

RESEARCH

Open Access



On performance evaluation and machine learning approaches in non-intrusive load monitoring

Christoph Klemenjak

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

Correspondence: christoph.klemenjak@aau.at
Alpen-Adria-Universität Klagenfurt,
Universitätsstraße 65-67, 9020
Klagenfurt, Austria

Abstract

Non-Intrusive Load Monitoring (NILM) is a set of techniques to gain deep insights into workflows inside buildings based on data provided by smart meters. In this way, the combined consumption needs only to be monitored at a single, central point in the household, providing advantages such as reduced costs for metering equipment. Over the years, a plethora of load monitoring algorithms has been proposed comprising approaches based on Hidden Markov Models (HMM), algorithms based on combinatorial optimisation, and more recently, approaches based on machine learning. However, reproducibility, comparability, and performance evaluation remain open research issues since there is no standardised way researchers evaluate their approaches and report performance. In this paper, the author points out open research issues of performance evaluation in NILM, presents a short survey of deep learning approaches for NILM, and formulates research questions related to open issues in NILM. An outline of future work is given including applied methodology and expected findings.

Keywords: Non-intrusive load monitoring, Performance evaluation, Machine learning, Deep learning

Motivation

The recent boost of smart meter installations in households and small businesses has led to increased interest in load monitoring techniques such as Non-Intrusive Load Monitoring (NILM). Based on smart meter data, these techniques provide deep insights into energy consumption and processes inside buildings. Furthermore, NILM allows occupancy detection for health-monitoring purposes (elderly care), enables prediction of maintenance windows for selected appliances, allows the optimisation of workflows inside industrial buildings, and aims to achieve cost reduction by providing (immediate) user feedback. Researchers find that a consensus regarding which performance metrics should be applied to measure and report performance has not been reached (Faustine et al., 2017; Pereira & Nunes, 2018). It has been pointed out repeatedly that standardising NILM performance metrics is one of the biggest research issues related to NILM (Faustine et al., 2017). Beside performance metrics, the used datasets for training and evaluation as well as the applied methodology influence if an objective

comparison of two candidate algorithms is possible or not (Nalmpantis & Vrakas, 2018). A requirements catalogue similar to the Zeifman requirements (Zeifman, 2012), a list of requirements that describe what characteristics a NILM algorithm should have, is likely to ease objective comparisons by providing clear guidelines how meaningful comparisons of several NILM approaches can be drawn. Recently, machine learning in NILM has gained popularity due to first promising research contributions, which indicate that machine learning algorithms have the potential to surpass existing HMM-based algorithms (Kelly & Knottenbelt, 2015; Bonfigli et al., 2018; Kim et al., 2017). In addition, these studies revealed one special aspect of machine learning approaches for load monitoring: good generalisation abilities. Hitherto, neither a comparison case study evaluating the generalisation abilities of existing NILM algorithms was conducted nor a machine-learning NILM algorithm was developed that shows acceptable performance on unseen smart meter data.

In this paper, the author highlights open research issues of performance evaluation in Non-Intrusive Load Monitoring (NILM), presents a short survey of deep learning approaches for NILM, formulates research questions related to the presented research problems, and gives an outline of future work.

Related work

Performance evaluation and comparison of NILM algorithms remain open research challenges for several reasons (Pereira & Nunes, 2018; Herrero et al., 2017). It is common practise that researchers evaluate their proposed NILM solutions on different datasets, with different criteria, and with the help of different metrics. From this follows that a direct comparison between two proposed algorithms is virtually impossible (Nalmpantis & Vrakas, 2018). To assess the validity of their proposed NILM approach, many researchers utilise the Zeifman requirements (Zeifman, 2012). These requirements serve to evaluate if a NILM method is applicable to home energy displays or smart meters and comprise requirements related to accuracy, real-time capabilities, need for training, scalability, etc. A requirements catalogue similar to the Zeifman requirements could serve as a guideline for a fair and meaningful performance reporting in the context of load disaggregation. To the best of our knowledge, such a requirements catalogue has not been proposed.

When comparing the performance of NILM algorithms, several aspects play an important role: datasets, metrics, and benchmarking tools. Energy consumption datasets are the outcome of measurement campaigns in households and industrial facilities. The aim is to not disrupt the everyday routines of the monitored space, so that the collected data resembles reality as close as possible (Pereira & Nunes, 2018). In order to enable reproducibility of results and comparison to other algorithms, researchers need to describe in detail the sections of the dataset that were used for training and evaluation and report the method applied to clean and pre-process the datasets (Makonin & Popowich, 2015). As in other data-driven approaches, the performance of NILM approaches highly depends on the datasets used for training and evaluation (Beckel et al., 2014). Therefore, detailed statistics of the utilised datasets should be made available alongside with a published approach. Beside commonly-mentioned aspects such as duration or number of appliances embedded in training data, researchers proposed reporting NILM-specific aspects. To the best of our knowledge, there is no exploratory study that

investigates in the suggestions made by researchers with respect to dataset statistics and standardised performance metrics in an extensive manner by applying several state-of-the-art NILM algorithms to multiple energy consumption datasets.

In the recent past, machine learning approaches for NILM have attracted a lot of attention due to breakthroughs in research disciplines such as computer vision. The authors of (Kelly & Knottenbelt, 2015) are the first to evaluate the application of deep neural networks for energy disaggregation. Three deep neural network architectures are adapted for energy disaggregation. Experiments are performed against unseen household consumption data and against data seen during training. In the presented case study, the deep neural networks achieved better F1 scores than two reference models and that all three networks achieve acceptable performance when applied to an unseen house (Kelly & Knottenbelt, 2015). The authors point out that there are many open issues such as overfitting or unsupervised pre-training. A feasibility study on the development of a generic disaggregation model is presented in (Barsim & Yang, 2018). The authors demonstrate that their generic deep disaggregation model is able to achieve similar performance as state-of-the-art load monitoring approaches for a selection of appliance types. For single-load extraction, a fully-convolutional neural network with a fixed architecture and set of hyper-parameters was applied. Investigations such as presented in (Beckel et al., 2014) don't consider machine learning approaches in their comparison case studies for NILM. Particularly with regard to recent suggestions of related work for improved comparability in NILM, the authors identify the need for an extensive comparison case study that considers well-established NILM algorithms based on Hidden Markov Models as well as novel machine learning approaches based on deep neural networks. Such an extensive comparison case study should evaluate the candidate NILM approaches on several datasets and consider suggestions made by related work such as the proposed performance evaluation strategy and disaggregation complexity.

Research questions

The research aims of current and future investigations are to identify requirements for a fair and meaningful comparison of NILM algorithms, to explore how and to what extent existing machine learning approaches for NILM can be enhanced, and to study under which circumstances an enhanced machine learning approach could be adapted for applications similar to NILM. As pointed out in related work, there is no consistent way researchers are measuring and reporting the performance of NILM algorithms. Furthermore, a recent review finds that drawing a direct comparison is virtually impossible at the moment. A requirements catalogue could serve as a guideline for fair and meaningful comparisons of several NILM algorithms by highlighting vital aspects that have to be considered such as dataset complexity, data noise, or bias.

RQ1: With regard to datasets and performance metrics, what requirements have to be met when comparing NILM algorithms and which factors might influence the outcome?

We hypothesise that on the basis of a requirements catalogue, a meaningful comparison of existing and future NILM approaches can be drawn, which is one of our objectives. In contrast to comparison studies carried out so far, our investigation aims to consider beside approved aspects also novel aspects that consider how complex the disaggregation problem included in dataset X is or how well algorithm Y performs on unseen data.

RQ2: In a comparison of selected existing NILM approaches, including novel approaches based on machine learning, which approach shows the highest accuracy and generalisation abilities across the data sets REDD, UK-DALE and Dataport? How does the novel requirements catalogue affect the outcome of a comparison of NILM approaches?

As related work indicates, novel machine learning approaches for Non-Intrusive Load Monitoring has the potential to surpass existing algorithms in this field but require further improvement to reduce the performance gap significantly.

RQ3: To what extent can existing machine learning approaches for NILM be enhanced for improved accuracy on seen and unseen scenarios?

Material and methods

NILM algorithms are trained and tested on energy consumption data sets. Such data sets include aggregate-level energy readings from smart meters as well as appliance-level energy readings from measurement equipment such as smart plugs. In the course of the years, a vast number of publicly-available data sets have been released. During the planned investigation, the author plans to use the data sets REDD, UK-DALE, and Dataport to train the algorithms and perform evaluations on, which were used in related work as well. In order to evaluate the performance of NILM implementations, adequate benchmarking toolkits are required to process the training data, train the algorithm, and perform evaluations. With NILMTK, an open-source toolkit was designed specifically to enable the comparison of energy disaggregation algorithms in a reproducible manner (Batra et al., 2014). NILMTK will serve as the testing environment and the authors aim to extend it with selected NILM algorithms and functionalities to evaluate the approaches. The author aims to conduct a literature survey to identify crucial requirements that enable a fair and meaningful comparison of NILM algorithms. The expected output of the literature survey is a requirements catalogue for comparing NILM algorithms and will answer research question 1. In order to address research question 2, the author plans to conduct a comprehensive case study on several real-world energy consumption data sets. The case study aims to determine the accuracy as well as the generalisation abilities of existing NILM algorithms on the data sets REDD, UK-DALE, and Dataport. In contrast to related work, the planned study takes into account novel aspects such as the metrics and evaluation approach of (Makonin & Popowich, 2015), the disaggregation complexity of (Egarter et al., 2015), and generalisation abilities (Nalmpantis & Vrakas, 2018). In order to answer research question 3, the author plans to apply a design science approach as research method. During the design science process, the author aims to improve accuracy and generalisation of a particular machine learning algorithm for NILM by re-designing the respective approach such that the existing performance gap of machine learning algorithms for the NILM problem can be reduced in order to make them applicable to real-world scenarios.

Conclusion

In this paper, the author presented motivation, research questions, and methodology related to his current and future investigations. A comprehensive overview of related work points where the author aims to contribute to the state of the art. In particular, planned research activities aim to contribute to the open research issue of comparability

in Non-Intrusive Load Monitoring (NILM). Additionally, the author aims to investigate in novel ways to enhance machine learning techniques for low-frequency NILM. Further, the author aims to examine how and to what extent obtained techniques are applicable to NILM-alike problems of Data Analytics for Smart Microgrids.

Abbreviations

HMM: Hidden Markov Models; NILM: Non-Intrusive Load Monitoring

Acknowledgements

The author would like to thank Professor Thorsten Staake for providing detailed feedback and valuable comments during the incremental revision process.

Funding

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

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

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>.

Author's contributions

CK analysed related work, identified open issues, and developed a research proposal related to his PhD project. The author have read and approved the final manuscript.

Competing interests

The author declares that he has 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

- Barsim KS, Yang B (2018) On the Feasibility of Generic Deep Disaggregation for Single-Load Extraction. arXiv preprint arXiv. <https://arxiv.org/abs/1802.02139>
- Batra N, Kelly J, Parson O, Dutta H, Knottenbelt W, Rogers A, et al. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In: Proceedings of the 5th International Conference on Future Energy Systems, e-Energy '14. New York: ACM; 2014. p. 265–276. Available from: <http://doi.acm.org/10.1145/2602044.2602051>
- Beckel C, Kleiminger W, Cicchetti R, Staake T, Santini S (2014) The ECO data set and the performance of non-intrusive load monitoring algorithms. In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings. New York: ACM, pp 80–89
- 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 and Buildings* 158:1461–1474
- Egarter D, Pöschacker M, Elmenreich W (2015) Complexity of power draws for load disaggregation. arXiv preprint arXiv. <https://arxiv.org/abs/1501.02954>
- Faustine A, Mvungi NH, Kaijage S, Michael K. A Survey on Non-Intrusive Load Monitoring Methodies and Techniques for Energy Disaggregation Problem. arXiv preprint arXiv:170300785. 2017;
- Herrero JR, Murciego AL, Barriuso AL, de la Iglesia DH, Gonzalez GV, Rodriguez JMC et al (2017) Non intrusive load monitoring (NILM): a state of the art. In: International Conference on Practical Applications of Agents and Multi-Agent Systems. Cham: Springer, pp 125–138
- Kelly J, Knottenbelt W. 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 BuildSys '15. New York: ACM. 2015. p. 55–64. Available from: <http://doi.acm.org/10.1145/2821650.2821672>
- Kim J, Le TTH, Kim H (2017) Nonintrusive load monitoring based on advanced deep learning and novel signature. *Comput Intell Neurosci* 2017, pp 1–22
- Makonin S, Popowich F (2015) Nonintrusive load monitoring (NILM) performance evaluation. *Energy Efficiency* 8(4):809–814
- Nalmpantis C, Vrakas D (2018) Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparison. *Artificial Intelligence Review*. p. 1–27. <https://link.springer.com/article/10.1007/s10462-018-9613-7>
- Pereira L, Nunes N (2018) Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p e1265
- Zeifman M (2012) Disaggregation of home energy display data using probabilistic approach. *IEEE Transactions on Consumer Electronics* 58(1):23–31