# eCAL 02132018 seminar LIT review

Teng Zeng

February 2018

## 1 A data-driven approach for characterizing the charging demand of electric vehicles: A UK case study

**Objective:**

- Due to the increasing number of electric vehicles, the impact of their charging on distribution network is being investigated using different load profiles. However, the majority of these load impact studies are mostly based on assumptions of probability distributions of electric vehicle charging demand profiles. In this paper, real EVs charging demand data were obtained from Plugged-in Midlands (PiM) project, one of the eight "Plugged-in Places" projects supported by the UK Office for Low Emission vehicles (OLEV). A data preprocessing and data mining model were developed to investigate the characteristics of EV charging demand in a geographical area. Then Fuzzy-Based model aggregates these characteristics and estimates the potential relative risk level of EVs charging demand among different geographical areas independently to their actual corresponding distribution network.

**Highlight:**

- 21.918 charging events from 255 different charging stations in UK were analyzed.

- charging event data:

  • connection time, disconnection time, energy drawn, user (unique ID), charging station

- charging station data:

  • charging station ID, lat, long, detailed addresses (road, post code, county), location category (public/private..), location sub-category (on street/ public parking garage..), ownership (dealership, hotel, train station..), host (name of the charging station host), NCR (whether the station is registered on the National Charging Registry, NCR of UK), manufacturer, supplier (operator of the station), charger type (power rate), Connector1/2 (socket pin type, e.g. 3 pin, 5 pin..), mounting type (ground, wall..)

- weather data:

  • max air temperature, min air temperature, mean air temperature, mean wind speed, max gust, rainfall, daily global radiation, daily sunny hours

- A data pre-processing methodology for dealing with EVs charging data was presented.

- A data mining model was developed to analyze and extract useful information from the data.

- A fuzzy logic decision model was developed to characterize the EVs charging demand.

**Result and output:**

- Risk Level index. The Fuzzy Based Characterization Model uses the outputs of the Clustering, Correlation, and Regression modules to calculate the index.

- Three geographical areas are studied:

Leicestershire: 34.1
Nottinghamshire: 34.8
West Midlands: 23.5

**Methodology:**

- Data Cleansing: EV dataset and weather data set were cleaned, removing missing and incorrect values, zero/negative energy were removed. Duplicated data entries were also removed.

- Data Mining:

- Clustering Module:

  Motivation: The most representative cluster centroid was used to create the typical daily EVs charging demand of an area.

  1. K-means clustering method (Davies-Boudin evaluation criterion was used to calculate the number of k cluster.

  2. Having the daily charging demand profile of an area, an index $\lambda$ was defined to express the proportion of EVs charging demand during peak hours (17:00 - 20:00)

- Correlation Module:

  Motivation: Weather affects road traffic congestion and the driving behavior of car owners.

  1. Pearson Correlation Coefficient (r) to measure the correlation between the weather attribute values and the daily peak power of EVs charging demand in a geographical area.

- Regression Module:

  Motivation: Investigate the monthly change of the EVs charging damand.

  1. Growth Ratio index is the ratio between the growth rate of EVs charging demand and the average monthly EVs charging demand (Y in kWh). X in months.

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{1}$$

$$GR = \frac{\beta_1}{E_{\text{month}}} \cdot 100\% \tag{2}$$

- Fuzzy (fuzzy-logic) Based Characterization Model:

Motivation: Characterize the EVs charging demand of a geographical area accodring to the information about the shape of the typical daily profile ($\lambda$ index), the predictability with respect to weather (r) and the trend of EVs charging demand (GR index). To this end, "risk level" was defined to aggregate the potential underlying risks from these characteristics.

Assumptions:

i. Magnitude and duration of the peak of the typical EVs charging demand profile ($\lambda$) are undelrying risk factors for the distribution network, as they affect the transformer/circuit loading and the voltage profile

ii. The change over time of EVs charging demand (GR index) affects the long term decision regarding the planning of the network reinforcement.

iii. The predictability of EVs charging demand with respect to weather in a geographical area (coefficient r) affects the accuracy of a forecasting model.

iv. The demand analysis will affect the mid-term normal operation of the distribution network of the area.

v.The risk index is not defined in a absolute term but is used to classify relatively the level of these risks among different geographical areas independently to their actual corresponding distribution networks.

**Take Away:**

- The clustering method to categorize different EVs charging demand profile.
- Pearson Correlation coefficient for weather impacts on the load.
- This paper focuses more on the risk analysis for district scale power system operation. What it does not do is a precise daily load demand profile generation.

# 2 Electric vehicle charging demand forecasting model based on big data technologies

**Objective:**

- This paper presents a forecasting model to estimate electric vehicle charging demand with historical traffic data and weather data of South Korea. Case studies for EVs during weekdays and weekends in summer and winter were presented to show the different charging load profiles of EVs in the residential and commercial sites. The proposed EV charging demand model can be the foundation for the research on the impact of charging EVs on the power system.

**Highlight:**

- The forecasting model uses historical real-world traffic data and weather data.

• traffice data: The data is collected by the Traffic Monitoring System (TMS) of the Ministry of Land, Infrastructure and Transport (MOLIT). The data includes daily traffic volumes and vehicle mileages by vehicle types collected in the highway, national route, and local roads of South Korea in every hour for a year.

Assumption: EV traffic patterns will be identical with the real traffic data of conventional vehicle.

• weather data: temperature, humidity, wind speed, and day types

- A battery charging starting time is determined by real-world traffic patterns.

• Assumes that the charging start time follows a Gaussian distribution (many real-world phenomena are normally distributed when their samples are large enough).

- The presented model considers charging demand for both electric cars and buses.

- Considered variables and characteristics: battery charging starting time determined by the traffic data, initial SOC (also Gaussian distribution, $\in [0.2, 0.8]$), type of battery, charging characteristics with different charging power classifications (L1 at work 3.7kW, L1 at work 3.7kW and L3 fast charging 115kW).

  Assumption: A car used a lithium-ion battery with a range of 148 km and a capacity of 27 kWh modeled by Kia Motors. A bus used a lithium-ion polymer battery, 85.9 kWh and a range of 69.8 km modeled by Hankuk Fiber E-Primus bus.

- The presented model considers both slow and fast charging classifications.

**Result and output:**

- The example case studies to anticipate EV charging demand in the residential and commercial sites on weedays and weekends in winter and summer sessions were presented. High EV charging demand in the residential sites was observed during the night weekends, because all cars were charged at home on weekends. In the commercial sites, high charging demand was observed during the non-operation hours due to its lesser charging start time interval in this period than that in the other charging periods.

- Allow utility operators to plan the operation and generation profiles in the residential and commercial sites.
- Contribute to decide investment and operation plans for adaptive EV charging infrastructures depending on EV charging demand.

**Methodology:**

- Charging Demand Foresting Model:

- Cluster Analysis:

  Motivation: Traffic volume consists of different traffic patterns resulted from various factors such as weather and day types. A hierarchical clustering technique was used to classify historical traffic data into several patterns.

  Four types of cars traffic pattern and two types of buses traffic patter are identified:

  1. Car cluster 1 is composed of event days such as festivals and international conferences.
  2. Car cluster 2 contains weekdays and most Saturdays.
  3. Car cluster 3 contains holiday.
  4. Car cluster 4 is mostly composed of Sunday.
  5. Bus cluster 1 contains most weekends.
  6. Bus cluster 2 contains most weekdays.

- Relational Analysis:

  Motivation: A grey relational analysis was used to determine important factors that have a significant influence on the traffic patterns.

  1. Data are pre-processed and normalized to a comparable time-series. In this study, the traffic volume was used as the reference series, the factors affecting the traffic volume were used for the comparative series. Then a grey relational coefficient is calculated between the two series.

  2. An averaged value of all the grey relational coefficient is calculated and named grey relational grade which was used to indicate the degree of the factor's influence on the traffic volume.

3. Influential Factors:
    1. Maximum temperature
    2. Average temperature
    3. average humidity
    4. average wind speed
    5. day type

- Decision Tree:

  Motivation: To establish the relationship between the formed clusters of traffic patterns and influential factors.

  1. The data of the influential factors (above fives) determine the specific car and bus clusters in which the forecasting day belongs.
  2. Once the clusters are obtained, charging demand can be estimated.

- Determination of a Charging Start Time:

  Assumptions:

  1. Electric car charged once a day. Either charged daytime at work or nighttime at home.
  2. Electric bus charged two to three times.
  3. Gaussian pdf of the charging time
  4.

     Mean charging time:

     1. Because the morning highest peak traffic volume occurred in the morning rush hour and the minimum traffic volume in the daytime happened at the end of morning rush hour, the midpoint time of these two traffic volume can be considered as the mean charging starting time in the daytime.
     2. Because the evening highest peak traffic occurred in the evening rush hour, the time in which the traffic volume decreased by the half of the peak was considered as the mean charging starting time in the evening.

     Standard deviation:

     1. Typical charging starting time interval was used to determine the sd that is defined by the half of its starting time interval.
     2. This paper assumed that the typical time interval for electric cars is 2hr in the daytime because a typical electric car is charged once its user arrives at the workplace.
     3. This paper assumed that the typical time interval for electric cars is 4hr in the nighttime because EV driver might not go home immediately after work.

**Take Away:**

- Apply decision tree method to determine the relationship between influential factors and the traffic volume/charging pattern.
- Too much assumptions made about EV charging, including starting time and initial SOC.
- No benchmark comparison as how well the model performs.

# 3 Modified Pattern Sequence-based Forecasting for Electric Vehicle Charging Stations

**Objective:**

- Three algorithm for the forecasting of energy consumption at individual EV charging outlets have been applied to real world data from the UCLA campus.

**Highlights:**

- k-Nearest Neighbor (k=1)
- ARIMA
- Pattern Sequence Forecasting
- Proposed new method, namely the Modified Pattern Sequence Forecasting. It could be interpreted as NN on integer valued data or as PSF with considering only the most recent neighbor to produce the output.
- Varies of effective days data (over 60 effective days) are being used for the total 15 charging outlets.
- First 90% of the data which makes the test set last 10%. Minimum training dtat was 30% of the training data (30% of 90%=27% of the whole) and validation data consist 70% of the training data.
- The only pre-processing was to force energy records that were mistakenly recorded as more than the physical maximum of the charging device ($E_{max}$) and lest than zero to the interval of [0, $E_{max}$]

**Result and output:**

- kNN (k=1, which is just nearest neighbor) has the least SMAPE (symmetric mean absolute percentage error).
- PSF is more robust towards noise by substituting the real valued time series with an integer valued one.
- MPSF has much better forecasting result in terms of the SMAPE. The mean SMAPE for PSF is about 43.29% whereas for MPSF, it is only 13.78%.
- The results were not sensitive to more or less minimum training data. Used 5 blocks in cross validation.

**Methodology:**

- K-Nearest Neighbor:

  - Based on KNN algorithm, each sample (training, test or validation) is composed of input and output pairs. In the application, the output is the energy consumption for the next 24 hours and the input is the concatenation of the consumption records for up to D prior days. The concatenation repeats for all days, if there are N days in the data set, there will be N-D+1 of these input-output pairs. The total number of data is n = 24N.

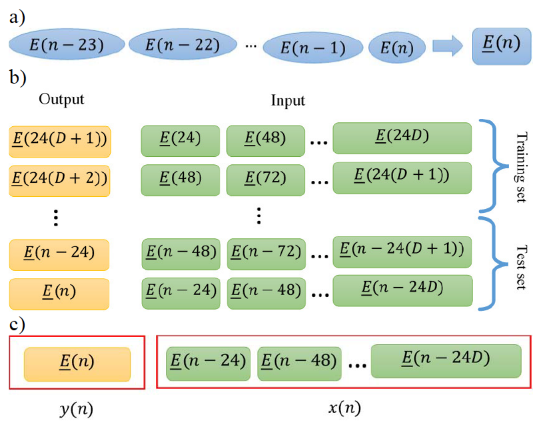
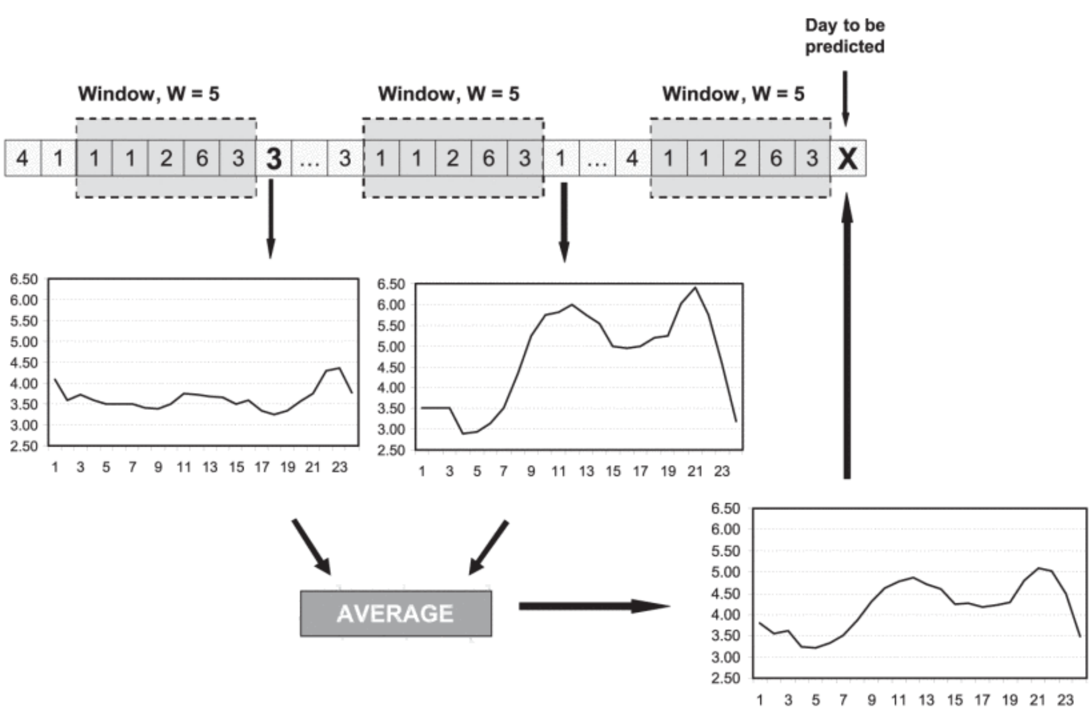



Figure 1. a) energy consumption vector (E) for 24 hours , b) input-output pairs and division of data into training and test sets, c) labeling inputs as x and outputs as y.

- ARIMA:

  • The Auto Regressive (AR) portion models the contribution of the past values of the variable, while the Moving Average (MA) portion models the contribution of noise terms. The Integrated (I) portion models the number of differences needed in order to transform the time series to a stationary time series. ARIMA (p,d,q) where p, d, q are the order of the AR, I, MA terms. The model forecast each hour based on the previously forecast hour(s) and the past actual values.

- Pattern Sequence-based Forecasting (PSF):




Citation: F. Martínez-Álvarez, A. Troncoso, J. C. Riquelme, and J. S. Aguilar- Ruiz, “Energy time series forecasting based on pattern sequence similarity”, IEEE Trans. Knowl. Data Eng., vol. 23, pp. 1230-1243, August 2011

• The idea is based on assigning each 24 hours set, i.e., a day, to a cluster and then the forecast is based on the cluster labels rather than actual values in each day. By clustering, the dimension of each day reduces to one (one label of the day) instead of 24. It also adds the robustness by substituting real values (e.g. power consumption) with integer numbers (cluster labels).

• Optimum number of clusters is found through the validity index. The silhouette index, SI. The number of clusters that maximizes the silhouette index is selected as optimum.

$$sil(i) = \frac{a(i) - b(i)}{\max\{a(i), b(i)\}} \quad (5),$$

where

$$a(i) = \frac{1}{n_c - 1} \sum_{j \neq i, j \in C} d(i,j)$$
$$b(i) = \frac{1}{N_{tr} - n_c} \sum_{j \notin C} d(i,j) \quad (6),$$

- Modified Pattern Sequence-based Forecasting (MPSF):

• For PSF, the silhouette index might give ill conditioned measurement when the data is really sparse. The nubmer of clusters that maximizes the SI might be only two (one for zero-days and one for non-zero days). So start k equal to 10% of distinct days to avoid degenerate clustering.

• The algorithm fails when there is no matching case even when the sequence length is 1. So the modified version is to assign the unmatched case with the most popular cluster.

- Modified Last Block Cross-validation:

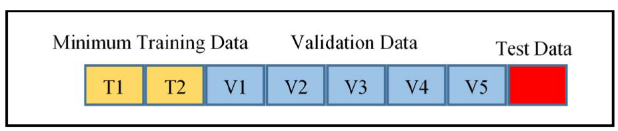



Figure 4. Modified blocked cross vailadtion. Training data is divided to minimum training data {T1,T2} and validation data {V1,…V5}. Model is first trained on minimum training data {T1,T2} and evaluated on V1, then it is trained on {T1,T2,V1} and evaluated on V2, up until training on {T1,T2,V1,…,V4} and evaluating on V5.

Dealing with time-series data, if use k-fold CV, it is very likely to mess up the temporal relationship among the data.

**Take away:**

• The proposed MPSF method has really decreased the prediciton error.

# 4 Fast Prediction for Sparse Time Series: Demand Forecast of EV Charging Stations for Cell Phone Applications

**Methodology:**

- Historical average data (with depth of the averaging)
- K-NN
- Weighted k-NN. The weights are defined based on Dudanis' weights.

$$w_p = \frac{\text{dis}[k+1] - \text{dis}[p]}{\text{dis}[k+1] - \text{dis}[1]}, \quad p = 1, \ldots, k \tag{6}$$

$$\boldsymbol{y}(t_{s^*}) = \frac{1}{\sum_{q \in idx} w_q} \sum_{p \in idx} w_p \boldsymbol{y}(p). \tag{7}$$