Smart Grid Analytics

Christoph M. Flath and Nikolai Stein

Learning Objectives

- Grasp the basic properties of big data and how they relate to smart grids
- Understand the different facets of business analytics in the context of the power industry
- Obtain a working knowledge of machine learning tasks and algorithms
- Apply analytics knowledge to smart grid use cases

The rapid digitization of the electricity sector has lead to a massive increase of data availability. Smart grids produce large volumes of data, thanks to IoT devices like smart meters. This necessitates appropriate Data Analytics capabilities to tap into the envisioned opportunities (Zhang et al. 2018). Against the backdrop of ubiquitous computing, companies are building up capabilities for data analysis as well as automatic decision-making. Such analytics platforms can analyze data to generate findings that lead to several benefits, like cost reduction and operational efficiency. Varaiya et al. (2011) note that sensors and smart meters will provide system operators with more detailed information on the power system state. This may even lead to several other improvements in grid optimization and customer engagement. With respect to the consumption data, electricity providers can use the revealed information patterns to improve their business processes.

C. M. Flath (✉) · N. Stein
Chair of Information Systems and Business Analytics, Julius-Maximilians-Universität Würzburg, Sanderring 2, 97070 Würzburg, Germany
e-mail: christoph.flath@uni-wuerzburg.de

N. Stein
e-mail: nikolai.stein@uni-wuerzburg.de

C. Dinther et al. (eds.), *Smart Grid Economics and Management*,
Lecture Notes in Energy 51, https://doi.org/10.1007/978-3-030-84286-4_7

The idea of "business analytics" generating business value from data has emerged in the last decade. Depending on the specific focus and impact of the application, one can distinguish descriptive, predictive, and prescriptive analytics (Lustig et al. 2010). Following this classification, *descriptive analytics* encompasses business intelligence tools and processes that use data to understand and analyze past business performance. It focuses on reporting and visualization of historical data and is thus decoupled from future decisions. In contrast to this backward-oriented application, *predictive analytics* aims at uncovering explanatory structures in the data to draw inferences about future instances. The most encompassing application is *prescriptive analytics* which integrates insights from descriptive and predictive analytics to determine appropriate actions or decisions by means of optimization.

Given the increasing adaption of smart grid technologies it is hardly surprising that researchers are underlining the central importance of data mining and analytics techniques for the transformation of the electricity system (Keshav and Rosenberg 2011; Ramos and Liu 2011).

1 Big Data in Smart Grids

Solving problems and answering questions through data analysis has emerged as standard practice across many industries undergoing the digitization trend. The energy industry is no exception to this general trend (Gust et al. 2017). There are numerous rapidly evolving technologies for analyzing data and building models. This section serves to establish fundamental concepts and relate them to the energy industry.

1.1 Properties of Big Data

While the attribute "big" is a very subjective characterization of a data source. However, there is a general consensus that there are four central attributes that characterize big data analytics. Frequently, these are referred to as the four Vs: volume, variety, velocity, and veracity (Demchenko et al. 2014; Dong and Srivastava 2013; Kaisler et al. 2013; Sagiroglu and Sinanc 2013).

The main characteristic that renders a dataset "big" is its size, i.e., *Volume*. It is important to stress that size cannot be referred to in absolute quantities since the information and the capability to process it is growing exponentially every year. Ultimately, the definition of big data depends on whether the data can be ingested, processed, and timely examined to meet an organization's requirements. In smart grids IoT devices are placed in different areas like the substations and consumer devices. These devices collectively produce petabytes of data across various timescales and aggregation levels. In turn, it is not possible to make sense of the data without smart grid analytics capabilities.

Variety is one the most interesting developments in technology as more and more information is digitized. Traditional data types (*structured data*) include things on a bank statement like date, amount, and time. These are things that fit neatly in a relational database. To put it generally, structured data is well-defined, featuring a set of rules which facilitate queries and reasoning on the data. (e.g., a database schema). This stands in stark contrast to *unstructured data*, which includes any data sources that cannot be described by such a meta model. Typical examples include text, audio, and visual information (images or video). This lack of structure of these data assets prevents straightforward analysis and in turn necessitates major pre-processing efforts. Handling unstructured data assets is a fundamental challenge in big data analytics. One of the goals of big data analytics is to use technology to take this unstructured data and make sense of it. In smart grids we encounter both structured and unstructured data. Structured data will typically include various forms of time series data (e.g., consumption, generation, prices) while unstructured data may emerge in the form of textual data (e.g., maintenance reports or customer requests), graph data (e.g., grid topologies or building plans), images (e.g., satellite images). Consequently, comprehensive smart grid analytics capabilities necessitate various forms of big data analysis (Zhang et al. 2018).

With so much data available, ensuring its relevance and quality differentiates between success and failure in big data analytics. This is reflected in the *Veracity* dimension which reflects the reliability of data assets. Can decision-makers trust the data to guide their course of action? In complex systems, various discrepancies may emerge—typical examples include duplicates, outliers, inconsistencies, or volatility. Smart grid nodes like smart meter are numerous and decentral and will in individually be exposed to service interruptions leading to unreliable aggregate data. Consequently, smart grid consumption data is likely to suffer from inconsistencies and missing data which necessitates appropriate detection and mitigation strategies (Chen et al. 2017).

Velocity is the frequency of incoming data that needs to be processed. Considering the sheer number of messages, file uploads, or website interactions taking place on a particular social network site offers a good appreciation of the importance of velocity. In the context of energy markets data velocity behaves in lock-step with the velocity of the underlying market transactions. Day-ahead trading means daily updates while real-time markets will necessitate continuous data exchange. The same reasoning applies to metering infrastructure where decision-makers are striking a balance between the investment cost burden, coordination capabilities, and privacy challenges of different smart metering time resolutions (Eibl and Engel 2014; Feuerriegel et al. 2016; Flath 2013).

1.2 Data Sources in the Smart Grid

Prior to dwelling on analytics applications, it is of central importance to assess the available data assets. This includes the systems involved (data sources) and the

properties of the data streams. A natural approach to structure these data assets is to align them with the electricity value chain (Zhang et al. 2018): Generation → Transmission and Distribution → Consumers. Thereby we highlight both data sources as well as potential data analytics use cases.

Generation Data

In the narrowest sense power generation data can be distinguished into output time series and capacity state information (location, technology, ramping characteristics) with varying aggregation level (ranging from generation unit to transmission system). Such information is key for any system-related analysis concerning, e.g., market prices, emissions, or availability. Notably, generation data pertains to all dimensions of the energy trilemma cost-sustainability-reliability (Heffron et al. 2015). In particular, the increased adoption of RES has resulted in previously unknown levels of short-term output fluctuations and volatility. Consequently, short-term power generation forecasts have become much more important. These play a key role in operational decisions of conventional power plants, inform trading decisions, as well as triggering demand response preparations. Given the strong correlation between RES output and the corresponding weather phenomena the relevant metrics can also be considered important generation data (e.g., solar irradiance, wind speed, water levels). More recently, researchers have even started to consider building orientation, shading conditions or roof surfaces to assess generation potentials (Fitriaty and Shen 2018).

Grid Data

Building upon a collaboration with a Swiss utility company Gust et al. (2016) highlight the different sources and forms of electricity grid data. On the one hand there is topographic and spatial data which is typically accessed through a geographic Information System (GIS). On the other hand there is investment and equipment information recorded in corporate asset databases embedded enterprise resource planning (ERP) systems. GIS data typically includes component locations of grid assets as well as information on the service area (e.g., infrastructure such as buildings and roads). Asset databases provide detailed information on asset types, their electrical specifications and thereby serve as a master data provider. Furthermore, investment cost and depreciation schedules are provided by ERP systems.

Consumption Data

With annual meter readings for many consumers electricity consumption data has traditionally been the blind spot of electricity systems. To compensate the absence of live data, planners and operators have relied on statistical models and static load profiles. These are primarily based on contracting data and the customers' historic consumption quantities (per billing period). Yet, in the last decade the idea of smart metering has increasingly challenged this myopic aggregation of (annual) consumption values. This new addition to the energy policy toolbox can have a marked effect on household energy management in turn creating change agents for the challenge of reducing (peak) demand and the integration of renewable generation (Darby 2010).

1.3 From Descriptive to Prescriptive Analytics

Analytics is defined by INFORMS as "the scientific process of transforming data into insight for making better decisions" (INFORMS 2020). With the rise of big data, analytics has received more attention. The most common categorization in the field is the distinction between descriptive, predictive, and prescriptive analytics (Lustig et al. 2010):

- **Descriptive analytics** relies on large amounts of historical data to identify patterns and to illustrate past behavior (den Hertog and Postek 2016). Here, a data set S^N containing information on quantities of interest (e.g., demand) $Y^N = \{\mathbf{y}^1, \ldots, \mathbf{y}^N\}$ as well as auxiliary data on associated covariates (features) $X^N = \{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$ where $\mathbf{x}^t$ is concurrently observed with $\mathbf{y}^t$ are aggregated and analyzed. These observations are used for reporting, dashboards, and data visualizations. These applications are summarized in the umbrella term Business Intelligence (Watson 2014; Delen and Ram 2018). Business Intelligence systems automatically retrieve and process data from various data sources to support decision-making.
- **Predictive analytics** aims at informing decision-makers by predicting future scenarios (Corea 2016). To this end, we seek to use statistical learning to find a function $h^*(\mathbf{x}) \in H$ that minimizes a given (typically symmetric) loss function $\mathscr{L}$. In predictive analytics this loss is broadly speaking a measure of the predictive power (the smaller the loss the better the predictiveness). The set of permissible functions H is governed by the selected learning algorithm:

$$h^*(\mathbf{x}) = \underset{h(\mathbf{x}) \in H}{\arg\min}\, \mathscr{L}\left(h(\mathbf{x}), \mathbf{y}\right). \tag{1}$$

  Decision-makers use predictive analytics to generate forecasts on future scenarios $s = h^*(\mathbf{x})$ that allow them to take informed decisions even in uncertain situations.
- **Prescriptive analytics** aims at providing decision support or decision automation. In many applications the way to derive optimal policies from forecasts is not straightforward. Here, prescriptive analytics has the potential to build on operations management techniques such as mathematical optimization and decision rules to determine optimized policies. To achieve this, one no longer chooses the loss function ℓ such that it tracks predictive power but rather it emerges from the business problem (e.g., lost profit or additional cost) underlying a decision $\mathbf{q}$. One way to establish such a loss function is by comparing the outcome of a prescribed action in a given scenario to the theoretically optimal profit $\Pi^*(\mathbf{q}^*, \mathbf{y})$:

$$\ell\left(\mathbf{q}, \mathbf{y}\right) = \Pi^*(\mathbf{q}^*, \mathbf{y}) - \Pi(\mathbf{q}, \mathbf{y}). \tag{2}$$

  One can than use the forecasts obtained from predictive analytics as input parameters for optimization models to find optimal policies sequentially. Alternatively, ML algorithms can be used to find a function $q^*(\mathbf{x})$ that maps directly from data to decisions by minimizing the expected lost profit (Eq. 2):

$$q^*(\mathbf{x}) = \underset{q(\mathbf{x})\in H}{\arg\min} \frac{1}{N} \sum_{(\mathbf{x},\mathbf{y})\in S} \ell\left(q(\mathbf{x}), \mathbf{y}\right). \tag{3}$$

In summary, predictive analytics applications deal with forecasting future scenarios whereas prescriptive analytics prescribes a decision in anticipation of the future (Bertsimas and Kallus 2020).

2 Machine Learning Fundamentals

Many algorithms and methods for big data analytics (predictive as well as prescriptive) originate from the field of machine learning. Such algorithms do not rely on hard-coded rules but instead *learn* the underlying patterns and regularities in a data-driven fashion from provided training data. By and large this "learning activity" corresponds to an optimization of the underlying loss function.

In this context one typically distinguishes two distinct learning tasks: unsupervised learning and supervised learning (Fig. 1). As the name suggests, unsupervised learning is machine learning tasks in which humans do not need to supervise the model. With respect to the training data unsupervised learning deals with unlabeled data, i.e., there are no annotations of a possible target variable. Instead, the model works independently to discover information and patterns that were previously undetected. In contrast, supervised learning is the task of inferring a function from labeled training data. Here, each training example is a pair consisting of an input object (in most cases a vector) and a output value. Problems with a continuous output space are further categorized under the term regression problems while classification describes problems with a discrete output space. In between these archetypes there are some further learning paradigms such as semi-supervised learning or reinforcement learning.[1] However, in the context of this book we focus on the two basic cases.

2.1 Unsupervised Learning

Unsupervised learning summarizes machine learning algorithms that aim at revealing hidden structures in unlabeled data (Hastie et al. 2009). The main goal of these algorithms is either dimensionality reduction or cluster analysis.

Dimensionality Reduction

Working in high-dimensional spaces can be undesirable for different reasons—in particular processing time and storage space requirements, multi-collinearity issues,

[1] Semi-supervised learning uses a small amount of labeled data bolstering a larger set of unlabeled data. In a reinforcement learning setting the system is not trained on labels but instead on rewards resulting from interaction with the environment.

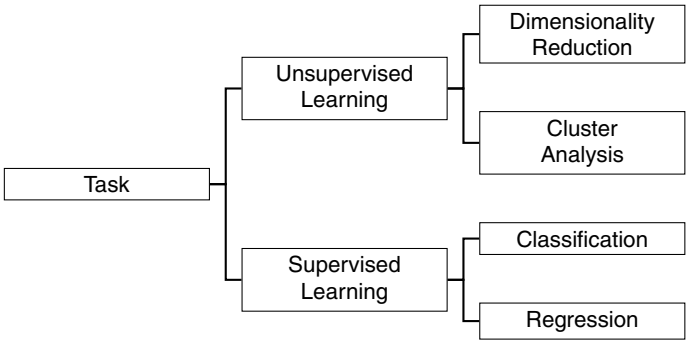

Fig. 1 Taxonomy of machine learning tasks

difficulty of visualizing and communicating data beyond two or three dimensions, and avoiding the curse of dimensionality. For these reasons dimensionality reduction is a standard step in data pre-processing. It is a transformation of data from a high-dimensional space into a lower dimensional space where the reduced representation retains the key properties of the original data. The principle component analysis (PCA) is a popular and well-studied method to transform high-dimensional datasets into low-dimensional datasets. PCA converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. To this end, it finds the n principal axes in the original m-dimensional space where the variance between the points is the highest. By selecting the axes that explain most of the variance, the number of variables is reduced from m to n. Thereby, the bulk of information is preserved as the new variables are combinations of the old variables (Hotelling 1933). However, PCA reaches its limitations if the relationships between the variables are non-linear. This shortcoming is tackled by more recent approaches (Maaten and Hinton 2008).

Clustering and Segmentation

Cluster Analysis is used to group and identify data-inherent structures. These algorithms assign data points to limited number of clusters to group similar data points. The basic premise of the clustering optimization task can be summarized as follows: data points assigned to the same cluster should be as similar as possible (within-cluster homogeneity) while data points from different clusters should not be similar (across-cluster heterogeneity). From this requirement it becomes obvious that one needs two components to formulate the underlying optimization problem:

- A suitable distance function that allows to measure the similarity of two data points (greater similarity corresponds to a shorter distance).
- A loss function that aggregates the distance measurements to evaluate the quality of the cluster configuration as a whole.

Any suitable metric can be used as a distance function. Typical candidates include Euclidean distance or cosine similarity. A straightforward measure for within-cluster

homogeneity h^w is given by the following expression:

$$h^w = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \, ||\mathbf{x}_n - \mu_k|| \, . \tag{4}$$

Here, N is the number of data points, K is the number of clusters, $||\cdot||$ is the chosen metric, and μ_k is the central point of cluster k (centroid). The binary variable r_{nk} indicates whether or not n is in cluster k. Between cluster heterogeneity h^b is easily measured by adjusting the h^w formula which measures the total distance of the cluster members to the centroid.

$$h^b = \sum_{i=1}^{N} \sum_{k=1}^{K} (1 - r_{nk}) \, ||\mathbf{x}_i - \mu_k|| \tag{5}$$

.

Overall clustering quality CQ can then quantified by the ratio of the two aggregates:

$$CQ = \frac{h^w}{h^b}. \tag{6}$$

Note that a lower CQ score indicates a better clustering result in the sense of a low h^w value (high within-cluster homogeneity) and a high h^b value (high between-cluster heterogeneity).

There are many different algorithms available in modern software packages to execute cluster analysis task. Depending on the clustering process, we can distinguish between hierarchical algorithms (Ward Jr 1963; Murtagh and Contreras 2012) and partitioning algorithms (MacQueen et al. 1967; Kaufman and Rousseeuw 2009). In the case of hierarchical clustering the various clusters form a nested hierarchy (individual clusters are subsets of their super-ordinate clusters). Conversely, partitioning algorithms yield non-overlapping clusters.

2.2 Supervised Learning

Supervised learning is the machine learning task of inferring a function from labeled training data (Mohri et al. 2012). In the typical supervised learning setting we have to predict unknown future variables (e.g., energy load, renewable generation) based on past observations to make informed decisions. While experienced decision-makers may have a good intuition on some factors (e.g., weather, holidays) influencing the unknown variable, it is hard for humans to quantify the impact of all possible factors. Supervised learning approaches provide a methodological way of quantifying such influences. Depending on the underlying quantity we distinguish between *regression* and *classification*:

Table 1 Confusion matrix for the evaluation of binary classifier

		Prediction: true	Prediction: false
Realization	true	true pos. TP	false neg. FN
	false	false pos. FP	true neg. TN

Regression

These problems aim to learn a mathematical relation $f : \mathbb{R}^n \mapsto \mathbb{R}$ explaining a continuous dependent output variable y in terms of n independent variables $\mathbf{x}$. The function f allows us to obtain an estimate for the dependent variable from its associated independent variables $\hat{y}_i = f(\mathbf{x}_i)$. In the supervised learning context the loss function in the training data is obtained from comparing estimates $\hat{y}_i$ with the true values y_i. Typical candidates include the mean squared error (MSE) $\sum_i \left(\hat{y}_i - y_i\right)^2$ or the mean absolute deviation (MAD) $\sum_i \left|\hat{y}_i - y_i\right|$. Typical regression use cases include forecasting of loads, RES generation, and power exchange prices.

Classification

Many quantities of interest (e.g., fault occurrence in smart grid systems, identification of household loads) cannot be represented by continuous output variables. In such classification problems, we want to predict an observation's class y based on the feature variables $\mathbf{x}$. Here, we try to find a function $f : \mathbb{R}^n \mapsto \{1, \ldots, k\}$ where n describes the number of features and k the number of different classes. The evaluation of classification performance is best illustrated using a confusion matrix for binary classification (Table 1). This matrix compares the predictions of a classifier against the known realizations. Depending on the prediction and the realization there are four possible outcomes—true positive (TP, the classifier correctly predicts a true), true negative (TN, the classifier correctly predicts a false), false positive (FP, the classifier incorrectly predicts a true), and false negative (FN, the classifier incorrectly predicts a false).[2]

Based on the confusion matrix we can derive some standard quality metrics of the classifier:

[2] For non-binary classification the confusion matrix generalizes in a straightforward manner to a $n \cdot n$ square where n is the number of cases.

- Accuracy—the percentage of correct predictions $ACC = \frac{TP+TN}{TP+FP+TN+FN}$
- True positive rate (sensitivity)—the proportion of positive cases that are correctly identified $TPR = \frac{TP}{TP+FN}$
- True negative rate (specificity)—the proportion of negatives cases that are correctly identified $TNR = \frac{TN}{TN+FP}$
- Balanced Accuracy—the average of sensitivity and specificity which provides a better overall assessment of classifier quality in settings with unbalanced data $\frac{TPR+TNR}{2}$
- Matthews Correlation Coefficient—As suggested this coefficient is indeed a correlation coefficient between the observed and predicted binary classifications. It is generally regarded as a balanced measure which can be used even if the classes are of very different sizes. $\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$.

Given the discreteness of the confusion matrix these quality metrics are typically non-differentiable which makes them unsuitable for serving as a machine learning loss. Instead, one typically leverages probability space and uses appropriate transformations like logistic or hinge loss.

Classification tasks in the power sector include asset reliability monitoring and prediction, customer churn analysis, or identification of appliances from meter readings.

Algorithms

We can distinguish between parametric and non-parametric machine learning algorithms to perform classification and regression tasks. Parametric algorithms are efficient to train and easy to interpret as they make strong assumptions on the form of the function. Therefore, this class of models is also referred to as white-box models. For regression tasks, multiple linear regression is the most prominent white-box algorithms. This method was established by Galton (1886) over 200 years ago and is probably the best studied form of statistical learning (Hastie et al. 2009). It rests on the assumption of a linear relationship between a set of input variables and a single output variable. The linearity assumption is simultaneously the major strength and weakness of this method. On the one hand, it renders the model very simple to understand and efficient to learn as well as robust to outliers. On the other hand, it constrains the predictive power of the model as many statistical relations are non-linear.

In contrast, non-parametric models do not make strong assumptions about the relationship between independent and dependent variables. Therefore, they are free to learn any functional form from the training data. While this flexibility oftentimes allows non-parametric algorithms to achieve a higher performance, more historical training data is required to avoid over-fitting. Non-parametric machine learning algorithms are also referred to as black-box models as the potentially higher performance comes at the cost of interpretability. The SVM is a prominent example of non-parametric algorithms. Here, each observation is viewed as a vector of all input variables. During the model training the hyper-plane that best separates the different output variables depending on the input vectors is determined. To incorporate non-linear relationships, the dimensionality of the input vector is augmented using

the kernel-trick if no hyper-plane separating all different output variables exists. A major benefit of support vector machines is their good generalization ability and therefore the low susceptibility to over-fitting even for small training data sets (Smola and Schölkopf 2004). In recent years, deep neural networks have shown remarkable performance across different tasks and therefore attracted attention in research and practice. This class of models is inspired by the distributed communication and information processing in biological systems. These models consist of several layers of artificial neurons. Each neuron is connected with many other neurons and processes incoming information and propagates the results to other neurons. In the third class of black-box model we introduce gradient boosting machines. This model class combines many weak learners to a strong predictive model in a sequential fashion. This way it is able to alleviate some of the shortcomings of the weak models while increasing the predictive power. On the downside, gradient boosting is prone to over-fitting (Hastie et al. 2009).

3 Application Examples

3.1 Load and Generation Forecasting

Hong and Fan (2016) from the very beginning of the electric power industry the problem of load forecasting has been central to the business. This is of course due to the fundamental balancing requirements of the power system. Hence, power output from conventional power plants was planned in accordance with forecast load. With increasing levels of generation from fluctuating RES an additional challenge in the energy mix is the stochasticity of the supply side. This results in the novel challenge of net load forecasting, that is the difference between intermittent renewable generation and fluctuating load (Wang et al. 2017). This quantity ultimately determines whether measures aimed at decreasing net load are needed (additional generation capacities have to be ramped, shifting of flexible loads) or alternatively measures to increase net load are to be performed (charging of storage facilities, expanding flexible consumption, ramping down generators).

Time Series Analysis Forecasting

The classic forecasting approach in power systems management is time series analysis (Amjady 2001; Espinoza et al. 2005). Time series analysis leverages past observations of a variable to predict its future values. A time series in this context is a sequence of values over time (daily, monthly,...) of a quantitative variable, $\langle y_t, y_{t-1}, y_{t-2}, ... \rangle$. Future realizations of the quantity of interest are assumed to be a function of the past realizations and some white noise random error term. Consequently future values can be predicted is $\hat{y}_t = f(y_{t-1}, y_{t-2}, y_{t-3} \ldots)$. The functional mapping should capture typical underlying patterns such as trends (growth or decline) and seasonalities. The classic Holt-Winters triple exponential smoothing approach provides a powerful yet

tractable framework for such problems (Chatfield 1978). The underlying logic of this procedure is decomposition of the time series into three distinct exponentially smoothed units, the level term E_t, the trend term T_t and the seasonality term S_t:

$$\hat{y}_{t+n} = (E_t + nT_t)\, S_{t+n-p}, \tag{7}$$

where

$$E_t = \alpha \left(\frac{Y_t}{S_{t-p}} \right) + (1 - \alpha)\, (E_{t-1} + T_{t-1}) \tag{8}$$

$$T_t = \beta\, (E_t - E_{t-1}) + (1 - \beta)\, T_{t-1} \tag{9}$$

$$S_t = \gamma \left(\frac{Y_t}{E_t} \right) + (1 - \gamma)\, S_{t-p} \tag{10}$$

Using this system of variables it is straightforward to choose the three parameters such that the training forecasting error is minimized. This approach is particularly suited for short-term forecasting applications (Amjady 2001). Forecasting using time series models has a long tradition in power systems management and is a robust and interpretable approach. A key disadvantage is the lack of non-time series features as these models rely solely on the information embedded in the historic time series but cannot readily incorporate other data sources. Furthermore, the relationships are restricted to the time series model and hence fairly restrictive (in particular with respect to functional relationships or the number of seasonal terms).

Machine Learning for Forecasting

The limitations of the time series analysis lend themselves to advanced analytics approaches. From a ML perspective forecasting tasks boil down to straightforward regression tasks where the target variable y_t corresponds to the load, generation, or net load at time t. Independent of the concrete use case typical features $\mathbf{x_t}$ include weather information, lagged y values (time series features), pricing data, or customer information. These are readily incorporated into standard algorithms. A particularly illustrative one is decision trees where the feature space is partitioned into groups through a sequence of (typically) binary decisions. An example in the load forecasting context is provided in Fig. 2. This example highlights how decision trees naturally incorporate rich relationships across Various features: the main driver of load in this case is the weather with high temperatures leading to a high load forecast (2,800MW, e.g., due to air conditioning) while in the case of lower temperatures other criteria like type of day or season of the year are embedded in the tree as well leading to differentiated forecasts (2,550MW for summer weekend, 2,300MW for other weekends, and 2,600MW for a working day).

Constructing a decision tree rests on a very simple procedure: Divide data using a beforehand chosen splitting rule into disjoint subsets. Repeat this algorithm recur-

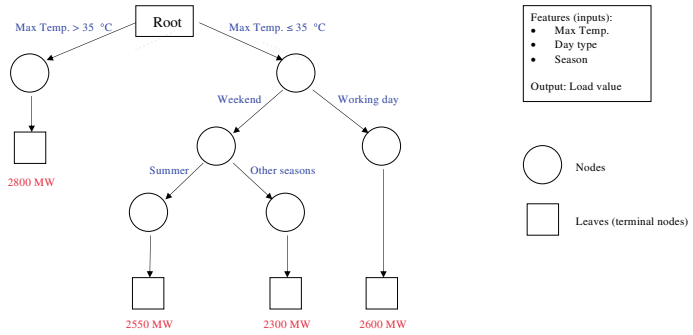

Fig. 2 Regression tree for load forecasting (Lahouar and Ben Hadj Slama 2015)

sively for each subset, stop the recursion when the leaves of the given subset are sufficiently pure.[3] Despite offering high interpretability decision trees suffer from a pronounced tendency to overfit to training data leading to subpar out of sample performance. To avoid this regularization techniques like limiting the tree depth or the number of leaves can be employed.

Going beyond a single tree ensemble methods which combine multiple decision trees helps to greatly improve the predictive performance. The two main ensembling approaches are *bootstrap aggregation* (bagging) and *boosting*:

- A bagging ensemble trains many parallel decision trees (resulting in so-called random forests Breiman 2001). The different trees are trained on datasets which are created from the original data through sampling with replacement. This parallelism increases the reliability of the predictions.
- A boosting ensemble trains many decision trees in sequence. Each tree is trained to minimize the loss taking into account the combined and weighted decision of all preceding trees. This sequence of weak learners is capable of greatly improving the overall result of the decision tree forecast (Chen and Guestrin 2016).

Recently, these machine learning approaches have seen ready adoption in the power systems literature (Persson et al. 2017; Su and Liu 2018; Guo et al. 2021).

Future Challenges in Load Forecasting

Historically, most models focused on point load forecasting. This was because there was sufficient slack in the system (due to controllable generation) which facilitated reliable matching between supply and demand. Recently, though network management has become more challenging given recent developments in the sector. Among others, probabilistic load forecasting is better suited to serve utilities companies in this challenging task (Hong and Fan 2016). This is another advantage of machine learning techniques which ultimately are inherently probabilistic. Given sufficient

[3] Purity can, for example, be measured using the misclassification error in a leaf. Alternative metrics include the Gini coefficient as well as the entropy.

training data there have also been successful applications of deep neural networks, in particular recurrent neural networks (Wang et al. 2019).

3.2 Clustering Load Profiles

The widespread deployment of smart metering has made household energy consumption data more available. This has paved the way for analyzing load characteristics using this advanced metering data. Such power usage data can be leveraged by system operators or energy retailers to better understand customer behavior. This information may help tailor different DR programs for different households in order to build a more robust grid design system, offer more effective energy reduction recommendations, and improve smart pricing models (Salah et al. 2017). These new sources of business value are crucial as smart meters are more costly than traditional metering equipment as they require large investments in the accompanying ICT infrastructure. Consequently, utilities companies are intensively looking for ways to recapture these investments. Faruqui et al. (2010) remark that utilities can profit from improved grid operations, lower costs for obtaining meter readings, or faster identification of outages and interruptions. Smart metering data can facilitate a customer segmentation based on dynamic load patterns instead of mere load totals. Flath et al. (2012) present a practice-oriented introduction to smart meter cluster analysis which we want to quickly outline this user case of unsupervised learning here.

Time Series Preparation

To be able to cluster a collection of smart meter time series (load profiles) we need to bring these data into a usable form. Traditional customer segments often focus on the total consumption as this data has always been readily available from billing systems. Clearly, this corresponds to a very coarse definition of similarity. A smart meter data series with a typical resolution of 15 min resolution provides $4 \cdot 24 \cdot 365 = 35,040$ data points per household. The central question is the underlying clustering objective, i.e., what measures similarity between two load profiles. To compare two load profiles side-by-side it is first necessary to determine the comparison scope and then obtain corresponding representative load profiles. Flath et al. (2012) put forward daily (differentiated by weekday and weekend days) and weekly profiles as natural candidates. In either case the quarter-hourly raw data are aggregated at the chosen level by means of averaging. This results in load profiles with 96 entries for daily and 672 entries for weekly profiles—for weekly profiles it may hence be advisable for weekly profiles to be based on coarser time resolution. To obtain scale-free comparability across load profiles one can normalize the values by division with the highest individual element.

The representative load profiles offer compact representation of the customer consumption records. Pairwise comparison of profiles requires furthermore a suitable metric. Given the time series character of the load data we can simply use a default metric such as the Euclidean metric.

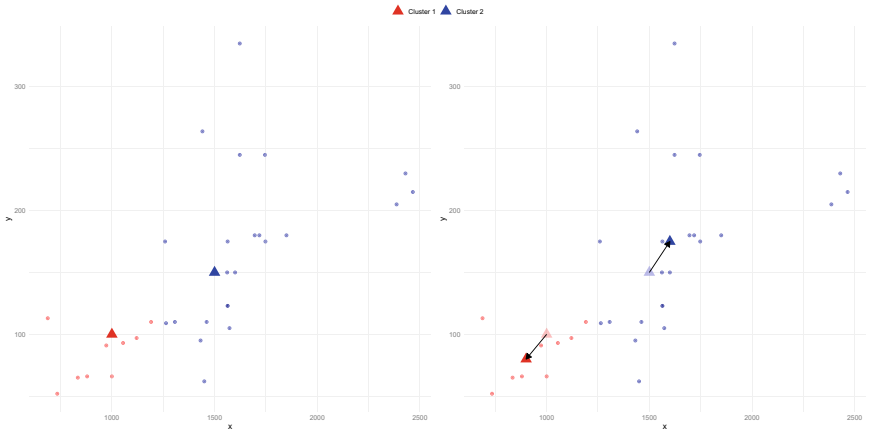

Fig. 3 Illustration of k-means clustering approach

Clustering Procedure

Equipped with representative load profiles and a suitable metric one can proceed to the actual clustering process. A very simple and commonly used algorithm for cluster analysis is the k-means method. This algorithm requires a pre-defined number of clusters, k. It then proceeds by picking k representative load profiles as cluster centers and subsequently assigns all remaining load profiles to the closest cluster. Subsequently, the cluster centers are recalculated and the cluster assignments are updated. This procedure is repeated until there are no more assignment changes which then yields the final clustering. See Fig. 3 for an illustration of the first two steps of the algorithm.

Given the random initialization and the greedy improvement procedure k-means can sometimes end up in sub-optimal clusterings. This issue can be mitigated by trying out multiple starting configurations. The other inherent issue of k-means is the necessity of specifying the number of clusters ex-ante. This can be overcome with additional ex-post metrics of cluster quality to compare across different k-values. Note that for this comparison the number of clusters has to be included as an additional distance penalty as the plain clustering quality is strictly increasing in the number of clusters. Note that oftentimes determining an "optimal" number of clusters is a somewhat theoretical question as practical considerations (cost of marketing additional products, marketing strategy) may anyway impose a fixed value for k.

Advanced Clustering Techniques

The above example is deliberately used the k-means approach as it is the most common and approachable technique for cluster analysis. However, it is somewhat limited with respect to unpacking underlying behavioral aspects of the consumers. To this end, Albert and Rajagopal (2013) propose using a hidden Markov model

to understand the inter-temporal consumption dynamics to appropriately segment the user population. They infer occupancy states from consumption time series data and characterize occupancy by magnitude, duration, and variability. Besides its non-probabilistic nature, the k-means approach also struggles to incorporate other data sources than the consumption time series as pointed out by Wang et al. (2018).

Putting Customer Segments to Use

More granular load profiles provide ample benefit potentials for the various stakeholders in smart grids (Beckel et al. 2014). For example, energy companies can better manage their assets and customer relationships. In particular, the cluster assignment can be a powerful feature in the context of load forecasting as described above. Furthermore, marketing activities can be better targeted with respect to both customers targeting as well as product design. Customers themselves may receive better recommendations and comparisons which may improve their consumption behavior. Finally, regulators may be able to turn richer insights into customer behavior into more effective legislation and rules for the energy market.

3.3 Non-intrusive Load Monitoring

A central theme of smart grids is the empowerment of more local units, i.e., pushing planning and control paradigms from the TSO to the DSO level. This coincides with greater visibility on individual consumers. Yet, these customers remain are still represented as a load profile which of course is a poor representation of the underlying energy end use. This is where load disaggregation techniques such as Non-intrusive load monitoring (NILM) come into play. NILM processes meter data in order to determine exactly what appliances have used the power and how much power each appliance has used during that time period (Hamid et al. 2017). This challenging class of smart grid analytics problems features elements of supervised multi-class classification (identification known appliance traces in an aggregate load profile) as well as unsupervised discovery of previously unseen loads. Weiss et al. (2012) highlight the particular challenge of reliably detecting inductive or capacitive loads. Unsurprisingly, research has so far been driven by the technology and less so by the resulting business opportunities (Welikala et al. 2017). It is also the main application area of advanced AI techniques as the sequence-based and super positional nature of aggregated load profiles is distinctly non-structured. Consequently, ANN approaches such as sequence (Zhang et al. 2018) or transfer learning (D'Incecco et al. 2019) have recently been used to tackle this problem.

If the NILM problem can be reliably solved the opportunities for future smart grid management would be significant. On the customer end there would be a great improvement of information transparency with regards to individual energy consumption. This may help identify out-dated appliances or bad consumption habits. Companies seeking to establish comprehensive DR programs could reliably target customers with certain appliance types or usage patterns. Similarly, customer

response and hence the effectiveness of RTP or ToU rate offerings could greatly benefit from knowing what customers actually do with the procured energy.

4 Summary

As discussed throughout this book the smart grid is a transformation of the power system both from the technological as well as the organizational perspective. Smart grid analytics assumes a similar perspective by highlighting the necessity of first turning data sources into information assets and subsequently leveraging those to generate new business value.

The increasing availability of smart grid data has lead to significant research activity in the smart grid analytics domain. However, so far a lot of research has focused on descriptive and predictive modeling tasks. Going forward one could expect a closer integration of these predictive ML models with smart grid optimization problems (e.g., Chen et al. 2013; Gärttner et al. 2018; Gust et al. 2021) to establish a truly data-driven prescriptive decision-making paradigm. An early pioneer in this area has been Rudin et al. (2011) who cooperated with a New York City utility company to schedule maintenance crews based on reliability predictions (predictive maintenance). The adoption of the weighted sample average approximation (Bertsimas and Kallus 2020) may provide a promising framework for tackling the prescriptive challenge in smart grid analytics more broadly.

NILM research has presented an avenue for deep learning applications. Another area where these deep neural networks are of particular relevance in smart grids is image recognition which can be used to extract information on grid assets (type, size or condition) or customer demand from aerial or satellite imagery (Yan et al. 2007; Malof et al. 2015; Gazzea et al. 2021).

Review Question

- Name the four Vs of big data and elaborate on each of them in the context of smart grid analytics.
- Explain the difference between prescriptive and predictive analytics in your own words.
- Random forests are often considered to be unpractical for large data sets due to their high memory usage. Explain this critique by means of a stylized example (1GB training data, 10 trees).
- Why is the CQ metric not useful for choosing the number of clusters?
- Discuss how electricity pricing and rate design can benefit from smart grid analytics.
- What type of machine learning task do the examples for image recognition correspond to?

References

Albert A, Rajagopal R (2013) Smart meter driven segmentation: what your consumption says about you. IEEE Trans Power Syst 28(4):4019–4030

Amjady N (2001) Short-term hourly load forecasting using time-series modeling with peak load estimation capability. IEEE Trans Power Syst 16(3):498–505

Beckel C, Sadamori L, Staake T, Santini S (2014) Revealing household characteristics from smart meter data. Energy 78:397–410

Bertsimas D, Kallus N (2020) From predictive to prescriptive analytics. Manag Sci 66(3):1025–1044

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Chatfield C (1978) The holt-winters forecasting procedure. J R Stat Soc Ser C (Appl Stat) 27(3):264–279

Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, KDD '16, pp 785–794

Chen W, Wang X, Petersen J, Tyagi R, Black J (2013) Optimal scheduling of demand response events for electric utilities. IEEE Trans Smart Grid 4(4):2309–2319

Chen W, Zhou K, Yang S, Wu C (2017) Data quality of electricity consumption data in a smart grid environment. Renew Sustain Energy Rev 75:98–105

Corea F (2016) What data science means to the business. In: Big data analytics: a management perspective, vol 21. Springer International Publishing, Cham, pp 5–17

Darby S (2010) Smart metering: what potential for householder engagement? Build Res Inf 38(5):442–457

Delen D, Ram S (2018) Research challenges and opportunities in business analytics. J Bus Anal 1(1):2–12

Demchenko Y, De Laat C, Membrey P (2014) Defining architecture components of the big data ecosystem. In: 2014 international conference on collaboration technologies and systems (CTS). IEEE, pp 104–112

den Hertog D, Postek K (2016) Bridging the gap between predictive and prescriptive analytics-new optimization methodology needed. Tilburg Univ, Tilburg, The Netherlands

D'Incecco M, Squartini S, Zhong M (2019) Transfer learning for non-intrusive load monitoring. IEEE Trans Smart Grid 11(2):1419–1429

Dong XL, Srivastava D (2013) Big data integration. In: 2013 IEEE 29th international conference on data engineering (ICDE). IEEE, pp 1245–1248

Eibl G, Engel D (2014) Influence of data granularity on smart meter privacy. IEEE Trans Smart Grid 6(2):930–939

Espinoza M, Joye C, Belmans R, De Moor B (2005) Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. IEEE Trans Power Syst 20(3):1622–1630

Faruqui A, Harris D, Hledik R (2010) Unlocking the 53 euro billion savings from smart meters in the EU: how increasing the adoption of dynamic tariffs could make or break the eu's smart grid investment. Energy Policy 38(10):6222–6231

Feuerriegel S, Bodenbenner P, Neumann D (2016) Value and granularity of ict and smart meter data in demand response systems. Energy Econ 54:1–10

Fitriaty P, Shen Z (2018) Predicting energy generation from residential building attached photovoltaic cells in a tropical area using 3d modeling analysis. J Clean Prod 195:1422–1436

Flath CM (2013) An optimization approach for the design of time-of-use rates. In: IECON 2013-39th annual conference of the ieee industrial electronics society. IEEE, pp 4727–4732

Flath CM, Nicolay D, Conte T, van Dinther C, Filipova-Neumann L (2012) Clusteranalyse von Smart-Meter-Daten. Wirtschaftsinformatik 54(1):1–10

Galton F (1886) Regression towards mediocrity in hereditary stature. J Anthropol Inst G B Irel 15:246–263

Gärttner J, Flath CM, Weinhardt C (2018) Portfolio and contract design for demand response resources. Eur J Oper Res 266(1):340–353
Gazzea M, Pacevicius M, Dammann DO, Sapronova A, Lunde TM, Arghandeh R(2021) Automated power lines vegetation monitoring using high-resolution satellite imagery. IEEE Trans Power Deliv
Guo W, Che L, Shahidehpour M, Wan X (2021) Machine-learning based methods in short-term load forecasting. Electr J 34(1):106884
Gust G, Brandt T, Mashayekh S, Heleno M, DeForest N, Stadler M, Neumann D (2021) Strategies for microgrid operation under real-world conditions. Eur J Oper Res 292(1):339–352
Gust G, Flath C, Brandt T, Ströhle P, Neumann D (2016) Bringing analytics into practice: evidence from the power sector. In: Proceedings of the 37th international conference on information systems
Gust G, Neumann D, Flath CM, Brandt T, Ströhle P (2017) How a traditional company seeded new analytics capabilities. MIS Q Exec 16(3):215–230
Hamid O, Barbarosou M, Papageorgas P, Prekas K, Salame C (2017) Automatic recognition of electric loads analyzing the characteristic parameters of the consumed electric power through a non-intrusive monitoring methodology. Energy Procedia 119:742–751
Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning, 2nd edn
Heffron RJ, McCauley D, Sovacool BK (2015) Resolving society's energy trilemma through the energy justice metric. Energy Policy 87:168–176
Hong T, Fan S (2016) Probabilistic electric load forecasting: a tutorial review. Int J Forecast 32(3):914–938
Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24(6):417
INFORMS (2020) Best definition of analytics
Kaisler S, Armour F, Espinosa JA, Money W (2013) Big data: issues and challenges moving forward. In: 2013 46th Hawaii international conference on system sciences. IEEE, pp 995–1004
Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis, vol 344. Wiley
Keshav S, Rosenberg C (2011) How internet concepts and technologies can help green and smarten the electrical grid. ACM SIGCOMM Comput Commun Rev 41(1):109–114
Lahouar A, Ben Hadj Slama J (2015) Day-ahead load forecast using random forest and expert input selection. Energy Convers Manag 103:1040–1051
Lustig I, Dietrich B, Johnson C, Dziekan C (2010) The analytics journey. Anal Mag, pp 11–13
Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. J Mach Learn Res 9:2579–2605
MacQueen J, et al (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Number 14. Oakland, CA, USA, , pp 281–297
Malof JM, Hou R, Collins LM, Bradbury K, Newell R (2015). Automatic solar photovoltaic panel detection in satellite imagery. In: 2015 international conference on renewable energy research and applications (ICRERA). IEEE, pp 1428–1431
Mohri M, Rostamizadeh A, Talwalkar A (2012) Foundations of machine learning. MIT press
Murtagh F, Contreras P (2012) Algorithms for hierarchical clustering: an overview. Wiley Interdiscip Rev Data Min Knowl Discov 2(1):86–97
Persson C, Bacher P, Shiga T, Madsen H (2017) Multi-site solar power forecasting using gradient boosted regression trees. Solar Energy 150:423–436
Ramos C, Liu C (2011) Ai in power systems and energy markets. IEEE Intell Syst 26(2):5–8
Rudin C, Waltz D, Anderson RN, Boulanger A, Salleb-Aouissi A, Chow M, Dutta H, Gross PN, Huang B, Ierome S et al (2011) Machine learning for the New York city power grid. IEEE Trans Pattern Anal Mach Intell 34(2):328–345
Sagiroglu S, Sinanc D (2013) Big data: a review. In: 2013 international conference on collaboration technologies and systems (CTS). IEEE, pp 42–47

Salah F, Flath CM, Schuller A, Will C, Weinhardt C (2017) Morphological analysis of energy services: paving the way to quality differentiation in the power sector. Energy Policy 106:614–624

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14(3):199–222

Su H-Y, Liu T-Y (2018) Enhanced-online-random-forest model for static voltage stability assessment using wide area measurements. IEEE Trans Power Syst 33(6):6696–6704

Varaiya P, Wu F, Bialek J (2011) Smart operation of smart grid: risk-limiting dispatch. Proc IEEE 99(1):40–57

Wang K, Qi X, Liu H (2019) Photovoltaic power forecasting based lstm-convolutional network. Energy 189:116225

Wang Y, Chen Q, Hong T, Kang C (2018) Review of smart meter data analytics: applications, methodologies, and challenges. IEEE Trans Smart Grid 10(3):3125–3148

Wang Y, Zhang N, Chen Q, Kirschen DS, Li P, Xia Q (2017) Data-driven probabilistic net load forecasting with high penetration of behind-the-meter pv. IEEE Trans Power Syst 33(3):3255–3264

Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58(301):236–244

Watson HJ (2014) Tutorial: big data analytics: concepts, technologies, and applications. Commun Assoc Inf Syst 34(1):65

Weiss M, Helfenstein A, Mattern F, Staake T (2012) Leveraging smart meter data to recognize home appliances. In: 2012 IEEE international conference on pervasive computing and communications. IEEE, pp 190–197

Welikala S, Dinesh C, Ekanayake MPB, Godaliyadda RI, Ekanayake J (2017) Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting. IEEE Trans Smart Grid 10(1):448–461

Yan G, Li C, Zhou G, Zhang W, Li X (2007) Automatic extraction of power lines from aerial images. IEEE Geosci Remote Sens Lett 4(3):387–391

Zhang C, Zhong M, Wang Z, Goddard N, Sutton C (2018) Sequence-to-point learning with neural networks for non-intrusive load monitoring. In: Proceedings of the AAAI conference on artificial intelligence, vol 32

Zhang Y, Huang T, Bompard EF (2018) Big data analytics in smart grids: a review. Energy Inform 1(1):8