Lessons Learnt from Mining Meter Data of Residential Consumers

Konstantinos Blazakis[1], Sima Davarzani[2], Georgios Stavrakakis[1], Ioana Pisica[2*]

Abstract

Tracking end-users' usage patterns can enable more accurate demand forecasting and the automation of demand response execution. Accordingly, more advanced applications, such as electricity market design, integration of distributed generation and theft detection can be developed. By employing data mining techniques on smart meter recordings, the suppliers can efficiently investigate the load patterns of consumers. This paper presents applications where data mining of energy usage can derive useful information. Higher demands, on one side, and the energy price increase on the other side, have caused serious issues with regards to electricity theft, especially among developing countries. This phenomenon leads to considerable operational losses within the electrical network. In order to identify illegal residential consumers, a new method of analysing and identifying electricity consumption patterns of consumers is proposed in this paper. Moreover, the importance of data mining for analysing the consumer's usage curves was investigated. This helps to determine the behaviour of end-users for demand response purposes and improve the reliability and security of the electricity network. Clustering load profiles for large scale energy datasets are discussed in detail.

Keywords

power theft, smart metering, data mining, non-technical losses, load profile, electric load clustering

[1] School of Electronic and Computer Engineering, Technical University of Crete, Greece

[2] College of Engineering, Design and Physical Sciences, Brunel University London

[*] Corresponding author, e-mail: Ioana.Pisica@brunel.ac.uk

1 Introduction

Advanced metering infrastructure enables automated load management and execution of demand response which is mainly designed to reduce peak demand, carbon emissions, and cost, providing a great opportunity for electricity market participations, improving voltage efficiency and distribution network automation [1]. However, network operators and suppliers should not only design and implement appropriate DR programs, but they should also be well-equipped to control the power network at all times, anticipate the future events to apply optimal solutions, while bidding in the electricity market to gain the maximum economic benefits.

Knowledge discovery in databases (KDD) for analysing the behaviour of different types of end-users in consuming energy have become of great interest in recent energy-related research. The evolution of smart meters is shifting the management of demand from supply-focused to supply & consumers-focused [2] which provides a great opportunity to activate demand response programs.

Having more knowledge about different customer's behaviour in how and when they use the electricity can help the network operators and energy suppliers for more correctly action in a liberalized and competitive electricity market in a dynamic form [3].

In recent years, various methods of electricity load data characterization have been introduced and applied. The design of such methods depends on the motivation of using that data and the volumes of data available [4-6].

Moreover, with the information extracted from smart meters, the power network would be able to cluster electricity consumers with monitoring their energy usages and data mining in their load profiles patterns during a determined time period. Such of these segmentations provides various higher-level applications such more accurate load forecasting, theft detection, determining required demand reduction and specific demand response programs etc.

The parameters that can affect residential usage patterns can be examined by identifying typical load profiles for each cluster of different dwellings. Comparing the different demand pattern can help utilities to design and apply demand response

programmes and tariffs schemes. Such demand response programmes generally imply load shifting or load shedding commands, aimed to increase energy efficiency and to reduce overall demand usage and costs.

The high level of losses in the entire electrical network can indicate non-technical losses due to electricity theft by illegal consumers mainly in distribution networks. Therefore, detection of illegal consumers is an extremely challenging problem.

With the growth of data in power system, data mining techniques have been developed dramatically in the past years. Using data mining in order to characterize the load profile provide valuable information on energy usage and controlling the demand. Basically, load profiles model different usage patterns which relate to various factors such as weather, time and consumers characteristics. Having the knowledge of where and when the energy is used, is an essential task to plan the power network and to detect any issues or illegal usage.

The reminder of the paper is structured as follows: Section 2 explains briefly the theft detection by data mining techniques including the used scenarios, the studied model, the simulation procedure and the result obtained from simulation analysis; Section 3 introduces the role of data mining in demand side management and presents some results of consumers load profile clustering; Section 4 presents the main conclusions of the work and suggests future developments.

2 Mining Data For Electricity Theft Detection

In general, operational losses can be divided into the two categories of technical and non-technical losses. The non-technical losses are due to electricity theft from illegal activities of consumers and are addressed in this section.

This paper presents a new method that uses energy consumption patterns to detect illegal consumers in a smart grid environment. With smart meters, real-time energy consumption information is collected from consumers several times per day. However, owing to the privacy and confidentiality of consumption data, a dataset with near real-time energy consumption patterns has been developed in this work. Then, a new clustering algorithm based on mean shift is proposed and implemented, which maps residential energy consumption patterns' into legal and illegal.

The most common and simplest way of electricity power theft is tapping energy directly from an overhead distribution feeder [7]. Factors that influence consumers to steal electricity depend upon various local parameters that fall into multiple categories, the most important of which is socio-economic [8-9].

The most important factors are:
- High electricity prices in contrast with income.
- The belief that it is dishonest to steal something from a neighbour, but not from a utility.
- Tax Purposes. In some countries, tax levels are higher for houses connected to electricity supply.
- Industrial consumers wishing to hide overall production.

The operational targets of utilities are generally to maintain good power factor and voltage profile and sufficient reactive power along the feeders. These targets may become difficult to achieve due to power theft, as un-metered energy flows through the networks. The most important consequences of power theft are:
- Electricity theft overloads the generation unit and decreases the frequency of the electricity network.
- The unpredictable amount of additional load may lead to brown outs and black outs during the peak load period.
- Utility companies expect to earn their money back from the customers for the service provided.
- Reactive power compensation is very difficult without complete information about the total load flow because of the theft.
- Utility companies are forced to impose excessive tariff on genuine customers as they cannot afford all these losses by themselves.
- Illegal tapping of electricity raises safety concerns like electric shocks and even the death of a person who operates it.

2.1 Power Theft Scenarios

Table 1 presents the power theft scenarios that we used for our experiments. We must mention for power theft scenario 3 that the consumer may not steal power from the electricity provider but steal tax from the state due to an unauthorized building.

Table 1 Power theft scenarios

Power Theft scenarios
1. Consumer with smart meter and steeling electricity fully due to power pass before smart meter.
2. Consumer with smart meter and steeling electricity partially due to power pass before smart meter.
3. Consumer with smart meter and increased consumption of electricity due to illegal activity or power delivery to unauthorized building.

For our experiments we consider two residential consumers with power theft scenario 1, three residential consumers with power theft scenario 2 and one residential consumer with power theft scenario 3.

2.2 Experimental model

The distribution line is presented in Fig. 1, where with black arrows are residential consumers and with blue arrows are commercial consumers. On the top of the distribution line there exists a sum meter for counting the total energy consumption. The distribution cables are of ACSR 50 R: 0.404 Ω/KM and X: 0.420 Ω/KM.
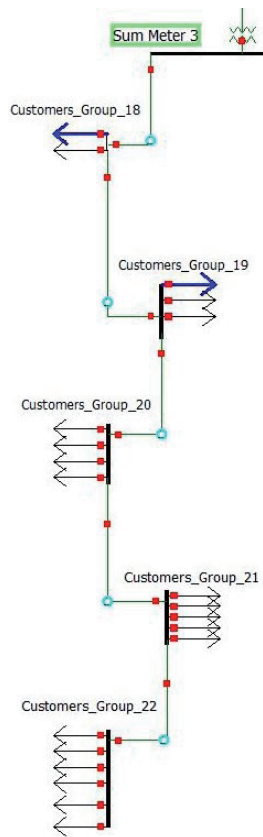
Fig. 1 Distribution line

2.1 Residential Patterns

In Fig. 2 and Fig. 3 we show residential patterns without and with power theft for power theft scenario 1 and power theft scenario 2. The data was examined in a period of 183 days from 1st April until 30th September 2015.
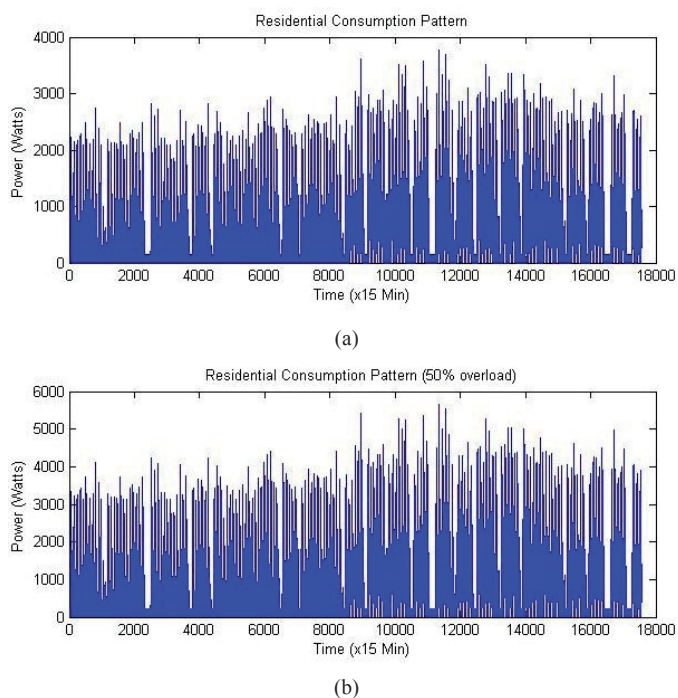

(a)

(b)

Fig. 2 Residential Consumption Pattern without (a) and with (b) power theft (overload)
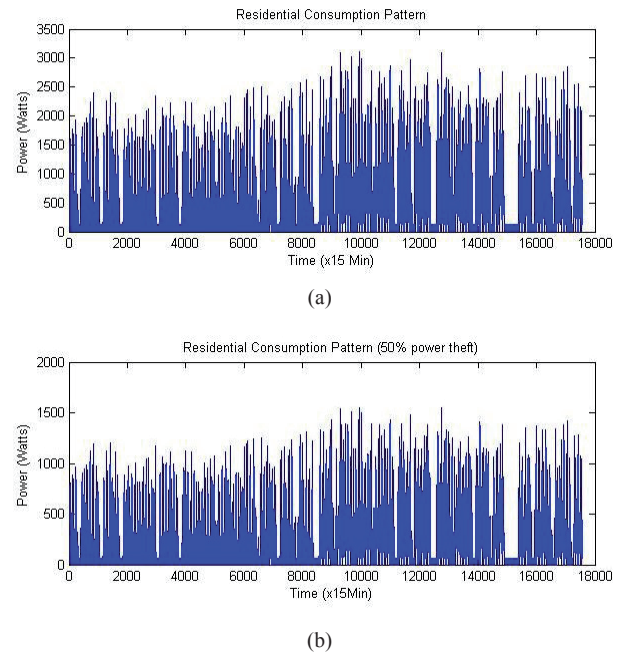

(a)

(b)

Fig. 3 Residential Consumption Pattern without (a) and with (b) power theft

2.2 Simulation procedure

Running our model we receive results as shown in Fig. 4. Subtracting from sum meter readings Fig. 4 (a) the smart meter readings of consumers corresponds to the line we estimated the total losses as shown in Fig. 4 (b). From total losses we subtract the technical losses (Fig. 4 (c)) as the software estimated and we ends up to the last diagram (Fig. 4 (d)) which shows the non-technical losses.

Next step is to separate commercial and residential consumers. From the residential consumers we throw out the consumers with zero consumption (power theft scenario 1 or uninhabitable house).

For residential consumers we keep smart meter readings only for weekdays and we throw out weekends due to the different habits that residents have especially on weekends. In this way we enhance the clustering procedure. Subsequently, before applying PCA (principal component analysis) it is standard practice to first perform mean normalization at feature scaling so that the features have zero mean and should have comparable range of values.

Having a matrix with residential consumers we apply PCA to our data so as to minimize the dimension of the data. Applying PCA we choose the appropriate number of principal components so as to keep the data variance to 95%.

2.3 Results and analysis

Applying mean shift to our processed data we receive the following results as shown in Fig. 5. The points P_2, P_3, P_4 corresponds to the three residential consumers with power theft scenario 2 and the point P_1 corresponds to one residential consumer with power theft scenario 3. Experimenting by

changing the number of power theft scenarios we have reached the conclusion for residential consumers with partial power theft up to 50% (power theft scenario 2) and for residential consumers with up to 40 % overload (power theft scenario 3) that we have 100% detection.
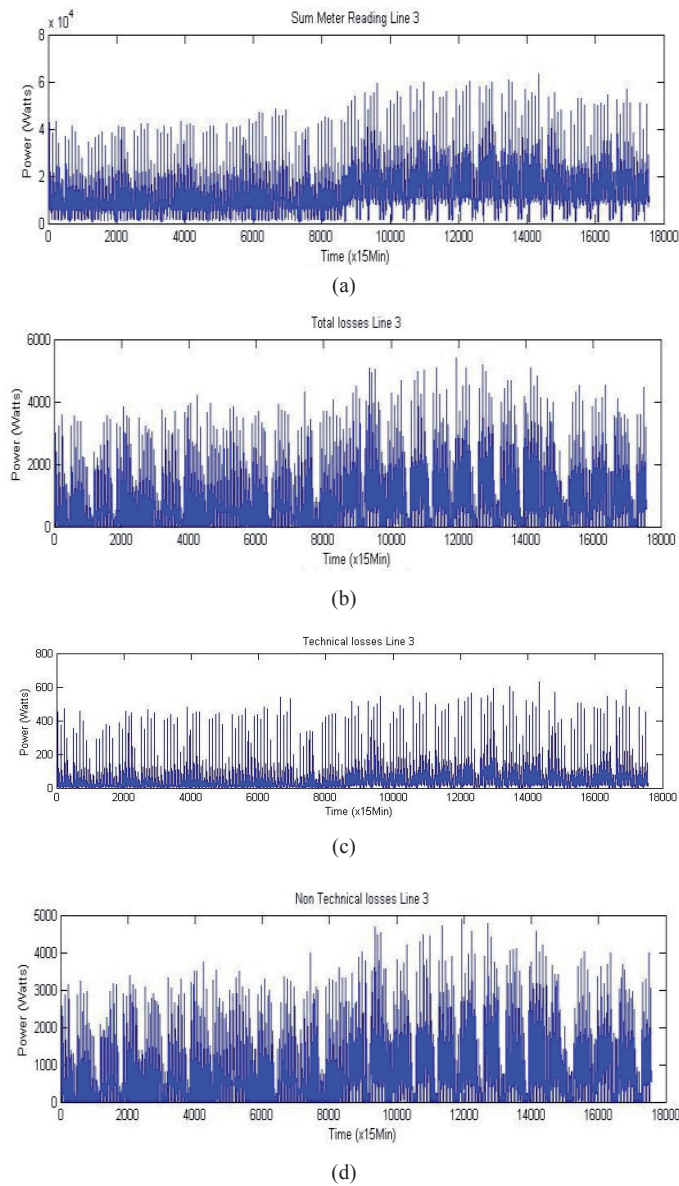

(a)

(b)

(c)
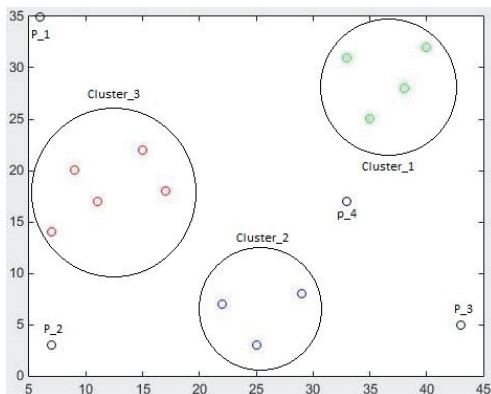

(d)

Fig. 4 Distribution Line Results

Fig. 5 Experimental Results

For the loads with zero consumption we check the permanent house of the residence. If the house that we examine is a permanent residence then there is a strong indication for power theft.

For commercial consumers, due to businesses are different from each other we propose to check them separately for those which are in problematic lines.

At last it is worth mentioning that the Cluster_1, Cluster_2 and Cluster_3 in Fig. 5 contains the rest residential consumers from distribution line. The categorization of the rest consumers in three clusters has to do with common habits that they have.

3 Data Mining For Assessing Demand Response Capacity

Various techniques have been proposed for data mining in load profiles characterizations. Such techniques enhance the accuracy of demand forecasting, which improves the power market design, and demand management.

One of the common methods for electrical load data characterization is using the historical data [5] of typical consumer's usage patterns which has been derived from electric load profiles over different period of time as daily, weekly, monthly or annually. Longer period of averaged load data creates less number of peaks in the load profile. The advantage of load profile characterization is that, it provides a good means of supply and demand comparison which allows the estimation for the level of coverage by different local suppliers and DGs.

However, in terms of domestic consumers, load profile analysis can be affected intensively by divergent factors mainly categorized to internal (occupancy features, social-economic factors, household features, etc.) and external (geographical location, weather condition, etc.) which are linked together in a complicated and close way.

3.1 Dataset and Pre-processing

The case study has been derived from ISSDA (Irish social science data archive) which was released on June 2013 by the Research Perspective Ltd on behalf of The Commission for Energy Regulation [10]. The selected dataset comprises recordings from 831 residential households in Ireland during 1st Jan. 2010 to 31st Dec. 2010.

The electricity consumption information was recorded at 30 minute intervals.

Losses of data are associated with losses of information, which can distort the results. The reduction of missing values for each consumer can be done by applying the means of existing values while examining the behaviour of the usage during the lost period of relevant time. For minimizing the percentage of errors in data reading due to inconsistent readings, meter faults, loss of supply or other interruptions, we run a pre-processing stage before any analysis or working on the data. It would enhance accuracy and quality of the result by replacing or removing any missed or error signals.

This beginning step which can be called the clearing and cleaning data phase consists of two pre-process filters removing the following data from dataset:

the negative and zero values

Dwellings whose readings had five or fewer different values for at least half of the total readings.

The daily load profile of selected data corresponds to the averaged data with thirty minute resolution during a day aggregating all working days of a specific dwelling. However, after the filtering process, some of the households did not have sufficient readings to be considered representative. Therefore, only load profiles that present values for at least 1438 of the 1440 min were used subsequently.

3.2 Creating Representative Load Profiles

Following the pre-processing and data cleaning steps, a demand curve for each individual household was created based on the following categories to enable data clustering and analysing based on time-varying factors:

- Average Load profile of each domestic per day
- Average load profile of each domestic per year

Figure 6 shows load profile of a typical customer per a winter week-day in 2009. The simulation and programming were performed in Matlab. For each category, the average load curves of each customer are normalized and transformed using the respective minimum and maximum loads of the period under study [11], obtaining Integer values between zero and one.
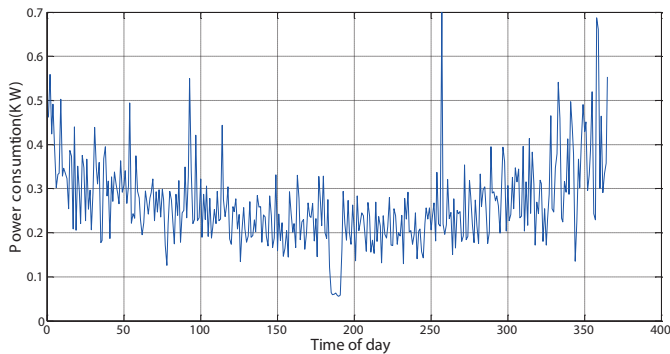

Fig. 6 Load profile of a typical household in a winter day

3.3 Clustering load profiles

In order to cluster similar households in terms of consumption data for each time-varying category, a typical representative load diagram for each domestic, was selected by averaging individual load profiles of all load curve during the period of each desired category.

The method used for clustering is based on Euclidean distance by K-means algorithm. The results are obtained for 8 clusters which are shown in Fig. 7. Each of the eight clusters reveals distinct behaviour. Of the 831 complete profiles, the number of households in each category is 150, 35, 99, 108, 146, 109, 30 and 154 respectively.

The centroid of each cluster is plotted in black marks from each of these load profile. A general and fundamental characteristic of power usage can be achieved by examining the centroid of each group. As we can see, each cluster presents a different load profile per year than the other group. This can be due to various factors and metadata is available such as levels of income, education, the awareness of different tariffs, facilities of installing smart meters, energy management systems, distributed generation, electrical vehicles etc.

For example, cluster 2 is representative for households with higher educational and social levels. The evening peak is very prominent and usage levels are high. Therefore, these groups have a higher potential for demand reduction and are an ideal group for participating in demand response programs. However, the other households can also benefit from demand reduction programs by employing time of use tariffs.

4 Conclusions

Data mining and clustering load profiles of customers enable demand response programmes on large scale. Management of energy consumption by tracking demand pattern in either power network or end-users will decrease the energy and therefore the cost for each group. It also helps to improve the reliability, security and operation of power network.

Power theft is a widespread phenomenon in developing countries, but energy-theft detection is a difficult problem in power grids. With the development of advanced metering infrastructure in smart grid, the energy theft identification has become more achievable and many new technologies are adopted in order to solve this problem. A novel methodology was proposed for automated detection of illegal use of electricity in the low voltage distribution networks. Smart meters are essential for this method.

The information extracted from mining the load profile provides a better understanding of usage pattern for both consumers and suppliers. Investigating the residential electricity load makes the power network operation more efficient. Data mining is generally used for analysing large scale of dataset which needs to apply clustering techniques for optimal load management. Each cluster represents a specific type of demand curve based on consumers' characteristics.
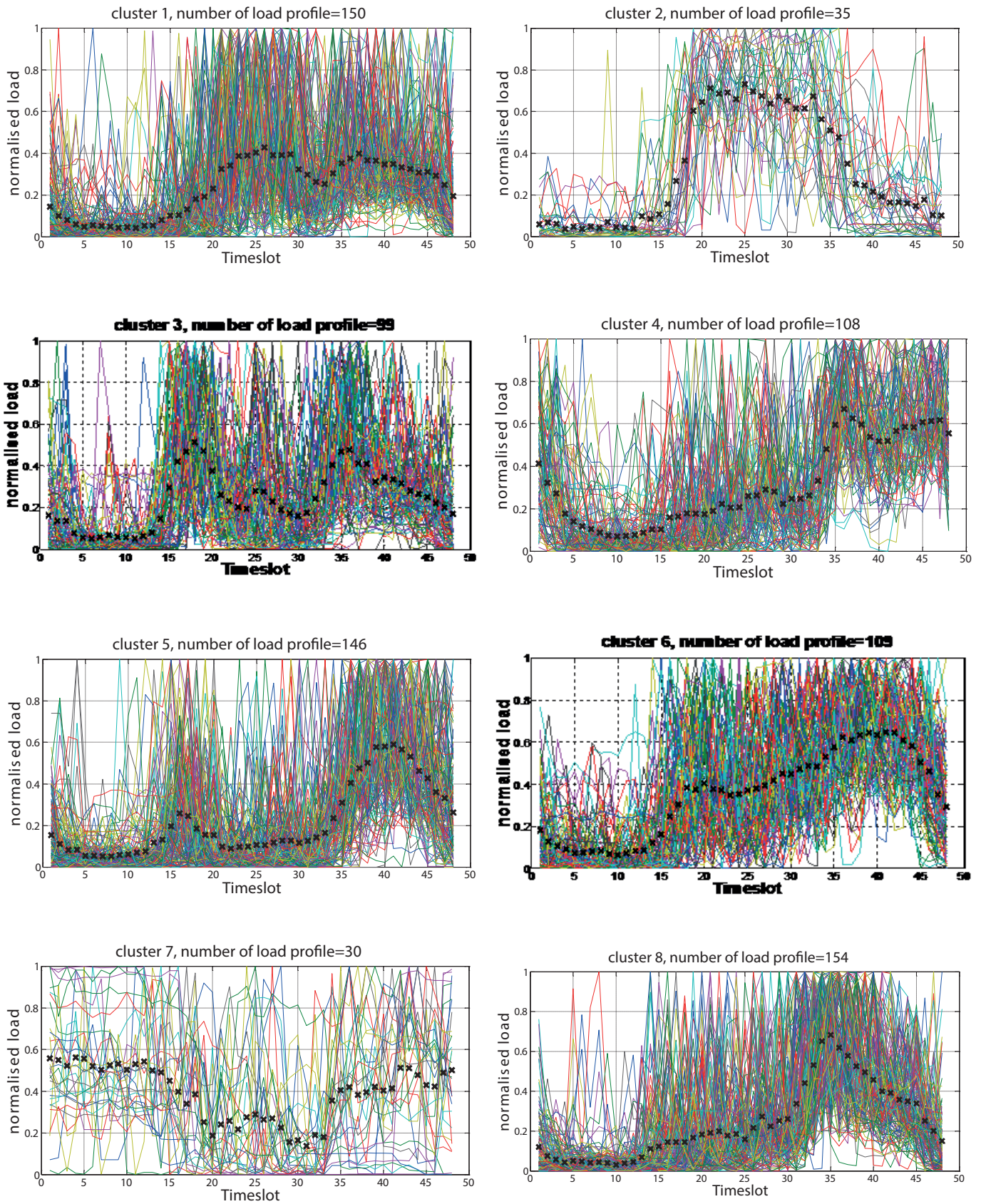
Fig. 7 Clustering load profile with K-means algorithm for 8 clusters

References

[1]   "Assessment of demand response and advanced metering." Federal Energy Regulatory Commission, Department of Energy. 2012. URL: http://www.ferc.gov/legal/staff-reports/12-20-12-demand-response.pdf

[2]   McKenna, E., Thomson, M. "Demand response behaviour of domestic consumers with photovoltaic systems in the UK: an exploratory analysis of an internet discussion forum." *Energy, Sustainability and Society*, 4(13), 2014. DOI: 10.1186/2192-0567-4-13

[3]   Frederiks, E. R., Stenner, K., Hobman, E. V. "Household energy use: Applying behavioural economics to understand consumer decision-making and behavior." *Renewable and Sustainable Energy Reviews*. 41, pp. 1385-1397. 2015. DOI: 10.1016/j.rser.2014.09.026

[4]   Zhang, Q., Li, J. "Demand response in electricity markets: A review." In: 2012 9th International Conference on the European Energy Market. Florence, May 10-12, 2012 , pp. 1-8. DOI: 10.1109/EEM.2012.6254817

[5]   Marques, D. Z., de Almeida, K. A., de Deus, A. M., da Silva Paulo, A. R. G., da Silva Lima, W. "A comparative analysis of neural and fuzzy cluster techniques applied to the characterization of electric load in substations." In: Transmission and Distribution Conference and Exposition: Latin America, 2004 IEEE/PES. pp. 908-913.
DOI: 10.1109/TDC.2004.1432503

[6]   Ramos, S., Vale, Z., Santana, J., Duarte, J. "Data mining contributions to characterize MV consumers and to improve the suppliers-consumers settlements." In: Power Engineering Society General Meeting, 2007. IEEE. Tampa, FL, USA, June 24-28, 2007, pp. 1-8. DOI: 10.1109/PES.2007.385996

[7]   Saikiran, B., Hariharan, R. "Review of methods of power theft in Power System." *International Journal of Scientific & Engineering Research.* 5(11), pp. 276-280. 2014. URL: http://www.ijser.org/researchpaper%5CReview-of-methods-of-power-theft-in-Power-System.pdf

[8]   Depuru, S. S. "Modeling, detection, and prevention of electricity theft for enhanced performance and security of power grid." (Electronic Thesis or Dissertation). 2012. URL: https://etd.ohiolink.edu/!etd.send_file?accession=toledo1341522225&disposition=inline

[9]   Nagi, J. "An intelligent system for detection of non-technical losses in tenaga nasional berhad (tnb) Malaysia low voltage distribution network." Vasa. 2009. URL: http://people.idsia.ch/~nagi/thesis/mee_thesis.pdf

[10]   Electricity Smart Metering Customer Behaviour Trials Findings Report. Commission for Energy Regulation. URL: http://www.cer.ie/docs/000340/cer11080(a)(i).pdf

[11]   Tsekouras, G. J., Hatziargyriou, N. D., Dialynas, E. N. "Two-stage pattern recognition of load curves for classification of electricity customers." *IEEE Transactons on Power Systems*. 22(3), pp. 1120-1128. 2007. DOI: 10.1109/TPWRS.2007.901287