Title
Nonintrusive load monitoring based on deep learning

ISBN
978-3-030-04303-2

Authors
Wang, K.
Zhong, H.
Yu, N.P.
et al.

Publication Date
2018-11-17

Peer reviewed

Nonintrusive Load Monitoring Based on Deep Learning

Ke Wang[1,2], Haiwang Zhong[1,2](✉), Nanpeng Yu[3], and Qing Xia[1,2]

[1] Department of Electrical Engineering, Tsinghua University, Beijing 100084, China
zhonghw@tsinghua.edu.cn
[2] State Key Laboratory of Control and Simulation of Power Systems and Generation Equipments, Tsinghua University, Beijing 100084, China
[3] University of California, Riverside, CA 92521, USA

Abstract. This paper presents a novel nonintrusive load monitoring method based on deep learning. Unlike the existing work based on convolutional neural network and recurrent neural network with fully connected layers, this paper develops a deep neural network based on sequence-to-sequence model and attention mechanism to perform nonintrusive load monitoring. The overall framework can be divided into three layers. In the first layer, the input active power time sequence is embedded into a group of high dimensional vectors. In the second layer, the vectors are encoded by a bi-directional LSTM layer, and the N encoded vectors are added up to form a dynamic context vector according to its weights calculated by the attention mechanism. In the third layer, an LSTM-based decoder utilizes the dynamic context vector to calculate the disaggregated power consumption at every time step. The proposed method is trained and tested on REFITPowerData dataset. The results show that compared to the state-of-the-art methods, the proposed method significantly increases the accuracy of the estimation for the disaggregated power value and decreases the misjudge rate by 10% to 20%.

Keywords: Nonintrusive load monitoring · Deep learning
Sequence-to-sequence model · Attention mechanism

1 Introduction

To increase energy efficiency [1] and enhance demand response [2] capabilities of end-use customers, it is crucial to inform household users of real-time electricity consumption of individual appliances in the buildings. Traditional load monitoring methods require a separate sensor for each individual appliance, which results in high implementation cost and low user acceptance. Nonintrusive load monitoring (NILM) is an emerging technology that estimates the electricity usage of individual loads from a single-point measurement of the combined power consumption.

In the literature, most of the research can be grouped into two categories [3]. The first is transient-state-based NILM approach and the second is steady-state-based NILM approach.

W. L. Woon et al. (Eds.): DARE 2018, LNAI 11325, pp. 137–147, 2018.
https://doi.org/10.1007/978-3-030-04303-2_11

Transient-state-based NILM approach utilizes high-frequency electricity features such as voltage and current waveform and its harmonics. Cox [4] conducts frequency spectrum analysis on transient voltage waveform, which reaches high accuracy identifying ON/OFF motion of a single appliance. Tsai [5] applies KNN (K-Nearest Neighbor) and BP-ANN (Back Propagate-Artificial Neural Network) on transient current waveform. Yun [6] carries out similarity matching between standardized template and extracted active and reactive power variation. These methods are based on high frequency sampling and have high requirements for metering devices and data storage systems. Thus, they are not appropriate for household load monitoring.

Steady-state-based NILM approach often utilizes active and reactive power consumption sampled at low frequency. Kolter [7] adopts FHMM (Factorial Hidden Markov Model), and is able to handle situations when multiple appliances are operating simultaneously. Kelly [8] grasps the upsurge of artificial intelligence, and proposed a deep learning model based on CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network). These methods are compatible with the existing smart meter infrastructure. However, as a result of low-frequency sampling, the existing approaches yield high misjudge rate and low accuracy of disaggregated power consumption estimation.

In this paper, we adopt the deep learning framework, and propose a NILM model based on sequence-to-sequence model and attention mechanism. The proposed model is trained and evaluated on the REFITPowerData [9] dataset. The testing results show that the proposed model has great potential in reducing misjudge rate and increasing accuracy of disaggregated power consumption.

The rest of the paper is organized as follows. In Sect. 2, the proposed model and NILM system are introduced. The data source and processing method are presented in Sect. 3. Experiment studies are reported in Sect. 4. Section 5 concludes the paper.

2 Nonintrusive Load Monitoring Model

2.1 Physical Model of Nonintrusive Load Monitoring

Suppose there are M electrical appliances, and the power consumption time series of the i^{th} electrical appliance is:

$$X_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,T})\, x_{i,t} \in R_+$$

Considering potential random noise in the smart meter measurements, the model can be represented as:

$$y_t = \sum_{i=1}^{M} x_{i,t} + \mu_t \, t = 1, 2, \ldots, T$$

where μ_t is Gaussian noise with zero mean and variance of σ^2.

The aim of nonintrusive load monitoring is to recover the power consumption of each electrical appliance from the aggregated power consumption data. Suppose Y is the aggregated power consumption with T sampling point, and:

$$Y = (y_1, y_2, \ldots, y_T)\, y_t \in R_+$$

2.2 Motivation for Adopting Deep Learning Framework

Deep learning is based on learning data representations, and has made great achievements in computer vision [10], speech recognition and natural language processing [11]. The universal approximation theorem proves that a feed-forward neural network containing finite number of neurons can approximate any continuous functions on compact subsets R^n. This characteristic makes deep learning a great candidate in approximating any real physical model. Hence, it should be suitable for the application of nonintrusive load monitoring.

2.3 Nonintrusive Load Monitoring Model

The overall framework of the proposed nonintrusive load monitoring model is shown in Fig. 1.

The proposed model only utilizes low sampling-rate aggregated active power consumption data, and the data is discretized to integers for simplicity. A deep learning framework with different parameter settings are trained to estimate the load consumption of M different target appliances.

2.3.1 Data Segmentation

Before the disaggregation process starts, data segmentation is conducted to split the long-range input data into pieces of preset length. This preset length is different from each other because different target appliance has different length of run time. For example, microwave's run time is typically less than ten minutes while a washing machine can operate for more than two hours.

2.3.2 Embedding

After data segmentation, an embedding process is used to map the integer value of aggregated power consumption to a high dimensional vector with an embedding matrix E:

$$E = [voc\_size, embedding\_size]$$

For each aggregated power consumption value i, it is mapped to vector $E[i]$, and after the data segmentation process, the $[N_i * 1]$ input sequence is transformed into $[N_i * embedding\_size]$ matrix Z.

2.3.3 Sequence-to-Sequence Model

Sequence-to-sequence model [11] is the encode-decode part in Fig. 1 which converts a sequence to another sequence, and both encoder and decoder are based on LSTM [12]

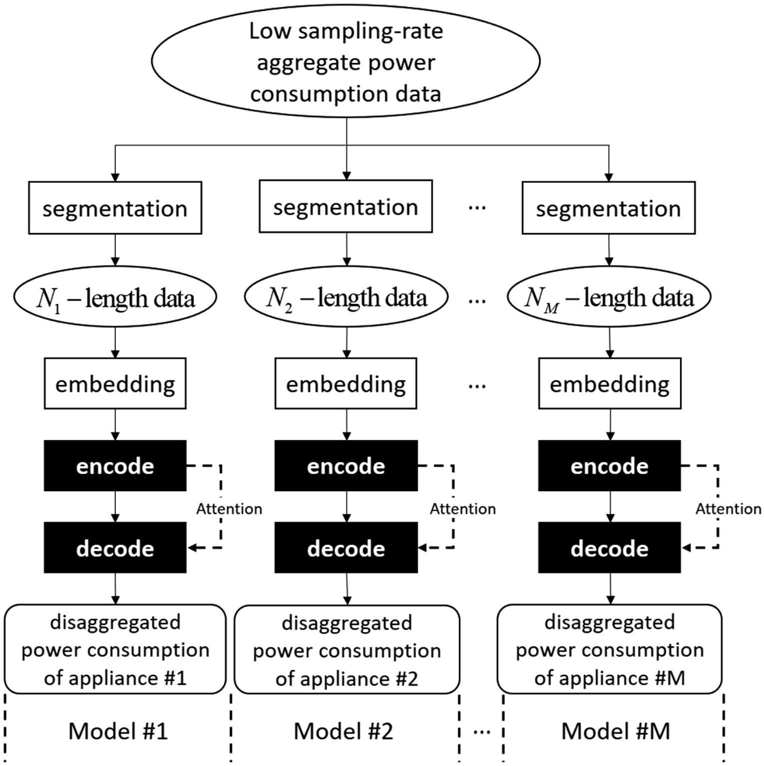

Fig. 1. Flowchart of the proposed nonintrusive load monitoring model

(Long short term memory). The deep learning architecture of sequence-to-sequence model is shown in Fig. 2.

At each time step t, the encoder calculates h_t, the hidden state of time t, from h_{t-1} and Z_t, the embedded input of time t.

$$h_t = f(Z_t, h_{t-1}) t = 1, 2, \ldots, N_i$$

where f is the inner computation rule of LSTM. After the encoding process is finished, we can get N_i hidden states, and the last hidden state h_T is assigned to context vector C.

In the decoding process, at each time step t, the decoder calculates s_t (to distinguish between encoder's hidden state h_t), the hidden state of time t, from s_{t-1}, Y_{t-1}, the output of time $t-1$, and context vector C.

$$s_t = f([Y_{t-1}, C], s_{t-1}) t = 1, 2, \ldots, N_i$$

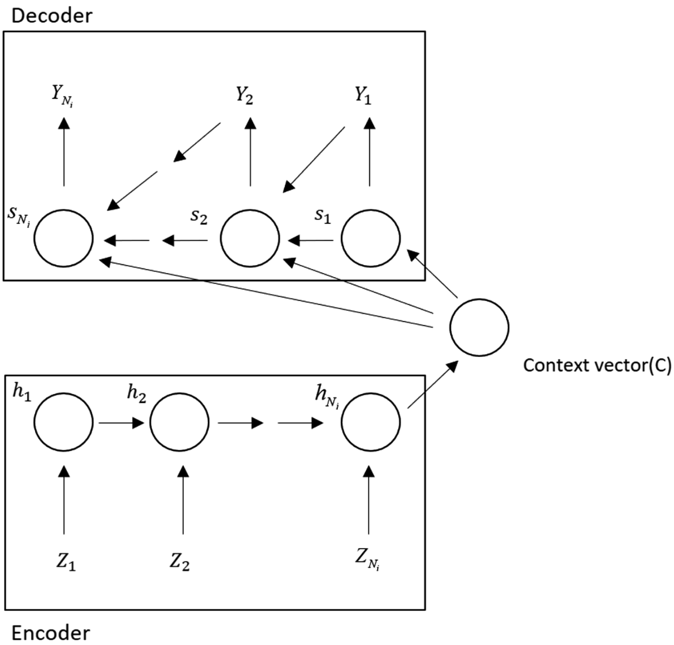

Fig. 2. Schematic of LSTM-based sequence-to-sequence model

Y_t is calculated from s_t:

$$Y_t = g(s_t)t = 1, 2, \ldots, T$$

where g is a projection layer.

2.3.4 Attention Mechanism

One drawback of the sequence-to-sequence model is that the context vector C is fixed during the decoding process. It means that all input information is restricted to a limited length, which might result in data leakage. The attention mechanism can address the problem.

The key component of attention mechanism [13] is to calculate $a_{t,i}$, the weight of the i^{th} hidden state of encoder at time t in decoding, and add the hidden states up according to their own weight to form a dynamic context vector C_t. In this way, the model can pay attention to data that is closely related to decoding at different times. The attention mechanism thus solves the data leakage problem and improves model's ability to extract useful information.

The attention mechanism works as follows:

$$e_{t,j} = V^T * \tanh(Ws_{t-1} + Uh_j) \quad t, j = 1, 2, \ldots, N_i$$

$$a_{t,j} = \frac{\exp(e_{t,j})}{\sum_{k=1}^{N_i} \exp(e_{t,k})} t,j = 1,2,\ldots,N_i$$

$$C_t = \sum_{j=1}^{N_i} a_{t,j} h_j t = 1,2,\ldots,N_i$$

$$s_t = f([Y_{t-1}, C_t], s_{t-1}) t = 1,2,\ldots,N_i$$

where V, W and U are trainable parameters that will be updated during the training process, and f is the inner computation rule of LSTM.

3 Dataset and Data Processing

3.1 Dataset

We adopt the REFITPowerData [11] dataset released in 2015 for model training and testing. The dataset includes aggregated power consumption data and single appliance power consumption data sampled every 8 s. The dataset is gathered from October, 2013 to May, 2015.

3.2 Selection of Appliances

In the case study, the fridge, TV, microwave, washing machine and dish washer are selected as the five target appliances. This is because they have different working patterns and make up majority of household electricity consumption. For example, fridge simply has on/off mode while washing machine has multiple operation modes with different power consumptions.

3.3 Data Processing

In order to synthetize the training data, we first need to extract load activation [8], which is the working period power consumption data of each target appliance. The parameter setting of the data extraction is shown in Table 1. The extracted load activations are stored properly.

After extracting load activations, the training data are synthetized in three steps.

First, create an all-zero sequence of length N_i, which is shown in Table 1. Then put one load activation of the target appliance into the sequence entirely with 50% probability. The remaining sequence is unchanged with 50% probability.

Second, for appliances except for the target appliance, put one load activation of each into the sequence with 25% probability, and this does not require the load activation to be put into the sequence entirely.

Third, repeat step one and two for K times, and make a training data that includes K pieces of N_i length sequences.

Table 1. Parameters for extracting load activations and sequence size

Load	Max Power (W)	Min Power (W)	Shortest operation time (point)	Longest time below threshold (point)	Sequence length (point)
Fridge	300	20	100	10	400
TV	300	20	100	10	2000
Microwave	NA	20	5	5	50
Washing machine	NA	20	100	10	1200
Dish washer	NA	20	100	10	1000

4 Experimental Results

In order to show the performance of the proposed model, a case with small data size and a case with large data size are both studied.

4.1 Case with Small Data Size

The case with small data size, which is synthetized randomly from load activations that have been extracted, is used to demonstrate the accuracy of disaggregated power consumption directly, and the results are shown in Fig. 3. It can be observed from Fig. 3 that the proposed model can reach a high level of accuracy in estimating the disaggregated power consumption. The model performs exceptionally well for appliances that have predictable work mode and high power consumption such as microwave, washing machine and dish washer. For appliances with low power consumption during working period such as TV and fridge, the predictions of start and end point are slightly less accurate. However, the accuracy of disaggregated power consumption forecast for appliance with lower power consumption is still at a high level.

4.2 Case with Large Data Size

To fully evaluate the performance of the proposed model, four metrics, accuracy, recall, F1-score and mean absolute error are used. These metrics can be calculated as follows.

$$PRE = \frac{TP}{TP + FP}$$

$$REC = \frac{TP}{TP + FN}$$

$$F1 = 2 * \frac{PRE * REC}{PRE + REC}$$

$$MAE = \frac{1}{T_1 - T_0} \sum_{t=T_0}^{T_1} abs(\tilde{y}_t - y_t)$$

PRE is accuracy, REC is recall and F1 is F1 score. TP is the number of true positives, FP is the number of false positives and FN is the number of false negatives. y_t is the actual power consumption of target appliance at time t, and $\tilde{y}_t$ is the disaggregated power consumption of target appliance calculated by the proposed model. MAE is the mean absolute error, which indicates the accuracy of disaggregated power consumption.

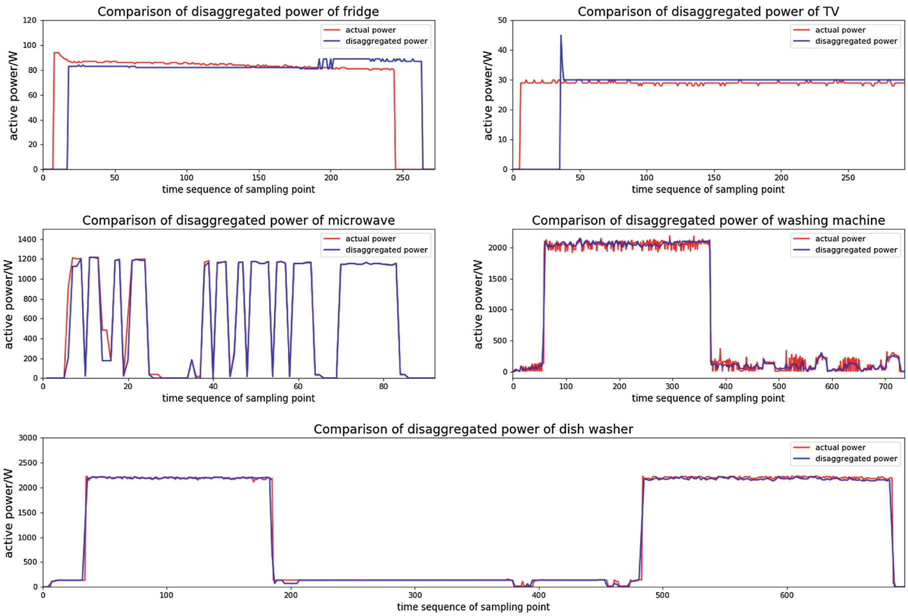

Fig. 3. Test result of case with small data size

To evaluate the performance and capability of generalization of the proposed model, the evaluation of this test case is split into two parts. The first is on houses selected in the training data set, and the second is on houses not included during training. In addition, Kelly's [8] deep learning model is replicated to serve as a benchmark.

4.2.1 Tests on Houses Involved in Training

The test results of houses involved in the training process are shown in Fig. 4.

Tests on houses involved in the training process do not mean the training and testing process is conducted on the same data. In the field of deep learning, the training dataset and testing dataset must be separated because the model may gain some memory of training dataset during the training process. In this work, the training process is on the first one million sampling points and the test process is on the last one

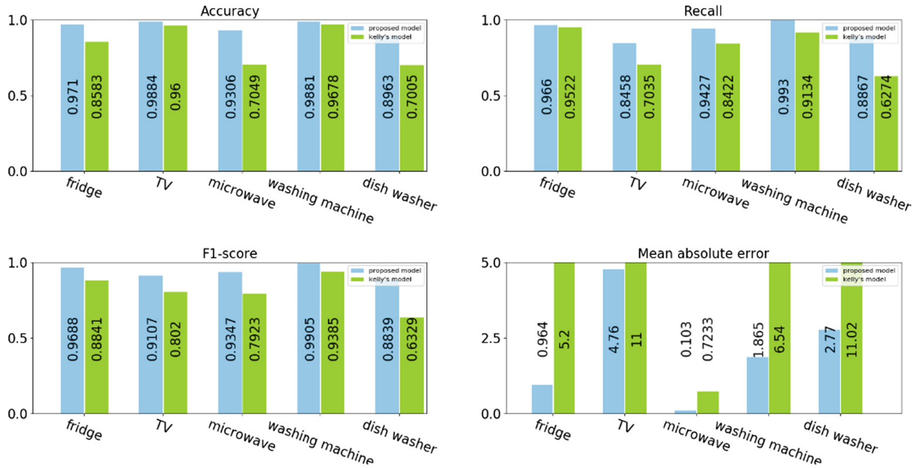

Fig. 4. Test result of large data size on houses involved in training

hundred thousand sampling points.

It can be observed from Fig. 4 that for all appliances, the proposed model can achieve good performance of power disaggregation. The MAE of the proposed method is much smaller than that of the Kelly's [8] model. This demonstrates that the proposed model has strong capability in utilizing data and identifying the working pattern of electrical appliances. Besides, the proposed model also increase accuracy, recall and F1-score by 10 to 20%.

4.2.2 Tests on Houses not Involved During Training

The test results of houses not involved during training are shown in Fig. 5.

Tests on houses not involved during training is of great significance because this is the only way to examine whether the model has learned the right pattern and whether the model has strong capability of generalization. Theoretically, for deep learning models, the more data they are trained with, the better the generalization ability is.

It can be observed from Fig. 5 that the performance of proposed model is slightly worse for houses not involved in training. However, the results are still decent, which demonstrates that the proposed model has strong capability of generalization. Compared to Kelly's [8] model, our proposed model improves accuracy, recall and F1-score by 10 to 20% and reduces the mean absolute error sharply.

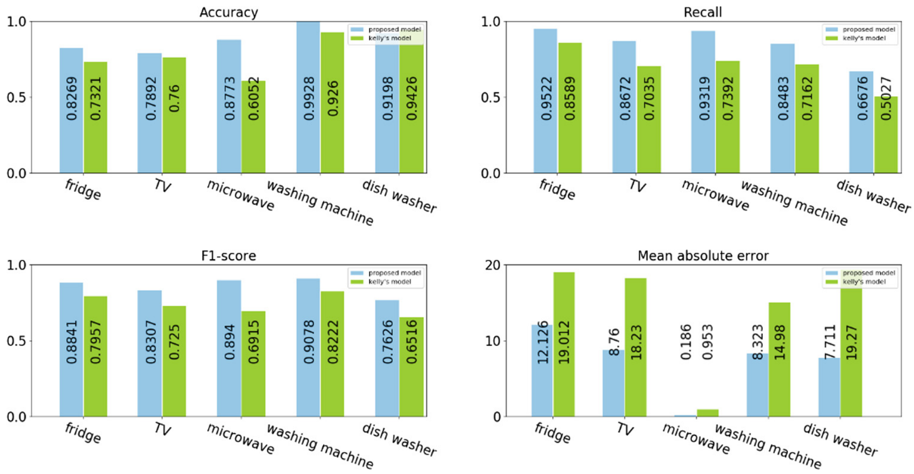

Fig. 5. Test result of large data size on houses not involved in training

5 Conclusion

We develop a deep learning framework based on sequence-to-sequence model and attention mechanism to perform nonintrusive load monitoring. The proposed model introduces the Encoder-Decoder architecture, and uses attention mechanism to extract the most relevant hidden states of encoder to guide the decoding process. These unique features enhance the proposed model's ability to extract and utilize information dramatically. Tests on houses involved in training and houses not involved in training demonstrate that the proposed model can increase accuracy, recall and F1-score by 10 to 20% and reduce mean absolute error dramatically compared to the deep learning models based on CNN and RNN. In the future, we plan to improve the model's capability of generalization and the model's applicability on low power consumption appliances.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China (51777102), in part by the Beijing Natural Science Foundation (3182017), and in part by the State Grid Corporation of China (5210EF18000G).

References

1. Zhao, J., Dong, C., Wen, F., et al. Data science for energy systems: theory, techniques and prospect. In: Proceedings of the CSEE 2017, vol. 41(04), pp. 1–11 (2017)
2. Zhang, D., Yao, L., Ma, W.: Development strategies of smart grid in China and Abroad. In: Proceedings of the CSEE, 2013, vol. 33(31), 1–15 (2013)
3. Fuqiu, Z., Juan, W.: Situations and suggestions of national DSM work. Power Demand Side Manag. **17**(2), 1–4 (2015)

4. Cox, R., Leeb, S.B., Shaw, S.R., et al.: Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion. In: Applied Power Electronics Conference and Exposition, New York, USA (2006)
5. Tsai, M., Lin, Y.: Modern development of an adaptive non-intrusive appliance load monitoring system in electricity energy conservation. Appl. Energy **96**, 55–73 (2012)
6. Yun, G., Honggeng, Y.: Household load measurement by classification based on Minkowski Distance. Electr. Meas. Instrum. **50**(569), 86–90 (2013)
7. Kolter, J.Z., Jaakkola, T.: Approximate inference in additive factorial HMMs with application to energy disaggregation. In: Artificial Intelligence and Statistics, La Palma, The Republic of Panama (2012)
8. 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. Seoul, South Korea (2015)
9. Murray, D., Stankovic, L., Stankovic, V.: An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. Sci. Data **4**, 160122 (2017)
10. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, USA (2016)
11. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. Adv. Neural Inf. Process. Syst. **27**, 3104–3112 (2014)
12. Hochreiter, S., Schrnidhuber, J.: Long short-term memory. Neural Compet. **9**(8), 1735–1780 (1997)
13. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: Proceedings of the 3rd International Conference on Learning Representations, Leuven, Belgium (2015)