Impact Factor (SJIF): 4.542

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444

Volume 4, Issue 6, June-2017

Survey on- Behaviour of load profiling in big data

[1]DIVYA S S, [2]NAVEEN KUMAR,

[1]PG student, Dept. of CSE, UBDTC, Davanagere

[2]Assistant Professor Dept. of CSE, UBDTCE, Davanagere

Abstract—The growing interest in discerning behaviors of electricity users in both the residential and commercial sectors the advent of high-resolution time-series power demand data through advanced metering, mining this data could be costly from the computational viewpoint. This work uses both raw data from household consumption data and synthetic profiles. The motivation for this work is to improve the clustering of electricity load profiles to help distinguish user types for tariff design and switching, fault and fraud detection, demand-side management, and energy efficiency measures.

Keywords- Clustering, load profile,big data, electricity consumption.

I. INTRODUCTION

Most meters will be capable of generating data with high temporal resolution, but customers and network operators have different requirements, expectations, and priorities for using this data. This paper addresses the question of how much information is sufficient. Data mining and pattern recognition techniques have been applied to power demand time series. The application of these techniques can be useful for many purposes such as detecting failures and fraudulent usage or grouping users that present similar consumption patterns.

Load profiling refers to the formulation of representative load curves for single consumers and groups of consumers. Based on certain criteria, the consumers are grouped together in a number of classes. Each class has a representative daily load curve which is the weighted average of the curves that belong to the cluster. According to this approach, the consumers are not only distinguished in macro-categories like residential, industrial, etc., but subcategories are formatted within the macro-categories. Such a detailed analysis entails advantages for the most market entities.

Load profiling enables the retailers to make lower risk settlement in spot market. Load profiles can be the basis for the design of flexible tariffs, a scheme that leads to increased profits. Moreover, it can be used effectively in short and medium-term load and energy forecasting. The forecasts can be done for a single consumer, for a class or even for the total demand that is served by the retailer. Various techno-economic studies for the evaluation of demand side management policies can be effectively carried out with the aid of load profiling.

II. FORECASTING FRAMEWORK AND CORE TECHNIQUES

The proposed forecasting framework includes five steps as shown in Steps 1 through 3 consist of machine learning techniques that aim to discover typical load patterns of historical load data and then find their critical influential factors to establish classification rules. Step 4 is a model training process, where parameter combinations for corresponding load patterns are chosen to build forecasting models. In Step 5, individual load forecasting results are added to arrive at the final system load.

A. Improved Hierarchical Clustering Technique

Since each individual load can consist of different types of load curves due to factors such as weather, dates, etc., a hierarchical clustering technique is used to classify historical load curves into several patterns. Using a hierarchical clustering algorithm represents a bottom-up approach, as shown in First, take each sample as a separate class. Second, calculate the distance between each sample. Third, merge classes that meet the distance requirements. Finally, repeat the above three steps until the number of classes meet the requirements. Proximity matrix A stores the distances between every two classes.

B. Identification of Critical Influencing Factors

Many factors, including temperature, humidity, day type, etc. have impact on loads. For different loads, critical influential factors are not the same. For example, temperature could be a major influential factor for residential loads. But temperature may have little impact on some temperature non sensitive industrial loads. Grey correlation analysis is applied to determine the critical influential factors of each individual load.

C. Establishment of Classification Rules

Based on cluster analysis and association analysis results a decision tree, which is based on CART algorithm is developed to establish relationships between clustering results and critical factors.

D. Selection of Appropriate Forecasting Models

For different load patterns, different support vector machine (SVM) models and parameters are developed to ensure the forecasting accuracy within the required limits. First proposed by Vapnik, SVM is an effective technique
for classification and regression problems. SVMs are supervised learning models with associated learning algorithms that analyze data and recognize patterns used for classification and regression analysis. SVM can efficiently perform a nonlinear classification using the kernel trick, implicitly mapping the inputs into high-dimensional feature spaces.

E. Forecasting Total Load

The total system load is forecasted based on aggregation of an individual load's forecasting results after the forecasting result of each user's load is obtained, the forecasted total load total can be calculated by adding all the
Fore casted individual loads, together with line loss.

III. CLUSTERING TECHNIQUE BACKGROUND

3. Clustering technique background

3.1. Clustering process

Classification of customers is obtained by applying a pattern recognition methodology which includes the following steps:

(1) Data selection: The first step is the selection of the data with more significance to the process. This selection is made according to the voltage level of the consumers. Separate studies must be conducted for different voltage levels.

(2) Customer classification in whole categories: In order to form the customer classes, the whole set of customers can be preliminarily partitioned into macro-categories defined on the
basis of global criteria.

(3) Data cleaning and preprocessing: In the cleaning phase, inconsistencies in the data are checked and outliers are removed. If necessary, a preliminary execution of a pattern recognition algorithm is carried out, in order to track bad measurements or network faults which if uncorrected, will reduce the number of the useful typical days for a constant number of clusters.

(4) Load curve clustering and selection of the typical load pattern for each class: The clustering methods are applied on
customer load curves and finally, the typical load pattern (TLP) of each class is selected. TLPs are then used for different programs such as demand side management (DSM),
designing suitable tariffs, etc. also, in this step clustering algorithms with better performance are selected by using adequacy measures.

IV. CASE STUDY

A. *Description of the Data Set*

The data set used in this paper was provided by Research Perspective, Ltd. and contains the electricity consumption of 6,445 customers (4,511 residents, 391 industries, and 1533 unknown) over one and a half years (537 days) at a granularity of 30 minutes. The whole data set consists of total 3.46 million daily load profiles. The bad load profiles are roughly identified by detecting the load profiles with missing values or all zeroes. Among these massive load data, we eliminated 6187 bad load profiles, which is a very small sample (approximately 0.18%) of the whole data set.

B. Modeling Dynamics of Electricity Consumption for Each Customer

According to the regular routine of electrical customers, we reasonably divide a day into four periods: Period 1 (00:00-06:30, 22:00-24:00, overnight period), Period 2 (06:30-11:30, morning period), Period 3 (11:30-17:00, daytime period), and Period 4 (17:00-22:00, night period). On this basis, the load data are transformed into PAA representations which also vary from 0 to 1.

C. Clustering for Full Periods

To obtain the typical dynamic characteristics of electricity consumption and to segment customers into several groups, CFSFDP is first applied to the full periods.

D. Clustering for Each Adjacent Periods

Sometimes, we may not be concerned with the dynamic characteristics of full periods and instead concentrate on a certain period of time. For example, to evaluate the demand response potential in noon peak shaving of each customer, the dynamics from Period 1 to Period 2 are much more important; to measure the potential to follow the change of wind power at midnight, the dynamics from Period 4 to Period 1 should be emphasized. Thus, it is necessary to conduct customer segmentation for different adjacent periods.

V. ACKNOWLEDGEMENT

I express my thanks to our guide naveen kumar, assistant professor of computer science, university BDT, for sharing his valuable knowledge.

VI. CONCLUSION

This work analyses the impact of the temporal resolution when clustering electricity load profiles. Several algorithms have been systematically tested by changing the resolution of the input data (of real household consumption). The results are evaluated with internal and external validity measures, and the

efficiency was computed using a large-scale synthetic data set. The implication of the work presented is that to be useful to electricity retailers in discerning differences between consumers time series power use data needs to be at a frequency of least 30 min and ideally 8 or 15 min. Distribution network operators may have recourse to higher resolution data for various applications, however, 8-min data would provide a useful basis. Data collected at frequencies slower than 30 min, and certainly 60 min, is not sufficiently reliable for most purposes.

VII. REFERENCES

[1] S.-J., Huang and K.-R. Shih, “Short-term load forecasting via ARMA model identification including non-Guassian process considerations,” IEEE Transactions on Power Systems, vol. 18, no. 2, pp. 673–679, May 2003.

[2] H.-T. Yang, C.-M. Huang, and C.-L. Huang, “Identification of ARMA model for short term load forecasting: an evolutionary programming approach,” IEEE Transactions on Power Systems, vol. 11, no. 7, pp. 403–408, Aug. 1996.

[3] K. Moshkbar-Bakhshayesh and M. B. Ghofrani, “Development of a robust identifier for NPPs transients combining ARIMA model and EBP algorithm,” IEEE Transactions on Nuclear Science, vol. 61, no. 4, pp. 2383–2391, Jul. 2014.

[4] P. R. Ji, D. Xiong, P. Wang, and J. Chen, “A study on exponential smoothing model for load forecasting,” in Proceedings of 2012 Power and Energy Engineering Conference (APPEEC), Shanghai, China, 2012, pp. 1– 4.

[5] R. Gob, K. Lurz, and A. Pievatolo, “Electrical load forecasting by exponential smoothing with covariates,” Applied Stochastic Models in Business and Industry, vol. 29, no. 6, pp. 629–645, Dec. 2013.

[6] Gerbec D, Gasperic S, Smon I, Gubina F. An approach to customers daily load profile determination. In: Proceedings of IEEE power engineering society summer meeting conference; 2002. p. 587–91.

[7] Gabriel SA, Genc MF, Balakrishnan S. A simulation approach to balancing annual risk and reward in retail electrical power markets. IEEE Trans Power Syst 2002;17(4):1050–7.

[8] Carrion M, Conejo AJ, Arroyo JM. Forward contracting and selling price determination for a retailer. IEEE Trans Power Syst 2007;22(4):1050–7.

[9] Green RJ. Retail competition and electricity contracts. CMI working paper series. <http://www.econ.cam.ac.uk/electricity/publications/wp/ep33.pdf>.

[10] Gabriel SA, Kiet S, Balakrishnan S. A mixed integer stochastic optimization model for settlement risk in retail electric power markets. Netw Spat Econ 2004;4(4):323–45.