Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 83 (2019) 670–674

www.elsevier.com/locate/procedia

11th CIRP Conference on Industrial Product-Service Systems

Non-intrusive Load Monitoring based on Convolutional Neural Network with Differential Input

Yuanmeng Zhang[a,b], Guanghua Yang[a,*], Shaodan Ma[b]

[a]Institute of Physical Internet, Jinan University (Zhuhai Campus), Zhuhai 519070, China.
[b]Department of Electrical and Computer Engineering, University of Macau, Macau 999078, China

* Corresponding author. Tel.: +86-756-8505687. *E-mail address:* ghyang@jnu.edu.cn

Abstract

Non-intrusive load monitoring (NILM) is a process for analyzing load in a building and deducing what appliances are working as well as their individual energy consumption. Compared with intrusive load monitoring, NILM is low cost, easy to deploy, and flexible. NILM installed in smart grids can provide information for decision making for energy management and therefore support energy-related industrial services. In this paper, we propose a NILM-based energy management system for appliance-level load monitoring service and a convolutional neural network based model with differential input. Experiment shows that the proposed model with differential input outperforms the existing models with raw input.

Peer-review under responsibility of the scientific committee of the 11th CIRP Conference on Industrial Product-Service Systems

Keywords: Energy Management System (EMS); Non-Intrusive Load Monitoring (NILM); Differential input; Convolutional neural network (CNN);

1. Introduction

With the widespread application of smart grid (SG), energy management system (EMS) which plays a key role in achieving the advantages of an SG, has attracted much interests. EMS monitors, controls and optimizes the performance of energy transmission, distribution, and consumption. For an efficient EMS, not only the aggregate building-level consumption needs to be monitored, but also the appliance-level consumption does. Traditionally, we deploy a sensor on each appliance to monitor its energy consumption, which is called intrusive load monitoring (ILM). ILM suffers from high cost of sensor deployment and needs the assistant of communication module with extra cost. A better way is to disaggregate the load (usually power) of each appliance from the aggregated load of a building, which is called non-intrusive load monitoring (NILM). NILM is promising in smart grid because of the benefits of low cost, easy deployment and more flexibility. Thanks to the NILM technology, service providers can play more roles in providing appliance-level services for stable and energy-efficient production.

1.1. EMS in industrial services

Some of the existing EMSs, such as PERSON [2], WattDepot [3], are the frameworks mainly designed for the customer domain. In this work, referring to the PERSON, we propose an EMS framework which is based on NILM technology and mainly focus on the service provider domain.

This EMS consists of four modules: Data Collector, Information Processer, Information Notifier and Remote Controller. The Data Collector collects the data from the customer, and the Information Processer processes and analyzes the collected information and then make decision for customized service. Finally, service provider provides services such as recommendations and alerts on energy use and equipment status for customers via Information Notifier. Meanwhile, provider can also conduct the control for customer via Remote Controller. The diagram for the proposed EMS is shown in Fig. 1.

The Information Processer of EMS makes use of NILM technology to monitor the equipment status at customer side and makes the optimization policy for the appliances.

2212-8271 © 2019 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 11th CIRP Conference on Industrial Product-Service Systems.
10.1016/j.procir.2019.04.110

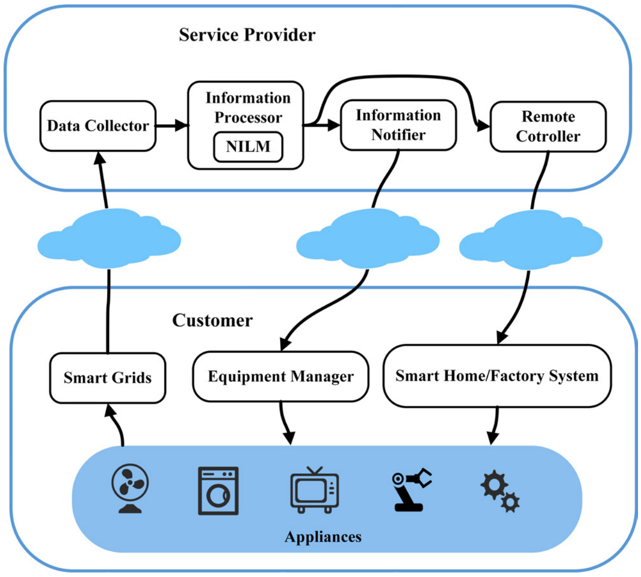

Fig. 1. The diagram of the proposed EMS.

One application is to monitor the equipment and alert when the equipment runs abnormally (overload and underload, etc). Since the load of the equipment reflects its operating status in some degree, it can be used to diagnose the health of the equipment. If only the building-level load is available, it is difficult to detect anomalies. Thanks to the NILM technology, appliance-level load tracking and diagnosis becomes feasible. In the proposed EMS, service provider alert and advise customers via the Information Notifier. Thus, the Equipment Manager at customer side can pay attention to or maintain the equipment before a more serious problem occurs. This service greatly reduces the fault rate of an equipment and improves efficiency and stability of the production.

Another application is demand side management to balance the power supply and load of customer or even its appliances during peak hours. Specifically, based on the results of NILM, Data Processer decides which appliances should be managed and how to manage them (shut down or properly reduce power, etc). In principle, only the appliance that have little impact on the service quality for customer should be managed. Remote Controller conducts the optimization of the appliance load for the customer, provided that the customer has authorized the control of equipment to the service provider. Economic incentives can be given to the customers who are willing to participate in the demand side management. An agent control can be applied in which the Remote Controller controls via the customer's own control system (such as smart home system or smart factory system), rather than directly control the customer's appliances. Usually the control system at customer side has already been established by the customer himself. Agent control costs less for service provider, intrudes less on the customer and makes the provider not affected by the addition or removal of the device at customer side. NILM plays a key role in the proposed EMS, since the optimization policy depends on the results of NILM.

The value of appliance-level data is not limited to this. In the commercial environment, appliance-level data is quite promising, and especially with the rise of artificial intelligence, more possibilities will be exploited.

1.2. NILM

NILM was first studied by Hart [1] in the 1980s. Hart models NILM as a combinatorial optimization problem. Later, some steady-state methods such as HMM [13,14] is proposed which considered the transition of steady states. However, steady-state methods ignore the transient information. Recently, the methods based on artificial neural network (ANN) have been applied with some success. These methods attempt to use neural networks to automatically extract rich power features to describe the behavior of the appliance. Kelly et al. [4] propose several models based on recurrent neural network (RNN) and Convolutional neural network (CNN), which proves the success of ANN-based method in NILM problem. Chaoyun et al. [6] propose a CNN-based model called sequence-to-point learning where the input is a window of the mains and the output is a single point of the target appliance. The above researches feed the raw aggregate data directly to the neural network as input. In this article, we propose to use the differential value of the raw data as the input of the network. We analyze the model theoretically and empirically, showing that the network using differential value as input outperforms that using raw value as input. As for the predictive model, we propose a model based on CNN as the disaggregator.

The rest of this paper is organized as follows: In section 2 we introduce the proposed model including differential input and network architecture. In section 3 the configuration of experiment is described and then the results are followed. Finally, we make a conclusion on the results and discuss the future work for improving NILM performance in section 4.

2. Proposed Model

In this section, convolutional neural network is introduced and then the input strategy is discussed. Finally, a model based on convolutional neural network is proposed.

2.1. Convolutional Neural Network

Convolutional neural networks have been successfully utilized in many fields. CNN is composed of several filters (also known as kernels) which can be regarded as feature detectors. Filters can identify specific short sequence and therefore be used to detect the specific power variation sequence coming from the target appliance and filter out that coming from the non-target appliance. The parameters of the filters are automatically learned by CNN during the training. For the appliance with more complex power dynamics such as multi-state or even time-varying power, the number of filters of the network should be increased in order to learn and extract more abundant power variation features.

2.2. Input Strategy

Existing approaches only disaggregate the target appliance, while other appliances are regarded as noise. The disaggregation can be regarded as a de-noising process. Since the number of other appliances and the power of each appliance are arbitrary, the distribution of the noise is indeterminate. But if we differentiate the raw aggregate power, the power change of target appliance will be distinguishable. This is because that

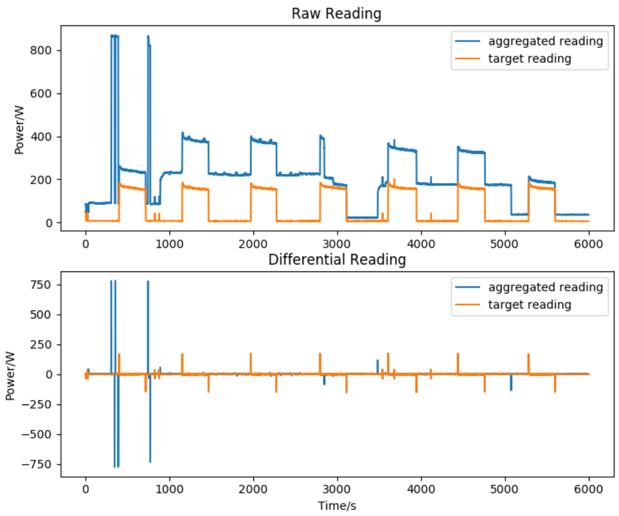

Fig. 2. (a) in the raw reading, the target appliance (Fridge) is "masked" by the appliance which makes it difficult to extract the feature of target appliance; (b) while in the differential reading, the power changes of the target appliance (orange) can be easily identified and the power changes of other appliances (blue) can be easily filtered out.

although the target appliance and the other appliances operate at the same time, their power changes do not occur simultaneously in most of time (see Fig. 2). Thus, theoretically the target appliance can be disaggregated from the differential aggregate data rather than the raw aggregate data. The reason raw data is effective as input in existing neural network based models is because the process of calculating the difference is carried out automatically in the neural network implicitly. It is inaccurate and computational expensive for such a simple and explicit differential operation to be conducted by neural network. Therefore, in this article, we propose to feed the differential of the raw data directly as input to the neural network, which makes it easier for the CNN to extract power change of target appliance. Besides, we also use statistics (mean, std, max, min) of raw aggregate data as the auxiliary inputs of the network to make full use of the raw aggregate data.

Below we analyze the effectiveness of differential operations. Define p_{t-1} and p_t as the power values at time $t-1$ and t, respectively. Then the difference at time t is

$$\Delta P_t = p_t - p_{t-1} = [p_{t-1}, p_t][-1,1] \tag{1}$$

In CNN, above differential operation can be calculated as a vector multiplication, where [-1,1] is the equivalent filter/kernel of CNN and needs to be learned by training. Normally, the equivalent kernel learned is somewhat fuzzy. For example, it may be [- 0.98, 1.02] or [- 1.01,0.97], etc.

Define the learned kernel is $[-1+n_1, 1+n_2]$, where n_1 and n_2 are errors with small value. Then, the difference operated by CNN is

$$\begin{aligned} \Delta\hat{P}_t &= [p_{t-1}, p_t][-1+n_1, 1+n_2] \\ &= p_t - p_{t-1} + n_2(p_t - p_{t-1}) + (n_1+n_2)p_{t-1} \\ &= \Delta P_t + e_t \end{aligned} \tag{2}$$

where $e_t = n_2\Delta P_t + (n_1+n_2)p_{t-1}$ is the difference operation error of CNN. Note that ΔP_t and p_{t-1} can be arbitrarily large and therefore the difference operation error cannot be ignored. This is why the difference operation of the neural network is not accurate.

2.3. Network Architecture

The proposed network structure is shown in Fig. 3. It has two inputs: 1d differential input and 1d auxiliary input consisting of simple statistics of raw aggregated data. The output of the network is the power sequence of the target appliance in the corresponding time. The length of differential input sequence depends on the target appliance. The differential input extracts the transient feature hierarchically through five convolutional layers and then concatenated with the auxiliary input. For each convolutional layer, a MaxPooling layer is followed in order to prevent overfitting. Finally, two fully connected layers (FC layer) are followed in which the last layer is used as the output layer.

2.4. Infer/Disaggregate

In the inference stage, sliding window method is applied. We put a window on the aggregated power sequence and slide on the sequence with stride of 1. The length of the window is the input size of the model for the target appliance. In each stride, a window of differential value of aggregate power sequence is given and the network outputs the corresponding power sequence of the target appliance with same length. Finally, we take an average of the overlapped power values as the predicted power. The process of averaging makes the disaggregated result more robust.

3. Experiments

3.1. Experiment Configurations

Since there is no industrial dataset but only household dataset is available, in this experiment we use the Reference Energy Disaggregation Dataset (REDD) [5] household dataset to verify the effectiveness of the proposed NILM. REDD records the power for 6 houses with sampling frequency of 1Hz for mains meter and 1/3Hz for appliance-channel meters. We use the data in house 2~6 for training and the data in house 1 for testing. In the experiment, refrigerator, dishwasher and microwave are selected to disaggregate. Those appliances are the most common household appliances, whose consumption accounts for most of a house's consumption.

In deep learning network, data normalization usually accelerates the convergence of networks. We have done some preliminary experiments and found that under the strategy of differential input normalization does not show much effect. But for comparison, we experiment our model both with data normalization (*norm*) and without data normalization (*no norm*). For the former one, we normalize the inputs and outputs by dividing with a handpicked maximum consumption value. Specifically, the power demands of fridge, microwave, and dishwasher are divided by 200, 2000, and 1400 W, respectively. Similar to [6, 8], to cover the entire operating cycle of most of appliances, the size of input window is set as 600.

Mean squared error (MSE) and mean absolute error (MAE) are used to evaluate the performance of NILM algorithm. Define y_t and $\hat{y}_t$ as the real power and the predicted power of

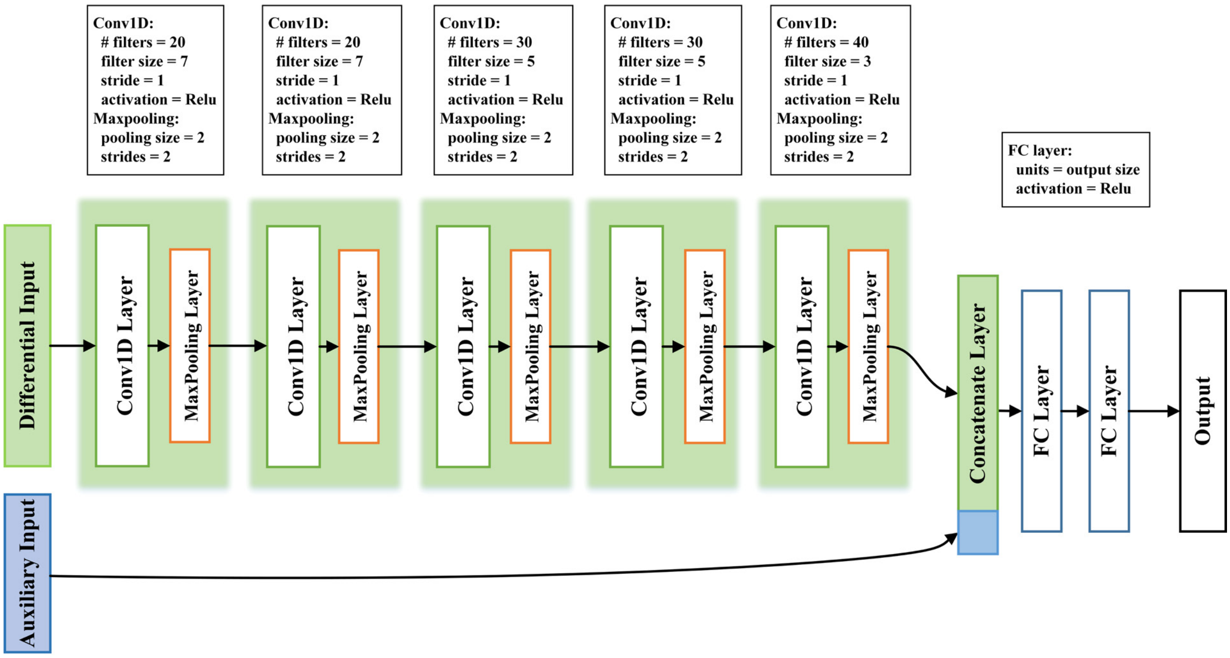

Fig. 3. Proposed CNN-based Network Architecture

Table 1. The MSE and MAE results for REDD dataset. Best results are shown in bold.

Metrics	Models	Fridge	Microwave	Dishwasher
MSE	*seq2point [6]*	2393.9	17483.5	15891.3
	seq2seq [6]	2151.9	19292.8	**14172.6**
	GLU-Res [8]	2197.4	25202.0	22301.1
	Proposed Model (no norm)	**1622.8**	**17037.9**	18658.5
	Proposed Model (norm)	1867.1	19195.6	15489.2
MAE	*seq2point [6]*	26.17	20.51	23.73
	seq2seq [6]	28.15	27.87	24.45
	GLU-Res [8]	23.52	28.41	33.37
	Proposed Model (no norm)	**21.76**	**18.32**	22.32
	Proposed Model (norm)	25.41	20.81	**20.12**

Table 2. Size of the parameters for each model.

Models	*# parameters*
seq2point [6]	29.2M
seq2seq [6]	29.8M
GLU-Res [8]	1.2M
Proposed Model	738K

the appliance at time t. MSE is represented as

$$\mathrm{MSE} = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} |\hat{y}_t - y_t|^2 \tag{3}$$

and MAE is represented as

$$\mathrm{MAE} = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} |\hat{y}_t - y_t| \, . \tag{4}$$

MAE intuitively represents the absolute errors at every moment.

For comparison, we also evaluate Chaoyun's seq2point model [6], seq2seq model [6] as well as Chen's GLU-Res model [8], in which the configurations follow the original papers. All the experiments are conducted on the same dataset. The experimental results are shown in Table 1 and the best results are shown in bold. Besides, the size of the parameters for each model is listed in Table 2.

3.2. Experiment Results

According to Table 1, we find that both the proposed models with differential input outperform the other three models [6, 8] on MSE and MAE for most of the appliances except for dishwasher on MSE. Furthermore, under the strategy of differential input, the proposed model without data normalization performs better than that with data normalization for fridge and microwave except for dishwasher. More importantly, the size of our model (738K) is one or two orders of magnitude less than other models. The results show that by adopting the differential input, a much smaller network achieves better results than existing models. Such advantages of the proposed system benefit the service providers as well as the customers.

4. Conclusion and Future work

In this paper, we propose a NILM-based EMS and an end-to-end NILM model based on convolutional networks using differential input. The EMS can help to achieve the promised advantages of smart grids and can be utilized in smart homes or smart factories. In this framework, service provider can provide customer much more appliance-level services and customized solutions, which will improve the efficiency of energy usage and stability of the production in manufacturing. For NILM, the experimental results demonstrate that the differential input design improves the disaggregation performance of the neural networks.

There are several further works worth taking. One point that requires more attention is the length of sliding window that is used as input. Similar to some other works, we use a default length of 600 samples for all appliances in the experiments. The appropriate sliding window sizes vary from different appliances and should be optimized accordingly in the future work. Secondly, the architecture of the network can be further optimized. A promising way is to introduce attention mechanism to the network so that the network can learn to focus on certain parts of the aggregate power sequence. In addition, current NILM algorithm is single-task and a separate model is needed for each appliance. A multi-task network is promising by disaggregating all the appliances though one network on the cost of more complicated network design.

Acknowledgements

We are very grateful for the support of Laboratory of Communication and Signal Processing, University of Macao and the support of Jinan University with the GPU computing platform for this research.

References

[1] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no. 12, pp.1870-1891, Dec.1992.

[2] G. H. Yang and V. O. K. Li. “Energy management system and pervasive service-oriented networks,” *Proceedings of 1st IEEE International Conference on Smart Grid Communications*, pp.1-6, 2010.

[3] R. Brewer and P. Johnson, “WattDepot: An Open Source Software Ecosystem for Enterprise-Scale Energy Data Collection, Storage, Analysis, and Visualization,” *IEEE Smart Grid Comm.*, 2010.

[4] J. Kelly and W. Knottenbelt, “Neural NILM: Deep Neural Networks Applied to Energy Disaggregation,” *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pp. 55-64, 2015.

[5] J. Z. Kolter and M. J. Johnson, “REDD: A public data set for energy disaggregation research,” *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, vol. 25, no. Citeseer, 2011.

[6] C. Zhang, M. Zhong, Z. Wang, et al., “Sequence-to-point learning with neural networks for non-intrusive load monitoring,” *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[7] K. Odysseas, C. Nalmpantis and D. Vrakas, “Sliding Window Approach for Online Energy Disaggregation Using Artificial Neural Networks,” *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, ACM, pp. 7, 2018.

[8] K. Chen, Q. Wang, Z. He, et al., “Convolutional sequence to sequence non-intrusive load monitoring,” *The Journal of Engineering*, vol. 2018, no. 17, pp. 1860-1864, 2018.

[9] F. Anthony, M. N. Henry, K. Shubi and M. Kisangiri, “A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem,” *arXiv preprint arXiv:1703.00785*, 2017.

[10] L. Mauch and B. Yang, “A new approach for supervised power disaggregation by using a deep recurrent LSTM network,” *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 63-67, 2015.

[11] A. Ridi, C. Gisler and J. Hennebert, “A survey on intrusive load monitoring for appliance recognition,” *22nd International Conference on Pattern Recognition*, IEEE, pp. 3702-3707, 2014.

[12] T. Sirojan, B. T. Phung, E. Ambikairajah, et al., “Deep Neural Network Based Energy Disaggregation,” *IEEE International Conference on Smart Energy Grid Engineering (SEGE)*, pp. 73-77, 2018.

[13] H. Kim, M. Marwah, M. F. Arlitt, et al., “Unsupervised disaggregation of Low Frequency Power Measurements,” *Proceedings of the 11th SIAM International Conference on Data Mining*, pp. 747-758, 2011.

[14] K. Zico, J. Tommi, and J. Z. Kolter, “Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation,” *In Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 1472-1482, 2012.