Chapter 7
Mining Smart Meter Data:
Opportunities and Challenges

Ayushi Tandon
Indian Institute of Management Ahmedabad, India

ABSTRACT

Metering side of electricity distribution system has been one of the prime focus of industry and academia both. The most recent advancement in this field is installation of smart meters. The installation of smart meters enables collection of massive amounts of data regarding electricity generation and consumption. The analysis of this data could help generate actionable insights for the supply side and provide the consumers demand management-related inputs. The problem addressed in this chapter is to identify suitable data mining algorithm for applications like: estimating the demand and supply of electricity, user and use profiling of commercial, and industrial customers, and variables suitable for these purposes. This chapter, on the basis of rigorous literature review, presents a taxonomy of smart meter data mining. It includes the summary of application of smart meter data analytics, characteristics of dataset used, and smart meter business globally. This chapter could help researchers identify potential research opportunities, and practitioners can use it for planning and designing a smart electricity system.

DOI: 10.4018/978-1-5225-3996-4.ch007

INTRODUCTION

Electricity is vital utility as it supports many critical activities and services in human life. Electric power system consists of generation, transmission and distribution networks. The majority of electricity generation at present is centralized operation relying on non-renewable energy sources like coal, natural gas and nuclear. Electricity supply as an industry is highly capital intensive and suffers heavily due to demand-supply mismatch and transmission losses. In the recent decade application of Information and Communication Technology (ICT) in the power Industry has given rise to a new smart grid and smart metering infrastructure. Attempts are being made to enhance energy efficiencies and increase the reliability of electricity generation, transmission, and distribution systems by implementing network intelligence, empowering customers and enabling grid flexibility with smart devices deployments.

The evolution from the first known electricity meter, patented by Samuel Gardiner in 1872, towards a distributed electricity grid model able to manage numerous generation and storage devices in efficient and decentralized manner(Uribe-Pérez, Hernández, de la Vega, & Angulo, 2016). The metering side of the distribution system has been the focus, and the most recent advancement is the installation of *Smart Meters.* During 90's in the developed economies, automated meter reading (AMR) systems were introduced in the distribution networks moving away from electrochemical meters (Farhangi, 2010). Recently the AMR systems are being replaced with smart metering infrastructure (SMI), with two-way communication capabilities. These smart meters can digitally send meter reading to suppliers and also provide consumption and pricing related feedback to the user. This bidirectional communication ability of smart meters enables collection of the massive amount of data regarding the electricity generation and consumption. One of the biggest challenges in smart metering is the gathering and analysis of this 1.9 Terabytes of data per user/year. But, this also presents the most significant opportunity since analysis of this data could help generate actionable insights for supply side and provide the consumers demand management related inputs to the consumer. This opportunity if realized could enable countries to achieve the 20/20/20 European Sustainability Objectives of 20% reduction in energy consumption and, greenhouse gas emissions while 20% increase in renewables uses by 2020. According to the recent Benchmarking Report by the European Commission, the European Union has set the target to achieve 72% coverage of households with smart meters by 2020. In May 2017, Bangladesh Power Development Board (less developed country) launched smart meter service in the capital city, realizing the potential of this technology.

To realize these digital opportunities, all the companies part of electricity generation and distribution ecosystem need to transform their operations. To begin, they must develop a digital transformation strategy incorporating the understanding of smart meter data analytics. With the commercialization of smart meters, a huge amount of data is generated, which should be utilized for price and demand forecasting. However, the storage, querying, and mining of such smart meter data streams poses a significant challenge. The application of data mining algorithms on smart meter data could help meet both consumer's and power distribution utility's objectives. In this chapter, we present a rigorous review of academic literature and provide the taxonomy of application of smart meter data analytics. The researchers could refer this chapter to identify potential research opportunities and practitioners can use these findings for planning and designing an intelligent electricity system.

The guiding questions for this literature review are:

1. What algorithms are applied on smart meter data and for what purpose?
2. What are the characteristics of smart meter datasets and other external input parameters used for data mining?
3. What are the challenges associated with mining smart meter data and how such challenges are addressed in academic literature?

Most of the data analytics-based research is done to address either electricity demand or supply side issues in silos, so this literature review aggregate these results in a systematic manner.

LITTERATURE SEARCH

As the primary focus of this chapter is the application of data mining related algorithms on smart metering data, so we used search strings "smart meter + data mining," "smart grid+ data mining" and "smart meter +load forecasting." We restricted our search only to the articles having search string as part of the article title, keywords, and abstract. Also, we limited our review to articles published from 2008 till 2017. The search is conducted in the Scopus database along with Google Scholar and IEEE Xplore Digital Library. We did a manual scan of the article by reading abstract and rejecting irrelevant articles. Finally, we have summarized 35 articles as part of this chapter.

ORGANIZING FRAMEWORK

Intelligent smart metering interface (SMI) is one of the assorted technological innovations. This generates a large amount of data that has the potential of solving many challenges in power infrastructure. The main advantage of SMI is that it is bidirectional in nature and hence provides utility in the form of full visibility over consumption data and pricing. Many research articles have been published recently using SMI data. Data mining algorithms are primarily applied on this data for doing individual consumer profiling, load forecasting, tariff determination and differentiated load monitoring. We have organized the chapter under these broad headings.

- Consumer data analytics
- Dataset characteristics
- Privacy
- Algorithms applied
- Implementation examples

CONSUMER DATA ANALYTICS

Customer Segmentation and Load Forecasting

Customer segmentation help power distributors to identify interesting groups of customers. It is used to determine the tariff rates and to perform an optimal allocation of resources to meet electricity demand. Customers segmentation based on the consumption pattern is the useful exercise for power grids, because of overhead costs involved in scaling up and down of the generator. This segmentation also has the potential of inducing a change in consumption pattern by enforcing differential pricing based on average consumption, such that use pattern distributions are statistically similar within groups and statistically different across groups.

Alzate and Sinn (2013) applied Kernel Spectral Clustering algorithms on experimental time series data for residential customers and, small and medium enterprises. The model was trained to predict customer segment using auto regression. K-means clustering approach was used to group the customers based on similarities in their consumption behavior. Then, load forecasting was conducted for each group by applying multilayer perceptron neural network. Many researchers showed less error in prediction of energy requirements using neural networks and support vector machines (Savio, Karlik, & Karnouskos, 2010). Time variant characteristics of load data could also be used for power consumption pattern prediction (Zhang, Grijalva, & Reno, 2014). They proposed a layered tree structure using load, time and other attributes to determine the branch of the decision tree. These forecasting techniques could be used to identify customers having peak demands and high variability in power consumption. This group of customer is the potential target for demand response management offers. They should be sensitized to shift activities (like using a washing machine, dish washing) to the time of day when charges are low. (Quilumba, Lee, Huang, Wang, & Szabados, 2015).

For load forecasting context based mining of electricity use pattern by dynamically monitoring devices that operate together is also useful (Liao, Liao, Liu, Fan, & Omar, 2016). Sequential pattern mining i.e. identifying the set of activities that occur in sequence and require significant power consumption is an alternative to consumption pattern mining (Kouzelis & Bak-jensen, 2014). This paper suggested short-term load forecasting by finding frequently occurring patterns and representing the load data as a sequence of these states. Past sequences, similar to the present one, were sought to define a Markov-like process. The sequence prediction algorithm was used to determine rules, and Davies-Bouldin Indicator (DBI) was used to get the cluster of such sequences. Thus, probabilities for the future states of the load were predicted using present cluster information as Markov state. Certain data points in the dataset showed unusual consumption patterns, these subgroups were identified to improve classification accuracy, and their consumption patterns were analyzed separately.

Knowing Customer

Re-identification of customer using electricity consumption data from the smart meter is possible. The feature selection for re-identification is done by informed guessing like using prior knowledge regarding the schedule of customer or number of occupants etc. Research showed that household and consumer identification using smart meter data is possible with nearly 60% accuracy. Aman et al. (2011) conducted an experiment at the University campus and proposed a model for campus daily energy use prediction, which could be extended to city scale. The model utilized disaggregated data collected at appliance level, and this improved the forecasting accuracy of the model. Appliances were associated with well-defined finite ON/OFF state and paper suggests that the timings of these events if detected could help in profiling of customers. The customer profiling like the number of members in the household, their age, education, etc. is useful for targeted advertising and creating energy use related awareness (Albert & Rajagopal, 2013). Fusco, Wurst, and Yoon (2012) suggested using power consumption data for the classification of households according to parameters like the presence of kids, ownership of specific appliances, employment status and education level of the residents. Many researcher articles are published on consumer profiling by using time series data of 24-hours power consumption (Beckel, Sadamori, & Santini, 2013; Lavin & Klabjan, 2015)

Beckel, Sadamori, and Santini (2014) also described how specific properties of a household – like its size or the number of individuals living in it – could be inferred from its electricity consumption profile. Profiling customer based on their electricity use pattern across day is done by applying techniques like K-means, K-medoid, and self-organizing maps (McLoughlin, Duffy, & Conlon, 2015). Mining smart meter could reveal consumer's income and education level (Viegas, Vieira, & Sousa, 2016). This technique could be applied to segment consumers based on their lifestyle like household activity time (Kwac, Flora, & Rajagopal, 2014). These results highlight need for research in the field of data security and anonymization of customer profile while transmitting smart meter data (Buchmann, Boehm, Burghardt, & Kessler, 2013).

Demand Management

The consumer of electricity is billed for its usage, but power generation is a complicated process with negative externality as it is heavily dependent on non-renewable fossil fuel. The main issue with power transmission and distribution is that the scaling up and down of the capacity is a costly process. Thus predicting demand and deciding the price is a challenge, which ICT-based smart power infrastructure is trying to address by applying data mining algorithms. Demand side management requires predicting the per-household energy consumption and informing same to the customer. This can be done easily and efficiently by using high-resolution data collected with the help of various sensors connected to appliances. Deployment of sensors in households provides high dimension data that can benefit standard electricity consumption prediction tools.

People get ready to leave for the office, school, etc. in the morning and get back at night in most parts of the world. The household consumption patterns show two peaks i.e. morning and evening time because of centralization of activities at these time. Marinescu, Harris, Dusparic, Clarke, and Cahill (2013), proposed a combination of several methods to predict better during these times and suggest to consumers a list of activities that can be shifted to some other time. This could peak off demand, by ensuring better load scheduling at the consumer end. To manage demand and influence customer behavior by generating useful insights the data needs to be stored efficiently and selectively. The predictive power of algorithm can be improved by doing variable selection that is an accurate representation of consumption pattern. The minimization of the number of the model input variable and forecasting error of model with the help of Genetic Algorithm- is suggested for simplified data-driven short-term forecasting of household electric loads (Niska, 2015).

Home appliance load modeling was done with the help of data collected using smart meters. The hidden markov model (HMM) was used to predict the state of appliances from aggregated power consumption data. Overall energy consumption due to the appliance of interest was calculated by using Gaussian probability distribution function in convention HMM (Guo, Wang, & Kashani, 2015). Article published in 2017 suggested using Gaussian Mixture Model for disaggregating household electricity consumption and identifying activities associated with appliances by approximating the distribution of power levels. The assumption in that paper was data points collected from meter were generated from a mixture of a finite number of Gaussian distributions and different modes of devices have different parameters associated with the Gaussian function (Cao, Wijaya, Aberer, & Nunes, 2017). These data mining and data visualization techniques could be applied for presenting each appliance related home energy consumption to the consumer in an easily comprehendible format. This way consumers would be encouraged to directly participate in demand response management which ultimately might lead to an overall energy consumption decrease.

Instead of using the multiple sensors to collect data, load disaggregating could be done from overall electricity consumption data to identify the start and stop time for most household appliances and certainty in their usage. The paper by Gajowniczek and Zabkowski applied C-means clustering and association rule mining to identify time-related patterns in appliance usage and the list of appliances used together (Gajowniczek & Zabkowski, 2015). Probability based standard kernel density estimation and conditional kernel density estimation (CKD) methods were applied to derive individual energy consumption function using previous week consumption data. This disaggregated data in combination with qualitative data collected from residents like: their perception towards dynamic pricing etc. could generate useful business insights. This information could be further utilized to provide meaningful feedback to the customers based on their energy time and appliance use profile (Stankovic, Stankovic, Liao, & Wilson, 2016).

This might help users in managing their demand during peak price hours as they will be informed charges associated with the use of the particular appliance by converting density estimates of consumption into density estimates of cost for different tariffs.

Suppliers could also use these demand estimates to devise innovative time-of-use pricing (Arora & Taylor, 2016). They could make data dependent decision making regarding the dynamic time of use pricing and this might encourage customers to shift their consumption resulting in significant savings. Time of use tariff is generally determined by clustering user group into three categories: commercial, industrial and residential. The tariff solution that flattens the demand curve is recommended because it is more efficient regarding energy generation (Zala & Abhyankar, 2014).

Distributors are also interested in integrating demand forecasts using smart meter data with renewable power generation data for efficient electricity distribution (Madureira, Gouveia, Moreira, Seca, & Lopes, 2013).

DATASET CHARACTERISTICS

With the commercialization of Smart meters, a huge amount of data regarding consumption pattern has been gathered. This is available in the form of the standard dataset for research purpose. Also, many sponsored projects are running at various universities, and data generated by such SMI is also used for research. The data could be logged using the smart meter at varying time intervals like every 15 minutes (most commonly used), every 30 minutes or hourly. The minimum sample size for data mining algorithms is 1000 data points.

Besides meter data, many exogenous factors like attributes from the university's academic calendar that indicate occupancy patterns, static knowledge of buildings such as surface area, and historical weather information were used as attributes in forecasting models by researchers (Aman et al., 2011). Research has shown that weather, seasonal variations, and house characteristics are the important determinant of house electricity consumption. Besides this building characteristics like stand alone or apartment, floor size of the building, the age of the residential colony is identified as factors that significantly affect household electricity consumption. Apartments equipped with energy-efficient lights and double-paned windows have significantly lower power consumption in comparison to households lacking both. House with pet consumes more energy and this consumption is high during summers because of air conditioner use to make home conformable for pets (Kavousian, Rajagopal, & Fischer, 2013). Thus data used for analysis needs to be identified by observing power consumption about time-of-day, the season of the year, house head activity, number of pets, etc.

Since forecasted weather impacts electricity usage hence Chakravorty, Rong, Evensen, & Wlodarczyk (2014) suggested its use to predict energy consumption along historical consumption data during similar weather conditions. They applied distributed unsupervised Gaussian-Means Clustering Algorithm in the paper. After including weather and pressure related data as predictors in energy consumption prediction models for buildings, the testing accuracy of four different algorithms was compared (Candanedo, Feldheim, & Deramaix, 2017). Selective load profiling i.e. identifying the relationship between outside air temperature and specific appliance electrical load in individual buildings for example AC electricity consumption based on external weather conditions is also useful (Dyson, Borgeson, Tabone, & Callaway, 2014). Jin et al. proposed a two-step method to forecast load at the household level by using local temperature, the day of week variable. Categorical variable indicating holiday is also used as the predictor in the model along with power consumption data (Jin et al., 2014). Many academic articles are published based on the assessment of the correlations between energy demand and the exogenous variables.

Hayes, Gruber, and Prodanovic, (2015) quantified the effect of variables which influence demand, at each level of aggregation into three broad categories: time-related (e.g. day, hour of day, and whether or not the day is a normal working day); historical (e.g. previous hour demand, previous week equivalent hour demand, previous 24 hour average); and weather-related (temperature has by far the greatest influence, but other weather factors such as humidity/precipitation, solar irradiation, and the wind can also have effects). External temperature and duration of the certain climatic condition like summer or winter influence the household electricity consumption. Also, some members and daily use device in the household is responsible for more energy consumption. Thus combining household characteristic along with temperature gradient could be used to predict power consumption and categorize customer. This could be used to recommend suitable demand management approach (Birt et al., 2012).

Dataset Related Challenges

As smart meter data is transferred digitally, noise is added to the signal. An important problem in load data analysis is therefore input validation, where the task is to distinguish between data corruption and a change in data pattern due to a random event or periodical patterns. The overall process of data validation, noise removal, and preparation for further analysis is often referred to as data cleansing.

Missing Values: The data might have many missing values that need to be estimated before using it for prediction. ARMA (autoregressive moving average) and ARIMA (autoregressive integrated moving average) are most commonly used for predicting missing data values. Chiky, Decreusefond, & Hebrail (2010) has suggested two approaches, (1) Based on the use of the past distribution of slopes in the time series; (2) A more sophisticated one based on a stochastic approach. These two methods are used for interpolating missing power consumption time series values and validated it using artificial dataset.

Data Storage: The data generated via SMI is huge and not all data attributes are used for analysis. Hence Data mining algorithms (like SOM) are used for data selection. Segmenting consumer is effective when besides using smart meter data, survey data like: weather, location, holidays, communities, etc. is also used (Wijaya, Ganu, Chakraborty, Aberer, & Seetharam, 2014). Thus smart meter data centers have to store appropriate consumption data along with other exogenous information variables available (Park, Ryu, Choi, Kim, & Kim, 2015). Only the segment identification information about consumers by clustering them based on load profile could be stored (Quilumba, Lee, Huang, Wang, & Szabados, 2014). This will save memory.

PRIVACY CONCERNS

There are many actors involved in smart meter ecosystem, and this adds to vulnerabilities of system design. The actors and factors part of this ecosystem are: customer, smart meter installation companies, smart meter reading staff, electricity generation units, electricity distribution companies (public and private entities), renewable energy devices at user end, climatic conditions, infrastructure of city, building layout and materials used, government policies and marketplace dynamics. External conditions (e.g., location and weather), physical characteristics of dwelling, appliance, and electronics stock, and occupants. The smart meter system design must be such that security concerns must be taken into account during each phase of the project, and even when the system becomes obsolete. The paper by Alabdulkarim and Lukszo deals in great detail about security requirements during system analysis, design, and implementation. It also recommends role base access of data collected to reduce malicious attack opportunities (Alabdulkarim & Lukszo, 2008).

Intrusion

Studies have shown that providing household's occupants with personalized breakdown of appliance energy consumption allows them to take steps towards reducing their total energy consumption. Putting sensors on each device and monitoring the sensor, could provide details about device activity. This is an intrusive method, and hence unsupervised training method for appliance load monitoring has been suggested (Parson, Ghosh, Weal, & Rogers, 2014). Researchers have developed methods for non-intrusive appliance signature extraction from turn ON and OFF events by detecting power spikes and clustering the suspect events (Dong, Meira, Xu, & Chung, 2013). Association rule mining technique could be used to prepare an electric signature database for various home appliances. This technique could be utilized for non-intrusive load monitoring (Dong et al., 2013). Power attributes like active power, reactive power, and harmonic content ranges, are appliance specific and hence helps locate appliance specific events. The on/off duration of an appliance is a critical variable to estimate the appliance's power consumption (Wang et al., 2012). Thus various methods are possible for load monitoring, and these methods have potential to reduce capital investment by the power supplier.

The commercialized smart meter's designs should be audited against hacking and attack by cyberterrorists. A detailed risk assessment should be carried out by the energy watchdog, and consumers should be informed to protect them from a range of fraudulent transactions for financial gain, remote disablement of critical device, etc.

ALGORITHMS USED

There are many algorithms and combination of algorithms in the academic literature for data mining. The researcher could do the comparative analysis of these algorithms for mentioned use case by varying the context and dataset characteristics like granularity of data captured. We have listed commonly used algorithm for the application mentioned above in the chapter as-

- **Data Pre-Processing:** Regression-based algorithms are most suitable for data cleansing and filling up missing values. Support vector machine, SVM is also used for data prediction.
- **Customer Segmentation:** Clustering is most commonly used for identifying customer segments. Spectral clustering, K-means clustering are commonly used as the base method and it shows performance improvement when heuristic based algorithms like genetic algorithm are applied to identify a number of clusters.

- **Load Disaggregation:** Probability density based methods are suitable for non-intrusive load monitoring. This application requires device energy usage and on/off state to be modeled using probabilistic functions like Gaussian function.
- **Feature Selection:** Artificial Neural Networks (ANN), and interview-based methods are most efficient in feature selection.

IMPLEMENTATION EXAMPLES

Texas, Austin is the hub of smart energy companies and university projects. Austin Energy, one of the largest electric utility companies in the US with more than 448,000 customer accounts. The company announced in October 2016 to install 2.5 lakh smart meters and these new meters will send information every 15 minutes, making it easier to diagnose outages.

Grid4C company a young startup, provides software products that make it possible to predict and optimize a smart electricity grid for both power producers, suppliers, and home customers.

Center- point Energy, a Houston-based utility company was awarded the 'International Smart Grid Action Network Award of Excellence 'and the 'Global Smart Grid Federation Best Smart Grid Project' for 2016. They use 50 days' historical data to determine how much wholesale supply is needed to serve the market. To integrate renewable off-grid energy production at consumer end CenterPoint planned the installation of new smart meters that could send power-off notifications to grid and signal when power demand was restored.

In Italy, Enel, the third- largest energy provider in Europe started deploying smart meters to about 27 million customers in 2015. The Enel's in-house analytics platform can optimize system-wide energy operations, including storage project, generation, transmission and distribution and micro grids.

In Korea, Korea Electric Power Corporation (KEPCO) started implementation of AMR based energy metering system for its industrial customers in 2000.

ElHub (Electricity Hub) will start operation in Norway, and its plan to facilitate the data exchange between market parties in Norway. ElHub aims at a standardization of data access to smart meter data for all eligible parties. This also presents research opportunity related to smart metering data storage and exchange format.

CONCLUSION

Application of information and communication technology (ICT) in the power industry has given rise to smart metering infrastructure (SMI). The smart meter ecosystem is a properly networked system which has two-way communication capabilities. The discussions about smart meter data management and data analytics in energy sector are driven by the smart meter roll-out and the need to enable customers to take energy saving decisions. Therefore, governments of most countries are at least discussing or have already implemented smart meter related policies. This in itself is expected to create a market of €6b+ for all IT vendors. The field offers varied research opportunities and many business organizations have been trying to extract meaningful and actionable insights from the large volumes of complex and high velocity SMI data streams.

Studies conducted on smart meter data show that smart metering system can help in peaking off the electricity demand and encouraging the use of off-grid renewable energy sources, by putting consumers in control of their energy use or by introducing time of use electricity tariff. With the commercialization of smart meters, a huge amount of standardized dataset is available easily, which can be utilized for price and demand forecasting. However, the storage, querying, and mining of such smart meter data streams poses significant challenge. The application of data mining algorithms on smart meter data could help meet both consumers' and power distribution utility's objectives. Consumers can also produce energy via various means such as solar and the wind, as well as distribute excess energy. However, due to the increasing share of renewables and the increasing number of new electricity devices (e.g. electric vehicles and battery storage) on the distribution grid level, power companies need to advance data analytics ability.

The smart meter has started transforming energy consumption as we know it and certainly brings plenty of exciting opportunities for customers and energy companies alike.

REFERENCES

Alabdulkarim, L., & Lukszo, Z. (2008). Information security assurance in critical infrastructures: Smart Metering case. *2008 1st International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future, INFRA 2008*. 10.1109/INFRA.2008.5439670

Albert, A., & Rajagopal, R. (2013). Smart meter driven segmentation: What your consumption says about you. *IEEE Transactions on Power Systems*, *28*(4), 4019–4030. doi:10.1109/TPWRS.2013.2266122

Alzate, C., & Sinn, M. (2013). Improved electricity load forecasting via kernel spectral clustering of smart meters. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 943–948. 10.1109/ICDM.2013.144

Aman, S., Simmhan, Y., & Prasanna, V. K. (2011). Improving energy use forecast for campus micro-grids using indirect indicators. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 389–397. 10.1109/ICDMW.2011.95

Arora, S., & Taylor, J. W. (2016). Forecasting electricity smart meter data using conditional kernel density estimation. *OMEGA-International Journal of Management Science, 59*(A, SI), 47–59. 10.1016/j.omega.2014.08.008

Beckel, C., Sadamori, L., & Santini, S. (2013). Automatic socio-economic classification of households using electricity consumption data. *Journal of Economic Psychology*, *3*(3–4), 75–86. doi:10.1016/0167-4870(83)90006-5

Beckel, C., Sadamori, L., Staake, T., & Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy, 78*(SI), 397–410. 10.1016/j.energy.2014.10.025

Birt, B. J., Newsham, G. R., Beausoleil-Morrison, I., Armstrong, M. M., Saldanha, N., & Rowlands, I. H. (2012). Disaggregating categories of electrical energy end-use from whole-house hourly data. *Energy and Buildings*, *50*, 93–102. doi:10.1016/j.enbuild.2012.03.025

Buchmann, E., Boehm, K., Burghardt, T., & Kessler, S. (2013). Re-identification of Smart Meter data. *Personal and Ubiquitous Computing, 17*(4, SI), 653–662. 10.100700779-012-0513-6

Candanedo, L. M., Feldheim, V., & Deramaix, D. (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, *140*, 81–97. doi:10.1016/j.enbuild.2017.01.083

Cao, H.-A., Wijaya, T. K., Aberer, K., & Nunes, N. (2017). Estimating human interactions with electrical appliances for activity-based energy savings recommendations. In *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016* (pp. 1301–1308). IEEE. 10.1109/BigData.2016.7840734

Chakravorty, A., Rong, C., Evensen, P., & Wlodarczyk, T. W. (2014). A distributed gaussian-means clustering algorithm for forecasting domestic energy usage. *Proceedings of 2014 International Conference on Smart Computing, SMARTCOMP 2014*, 229–236. 10.1109/SMARTCOMP.2014.7043863

Chiky, R., Decreusefond, L., & Hébrail, G. (2010). Aggregation of asynchronous electric power consumption time series knowing the integral. In *Advances in Database Technology - EDBT 2010 - 13th International Conference on Extending Database Technology, Proceedings* (pp. 663–668). Academic Press. 10.1145/1739041.1739122

Dong, M., Meira, P. C. M., Xu, W., & Chung, C. Y. (2013). Non-intrusive signature extraction for major residential loads. *IEEE Transactions on Smart Grid*, *4*(3), 1421–1430. doi:10.1109/TSG.2013.2245926

Dyson, M. E. H., Borgeson, S. D., Tabone, M. D., & Callaway, D. S. (2014). Using smart meter data to estimate demand response potential, with application to solar energy integration. *Energy Policy*, *73*, 607–619. doi:10.1016/j.enpol.2014.05.053

Farhangi, H. (2010). The Path of the Smart Grid 18. *IEEE Power & Energy Magazine*, *8*, 18–28. doi:10.1109/MPE.2009.934876

Fusco, F., Wurst, M., & Yoon, W. J. (2012). Mining Residential Household Information from Low-resolution Smart Meter Data. *21st International Conference on Pattern Recognition (ICPR 2012)*, 3545–3548.

Gajowniczek, K., & Zabkowski, T. (2015). Data mining techniques for detecting household characteristics based on smart meter data. *Energies*, *8*(7), 7407–7427. doi:10.3390/en8077407

Guo, Z., Wang, Z. J., & Kashani, A. (2015). Home Appliance Load Modeling From Aggregated Smart Meter Data. *IEEE Transactions on Power Systems*, *30*(1), 254–262. doi:10.1109/TPWRS.2014.2327041

Hayes, B., Gruber, J., & Prodanovic, M. (2015). Short-Term Load Forecasting at the local level using smart meter data. *IEEE Eindhoven PowerTech*, 1–6. 10.1109/PTC.2015.7232358

Jin, N., Flach, P., Wilcox, T., Sellman, R., Thumim, J., & Knobbe, A. (2014). Subgroup discovery in smart electricity meter data. *IEEE Transactions on Industrial Informatics*, *10*(2), 1327–1336. doi:10.1109/TII.2014.2311968

Kavousian, A., Rajagopal, R., & Fischer, M. (2013). Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy*, *55*, 184–194. doi:10.1016/j.energy.2013.03.086

Kouzelis, K., & Bak-jensen, B. (2014). *A Simplified Short Term Load Forecasting Method Based on Sequential Patterns*. Academic Press.

Kwac, J., Flora, J., & Rajagopal, R. (2014). Household Energy Consumption Segmentation Using Hourly Data. *IEEE Transactions on Smart Grid*, *5*(1), 420–430. doi:10.1109/TSG.2013.2278477

Lavin, A., & Klabjan, D. (2015). Clustering time-series energy data from smart meters. *Energy Efficiency*, *8*(4), 681–689. doi:10.100712053-014-9316-0

Liao, Y.-S., Liao, H.-Y., Liu, D.-R., Fan, W.-T., & Omar, H. (2016). Intelligent Power Resource Allocation by Context-Based Usage Mining. In *Proceedings - 2015 IIAI 4th International Congress on Advanced Applied Informatics, IIAI-AAI 2015* (pp. 546–550). IIAI. 10.1109/IIAI-AAI.2015.165

Madureira, A., Gouveia, C., Moreira, C., Seca, L., & Lopes, J. P. (2013). Coordinated management of distributed energy resources in electrical distribution systems. *2013 IEEE PES Conference on Innovative Smart Grid Technologies, ISGT LA 2013*. 10.1109/ISGT-LA.2013.6554446

Marinescu, A., Harris, C., Dusparic, I., Clarke, S., & Cahill, V. (2013). Residential electrical demand forecasting in very small scale: An evaluation of forecasting methods. *2013 2nd International Workshop on Software Engineering Challenges for the Smart Grid, SE4SG 2013 - Proceedings*, 25–32. 10.1109/SE4SG.2013.6596108

McLoughlin, F., Duffy, A., & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, *141*, 190–199. doi:10.1016/j.apenergy.2014.12.039

Niska, H. (2015). *Evolving Smart Meter Data Driven Model for Short- Term Forecasting of Electric Loads*. Academic Press.

Park, S., Ryu, S., Choi, Y., Kim, J., & Kim, H. (2015). Data-driven baseline estimation of residential buildings for demand response. *Energies*, *8*(9), 10239–10259. doi:10.3390/en80910239

Parson, O., Ghosh, S., Weal, M., & Rogers, A. (2014). An unsupervised training method for non-intrusive appliance load monitoring. *Artificial Intelligence*, *217*, 1–19. doi:10.1016/j.artint.2014.07.010

Quilumba, F. L., Lee, W. J., Huang, H., Wang, D. Y., & Szabados, R. (2014). An overview of AMI data preprocessing to enhance the performance of load forecasting. *2014 IEEE Industry Application Society Annual Meeting, IAS 2014*, 1–7. 10.1109/IAS.2014.6978369

Quilumba, F. L., Lee, W.-J., Huang, H., Wang, D. Y., & Szabados, R. L. (2015). Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities. *IEEE Transactions on Smart Grid*, *6*(2), 911–918. doi:10.1109/TSG.2014.2364233

Savio, D., Karlik, L., & Karnouskos, S. (2010). Predicting energy measurements of service-enabled devices in the future smartgrid. *UKSim2010 - UKSim 12th International Conference on Computer Modelling and Simulation*, 450–455. 10.1109/UKSIM.2010.89

Stankovic, L., Stankovic, V., Liao, J., & Wilson, C. (2016). Measuring the energy intensity of domestic activities from smart meter data. *Applied Energy*, *183*, 1565–1580. doi:10.1016/j.apenergy.2016.09.087

Uribe-Pérez, N., Hernández, L., de la Vega, D., & Angulo, I. (2016). State of the Art and Trends Review of Smart Metering in Electricity Grids. *Applied Sciences*, *6*(3), 68. doi:10.3390/app6030068

Viegas, J. L., Vieira, S. M., & Sousa, J. M. C. (2016). Mining consumer characteristics from smart metering data through fuzzy modelling. *Communications in Computer and Information Science*, *610*, 562–573. doi:10.1007/978-3-319-40596-4_47

Wang, Y., Hao, X., Song, L., Wu, C., Wang, Y., Hu, C., & Yu, L. (2012). Tracking states of massive electrical appliances by lightweight metering and sequence decoding(重点). *Proceedings of the Sixth International Workshop on Knowledge Discovery from Sensor Data*, 34–42. 10.1145/2350182.2350186

Wijaya, T. K. b, Ganu, T., Chakraborty, D., Aberer, K., & Seetharam, D. P. (2014). Consumer segmentation and knowledge extraction from smart meter and survey data. In *SIAM International Conference on Data Mining 2014, SDM 2014* (*Vol. 1*, pp. 226–234). SIAM. 10.1137/1.9781611973440.26

Zala, H. N., & Abhyankar, R. (2014). A novel approach to design time of use tariff using load profiling and decomposition. *2014 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*, 1–6. 10.1109/PEDES.2014.7042027

Zhang, X., Grijalva, S., & Reno, M. J. (2014). A time-variant load model based on smart meter data mining. *IEEE PES General Meeting / Conference & Exposition*, 1–5.

KEY TERMS AND DEFINITIONS

Data Mining: It is practice and process of identifying patterns in large datasets with the help of computing algorithms.

Feature Selection: Identifying attributes or variables relevant for particular prediction model is called as feature selection or attribute selection or variable selection.

K-Means Clustering: This is unsupervised process of partitioning data into k-clusters, so as to minimize the distance between point and nearest cluster mean.

Probability Density Function: A function of a continuous random variable, whose integral across an interval gives the probability that the value of the variable lies within the same interval.

Self-Organizing Maps: It is a data visualization technique which helps to understand high dimensional data by reducing the dimensions of data to a map.

Smart Meter: It is advance meter that displays real time information about electricity consumption and costing to the consumer. It is capable of sending and receiving information using radio waves, like mobile phones and TVs.

Support Vector Machine: It is used to categories new unseen data points into separate groups based on their properties and a set of known examples, which are already categorized. It works by representing the features with the help of finite dimension vector space.