Non-Intrusive Load Disaggregation Based on Residual Gated Network

Hui Cao\textsuperscript{a}, Liguo Weng\textsuperscript{b}, Min Xia* and Dezheng Zhang\textsuperscript{c}

School of Automation, Nanjing University of Information Science & Technology, Nanjing, China

*Corresponding author e-mail: xiamin@nuist.edu.cn, \textsuperscript{a}2792774751@qq.com, \textsuperscript{b}37583087@qq.com, \textsuperscript{c}980292383@qq.com

Abstract. Non-intrusive load disaggregation is designed to estimate the power consumption of each appliance based on the total power of the appliance in the household. Conventional machine learning algorithms cannot accurately extract semantic information from time series data, which motivates us to implement non-intrusive load disaggregation using residual gated recurrent neural networks model (Res-GRU). First, the networks model use multi-scale convolution kernels networks model extract time series data features, and will get multiple map fusions. Secondly, the networks model use residual learning to deepen the network to extract deep load features. Finally, the networks model use the gated recurrent unit to reset and update high level features. In this way, we can get the output power value of the target appliance. The experimental results show that the proposed network model has a good disaggregation effect.

1. Introduction

Concept of non-invasive load monitoring (Non-Intrusive Load Monitoring, NILM) was first proposed by Hart [1]. The essence of Non-Intrusive Load Disaggregation (NILD) is energy disaggregation, which estimates the power consumption of each appliance based on the aggregated power, so as to better guide users to make reasonable electricity consumption and reduce the cost of electricity for users [2]. Lin [3] and Chung [4] used particle swarm optimization to perform non-intrusive load disaggregation experiments on a small number of appliances. The algorithm can simultaneously decompose the total power data to each electrical device, but the disaggregation is obtained. The resulting error is still large. For the problem of large disaggregation error, Pipa [5] proposed a sparse optimization algorithm for non-intrusive charge load disaggregation algorithm, which reduces the disaggregation error to some extent. Other algorithms such as Adaboost algorithm [6], K- nearest neighbor algorithm [7], support vector machine algorithm [8], fuzzy algorithm [9, 10, and 11], neural network algorithm and other algorithms in non-intrusive charge loading some achievements have been made in the disaggregation task. As we know, disaggregate a variety of electrical appliances with varying states is an important issue, in this work the residual gated recurrent neural networks (Res-GRU) is used for non-intrusive disaggregation.
2. Residual gated recurrent neural network model

2.1. Residual network structure
In 2015, He [12] proposed Residual Networks (ResNet), which solved the model degradation and vanishing gradient of the model after the depth of the network became deeper, and finally improved the accuracy of the model. Therefore, the deep residual network has been widely used. The residual network structure adopts the idea of cross-layer connection. In the residual block, the feature maps of different layers are subjected to feature map fusion processing. Suppose that in a single residual block, the output of the residual block is \( H(x) \). An arbitrary mapping can be fitted by a base map of multiple hidden layer stacks. So we can fit with stacked layers \( H(x)-x \). And plus \( x \), you get \( H(x) \). We put \( H(x) - x \) defined as \( F(x) \). The residual block is divided into two layers, \( x \) expressed as the input to the residual block. \( W \) representing the weight vector, the expression of the forward neural network in the residual block is as follows, where \( \sigma \) represents a nonlinear activation function:

\[
F(x) = W_2 \sigma (W_1 x) \tag{1}
\]

Then, through the residual connection, the output of the residual block is obtained.

\[
y = F(x, \{W_i\}) + x \tag{2}
\]

Among them, \( F(x, \{W_i\}) \) represents the residual mapping function that needs to be learned, \( W_i \) represents the hidden layer weight matrix.

2.2. Gated recurrent unit
The internal structure of the Gated Recurrent Unit (GRU) is simple. There are only two doors inside the GRU, namely the Forgotten Gate and the Reset Gate. In the structure of the GRU, the first step is to determine how much of the state information at the previous moment is brought into the current state. This step is achieved by updating the door.

The function of the update gate is to control the extent to which the status information at the previous moment is brought into the current state. The update gate value ranges from \((0, 1)\). The larger the value of the update gate, the more state information is brought in at the previous moment.

In the structure of the GRU, after updating the role of the gate, it is also necessary to ignore the state information of the current time. This step is achieved by resetting the door. The reset gate value ranges from \((0, 1)\). The smaller the value of the reset gate, the more the status information at the previous moment is ignored.

2.3. Residual gated recurrent unit neural network model
In this paper, non-intrusive load disaggregation is implemented by residual gated recurrent unit neural network (Res-GRU). The source sequence first obtains the multi-scale feature map through multi-scale convolution processing, and then obtains the residual feature map through the residual block processing. Finally, the target sequence is obtained through the processing of the update gate and the reset gate in the gated recurrent unit. We use three different convolution kernels to process source sequence information. The size of these three different convolution kernels are 1x1, 3x1 and 5x1, respectively. Multi-scale convolution kernels can capture feature information for different local ranges. The feature maps obtained by multi-scale convolution kernel processing are fused to obtain a wider range of feature information. The number of convolution kernels with different scales is small. This kind of processing is beneficial to the smooth extraction of features and the explosive growth of convolution structure training parameters after multi-scale convolution. So we use the multi-scale convolution method to broaden the network structure, which also enhances the generalization ability of the network model.
After the multi-scale convolution processing of the source sequence, the feature map enters the residual block. The residual block consists of two stacked 3x3 convolution layers and a shortcut connection which added to the output of two stacked layers. The residual block is connected by residuals, which can increase the nonlinearity of the network, and makes the deep neural network easier to optimize and has better performance. In the residual block, the feature map is subjected to batch normalization, nonlinear activation, and convolution processing in sequence. As the network deepens, the residual network's ability to extract deep features is gradually enhanced. In the residual block, the sequence relationships in the longer time series feature map are captured. The feature map of the residual block output is input to the gated recurrent unit. The gated recurrent unit further processes the feature map to obtain an output sequence. The network model of the Res-GRU is shown in Figure 1.

![Res-GRU network model structure.](image)

3. Experiment

3.1. Data set

The experimental data used in this paper is from the public dataset WikiEnergy. WikiEnergy data set is the power of data Pecan Street Inc. released by WikiEnergy project. The data contains electricity usage data for more than 600 households, including the total household power supply and the data for each individual appliance power supply. The data set has a sampling period of 60 seconds. We divided the data set into two parts, randomly selecting 80% of the data set as the training set and 20% of the data set as the test set. Air conditioning, refrigerator, Wash machine, Microwave oven, Dishwasher are selected for our experiments. The input data of the network model needs to be normalized and the prediction needs to be denormalized.

We use the mean absolute error (Mean Absolute Error, MAE) for the evaluation. We use K- nearest neighbors (KNN), Random forests (RF), and convolutional neural network (CNN) to compare with the network model of this paper.

We carried out the load non-intrusive disaggregation experiments on WikiEnergy household data sets, this experiment selected WikiEngry dataset 18 household electrical power data.
3.2. Experimental results

Figure 2 shows the overall effect of non-intrusive load disaggregation of five types of appliances for WiKiEnergy18 users. The paper uses KNN, Random forest, CNN and other methods to compare with the Res-GRU algorithm.

The comparison results of the evaluation indexes of various algorithms are shown in Table 1. As can be seen from Table 1, among the MAE indicators of five types of appliances, such as air conditioners, refrigerators, washing machines, microwave ovens and dishwashers, Res-GRU has shown obvious advantages in MAE indicators. Res-GRU performs better than other algorithms on five types of electrical appliances.

| index | method     | air conditioning | refrigerator | wash machine | Micro-oven | dishwasher |
|-------|------------|------------------|--------------|--------------|------------|------------|
| MAE   | KNN        | 35.418           | 33.381       | 38.421       | 3.968      | 21.255     |
|       | Random forest | 38.091           | 29.060       | 37.862       | 6.222      | 22.543     |
|       | CNN        | 31.843           | 28.636       | 50.306       | 3.974      | 25.487     |
|       | Res-GRU    | 20.424           | 18.391       | 20.625       | 2.378      | 8.031      |

It can be seen from Fig. 3 that the four algorithms have better disaggregation effects on the air conditioner. Next, we mainly analyze the disaggregation effects of various algorithms in the refrigerator, washing machine, microwave oven and dishwasher. Figure 4 shows the partial disaggregation effect of the four types of appliances in the No. 18 household.
Figure 3. Local decomposition results of the 18th family in the WikiEngry data set.

It can be seen from the figure that the non-intrusive load disaggregation effect of the KNN algorithm on the latter three appliances is the worst, and it is difficult to achieve accurate disaggregation of the mutation point. For the air conditioner, which has obvious regularity and high frequency of use, all the above algorithms can basically achieve effective load disaggregation. In the load disaggregation diagram of the refrigerator, CNN is superior to the other two algorithms in comparison with the Res-GRU algorithm. Since the latter three appliances are used at a lower frequency and the regularity is not obvious, the load disaggregation of the KNN and Random forest methods in the low power consumption region fluctuates greatly. Compared with the load disaggregation effects of the above methods, Res-GRU performs more prominently on each type of electrical appliance, and can achieve more accurate load disaggregation for appliances with lower frequency of use.

4. Conclusion
We propose that the Res-GRU model can solve the problem of gradient disappearance in neural networks. The experimental results show that the Res-GRU model has good performance when dealing with appliances with lower frequency of use. However, the disaggregation of low-frequency electrical appliances is more susceptible to noise interference, and it is difficult to achieve accurate load disaggregation, which requires further research, and further optimization of the model in future research.
Acknowledgments

This work is supported in part by, the National Natural Science Foundation of PR China (61773219) and State Grid Corporation of China Project ‘Fundamental Theory of Dynamic Demand Response Control Based on Large-Scale Diversified Demand Side Resources’.

References

[1] Froehlich, J., Larson, EC, Gupta, S., Cohn, G., Reynolds, MS, Patel, SN: DisAggregate end-use energy sensing for the smart grid. IEEE Pervasive Computing 10 (2011).

[2] Paterakis NG, Erdin O, Bakirtzis AG, et al. Optimal Household Appliances Scheduling Under Day-Ahead Pricing and Load-Shaping Demand Response Strategies. IEEE Transactions on Industrial Informatics, 2017, 11 (6): 1509-1519.

[3] Lin YH, Tsai M S. Development of an Improved Time-Frequency Analysis-Based Nonintrusive Load Monitor for Load Demand Identification. IEEE Transactions on Instrumentation and Measurement, 2014, 63 (6): 1470-1483.

[4] Chang HH, Lin LS, Chen N, et al. Particle-Swarm-Optimization-Based Noninvasive Demand Monitoring and Load Identification in Smart Meters. IEEE Transactions on Industry Applications, 2013, 49 (5): 2229-2236. Education Forum, 2004: 690-694.

[5] Piga D, Cominola A, Giuliani M, et al. Sparse Optimization for Automated Energy End Use Disaggregation. IEEE Transactions on Control Systems Technology, 2016, 24 (3): 1044-1051.

[6] Hassan T, Javed F, Arshad N. An Empirical Investigation of VI Trajectory Based Load Signatures for Non-Intrusive Load Monitoring. IEEE Transactions on Smart Grid, 2014, 5 (2): 870-878.

[7] Tsai MS, Lin Y H. Modern development of an Adaptive Non-Intrusive Appliance Load Monitoring system in electricity energy conservation. Applied Energy, 2012, 96 (8): 55-73.

[8] Saitoh T, Osaki T, Konishi R, et al. Current Sensor Based Home Appliance and State of Appliance Recognition. Sice Journal of Control Measurement and System Integration, 2010, 3 (2): 86-93.

[9] Lin YH, Tsai M S. Non-Intrusive Load Monitoring by Novel Neuro-Fuzzy Classification Considering Uncertainties. IEEE Transactions on Smart Grid, 2014, 5 (5): 2376-2384.

[10] Welikala S, Dinesh C, Ekanayake MPB, et al. Incorporating Appliance Usage Patterns for Non-Intrusive Load Monitoring and Load Forecasting. IEEE Transactions on Smart Grid, DOI:10.1109/TSG.2017.2743760.

[11] Du L, Restrepo JA, Yang Y, et al. Nonintrusive, Self-Organizing, and Probabilistic Classification and Identification of Plugged-In Electric Loads. IEEE Transactions on Smart Grid, 2013, 4 (3): 1371-1380.

[12] Kaiming He, Xiangyu Zhang, Shaqing Ren, and Jian Sun. Deep residual learning for image recognition. CVPR, 770–778, 2016.