

RESEARCH

Open Access



Prediction of domestic appliances usage based on electrical consumption

Patrick Huber^{*}, Mario Gerber, Andreas Rumsch and Andrew Paice

From The 7th DACH+ Conference on Energy Informatics
Oldenburg, Germany. 11-12 October 2018

^{*} Correspondence:
patrick.huber@hslu.ch
Engineering and Architecture,
Lucerne University of Applied
Sciences and Arts, Technikumstrasse
21, 6048 Horw, Switzerland

Abstract

Forecasting or modeling the on-off times of domestic appliances has gained increasing attention in recent years. However, comparing currently published results is difficult due to the many different data-sets and performance measures employed. In this paper, we evaluate the performance of three increasingly sophisticated approaches within a common framework on three data-sets each spanning 2 years. The approaches forecast the future on-off times of the appliances for the next 24 h on an hourly basis, solely based on historic energy consumption data. The appliances investigated are driven by user behavior and consume a significant fraction of the household's total electrical energy consumption. We find that for all algorithms the average area under curve (AUC) in the receiver operating characteristic (ROC) is in the range between 72% and 73%, i.e. indicating mediocre prediction quality. We conclude that historic consumption data alone is not sufficient for a good quality hourly forecast.

Keywords: Load forecasting, Shiftable loads, Domestic appliance, Experimental comparison

Introduction

Forecasting or modeling the expected on-off times of domestic appliances is motivated from two directions: (i) generation of electrical load profiles and (ii) learning and predicting user behavior. Artificial load profile generation (Pflugradt 2016) can be helpful if large numbers of profiles spanning extended durations are required because their collection typically involves arduous measurement campaigns. While the precise prediction of the switch-on/off times is in this case not the main concern, it is an essential part in applications targeting demand response systems: Learning the usage pattern of appliances and therefore, knowledge of the user behavior is a vital input to optimally plan energy usage (Chrysopoulos et al. 2014; Holub and Sikora 2013). While different prediction approaches have been published (Chrysopoulos et al. 2014; Holub and Sikora 2013; Truong et al. 2013; Barbato et al. 2011), an outstanding matter in adequately addressing the forecast of domestic appliance usage is a comparison of the available approaches: Published results are difficult to compare because of diverse performance metrics, different predicted appliances and the large variety of employed datasets, either measured at different geographic locations or even simulated. It is

therefore unclear how well a method generalizes (i) over extended time periods and (ii) to other datasets with different attributes such as appliances, number of inhabitants, user habits and behavior.

In this work, we compare published the approaches we are aware of (Chrysopoulos et al. 2014; Truong et al. 2013; Barbato et al. 2011) and extensions from these on three datasets measured over 2 years in households located in Switzerland, Canada and the UK. We implemented these approaches into a common framework and compare their fitness in predicting the usage patterns of the appliances. In doing so, we focused on appliances, whose usage is mainly driven by user behavior and whose switch-on time is flexible. In the relevant literature such appliances are commonly referred to as “shiftable loads”. Examples for such loads are washing machine, dish washer or tumble dryer. The Python source code for the experiments can be obtained from the authors upon request.

Algorithms

The following subsections shortly discuss the main characteristics of the three implemented algorithms. All algorithms have been used to predict the on-off times of appliances with a resolution of 1 h.

Histogram algorithm

Assuming that household activities follow a weekly pattern, one can build up a histogram of on-times of an appliance for each weekday based on the training data (Chrysopoulos et al. 2014; Holub and Sikora 2013). The approach used in this work is shown in Eq. (1). It conditions relevant day-profiles with a Gaussian weighting around the time of interest. In this manner we allow on-events in the past that are not precisely aligned with the time of interest to influence the prediction. Based on the preceding N days each subdivided into T time intervals, the probability that on day n at time t appliance l is running is calculated as

$$p(x_{nlt}) \propto \sum_{m \in N} \sum_{\tau \leq T} w_{nm} e^{-\frac{(t-\tau)^2}{2\sigma^2}} x_{m\tau l} \quad (1)$$

where $x_{m\tau l} = 1$ if appliance l was running during the interval τ on weekday m and $x_{m\tau l} = 0$ otherwise. $w_{nm} = 1$ if $n = m$, and $w_{nm} = 0$ otherwise. The variance σ is a model parameter that was set experimentally, see results section.

Pattern search algorithm

Whereas the histogram-based approaches assume the weekdays to be the governing pattern defining the weights w_{nm} , see Eq. (1), the approach by Barbato (Barbato et al. 2011) tries to identify these patterns. It does so by relying on the redundancy of variably sized day-patterns. To this end, one maps the N days preceding the day to be predicted, n , to a binary array of the form $[\delta_{n-N}, \delta_{n-(N-1)}, \dots, \delta_{n-1}]$ of length N with $\delta_i = 1$ when the appliance was running on day i , and $\delta_i = 0$ otherwise. The sub-array $S_n(i)$ is then defined as $S_n(i) = [\delta_{n-i}, \delta_{n-(i-1)}, \dots, \delta_{n-1}]$ for a given length $1 \leq i < N/2$. The occurrences of both the sub-pattern $S_n(i)$ as well as $[S_n(i), 1] = [\delta_{n-i}, \delta_{n-(i-1)}, \dots, \delta_{n-1}, 1]$ is counted in the original array and the probability of a pattern of length i followed by a conjectured on-day is calculated as

$$s_n(i, 1) = \begin{cases} 0 & \text{if } \#[S_n(i)] = 1 \\ \frac{\#[S_n(i), 1]}{\#[S_n(i)] - 1} & \text{else} \end{cases} \quad (2)$$

and correspondingly $s_n(i, 0)$ for a conjectured off-day (note that $s_n(i, 1) + s_n(i, 0) = 1$ by construction). Now i is increased until either $s_n(i, 1)$ or $s_n(i, 0)$ equals 1. In the latter case, a day without any appliance usage is predicted. Whereas in the former, the days following the occurrences of the pattern $S_n(i)$ define the relevant days used for forecasting. They replace the days with identical weekday as used in the Histogram algorithm. It turns out that for the investigated data, patterns are not as obvious as in (Barbato et al. 2011), i.e. there is typically not an optimal pattern length i resulting in either $s_n(i, 1)$ or $s_n(i, 0)$ being 1. We therefore extended the original approach as can be seen in Eq. (3). Day n is predicted by the sum of the K most probable patterns weighted with the probabilities $s_n(i, \alpha)$.

$$p(x_{nlt}) \propto \sum_{i, \alpha} s_n(i, \alpha) \delta_{\alpha 1} \sum_{m \in N_i, s \leq T} e^{-\frac{(t-s)^2}{2\sigma^2}} x_{msl} \quad (3)$$

where $\sum_{i, \alpha}$ goes over the K most relevant patterns. The Kronecker Delta $\delta_{\alpha 1}$ leads to a zero contribution of the patterns predicting a day with no appliance usage.

Bayesian inference algorithm

The third investigated method (Truong et al. 2013) uses Bayesian inference, which differs fundamentally from the previous approaches. It uses a Markov-Chain Monte-Carlo approach to sample the posteriori distribution of the model parameters. The key elements of the model are the latent day-types k . They are used to create day profiles and to record correlations between the use of individual appliances. In summary, the probability $p(x_{nlt})$ of appliance l running at time t on day n is calculated as

$$p(x_{nlt}) \propto \sum_K p(k | n) \mu_{kl}(t) \quad (4)$$

where k goes over all K day-types and $p(k | n)$ is the probability of day n being described by day-type k . One of the advantages of this approach is that it infers the parameters for each appliance l from the data of all appliances resulting with an effective training set of $N \cdot L$ data points, L being the total number of appliances.

Data and methods

Test data

Various datasets containing electrical consumption data of individual households are available (Murray et al. 2017). The three datasets employed in this investigation are GH9, collected by the authors, AMPds2 (Makonin et al. 2016) and REFIT, House 5 (Murray et al. 2017). They all cover at least two continuous years of data records from a single-family house, stem from Switzerland, Canada and UK respectively, and include sub-metered data for dishwasher, washing machine, and tumble dryer. In order to produce hourly on- and off-times of the appliances, the measurement data was preprocessed by imposing i) minimal on- and off-times i.e. removing noise spikes and preventing double-counting due to intra-cyclic pauses, ii) as well as a minimal power levels. It was then downsampled to hourly intervals.

Performance metric

Binary classifiers can be assessed with a variety of performance metrics. We compare the predictive quality of the tested algorithms on the basis of the so called ReceiverOperator-Characteristics (ROC) curves because of their independence from the relative weight of the ground-truth's classes. The ROC method is well suited for a posteriori measure of the prediction quality but for an actual predictive algorithm a single working point along the curve (i.e. a single fixed threshold) must be chosen in advance. To average the ROC curves over individual samples each predicting the on-off behavior of an appliance during 1 week and estimate the resulting statistical variance, methods described in (Macskassy and Provost 2004) are employed. To allow for simple comparison with other experiments, the ROC curve is integrated, resulting in the area under curve AUC.

Results

Where not otherwise mentioned, results stem from an average over 90 samples, where each individual sample predicted on-off behavior of an appliance during 1 week based on the eight preceding weeks, hence covering in total roughly 2 years of data.

Histogram algorithm

The basic histogram method was tested on all three datasets with the model parameter σ (variance) varying between 0 and 2. Overall best performance with respect to AUC was achieved with $\sigma = 1.3$ which was used for all further experiments. The average performance improves by increasing the training window, i.e. increasing the individual train-sets, but saturates for lengths above about 3 months. As a trade-off between prediction quality and a quickly increasing computational effort for the more elaborate algorithms, a training window of 8 weeks was chosen to ensure comparability of the results. Table 1 summarizes the results. The algorithm generally performs in a medium quality range with AUC-values around 0.7. Differences in the AUC of different

Table 1 Area Under Curve for predicting 1 week based on the preceding 8 weeks. The results are averaged over 2 years (90 samples) with corresponding standard deviation

	Histogram		Pattern Search		Bayesian Inference	
GH9						
Dish-Washer	0.64	± 0.15	0.72	± 0.10	0.72	± 0.10
Tumble Dryer	0.74	± 0.21	0.78	± 0.14	0.71	± 0.08
Washing Machine	0.84	± 0.10	0.75	± 0.09	0.73	± 0.08
AMPds2						
Dish-Washer	0.68	± 0.12	0.66	± 0.10	0.74	± 0.08
Tumble Dryer	0.72	± 0.11	0.73	± 0.08	0.77	± 0.09
Washing Machine	0.63	± 0.17	0.81	± 0.08	0.73	± 0.10
REFIT						
Dish-Washer	0.79	± 0.14	0.72	± 0.15	0.72	± 0.06
Tumble Dryer	0.76	± 0.14	0.66	± 0.13	0.69	± 0.07
Washing Machine	0.74	± 0.09	0.68	± 0.10	0.77	± 0.08

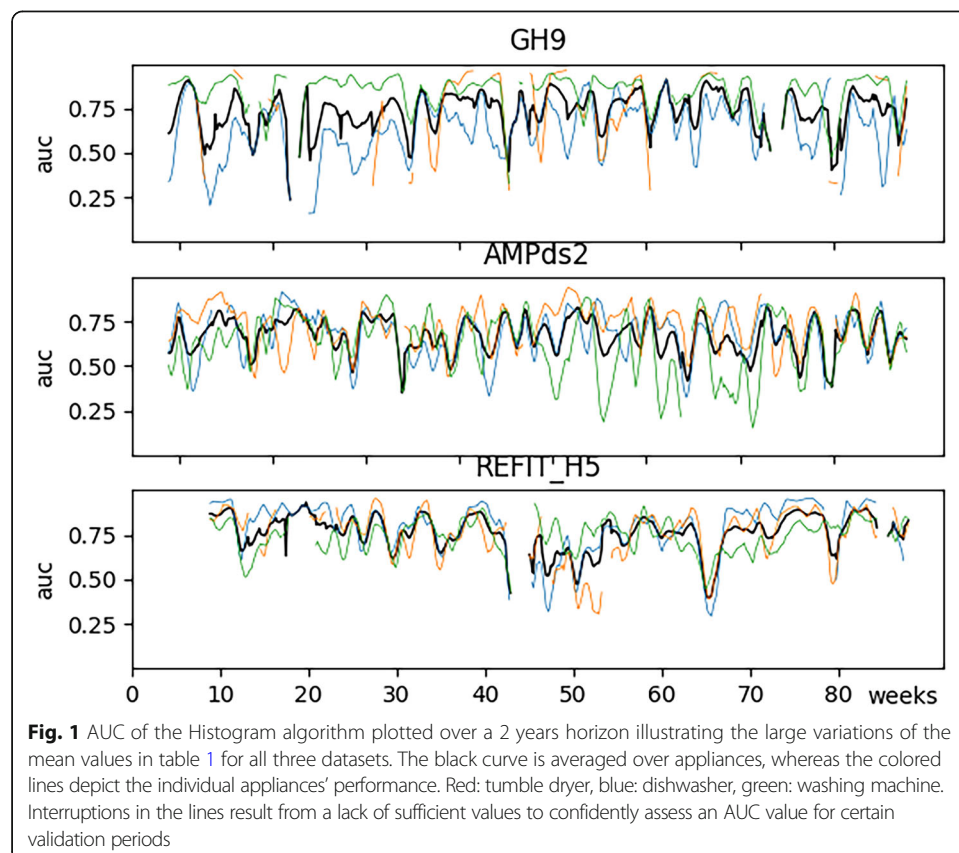
appliances and datasets are large but are not significant due to the large uncertainty as illustrated in Fig. 1.

Pattern search algorithm

Whereas the Histogram algorithm and the Bayesian Inference algorithm can ‘predict’ arbitrarily far into the future, the Pattern Search algorithm adapted from (Barbato et al. 2011) is only able to predict the day immediately following the training set. Thus for the latter, the window of training- and validation-set was not shifted by intervals of a week but day by day. Seven day-predictions were then summarized to a 1 week-prediction. Barbato’s approach has been modified to include the K most relevant patterns. Experimentally we found that $K = 14$ leads to satisfactory performance. As can be seen in Table 1, the AUC values are around 0.7 as for the Histogram algorithm but the standard deviation for the Pattern Search algorithm is mostly reduced.

Bayesian inference algorithm

In contrast to the results discussed so far, results from the Bayesian Inference algorithm originate from averaging not only over 90 samples i.e. validation weeks, but in addition each sample was obtained by averaging the prediction of ten independent Markov Chains. Tests showed a fast convergence of the individual Markov Chains independent of the initialization. With a burn-in period of 500 steps, the individual prediction was calculated by averaging over 2000 Gibbs iterations. Results are summarized in Table 1.



Discussion and conclusions

The results summarized in Table 1 lead to the following observations: i) The overall performance of the three algorithms is essentially the same: averaging all appliances in all houses leads to the following values: 0.72 for Histogram and 0.73 for both Pattern Search and Bayesian Inference algorithms. ii) With increasing complexity of the algorithm not only the variances over 2 years decrease, but also the prediction quality across appliances and data-sets becomes more similar. iii) An algorithm's (relative) performance for a given appliance and data-set does not necessarily relate to another algorithm's performance on the same data-set. That is, an algorithm performing particularly well on a given appliance of a given data-set does not necessarily imply other algorithms to perform similarly, i.e. the governing reason for the large differences does not seem to be the underlying data. iv) Similarly, no statement is possible about certain appliances performing markedly worse or better across all datasets for a specific algorithm, i.e. no particular algorithm is especially good at predicting a certain appliance. As discussed above the results imply that the mean predictive performance is not affected by the choice of the employed model. Because of the complexity of the implementation and computational considerations this would favor the Histogram algorithm over the two others. It performs at least 2–3 orders of magnitude faster than the algorithm based on Bayesian inference. For most real-world applications, it is, however, not the mean performance that counts most but a reliable performance on any given data. Here the reduced variance of the Bayesian algorithm with respect to the Histogram and Pattern Search approach speaks in favor of the former.

From our viewpoint, a limitation of the Bayesian algorithm in its current form is the fact that it strongly relies on weekly patterns despite its introduction of the latent day-types. This could be addressed by making minor changes to include more data such as weather or schedules (Truong et al. 2013). An alternative could be to combine the Bayesian and the Pattern Search algorithms so that day n would also be predicted based upon the inferred day-types of the immediately preceding days.

One aim of this study was to investigate if the on-off times of domestic appliances can be predicted solely based on electrical usage data. From our results, we tend to negate this hypothesis: On a coarse-grained timescale of 1 h, we achieved on average a mediocre prediction performance with a large variance. However, we believe that for domestic load optimization an improved performance and, in particular, a smaller variance of the prediction would be desirable if not necessary. Our choice of algorithms is far from exhaustive and one can think of various improvements of the examined algorithms. Nevertheless, from our point of view, the presented results based on 2 years of data from three different households reflect a general limit for the hourly predictability of an individual household's electrical appliances. We conclude that taking solely electrical data of a single family into account, every stochastic approach must suffer from a lack of information, independently of its complexity.

Abbreviations

AUC: Area under curve; ROC: Receiver operating characteristic

Funding

This work has been financially supported through the Swiss Competence Centers for Energy Research – Future Energy Efficient Buildings and Districts. Publication costs for this article were sponsored by the Smart Energy Showcases - Digital Agenda for the Energy Transition (SINTEG) programme.

Availability of data and materials

The datasets analyzed during the current study are available from (AMPds2) Harvard Dataverse at <https://doi.org/10.7910/DVN/FIE05> and (REFIT) University of Strathclyde's PURE data repository at <https://doi.org/10.15129/9ab14b0e-19ac-4279-938f-27f643078cec>. The datasets GH9 is for the moment only available from the corresponding author on reasonable request.

About this supplement

This article has been published as part of *Energy Informatics* Volume 1 Supplement 1, 2018: Proceedings of the 7th DACH+ Conference on Energy Informatics. The full contents of the supplement are available online at <https://energyinformatics.springeropen.com/articles/supplements/volume-1-supplement-1>.

Authors' contributions

MG implemented the algorithms, and analyzed and interpreted their performance on the different datasets. The manuscript was written jointly by PH and MG and critically revised by both AR and AP. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published: 10 October 2018

References

- Barbato A, Capone A, Rodolfi M, Tagliaferri D (2011) Forecasting the usage of household appliances through power meter sensors for demand management in the smart grid. In: 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm). IEEE, Brussels, pp 404–409
- Chrysopoulos A, Diou C, Symeonidis AL, Mitkas PA (2014) Bottom-up modeling of small-scale energy consumers for effective demand response applications. *Eng Appl Artif Intell* 35:299–315
- Holub O, Sikora M (2013) End user models for residential demand response. In: Innovative Smart Grid Technologies Europe (ISGT Europe) 4th IEEE PES, pp 1–4
- Macskassy S, Provost F (2004) Confidence bands for ROC curves: methods and an empirical study. In: ROC Analysis in AI, First Int Workshop (ROCAI-2004) Valencia Spain
- Makonin S, Ellert B, Bajić IV, Popowich F (2016) Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Sci Data* 3:160037
- Murray D, Stankovic L, Stankovic V (2017) An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Sci Data* 4:160122
- Pflugradt ND (2016) Modellierung von Wasser und Energieverbräuchen in Haushalten. TU Chemnitz
- Truong NC, McInerney J, Tran-Thanh L, Costanza E, Ramchurn SD (2013) Forecasting multi-appliance usage for smart home energy management. *Proc Twenty-Third Int Jt Conf Artif Intell Beijing China* 4:3–9

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
