Identifying electric vehicles from smart meter recordings

Ali Babalhavaeji
school of engineering
RMIT University
Melbourne, Australia
ali.babalhavaeji@student.rmit.edu.au

Hui Song
school of engineering
Hangzhou Dianzi University
Hangzhou, China
hui.song@hdu.edu.cn

Mohammadreza Radmanesh
school of engineering
RMIT University
Melbourne, Australia
s3697007@student.rmit.edu.au

Mahdi Jalili
school of engineering
RMIT University
Melbourne, Australia
mahdi.jalili@rmit.edu.au

***Abstract*—Electrification of transport systems is a key to decarbonize the economy and mitigate climate change. Electric Vehicles (EVs) are used to electrify the transport systems. But charging of EVs raises concerns for power grids, such as load unbalance and voltage fluctuations. Therefore, distributed network operators need to have full visibility on the EVs in the network. This requires identifying which consumer has an EV. Smart meter devices collect electricity consumption of consumers, which can be potentially used to identify whether own an EV that is charge at home. Currently, the EV owners are in minority and data is heavily unbalanced meaning that the number of EV owners is much less than those without. In this paper, we propose an outlier detection-based method that does not use a supervised learning to identify EV owners. The proposed model uses temporal information from smart meter time series. The performance of the proposed method is compared with state-of-the-art resampling-based methods in terms of F1 measure, precision, recall, and accuracy. We use real-world data to evaluate the performance.**

***Index terms*-- Electric vehicles, Outlier detection, Recurrent neural networks, Convolutional Neural Networks,**

I. INTRODUCTION

Greenhouse gas is a concerning result of burning fossil fuels for energy and transport demands, which has resulted in destructive climate change [1]. The International Energy Agency (IEA) determined targets to control the increase of global temperature uptake to two degrees Celsius. This has initiated movement in many countries to put zero emission targets to reduce their carbon footprints [2]. Electrification of the transport systems through Electric Vehicles (EVs) is a promising way to decarbonize the transport sector and reduce the emissions [3]. However, EV adoption may introduce some new challenges to the electricity grid. Our grid has not been originally designed to support mass EV uptake and uncoordinated/unmanaged EV charging may raise concerns like load unbalance, voltage fluctuation, and instability of the grid [4].

Due to the lower price and also convenience, many consumers prefer to charge their EVs at home [5]. EVs are much larger load compared to other appliances and they must be carefully considered in distribution network operation and planning. Distribution Network Operators (DNOs) often require full visibility of EVs in their network to enable them properly studying their impact [6]. Many customers charge their EVs from standard wall socket and in most cases, DNOs are largely unaware of their presence. Therefore, they need to employ effective approaches to identify customers owning and EV (who are charging it at home). Smart meter data is recorded at scale and can be used to identity EV loads.

The problem of identifying EV from smart meter recordings has recently received attention in the literature. Some studies have approached the problem as a classification task; customers are classified into two classes, those with an EV and those without. The main issue of this problem is the imbalance issue [7]. This issue means one class, the bigger one in terms of size, in our case customers without EVs (nonEV), has much more records than the other one, in our case EV part [8]. In a recent work [5], the author develops a resampling algorithm to address the imbalance issue. After resampling, the new training data is given to classifiers like Support Vector Machine (SVM) [9] or Random Forest (RF) [10] to learn the classification boundaries. The resampling methods are of two types: oversampling and undersampling. The oversampling means creating more records from the minority class and undersampling means removing records from the bigger class [11]. [5] takes an oversampling approach and proposes an ensemble classifier. [12] take a resampling approach as well approach and propose a RF to identify EVs.

Outlier detection is another approach to deal with imbalance issues. It is the case where a few records differ significantly from the distribution of the rest of the records [13]. In this paper, we propose an outlier detection-based approach to solve the EV identification problem. Our method is a prediction-based method that uses temporal information of consumption time series. Our experiments on real smart meter recordings show that the proposed method is more effective than supervised classification approaches previously proposed for EV identification.

II. Methodology

A. Problem Definition

Our problem is to use electricity consumption time series recorded by smart meters and to identify consumers who have an EV (and charge it at home). EV users are still in the minority in almost all countries including Australia (where the data is recorded). Therefore, outlier detection methods can be developed to identify them from non-EV users. We formulate the problem as an outlier detection problem, where non-EV users correspond to normal data and EV users correspond to abnormal data. Let's denote the input with $\chi_t = (x_1, \cdots, x_t)$ and the output with $\Omega \in \{0,1\}$, where 0 represents nonEVs and 1 represents EVs.

For a record to be an outlier, it should have two main properties: 1) scarcity and 2) deviation from the distribution of the normal part. Our data analysis showed that EV users' patterns hold the two properties. Our problem turns to derive $\Omega: \chi^W \rightarrow \{1,0\}$ where W is the size of the time window which slides over the electricity consumption time series., we express the problem as follows:

$$\Omega(x) \rightarrow \begin{cases} 1 & if e > \omega \\ 0 & o.w \end{cases} \quad (1)$$

where $x \in \chi_t \subseteq R^W$, ω is a threshold that determines the boundary between non-EV and EV users and e is residual error.

B. Proposed Method

We propose a deep learning-based outlier detection method, which includes training and detection steps. The training step includes a deep learning module and a threshold module. Fig 1 shows the framework of the proposed model, called ODEV (Outlier Detection EV identifier). Before training the deep neural network, the method slides a window on time series to create time windows. Then, the windows are given to the deep learning module, to extract temporal information. The deep learning module uses Convolutional Neural Networks (CNN) and Gated Recurrent Unit (GRU). The ODEV continues training on non-EV data until the neural network converges. After convergence, the method derives a threshold ω. The threshold is a boundary that distinguishes EV and non-EV users.

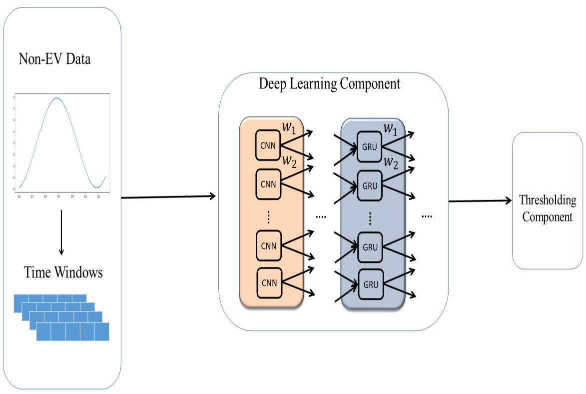

Fig 1: Framework of the ODEV.

C. Training Step

1) Deep Learning Module

This module uses two kinds of deep neural networks, CNN and GRU. CNN consists of two kinds of layers, convolutional, and pooling layers. The convolutional layer maps the input into a new feature space with lower dimensions, named feature map, by sliding filters on input. the input of the convolutional layer is a window of the consumption time series. The pooling layer outputs a new representation by applying statistical operations on the feature map. Then, the GRU layer takes the output of the pooling layer to extract temporal information. The GRU cells include two types of gates, the reset gate which decides if the last hidden state should be ignored, and the update gate which decides if the output of the current state should get updated. We minimize the mean squared loss function between the next timestamp, χ_{t+1} and its associated prediction generated by deep neural networks, $\hat{\chi}_{t+1}$ as follows:

$$\mathcal{L} = \|\chi_{t+1} - \hat{\chi}_{t+1}\|_2^2 \quad (2)$$

2) Threshold Module

This module uses $\mathcal{L}$ over the training data to derive the ω. $\langle \chi_{tr}, Z_{tr} \rangle$ is the training data where $\chi_{tr} = \{\chi_1, \ldots, \chi_N\}$ and $Z_{tr} = \{Z_1, \ldots, Z_N\}$is next step values. The ODEV method uses $\mathcal{L}$, which calculates the prediction residual error. We express the threshold as follows:

$$\omega = \max\left(\mathcal{L}(Z_{tr}, \hat{\chi}_{tr})\right) \quad (3)$$

D. Detection Step

In this step, ODEV uses ω to detect EVs. We have test data $D_{ts} = \langle \chi_{ts}, Z_{ts} \rangle$. ODEV calculates $\mathcal{L}(Z_{ts}, \hat{\chi}_{tr})$for each test data records as follows:

$$e_{ts} = \{\mathcal{L}(Z_{ts}, \hat{\chi}_{tr}) | \langle X_i^{ts}, Z_i^{ts} \rangle \in D_{ts}\} \quad (4)$$

Finally, the detection operation, denoted by $\Omega(X_i^{ts})$ is run as follows:

$$\Omega(X_i^{ts}) = \begin{cases} 0 & if\ e_i \leq \omega \\ 1 & e.t \end{cases} \quad (5)$$

III. Experiments

A. Data Description

The dataset includes electricity consumption records for consumers in Victoria, Australia. The data is collected from March 2021 to May 2021 with a time interval of 30 minutes. It consists of 1047 users (39 EV users and 1008 non-EV users).

B. Experiments

We compare ODEV with baselines including ECM [5], and RF [12].

C. Evaluation Metric

We use four main metrics as follows:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

$$Accuracy = \frac{TP + TN}{N} \quad (9)$$

where TP, FP, and FN stand for true positive, false positive, and false negative, respectively. True positive means EV users who are correctly identified as EVs, false positive means nonEV users that are falsely identified as EVs, and false negative means EV users who are falsely identified as nonEV users. N is the number of users. In cases where imbalance issues hold, the F1 measure is the most important.

D. Results and Discussion

We conduct a grid search over the hyperparameters: the number of neurons, the number of layers, the optimizer, and the activation function. Table 1 shows that for almost all metrics ODEV outperforms. In cases where the imbalance issue holds, F1 is the most important metric. ODEV outperforms all baselines in terms of F1 with F1 value of 0.5519. the second best in terms of F1 is ECM with 0.4608 which has a big gap with ODEV. Our method enjoys the abundance of the normal population, i.e. the non-EV users in our case, and calculates a good threshold to separate EV users and non-EV users. We use 80% of the non-EV records for the training module and then we use a mix of EV and non-EV data for the detection step.

TABLE 1:PERFORMANCE COMPARISON

Baselines	Accuracy	Precision	Recall	F1
RF	0.7836	0.3065	0.8538	0.4509
ECM	0.8224	0.3374	0.7308	0.4608
ODEV	**0.8326**	**0.5102**	0.6012	**0.5519**

ODEV outperforms by a great margin in terms of precision, and most importantly F1. The accuracy of ODEV is 0.8326. The second-best accuracy is ECM method with an accuracy of 0.8224. The best recall result is gained by RF with 0.8538, whose precision result is poor, a precision of 0.3065, leading to a lower F1 of 0.4509. The goal of the recall is to reward a method when it identifies true positives correctly and penalize it when it outputs false negatives. None of the baselines capture temporal information of electricity consumption time series, which is the reason for a lower F1 score. Regarding precision, which penalized false positives, ODEV archives 0.5102, followed by ECM with 0.3374.

IV. CONCLUSION

In this paper, we proposed an outlier detection algorithm, named ODEV that uses the capabilities of recurrent neural networks for capturing temporal information, for the EV identification problem. We first train the model on the electricity consumption data of users without an EV. Then, ODEV calculates a threshold for non-EV consumption patterns. Using the threshold and the trained model, we performed the identification on a mix of EV and non-EV users. We showed that ODEV outperforms baselines using a real-world data collected in Victoria, Australia.

V. ACKNOWLEDGMENTS

This research was partially supported by Australian Research Council through project number LP180101309.

REFERENCES

[1] Yong, J.Y., et al., *A review on the state-of-the-art technologies of electric vehicle, its impacts and prospects.* Renewable and sustainable energy reviews, 2015. **49**: p. 365-385.

[2] Kobashi, T., et al., *On the potential of "Photovoltaics+ Electric vehicles" for deep decarbonization of Kyoto's power systems: Techno-economic-social considerations.* Applied Energy, 2020. **275**: p. 115419.

[3] Li, H., et al., *Optimal cost minimization strategy for fuel cell hybrid electric vehicles based on decision-making framework.* IEEE Transactions on Industrial Informatics, 2020. **17**(4): p. 2388-2399.

[4] Rizvi, S.A.A., et al. *Electric vehicles and their impacts on integration into power grid: A review.* in *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*. 2018. IEEE.

[5] Song, H., et al., *Ensemble Classification Model for EV Identification from Smart Meter Recordings.* IEEE Transactions on Industrial Informatics, 2022.

[6] Pagani, M., et al., *User behaviour and electric vehicle charging infrastructure: An agent-based model assessment.* Applied Energy, 2019. **254**: p. 113680.

[7] Kaur, H., H.S. Pannu, and A.K. Malhi, *A systematic review on imbalanced data challenges in machine learning: Applications and solutions.* ACM Computing Surveys (CSUR), 2019. **52**(4): p. 1-36.

[8] Krawczyk, B., et al., *Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy.* Applied Soft Computing, 2016. **38**: p. 714-726.

[9] Pisner, D.A. and D.M. Schnyer, *Support vector machine*, in *Machine learning*. 2020, Elsevier. p. 101-121.

[10]. Belgiu, M. and L. Drăguţ, *Random forest in remote sensing: A review of applications and future directions.* ISPRS journal of photogrammetry and remote sensing, 2016. **114**: p. 24-31.

[11] Cordón, I., et al., *Imbalance: Oversampling algorithms for imbalanced classification in R.* Knowledge-Based Systems, 2018. **161**: p. 329-341.

[12] Jaramillo, A.F.M., et al. *Supervised non-intrusive load monitoring algorithm for electric vehicle identification*. in *2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. 2020. IEEE.

[13] Pang, G., et al., *Deep learning for anomaly detection: A review.* ACM Computing Surveys (CSUR), 2021. **54**(2): p. 1-38.