Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 142 (2017) 2236–2242

Energy Procedia

www.elsevier.com/locate/procedia

9th International Conference on Applied Energy, ICAE2017, 21-24 August 2017, Cardiff, UK

Smart meter data clustering using consumption indicators: responsibility factor and consumption variability

Maher Azaza*[a], Fredrik Wallin[a]

Future Energy Center
Department of Energy, Building and Environment
Mälardalens University, Västerås, Sweden

Abstract

The wide spread of smart metering roll out enables a better understanding of the consumer behavior and tailoring demand response DR programs to achieve cost-efficient energy savings. In the residential sector smart metering allows detailed readings of the power consumption in the form of large volumes time series that encodes relevant information for distribution network operators DNOs to manage in optimal ways low-voltage networks. Further, DNOs may leverage the smart meter data to identify customer group for energy efficiency programs and demand side response DSR (e.g., dynamic pricing schemes). In this paper, we outline the application of smart meter data mining to identify consumers who are more responsible for the peak system using responsibility factor and consumption variability. Identification of consumers having higher responsibility to the peak system may yield to better enhance energy reduction recommendations and enable more tailored dynamic pricing plans depending on the consumer's influence on the utility peak. Responsibility factor and consumption variance have been investigated as input features of the clustering algorithms. Two clustering techniques, hierarchical clustering and self-organising map SOM, have been used to study the resulting customer groups and to have an effective graphical visualization of the customer's cluster distribution on the input feature space.

Peer-review under responsibility of the scientific committee of the 9th International Conference on Applied Energy.

Keywords: Demand response, Smart metering, Consumers clustering, Responsibility factor, Variabiliy,

* Corresponding author. Tel.: +46 73-6621308.
E-mail address: maher.azaza@mdh.se.

1876-6102 © 2017 The Authors. Published by Elsevier Ltd.
Peer-review under responsibility of the scientific committee of the 9th International Conference on Applied Energy .
10.1016/j.egypro.2017.12.624

1. Introduction

Smart grids are specific application of the smart energy systems that [1], through digital information and communication technologies ICTs, adjust the electricity flow between suppliers and consumers. By collecting information on the state of the network, smart grids contribute to matching the production, the distribution and the consumption [1]. The electrical system, then becomes predictive, communicative and controllable. The aim is to make the power grids flexible and capable of integrating more finely the behaviors and actions of all connected users (producers, consumers, and users) in order to ensure synchronization between production and consumption at lower cost, while favoring the efficient integration clean technologies (renewable energies, electric vehicles, etc.).

At the consumer level, smart meter is an enabling technology of the smart grids. In recent years, advanced metering infrastructure (AMI) has been widely deployed enabling detailed measurements of the power consumption at a higher frequency [2]. That yielded to the large volume of smart meter data that reveals more insights on the consumer behavior and enables extracting site-specific information, allowing DNOs to introduce different price schemes for consumption based on the time of day and the season [3]. The network between the smart meters and the business systems enables collection and communication of information to the system components such as service providers, DNOs and customers. This leads to more system component integration toward demand response services. For instance, consumers can benefit from the information provided by the system to change their energy usage behavior to take advantage of lower electricity tariffs. In fact, dynamic pricing models are gradually replacing conventional demand-side management measures [4] and may be seen as a key enabler of the smart grid business system and a key solution to curb the demand consumption growth during peak hours.

One of the main challenges in smart metering is how to extract useful information from the large data volume collected at the distribution network terminals (e.g. households, office buildings). For example, a distribution network equipped with one million smart meters and a meter reading frequency every 15 min lead to 35.04 billion measurements and 2920 Tb of data that will be recorded during one year. To tackle this problem, machine learning technique is a powerful set of algorithms that make it easier the analysis, the interpretation and discovering hidden structures in the smart metering data. In particular, data clustering is used to find similar groups having the same consumption patterns. Smart meter data clustering has been widely investigated toward consumers grouping and revealing their energy usage behavior which leads to tailored energy efficiency programs for specific users.

Smart meter data clustering or consumers' loads patterns profiling depend basically on the goals or the purposes of the analysis, including energy efficiency programs, demand response, load forecasting and tariff design. For demand response and demand side management, load profiling allows targeting and identifying consumer groups who exhibit specific consumption patterns (e.g. higher peak load, high system peak share, lower variability) toward energy reduction recommendations [5-8]. For instance, it might be easier to identify end-users with high variability for behavioral change recommendations and energy efficiency programs [9], and those with lower variability for DR programs [10]. Moreover, customer's load pattern clustering can make it easier the assessment of DR and energy efficiency program's impact on achieving energy reduction or influencing end-user's behavior [11]. Currently DNOs companies associate other social, economics and building owners' data (e.g. income, building area, family size, etc.) as the basis for consumer classification, clustering and designing electricity tariff [11]. However, the widespread of smart metering yield to new possibilities in defining new basis for designing electricity tariff exploiting consumption characteristics in forms of indicators (e.g. load factor, duration of peak demand) derived from mining load curves of consumers.

In this paper, we focus on mining the smart meter data toward grouping and identifying the consumers who are more likely responsible for the network peaks. In terms of features, we propose to use the responsibility factor and the consumption variance as inputs for the clustering algorithms instead of the raw consumption series. Two clustering approaches have been investigated hierarchical clustering and SOM to assess the characteristics of the customer groups and to leverage the graphical representation capability of the SOM to visualize the cluster in an effective and useful way.

2. Related work

In the literature, different approaches have been proposed for load pattern clustering, which differ mainly by the targeted application and the set of features used as inputs for the clustering algorithm. The most commonly used features are the daily smart meter records themselves [12] so called time domain data. The data are basically preprocessed (e.g. cleaning, normalization) prior to the clustering step. Cleaning the raw data removes the noisy and possible frozen meter records that may occur due to some faults [2]. Then the data is normalized to improve the data integrity and to emphasize the load shape patterns rather than the amplitude absolute value [9,13]. However, processing smart meter data often requires special computing equipment and software. To simplify manipulating the large data volume, different profile indicators may be formulated and extracted to reduce the complexity of data manipulating and simplify the data mining process. Different indicators derived from load profiles are proposed in the literature as input features for clustering purpose as summarized in Table 1. Basically, these indicators provide a descriptive summary of the consumer load, including; statistical features (e.g. mean, variance, standard deviation, etc.) and key events within consumption pattern including load magnitude (e.g. maximum, minimum, base load, peak to off-peak loads, etc.). Using such a set of descriptive indicators yields to considerably reduce the large time series data to a household-by-household basis instead which allows removing redundant and atypical meter data.

Previous studies using such indicators have attempted to group distinct customer segments having a similar load pattern leading to a customized DR program. However, a customer group having a peaky load pattern, resulting from the clustering step and using the aforementioned indicators as inputs, may not reflect the real characteristics of that customer segment since some other questions related to the group behavior may remain unanswered. For example, how each consumer group contributes to the peak system? when each group is more responsible for the system peak? How can we select the consumer group for specifics DRs programs? Our study instead proposes using a responsibility factor elaborated at the system peak hours and the consumption variance as major driving features for consumer profiling. We evaluate the potential value of these indicators as a key characteristic of customer clustering through two clustering techniques Hierarchical clustering and SOM.

Table 1. Profile indicators as input features for clustering loads pattern.

References	Indicators	Description
[14]	Total consumption power	Total consumed power during period of the study (yearly, monthly, weekly daily)
[15,16]	Mean load	Mean load´ during defined time period
[16,17]	Load factor	Ratio of the average load during predefined time period to the maximum load during that period
[18]	Base load	Mean load from 2 am – 5 am
[19]	Morning maximum	The maximum consumption observed between 6 am – 10 am
[16]	Midday load	Mean observed between 12 am and 2 pm
[20]	Evening consumption load	Mean observed load during evening peak relative the mean load during the rest of the day.

3. Methods

3.1. Datasets pre-processing

This study relies on the publicly available data set CER [21], it consists of half hourly records of electricity consumption collected from 4232 households for 75 weeks starting from July 2009. For this study only one full year consumption data have been used. The data set was first cleaned and pre-processed to normalize the data. The data was normalized by dividing every consumer daily profile by its total consumption. The data set is transformed in the form of a N×M matrix; N rows for the number of households and M columns corresponding to the meter readings.

3.2. Feature definition

In this section we introduce customer profile indicators used as key features for customer profiling. These indicators allow DNOs to identify customers having a higher peak share and contributing more to the peak system during the peak hours. In fact, smart metering has enabled acquiring consumption data with higher resolution, therefore it becomes more interesting to have more insights on the customer behavior through more detailed analysis. For instance, descriptive profile indicators may be elaborated at more grained level (e.g. hourly or daily) instead of aggregated monthly analysis. One key indicator of customer profiling is the responsibility factor *RF* defined as the load of an individual component L_i at the peak system time divided by the peak load of this individual component P_i, expressed as follows:

$$RF_i = \frac{L_i\,(at\ system\ peak)}{P_i} \tag{1}$$

The RF is defined at the system peak hours; morning peak (06:00 -10:00), midday peak (11:00 -13:00), evening peak (19:00-01:00) and the night period (01:00 - 06:00). The RF tells how much an individual component is contributing to the system peak, it ranges between 0 and 1. If a component peaks at the same time as the system then its RF is close to 1. The variance of the daily consumption has been computed as well as an indicator of the consumer behavior variability.

3.3. Clustering Algorithms

Based on the defined features, clustering algorithms are widely used to perform customer pattern profiling. In this study two clustering techniques have been investigated: hierarchical clustering and the self-organising map SOM.

4. Results and discussions

We assess the clustering result using the responsibility factor and the consumption variance for a typical peak system period. One-year daily data of 4225 residential consumers are processed and clustered via hierarchical clustering and SOM algorithm. Hierarchical clustering produced five clusters representing 5 customer groups with different patterns as shown in Fig.1. a. The mean of each cluster is represented in the Fig.1. b. Table 2 summarizes each cluster information in terms of cluster's center and size as a percentage of the total observations. It could be seen that the majority of clusters has the same pattern shape with a dominant broad evening peak and a secondary peak during the morning hours.

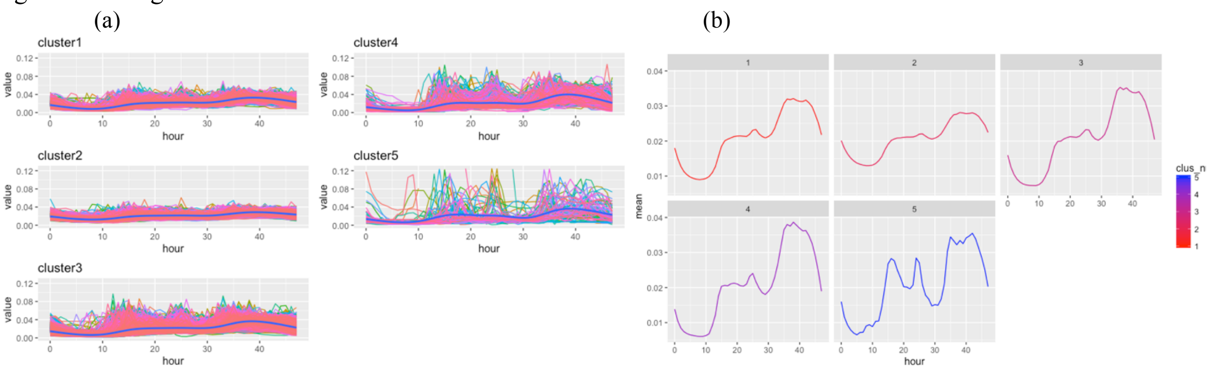

Fig. 1. (a) Clusters in time domain; (b) Representative cluster mean.

Table 2. Clusters information

Cluster	RF Morning	RF Midday	RF Evening	RF Night	Variance	Cluster Size
1	0.44	0.53	0.74	0.25	7.7	29.7 %
2	0.55	0.63	0.78	0.41	3.9	25.8 %
3	0.36	0.44	0.66	0.17	12.4	35.5 %
4	0.29	0.33	0.54	0.1	20.5	6.9 %
5	0.24	0.25	0.36	0.09	36.2	1.7 %

Looking at the summary table, we can see that the different customer classes have different characteristics even they share the same shape. For instance, cluster 1 and cluster 2 have almost same evening RF but they slightly differ from the rest of variables with clear different variance value. Cluster 3 and cluster 4 have lower responsibility factor at different peak system period but they are quite characterized by a remarkable variance difference. Cluster 5 shows a specific pattern with 3 distinct peaks, lower responsibility factors and highest variance (36 %). Basically, it could be seen that customers with higher responsibility factors tends to have the lower consumption variability. That may be further assessed by plotting the correlation matrix of the input features, as shown in Fig.2. It could be seen that the consumption variance is inversely correlated to the responsibility factor. Consumers having higher responsibility factor on the peak system (> 0.7) tend to have lower variance (< 8 %) which means less unchanged behavior. Merging cluster 1 and cluster 2 to a same class may lead that more than 55 % of customers have low variance. Therefore, it could be worthwhile to target those consumers having a higher contribution to the peak system and lower variability in their behavior by DR programs.

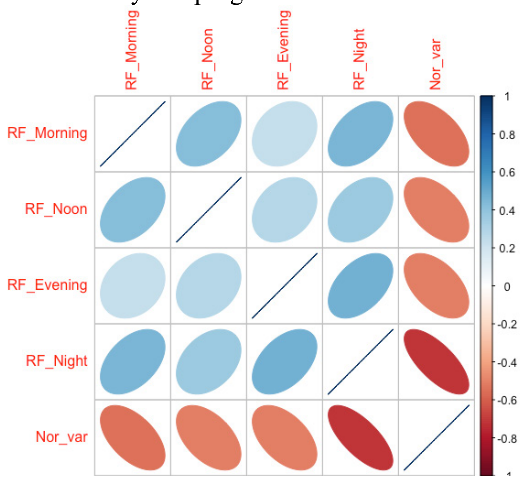

Fig. 2. Correlation matrix of the input features.

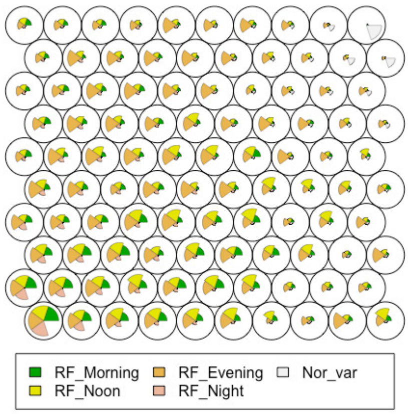

Fig. 3. SOM: Input variables magnitude representation.

To have a graphical projection of the consumers on the input variable's dimensions, SOM as unsupervised clustering technique has interesting and useful graphical representation capability that enable effective visualization of similar customer groups having similar characteristics on a bi-dimensional map. Fig.3 and Fig.4 show the SOM clustering results and the heatmaps of observations distribution according to each input dimension. That allows to see the hidden patterns in the data and identify observations linked to each specific pattern. For instance, it could be seen in Fig.3 consumers having a higher responsibility factor magnitude for each day time period are aggregated in the same area of the SOM (left-bottom of the SOM). Consumers with higher consumption variability are aggregated on the top-right area of the map. Fig.4 visualize further the distribution of observations on the SOM in relation to each input variable.

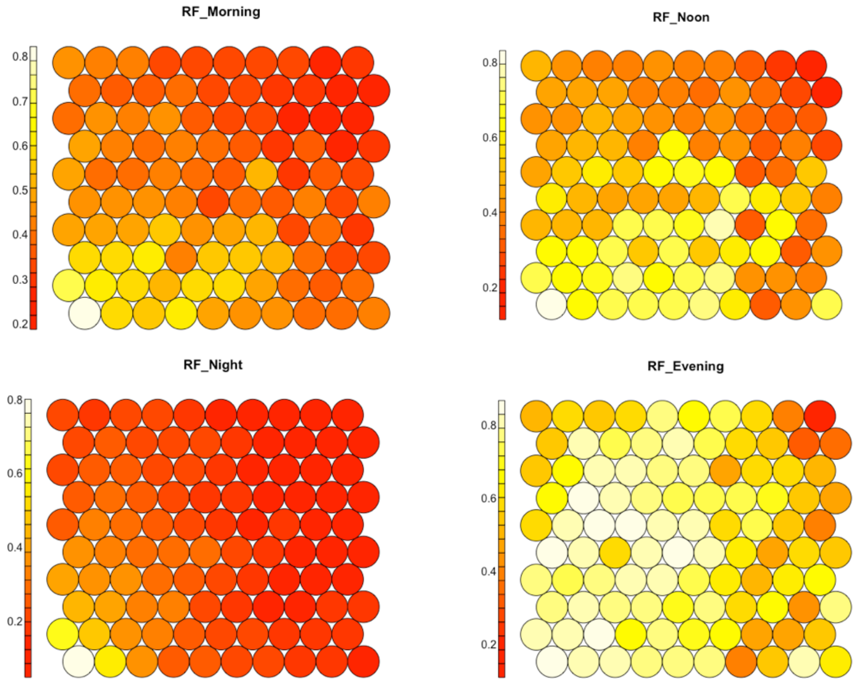

Fig. 4. Heatmaps of clusters distribution vs input features.

5. Conclusions and future work

This paper presented a mining approach of smart meter data toward identification of consumer group having a higher contribution to utility system peak. RF derived from the smart meter data at the system peak periods was used to describe the contribution of each customer to the system peaks. Further, the consumption variance was used as an indicator of behavior variability or customers' behavior changes. Two clustering techniques, hierarchical clustering and SOM, have been investigated to have more insights on the consumer group using the RF and the variance as input features to the clustering algorithms. Five clusters may be distinguished through the hierarchical clustering having almost the same shape, but with highly different characteristics in terms of RF and variance. The analysis of clustering result and the correlation analysis of the input variables has shown that more than 55 % of the customer having higher responsibility factor tend to have less changed behavior. The SOM provides an effective and simple visualization of customer group distribution through a bi-dimensional graphical representation that represent an effective way to identify specific groups sharing similar profile indicator to be addressed by a specific DR programs.

Future work will continue to investigate the key consumption indicators of customers considering the seasonal and calendar effects. The resulting clusters may be characterized further by correlating the socio-economics data to build a consumer identification model that may be used for DR programs. Simulation scenarios of different pricing models based on the resulting customer group may then be designed to evaluate the potential of specific electricity tariff model to the targeted customer group having a high contribution to the peak system.

Acknowledgements

The authors would like to thank the Knowledge Foundation (KK-stiftelsen) for the financial support.

References

[1] Zhou K, Fu C, Yang S. Big data driven smart energy management: From big data to big insights. Renewable and Sustainable Energy Reviews 2016;56:215–25.
[2] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy 2012;42:68–80.
[3] Sinopoli J. Smart building systems for architects, owners, and builders. Amsterdam: Butterworth-Heinemann; 2010.
[4] Walawalkar R, Fernands S, Thakur N, Chevva KR. Evolution and current status of demand response (DR) in electricity markets: Insights from PJM and NYISO. Energy 2010;35:1553–60.
[5] I. Dent, T. Craig, U. Aickelin, and T. Rodden, "An approach for assessing clustering of households by electricity usage," in Proc. 12th Annu. Workshop Comput. Intell. (UKCI), Edinburgh, U.K., Sep. 2012, pp. 1–4.
[6] B. A. Smith, J. Wong, and R. Rajagopal, "A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting," in Proc. ACEEE Sum. Study Energy Efficiency Build., 2012, pp. 374–386.
[7] H.-A. Cao, C. Beckel, and T. Staake, "Are domestic load profiles stable over time? An attempt to identify target households for demand side management campaigns," in Proc. IEEE Ind. Electron. Soc. (IECON), Vienna, Austria, Nov. 2013, pp. 4733–38.
[8] A. Albert, T. Gebru, J. Ku, J. Kwac, J. Leskovec, and R. Rajagopal, "Drivers of variability in energy consumption," in Proc. ECML-PKDD Data Anal. Renew. Energy Integr. (DARE), Prague, Czech Republic, Sep. 2013, pp. 1–12.
[9]Tsekouras GJ, Hatziargyriou ND, Dialynas EN. Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers. IEEE Transactions on Power Systems 2007;22:1120–8.
[10] Panapakidis IP, Papadopoulos TA, Christoforidis GC, Papagiannis GK. Pattern recognition algorithms for electricity load curve analysis of buildings. Energy and Buildings 2014;73:137–45.
[11] Tsekouras G, Kotoulas P, Tsirekis C, Dialynas E, Hatziargyriou N. A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. Electric Power Systems Research 2008;78:1494–510.
[12] Stephen B, Mutanen AJ, Galloway S, Burt G, Jarventausta P. Enhanced Load Profiling for Residential Network Customers. IEEE Transactions on Power Delivery 2014;29:88–96.
[13] Kwac J, Flora J, Rajagopal R. Household Energy Consumption Segmentation Using Hourly Data. IEEE Transactions on Smart Grid 2014;5:420–30.
[14] Mcloughlin, F. (2013). Characterising Domestic Electricity Demand for Customer Load Profile Segmentation [THESIS]. Dublin: Dublin Institute of Technology.
[15] Beckel C, Sadamori L, Santini S. Towards automatic classification of private households using electricity consumption data. Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings - BuildSys 12 2012.
[16] Carroll J, Lyons S, Denny E. Reducing household electricity demand through smart metering: The role of improved information about energy saving. Energy Economics 2014;45:234–43.
[17] Mcloughlin F, Duffy A, Conlon M. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. Energy and Buildings 2012;48:240–8.
[18] Yohanis YG, Mondol JD, Wright A, Norton B. Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use. Energy and Buildings 2008;40:1053–9.
[19] Chicco G, Napoli R, Postolache P, Scutariu M, Toader C. Electric energy customer characterisation for developing dedicated market strategies. 2001 IEEE Porto Power Tech Proceedings.
[20] Powells G, Bulkeley H, Bell S, Judson E. Peak electricity demand and the flexibility of everyday life. Geoforum 2014;55:43–52.
[21] Sustainable Energy Authority of Ireland. Available: http://www.seai.ie.