Energy Informatics

RESEARCH

Open Access

CrossMark

Enhancing neural non-intrusive load monitoring with generative adversarial networks

Kaibin Bao[*], Kanan Ibrahimov, Martin Wagner and Hartmut Schmeck

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

* Correspondence: bao@kit.edu
Karlsruhe Institute of Technology
(KIT), Kaiserstraße 12, 76131
Karlsruhe, Germany

Abstract

The application of Deep Learning methodologies to Non-Intrusive Load Monitoring (NILM) gave rise to a new family of Neural NILM approaches which increasingly outperform traditional NILM approaches. In this extended abstract describing our ongoing research, we analyze recent Neural NILM approaches and our findings imply that these approaches have difficulties in generating valid, reasonably-shaped appliance load profiles. We propose to enhance Neural NILM approaches with appliance load sequence generators trained with a Generative Adversarial Network to mitigate the described problem. The preliminary results of our experiments with Generative Adversarial Networks show the potential of the approach, albeit there is no strong evidence yet that this approach outperforms the examined end-to-end-trained Neural NILM approaches. In the progress of our investigations, we generalize energy-based NILM performance metrics and establish the complete classification confusion matrix based on the estimated energy in appliance load profiles. This enables the adaption of all known classification scores to their energy-based counterparts.

Keywords: Non-intrusive load monitoring, Generative adversarial networks, Neural NILM, Generative modeling, Deep Learning

Introduction

Non-Intrusive Load Monitoring (NILM) (Hart, 1992) describes a source separation problem: the energy usage of single appliances is inferred from the aggregated load of the household measured at the household connection point (mains) (Mauch & Yang, 2016). Another term for NILM is *energy disaggregation* and in this abstract, we call a technique that implements NILM a *disaggregator*. Visualizing energy usage using NILM techniques raises awareness of the energy consumption, without the need of individual meters for each household appliance. However, whether this facilitates energy efficiency and reduces energy cost is disputed (Kelly & Knottenbelt, 2016).

Inspired by the successes of Deep Neural Networks (DNNs) in the fields of computer vision, audio, and natural language processing, DNNs have been applied to NILM (Kelly & Knottenbelt, 2015a; Mauch & Yang, 2015; do Nascimento, 2016; Zhang et al., 2016; Barsim & Yang, 2018), which Kelly coined as *Neural NILM* (Kelly & Knottenbelt,

Bao *et al. Energy Informatics* 2018, **1**(Suppl 1):18

Page 296 of 428

2015a). Recently, Bonfigli (Bonfigli et al., 2018) showed that Kelly's Neural NILM approach is able to outperform state-of-the-art NILM approaches which are not based on DNNs like Additive Factorial Approximate Maximum A Posteriori estimation (AFA-MAP) by Kolter and Jaakkola (Kolter & Jaakkola, 2012).

(Fig. 1) depicts how Neural NILM disaggregation is performed: Assume we have recorded c electrical features (channels) from mains with a fixed temporal resolution for a limited period of time such that we obtain a history of T measurements. Consequently, the measured values $L_M \in \mathbb{R}^{c \times T}$ form a time series with c channels. Current Neural NILM approaches split this time series into segments of fixed length S and run the disaggregation once for each segment, respectively. Later, the partial disaggregation results for each segment have to be merged to form the final result. Neural NILM approaches usually perform the splitting with overlapping sliding windows.

For each appliance type a, a specific disaggregator Y_a is used. This is in contrast to traditional NILM approaches (cf. (Kolter & Jaakkola, 2012; Zeifman & Roth, 2011; Zoha et al., 2012)) where appliance models are merged into a household model before disaggregation is conducted.

Analysis

The quality of NILM approaches can be assessed in two ways. Firstly, whether the disaggregator can correctly *detect* the time intervals when the target appliance consumes energy. Secondly, the degree of precision with which the disaggregator *reproduces* the shape of the target appliance load.

With regard to the first criterion, Kelly's denoising autoencoder (Kelly & Knottenbelt, 2015a) already achieves good results. In most cases, his approach can correctly identify and localize the energy consumption of the target appliance within the aggregated load sequence. However, with regard to the second criterion, the autoencoder has noticeable difficulties.

(Fig. 2) shows the disaggregation result for the autoencoder of the washing machine on a test data window. We show load sequences of the washing machine, as they are complex and consist of multiple stages (heating, washing, spinning, rinsing). Kelly's approach uses a sliding window with a stride of 16 samples in order to split mains into input sequences and applies the autoencoder on each sequence (cf. (Fig. 1)). In (Fig. 2), we see that the disaggregated estimate (left plot) differs from reasonably-shaped appliance load sequences like the measured appliance load. Kelly uses averaging to merge partial disaggregation results (sliding windows). Zhang et al. (Zhang et al., 2016)
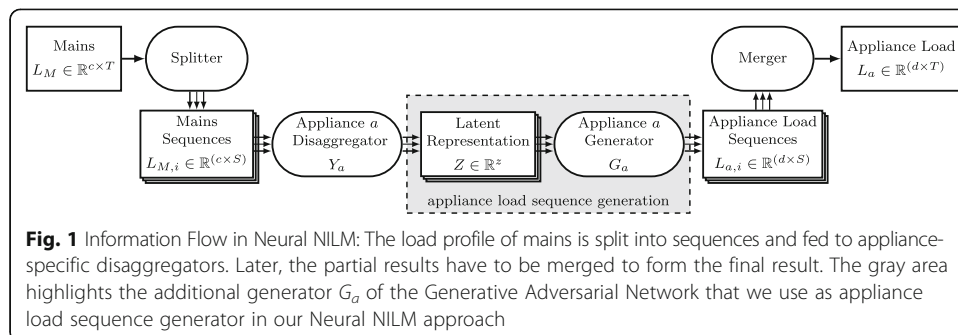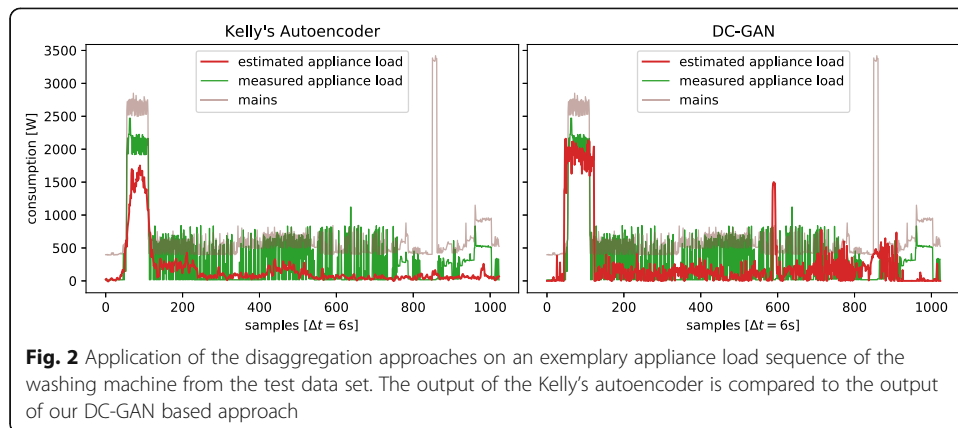

Fig. 1 Information Flow in Neural NILM: The load profile of mains is split into sequences and fed to appliance-specific disaggregators. Later, the partial results have to be merged to form the final result. The gray area highlights the additional generator G_a of the Generative Adversarial Network that we use as appliance load sequence generator in our Neural NILM approach

Fig. 2 Application of the disaggregation approaches on an exemplary appliance load sequence of the washing machine from the test data set. The output of the Kelly's autoencoder is compared to the output of our DC-GAN based approach

criticize this practice and propose that the DNN should only estimate single time points (Sequence-to-Point) instead for a whole target sequence (Sequence-to-Sequence). This eliminates the need of merging multiple estimates for one point in time.

To conclude our analysis, we observe that Kelly's Neural NILM approach is successful at deciding whether the target appliance is active in the aggregate load and is able to localize it, whereas it shows poor performance when the exact appliance load must be estimated. From the human perspective, the result does not seem to be a reasonably-shaped and valid appliance load sequence.

Concept

We propose to mitigate the problem stated in the previous section by using a *generative neural model* for appliance load sequence generation. We pre-train this model using a Generative Adversarial Network (GAN) (Goodfellow et al., 2014) architecture and integrate it into the Neural NILM disaggregation process.

The functional principle of GAN is depicted in (Fig. 3). GAN consists of two neural networks, a generator G and a discriminator D. During disaggregation, we want G to generate load sequences L_a of a specific appliance a. Thereby, the distribution of the generated appliance load sequences L_a should match the distribution of measured appliance load sequences L_a^M as close as possible. For the generation process, G uses a source of randomness Z to express the variations in the distribution of L_a^M. The dimensionality of Z should be high enough to portray all the variations that real appliance load sequences may exhibit. We empirically choose $z = 100$ as an upper bound for the

Fig. 3 A Generative Adversarial Network to generate appliance load sequences
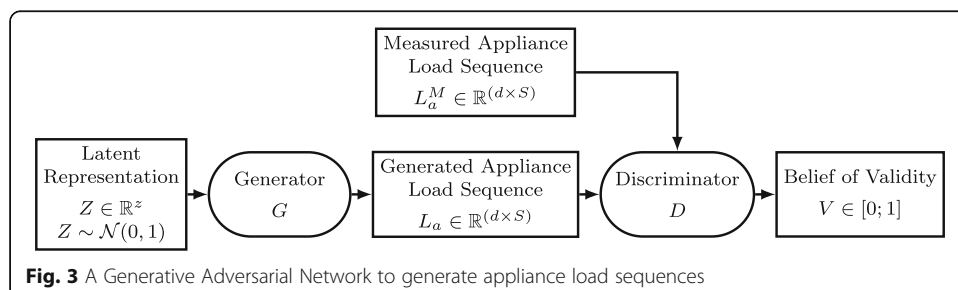
number of variance dimensions. During training, the input for the discriminator D are real appliance load sequences observed in the training data (L_a^M) as well as appliance load sequences generated by G (L_a). D's objective is to determine whether the load sequences were drawn from the training data ($V \coloneqq 1$) or generated by G ($V \coloneqq 0$).

If the GAN training converges, both D and G internalize the distribution of the training data implicitly. Then, Z can be interpreted as a latent representation of an appliance load sequence. G and D are trained simultaneously in an unsupervised manner, where they play a minimax game against each other, hence the name *Adversarial Networks*. The objective of G is to deceive D, i.e. to generate data samples which make D believe that they were drawn from the real data set. D, on the contrary, strives to classify the data samples generated by G as *fake* samples and the data samples drawn from the training data set as *real* samples.

To provide an intuition for the proposed approach, we apply the *manifold assumption* for appliance load sequences: We assume that reasonably-shaped appliance load sequences span a connected low-dimensional subspace (manifold) embedded in $\mathbb{R}^S$, where S is the length of the load sequences we want as output from each disaggregation step.

The training of the generator in the GAN architecture ensures that the output of the generator is located on the manifold of appliance load sequences with high probability. As we integrate the pre-trained generator to the disaggregation process, we force the output of the disaggregator to be located on the manifold of reasonably-shaped load sequences.

As depicted in (Fig. 1), our approach consists of two main components, a disaggregator Y_a and generator G_a for a specific appliance a. During training, G_a learns a self-defined latent representation of the variations in the appliance load sequences. G_a is used to map from that latent representation into the space of reasonably-shaped appliance load sequences.

Compared to previous Neural NILM approaches, the disaggregator Y_a is relieved from the task to generate appliance load sequences. It can focus on the detection and representation tasks, which are already performed sufficiently well by the existing Neural NILM approaches.

In contrast to the works of Barker et al. (Barker et al., 2013) and Buneeva and Reinhardt (Buneeva & Reinhardt, 2017), this approach does not need manual engineering of the characteristics of appliance load sequences. Instead, our approach relies on the ability of DNNs to find load sequence characteristics automatically.

Energy-based performance evaluation metrics

To compare different NILM approaches, we need to define informative metrics that capture specific performance aspects of these approaches. Binary classification metrics are very commonly used in NILM literature (Kelly & Knottenbelt, 2015a; Barsim & Yang, 2018; Bonfigli et al., 2015; Makonin & Popowich, 2015; Faustine et al., n.d.). The practice is to quantize both the appliance load ground truth and the estimate using appliance-specific on/off-thresholds. Unfortunately, these parameters allow to trade-off recall with precision and lead to hardly-comparable results between various NILM approaches. Also, because of the quantization, the information of the detailed load shape

Bao *et al. Energy Informatics* 2018, **1**(Suppl 1):18

Page 299 of 428

gets lost. The metric does not take into account that the shape of the estimated load should match the shape of the ground truth. Therefore, Bonfigli et al. (Bonfigli et al., 2018) propose energy-based precision and recall scores based on the correctly estimated amount of energy in each time interval. We generalize this idea and establish the complete energy-based binary confusion matrix in the following way:

Let $y^{max} > 0$ be the upper load limit of the appliance, $y(t) \geq 0$ be the true appliance load at time t and $\hat{y}(t) \geq 0$ be the load estimate at time t. Then the elements of the confusion matrix are:

$$TP^E = \sum\nolimits_{t=1}^{T} \min(\hat{y}(t), y(t)) \quad , \qquad FP^E = \sum\nolimits_{t=1}^{T} \max(\hat{y}(t) - y(t), 0) \quad ,$$
$$FN^E = \sum\nolimits_{t=1}^{T} \max(y(t) - \hat{y}(t), 0) \quad , \quad TN^E = \sum\nolimits_{t=1}^{T} \min(y^{max} - \hat{y}(t), y^{max} - y(t)) \quad .$$

Now we can define arbitrary energy-based binary classification metrics which do not need an appliance-specific on/off-threshold. Energy-based precision P^E, recall R^E and F_1-score can be determined as follows:

$$P^E = \frac{\sum_{t=1}^{T} \min(\hat{y}(t), y(t))}{\sum_{t=1}^{T} \hat{y}(t)} \quad , \qquad R^E = \frac{\sum_{t=1}^{T} \min(\hat{y}(t), y(t))}{\sum_{t=1}^{T} y(t)} \quad , \qquad F_1^E = 2 \cdot \frac{P^E \cdot R^E}{P^E + R^E} \quad .$$

As Barsim (Barsim & Yang, 2018) points out, the F_1-score does not account for the true negatives and they propose to use Matthews Correlation Coefficient (MCC). An energy-based pendant of MCC can be derived analogously.

Another metric that is able to cope with data imbalances is the balanced accuracy (BACC). Energy-based BACC is defined as follows:

$$BACC^E = \frac{1}{2} \cdot \left(\frac{TP^E}{TP^E + FN^E} + \frac{TN^E}{TN^E + FP^E} \right) \quad .$$

Results

We evaluate our approach using the UK-DALE data set (Kelly & Knottenbelt, 2015b) which consists of electric meter recordings of up to 1.8 years duration from 5 households, sampled at 1/6 Hz. We use the same pre-processing, artificial data augmentation approach, and data partitioning into train, validation and test data folds as described in (Kelly & Knottenbelt, 2015a). Based on Kelly's own re-write of his denoising autoencoder,[1] we re-implemented the neural networks using PyTorch.[2] Our first GAN implementation is based on the Deep Convolutional GAN topology (DC-GAN) by Radford et al. (Radford et al., 2015). The generator and discriminator networks contain five convolutional layers and one fully-connected layer each. The generator uses transposed convolutional layers, which reflects the convolutions of the discriminator. For the disaggregator's topology, we replaced the last layer of Kelly's autoencoder (Kelly & Knottenbelt, 2015a) in order to map to the latent space $\mathbb{R}^z$. The loss function is binary cross entropy for the discriminator and mean squared error for the disaggregator. We use the Adam optimizer (Kingma & Ba, 2014) when training the generator and discriminator. For the disaggregator, we use Stochastic Gradient Descent with Nesterov Momentum.

At first, we tried to train DC-GAN with appliance load data, where each training sample contained an arbitrarily placed load sequence. The training did not converge

Bao *et al. Energy Informatics* 2018, **1**(Suppl 1):18

Page 300 of 428

properly and the DC-GAN could only output sequences with zero load. To mitigate this mode collapse, we trained the DC-GAN only on load sequences which contained a complete appliance activation cycle.

(Fig. 2) shows an example output of our DC-GAN-based disaggregator compared with Kelly's autoencoder (Kelly & Knottenbelt, 2015a), both evaluated on a single observation window. As can be seen, our approach has the potential to reproduce appliance load sequence more accurately than the autoencoder. Because the generator has learned to solely output valid load sequences, its output is more consistent. However, when we compare the F_1 and BACC metrics in (Fig. 4), the overall performance of our DC-GAN-based disaggregator is worse than the autoencoder. As we were forced to train DC-GAN with complete appliance activation cycles, a cause for the worse performance is the inability of DC-GAN to output sequences with zero load. To solve this problem, we applied Auxiliary Classifier GAN (AC-GAN) (Odena et al., 2016). AC-GAN is an extension of GAN, where the generator is conditioned to additional class information. We supply the additional information whether the load sequence has zero load. The F_1-score in (Fig. 4) shows that our approach based on an AC-GAN can improve disaggregation on washing machines in building 2 and 5. Disaggregation in building 1, however, did not outperform Kelly's autoencoder. Also, the balanced accuracy scores do not show a clear advantage of our approach.

Conclusion

In this work, we analyzed Kelly's Neural NILM approach and noticed that it has difficulties in the reproduction of reasonably-shaped appliance load sequences. Based on this insight, we proposed to integrate the generator of a Generative Adversarial Network into the Neural NILM disaggregation process to support a more accurate reproduction of appliance load sequences. To this end, we stated the manifold hypothesis for appliance load sequences and provided a generalization of energy-based NILM performance metrics by defining the complete energy-based confusion matrix. We showed the preliminary results of our ongoing research, which do not yet provide strong evidence that our approach effectively improves Neural NILM. However, we identify promising indications of the potential of the proposed approach.
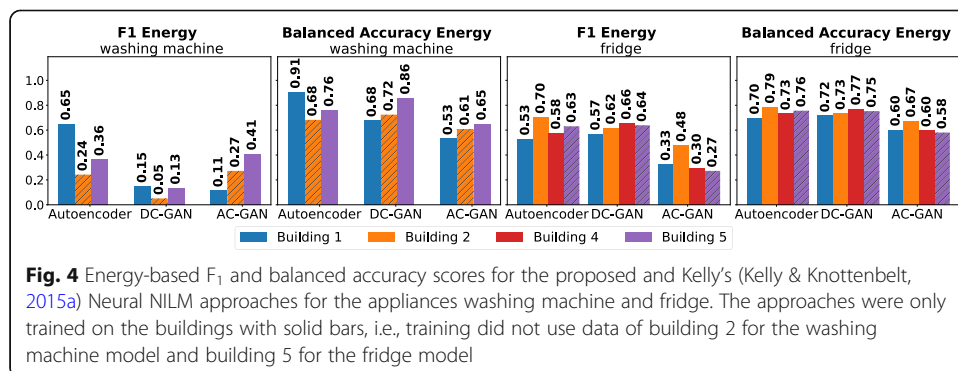

Fig. 4 Energy-based F_1 and balanced accuracy scores for the proposed and Kelly's (Kelly & Knottenbelt, 2015a) Neural NILM approaches for the appliances washing machine and fridge. The approaches were only trained on the buildings with solid bars, i.e., training did not use data of building 2 for the washing machine model and building 5 for the fridge model

Bao *et al. Energy Informatics* 2018, **1**(Suppl 1):18

Page 301 of 428

Endnotes

[1] https://github.com/JackKelly/neuralnilm

[2] https://pytorch.org

Availability of data and materials
The data set analyzed during the current study is available in the UK Energy Research Centre repository, https://doi.org/10.5286/UKERC.EDC.000001

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
KB introduced the idea to use a Generative Adversarial Network to model appliance load profiles in NILM and the idea to complement the energy-based confusion matrix. He implemented the experimentation framework in PyTorch (based on Kelly's Data Pipeline) and drafted most of the manuscript. KI analyzed current Neural NILM approaches, implemented the GAN-based disaggregation approaches, proposed the AC-GAN-based approach and conducted the experiments. MW helped to write the manuscript, provided the result plot and assisted in the execution of the experiments. HS provided supervision, organization of funding and resources for this work. He also helped to write the final version of this publication. 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
Barker S, Kalra S, Irwin D, Shenoy P (2013) Empirical characterization and modeling of electrical loads in smart homes. In: 2013 international green computing conference proceedings, pp 1–10. https://doi.org/10.1109/IGCC.2013.6604512

Barsim, K.S., Yang, B.: On the Feasibility of Generic Deep Disaggregation for Single-Load Extraction (2018). 1802.02139

Bonfigli R, Felicetti A, Principi E, Fagiani M, Squartini S, Piazza F (2018) Denoising autoencoders for non-intrusive load monitoring: improvements and comparative evaluation. Energy Buildings 158:1461–1474

Bonfigli R, Squartini S, Fagiani M, Piazza F (2015) Unsupervised algorithms for non-intrusive load monitoring: an up-to-date overview. In: Environment and Electrical Engineering (EEEIC), 2015 IEEE 15th International Conference on, pp 1175–1180 IEEE

Buneeva N, Reinhardt A (2017) AMBAL: Realistic load signature generation for load disaggregation performance evaluation, pp 443–448. https://doi.org/10.1109/SmartGridComm.2017.8340657

do Nascimento, P.P.M.: Applications of deep learning techniques on NILM. Phdthesis, Universidade Federal do Rio de Janeiro (2016)

Faustine, A., Mvungi, N.H., Kaijage, S., Michael, K.: A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem.(n.d.) 1703.00785

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative Adversarial Nets. In: Advances in neural information processing systems, pp 2672–2680

Hart GW (1992) Nonintrusive appliance load monitoring. Proc IEEE 80(12):1870–1891

Kelly J, Knottenbelt W (2015a) Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, pp 55–64 ACM

Kelly J, Knottenbelt W (2015b) The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. Scientific data 2:150007

Kelly, J., Knottenbelt, W.: Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature (2016). 1605.00962

Kingma DP, Ba J (2014) Adam: A method for stochastic optimization 1412.6980

Kolter JZ, Jaakkola T (2012) Approximate inference in additive factorial HMMs with application to energy disaggregation. In: Lawrence ND, Girolami M (eds) Proceedings of the fifteenth international conference on artificial intelligence and

Bao *et al. Energy Informatics* 2018, **1**(Suppl 1):18

Page 302 of 428

statistics. Proceedings of machine learning research, vol. 22. PMLR, La Palma, Canary Islands, pp 1472–1482 http://proceedings.mlr.press/v22/zico12.html

Makonin S, Popowich F (2015) Nonintrusive load monitoring (NILM) performance evaluation: A unified approach for accuracy reporting. Energy Efficiency 8(4):809–814

Mauch L, Yang B (2015) A new approach for supervised power disaggregation by using a deep recurrent LSTM network. In: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp 63–67 IEEE

Mauch L, Yang B (2016) A novel dnn-hmm-based approach for extracting single loads from aggregate power signals. In: Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference On, pp 2384–2388 IEEE

Odena, A., Olah, C., Shlens, J.: Conditional Image Synthesis With Auxiliary Classifier GANs (2016). 1610.09585

Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks (2015). 1511.06434

Zeifman M, Roth K (2011) Nonintrusive appliance load monitoring: review and outlook. IEEE Trans Consum Electron 57(1):76–84

Zhang, C., Zhong, M., Wang, Z., Goddard, N., Sutton, C.: Sequence-to-point learning with neural networks for nonintrusive load monitoring (2016). 1612.09106

Zoha A, Gluhak A, Imran MA, Rajasegarar S (2012) Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. Sensors 12(12):16838–16866