


Article

Clustered Based Prediction for Batteries in the Data Centers

Syed Naeem Haider , Qianchuan Zhao*  and Xueliang Li¹ haider.gillani12@gmail.com

* Correspondence: zhaoqc@tsinghua.edu.cn

† Center for Intelligent and Networked Systems (CFINS), Department of Automation and BNRist, Tsinghua University, China, Beijing 100084

Abstract: This paper proposes an ARIMA approach to battery health forecasting with accuracy improvement by K shape-based clustered predictors. The health prediction of the battery pack is an important function of a battery management system in data centers. Accurate forecasting of battery life turns out to be very difficult without failure data to train a good forecasting model in real life. The conventional ARIMA model is compared with total and clustered predictors for battery health forecasting. Results show that the forecasting accuracy of the ARIMA model significantly improved by utilizing the results of the clustered predictors for 40 batteries in a real data center. One year of actual historical data of 40 batteries of large scale datacenter is presented to validate the effectiveness of the proposed methodology.

Keywords: forecasting; clustering ; energy systems ; classification

1. Introduction

Battery uninterrupted power supply (UPS) is an integral part of the data centers, which ensures the stable performance of the data center during transitional fail-over mechanisms between power grids and diesel generators [1]. Steady power is required by the data centers, which is managed by the battery uninterrupted power supply (UPS). UPS is installed between the main power grid and servers [2]. Since the electricity bill of a data center constitutes a significant portion of its overall operational costs, data centers are now major consumers of electrical energy[3]. In 2013, data centers in the U.S. consumed 91 billion kilowatt-hours of electricity, and this is expected to continue to rise[4]. In 2017, nearly 8 million data centers required an astronomical 416.2 terawatt-hours of electricity[5,6]. The layout of the data centers design is illustrated in Figure 1.

The health assessment and remaining cycle life estimation of battery is a challenge when the battery system participates in data centers. Despite the increasing improvements in battery manufacturing and storage technology [7]. Not surprisingly, many traditional battery management methodologies have been studied to develop battery life prediction of the battery packs, such as voltage fault diagnosis, charge regimes, and state of health (SOH) estimation. Kristen et al [8] used the discharge voltage curve of 124 batteries to demonstrate a data-driven model to predict the battery life cycle before degradation. Tang et al. [9] predict the battery voltage, power with the model-based extreme learning machine for the electric vehicles. Xiaosong et al. [10] utilized the sample entropy of short voltage the sequence is used as an effective signature of capacity loss. Advanced sparse Bayesian predictive modeling (SBPM) methodology is employed to capture the underlying correspondence between capacity loss and sample entropy. You et al. [11] proposed a data-driven approach to trace battery SOH by using data such as current, voltage, temperature as well as historical distributions. Song

et al. [12] Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery by fusing IND-AR model and empirical model via the state-space model in RPF. Iterative updating is used to improve the prediction capability of ND-AR. Yape et al. [13] combined empirical mode decomposition (EMD) and autoregressive integrated moving average (ARIMA) model is for prediction of lithium-ion batteries remaining useful life (RUL) in battery management system (BMS) used in electric vehicles. Luping et al. [14] propose a hybrid approach combining Variational mode decomposition (VMD) denoising technique, auto-regressive integrated moving average (ARIMA), and GM(1,1) models for battery RUL prediction.

The autoregressive integrated moving average (ARIMA) model has been one of the most widely used models in time series forecasting [15]. RG Kavasseri et al. [16] examines the use of fractional-ARIMA or f-ARIMA models to model, and forecast wind speeds on the day-ahead (24 h) and two-day-ahead (48 h) horizons. A hybridization of artificial neural networks and ARIMA model is proposed by Mehdi et al [17] to overcome the mentioned limitation of ANNs and yield a more general and more accurate forecasting model than traditional hybrid ARIMA-ANNs models. The annual energy consumption in Iran is forecasted using 3 patterns of ARIMA–ANFIS model by Sasan et al [18].

ARIMA has found applications in forecasting social, economic, engineering, foreign exchange, and stock problems. It predicts future values of a time series using a linear combination of its past values and a series of errors [19,20]. Since batteries in the data center are always on charging mode thus deep discharge is a rare occurrence for batteries and their distinctive internal chemistry causes different behaviors like stationary or stochastic for each battery. Also, failure data is not available in real life which makes it a challenge to accurately predict the battery health before it fails. In this paper, we develop a cluster assisted ARIMA model to improve the accurate prediction of battery health. Our groundwork for the cluster consistency with battery data is demonstrated in [21]. Clustering is applied in many applications to improve the model forecasting accuracy. The proposed K shape-based cluster assisted forecasting results are compared with actual battery data and without clustered ARIMA forecasting [22,23].

The rest of the paper is organized as follows: Section 2 describes the features of the data center and data set used for the study. Section 3 describes data preprocessing and explain the methodology by introducing the algorithms for cluster consistency and clustered ARIMA forecasting. Section 4 shows the steps to implement the proposed clustered forecasting method. Section 5 demonstrates the battery cluster consistency detection results and cluster assisted ARIMA forecasting discusses the effectiveness of the method by comparing the results with actual data and without cluster assisted forecasting ARIMA model. Section 6 concludes this work.

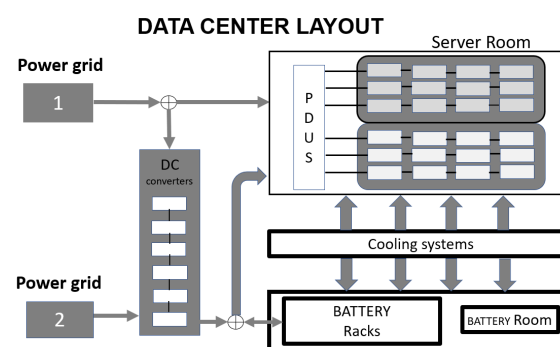


Figure 1. Data center Layout

2. Overview of the data set

In this paper, data is collected from a large-scale social media company located in China. One year data is used for research with 407266 data points and a sampling interval of 1 min. This data set includes the variables of datacenter main power, transmission units, battery units, cooling systems, and DC load values. Data set variables are shown in Table 1

Our objective is to develop an accurate and scalable method for the battery life degradation forecasting model. Voltage is utilized in the simplest of BMS of small vehicles to large scale data centers. Our data has 40 batteries voltages. Battery aging features are selected from domain knowledge of batteries [8].

Table 1. Data center data set with all feature instances 470226

Data Center features	Type	Attributes
Phase current/active/factor	Power	12
3-Phase active/power/factor	Power	6
HVDC module load/voltage	Transmission units	11
HVDC DC module current/volt	Transmission units	12
Battery group current/state	Battery units	4
Voltage/Resistance/Temperature	Battery units	120
PDU branch current	server units	24
AC supply/return temperature	Cooling system	24
Humidity	Cooling system	19
AC coil temperature	Cooling system	6
Up/Down front temperature	Cooling system	24
DC meter volt/current	DC unit	4

3. Methodology

Figure 2 shows the flowchart of the proposed method and the steps of the proposed method are given as follows.

Step 1 Data Preprocessing: First, separate the battery voltage data from the data set. Extract the historic values of first-month battery voltages and keep updating the real-time voltage values.

Step 2 Cluster consistency: Apply the K shape-based clustering to first month data and updated real time data separately. Compare first month and updated month clustering for cluster consistency. If a clusters are inconsistent, then go to step 3.

Step 3 Clustered arima forecasting: Chose a single battery from the cluster and cluster member as a predictor to fit an ARIMA model. If clustering outlines a single battery as a cluster fit an ARIMA model without a predictor to predict battery health.If a declining trend is predicted then it is a degrading battery. otherwise, go back to step 2

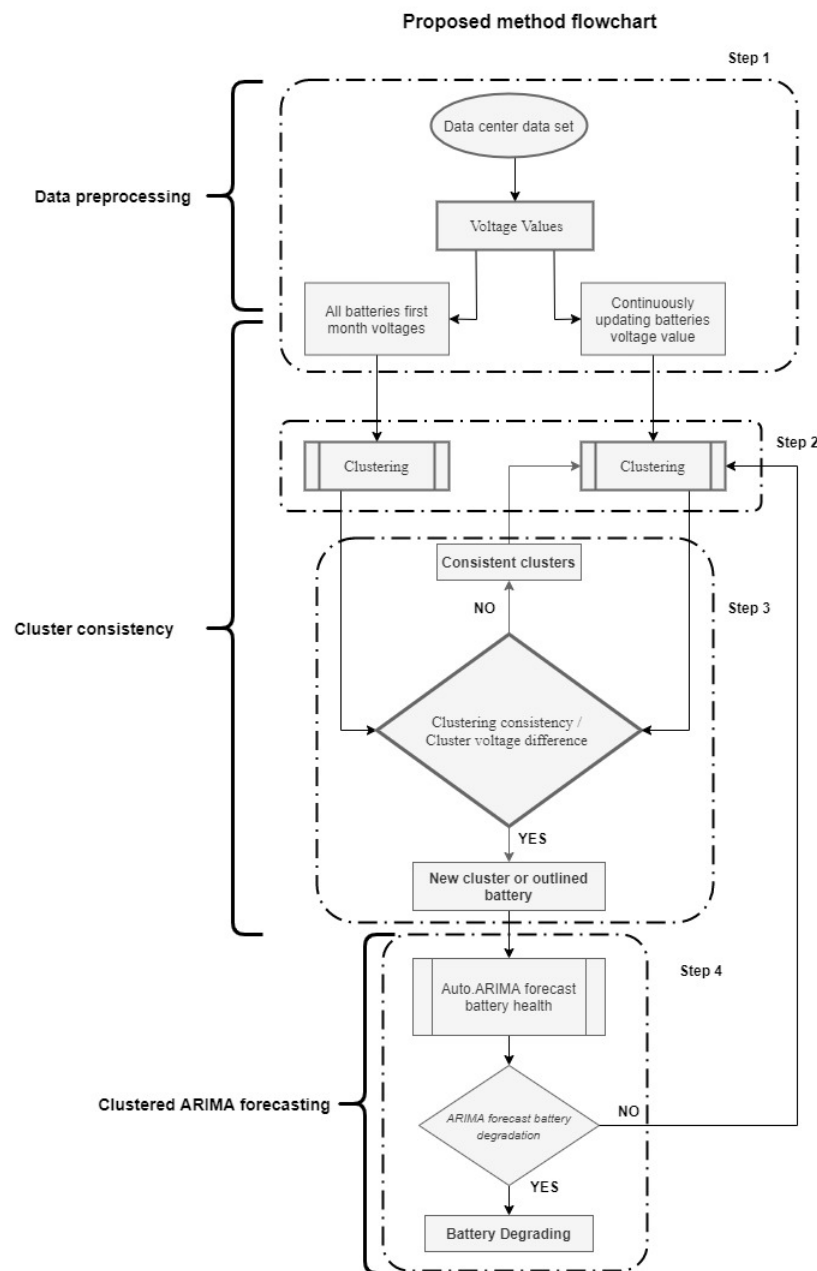


Figure 2. Proposed method flowchart

3.1. Data preprocessing

Data cleaning is the initial step in the data preprocessing step, identifying the missing values and correcting the raw data for analysis. See [section 2](#) for multiple features of the data set. Battery voltage data is utilized to forecast battery health. First month data is extracted from the data set and used as a baseline for comparing clustering and voltage status with real-time updated data. Data centers real-time data continuously update itself, we used one year of data and divide the data in 12 months to update on each iteration of clustering. See [subsection 4.1](#).

3.2. Cluster consistency

We now present our proposed algorithm 1 based on K shape clustering for battery cluster consistency for data centers. For a detail description of the K shape-based clustering algorithm (see [Section 3.2.1](#)).

K shape-based algorithm expects as input the battery voltage data set V_{ij} to be the first-month historic voltage data set and LV_{ij} the latest and updated voltage data set, where (i) is the time and (j) is the total number of batteries. B is the set of batteries when clustering is applied in the first month. LB is the set of batteries when clustering is applied in the latest month. DA is the set of inconsistent batteries cluster resulting in a comparison of the latest and first month LB , B . If DA is not equal to \emptyset , its an inconsistent or outlined battery cluster. MB, MC are the first and latest month clusters mean voltage respectively. These sets also represent cluster voltage status relative to other clusters. The difference between the MB, MC gives us DM . If DM is not equal to \emptyset , cluster voltage status is changed.

Algorithm 1: Cluster consistency and voltage status estimation

Input: V_{ij} and LV_{ij} are the first and latest month input data set, i =time and j =total number of batteries
Output: Outlined battery cluster DA

```

1 initialize
2 for first month clustering do
3    $B = V_{ij} \leftarrow$  clustering applied to input data set
4 return Set of initial clustering  $FB$ 
5 for first month cluster voltage status do
6    $FB \leftarrow$  Mean estimation applied to clusters
7 return First month cluster voltage status  $MB$ 
8 for latest month clustering do
9    $LB = LV_{ij} \leftarrow$  clustering applied to input data set
10   $LB$  // set of latest month clustering
11   $MC = LB \leftarrow$  Mean estimation applied to latest clusters
12   $MC$  // set of latest month clusters voltage status
13   $DA = LB - FB$  // Difference of initial clustering set and latest clustering set
14
15   $DM = MC - MB$  // Initial clustering set and latest clustering mean voltage difference
16
17  if  $DM = \emptyset$  then
18    No change in cluster voltage status
19  else
20     $DM \neq \emptyset$ 
21    Change in cluster voltage status
22     $DM$  is the set of changed voltage status batteries
23    if  $DA = \emptyset$  then
24      Consistent clusters
25    else
26       $DA \neq \emptyset$ 
27      Inconsistent cluster
28       $DA$  is the set of odd batteries
29    return New cluster or outlined batteries  $DA$ 
30 return Changed cluster voltage status  $DA$ 
  
```

3.2.1. K shape-based clustering

K shape clustering is an iterative refinement algorithm to isolate each cluster with keeping the shapes of time-series data. In K shape, cross-correlation measures are implemented to calculate the centroid of all clusters, and then update the members of each cluster [24].

In the assignment step, the algorithm updates the cluster memberships by matching all time series with all calculated centroids and by appointing each time series to the cluster of the closest centroid as shown in Equation 1. In the refinement step, the cluster centroids are updated to show the shift in cluster memberships in the previous step. The algorithm repeats these two steps until either no change in cluster membership occurs or the maximum number of iterations allowed is reached.

$$SBD(\vec{x}, \vec{y}) = 1 - \max_w \left(\frac{CC_w(\vec{x}, \vec{y})}{\sqrt{R_o(\vec{x}, \vec{x}) \cdot R_o(\vec{y}, \vec{y})}} \right) \quad (1)$$

3.3. Clustered ARIMA forecasting

A clustered ARIMA forecasting method is proposed for cluster consistency and cluster voltage status change in *DM* and *DA*. Algorithm 2 is proposed for forecasting battery health using cluster members as predictors to improve the forecasting accuracy than a simple ARIMA forecasting. ARIMA is widely used in numerous applications including finance, engineering, social sciences, and agriculture. ARIMA models are the integration of Autoregressive models (AR) and Moving Average models (MA). ARIMA models give good accuracy in forecasting relatively stationary time-series data [25]. *DA* or *DM* either can be input set. Extracting a battery element from the set v_j makes a new set *DC*. Extracting another element from *DA* from the remaining elements after v_j extraction results in *R*. *R* is the set of predictors used to forecast the battery element in *DC*. Fit an ARIMA model with *R* predictors to forecast *DC*. *AF* is the battery forecasted voltage values. If a declining trend is predicted in *AF* than the battery is degrading and if stable trend is predicted than the battery is stable

Algorithm 2: K shape clustered ARIMA forecasting

Input: *DA* set of outlined batteries
or *MA* voltage status changed
Output: *AF* ARIMA forecasts voltage behaviour

```

1 initialize
2  $DC = DA - \{v_j\} = \{v_j : v_j \in DA\}$            // Select a battery from the cluster set
3 if  $DC = \emptyset$  then
4    $DA \leftarrow$  Fit ARIMA model to DA
5    $AF \leftarrow$  Forecast with fitted ARIMA model
6 else
7    $DC \neq \emptyset$ 
8    $DR = DA - DC$   $R = DR - \{v_j\} = \{v_j : v_j \in DR\}$            // Select predictors
9    $DC \leftarrow$  Fit ARIMA model with R
10   $AF \leftarrow$  Forecast with fitted ARIMA model and R
11 return Battery voltage forecast AF
12 while AF is the voltage forecast do
13   Battery voltage forecast status check
14   if AF Decline in battery voltage then
15     AF is the set of degrading batteries
16   else
17     AF is the set of stable batteries
18 return AF Voltage forecast

```

4. Software implementation

4.1. Cluster consistency detection

Import the time-series data transformed into CSV format in the data preprocessing step for R programming. Dtwclust package is used for time series clustering in R. For K shape-based clustering battery data frame should be converted into a matrix see [Section 3.2.1](#)). Plot the K shape-based clustering with Plot function is used to visualize the results. Repeat this process every month until an inconsistent cluster is detected and perform clustered ARIMA forecasting See [subsection 4.2](#). An overview of the clustering inconsistency detection procedure is shown in [Figure 3](#).

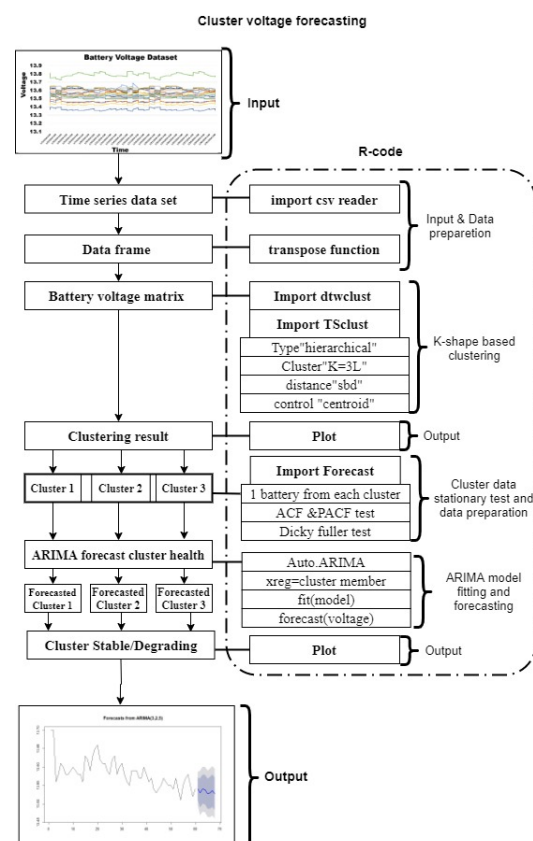


Figure 3. Battery cluster inconsistency and battery degradation forecast method

4.2. Implementing clustered ARIMA forecasting

The objective of this procedure is to forecast the battery voltage detected by cluster consistency. An overview of the method is shown in [Figure 3](#). Import "Forecast" package in R. Select a battery from the inconsistent cluster to forecast. Perform ACF, PACF and dicky fuller test to check the data stationarity. Use auto.Arima function to build the fitting model for the selected battery. Select cluster predictors for "Xreg" function in the fitting model, If the cluster contains only one battery than "Xreg" function is not required. Use the "forecast" function to forecast the battery voltage. If the declining trend is shown, the cluster is degrading and if the trend is stable than the battery will be stable in the future as well.

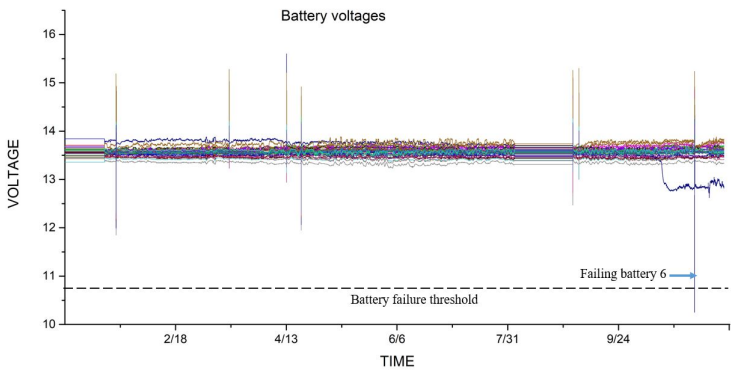


Figure 4. 1 year battery voltages in data center

5. Result and discussion

5.1. Data center battery setup

40 VRLA batteries are installed in the battery room with 20 batteries in each rack. Voltage data is collected in the BMS of the data center. There are four Discharge cycles and three power surges during one year of battery life in the data center as shown in Figure 4.

5.2. Battery voltage time series clustering

K shape-based clustering on the first month of the battery dataset results in 3 clusters see Figure 5. Cluster consistency is compared every month. Figure 6 shows consistent cluster members from 1st to 8th. Inconsistent cluster is shown in Figure 7 after 9 months. Battery 6 is now separated by battery 36 and 39, which was originally in the same cluster from the first month. This can also be interpreted as battery 6 is outlined from its original cluster.

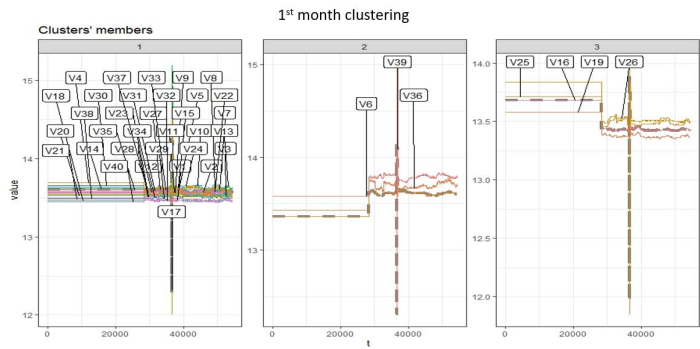


Figure 5. K shape-based 1st month clusters

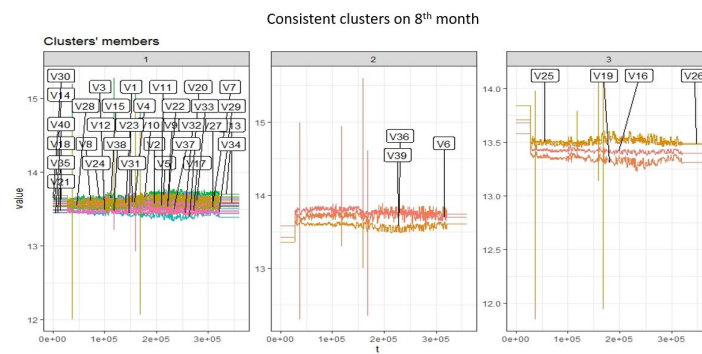


Figure 6. Consistent clusters after 8 months



Figure 7. Cluster inconsistency encounter after 9 months

This change in cluster consistency is an indication of a change in battery voltage behavior. To utilize this new information and predict the battery health from each cluster an improved accuracy forecasting model is discussed see [subsection 5.3](#)

5.3. ARIMA forecasting

The proposed clustered ARIMA approach with improved prediction accuracy model is evaluated by comparing actual voltage with clustered predictors(predictors within the cluster) , single predictors(without clustering), and total predictors(complete data).The metrics used are root mean square error (RMSE), mean average error (MAE), mean average percentage error (MAPE).

Batteries 6,15 and 36, one battery from each cluster is selected for demonstration. Cluster inconsistency was detected on 9th month, data of 9th month is transformed for the forecasting model. ACF and PACF for the transformed data are shown in [Figure 8](#). [Table 2](#) shows the augmented Dickey-Fuller test of the selected batteries. Batteries are selected from different clusters each shows different voltage behavior, which will require a different fitting model for each battery. The forecast package uses (auto.ARIMA) function to automatically select the best-fitted model by comparing other models. AIC and BIC are both penalized-likelihood criteria that are used for fit criteria[26]. [Table 3](#) shows the AIC and BIC values of the best-fitted model on the batteries for total, single and clustered predictors scenario.

Battery 6 (cluster 2) is a single member in cluster 2, battery 6 has zero external predictors in the cluster at the point of cluster inconsistency detection by K shape clustering. This makes battery 6 (cluster 2) a special case because clustered predictors and single predictor case is equal for battery 6. Prediction results of battery 6 with single or clustered predictors have better accuracy than Total predictors. This argument is further verified for battery 15 (cluster 1) and battery 36 (cluster 3) with the metrics comparison of the clustered predictors, single predictors and total predictors in [Table 4](#).

Table 2. The Augmented Dickey-Fuller

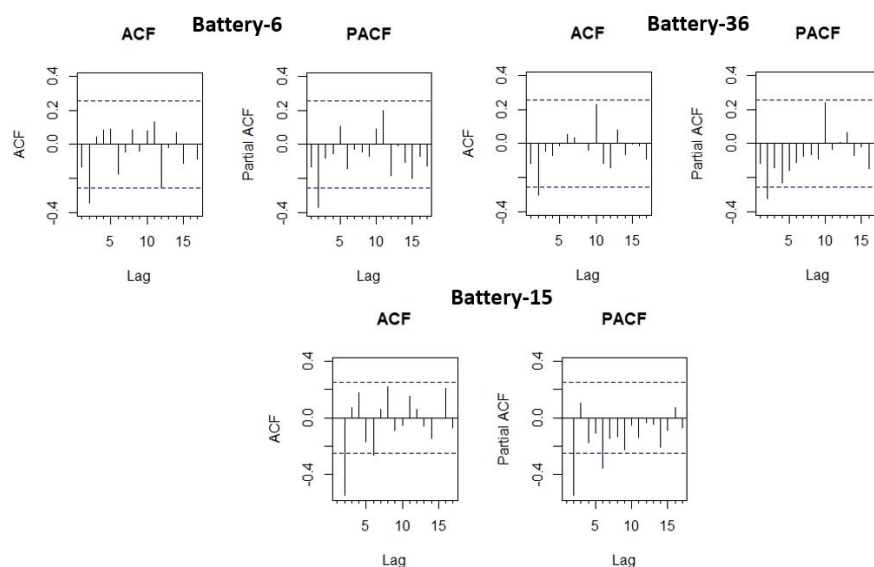
	Dickey-Fuller	Lag order	p-value
Battery 6	-5.6687	3	0.01
Battery 15	-4.8736	3	0.01
Battery 36	-7.1468	3	0.01

Table 3. Fitted models AIC and BIC values

	Battery6		Battery36			Battery15		
	T-predictor	S=C-predictor	C-predictor	S-predictor	T-predictor	C-predictor	S-predictor	T-predictor
AIC	-251.59	-268.59	-244.52	-228.09	-234.49	-297.4	-220.57	-255.08
BIC	-249.5	-206.81	-229.86	-207.31	-217.01	-288.96	-206.27	-252.99

The overall prediction accuracy after the application of K shape clustering is improved when clustered predictors are utilized.

Battery voltage is compared against actual voltage, Clustered predictors and single predictor in Figure 9, Figure 10 and Figure 11. Battery 6 only has one comparison with clustered predictor because it's the only battery in cluster 2. Battery 15 and 36 are compared with single predictor, clustered predictors and actual voltage in Figure 9 and Figure 11. Actual voltage is plotted against total predictors forecasting for battery 6, 15 and 36 in Figure 12, Figure 13 and Figure 14. It is clear from Table 4 and these figures that clustered predictor model is better fit for the battery voltage data.

**Figure 8.** Auto correlation and partial correlation of the selected battery data

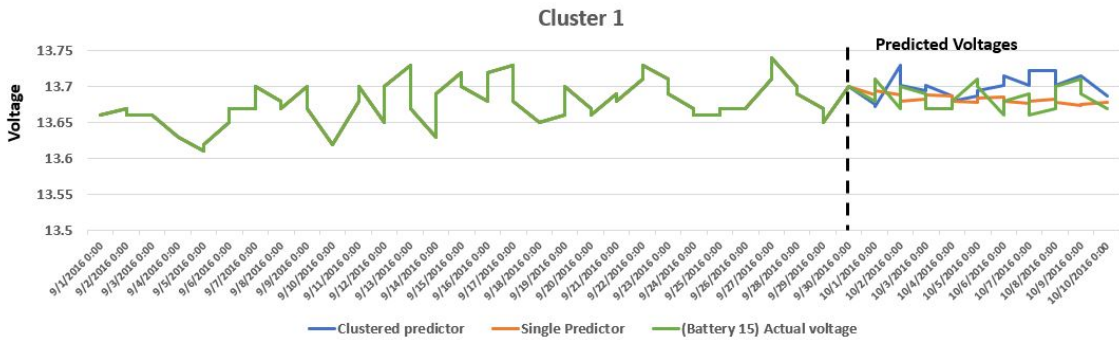


Figure 9. Battery-15 from cluster 1 , Comparison of measured and ARIMA forecasted voltage with Clustered predictors and single predictor

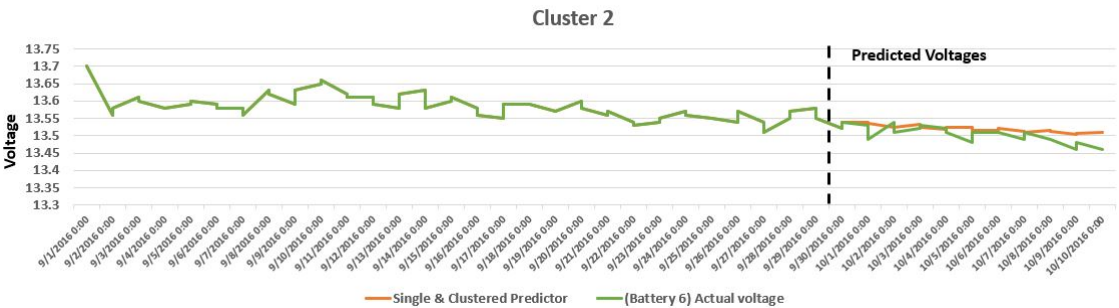


Figure 10. Battery-6 a single member of cluster 2 , Comparison of measured and ARIMA forecasted voltage with Clustered (single) predictor

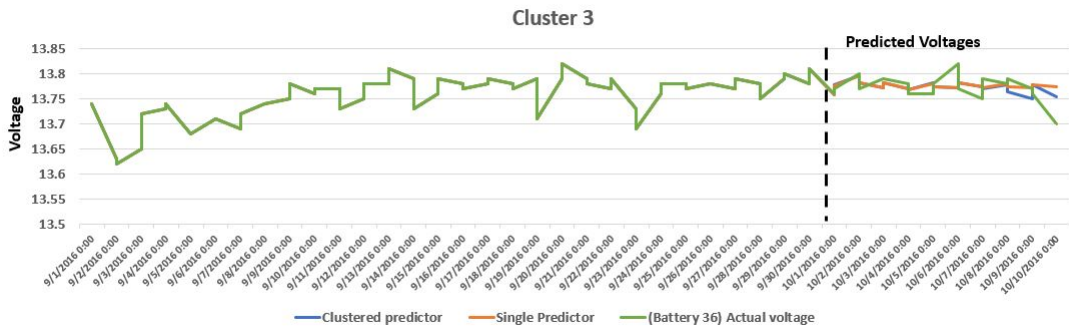


Figure 11. Battery-36 from cluster 3 , Comparison of measured and ARIMA forecasted voltage with Clustered predictors and single predictor

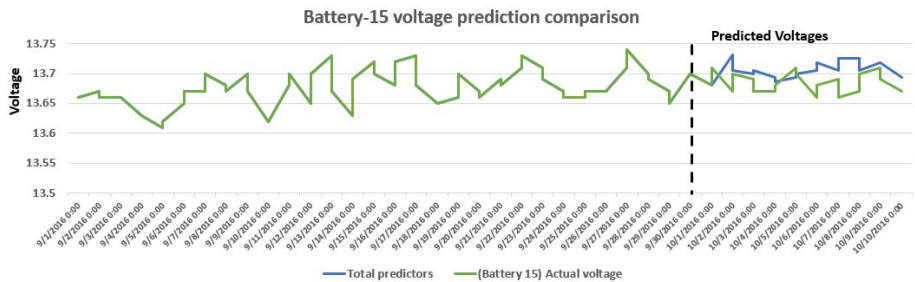


Figure 12. Battery-15 comparison of measured and ARIMA forecasted voltage with Total predictor

Table 4. K shape-based Clustered predictors (C), Single predictor(S) and Total predictors(T) ARIMA performance comparison

	Battery6		Battery36		Battery15			
	T-predictor	S=C-predictor	C-predictor	S-predictor	T-predictor	C-predictor	S-predictor	
RMSE	0.0253	0.0224	0.0252	0.0285	0.0283	0.0180	0.0282	0.0243
MAE	0.0206	0.0167	0.0186	0.0204	0.0219	0.0149	0.0225	0.0191
MAPE	0.1523	0.1233	0.1358	0.1489	0.1597	0.1096	0.1646	0.1398

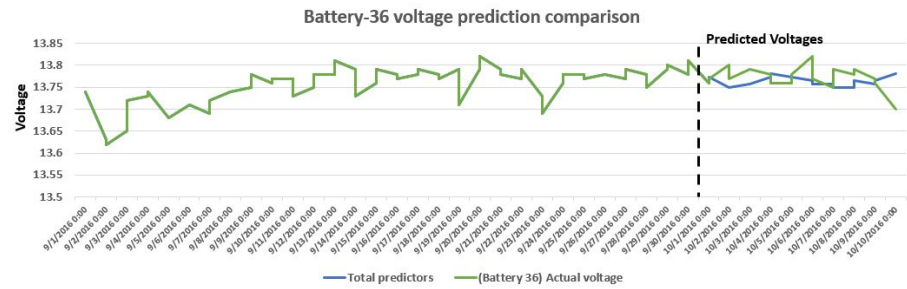


Figure 13. Battery-36 comparison of measured and ARIMA forecasted voltage with Total predictor

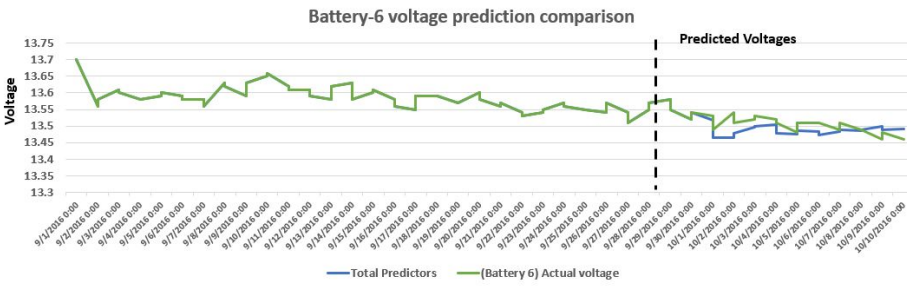
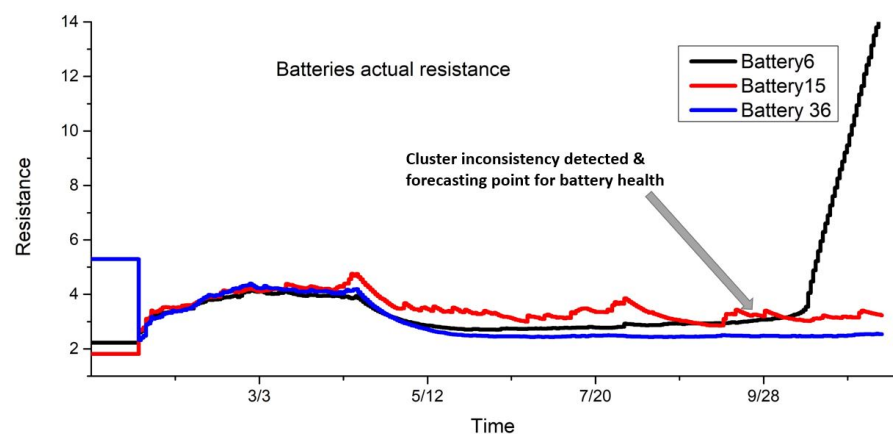


Figure 14. Battery-6 comparison of measured and ARIMA forecasted voltage with Total predictor

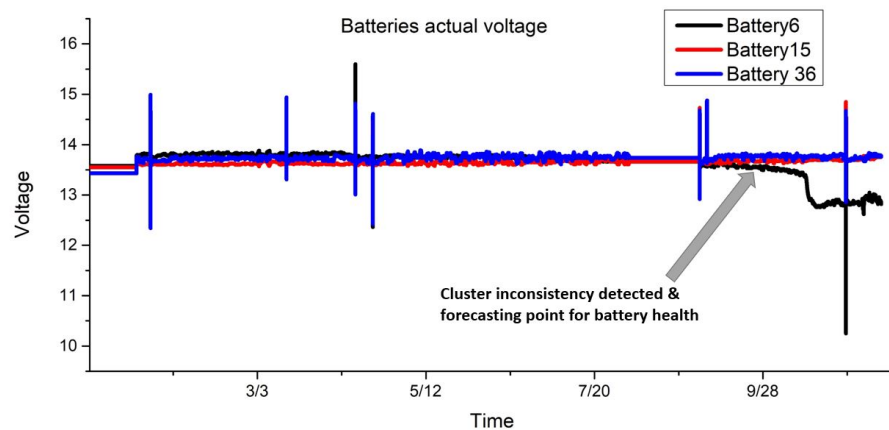
5.4. Effectiveness of clustered ARIMA approach

Forecasting model predicted declining trend in battery 6, this can be confirmed in Figure 15 with actual resistance data. Battery-6(cluster 2) shows an exponential increase in resistance after nine months. This degradation was accurately predicted by our proposed method. Similarly, stable resistance of battery 15 (cluster 1) and battery 36 (cluster 3) verify the predicted results. Actual voltages of the battery pack after the cluster inconsistency and predicted results are shown in Figure 16. A clear drop in battery 6 voltage can be seen which was predicted by the proposed method. Furthermore, battery 15,36 from clusters 1 and 3 predicted to be stable batteries as shown in Figure 16.



resistance-validate-v2.JPG

Figure 15. Actual resistances, Resistance rise in battery 6 and stable resistance for battery 15, 36 validate the proposed method



voltage-validate-v2.JPG

Figure 16. Actual voltages, Voltage drop in battery 6 and stable voltages for battery 15, 36 validate the proposed method

6. Conclusion

ARIMA forecasting with clustering predictors is proposed to predict the battery health. Forecasting accuracy of ARIMA model with the 40 batteries in the data center is improved. It is observed that the K shape-based clustering assisted result can significantly improve the ARIMA forecasting accuracy compared with the single predictor and total data predictors. A few challenges with our data-driven technique implications are the cleaning and preparation of data set, loss of data and missing values that have to be addressed to apply the proposed method.

Acknowledgement

This work received supports from National Key Research and Development Project of China under Grant 2017YFC0704100 and Grant 2016YFB0901900, in part by the National Natural Science Foundation of China under Grant 61425027, Tencent Inc., and in part by the 111 International Collaboration Program of China under Grant BP2018006, and BNRist Program (BNR2019TD01009).

References

1. Rahul, Urgaonkar, Bhuvan Urgaonkar, Michael J. Neely, and Anand Sivasubramaniam. "Optimal power cost management using stored energy in data centers." In Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems 2011, 221-232.

2. Miyuru, Dayarathna, Yonggang Wen, and Rui Fan. "Data center energy consumption modeling A survey." *IEEE Communications Surveys and Tutorials* 18, **2015**, 732-794.
3. Joao, Ferreira, Gustavo Callou, and Paulo Maciel. "A power load distribution algorithm to optimize data center electrical flow." *Energies* 6, no. 7, **2013**, 3422-3443.
4. Ferreira, Joao, Gustavo Callou, Dietmar Tutsch, and Paulo Maciel. "PLDAD-An Algorithm to Reduce Data Center Energy Consumption." *Energies* 11, no. 10, **2018**, 2821.
5. Callou, Gustavo, Joao Ferreira, Paulo Maciel, Dietmar Tutsch, and Rafael Souza. "An integrated modeling approach to evaluate and optimize data center sustainability, dependability and cost." *Energies* 7, no. 1, **2014**, 238-277.
6. Wu, Qiang, Qingyuan Deng, Lakshmi Ganesh, Chang-Hong Hsu, Yun Jin, Sanjeev Kumar, Bin Li, Justin Meza, and Yee Jiun Song. "Dynamo: facebook's data center-wide power management system." *ACM SIGARCH Computer Architecture News* 44, no. 3, **2016**, 469-480.
7. Hosseini, Shadi, Khalil Farhadi, and Sepideh Banisaeid. "Improving particle size of BaSO₄ with a unique glycerol base method and its impact on the negative active material of the lead-acid battery." *Journal of Energy Storage* 21, **2019**, 139-148.
8. Kristen A., Severson, Peter M. Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H. Chen et al. "Data-driven prediction of battery cycle life before capacity degradation." *Nature Energy* 4, no. 5, **2019**, 383-391.
9. Tang Xiaopeng, Ke Yao, Boyang Liu, Wengui Hu, and Furong Gao. "Long-term battery voltage, power, and surface temperature prediction using a model-based extreme learning machine." *Energies* 11, no. 1, **2018**, 86.
10. Hu, Xiaosong, Jiuchun Jiang, Dongpu Cao, and Bo Egardt. "Battery health prognosis for electric vehicles using sample entropy and sparse Bayesian predictive modeling." *IEEE Transactions on Industrial Electronics* 63, no. 4, **2015**, 2645-2656.
11. You, Gae-won, Sangdo Park, and Dukjin Oh. "Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach." *Applied Energy* 176, **2016**, 92-103.
12. Song, Yuchen, Datong Liu, Chen Yang, and Yu Peng. "Data-driven hybrid remaining useful life estimation approach for spacecraft lithium-ion battery." *Microelectronics Reliability* 75, **2017**, 142-153.
13. Zhou, Yapeng, and Miaohua Huang. "Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model." *Microelectronics Reliability* 65, **2016**, 265-273.
14. Luping, Chen, Liangjun Xu, and Yilin Zhou. "Novel approach for lithium-ion battery on-line remaining useful life prediction based on permutation entropy." *Energies* 11, no. 4, **2018**, 820.
15. Box, G. E., and G. M. Jenkins. "Time series analysis, forecasting and control-Segunda Edição." **1976**.
16. Kavasseri, Rajesh G., and Krithika Seetharaman. "Day-ahead wind speed forecasting using f-ARIMA models." *Renewable Energy* 34, no. 5, **2009**, 1388-1393.
17. Khashei, Mehdi, and Mehdi Bijari. "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting." *Applied Soft Computing* 11, no. 2, **2011**, 2664-2675.
18. Barak, Sasan, and S. Saeedeh Sadegh. "Forecasting energy consumption using ensemble ARIMA-ANFIS hybrid algorithm." *International Journal of Electrical Power and Energy Systems* 82, **2016**, 92-104.
19. Tseng, Fang-Mei, Gwo-Hshiung Tzeng, Hsiao-Cheng Yu, and Benjamin JC Yuan. "Fuzzy ARIMA model for forecasting the foreign exchange market." *Fuzzy sets and systems* 118, no. 1, **2001**, 9-19.
20. Zhang, G. Peter. "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing* 50, **2003**, 159-175.
21. Syed Naeem Haider, Q. Zhao, X. Li, Data driven battery anomaly detection based on shape based clustering for the data centers class, *Journal of Energy Storage*, **2019**, under review.
22. Yang, Junjing, Chao Ning, Chirag Deb, Fan Zhang, David Cheong, Siew Eang Lee, Chandra Sekhar, and Kwok Wai Tham. "k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement." *Energy and Buildings* 146, **2017**, 27-37.
23. Shahzadeh, Abbas, Abbas Khosravi, and Saeid Nahavandi. "Improving load forecast accuracy by clustering consumers using smart meter data." In *2015 international joint conference on neural networks (IJCNN)*. **2015**, 1-7.
24. Paparrizos, John, and Luis Gravano. "k-shape: Efficient and accurate clustering of time series." In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, **2015**, 1855-1870.

25. Wang, Huan, Jiejun Huang, Han Zhou, Lixue Zhao, and Yanbin Yuan. "An Integrated Variational Mode Decomposition and ARIMA Model to Forecast Air Temperature." *Sustainability* 11, no. 15, **2019**,4018.
26. Burnham, Kenneth P., and David R. Anderson. "Multimodel inference: understanding AIC and BIC in model selection." *Sociological methods and research* 33, no. 2, **2004**,261-304.