Probabilistic Forecasting Model for Non-normally Distributed EV Charging Demand

Daisuke Kodaira
Department of Electrical Engineering
Tokyo University of Science
Chiba, Japan
daisuke.kodaira03@gmail.com

Junji Kondoh
Department of Electrical Engineering
Tokyo University of Science
Chiba, Japan
j.kondoh@rs.tus.ac.jp

***Abstract*—A method for probabilistic electric vehicle (EV) demand forecasting is proposed in this paper. The EV demand in a certain area is forecasted by an ensemble forecasting model. The forecast result includes a deterministic forecasting and prediction interval that indicates the probability of deviation from deterministic forecasting. In the case study, an actual observed dataset from the UK is used to verify the proposed algorithm. The results show that the target EV charging demand to be forecasted shows 80% prediction interval coverage for 27 days out of 30 simulated days.**

***Keywords*—probabilistic forecasting, prediction interval, ensemble forecasting, electric vehicle, demand forecasting**

I. Introduction

The number of electric vehicles (EVs) is rapidly increasing every year. The deployment of EVs in certain countries increased from three million in 2017 to five million in 2018 [1]. Supplying the charging demand for such a huge number of EVs has an impact on the low-voltage distribution network [2]. Studies have proposed control strategies to avoid high charging demand in buildings or certain areas owing to EV charging [3]–[5]. The voltage deviation in the network is a another major problem, which is addressed in [6]–[8]. All these strategies propose charge and/or discharge schedules for EVs. These schedules must be defined before the system operates to calculate the optimal solution. Forecasting EV charging demand is an essential aspect of leveraging EVs in distribution networks. Numerous works have proposed forecasting methods for EV charging demand. Recent studies mainly adopted deterministic forecasting [5][9][10]. In [9], ARIMA and GARCH are ensembled. The weight for each algorithm is comprehensively analyzed. In [10], both historical and real-time data obtained via the global positioning system are utilized to forecast EV charging demand. The challenge in deterministic forecasting is risk management. EV demand in inevitably erratic. For example, if the objective of EV demand control were to reduce the peak-load of buildings or the local grid, another reliable energy storage and control system such as large-scale batteries would be required as a backup energy resource. The backup resource bolsters the reliability in case the forecasting of EV demand has an unexpected error, but the investment cost also increases. Probabilistic EV demand forecasting is more practical for demand-side energy management. A few studies adopted probabilistic forecasting [11]–[13]. In [11], the state-of-charge (SOC) demand is forecasted with a probabilistic method. The distribution of the required SOC in [11] is assumed to follow a normal distribution consistently, but this assumption is not realistic, as shown in Section II-A of this paper. In [13], the hourly number of electric buses for battery swapping and its uncertainty is forecasted with a 90% prediction interval (PI). In [12], the probabilistic one-hour-ahead charging load is forecasted. Traffic flow is utilized to generate EV charging demand, and a probabilistic queuing model is also proposed.

In this paper, a method for one-day-ahead EV demand forecasting is proposed. The proposed method can forecast EV demand with a PI implying prediction error risk. First, the deterministic EV demand forecast is performed by the proposed ensemble forecasting model. Second, the error between the target and forecasted EV demand accumulates with respect to a specific duration. Finally, the ensemble deterministic forecasting and the obtained error distribution lead to the PIs. The charging history of the local chargers distributed in a city is utilized for the forecasting data. The data do not follow a normal distribution, but the proposed method can generate reliable PIs in the case study. The simulation codes and dataset are available on GitHub [14].

The remainder of this paper is organized as follows. Section II analyzes the original dataset and explains the proposed methodology. In Section III, the proposed probabilistic forecasting algorithm is verified by the simulation with the actual dataset. Finally, the conclusion is drawn in Section IV.

II. Forecasting Method and Data

A. Dataset analysis

The dataset is obtained from Dundee, Scotland [15]. The accumulated data contain the charging event number, user ID, plug-in time, unplugged time, total charged energy during connection, charging points, and charger type. Fig. 1 shows the geographical locations of the EV charging points. The charging points are indicated by red markers on the map. The charging points are radially distributed from the central area of the city. The original observed data include every plug-in and unplug as individual events. However, the energy transaction in the process of charging is not obtained. In this work, the original observed data are converted into a fifteen-minute basis. Fig. 2 shows an example of the data conversion from observed data to input data for the proposed forecasting model. In Fig. 2, the upper table shows the original data format. User 1 plugs in at 10 am and unplugs at 11 am. The charging amount during this period is 20 kWh. In the data conversion process, the 20

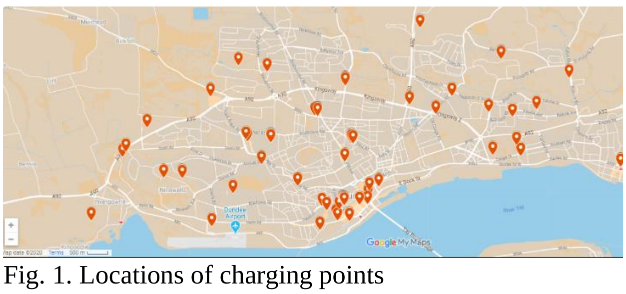

Fig. 1. Locations of charging points

978-1-7281-8550-7/20/$31.00 ©2020 IEEE

kWh is equally distributed as shown in the lower table of Fig. 2. User 2 also charges from 10:15 to 10:45. The charging amount of 10 kWh is distributed as for User 1. As a result, the energy transactions from User 1 and User 2 are accumulated from 10:00 to 10:45 as shown in the lower table of Fig. 2.

	Start time		End time
User1	10:00	→ Charing **20 kWh**	11:00
User2	10:15	→ Charing **10 kWh**	10:45

Convert

Quarter	Charging Demand [kWh]	Participants
10:00-10:15	5	user1
10:15-10:30	5+5=10	user1, user2
10:30-10:45	5+5=10	user1, user2
10:45-11:00	5	user1

Fig. 2. Example of data conversion from observed data to training data for forecasting models

In this study, the total EV demand is forecasted. All the energy transactions at the charging points are combined into fifteen-minute intervals. Fig. 3 shows a histogram of the EV charging demand from 5 am to 6 am. The total charging demand in this period does not follow a normal distribution. The difference between the actual and normal distributions is shown in Fig. 4 as a quantile–quantile plot. In Fig. 4, the red dotted line is an ideal line corresponding to the observed samples with a normal distribution. The blue symbols indicate the observed total charging demand samples. The blue and red plots do not correspond. This gap between the samples and the normal distribution is observed in the other durations as well. The method proposed in this work handles non-normally distributed data to generate a probabilistic forecast.

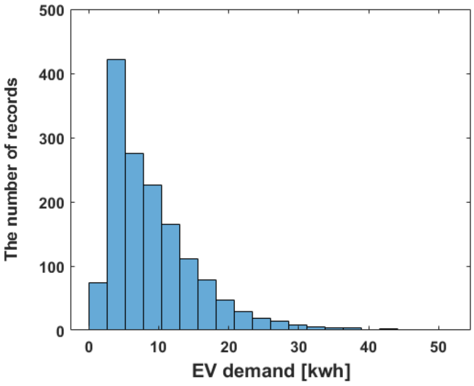

Fig. 3. Example of a histogram of the EV charging demand from 5 am to 6 am

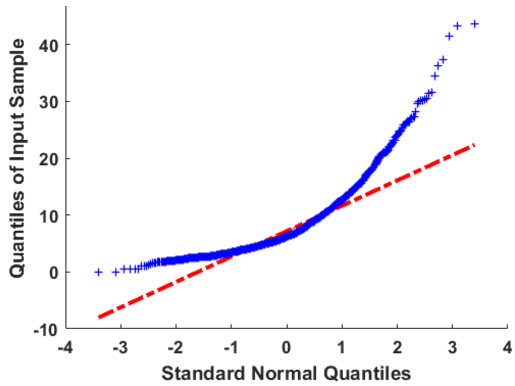

Fig. 4. Example of quantile–quantile plot of EV charging demand from 5 am to 6 am

B. Forecasting models

Before forecasting the future (unknown) EV charging demand, the forecasting model must be trained for the proposed method. In the training process, the proposed forecasting method consists of three steps: 1) perform deterministic forecasting by multiple forecasting models, 2) optimize weights for each deterministic forecasting model to be ensembled, 3) calculate the error between the observed EV demand and that forecasted by the ensemble model. After the training process, we trained the ensemble model and error distribution derived from 3). Using the trained ensemble model and error distribution, the unknown EV demand is forecasted with a PI. The data are configured as described in Fig. 5. Two groups are arranged for training and forecasting: long-term past data for training and validation, and forecast data. The long-term data contain predictors (year, month, day in a month, day of the week, and holiday or not) and target (EV charging demand); the forecast data contain only predictors. Training data are utilized as a training dataset for the construction of the k-means [16] and neural network [17] forecast models. The long-term past data preferably contain at least one year's worth of data to capture seasonal features. In the following case study, the long-term past data consist of 34,921 records collected every 15 min from 1st September 2017 to 31st August 2018. Validation data are selected in sets of arbitrary length from the long-term past data. The validation data are utilized to determine the optimal weight for the ensembled forecast model. The proposed algorithm works as one-day-

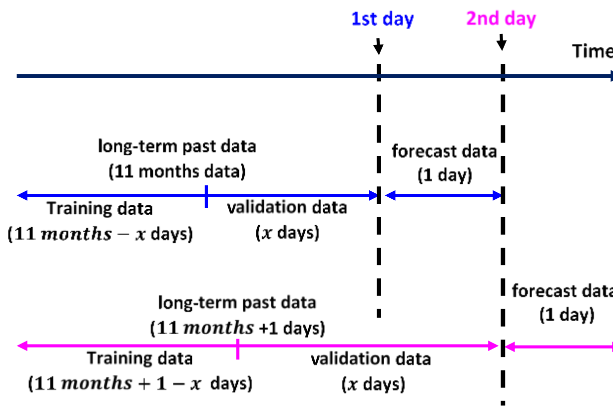

Fig. 5. Arrangement of a given dataset for EV demand forecasting

ahead forecasting. The forecast data contain predictors for the next 24 hours.

In this paper, an ensemble forecasting model is proposed. Optimal coefficients are arranged for two independent forecasting methods. The ensemble method is referenced from [18]. An ensemble prediction model is constructed by combining these two prediction models with weights, as shown in Eq. (1):

$$\hat{y}_t^i = p_t \widehat{K}_t^i + q_t \widehat{F}_t^i \tag{1}$$

where $\hat{y}_t^i$ is the ultimate deterministic predicted value of the load for time t on the *i-th* day, and $\widehat{K}_t^i$ and $\widehat{F}_t^i$ are the deterministically predicted values by k-means and neural network models at time t, respectively. The coefficients p_t and q_t are the weights of each prediction. The values of the weights are time-dependent values sought by the particle swarm optimization algorithm that minimizes the error between the observed and predicted loads, as shown in Eq. (2):

$$\underset{p_t, q_t}{\boldsymbol{arg}\ \mathbf{min}} \left\| \mathbb{Y}_i - \widehat{\mathbb{Y}}_i \right\|_2 \tag{2}$$

where

$$\begin{aligned} \mathbb{Y}_i &:= \{y_1^i, y_2^i, y_3^i \dots y_{96}^i\} \\ \widehat{\mathbb{Y}}_i &:= \{\hat{y}_1^i, \hat{y}_2^i, \hat{y}_3^i \dots \hat{y}_{96}^i\} \end{aligned} \tag{3}$$

$\mathbb{Y}_i$ is the set of observed data y_t^i corresponding to the predicted load $\hat{y}_t^i$ at time t on the *i-th* day, and $\widehat{\mathbb{Y}}_i$ is the set of predicted loads $\hat{y}_t^i$. In this work, the observed data are composed of 15-minute intervals. t comprises 96 instances during the day. Deterministic forecasting by the ensemble model is performed for past data for a specific time duration, such as the period of one year (i = 1, 2, …, 365).

Once the optimal p_t and q_t are obtained, future EV demand is forecasted with the trained ensemble model. The boundaries of PIs are subsequently calculated. The absolute error set $\mathbb{E}_t$ for the specific time t is derived by comparing the forecasted and the observed data through validation data. A series of errors and the error sets are expressed as follows:

$$e_t^i = y_t^i - \hat{y}_t^i \tag{4}$$

$$\mathbb{E}_t := \{e_t^1, e_t^2, e_t^3 \dots e_t^i \dots e_t^{days}\}, 1 \le t \le 96 \tag{5}$$

where e_t^i is the prediction error for the *i-th* day at time t ($1 \le t \le 96$). The value of *days* indicates how many days are included for the validation dataset. Each time t has an error record for *days*. The set $\mathbb{E}_t$ forms a histogram for each time t and is herein referred to as the error distribution. After the error distribution is formulated in the model training process, the deterministic EV demand is forecasted for the next 24 hours. The deterministic forecasting and error distribution are added together into a series of histograms. A set of deterministic forecasting with errors for the time t is defined as follows:

$$\mathbb{D}_t := \{\hat{y}_t + e_t^1, \hat{y}_t + e_t^2, \hat{y}_t + e_t^3 \dots \hat{y}_t + e_t^{days}\} \tag{6}$$

After defining the set $\mathbb{D}_t$, the boundaries are defined by a fitting sample-based method [18]. For the sample-based PIs for the set $\mathbb{D}_t$, a certain percentage of past data is considered from the entire dataset. The upper and lower boundaries of PIs are chosen from the set $\mathbb{D}_t$. The samples in the set $\mathbb{D}_t^i$ are arranged in ascending order. The record corresponding to an arbitrary percentage such as 95% is selected from the set $\mathbb{D}_t^i$ as the upper and lower boundaries.

III. Case Study

In the case study, EV charging demand is forecasted for a continuous period of 30 days in a day-ahead manner. Fig. 6 shows the forecasting result on 21st of August, which has the best PI coverage. The X-axis indicates the time instances for every 15 minutes. There are 96 time instances in 24 hours. As shown in Fig. 6, the long-term components are forecasted properly. On the contrary, the fluctuation of short-term components is not forecasted well, but the PI is designed to cover the fluctuation. The PI coverage rate is 99% in the day and only one observed data point is outside of the PI. Fig. 7 shows the worst PI coverage day, the 15th of August. From 0 to 40, which is equivalent to 12 am to 10 am, the true (observed) data are markedly outside of the PI. The observed EV demand on the worst day has a different propensity from the trained past data, such that even the PI cannot cover the observed data perfectly. The PI coverage rate for the worst day is 68%.

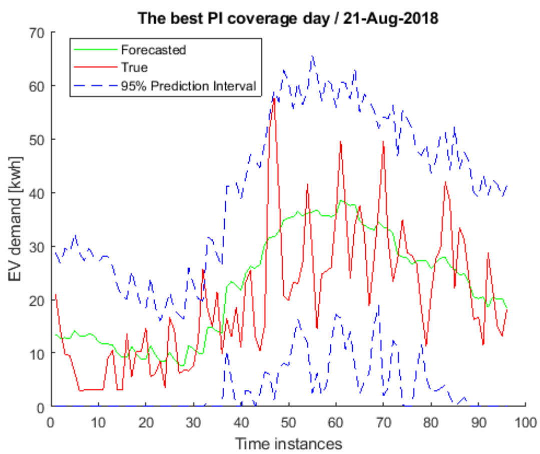

Fig. 6. Forecasting result on the day showing the best prediction interval coverage rate

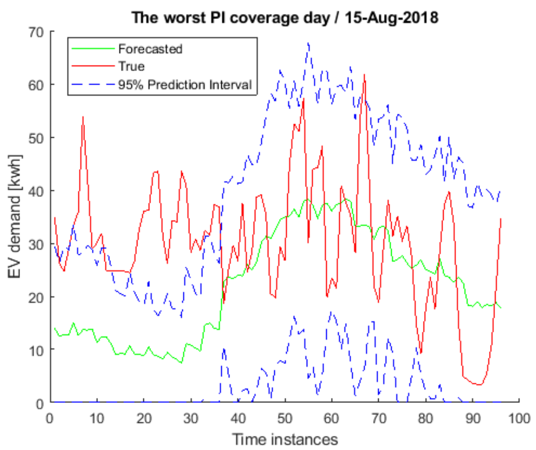

Fig. 7. Forecasting result on the day showing the worst prediction interval coverage rate

Fig. 8 shows the PI coverage rate during the forecasted 30 days. Of this period, 27 days show more than 80% PI coverage rate. Fig. 9 shows the root mean square error (RMSE) during the forecasted 30 days. All the days have more than 5 kWh residual error. This certain residual error is derived from the short-term fluctuations, which is observed even on the best day in terms of RMSE.

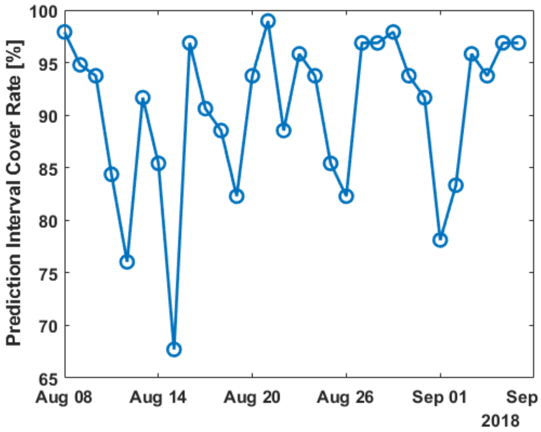

Fig. 8. Prediction interval coverage rate during the forecasting period

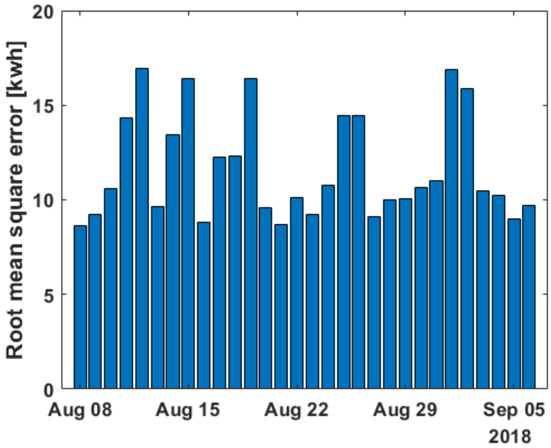

Fig. 9. Root mean square error during the forecasted 30 days

IV. Conclusion

EV charging demand forecasting for buildings or in a local area is essential to control the distribution network load and voltage. EV charging demand is erratic; therefore, probabilistic forecasting with a prediction interval is more flexible than a deterministic method for cooperation with other resources. In this paper, an ensemble model for probabilistic forecasting is proposed. The proposed model is validated with respect to an actual dataset that does not follow the normal distribution.

References

[1] B. Till *et al.*, "Global EV Outlook 2019 to electric mobility," *OECD iea.org*, p. 232, 2019.

[2] E. Veldman and R. A. Verzijlbergh, "Distribution grid impacts of smart electric vehicle charging from different perspectives," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 333–342, 2015.

[3] K. Mahmud, M. J. Hossain, and G. E. Town, "Peak-Load Reduction by Coordinated Response of Photovoltaics, Battery Storage, and Electric Vehicles," *IEEE Access*, vol. 6, pp. 29353–29365, 2018.

[4] G. Zhang, S. T. Tan, and G. Gary Wang, "Real-Time Smart Charging of Electric Vehicles for Demand Charge Reduction at Non-Residential Sites," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4027–4037, 2018.

[5] M. Pertl, F. Carducci, M. Tabone, M. Marinelli, S. Kiliccote, and E. C. Kara, "An Equivalent Time-Variant Storage Model to Harness EV Flexibility: Forecast and Aggregation," *IEEE Trans. Ind. Informatics*, vol. 15, no. 4, pp. 1899–1910, 2019.

[6] J. Zhao *et al.*, "Distribution Network Electric Vehicle Hosting Capacity Maximization: A Chargeable Region Optimization Model," *IET Gener. Transm. Distrib.*, vol. 12, no. 5, pp. 4407–4418, 2017.

[7] S. Pirouzi, J. Aghaei, M. A. Latify, G. R. Yousefi, and G. Mokryani, "A robust optimization approach for active and reactive power management in smart distribution networks using electric vehicles," *IEEE Syst. J.*, vol. 12, no. 3, pp. 2699–2710, 2018.

[8] S. Rezaee, E. Farjah, and B. Khorramdel, "Probabilistic analysis of plug-in electric vehicles impact on electrical grid through homes and parking lots," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 1024–1033, 2013.

[9] C. Gómez-Quiles *et al.*, "A novel ensemble method for electric vehicle power consumption forecasting: Application to the Spanish system," *IEEE Access*, vol. 7, pp. 120840–120856, 2019.

[10] T. Yang, X. Xu, Q. Guo, L. Zhang, and H. Sun, "EV charging behaviour analysis and modelling based on mobile crowdsensing data," *IET Gener. Transm. Distrib.*, vol. 11, no. 7, pp. 1683–1691, 2017.

[11] M. S. Islam, N. Mithulananthan, and D. Q. Hung, "A Day-Ahead Forecasting Model for Probabilistic EV Charging Loads at Business Premises," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 741–753, 2018.

[12] X. Zhang, K. W. Chan, H. Li, H. Wang, J. Qiu, and G. Wang, "Deep-Learning-Based Probabilistic Forecasting of Electric Vehicle Charging Load With a Novel Queuing Model," *IEEE Trans. Cybern.*, pp. 1–14, 2020.

[13] Q. Dai, T. Cai, S. Duan, and F. Zhao, "Stochastic modeling and forecasting of load demand for electric bus battery-swap station," *IEEE Trans. Power Deliv.*, vol. 29, no. 4, pp. 1909–1917, 2014.

[14] "daisukekodaira/EV-forecasting." https://github.com/daisukekodaira/EV-forecasting (accessed Jul. 06, 2020).

[15] "Electric Vehicle Charging Sessions Dundee." https://data.dundeecity.gov.uk/dataset/ev-charging-data (accessed Jun. 28, 2020).

[16] "k-means clustering - MATLAB kmeans." https://www.mathworks.com/help/stats/kmeans.html?s_tid=srchtitle (accessed Jul. 04, 2020).

[17] "Function fitting neural network - MATLAB fitnet." https://www.mathworks.com/help/deeplearning/ref/fitnet.html (accessed Jul. 04, 2020).

[18] D. Kodaira, W. Jung, and S. Han, "Optimal Energy Storage System Operation for Peak Reduction in a Distribution Network Using a Prediction Interval," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2208–2217, 2020.