Data augmentation for dealing with low sampling rates in NILM

Tai Le Quy, Sergej Zerr, Eirini Ntoutsi, Wolfgang Nejdl
Leibniz University Hannover & L3S Research Center, Hannover, Germany
{tai, szerr, ntoutsi,nejdl}@l3s.de

***Abstract*—Data plays a crucial role in evaluation the performance of NILM algorithms. The best performance of NILM algorithms is achieved with high quality evaluation data. However, many existing real world data sets come with a low sampling quality, and often with gaps, lacking data for some recording periods. As a result, in such data, NILM algorithms can hardly recognize devices and estimate their power consumption properly. An important step towards improving the performance of these energy disaggregation methods is to improve the quality of the data sets. In this paper, we carry out experiments using several methods to increase sampling rate of low sampling rate data. Our results show that augmentation of low frequency data can support the considered NILM algorithms in estimating appliances' consumption with a higher F-score measurement.**

***Index Terms*—NILM, data augmentation, energy disaggregation, temporal disaggregation**

I. Introduction

Non-Intrusive Load Monitoring (NILM) or energy disaggregation is a state-of-the-art technology to disaggregate and estimate the power consumption of individual appliances from the aggregated signal in households or companies. It is preferred over intrusive approaches due to bounded costs compared to the monitoring of each device separately. From the first work of Hart [1], a number of NILM techniques have been proposed [2]. NILM research requires large amounts of high-quality aggregated data [3] [4]. It has been shown that NILM algorithms work efficiently with sampling rate at 1 Hz because at this granularity the data can provide several electricity measurements such as active and reactive power [3] [4]. However, many data sets exist with *lower sampling rates*, like Pecan Street Sample [5] at 1/60 Hz or HES [6] at 1/120 and 1/600 Hz. Recognising devices in such data sets is extremely difficult for the existing NILM algorithms. Besides the low sampling rates, the algorithms have often to deal with *data incompleteness* like missing aggregated signal for some time points or even, for certain time periods. Furthermore, recently, deep learning techniques have shown their potential for NILM, e.g., [7] [8], however, they require huge amounts of training data. As a result, constructing higher sampling rate data even from small amount of the lower samples is an important direction in order to improve the performance of NILM algorithms.

In this research we investigate and propose data augmentation methods in order to construct high sampling rate data from the low sampling rate signal that can be used for NILM. To this end we apply interpolation techniques such as Denton-Cholette method for temporal disaggregation, stepwise method and Cubic spline interpolation on power consumption data of two selected data sets (ECO [3] and iAWE [9]) in order to generate a higher sampling rate data. We then report the results of the performance of two NILM methods (Weiss's algorithm [10] and Parson's algorithm [11]).

TABLE I
DATA SETS FOR NILM

Data set	Institution	Sampling rate
REDD (2011)	MIT	1 Hz and 15 kHz
BLUED (2012)	CMU	12 kHz
Smart* (2012)	UMass	1 Hz
Sample (2013)	Pecan Street	1/60 Hz
HES (2013)	DECC DEFRA	1/120 Hz and 1/600 Hz
iAWE (2013)	III Delhi	1 Hz
ECO (2013)	ETH	1 Hz
UK-DALE (2014)	Imperial College	1/6 Hz and 16 kHz

The rest of the paper is structured as follows: In Section II, we overview the related work. In Section III, we describe the augmentation techniques for generating high sampling rate data from the lower samples. In Section IV, we present our experiment results as the effect of augmentation on the performance of two well-known NILM algorithms. Conclusions and outlook are finally presented in in Section V.

II. Related work

The first NILM method has been introduced by Hart [1] to extract device consumption profiles called signatures. Following this work, different methods have been proposed relying on state analysis (e.g., tracking on/off operation by using real power and reactive power), utilize learning methods or different data granularities [2], [4]. Parson et al. [11] proposed a semi-supervised approach using factorial HMMs which was evaluated on data sets at sampling rate 1/60 Hz. Weiss et al. [10] proposed a supervised approach to extract switching events and find best match in a signature database by using real and reactive power information with granularity 1 Hz.

Unfortunately, the publicly available data sets do not always come with the desired granularity. In fact, data signal collection is a very costly process, in terms of the time required to collect reasonably large data sets but also due to other reasons like privacy. In Table I, we survey the data sets that are often used for NILM evaluation. As depicted, different datasets come with different granularities and for some cases the sampling rate is too low.

In contrast to existing works focussing on better NILM methods to cope with real world electricity data, we follow an approach by proposing data augmentation for NILM in order to generate high granularity samples from low granularity ones. Except for the low sample case, such an augmentation can also help with data incompleteness. Our method-independent approach can work with a variety of different NILM algorithms, because it is applied at the data level.

Data augmentation is summarizing techniques for dealing with data sparsity by deducing missing values using historical information or third party information sources. Recently, it has gained a lot of attention as a way to cope with the huge data demands of complex learning models such as deep neural networks, and it has been proven that for many architectures and different applications, it improves the robustness and the generalization power of the models. Data augmentation techniques increase the volume of the data by generating new instances from the existing instances. In the image domain, for example, this is achieved by applying label-preserving transformations like rotation and illumination [12] to teach a machine model and achieve higher accuracy. Similarly, in the audio domain augmentation is achieved by applying transformations like adding background noise or changing the pitching level [13]. In case of signal data being undersampled, augmentation is related to interpolation, which is used traditionally for e.g., timeseries and trajectory data [14]. In trajectory data, for example, interpolation between successive moving object positions is used in order to simulate the continuous nature of the actual movement.

To our best knowledge, data augmentation for NILM has not been investigated thus far, however our results show that it comprises a promising approach.

III. Generating high sampling rate data from low rate samples

In our work, these augmentation methods are used to generate augmented time series data in between every time-stamp t_i and t_{i+1}.

Stepwise interpolation

Augmented data in between every time-stamp t_i and t_{i+1} is generated by dividing the time-gap of t_i and t_{i+1} into k parts, then we create data for each part by the following formula: $data(part_j) = data(t_i) + \frac{data(t_{i+1}) - data(t_i)}{k} * j,\ j = 1..k$.

Cubic Spline interpolation

This is a form of interpolation using a special type of piecewise polynomial called a spline. In our work, we use a spline function in *Numerical Recipes in C* [1], that returns interpolated values between data at time-stamp t_i and t_{i+1}, the distance to the next interpolated point is calculated by a parameter *t* with range from 0 for time-stamp t_i to 1 for time-stamp t_{i+1}.

Denton-Cholette interpolation is a method for temporal disaggregation, it can disaggregate a low frequency time series data with or without high frequency indicator series. This method is primarily concerned with movement preservation. Augmented data that is similar to the indicator is generated by considering the most common case in the indicator. Mathematical techniques are use to distribute low frequency data in high frequency series when the indicator has a poor quality. In this paper, we use an implementation of this method in a R package by Christoph et al. [15].

Device interpolation

In this method, the changes in the power consumption of devices with high sampling rate are used as indicators to estimate the augmented values for aggregated signal. For every time points in between two time-stamps t_i and t_{i+1}, if the power consumption of any appliance changes beyond a threshold value (in our experiment, we set threshold value from 5 to 10W), we will increase or decrease the aggregated value to the power level of time-stamp t_{i+1}. Otherwise we set the same value for every time series data in between t_i and t_{i+1}.

Fig. 1 visualizes 10 minutes of aggregated data on 01/06/2012 in house 2 of ECO data set with different augmentation methods.

IV. Experiments

We evaluate the effect of data augmentation for NILM on two real data sets (Section IV-A), the evaluation is presented in Section IV-C.

A. Datasets

The **iAWE data set** [9] contains ambient, water and electricity data from a single house in Delhi over a period of 73 days (May-August 2013), 10 jPlugs are used to measure 10 appliances with multiple parameters including voltage, current, phase and frequency. It consists of almost 15M measurements at sampling rate 1 Hz. The average of missing values in this data set is 31%.

The **ECO data set** [3] contains more than 650M measurements from 45 smart plugs in 6 households in Switzerland over a period of 8 months (June 2012-January 2013). The data set is collected at sampling rate 1 Hz with the aggregated consumption data and appliances' consumption. Each of the measurement contains information of power consumption, voltage, current and the phase shift between voltage and current. For each household, there are 6-10 devices connected to smart plugs to measure power consumption in order to obtain ground truth data for analysis. The average of missing values in smart meter data is 0.8%. The proportion of appliance's consumption that is measured by plug meters in household varies from 10% to 80%.

Both datasets contain both real and reactive power measurements which are a prerequisite for several NILM algorithms, e.g., Weiss [10]. For the iAWE dataset, we use the whole observation period (73 days). We evaluate the performance on five appliances: two air conditionals, fidge, laptop-PC and TV. In the ECO data set, each house has a different observation period. Therefore, we select 30 days for each house as follows: House 1, 2: from 01/06/2012 to 30/06/2012, House 3: from 06/12/2012 to 04/01/2013, House 4, 5, 6: from 01/07/2012 to 30/07/2012. We evaluate the performance on 15 devices: dryer, freezer, fridge, water kettle, PC, laptop, dish washer,

[1] http://www.aip.de/groups/soe/local/numres/bookcpdf/c3-3.pdf

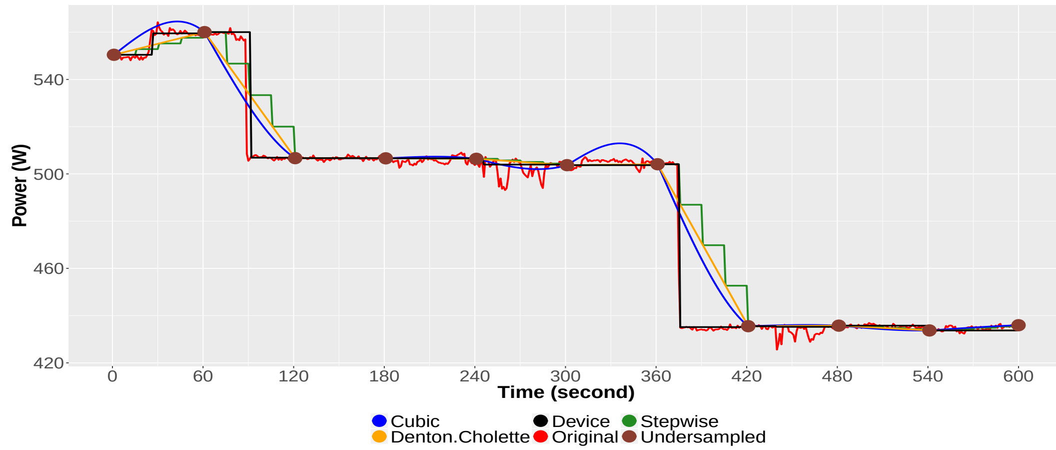

Fig. 1. A sample augmented data with different methods

lamp, microwave, stove, stereo system, TV, coffee machine, entertainment system and fountain.

For evaluation we took the first 2/3 of the recording period from each data set for training and used remaining segment for testing. Resulting in 20 and 50 days training data for ECO and iAWE data sets respectively.

B. Experiment setup

We evaluate the performance, in terms of the F-score on the estimation of power consumption, of two well known NILM algorithms, namely Parson [11] and Weiss [10]. For the experiments, we used the *NILM-Eval* framework [3].

To evaluate the effect of data augmentation for creating high sampling data from low sampling ones, we first *down-sample* the original data sets to the 1/60 Hz granularity by keeping the first second data for each minute. For Parson algorithm, because this algorithm is designed to work with data at sampling rate 1/60 Hz, we down-sample the data to granularity of 1/600 Hz (10 minutes) before evaluate the performance of this algorithm. We refer to the under sampled data set as "undersampled" and to the original data set as "original". For the experiment of Weiss algorithm, we generate the augmented data at granularity 1Hz from the "undersampled" data at 1/60Hz. Parson algorithms use the data at sampling rate 1/60Hz which is reconstructed from the "undersampled" data at granularity 1/600Hz. We also do several experiments to find the best parameter for our augmentation methods. In "stepwise" method, we carried out the experiments with different values of k ($k = 2, 3, 4, 6, 10$) and we found that with $k = 4$ our stepwise interpolation shows the best results across the available datasets. We then compare the performance of the NILM algorithms for the different data sets: "original", "undersampled" and several proposed augmented methods. Our goal is to investigate, how augmentation helps the undersampled data set to reach a performance close to the original high sampling data set.

C. Performance evaluation

Table II describes the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) that shows the difference between the augmented data generated by our augmentation methods and the original data on two data sets. We calculate the average values of MSE and RMSE of 6 households in the ECO data set. These measurements are calculated on the data set at granularity 1Hz.

TABLE II
MSE AND RMSE OF AUGMENTATION METHODS

Methods	ECO data set		iAWE data set	
	MSE	RMSE	MSE	RMSE
Stepwise	85374.7	286.7	137600.8	370.9
Cubic	247425.5	473.7	100464.6	316.9
Denton-Cholette	34921.3	181.9	125423.4	354.1
Device	61021.8	238.6	330394.8	574.8

Comparison of the NILM Algorithm performance on original data: During our experiments we noticed the difference in performance and properties of the used NILM techniques. Whilst Parson algorithm was showing a good performance for smaller period data sets, it was lagging behind for data sets comprising large time periods. The possible explanation is that Parson uses pre-trained models, whilst Weiss needs more time to train, but on the long run it is able to identify more devices, especially outperforming at the "outlier-devices" such as Fountain and Dryer. For similar reasons the smaller **iAWE data set** was dominated by Parson and the larger **ECO data set** by Weiss. In this context, whilst Parson showed high precision for a small number of devices, Weiss was able to identify more devices with a moderate precision.

Down-sampling effect: Separately for each method we measured the effect of the down-sampling. Down-sampling has shown a strong effect on the Weiss method, while it had a light effect on Parson algorithm. This can be explained with the fact that the Weiss method requires information about active power and reactive power as shown for the ECO data set in Figure 2.

Comparison of the augmentation methods: The performance of the augmentation methods (measured by F-score value) on **iAWE data set** are shown in Table III and Figure 3, the results on **ECO data set** are presented in Table IV and Figure 2. Comparing the different augmentation techniques, device interpolation shows the best performance for most of

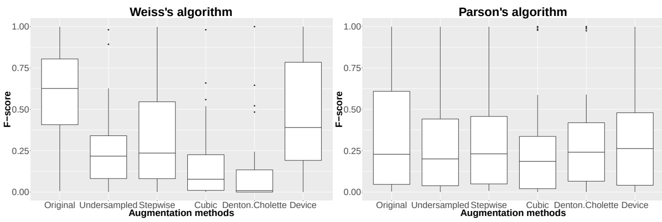

Fig. 2. ECO: Performance of NILM algorithms

TABLE III
iAWE: PERFORMANCE OF NILM ON AUGMENTATION METHODS

Augmentation methods	Weiss's algorithm		Parson's algorithm	
	Avg.	Dev.	Avg.	Dev.
Original	0.736	0.136	0.866	0.007
Undersampled	0.206	0.178	0.877	0.048
Stepwise	**0.416**	0.217	**0.793**	0.099
Cubic	0.145	0.136	0.775	0.085
Denton-Cholette	0.305	0.227	0.752	0.136
Device	0.275	0.175	0.718	0.114

the devices in both Weiss and Parson algorithm. The stepwise method follows, whilst the cubic interpolation shows the worse performance among the remaining augmentation techniques. One of the reasons, as we noticed from the augmented signal results, is the introduced smoothness of the produced timeseries, which can hinder event detection and disturb the inferring power consumption of the appliances.

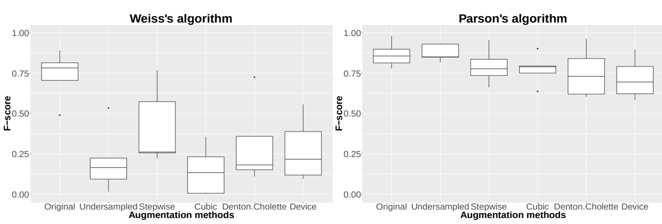

Fig. 3. iAWE: Performance of NILM algorithms

Performance with respect to device type: We categorize devices into five groups based on their function and characteristics: Cooling devices (Freezer, Fridge), Cooking devices (Coffee Machine, Microwave, Water kettle, Stove), Entertainment (TV, Stereo), Computer (PC, Laptop), Lighting device (Lamp). A summary of results for groups of appliances in ECO data set is presented in Table V and Table VI. Device interpolation method showed the best performance among the four methods, although in some cases it failed for Lighting devices, as such appliances consume low amount of electric power. Another observation is that Parson's algorithm can work well with data generated by Device interpolation and Stepwise methods for cooling devices, entertainment and computer, as these appliances do not have significant changes in power consumption over time.

TABLE IV
ECO: PERFORMANCE OF NILM ON AUGMENTATION METHODS

Augmentation methods	Weiss's algorithm		Parson's algorithm	
	Avg.	Dev.	Avg.	Dev.
Original	0.604	0.257	0.359	0.331
Undersampled	0.269	0.24	0.27	0.274
Stepwise	0.332	0.278	0.328	0.307
Cubic	0.176	0.236	0.277	0.323
Denton-Cholette	0.128	0.234	0.325	0.305
Device	**0.467**	0.331	**0.331**	0.31

TABLE V
ECO: PERFORMANCE OF WEISS'S ALG. WITH APPLIANCE'S GROUPS

Methods	Cooling	Cooking	Ent.	Computer	Lighting
Original	0.711	0.613	0.585	0.357	0.281
Undersampled	0.318	0.101	0.396	0.208	0.174
Stepwise	0.482	0.231	0.312	**0.045**	0.157
Cubic	0.273	0.124	0.246	0.021	0.078
Denton-Cholette	0.239	0.037	0.181	0.014	0.06
Device	**0.744**	**0.3**	**0.416**	0.043	**0.193**

TABLE VI
ECO: PERFORMANCE OF PARSON'S ALG. WITH APPLIANCE'S GROUPS

Methods	Cooling	Cooking	Ent.	Computer	Lighting
Original	0.667	0.042	0.41	0.454	0.03
Undersampled	0.504	0.021	0.326	0.306	0.028
Stepwise	0.596	0.048	0.306	**0.485**	**0.037**
Cubic	0.482	0.038	**0.319**	0.373	0.023
Denton-Cholette	0.574	0.047	**0.319**	0.472	0.036
Device	**0.62**	**0.049**	0.309	0.456	**0.037**

V. CONCLUSIONS AND OUTLOOK

In this work we presented an attempt to assist NILM by improving the quality of low sampling rate energy consumption data sets through data augmentation by using several interpolation techniques. Our approach works at the data level and therefore it is method-independent and applicable for a variety of different NILM algorithms. Augmentation was also shown to be helpful for data-intensive machine learning models like deep neural networks which recently have been successfully used also for NILM [7] [8]. Our results show that data augmentation is applicable for increasing the sampling rates of a data set. We believe that this is a promising direction for NILM and further research should be carried, in parallel to the development of new, more sophisticated methods.

In this preliminary work, we adapted simple augmentation techniques, which nevertheless yield improvements over the non-augmented low-sample data sets. As part of our ongoing work, we are investigating more sophisticated augmentation approaches which take into account the consumption profile of the household as well as the profiles of individual devices and do not require the high sampling rate data of appliances. Deep learning technique is also a potential approach that can learn from devices' using pattern in order to construct the aggregated data. Such "informed"-augmentation approaches are expected to yield better augmented data and therefore, better predictions.

REFERENCES

[1] G. W. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
[2] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, pp. 16838–16866, dec 2012.
[3] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "The ECO data set and the performance of non-intrusive load monitoring algorithms," in *BuildSys'14*, 2014.
[4] C. Klemenjak and P. Goldsborough, "Non-intrusive load monitoring: A review and outlook," 2016.
[5] C. Holcomb, "Pecan street inc.: A test-bed for nilm," in *International Workshop on Non-Intrusive Load Monitoring*, pp. 271–288, 2007.

[6] P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, and E. C, "Household electricity survey. a study of domestic electrical product usage," tech. rep., DEFRA, 2012.
[7] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," 2015.
[8] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent lstm network," in *GlobalSIP*, (Piscataway, NJ), pp. 63–67, 2015.
[9] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava, "It's different: Insights into home energy consumption in india," in *ESEEB*, 2013.
[10] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, "Leveraging smart meter data to recognize home appliances.," in *PerCom*, 2012.
[11] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," 2012.
[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012.
[13] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.
[14] J. Macedo, C. Vangenot, W. Othman, N. Pelekis, E. Frentzos, B. Kuijpers, E. Ntoutsi, S. Spaccapietra, and Y. Theodoridis, "Trajectory data models," in *Mobility, Data Mining and Privacy*, pp. 123–150, 2008.
[15] C. Sax and P. Steiner, "Temporal disaggregation of time series," *The R Journal*, vol. 5, pp. 80–87, Dec. 2013.