Energy Time-Series Features for Emerging Applications on the Basis of Human-Readable Machine Descriptions

Michael Vollmer[*]
michael.vollmer@kit.edu

Holger Trittenbach[*]
holger.trittenbach@kit.edu

Shahab Karrari[*]
shahab.karrari@kit.edu

Adrian Englhardt[*]
adrian.englhardt@kit.edu

Pawel Bielski[*]
pawel.bielski@student.kit.edu

Klemens Böhm[*]
klemens.boehm@kit.edu

[*]Karlsruhe Institute of Technology (KIT), Germany

ABSTRACT

Feature extraction from energy time series gives way to use data mining methods that require static vectors as an input. However, there is a plethora of feature extraction methods, and selecting a good set of features for energy time series is difficult. In this article, we make some strides towards the long-term vision to guide feature selection for emerging applications in the energy domain. To this end, we study the issue of extracting features from energy time series for a novel use case: Deriving human-understandable descriptions for smart-meter measurements of industrial production machines. We first categorize existing feature extraction based on their technical specifications and on their usefulness with our application. Based on it, we select features suitable for our use case to derive machine descriptions for an industrial production facility. Our experimental results show that our overview and categorization are useful to select features for a novel use case.

CCS CONCEPTS

• **Computing methodologies → Feature selection**; • **Hardware → Best practices for EDA**; • **Information systems → Extraction, transformation and loading**; *Summarization.*

KEYWORDS

Feature Extraction, Energy Status Data, Time Series, Machine Characterization

ACM Reference Format:
Michael Vollmer, Holger Trittenbach, Shahab Karrari, Adrian Englhardt, Pawel Bielski, and Klemens Böhm. 2019. Energy Time-Series Features for Emerging Applications on the Basis of Human-Readable Machine Descriptions. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems (e-Energy '19), June 25–28, 2019, Phoenix, AZ, USA*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3307772.3331022

1 INTRODUCTION

Energy data often are multivariate time series of measurements. However, many knowledge extraction methods require static data as input. A common way to deal with this antagonism is *feature extraction*, i.e., map each time series to a static vector representation. This requires a good set of features for the use case at hand.

Since feature extraction is so commonplace, there is a plethora of feature extraction methods which have been used with energy time series [21, 25, 54, 55]. However, this variety also is a burden. Finding a good set of features is difficult, because it depends both on the application as well as on characteristics of the data. For instance, the feature "voltage-current trajectory" [31] is useful for non-intrusive appliance load monitoring (NIALM), but may not be particularly useful for power quality applications. Another example is the "voltage-current trajectory" which only is feasible if measurements are high-resolution. Existing summaries on feature extraction tend to focus on specific applications, such as NIALM [25], power-quality detection [36] and exploration of household data [23]. However, with the rise of energy data analysis, new use cases emerge frequently. Usually, these use cases need to gain some prominence before a study of feature extraction methods become worthwhile. In consequence, the feature extraction for each new application is performed in isolation, without a reliable starting point. This gives way to a long-term vision, namely to provide guidance on feature selection for any novel use case in the energy domain. While such guidance clearly is useful, it is difficult to achieve.

Contributions. In order to make some strides towards that vision, we look at a use case which has not received much attention so far: finding human-understandable distinctive descriptions of industrial machines based on characteristic smart-meter measurements. Specifically, we use Explanation Tables [13], an approach for concise and human-interpretable summaries of data. Explanation Tables are an exploratory tool to support humans in understanding complex patterns. They are exemplary for a broader category of finding descriptions in large data sets, an important area with databases literature [30, 37]. Thus, the objective with ET is different to existing use cases that focus on device monitoring without human interaction, like NIALM. Explanation tables work on static data, i.e., they require feature extraction from energy time series.

Given this use case, we make three specific contributions. (i) We propose an approach to select a good set of features for a specific, often new application: In a nutshell, we introduce categories of features for energy time series. We then use this categorization to narrow down the search space to features that are applicable in a specific use case. (ii) We study the usefulness of these features with Explanation Tables. Our data are multivariate time series of smart meter measurements from ten production machines and the main terminal, with high volume and high dimensionality [6].

e-Energy '19, June 25–28, 2019, Phoenix, AZ, USA

Our experiments show that Explanation Tables with the selected features are compact and characterize the machines almost entirely. This shows the effectiveness of our selection approach, and suggests that our approach may also be useful as general guidance. (iii) We make our code for feature extraction and the resulting data sets publicly available.[1] They can serve as a starting point for other applications which require feature extraction for energy time-series data. The data set may also serve as a benchmark to find compact, human-understandable descriptions of characteristic machine behaviour.

2 RELATED WORK

Feature extraction from energy time series has been studied with various applications in mind. In this section, we review work on feature extraction for the following applications: power quality monitoring, NIALM, and data exploration for household data, e.g., to predict energy efficiency from characteristics of households.

Power Quality Monitoring. Power quality disturbances, like voltage sags, flicker, supply interruption and frequency disturbances, are electromagnetic phenomena, with durations from microseconds to minutes. Researchers have focused on data mining to detect disturbances beyond simple threshold-based detection. Feature extraction is important here, see [36] for an overview. The features used depend on specifics, like the type of disturbance and the time resolution of measurements [7, 35, 50]. Many features used require very high resolution with sampling rates of several kHz.

NIALM. The objective of NIALM is to disaggregate energy data into appliance specific signals [18], to design demand response algorithms [19], and to identify devices [25, 39, 43]. Existing NIALM applications mainly focus on household data. Here, several categorizations of features have been proposed: Steady-states vs. transient-states features [18], traditional power metric-based, vector-based, and voltage-current (V-I) trajectory-based features [31], as well as restrictions on sampling rates to distinguish between micro- and macro-level device signatures [32]. [25, 46, 54, 55] are comprehensive overviews for low- and high-resolution data.

Exploration of Household Data. There are several approaches that analyze consumption data from households based on low-resolution data from smart meters. This includes classification of households and predicting energy efficiency related household properties [23]. Features used in this context are often handcrafted, e.g., consumption figures, ratios, temporal properties and statistical properties [4, 5, 20, 21, 23, 48].

So there is a variety of summaries on feature extraction for energy data. However, they are application specific and it is unclear to which extent insights from such summaries also transfer to other use cases.

3 TIME-SERIES FEATURE REPRESENTATION

This section focuses on the extraction of features from smart-meter measurements from an industrial production site. We structure feature extraction as a two-step process: the pre-processing of sequences and the mapping of sequences to static vectors. There are several design decisions one has to take along this process with any use case. We elaborate on the options one has with these decisions, and discuss how data characteristics and application restrict them.

In the following, we first describe the raw smart-meter data and motivate our use case. We then discuss sequence pre-processing and feature extraction. To this end, we first present the general options, and then elaborate how we decide on them in our specific case.

3.1 Data and Use Case

The starting point for this work is smart meter data from an industrial production site, the HIPE [6] data set. It contains smart meter measurements from 10 machines over 3 months with a time resolution of 5 s. Each meter stores stores 112 values per time step if the machine is connected with 1 phase and 360 values for machines with 3 power phases. This abundance of data poses challenges for human understanding and interpretation.

Commonly, humans interpret only small portions of the data, like the power consumption on a coarse time scale. Further, they may "feed" the data to machine learning methods that produce results untraceable for humans. Our use case, however, are human interpretable descriptions of the machines based on the data available. Describing the machine behaviour is different from NIALM, where models often are used as a "black box".

The summaries sought are Explanation Tables (ET) [13] with some modifications to allow for numerical attributes [51]. However, as many methods in knowledge discovery, ET have not explicitly been developed for time-series data, but currently are confined to static input. This requires a transformation of time series to a feature-based representation. Unlike other machine-learning methods, ET requires features that humans can understand, since ET describe the machines using solely these features to human interpreters.

3.2 Sequence Pre-Processing

Selecting a good set of features from energy time series requires to decide on what constitutes an observation, i.e., the entity of measurements a feature is calculated on. This step is called *segmenting*. To illustrate, our data are time series over three months from several machines. On the one hand, one could define the full three-month time series of each machine as one observation. On the other hand, comparing the consumption pattern of one machine over the course of a day requires to extract several sequences; so each sequence is one observation. As an optional step, one can then use *filtering* to select a subset of the sequences of interest.

Segmentation. The first step of pre-processing is to segment the time series. In general, this is a trade-off whether one is interested in "global" or "local" behaviour of the time series, as well as adherence to technical limits. In our case, the technical minimum sequence length is about 5 s. This would result in one observation per sequence. However, this corresponds to interpreting the raw time-series vectors and would remove any temporal dependencies from the data. The technical maximum sequence length is the minimum number of remaining observations in the segmented data set. For instance, with three months worth of data, weekly segments would result in only 12 observations per machine. With so few observations, a small number of anomalies in the measurements can

[1] https://www.energystatusdata.kit.edu/hipe-features.php

distort the description of "usual" behaviour. Besides these restrictions, our choice is arbitrary: We choose 15 min, the most common interval used in literature [52], and 1 h intervals as comparison. The extracted sequences are non-overlapping, since overlapping sequences are prone to detecting spurious effects [28]. In general, one may also extract sequences of dynamic length, e.g., based on on/off periods of machines. However, this requires assumptions on how to find events that define the start and the end of a sequence [2, 18]. Since we do not make any assumption to this end, we rely on static sequence lengths.

Filtering. Filtering allows to focus on segments of interest, based on any selection predicate. Naturally, this choice is application-dependent. Literature filters attributes [32] or features after the extraction [22]. In our case, we use filtering to remove observations for all attributes where machines are in an off state, i.e., a current of 0. For comparison, we also use data without any filtering.

3.3 Feature Extraction from Sequences

There is no general categorization of features from energy time series that we can rely on to select features for our use case. Instead, we list several characteristics that we deem useful to categorize features from literature. One can then use this categorization to select a subset of features relevant to a specific use case. These are *Interpretability*, *Temporal Scope*, *Number of Values*, *Additional Parameters*, and *Data Requirements*. We first explain these characteristics and then use them to select features for our use case.

Interpretability. Intuitively, human-understandable descriptions of energy consumption patterns require features to be interpretable. Any definition of interpretable is subjective to some extent, and there is no single way to define it [34]. In this current article, we deem a feature interpretable if it describes a property of the time series that has a real-world correspondence, and the meaning of the features is independent of a specific application. For instance, the auto-correlation of a time series is interpretable. It characterizes the inter-dependence of subsequent measurements, whether this is for, say, voltage curves in households or reactive power of production machines. In contrast, features from auto-encoders are not interpretable. The output of an auto-encoder has no immediate real-world expression and depends on the training data. Thus, auto-encoded time series values have no analogue meaning across applications. In our use case, we require features to be interpretable.

Temporal Scope. The temporal scope of a feature describes the granularity of the feature calculation, and whether the order of measurements matter. We discern between three temporal scopes: *a-temporal*, *local*, and *global*. *A-temporal* features are independent of the measurement order. Examples are descriptive statistics like "maximum" or "mean". A *global* feature is a compression of the full time series that depend on the order of measurements. It often is a lossy compression, e.g., the Fourier coefficients up to a certain degree. *Local* features are more fine-granular, and consider individual positions or small subsequences of measurements in the time-series. Examples are the position of the first maximum or the sum of differences between consecutive values; both depend on the order of measurements. With our use case, we deem all of the above reasonable choices, as they capture different aspects of the data. In contrast, other applications may impose a stronger focus on the sequence of values or on individual measurements.

Number of Values. Features differ by their dimensionality. In the extreme cases, features either reduce a sequence to a single value, or they output more values than in the original sequence. In general, machine learning becomes difficult with large numbers of features, also known as the curse of dimensionality. So feature with few values tend to be preferred. We use the categories *single*, *multiple* and *variable* number of values. The difference between the last two categories is that for 'variable' the number of output values is defined before computing the feature and affects the feature values. For instance, Fourier transformation returns multiple values, the coefficients. From these values, one can then choose the largest ones. However, piece-wise aggregate approximation aggregation is variable, since one has to decide on the number of aggregation intervals a priori, because feature computation depends on this number. This is useful, for instance, to evaluate different aggregation levels by comparing analysis results based on different number of output values [49]. In general, features that result in more than a single value are difficult to configure with energy data. They require to reason which number captures sufficient information from the original time series. To limit the dimensionality, we use only single-value features with our use case.

Additional Parameters. Some features are parameterized. For instance, the feature "number of peaks" depends on the number of consecutively increasing values that define a peak. In our categorization, we only differentiate whether features require to set additional parameters (*yes* and *no*). Choosing parameter values is difficult since there often is no intuition which values are suitable for energy data. Extracting features for multiple parameter values is not a realistic option, since this increases the dimensionality of the feature space. In our use case, we do not use any features which are parameterized due to the various problems.

Data Requirements. Features can also have requirements on the input data. This includes restrictions to a specific *attribute* and constraints on the *sampling* rate. For instance, voltage-current trajectories are useful to distinguish between appliances. However, they require measurements of two attributes, voltage and current, with high sampling rates, to extract the trajectory from single power cycles. In our use case, we are restricted to features with a required sampling rate of 5 s or lower.

3.4 Selecting Features

The categorization gives way to identify a set of features from literature that is relevant for a specific use case. Based on the categorization, one has to filter for features with specific characteristics, e.g., features that are interpretable. Table 1 is an overview of the features considered here. In this table, the features are grouped by their characteristics. In other words, features are in the same group if they have the same interpretability, temporal scope, etc. Note that we omit many time-related features such as "Mean consumption in the morning" [23] since these are essentially covered by segmentation and filtering choices.

Features	Interpretable	Scope	# Values	Parametrized	Requirements
On-Off-Ratio [18]	Yes	a-temporal	single	no	attribute
Count-Above-Mean [23], Count-Below-Mean [9], Num-Max [46], Any-Duplicate [9], Max-Duplicate [9], Min-Duplicate [9] Kurtosis [42], Length [53], Non-Zero-Ratio [53], Num-Min [46] Max [23], Min [23], Mean [23], Median [23], Crest-Factor [11], Ratio-Recurring-Values [9], Ratio-Recurring-Datapoints [9], Num-States [12], Skewness [42], Standard-Deviation [42],	Yes	a-temporal	single	no	none
Binned-Entropy [24], Quantiles [9], Range-Count [9]	Yes	a-temporal	single	yes	none
Abs-Diff-Sum [9], Mean-Second-Derivative [9], Mean-Diff [9], Mean-Abs-Diff [9], Num-Mean-Crossing [26], Pos-First-Max [23], Pos-First-Min [23], Pos-Last-Max [9], Pos-Last-Min [9], Max-Streak-Above-Mean [9], Max-Streak-Below-Mean [9]	Yes	local	single	no	none
Change-Quantile [9], Pos-Mass-Quantile [9], Peak-Count [23]	Yes	local	single	yes	none
Augmented-Dickey-Fuller-Test [10]	Yes	local	multiple	no	none
Duration-Distribution [18]	Yes	local	multiple	no	attribute
Linear-Weighted-Average [53], Trend-Slope [9], Quadratic-Weighted-Average [53], Sample-Entropy [44], Complexity-Estimate [3]	Yes	global	single	no	none
Aggregated-Autocorrelation [17], Autocorrelation [23] Approximate-Entropy [44], C3 [47],Partial-Autocorrelation [9]	Yes	global	single	yes	none
Time-Reversal-Asymmetry [16]	Yes	global	single	yes	none
Aggregated-FFT [9], FFT-Coefficients [29], Poles and residues [14]	Yes	global	multiple	no	sampling
Voltage-Current-Trajectories [31]	Yes	global	multiple	no	attribute, sampling
Aggregated-Linear-Trend [9]	Yes	global	multiple	yes	none
Energy [42], Maximal-Lyapunov-Exponent [42], Structure-Detectors [40]	Yes	global	multiple	no	none
Max-Langevin-Fixed-Point [15]	No	global	single	yes	none
Autoregression-Coefficients [1], Friedrich-Coefficients [15] Adaptive-Piecewise-Constant-Approximation [27]	No	global	variable	no	none
Autoencoders [8], Restricted Boltzmann Machine [38] Symbolic-Aggregate-Approximation [33]	No	global	variable	yes	none
Ricker-Wavelet-Coefficients [45], Stockwell-Matrix-Coefficients [7]	No	global	variable	yes	sampling

Table 1: List of time-series features grouped by characteristics.

For our use case, we are interested in features that are interpretable, have a single value, no parameters and no sampling requirements, which typically require sampling rates above 1000 Hz. While this eliminates some of the groups in Table 1, the resulting set of 36 features remains sizeable. Because of technical restrictions on the sequence length (cf. Section 3.2), we extracted these features for the HIPE data set with a segmentation of 15 min and 1 h intervals. We restrict ourselves to these two intervals for brevity while still providing some evidence on the impact of segmentation. An extensive evaluation regarding the optimal interval length in light of different data sets, features and data sets would be interesting, however this is not focus of this work. Since HIPE contains both 1- and 3-phased machines with different attributes, we used the average current, line voltage, active power, reactive power, apparent power, frequency, power factor of the 3 phased machines. The data set is filtered from HIPE data when the respective machine was turned on, as determined by a positive current. This results in four different feature representations of HIPE for each combination of the two segmentation choices and two filtering choices, which we refer to as *Features-1H-Full*, *Features-1H-On*, *Features-15Min-Full* and *Features-15Min-On*. These data sets and our code for this feature extraction are available at https://www.energystatusdata.kit.edu/hipe-features.php.

4 UNDERSTANDABLE DESCRIPTIONS

In this section we build Explanation Tables (ET) [13] for the HIPE data by using the feature representation discussed in the previous section. We also compare ET for our four extracted feature data sets to assess the impact of our segmentation and filtering choices on the quality of machine characterizations.

Row	I_Max	I_Mean	I_Min	Count	Machine	Correct
1	0.26 − 734	⋆	0.26 − 1.83	8842	PnPUnit	99.92%
2	⋆	⋆	0 − 0.26	69165	Terminal	0.01%
3	⋆	⋆	428 − 452	150	ChipSaw	96.0%
4	4.38 − 4.87	⋆	⋆	333	V.Pump2	99.4%

Table 2: Exemplary ET based on electric current.

4.1 Explanation Tables

Explanation Tables [13] summarize the relationship between a designated "outcome" attribute and "explanatory" attributes. Here, we are interested in the relationship between different machines and the feature representation. That is, the *machine name* is the outcome and the feature vectors are the explanatory attributes. This way, an ET describes patterns in the smart meter data that allow to distinguish the machines that have generated the data. To fit the smart-meter data, we use adapted ET [51] to accommodate numerical attributes, without changing the theoretic framework, functionality and quality evaluation.

Table 2 is an example of an ET with features *min*, *mean* and *max* from the electric current as explanatory attributes. A row consists of two parts. The left part is a conjunctive pattern that describes a subset of the data. Each clause in the conjunction defines an interval on an attribute to be matched. We display intervals matching exactly one value as a constant, and intervals matching the full attribute domain as a wildcard ⋆. For example, the first row of Table 2 matches all instances where the measured current has its maximum between 0.26 A and 734 A and its minimum between 0.26 A and 1.83 A, for all current means. The right part is a machine label with summary statistics. These are the number of instances the pattern matches ("Count") and the percentage of correct matches ("Correct"), i.e., the share of Count that indeed matches the "Machine" label. The first row of Table 2 shows that 8842 instances match the pattern, and 99.92 % of them are the pick and place unit ("PnPUnit"). So a user knows how many instances the row affects, and how decisive the pattern is for this machine. In essence, each row "explains" the association of one type of machine behaviour with one machine. Note that these explanations can also happen with negative examples, as seen in the second row. There are 69165 instances, which are over 70% of all instances in the data set, with low currents below 0.26 A, but these never are from the "MainTerminal". Finally, patterns need not be mutually exclusive, i.e., an instance may fit several rows, e.g., the first and the fourth row of Table 2. This is because ET provide interesting and informative patterns for human interpreters instead of focusing on exact classification results. That is, the first and fourth row are interesting patterns of the "PnPUnit" and "Vacuum Pump 2". Using ET for classification, i.e. assigning machine labels to different feature vectors requires intricate considerations. For instance consider an entry with I_Max = 4.38 A and I_Min = 1.83 A that matches row 1 and 4 in Table 2. It is nearly four times more like to be measured by the "Vacuum Pump 2" than the "PnPUnit" because there are $8842 \cdot 0.0008 = 7$ entries matching row 1 with a different machine label compared to row 4 with $333 \cdot 0.006 = 2$ such entries.

To provide a formal measure for the overall informativeness of the patterns presented by an ET, an extensive model is used. Although we defer the specifics to [13, 51], in essence this information content quantifies how well the patterns of an ET describes the distribution of machine labels across all instances in the data. Since this information content of an ET is not normalized and scales with data size and information content, i.e. complexity, of the input data, we normalize the quality scores for better comparability. That is, we compare the explanatory power to the absolute maximum for this data, to obtain a relative score in percent, which we call *completeness*. That is, *completeness* quantifies how well all patterns of one ET describe the entire data set. For instance, an ET with 100% completeness allows a perfect reconstruction of the machine labels given the explanatory attributes ("features"). As a result is it particularly interesting how much completeness ET with a limited number of patterns, e.g. 50 rows, can achieve.

4.2 Experiments

We now generate ET for the feature representation of HIPE. Since ET do not scale well with high dimensional feature vectors [13, 51], a restriction on the number of input features is necessary. We choose 18 features we deem a good coverage and diversity in a manual selection step. In detail, we use the following features from Table 1: *Complexity-Estimate*, *Count-Above-Mean*, *Crest-Factor*, *Kurtosis*, *Linear-Weighted-Average*, *Trend-Slope*, *Max*, *Mean*, *Mean-Abs-Diff*, *Mean-Second-Derivative*, *Min*, *Num-Mean-Crossing*, *Num-States*, *Non-Zero-Ratio*, *Ratio-Recurring-Datapoints*, *Sample-Entropy*, *Skewness* and *Standard-Deviation*. To avoid the large vectors yielded by computing each of these features for each measured attribute, we compare the completeness of ET using different features and attributes. We also consider this evaluation separately for *Features-1H-Full*, *Features-1H-On*, *Features-15Min-Full* and *Features-15Min-On*. This comparison indicates which features computed on which attributes yield ET with high completeness depending on segmentation and filtering.

Figure 1 graphs the completeness of ET with 50 rows based on different attributes. That is, for each attribute we build a distinct ET where the input is our set of features computed for that attribute. The figure graphs the completeness for each attribute, marked at the x-axis, and data set. As one might expect, frequency and voltage offer little insights related to individual machines, as these attributes almost entirely depend on the grid. The remaining attributes, however, yield tables with high completeness across the different data sets. The only exception is the power factor for unfiltered hourly data, which yields an ET with noticeably lower completeness. Otherwise, the only significant difference between the different data sets is that frequency and voltage yield better, though still very incomplete, ET with unfiltered data. The reason is that "Vacuum Pump 2", where the smart meter is installed behind the power switch of the machine, produces measurements of 0 V and 0 Hz when the machine is turned off. These measurements are unique to this machine and thus provide identifying patterns.

Next, we evaluate which features enables ET with high completeness. That is, we compute each feature on each attribute and evaluate the completeness of the resulting ET. Figure 2 graphs the completeness of ET with 50 rows based on different features

Row	I_NZ	L_Min	L_NZ	P_Min	P_NZ	Q_Max	Q_Min	Q_NZ	S_Max	S_Min	S_NZ	Count	Machine	Correct
1	★	-1	★	★	0	★	0	★	★	★	★	8511	V.Pump2	100%
2	★	★	★	★	★	-0.07	★	★	★	★	0	7970	PnPUnit	100%
3	★	★	★	★	★	★	★	★	0.01 − 0.04	0.01 − 0.02	0	4261	WashM.	100%
4	★	★	★	★	0	★	★	0	★	★	★	23885	Terminal	36.89%
5	★	★	★	0.33 − 0.34	★	★	★	★	★	★	0 − 0.006	675	S.Printer	98.96%
6	★	★	0	★	★	★	0.06	★	★	★	★	601	H.T.Oven	100%
7	★	★	★	★	★	★	★	0	★	0.15 − 0.19	★	461	ChipSaw	94.14%
8	0	★	★	★	0	★	★	0	★	★	★	23874	V.Pump2	01.36%
9	★	★	★	0.08 − 0.12	★	★	★	★	★	0.2 − 0.22	0 − 0.006	274	PnPUnit	100%
10	★	★	★	★	★	★	★	0	★	0.04 − 0.05	★	1047	V.Oven	39.06%
11	★	★	★	★	★	★	★	★	1.06 − 1.24	★	★	1363	V.Pump1	93.84%
12	★	-0.999 − -0.981	0	★	★	★	★	★	★	★	★	725	S.Oven	53.10%
13	★	★	0	★	★	★	★	★	★	0.05	★	553	H.T.Oven	98.55%
14	★	★	★	★	0	★	★	★	★	★	★	24357	H.T.Oven	06.76%
15	★	★	★	★	★	0.7	★	★	★	★	0	144	V.Pump2	46.53%

Table 3: ET for HIPE machines, based on the *Min*, *Max* and *Non-Zero-Ratio* (NZ) features extracted in 15 minute segments without filtering. The data set contains 97112 feature vectors in equal parts from each of the 11 smart-meters.

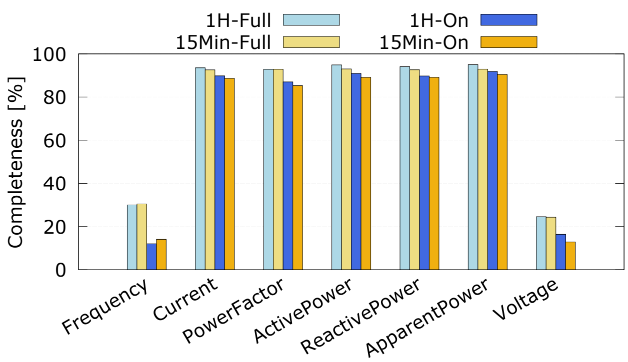

Figure 1: Completeness of ET based on single attributes.

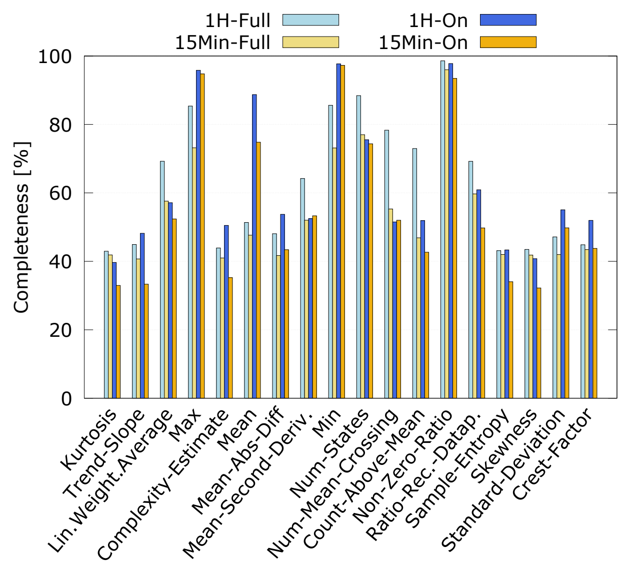

Figure 2: Completeness of ET based on single features.

for different data sets. Overall, most features yield similar completeness for all data sets with a few singular exceptions, like the *Count-Above-Mean*. Only the *Mean*, and to a lesser extent *Min* and *Max*, show a significant difference between filtered and unfiltered data. This is surprising, as one might expect stronger distinctions between filtered and unfiltered data as machines usually have more unique consumption patterns while they are running. Additionally, the features resulting in the highest completeness are a-temporal, indicating that the sequence of measurements are not very characterizing for the machines. One reason might be that the noise with these time series obfuscates distinctive sequences. Going forward, the features *Non-Zero-Ratio*, *Min* and *Max* appear to be the best across the different data sets. Note that we chose this application specific method for selection features and attributes as it yielded direct results for interpretation. Although out of scope for this work, an interesting question is what differences arise by using conventional feature selection methods like Min-Redundancy-Max-Relevance [41]. That is, whether the same features are selected and if the final ET has a higher or lower completeness.

Finally, Table 3 is an ET for the unfiltered data, segmented to 15 min intervals. For brevity, we show only the first 15 rows, and we omit I_Max, I_Min, L_Max and P_Max, which have a ★ wildcard for all rows. While the corresponding ET with 50 rows is 98% complete, the 15 rows presented in Table 3 already are 85% complete. This table presents characteristics of smart meter data for some machines. For instance, if the reactive power is zero, and if active power has no non-zero values, and if a power factor of −1 occurs in the 15 minute intervals, the measurements are guaranteed to come from "Vacuum Pump 2". The lack of non-zero values for active and reactive power stems from the specific smart meter configuration of this machine mentioned in the discussion of Figure 1. As another example, the second row describes "Pick and Place Unit" by the reactive power of at most −0.07 kvar and no non-zero values for apparent power. Overall, Table 3 is a compact description of characteristic patterns of the machines while avoiding overly complicated patterns with many clauses.

To conclude this section, we briefly summarize our findings. We achieved high completeness with ET for the the smart meter data from production machines. That is, we obtained compact descriptions of measurements characteristic to different machines, as evidenced by Table 3. In addition to the immediate insights from the unique patterns of individual machines, this also demonstrates the usefulness of our feature extraction procedure. The features we chose because they fit the categories relevant to our use case directly enabled good ET, as seen in Figure 1 and Figure 2. Although out of scope for this work, an interesting direction for future work is to study the influence of the data set on the features selected. Naturally, it is also impossible to prove the usefulness for any emerging application. However, we deem our categorization and selection procedure a good starting point.

5 CONCLUSIONS

In this article, we study how feature extraction from energy time series tailored to a specific application should look like. Although feature extraction is common, there is a lack of guidance on how to proceed with novel use cases. In this article, we focus on a specific use case, namely creating human-understandable descriptions of machines based on characteristic smart-meter measurements, and we study feature selection to this end. For a systematic selection, we propose a categorization of the many features proposed earlier and classify related work. This categorization has turned out to be useful to select a good set of features for our use case. In particular, the features selected give way to a compact description of the factory machines based on characteristic measurements. This is a stride towards the long-term vision to provide guidance on feature selection for new applications.

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG) as part of the Research Training Group GRK 2153: Energy Status Data – Informatics Methods for its Collection, Analysis and Exploitation.

REFERENCES

[1] Beth Andrews, Matthew Calder, Richard A Davis, et al. 2009. Maximum likelihood estimation for α-stable autoregressive processes. *The Annals of Statistics* 37, 4 (2009), 1946–1982.
[2] Michael Baranski and Jürgen Voss. 2004. Genetic algorithm for pattern detection in NIALM systems. In *IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4. 3462–3468.
[3] Gustavo EAPA Batista, Eamonn J Keogh, Oben Moses Tataw, and Vinicius MA De Souza. 2014. CID: an efficient complexity-invariant distance for time series. *Data Mining and Knowledge Discovery* 28, 3 (2014), 634–669.
[4] Christian Beckel, Leyna Sadamori, and Silvia Santini. 2012. Towards automatic classification of private households using electricity consumption data. *Workshop on Embedded Systems for Energy Efficiency in Buildings (BuildSys '12)*, 169–176.
[5] Christian Beckel, Leyna Sadamori, Thorsten Staake, and Silvia Santini. 2014. Revealing household characteristics from smart meter data. *Energy* 78 (2014), 397–410.
[6] Simon Bischof, Holger Trittenbach, Michael Vollmer, Dominik Werle, Thomas Blank, and Klemens Böhm. 2018. HIPE–An Energy-Status-Data Set from Industrial Production. *Proceedings of ACM e-Energy (e-Energy '18)*.
[7] Birendra Biswal, P.K. Dash, and B.K. Panigrahi. 2009. Power quality disturbance classification using fuzzy C-means algorithm and adaptive particle swarm optimization. *IEEE Transactions on Industrial Electronics* 56, 1 (2009), 212–220.
[8] Roberto Bonfigli, Andrea Felicetti, Emanuele Principi, Marco Fagiani, Stefano Squartini, and Francesco Piazza. 2018. Denoising autoencoders for non-intrusive load monitoring: improvements and comparative evaluation. *Energy and Buildings* 158 (2018), 1461–1474.
[9] Maximilian Christ, Andreas W Kempa-Liehr, and Michael Feindt. 2016. Distributed and parallel time series feature extraction for industrial big data applications. *arXiv preprint arXiv:1610.07717* (2016).
[10] David A Dickey and Wayne A Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association* 74, 366a (1979), 427–431.
[11] Liang Du, Jose A. Restrepo, Yi Yang, Ronald G. Harley, and Thomas G. Habetler. 2013. Nonintrusive, self-organizing, and probabilistic classification and identification of plugged-in electric loads. *IEEE Transactions on Smart Grid* 4, 3 (2013), 1371–1380.
[12] Liang Du, Yi Yang, Dawei He, Ronald G Harley, and Thomas G Habetler. 2015. Feature extraction for load identification using long-term operating waveforms. *IEEE Transactions on Smart Grid* 6, 2 (2015), 819–826.
[13] Kareem El Gebaly, Parag Agrawal, Lukasz Golab, Flip Korn, and Divesh Srivastava. 2014. Interpretable and informative explanations of outcomes. *VLDB Endowment* 8, 1 (2014), 61–72.
[14] Khalil El Khamlichi Drissi, Kamal Kerroum, Claire Faure, Alioune Diop, Hala Najmeddine, Michel Michou, Thierry Jouannet, and Christophe Pasquier. 2010. Smart metering by using Matrix Pencil. *International Conference on Environment and Electrical Engineering* (2010), 238–241.
[15] Rudolf Friedrich, Silke Siegert, Joachim Peinke, St. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister. 2000. Extracting model equations from experimental data. *Physics Letters A* 271, 3 (2000), 217–222.
[16] Ben D. Fulcher and Nick S. Jones. 2014. Highly comparative feature-based time-series classification. *IEEE Transactions on Knowledge and Data Engineering* 26, 12 (2014), 3026–3037.
[17] Andrew Hall, John Louis, and David Lamb. 2001. A method for extracting detailed information from high resolution multispectral images of vineyards. In *International Conference on Geocomputation*. 24–26.
[18] G. W. Hart. 1992. Nonintrusive appliance load monitoring. *Proc. IEEE* 80, 12 (1992), 1870–1891.
[19] Dawei He, Weixuan Lin, Nan Liu, Ronald. G. Harley, and Thomas. G. Habetler. 2013. Incorporating non-intrusive load monitoring into building level demand response. *IEEE Transactions on Smart Grid* 4, 4 (2013), 1870–1877.
[20] Konstantin Hopf. 2018. Mining volunteered geographic information for predictive energy data analytics. *Energy Informatics* 1, 1 (2018), 4.
[21] Konstantin Hopf, Mariya Sodenkamp, Ilya Kozlovkiy, and Thorsten Staake. 2014. Feature extraction and filtering for household classification based on smart electricity meter data. *Computer Science - Research and Development*.
[22] Konstantin Hopf, Mariya Sodenkamp, Ilya Kozlovkiy, and Thorsten Staake. 2016. Feature extraction and filtering for household classification based on smart electricity meter data. *Computer Science-Research and Development* 31, 3 (2016), 141–148.
[23] Konstantin Hopf, Mariya Sodenkamp, and Thorsten Staake. 2018. Enhancing energy efficiency in the residential sector with smart meter data analytics. *Electronic Markets* 28, 4 (2018), 453–473.
[24] Mikko Juusola and Andrew S French. 1997. The Efficiency of Sensory Information Coding by Mechanoreceptor Neurons. *Neuron* 18, 6 (1997), 959 – 968.
[25] Matthias Kahl, Anwar Ul Haq, Thomas Kriechbaumer, and Hans-Arno Jacobsen. 2017. A Comprehensive Feature Study for Appliance Recognition on High Frequency Energy Data. In *Proceedings of ACM e-Energy (e-Energy '17)*. 121–131.
[26] Benjamin Kedem and Sidney Yakowitz. 1994. *Time series analysis by higher order crossings*. IEEE press New York.
[27] Eamonn Keogh, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. 2001. Locally adaptive dimensionality reduction for indexing large time series databases. *ACM Sigmod Record* 30, 2 (2001), 151–162.
[28] E Keogh, J Lin, and W Truppel. 2003. Clustering of time series subsequences is meaningless: implications for previous and future research. In *IEEE International Conference on Data Mining*. 115–122.
[29] Andreas Krause, Matthias Ihmig, Edward Rankin, Derek Leong, Smriti Gupta, Daniel Siewiorek, Asim Smailagic, Michael Deisher, and Uttam Sengupta. 2005. Trading off prediction accuracy and power consumption for context-aware wearable computing. In *IEEE International Symposium on Wearable Computers (ISWC'05)*. 20–26.
[30] Himabindu Lakkaraju, Stephen H Bach, and Jure Leskovec. 2016. Interpretable decision sets: A joint framework for description and prediction. In *ACM SIGKDD international conference on knowledge discovery and data mining*. 1675–1684.
[31] H. Y. Lam, G. S. K. Fung, and W. K. Lee. 2007. A Novel Method to Construct Taxonomy Electrical Appliances Based on Load Signatures. *IEEE Transactions on Consumer Electronics* 53, 2 (2007), 653–660.

[32] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng. 2010. Load Signature Study-Part I: Basic Concept, Structure, and Methodology. *IEEE Transactions on Power Delivery* 25, 2 (2010), 551–560.

[33] Jessica Lin, Eamonn Keogh, Li Wei, and Stefano Lonardi. 2007. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and knowledge discovery* 15, 2 (2007), 107–144.

[34] Zachary C Lipton. 2016. The Mythos of Model Interpretability. In *ICML Workshop on Human Interpretability in Machine Learning*.

[35] Om Prakash Mahela and Abdul Gafoor Shaik. 2017. Power quality recognition in distribution system with solar energy penetration using S-transform and Fuzzy C-means clustering. *Renewable Energy* 106 (2017), 37–51.

[36] Om Prakash Mahela, Abdul Gafoor Shaik, and Neeraj Gupta. 2015. A critical review of detection and classification of power quality events. *Renewable and Sustainable Energy Reviews* 41 (2015), 495–505.

[37] Michael Mampaey, Nikolaj Tatti, and Jilles Vreeken. 2011. Tell me what i need to know: succinctly summarizing data with itemsets. In *ACM SIGKDD international conference on Knowledge discovery and data mining*. 573–581.

[38] Elena Mocanu, Phuong H Nguyen, and Madeleine Gibescu. 2016. Energy disaggregation for real-time building flexibility detection. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 1–5.

[39] M.W. Mustafa, S.N. Khalid, I. Abubakar, M. Mustapha, and Hussain Shareef. 2016. Application of load monitoring in appliances' energy management – A review. *Renewable and Sustainable Energy Reviews* 67 (2016), 235–245.

[40] Robert T Olszewski. 2001. *Generalized feature extraction for structural pattern recognition in time-series data*. Ph.D. Dissertation. Carnegie Mellon University, PA, School of Computer Science.

[41] Hanchuan Peng, Fuhui Long, and Chris Ding. 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 8 (2005), 1226–1238.

[42] Teemu Räsänen and Mikko Kolehmainen. 2009. Feature-Based Clustering for Electricity Use Time Series Data. *Lecture Notes in Computer Science* 5495, 401–412.

[43] Andreas Reinhardt, Paul Baumann, Daniel Burgstahler, Matthias Hollick, Hristo Chonov, Marc Werner, and Ralf Steinmetz. 2012. On the accuracy of appliance identification based on distributed load metering data. In *SustainIT*. IEEE, 1–9.

[44] Joshua S Richman and J Randall Moorman. 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology* 278, 6 (2000), H2039–H2049.

[45] Norman Ricker. 1953. The form and laws of propagation of seismic wavelets. *Geophysics* 18, 1 (1953), 10–40.

[46] Nasrin Sadeghianpourhamami, Joeri Ruyssinck, Dirk Deschrijver, Tom Dhaene, and Chris Develder. 2017. Comprehensive feature selection for appliance classification in NILM. *Energy and Buildings* 151 (2017), 98–106.

[47] Thomas Schreiber and Andreas Schmitz. 1997. Discrimination power of measures for nonlinearity in a time series. *Physical Review E* 55, 5 (1997), 5443.

[48] Gan Sun, Yang Cong, Dongdong Hou, Huijie Fan, Xiaowei Xu, and Haibin Yu. 2017. Joint Household Characteristic Prediction via Smart Meter Data. *IEEE Transactions on Smart Grid* (2017).

[49] Holger Trittenbach, Jakob Bach, and Klemens Böhm. 2018. On the Tradeoff Between Energy Data Aggregation and Clustering Quality, In Proceedings of the Ninth International Conference on Future Energy Systems. *Proceedings of ACM e-Energy (e-Energy '18)*, 399–401. https://doi.org/10.1145/3208903.3212038

[50] V Vega, C Duarte, and G Ordóñez. 2006. Automatic power quality disturbances detection and classification based on Discrete Wavelet Transform and Artificial Intelligence. In *2006 IEEE/PES Transmission & Distribution Conference and Exposition: Latin America*. 1–6.

[51] Michael Vollmer, Lukasz Golab, Klemens Böhm, and Divesh Srivastava. 2019. Informative Summarization of Numeric Data. In *International Conference on Scientific and Statistical Database Management (SSDBM '19)*.

[52] Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang. 2018. Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid* (2018).

[53] Jenna Wiens, Eric Horvitz, and John V Guttag. 2012. Patient risk stratification for hospital-associated c. diff as a time-series classification task. In *Advances in Neural Information Processing Systems*. 467–475.

[54] M. Zeifman and K. Roth. 2011. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics* 57, 1 (2011), 76–84.

[55] Ahmed Zoha, Alexander Gluhak, Muhammad Imran, and Sutharshan Rajasegarar. 2012. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors* 12, 12 (2012), 16838–16866.