Dissecting the Problem of Individual Home Power Consumption Prediction using Machine Learning

Enrico Casella Eleanor Sudduth Simone Silvestri
Department of Computer Science
University of Kentucky
Lexington, KY, USA
enrico.casella@uky.edu, eleanor.sudduth@uky.edu, silvestri@cs.uky.edu

2022 IEEE International Conference on Smart Computing (SMARTCOMP) | 978-1-6654-8152-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/SMARTCOMP55677.2022.00037

***Abstract*—The growth and widespread diffusion of Internet-of-Things devices and advanced metering infrastructure allows to closely monitor appliances in a user home. Although only few works have focused on the issue of individual home power consumption predictions, recent efforts have unveiled the complexity of this task. As opposed to building-level power predictions, the finer granularity of single home predictions is characterized by the high impact that individual user actions have on the power consumption. As a matter of fact, the current state of the art shows inadequate prediction performance. In this work, we investigate the issue of single home power prediction by analyzing a recent dataset of real power consumption data. We carry out a profound analysis of several processing parameters and environmental parameters that make this task so challenging, thus providing meaningful insights that can guide future research on individual home power consumption predictions. Results show an overall low daily error, and very accurate hourly predictions when less variable usage patterns occur.**

I. Introduction

Many studies exist that focus on the prediction of aggregated power consumption for multiple residential and commercial buildings. These scenarios allow quite accurate inference, since the unpredictability of human behaviors is balanced out by the aggregated data [1]. The task of single-home power consumption prediction represents an important milestone to further improve demand response and power conservation programs [2]. However, this poses many more challenges, as proven by the low number of works investigating this issue, and related unsatisfactory results.

For instance, in [3], a LTSM model is used to predict hourly and daily power consumption of Heating Ventilation and Air Conditioning (HVAC). However, results measured with R^2 scores are well below 0.5[1]. Similarly, [4] focuses on HVAC using LSTM Neural Networks but does not achieve satisfactory prediction errors. In [5], the focus shifts on the net total power consumption with SVR models, with highly variable performance that show better daily predictions over hourly ones. Nonetheless, other appliances are not investigated. Clearly, most efforts are directed to hourly and daily HVAC or net total prediction, while the power prediction of other appliances and the net total power consumption have not been considered. Overall, the current state of the art does not offer a comprehensive study that analyzes the impact of several factors that provide meaningful insights to guide future research on single home power predictions. Hence, additional investigation is required to investigate the characteristics that make this task so complicated.

In this work, we investigate the potential of machine learning in predicting the power consumption of an individual home. We take a holistic approach that considers multiple dimensions and data processing techniques, including several machine learning algorithms, multiple appliances, time granularity, learning rate, feature impact, etc. We validate our results using a real dataset from Honda Smart Home US [6], an innovative smart home project designed to gather deeper understanding of efficient home building with real tenants. Data is collected at 60 samples per hour from a variety of appliances, beyond just HVAC and net total. Results confirm that predicting individual home power consumption is non-trivial. Nevertheless, we provide several important insights on the ability of machine learning algorithms to solve this problem. To the best of our knowledge, this is the first study that focuses on providing such a comprehensive overview of the problem, and guide future research on the development of single home power prediction systems.

II. Methodology

In Fig. 1, an overview of our methodology is depicted. The main system is represented by a smart house, equipped with widely available smart meters to collect readings from all appliances. A hub, known as the Home Energy Management System (HEMS), collects the information and processes the data. Data is collected at a certain *sampling rate* from all smart meters, say one sample per minute, and stored, before being processed at a certain granularity, e.g. hourly or daily. As samples are collected, data cleaning procedures check for any missing data which may occur due to malfunctioning of the sensors. When a request for power prediction occurs, a time

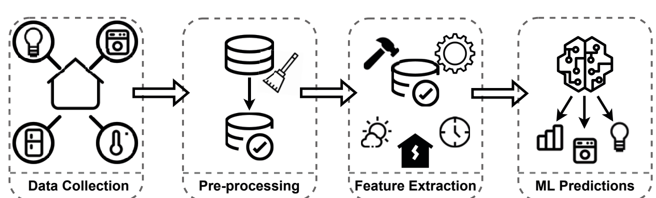


Fig. 1: Pipeline overview of our methodology

[1] $R^2 \in [0, 1]$, higher score represents a better prediction

2693-8340/22/$31.00 ©2022 IEEE
DOI 10.1109/SMARTCOMP55677.2022.00037

frame (i.e., the *hour grouping*) is specified, declaring how far ahead such prediction needs to occur. Hour grouping, sampling rate, and other parameters are set prior to the third step, feature extraction. In fact, at this stage, a processing step called aggregation fits the data into windows with characteristics determined by such parameters. Windows are then passed to the feature extraction step, in which weather information, time information, and power consumption history are processed further before being fed to machine learning models.

Specifically, weather information includes the mean and standard deviation of sun altitude, outdoor temperature, and incident solar radiation, along with the inside temperature of the house. Time information include the hour of the day, day of the week, and a binary value to indicate day light. Finally, power history information consists on the processing of net total, HVAC, lighting, and an additional group that considers all other appliances. For each one of these quantities, we create a brief history of data consisting of six past moments, resulting from a combination of the power consumption during previous hour groupings and previous days. Before feeding this data to the machine learning models for training and testing, additional processing steps such as standardization and quantile transformation are applied to further help the models make accurate predictions.

III. Experimental Section

In this section we provide details about all the information and details related to the experiments, such as the dataset used, and parameters necessary to replicate the experiments.

A. Dataset

The dataset used in this work is provided by Honda Smart Home US [6]. It is a project of a smart home designed for the specific purpose of studying and understanding the complex dynamics of power consumption in a regular home. The house is provided with plenty of sensors and smart meters that allow the collection of data for such scientific purposes. Power data is collected from the beginning of spring, i.e., $03/20/2018$, until august $08/10/2018$, at a sampling rate of 60 samples an hour. Raw data consist of $207,359$ rows and 10 columns.

B. Experimental setup and Preliminary Results

In this work we adopt overlap values of 1, 10, 30, and 60 minutes, and hour groupings of 1, 2, 4, 12, and 24 hours. Furthermore, quantile transformation is characterized by 200 quantiles, and the number of folds for cross validation is set to 10, a value that was chosen to create testing sets consisting of exactly 2 weeks of data. The impact of different hour groupings is shown in Fig. 2, in which we also test performance of a few baseline machine learning algorithms by means of Mean Average Percentage Error (MAPE). A clear trend is depicted in the figure, which shows a decreasing error of predictions as the hour grouping increases. This is due to the fact that, while sudden actions of the house tenants can have quite a big impact at finer granularity, predicting the overall power consumption during a whole day is just slightly

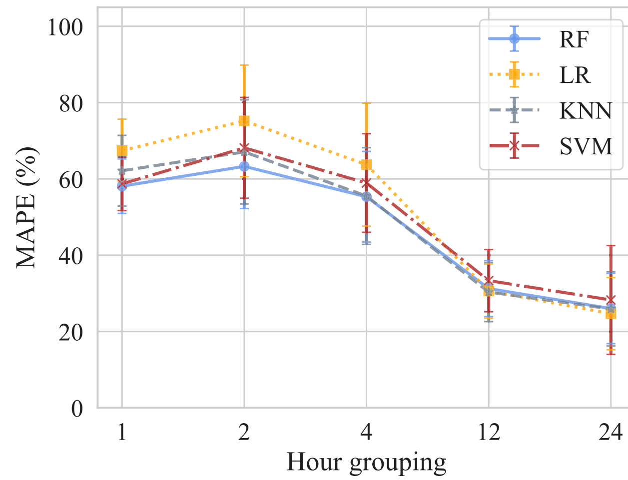
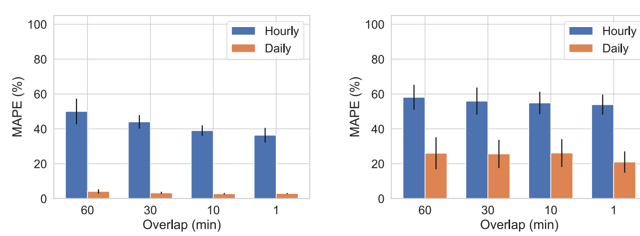


Fig. 2: Impact of hour grouping on net total prediction

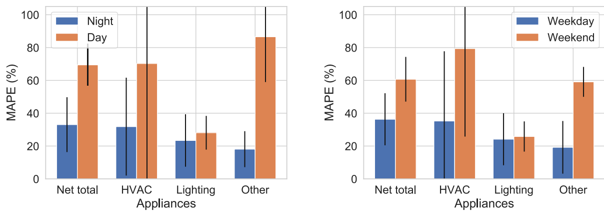
less affected by these events. However, all algorithms seem to behave similarly, with slightly worse performance from Linear Regression (LR). Although some of these predictions show to be not satisfactory, especially on the hourly, it is important to mention that, when the power consumption is close to 0, a very small absolute error can lead to a very big percentage error. When it comes to the daily prediction however, i.e., the 24 hour grouping, our error is in line with the current state of the art. Furthermore, since RF has the best overall performance, we will use it as the baseline algorithm of the following experiments.

Another important detail of the experimental setup concerns the adopted K-fold cross validation, which does not use data randomization. This is an important detail that allows us to create realistic experiments because each testing set contains a set of consecutive data points that never overlaps with the training set. In fact, randomization may lead to a training set that contains several samples adjacent to testing samples, potentially leading to unfairly accurate results. In Fig. 3, we show the impact of overlap on the predictions, and further analyze performance when the training and testing sets are created randomly. As anticipated, randomization widely benefits performance. As shown in Fig. 3a, daily predictions have the lowest error, well within 5% across all overlap values. On the other hand, the impact of overlap on the hourly shows a



(a) Randomized K-Fold (b) Non-randomized K-fold

Fig. 3: Impact on net total prediction of overlap and randomization of samples in K-fold cross validation



(a) Day vs night

(b) Weekday vs weekend

Fig. 4: Prediction performance at different time slices with heterogeneous variability of human behavior

15% improvement in the error prediction. Furthermore, hourly performance in the case of randomization are also better compared to Fig. 3b.

The results depicted in Fig. 3b show that overlap does not have quite the same impact in the performance. Hourly and daily prediction improve, respectively, by 4.1% and 6% when going from 60 minute overlap to 1 minute overlap. Overall, while overlap does have an impact, it is a parameter that relies on the sampling rate of smart meters. For instance, if smart meters collect data at a sampling rate of 4 samples per hour, the most frequent overlap would have to be a 15-minute overlap. Hence, it is worth evaluating whether the HEMS has the capacity to store data at a higher sampling rate, and the processing power to perform timely predictions on the bigger dataset that results from more frequent overlaps. In fact, dataset size may vary largely with sampling rate of 60 samples per hour, and window overlap of one minute.

C. Impact of Environmental Parameters

The following experiment depicted in Fig. 4 shows the prediction performance in a different light. We fix the hour grouping to 1 hour. On the x-axis, we vary the type of appliance that is being recognized, while on the y-axis we show the performance in two different scenarios. In Fig. 4a, we show the performance when the prediction occurs in the day time and in the night time. While in Fig. 4b, we analyze the performance of predicting weekdays and weekends. Unexpectedly, results suggest that time frames characterized by more spontaneous behaviors are affected by a greater error and variability.

Overall, lighting is well predicted in all scenarios. It is clear, however, that the prediction of power consumption at night and during the weekdays is far better than their counterparts. Specifically, HVAC, net total, and appliances categorized as “Other” have roughly 50% (or more) improvement on performance when we predict their consumption during times with more predictable behavior. This further proves that having more detailed information about the user behavior can highly help to make better predictions. However, the current state of the art has not yet found a smart and pervasive solution to solve this problem. Although the patterns and habits of families with extremely fixed routines could be well detected by these algorithms, it is clear that the currently publicly available datasets do not have sufficient information to make accurate predictions, as it is further proven by the performance of the current state of the art.

IV. Conclusions and Future Work

Single home power consumption prediction is still an open issue with many challenges, as proven by the limited research and related unsatisfactory results. Power consumption is highly affected by daily sporadic actions that any of the house tenants may perform. As a consequence, our results, as well as the state of the art, show higher predictability at coarser granularity, such as daily predictions, compared to the finer granularity of hourly predictions. This is the equivalent reason according to which building level power predictions achieve better results than single home power predictions.

Our solution achieves performance in line with the current state of the art. The high impact of user behavior is shown by the higher error that occurs in hourly prediction, as opposed to daily predictions. However, the error percentage on the hourly prediction can be affected by smaller absolute errors. Furthermore, we offer an unprecedented perspective, by investigating and analyzing the impact of a wide set of unexplored parameters. We find that randomizing the selection of training and testing samples has a positive impact on the performance, and that a high frequency of window overlap can help in having slightly better predictions. This is further confirmed by the significantly better performance achieved during period of times with more regular behaviors, such as weekdays rather than weekends, and night time rather than day time.

Hence, in our future research efforts, we aim to investigate the behavioral sphere of users by tracking their routines, as well as the house occupancy. Furthermore, we want to prove the importance of the user impact by performing experiments on the same datasets with and without the behavioral model, and evaluate performance accordingly.

Acknowledgments

This work is partially supported by the National Institute for Food and Agriculture (NIFA) under the grant 2017-67008-26145; the NSF grants EPCN-1936131, CPS-1545037, and CNS-2008878; and the NSF CAREER grant CPS-1943035.

References

[1] K. Amasyali and N. M. El-Gohary, “A review of data-driven building energy consumption prediction studies,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1192–1205, 2018.

[2] E. Casella, A. R. Khamesi, S. Silvestri, D. A. Baker, and S. K. Das, “Hvac power conservation through reverse auctions and machine learning,” in *2022 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 2022.

[3] R. E. Alden, H. Gong, E. S. Jones, C. Ababei, and D. M. Ionel, “Artificial intelligence method for the forecast and separation of total and hvac loads with application to energy management of smart and nze homes,” *IEEE Access*, vol. 9, pp. 160 497–160 509, 2021.

[4] R. Sendra-Arranz and A. Gutiérrez, “A long short-term memory artificial neural network to predict daily hvac consumption in buildings,” *Energy and Buildings*, vol. 216, p. 109952, 2020.

[5] X. M. Zhang, K. Grolinger, M. A. Capretz, and L. Seewald, “Forecasting residential energy consumption: Single household perspective,” in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 110–117.

[6] Honda. Honda smart home US. [Online]. Available: https://www.hondasmarthome.com