other studies have been done on degradation estimation using partial curves, for instance in [13] a portion of the charging curve of batteries is taken and the degradation is estimated by similarity, finding the 2 batteries in the training set with the same portion of the curve being the most similar.

In this work specifically, the extracted feature is the time required by the battery to recharge by some little voltage step, defined within a voltage interval:

$$
\begin{cases}
V_1 = V(t_1) \to (V_1 + V_s) = V(t_{s_1}) \\
feature_1 = t_{s_1} - t_1 \\
(V_1 + V_s) = V(t_{s_1}) \to (V_1 + 2V_s) = V(t_{s_2}) \\
feature_2 = t_{s_2} - t_{s_1} \\
. \\
. \\
. \\
(V_1 + (n-1)V_s) = V(t_{s_{n-1}}) \to (V_1 + nV_s) = V(t_{s_n}) \\
feature_n = t_{s_n} - t_{s_{n-1}}
\end{cases} \tag{4}
$$

.

In Fig. 1 it can be seen, as expected, that the time of recharge decreases while the battery ages, and this is because of the global capacity reduction due to the degradation. We

can actually see how the charging time is halved near the final cycles, with respect to the first.

It can also be seen that the first part of the charging phase is almost vertical, which means that voltage raises rapidly in time, and it is hard to appreciate the time differences between different cycles. The middle phase instead is the more suitable to extract time features since it is the most extended over time.
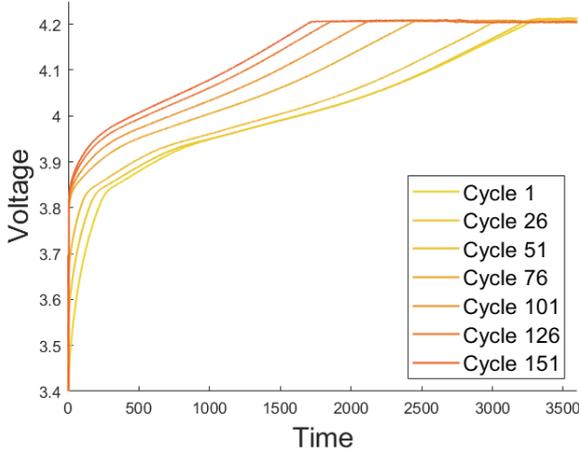

Fig. 1. Charging curves at different cycles, B0005

$$\begin{cases} R^2 = 1 - \dfrac{SS_r}{SS_t} \\ SS_r = \displaystyle\sum_i (y_i - f_i)^2 \\ SS_t = \displaystyle\sum_i (y_i - \bar{y})^2 \\ \bar{y} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i \end{cases} \qquad (5)$$

where:
SS_r - residual sum of squares
SS_t - total sum of squares
y_i - target value
f_i - estimated value
f_i - mean of the target values

Table 1 shows the results of the cross-validation procedure of the proposed models. Features sets not containing the 3.9 V – 4 V phase give the worst R^2 values, while the best validation R^2 is given by the 3.9 V – 4 V voltage window with a 0.05 V voltage step, which is a set of 2 features.

5. RESULTS

A first selection of the voltage window and the voltage step has been performed heuristically, by observing the plots of the features obtained with different set ups. In Fig. 2 the features are computed over the entire voltage range available, with voltage steps of 0.1 V. The charge time features computed for the lowest voltage values, from 3.6 V to 3.8 V, appeared to be almost flat curves, containing no variance thus no information about the data. One the other hand, the features computed from 3.8 V to 4.1 V have more variance hence are more descriptive of the data over the number of cycles. The badness of the 3.6 V – 3.8 V voltage windows has been assessed by a cross-validation procedure.

To find the optimal selection of parameters and features, 6 models have been developed, with different subset of features computed within the 3.9 V – 4.1 V voltage window.

Leave-one-out cross-validation (**LOOCV**) procedure has been applied on the 3 batteries of the dataset to find the best among the selected models.

The fitting accuracy of the model was assessed through the R^2 value: also called the coefficient of determination, it is a measure used in statistics, indicating how much a hypothesis describes the variance of the data. In other words, it is used to describe how well a model can fit observed data. R^2 is described as:
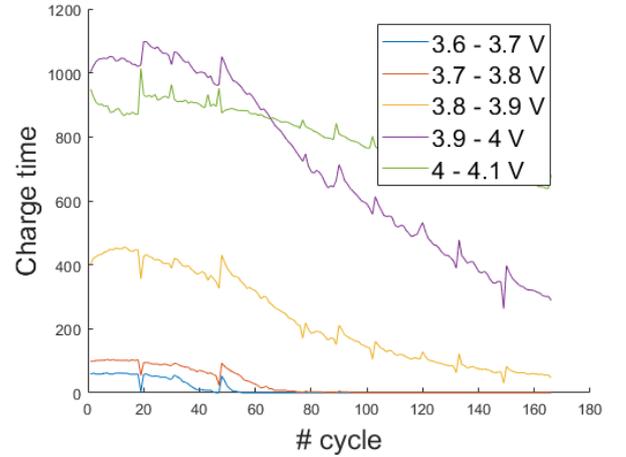

Fig. 2. PCC features computed over the 3.6 V - 4.1 V voltage window, with a voltage step of 0.1 V, and plotted over the number of cycle, for battery B0005.

Models containing more features like model 4, give slightly worse results since they are too complicated for the amount of data contained in the dataset. If bigger datasets were available, bigger models would likely give better results.

111

Table 1. Leave-one-out cross-validation results for feature optimization

	Features #	Features	Validation R^2
Model 1	1	Charge time of 0.1 V steps from 3.8 to 3.9 V	0.918
Model 2	1	0.1 V steps from 3.9 to 4 V	0.932
Model 3	2	0.05 V steps from 3.8 to 3.9 V	0.939
Model 4	2	0.05 V steps from 3.9 to 4 V	0.950
Model 5	4	0.05 V steps from 3.8 to 4 V	0.945
Model 6	4	0.025 V steps from 3.9 to 4 V	0.935

SVR hyperparameters have been initially tuned with the Matlab built-in function for SVR model, using Bayesian optimization algorithm, run for 500 iterations, to define a good starting range for the hyperparameters.

- **Box constraint:** factor C that weights the slack variables in (1). It helps controlling overfitting.
- **Kernel scale:** rescales the predictors. Each value in the predictors is divided by the kernel scale value.
- **Epsilon:** is the value ε which defines the radius of the epsilon tube where the algorithm tries to contain the points, or in other words, is the maximum error allowed.

Further tuning of the hyperparameters has been done during the validation procedure. Table 2 shows the tuned values of the hyperparameters. The kernel function has been chosen to be linear since the features are quite proportional to the target value to estimate, and different kernel functions led to lower validation accuracy.

Table 2. Tuned hyperparameters value for the considered SVR model

	Hyperparam.	Value
1	Box constraint	0.1989
2	Kernel scale	11.55
3	Epsilon	0.030
4	Kernel function	Linear

In Fig. 4 are shown the results of the validation of the final selected model. 3 equal models were created, each trained on 2 batteries, and tested on the remaining one. For instance, the estimation of battery B0005 has been done with a model trained on B0006 and B0007.
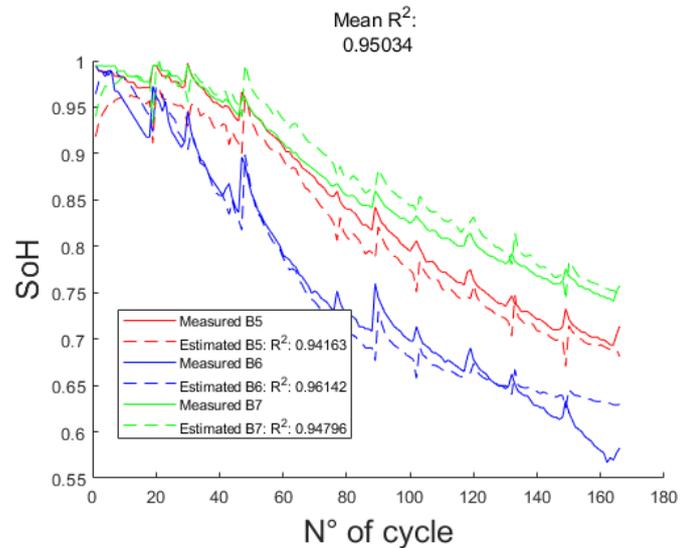

Fig. 3. Results of the proposed model. Each estimation has been made by training on the other 2 batteries.

The continuous line is the measured SoH for the battery, while the dashed line is the estimated SoH, over the number of cycles. The approach proves to be able to also model the peaks in the SoH function which are due to the rest time of the battery.

The model gives the best estimation accuracy on B0006, because it is the battery showing the most SoH degradation pattern over the time. After SoH of B7 reaches 0.7 though, the estimation diverges from the measurement, because the B5 and B7 which are used for training don't contain data about SoH lower than 0.7.

The overall R^2 value of 0.95 is a very satisfactory result, especially if we consider the scarcity of data available.

6. CONCLUSIONS

The SVR algorithm had a great importance in the quality of the results obtained, since it can work even with little data, and provides faster training and inference time with respect to other techniques.

The partial charging time proved to be a good source of information about the aging of the batteries, if measured over the proper partial interval of voltage, as long as the charging phase of the involved batteries is constant and equal for all.

To get more sturdy results and strongly validate this approach, a bigger dataset is required, containing batteries with charging characteristics similar to the NASA dataset. Also, partial discharge curves can be considered for features extractions, and the application of techniques which can exploit the time-series structure of batteries cycling, like recurrent neural networks.

REFERENCES

[1]   Tobias Werling, Peter Geuting, Patrick Höschele, Christian Ellersdorfer, Wolfgang Sinz, "Investigation of the electro-mechanical behavior of automotive high voltage busbars under combined electrical load with varying indenter geometry and

environmental conditions", Journal of Energy Storage, Volume 32, 2020, 101861, ISSN 2352-152X

[2] C. Campestrini, M. F. Horsche, I. Zilberman, T. Heil, T. Zimmermann, and A. Jossen, "Validation and benchmark methods for battery management system functionalities: State of charge estimation algorithms," Journal of Energy Storage, vol. 7, pp. 38–51, 2016, doi: 10.1016/j.est.2016.05.007.

[3] Z. He, M. Gao, C. Wang, L. Wang, and Y. Liu. "Adaptive state of charge estimation for li-ion batteries based on an unscented kalman filter with an enhanced battery model". Energies, 6:4134–4151, 08 2013. doi: 10.3390/en6084134.

[4] S. Barcellona, L. Cristaldi, M. Faifer, E. Petkovski, L. Piegari and S. Toscani, "State of Health Prediction of Lithium-ion Batteries," *2021 IEEE* International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4.0&IoT), 2021, pp. 12-17, doi: 10.1109/MetroInd4.0IoT51437.2021.9488542.

[5] Eddahech, Akram & Briat, Olivier & Bertrand, Nicolas & Deletage, Jean-Yves & Vinassa, Jean-Michel. (2012). "Behavior and State-of-Health Monitoring of Li-ion Batteries Using Impedance Spectroscopy and Recurrent Neural Networks". International Journal of Electrical Power & Energy Systems. 42. 487-494. 10.1016/j.ijepes.2012.04.050.

[6] M. Catelani, L. Ciani, R. Fantacci, G. Patrizi, and B. Picano. "Remaining useful life estimation for prognostics of lithium-ion batteries based on recurrent neural network". IEEE Transactions on Instrumentation and Measurement, 70:1–11, 2021. doi: 10. 1109/TIM.2021.3111009.

[7] Z. Chen, M. Sun, X. Shu, J. Shen, and R. Xiao. "On-board state of health estimation for lithium-ion batteries based on random forest". 2*018* IEEE International Conference on Industrial Technology (ICIT), 1754-1759.

[8] B. Saha and K. Goebel. Nasa ames prognostic data repository. NASA Ames, Moffet field, CA, USA, 2007. URL http://ti.arc.nasa.gov/project/prognosticdata-repository

[9] Z. Chen, X. Xia, M. Sun, J. Shen, and R. Xiao. "State of health estimation of lithium-ion batteries based on fixed size ls-svm". *2018 IEEE Vehicle Power and Propulsion Conference (VPPC)*, 2018, pp. 1-6, doi: 10.1109/VPPC.2018.8605041.

[10] V. Klass, M. Behm, G. Linbergh. "A support vector machine-based state-of-health estimation method for lithium-ion batteries under electric vehicle operation", Journal of Power Sources, vol. 270, pp. 262–272, 2014, doi.org/10.1016/j.jpowsour.2014.07.116.

[11] T. Evgeniou and M. Pontil. Support vector machines: Theory and applications. 01 2001.

[12] Matlab, Understanding support vector machine regression. https://it.mathworks.com/help/stats/understanding-support-vector-machine-regression.html

[13] X. Feng, C. Weng, X. He, X. Han, L. Lu, D. Ren, and M. Ouyang. "Online state-of-health estimation for li-ion battery using partial charging segment based on support vector machine". in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 8583-8592, Sept. 2019, doi: 10.1109/TVT.2019.2927120.