A Non-intrusive Load monitoring Algorithm Based on Seq2point

Zhen Zhang¹, Junqiang Ma¹*, Menghou Li¹, Haofang Li¹
¹Zhengzhou University School of Electrical Engineering, Zhengzhou 450001, China
*Corresponding author's e-mail:junqiang@gs.zzu.edu.cn

Abstract. Aiming at the problems of long training time and poor decomposition performance of neural network in the application of Sequence-to-point deep learning method in the field of non-intrusive load monitoring, a non-intrusive load monitoring optimization model based on Seq2point was proposed. In this model, the training speed of the neural network is optimized by the forward gated recurrent unite, and the decomposition performance is improved by the median filter and the standardized data pre-treatment method. Experimental results on public data set AMPDS2 show that this algorithm can effectively improve the accuracy of load decomposition and shorten the training time of network.

1. Introduction
In the process of achieving the goal of "peak carbon dioxide emissions" and "carbon neutrality", the energy conservation and emission reduction of the power industry plays a very important role. Studies have shown that monitoring household electrical equipment and giving feedback on specific energy consumption information can help residents improve their power consumption habits and save up to 20% of electricity costs [1]. The traditional load monitoring method needs to install sensors on all electrical equipment to obtain energy consumption information, so it has the problems of high implementation cost and low user acceptance. However, non-intrusive load monitoring (NILM) does not depend on the monitoring sensor of each device. It only needs to collect the total load data of power users, and the sub-load data can be decomposed from the total load data with the help of relevant technologies. Therefore, the research of non-intrusive load monitoring technology has very important significance.

With the rapid development of artificial intelligence, data science and other frontier fields, Jack Kelly [2] proposed in 2014 to build a non-intrusive load monitoring model based on sequence to sequence (Seq2seq) by using convolutional neural network and recurrent neural network. The decomposition effect is better than the optimal combination model and the factor hidden Markov model. On this basis, Literature [3] proposed a load decomposition model based on Seq2seq and the attention mechanism, which improved the accuracy of power decomposition value by increasing the attention mechanism. The literature [4] and [5] proposed the convolutional neural network and recurrent convolutional neural network model based on Sequence to point (Seq2point), which reduced the mean absolute error but increased the training time of network.

In order to further improve the decomposition performance of the non-intrusive load monitoring model based on Seq2point and reduce the training time of the neural network, this paper takes refrigerators, washing machines, dishwashers and lights as research objects and proposes a non-intrusive load decomposition optimization model based on Seq2point. And compared with literature [4] and literature [5] on AMPDS2 dataset, the superiority of the algorithm proposed in this paper will be verified.
2. Non-intrusive monitoring model based on Seq2point

2.1. Non-intrusive load monitoring
Non-intrusive load monitoring is a technique to recover source appliances from only the recorded mains in a household. The schematic is shown in Figure 1. It can be seen that the essence of non-intrusive load monitoring problem is load decomposition.

Firstly, we summarize the problem of non-intrusive load monitoring. Suppose that the mains \( Y(t) \) are the aggregation of all the active power consumption of the individual appliances in a household. Note that \( Y(t) \) denotes the mains reading at time \( t \). Then the mains could be represented as the following formula,

\[
Y(t) = \sum_{i}^{N} X_i(t) + \epsilon(t)
\]

where \( X_i(t) \) represents the power reading of the appliance \( i \) at time \( t \). \( N \) is the number of electrical appliances and \( \epsilon(t) \) is the variable representing the model noise which can be ignored.

2.2. Mathematical model
Recently, many household electricity data have been published. These data have both the mains and correspondingly the appliance power readings, which makes it possible to formulate NILM as a supervised learning. Precisely, we have observed the pairs of the data \((X_i,Y_i)\) where \( X_i \) and \( Y_i \) denote respectively the power reading of an appliance and the mains at time \( t \). Because there are plenty of observations, it is possible to train a model to represent the relationship between \( X \) and \( Y \), which can be viewed as a non-linear regression problem\[^6\]. Now, both sequence-to-sequence and sequence-to-point learning methods were proposed to deal with NILM. Suppose both the seq2point and seq2seq learning have the same architecture, the former works better\[^4\]. In this paper, the seq2point learning method was adopted. The seq2point architectures define a neural network \( F \) which maps sliding windows \( Y_{t-w+1}^{t} \) of the input to the last point \( x_t \) of the corresponding windows \( X_{t+w-1}^{t} \) of the output. The model is \( x_t = F(Y_{t-w+1}^{t}), t = 1, 2, \ldots, n+1-W, W \) is the sliding windows of length.

2.3. Process flow
The process of the non-intrusive load monitoring algorithm based on Seq2point is shown in Figure 2.
1) The mains of active power are collected from the power bus and processed by median filter;
2) Standardize the mains of active power data and convert it into the input format required by the neural network;
3) The processed active power data of the mains are sent to each electrical model for decomposition in turn to obtain the standardized decomposition value of each target electrical appliance;
4) Obtain the real active power data of each target electrical appliance through anti-standardization operation.

2.4. Network architecture
The network architecture is optimized on the basis of literature [5] and mainly consists of forward GRU layer and full connection layer. The optimization part mainly consists of the following two points: First, the operation of the convolutional layer is removed, and the GRU layer, which is more suitable for processing time series data, is directly used to remember the most original features. Second, changing the Bi-directional GRU layer into the forward GRU layer can not only reduce the training time of the network, but also improve the accuracy of model decomposition. The specific network architecture is shown in Figure 3.

1) Input layer: the tensor of the input layer is (32, 1), which means 32 sampling points are input to the neural network each time.
2) Forward GRU layer: the main function is to mine the change rules inside the input layer data.
The input data tensor of the first forward GRU layer is \((32,1)\), the output data tensor is \((32,64)\), and the size of the hidden state is 64. The second forward GRU layer input data tensor is \((32,64)\), output data tensor is \((1,128)\), and the size of the hidden state is 128.

3) Full connection layer: the main function is to map the high dimension to the low dimension, and extract and integrate the effective feature information. The number of neurons in the full connection layer is 64. The input data dimension of this layer is \((1,128)\), and the output dimension is \((1,64)\). The rectified linear unit is used as the activation function.

4) Output layer: Corresponding to the standardized power value of the target electrical appliance. The input data dimension of this layer is \((1,64)\), and the output dimension is \((1,1)\). Linear is adopted as the activation function.

2.5. Loss and Optimization Function
When training the model, the deviation of the current state of the model must be repeatedly estimated in order to update the weight in the next evaluation to reduce the loss. In this paper, we use the Mean Square Error (MSE) as the loss function.

After the loss function value of the model is obtained through forward calculation, the reverse optimization algorithm is needed to update the model parameters and finally make the model converge. Literature \(^7\) shows that the Adaptive Moment Estimation (Adam) algorithm has better performance than other optimization algorithms, so the Adam optimization algorithm was adopted.

3. Experiments and result

3.1. Dataset and Pre-processing
Since data is the basis of load decomposition, and the reliability and accuracy of data have a great impact on the accuracy of the results. we trained all of the models using AMPDS2\(^1\) dataset with a sampling period of 1 minute. This dataset has the advantages of long monitoring cycle, low sampling frequency and rich monitoring data, which is very suitable for the research of non-intrusive load decomposition based on deep learning method.

Fridge, dishwasher, washing machine and light were chosen as research subjects. The training set is the data from April 1, 2012 to March 31, 2013 in the data set, which basically covers the various operating states of electrical appliances and the connections between each electrical appliance. The testing set consist of data from April 1, 2013 to May 30, 2013.

In order to enable the model to read the characteristics of the signal quickly and accurately, the original data was firstly processed by median filtering. Then, for all experiments, both the input windows and targets were processed by subtracting the mean values and dividing by the standard deviations.

3.2. Metrics
Three metrics were employed in this paper to evaluate the algorithms. the first metric is the mean absolute error (MAE), which evaluates the absolute difference between the prediction \(\hat{y}_t\) and the ground truth \(y_t\) at every time point and calculates the mean value.

\[
\text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |\hat{y}_t - y_t|
\]

The second metrics is the normalised signal aggregate error (SAE), which indicates the relative error of the total energy. Denote \(e\) as the total energy consumption of the appliance and \(\hat{e}\) the predicted total energy, then SAE is defined as

\[
\text{SAE} = \frac{|\hat{e} - e|}{e}
\]

The third metric is a batch size of training time, which used as an index to evaluate the training speed of the network.
3.3. Visualization of results
In order to show the decomposition effect of the algorithm in this paper more intuitively, the prediction and the ground truth of four electrical appliances in a certain period of time are visualized, and shown in Figure 4.

![Fig4 Visualization of the decomposition effect of each electrical appliance](image)

As can be seen from Figure 4, the ground truth curves and prediction curves of fridge and dishwasher basically coincide. There is little difference between the ground truth curve and the prediction curve of the washing machine. Accurate prediction can be achieved in both the working interval and the non-working interval. Electric lights are the worst at decomposing, occasionally by mistake.

3.4. Algorithm Comparison and Analysis
In order to show the superiority of the algorithm presented in this paper, two algorithms [6] and [7] are reproduced on the same data set. The results are shown in Table 1.

| Metrics        | MAE   | SAE   | Time  |
|----------------|-------|-------|-------|
| Fridge         | 5.39W | 0.027 | 169ms |
| CNN[4]         |       |       |       |
| CNN+BIGRU[5]  | 3.58W | 0.006 | 401ms |
| GRU            | 2.14W | 0.009 | 135ms |
| CNN            | 4.69W | 0.270 | 172ms |
| Dishwasher     | 1.13W | 0.040 | 405ms |
| CNN+BIGRU[5]  | 0.72W | 0.013 | 136ms |
| GRU            |       |       |       |
| CNN            | 2.46W | 0.340 | 171ms |
| Washing machine| 1.35W | 0.100 | 408ms |
| CNN+BIGRU[5]  | 1.03W | 0.070 | 141ms |
| GRU            |       |       |       |
| CNN            | 0.42W | 0.560 | 175ms |
| Washlight      | 0.54W | 1.000 | 411ms |
| CNN+BIGRU[5]  | 0.34W | 0.390 | 143ms |
| GRU            |       |       |       |

Taking the washing machine as an example, the MAE of CNN, CNN+GRU and GRU is 2.46W, 1.35W and 1.03W respectively, indicating that the mean absolute error is not high as a whole, and the
improved model is smaller. It is found that the mean absolute error of the other three electrical appliances before improvement is less than 6W, and the mean absolute error of the improved model is less than 3W. It can be seen that the anti-interference ability of the improved model has been improved.

Analysing the model from the perspective of relative error of total energy consumption, it can be found that the models with the smallest relative error of total energy consumption of refrigerator, dishwasher and washing machine are CNN+BIGRU, GRU and GRU, which are 0.006, 0.013 and 0.070, respectively. For two of the three appliances, the model with the smallest relative error in total energy consumption is the GRU. As can be seen from Table 1, when the GRU model is used to decompose the refrigerator, the relative error of the total energy consumption is 0.009, and the difference is only 0.003 from CNN+BIGRU. It can be seen that the accuracy of the improved model in the estimation of the total energy consumption has been further improved.

From the perspective of the training efficiency of the model, the training time of CNN, CNN+BIGRU and GRU is about 170ms, 405ms and 137ms, respectively. It can be seen that the training time of the improved GRU model is significantly shortened compared with the other two models. In particular, the training time increased by 66% compared with CNN+BIGRU.

For electric light, however, none of the three methods worked so well. By comparison, the algorithm in this paper is the best. This may be because low-power appliances are vulnerable to interference from other appliances.

4. Conclusion

Compared with other mainstream algorithms, it is found that non-intrusive load monitoring algorithm based on Seq2point designed in this paper can effectively memorize the timing sequence information of the total load by adopting forward GRU network, and shorten the network training time by 66%. Standardized data pre-processing can enhance the performance of training data and enhance the effectiveness of model training. This provides a good model foundation for the further study of non-intrusive load monitoring.

Acknowledgments

This work was supported by Baofeng County Digital Government "Fourteenth Five-Year Plan" (20200108A).

References
[1] Aydin E, Brounen D, Kok N. Information provision and energy consumption: Evidence from a field experiment[J]. Energy Economics, 2018,71:403-410.
[2] Kelly J,Knottenbelt W. Neural NILM: deep neural networks applied to energy disaggregation[C]. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments. Seoul, South Korea:ACM,2015:55-64.
[3] Wang Ke, Zhong Haiwang, Yu Nanpeng, et al. Nonintrusive load monitoring based on sequence-to-sequence model with attention Mechanism[J]. Proceedings of the CSEE,2019,39(1): 75-83.
[4] Zhang C Y, Zhong M J,Wang Z Z, et al.Sequence-to-point learning with neural networks for nonintrusive load monitoring[C]. Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18).USA:AAAI,2018:2604-2611.
[5] Krystalakos O, Nalmpantis C, Vrakas D. Sliding window approach for online energy disaggregation using artificial neural networks[C]. Proceedings of the 10th Hellenic Conference on Artificial Intelligence. ACM, 2018: 1-6.
[6] D’Incecco M, Squartini S, Zhong M. Transfer Learning for Non-Intrusive Load Monitoring[J]. IEEE Transactions on Smart Grid, 2020,11(2):1419-1429.
[7] XU Xiaohui, ZHAO Shutao, CUI Kebin. Non-intrusive Load Disaggregate Algorithm Based on Convolutional Block Attention Module [J/OL]. Power System Technology:1-8
[8] Makonin S , Popowich F,Bartram L,et al. AMPds:a public dataset for load disaggregation and eco-feedback research[C].Electrical Power & Energy Conference.IEEE,2014:1-6.