Detecting Electric Vehicle Battery Failure via Dynamic-VAE

Haowei He*,† hhw19@mails.tsinghua.edu.cn

Jingzhao Zhang*,† jzhzhang@mit.edu

Yanan Wang*,‡ wangyn7@mail.tsinghua.edu.cn

Shaobo Huang § huangshaobo@thinkenergy.net.cn

Chen Wang ¶ wangchen512@buaa.edu.cn

Yang Zhang § zhangyang@thinkenergy.net.cn

Dongxu Guo ‡ guodongxu@tsinghua.edu.cn Guannan He^{||,**}

Minggao Ouyang^{‡,**} ouymg@tsinghua.edu.cn

gnhe@pku.edu.cn

Abstract

In this note, we describe a battery failure detection pipeline backed up by deep learning models. We first introduce a large-scale Electric vehicle (EV) battery dataset including cleaned battery-charging data from hundreds of vehicles. We then formulate battery failure detection as an outlier detection problem, and propose a new algorithm named Dynamic-VAE based on dynamic system and variational autoencoders. We validate the performance of our proposed algorithm against several baselines on our released dataset and demonstrated the effectiveness of Dynamic-VAE.

Introduction

As the popularity of electric vehicles continues to rise, safety problems with on-board lithium-ion batteries have become increasingly important. Electric vehicle batteries not only constitutes a major fraction of the vehicle cost, its operation is also crucial to the safety of passengers and owners. Detecting an EV battery failure in a manner that is implementable and effective needs to take practical social factors into account. Traditional methods including inspection of the battery's physical structure and chemical composition usually requires the invasive procedure upon the battery. In this note, we describe an alternative way to provide early warning of battery failure is to analyze battery charging status information. Detecting abnormal data points in the sequential information is called time series anomaly detection in deep learning community.

Though our released dataset follows a standard time-series format, detecting battery anomalies differs from detecting anomalous events in time series in at least two aspects. On the one hand, there

^{*}Equal contribution.

[†]IIIS, Tsinghua University

[‡]State Key Laboratory of Automotive Safety and Energy, Tsinghua University

[§]Beijing Circue Energy Technology Co., Ltd.

[¶]School of Automation Science and Electrical Engineering, Beihang University

Department of Industrial Engineering and Management, Peking University

^{**}Corresponding author.

is no exact distinction between normal and abnormal batteries. In fact, according to the physical structure of the battery, the exact time of failure cannot be determined. On the other hand, in practical application scenarios, we can only find the battery failure when the vehicle fails. This means that in the multiple charging records of a faulty battery, the abnormal charging records may not be adjacent, nor may it appear only once. This is different from the existing public time series anomaly detection in deep learning community. In fact, existing studies do not reveal the power of deep learning for EV battery anomaly detection with large-scale publicly-available EV battery charging datasets, nor do they discover how practical factors should inform algorithm design.

To facilitate the advancement of research in this field, we release a large-scale EV battery charging dataset and propose a new model based on variational autoendocer. Our released dataset has 248 EVs in total, 46 of which are abnormal vehicles (caused by battery failure). The dataset has a clear hierarchy strucutre: the anomaly labels are at vehicle level rather than charging snippet level. To the best of our knowledge, no previous work has published a large-scale dataset of EV battery charging records. Meanwhile, we apply the deep learning framework and train the Dynamic Variational Autoencoder (VAE) algorithm on the dataset. Compared with conventional physics-based methods and common deep learning models, our algorithm produces a dominating ROC curve in average in predicting battery anomaly.

2 Related works

2.1 Time series datasets

There are common public time series anomaly detection datasets widely used in the analysis of time series anomaly detection algorithms. The SMAP (Soil Moisture Active Passive satellite) dataset is collected by a NASA's Earth Environment Satellite Observation Satellite [OENK10]. The MSL (Mars Science Laboratory rover) collects data sequences to determine if Mars was ever able to support microbial life. The water treatment physical test-bed datasets, SWaT (Secure Water Treatment) [GAJM16], and WADI (Water Distribution) [APM17], are sensor data recording simulated attack scenarios of real-world water treatment plants. For these public datasets, the anomaly events are exactly defined according to different criterion. The TSA (Time Series Anomaly detection system) contains time-series data from Flink⁸ [ZWD⁺20]. All the anomaly labels in these time series datasets are labeled at piece level. However, as we mentioned above, piece-level labels cannot be obtained in EV battery failure detection. We can only observe vehicle breakdowns due to battery failure.

2.2 Multivariate time series anomaly detection

Several recent works focus on multivariate time series anomaly detection. [MRA⁺16] propose to model reconstruction probabilities of the time series with an LSTM-based encoder-decoder network and use the reconstruction errors to detect anomalies. [HCL⁺18] leverage the prediction errors of the LSTM model to detect telemetry anomaly data. [SZN⁺19] propose OmniAnomaly to find the normal patterns through stochastic recurrent neural network and use the reconstruction probabilities to determine anomalies. [ZWD⁺20] capture multivariate correlations by considering each univariate series as an individual feature and including two graph attention layers to learn the dependencies of multivariate series in both temporal and feature dimensions. [DH21] apply a structure learning approach with graph neural networks to learn the inter-variable relationships. All the mentioned algorithms achieve high performance on piece level time series anomaly detection tasks.

3 A large EV battery charging dataset

3.1 Dataset

We release a large EV battery charging dataset for researchers to evaluate current anomaly detection algorithms and develop new ones. The dataset contains battery charging snippets collected from 248 cars and then cleaned by experienced engineers and data scientists. Among all the vehicles, 46 of them suffers from battery failure. There are over 650k snippets in total. Each charging snippet contains 128 timestamps sampled every 30 seconds (64 minutes record in total). The multiple features

⁸https://flink.apache.org/

are current, voltage, temperature and SOC information. We deliberately scaled and shifted each feature dimension by a random float to normalize the value and remove the sensitive information. Notice that the EV battery failure labels are on car level rather than snippet level, which is one of the main difference from the previous multi-variable time series anomaly detection datasets.

At the same time, we divided all the vehicles into two groups according to the brand of EVs. The first group of EVs come from the same EV brand, with a total of 198 vehicles, including 33 abnormal vehicles. The second group contains 49 vehicles, including 16 abnormal vehicles. We encourage researchers to consider transfer learning using these two datasets.

Our datasets are available through Vloong battery data platform⁹, and example usage can be found in our public code repository ¹⁰.

3.2 Evaluation pipeline

Since the battery failure labels are at vehicle level and the data are collected at charging snippet level, the evaluation pipeline is different from previous time series anomaly detection tasks. First, we split the data into training and test sets by vehicle. After the model is trained on the training set, we calculate the mean of the anomaly scores of all charging snippets for each vehicle in the test set. With ground truth vehicle anomaly labels, we can get the final receiver operating curve (ROC), which is a common evaluation metric in anomaly detection.

Remark Assigning vehicle level anomaly labels to each snippet may contrary to the facts, even though we can get finer labels. However, since the vehicle number is limited, the ROC curve may not be smooth which will make it hard to evaluate different algorithms. To compensate, we apply the five-fold cross validation to compute an averaged ROC curve in all our experiments.

4 Dynamic-VAE

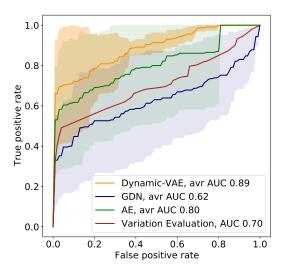


Figure 1: Interpolated averaged ROC curves of several algorithms on our EV battery dataset. Shaded area represents five-fold variance.

We develop a new detection algorithm based on the canonical variational autoencoder-decoder (VAE) model used in outlier detection.

Our model aims to focus on the state of the battery rather than the charging snippet itself. We provide a dynamic system perspective to the VAE pipeline, and hence we name the model dynamicVAE. That means, our model is trying to retrieve partial dimensions of the raw input sequence instead

⁹http://82.156.209.173/s/6Saazbbxq92iez7/download

¹⁰ https://github.com/thinkenergy/dynamic_vae

of compressing and retrieving the whole sequence like traditional VAEs. Specifically, we divide the input data into control inputs (current & soc) and observations (voltage & temperature). And tho control inputs together with the latent variable serve as input to the decoder of the VAE. Now, the decoder serves as a dynamic system parameterized by the latent variable and the hidden neural network weights. Formally, the reconstruction loss for a single data snippet is defined as

$$l(\theta, \zeta, x_0, x_1) = \text{MSE}(\text{Decoder}_{\theta}(z, x_0), x_1), \text{ where } z = \text{Encoder}_{\zeta}(x_0, x_1).$$
 (1)

Here θ and ζ denote neural network parameters in encoder and decoder respectively, whereas x_0 denotes the control input and x_1 denotes the observations. In addition to the reconstruction loss, we also added a regularization loss as in standard VAE model to limit model capacity, as well as a prediction head to predict cell mileage to guide learning.

Besides the "averaging all snippets for each vehicle" strategy, we also propose a procedure to robustly generate vehicle level predictions from the charging snippets level predictions if we have abnormal vehicles in the training set. We predict whether a charging snippet is abnormal by thresholding at the reconstruction error τ , and then predict whether a vehicle is abnormal by averaging the top p percentile errors. Both τ and p are finetuned with the additional abnormal vehicles in the training set.

5 Experiments

Besides our dynamic VAE, we also evaluate several traditional statistic methods used in battery failure detection and advanced time series anomaly detection algorithms on our dataset. As mentioned above, we perform five-fold cross validation for all the experiments. Specifically, 201 normal vehicles are divided into five folds. For each trial, we pick four folds of the normal vehicles as the training set and the rest one fold together with all the abnormal vehicles as the test set. The detailed split information can be found in our released code.

Interpolated averaged ROC curves are shown in Figure 4. We can see that our algorithm achieves the highest AUC value compared to two traditional algorithms, variation evaluation and risk early warning [LLH⁺20], a graph based algorithm [DH21] and the canonical autoencoder [Agg13]. Meanwhile, we notice that there are still room for researchers to improve the detection of EV battery failure. Limited to time and computing resources, we only implement these baseline algorithms. We also welcome researchers to test performance on our released dataset to improve electric vehicle safety.

6 Conclusion

In this note, we try to promote the detection of defective electric vehicle battery in the following two ways: (1) We release a large-scale EV charging dataset. It contains hundreds of vehicles with over 650k battery charging snippets. Anomaly labels are at vehicle level rather than snippet level, which is a main difference from the previous time series anomaly detection datasets. (2) We propose a new algorithm dynamic VAE to handle the multivariate time series anomaly detection problem from the perspective of dynamic system. The five-fold cross-validation experiments show a dominating detection performance over other baseline algorithms.

References

- [Agg13] Charu C. Aggarwal. Outlier Analysis. Springer, 2013.
- [APM17] Chuadhry Mujeeb Ahmed, Venkata Reddy Palleti, and Aditya P. Mathur. WADI: a water distribution testbed for research in the design of secure cyber physical systems. In Panagiotis Tsakalides and Baltasar Beferull-Lozano, editors, *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, CySWATER@CPSWeek 2017, Pittsburgh, Pennsylvania, USA, April 21, 2017*, pages 25–28. ACM, 2017.
- [DH21] Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 4027–4035. AAAI Press, 2021.

- [GAJM16] Jonathan Goh, Sridhar Adepu, Khurum Nazir Junejo, and Aditya Mathur. A dataset to support research in the design of secure water treatment systems. In Grigore M. Havârneanu, Roberto Setola, Hypatia Nassopoulos, and Stephen D. Wolthusen, editors, Critical Information Infrastructures Security - 11th International Conference, CRITIS 2016, Paris, France, October 10-12, 2016, Revised Selected Papers, volume 10242 of Lecture Notes in Computer Science, pages 88–99. Springer, 2016.
- [HCL⁺18] Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Söderström. Detecting spacecraft anomalies using 1stms and nonparametric dynamic thresholding. In Yike Guo and Faisal Farooq, editors, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, pages 387–395. ACM, 2018.
- [LLH+20] Yifan Lu, Kai Li, Xuebing Han, Xuning Feng, Zhengyu Chu, Languang Lu, Peifeng Huang, Zhi Zhang, Yongsheng Zhang, Fuqiang Yin, et al. A method of cell-to-cell variation evaluation for battery packs in electric vehicles with charging cloud data. eTransportation, 6:100077, 2020.
- [MRA⁺16] Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection. CoRR, abs/1607.00148, 2016.
- [OENK10] Peggy O'Neill, Dara Entekhabi, Eni G. Njoku, and Kent H. Kellogg. The NASA soil moisture active passive (SMAP) mission: Overview. In *IEEE International Geoscience & Remote Sensing Symposium, IGARSS 2010, July 25-30, 2010, Honolulu, Hawaii, USA, Proceedings*, pages 3236–3239. IEEE, 2010.
- [SZN+19] Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis, editors, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019, pages 2828–2837. ACM, 2019.
- [ZWD+20] Hang Zhao, Yujing Wang, Juanyong Duan, Congrui Huang, Defu Cao, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, and Qi Zhang. Multivariate time-series anomaly detection via graph attention network. In Claudia Plant, Haixun Wang, Alfredo Cuzzocrea, Carlo Zaniolo, and Xindong Wu, editors, 20th IEEE International Conference on Data Mining, ICDM 2020, Sorrento, Italy, November 17-20, 2020, pages 841–850. IEEE, 2020.