

A non-intrusive load identification algorithm based on deep learning and a compound feature

Tong Jiang¹, Ruyu Bai^{1*}

¹School of electrical and Electronic Engineering, North China Electric Power University, Beijing, 102206, China

Abstract. Aiming at the limitations of using a single feature for load identification, a non-intrusive load identification algorithm based on deep learning and compound features is proposed. The pixelated V-I trajectory characteristics and current harmonic characteristics are extracted by analyzing the load data under high-frequency sampling. Using the feature extraction capabilities of neural networks, the combination of pixelated V-I trajectory features and current harmonic features is realized. Finally, the composite feature is used as the new load feature to train the neural network for non-invasive load identification. The experimental results show that the two-layer neural network constructed by the algorithm can take advantage of the complementarity between the two features, thereby improving the load identification ability.

1 Introduction

In recent years, research on smart grid has gradually refined and deepened. Among them, the research of non-intrusive load monitoring technology is an important part of improving the construction of smart grid. Non-intrusive load monitoring means that the types and operating conditions of each load in the load cluster can be analyzed by monitoring the total load data such as voltage and current at the user's power entrance. Non-intrusive load monitoring has a wide range of application prospects in smart grid[1], which is very beneficial to users and power systems. For users, they can understand their own energy consumption behaviors, so as to realize the management of users' energy consumption behaviors and the optimization of power usage schemes, and ultimately achieve the goals of users' rational use of electricity and improvement of power utilization efficiency[2-4]. For the power system, it is possible to understand the proportion of all kinds of loads in the power system, as well as the operation status and power consumption period of each type of load, which provides an important reference for the reasonable arrangement of energy access and power generation plan of the power system, and helps the power system to achieve efficient power dispatch.

In view of the importance of non-intrusive load monitoring, more and more scholars have entered the field to study non-intrusive load identification algorithms. The construction of non-intrusive load identification algorithms is divided into two aspects, one is the selection of features, and the other is the study of classification algorithms used by features. The selection of characteristics can be divided into transient

characteristics and steady-state characteristics in general, and the characteristics can be selected from current, power, harmonics, and so on. Classification algorithms include various algorithms such as cluster analysis algorithms and deep learning algorithms.

Scholars at home and abroad have done a lot of research on non-intrusive load identification algorithms. Paper [5] proposed a method of clustering statistics through the segmented normalized mean-shift clustering method, and finally using the Bayesian model to identify the load. This method uses load power characteristics and integrates the statistics of the time characteristics of the load, and has a good identification effect. Paper [6] proposed a load identification method based on the Relational Recurrent Neural Network (RNN) model. This method uses harmonic components as load characteristics, memorizes historical input characteristics, and establishes a time series input-oriented RNN load identification method. Paper [7] proposed a classification algorithm based on multi-layer perceptron, using the amplitude and phase angle of the 1st, 3rd and 5th harmonics of the steady-state current as load characteristics. The result of load identification is not very good for multi-state loads. Paper [8] compares the identification effects of various common features under high-frequency sampling, and the experimental results show that the V-I trajectory characteristics have the best load identification performance among all the common high-frequency features. Paper [9] shows through research that under steady-state conditions, the change of VI trajectory characteristics caused by the change of load working status is much smaller than the feature difference between two different devices, which proves that V-I trajectory can effectively identify multi-state load.

E-mail: bairuyu@ncepu.edu.cn

Comprehensive analysis of the above papers shows that the load identification performance of the VI trajectory characteristics are the best, and they also have good effect on the identification of multi-state loads. However, a disadvantage of the VI trajectory characteristics is that it is drawn by the normalized voltage and current values, leading to the load identification cannot be effective for loads with similar shapes of VI trajectories but large difference in current amplitudes. If the beneficial effects of current amplitudes can be considered, the accuracy of load identification still has room for further improvement. Although the harmonic components of steady-state current are not very effective in identifying multi-state load, they have complementary effect with the V-I trajectory characteristics drawn by the normalized value because they contain information that can reflect the magnitude of the current amplitude.

Aiming at the problems of the different features analyzed above, and considering the complementarity between the two features, this paper proposes a non-intrusive load identification algorithm based on deep learning and using the combination of pixelated V-I track characteristics and current harmonic characteristics. In this paper, the pixelated V-I track characteristics and current harmonic characteristics are extracted from the high-frequency sampled data of the load, and then the convolutional neural network and backpropagation (BP) neural network of the first layer are used to process the two features respectively to form a compound feature, and the compound feature is used as the new load feature to train the second layer of BP neural network for non-

invasive load identification. Finally, the data from the PLAID data set is used to evaluate and verify the identification effect of the algorithm.

2 Fundamental

2.1 Process of the non-intrusive load identification algorithm

The process of the non-intrusive load identification algorithm in this paper includes five steps: data collection, feature extraction, data processing, classification algorithm design and training, and load identification. The flow of load identification algorithm is shown in Figure 1. The focus of this paper is on the two steps: feature extraction of load steady-state waveforms and design and training of the classification method.

The first step of the algorithm is data collection, which obtains the voltage and current sampling data under the specified sampling frequency at the entrance of the total load, and separates the voltage and current data of each individual load from the total load data through the load switching event detection rules. The detection of load switching events is a problem worthy of separate study, which is not discussed in this paper, as long as the data of each load is obtained. Assuming that only one load is switched on and off at a time point, the voltage and current waveforms of a single load can be obtained

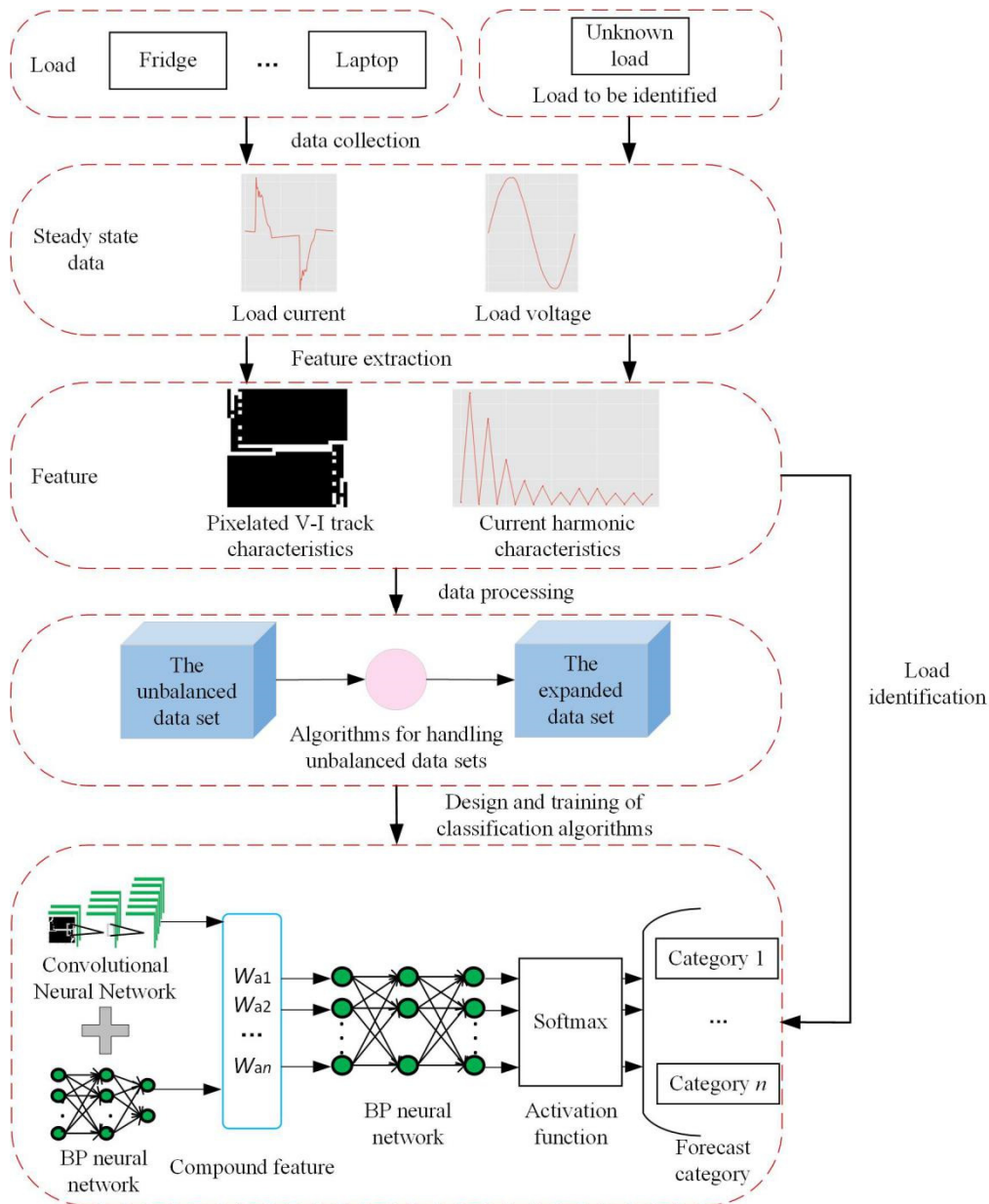


Figure 1. Flow of load identification algorithm.

by calculating the steady-state voltage and current difference before and after the load switching event[10], so as to obtain the data of each load.

The second step of the algorithm is feature extraction. The feature data needed for load identification is extracted from the data of each load by each feature extraction method. This paper will extract the pixelated V-I track characteristics and current harmonic characteristics from the steady-state voltage and current data of the load.

The third step of the algorithm is data processing. When creating the data set used for training the next classification algorithm, due to the difference in the number of samples collected on load of each type, there is an imbalance problem in the data set. Moreover, in order to obtain more training data, it is necessary to expand the original data set by a certain method of processing the unbalanced data set.

The fourth step of the algorithm is the design and training of the classification algorithm. Firstly, select, improve or create new classification algorithm from various existing algorithms, then train the algorithm with the data set, and adjust the parameters of the algorithm to make this classification algorithm meet the requirements of load identification. In this paper, a new classification algorithm is created based on deep learning algorithm, which has a two-layer network. The first layer has two parallel neural networks, a convolutional neural network and a BP neural network. The data of two features are input to two neural networks of the first layer, pixelated V-I track characteristics are input to convolutional neural network, and current harmonic characteristics are input to BP neural network, and then the output vectors of the two neural network hidden layers are combined to form a compound feature. The second layer network is a BP neural network. The compound feature output by the first

layer network are used as the input feature of the second layer network, and the output is the load category.

The last step of the algorithm is load identification. The voltage and current data of the load to be identified are input to the first step of the load identification algorithm, and then the features are automatically extracted by the feature extraction method created by the algorithm, and the feature data will be transmitted to the previously trained classification algorithm for load category prediction, that is, load identification.

2.2 Feature extraction

2.2.1 Pixelated V-I track characteristics

V-I trajectory characteristics are the feature extracted from the steady-state voltage and current data of the load under high-frequency sampling. The original V-I trajectory refers to the graph formed by directly taking the voltage and current as the horizontal and vertical coordinates, and drawing points according to the coordinate values of each sampling point. This paper uses pixelated V-I track characteristics, inspired by handwritten digit recognition in image recognition technology, and maps the original V-I trajectory into an image with a certain resolution and only 0 and 1 values for each pixel. The pixelated V-I track can not only reflect the shape of the original V-I trajectory to a high degree, but also the calculation is much simpler. Therefore, this paper adopts the pixelated V-I track characteristics. The method of extracting pixelated V-I track characteristics is as follows:

- Obtain the load voltage and current data in a steady-state cycle under high-frequency sampling. Assuming that the current frequency is f and the sampling frequency is fs , the number of sampling points in a cycle is:

$$N = \frac{fs}{f} \quad (1)$$

N is the number of sampling points in a cycle.

- Assuming that the resolution of the pixelated V-I track characteristics is d , their values are composed of a $d \times d$ order matrix. The voltage and current value of the load is linearly converted into an integer between 0 and d . The current and voltage conversion formula for each sampling point is:

$$i_m = \left\lfloor \frac{I_m - \min I}{\max I - \min I} d \right\rfloor \quad m = 1, 2, \dots, d \quad (2)$$

$$v_m = \left\lfloor \frac{V_m - \min V}{\max V - \min V} d \right\rfloor \quad m = 1, 2, \dots, d \quad (3)$$

I_m and V_m are the original current and voltage values of the m th sampling point; I_m and V_m are the current and voltage values of the m th sampling point after linear conversion; $\min I$ and $\min V$ are the minimum values of current and voltage in a steady-state cycle; $\max I$ and $\max V$ are the maximum values of

current and voltage in a steady-state cycle; d is the order of the matrix; $\lfloor \cdot \rfloor$ is the symbol of rounding down.

- Create a $d \times d$ order matrix with all elements being 0. From the first sampling point of the steady-state cycle to the end of the N th sampling point, when the m th sampling point is reached, assign the value of the element in the I_m th row and the V_m th column of the matrix to 1 to obtain the pixelated VI track characteristics formed by the $d \times d$ order matrix.

2.2.2 Current harmonic characteristics

Current harmonic characteristics are extracted from the load steady-state current data. Since the current is a discrete time-domain signal obtained under high-frequency sampling, it is necessary to convert the time-domain data to the frequency domain to obtain the current harmonics of the load. This paper uses Fast Fourier Transform (FFT) to extract the current harmonic characteristics of the load[11].

The current signal in a steady-state cycle in this paper has N sampling points, $x(n)$ is the n th sampling point of the signal, $X(k)$ is the component of the k th harmonic of the time-domain signal, and the FFT formula of the time-domain signal is:

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{nk} \quad k = 1, 2, \dots, N-1 \quad (4)$$

W_N^{nk} is the rotation factor, and its calculation formula is:

$$W_N^{nk} = e^{-j\frac{2\pi}{N}nk} \quad n = 0, 1, \dots, N-1 \quad (5)$$

Suppose the k th frequency domain signal after FFT is $X(k) = a(k) + jb(k)$, and its modulus and phase angle are $|X(k)| \arg[X(k)]$, then the effective value of each current harmonic is:

$$I_k = \frac{|2X_i(k)|}{\sqrt{2N}} \quad k = 1, 2, \dots, N-1 \quad (6)$$

I_k is the effective value of the k th current harmonic.

3 The classification algorithm

3.1 Design of the classification algorithm

The pixelated V-I track characteristics are an image feature, and convolutional neural network is suitable for image recognition. Therefore, this paper uses the convolutional neural network structure commonly used to process images to process the pixelated V-I track characteristics, and further adjust the structure and parameters of the convolutional neural network model according to the characteristics. Both the processing of current harmonic characteristics and the classification using the compound feature adopt the BP neural network model, and the network structure is mainly determined

according to the dimension of the input vector. The specific network structure and parameters of the classification algorithm are shown in Table 1.

3.2 Training of the classification algorithm

Since non-intrusive load identification is a multi-classification problem, and the suitable activation function for multi-classification problems is Softmax, the activation functions of the output layers of the three networks in the classification algorithm in this paper are all set to Softmax. The label output by Softmax is a

multi-dimensional vector, the dimension is the same as the number of load categories, and the sum of the vector elements is 1, where the value of each element represents the probability that the prediction result is the i th type of load, then the i with the largest output element value is the predicted load category.

Since the classification algorithm in this paper is composed of three independent neural networks, each network can be trained separately when the algorithm is trained. This modular idea realizes the

Table 1. Network structure and parameters of the classification algorithm.

Network type	Layer type	Output dimension	Kernel function dimension	Activation function
Convolutional neural network of the first layer	Input layer	(32,32,1)	-	-
	Convolution layer	(32,32,10)	3×3(10)	Relu
	Max-pooling layer	(16,16,10)	2×2	-
	Convolution layer	(16,16,18)	3×3(18)	Relu
	Flatten layer	4608	-	-
	Dense layer	256	-	Relu
BP neural network of the first layer	Dense layer	11	-	Softmax
	Input layer	11	-	-
	Dense layer	32	-	Relu
	Dense layer	128	-	Relu
BP neural network of the second layer	Dense layer	11	-	Softmax
	Input layer	384	-	-
	Dense layer	128	-	Relu
	Dense layer	11	-	Softmax

decoupling of the classification algorithm network structure, and provides convenience for the replacement and expansion of the various modules of the algorithm in the follow-up research. The training steps of the classification algorithm are as follows:

- Extract the pixelated V-I track characteristics and current harmonic characteristics of the load from the steady-state voltage and current data of the load.
- Construct the first layer of the classification algorithm, and then use the pixelated V-I track characteristics and the current harmonic characteristics of the load as the input of the two neural networks of the first layer, and use the vectorized load categories with the same dimension as the output label of the output layer as the labels to train two networks for load identification.
- After the training of the two networks of the first layer is completed, combine the output vectors of the two hidden layers of the neural networks. The output vector of the hidden layer of the first network has a dimension of 256, and the output vector of the hidden layer of the second network has a dimension of 128, and the combination becomes a vector of dimension 384, which is the compound feature.
- Construct the second layer of the classification algorithm, use the compound feature as the input of the neural network of the second layer, and use the

vectorized load categories with the same dimension as the output label of the output layer as the labels to train the neural network of the second layer for load identification.

4 Instance verification

The use of this section to evaluate and verify the load identification algorithm proposed in this paper through specific instance data.

4.1 Source of the instance data

The data used by the instance comes from the PLAID data set. The PLAID data set is a public data set that contains the voltage and current data of 11 types of household appliances. The data are collected from more than 60 households in the United States, and the sampling frequency is 30kHz. There are a total of 1793 sample data of 312 electrical appliances in the data set, and each sample data contains data within a few seconds from the transient state to the steady state. The PLAID data set is often used in the research of load identification for analysis. This paper selects the voltage and current data of the last 20 steady-state cycles of all samples in

the PLAID data set, and takes the average value as the voltage and current data of each sample.

4.2 Feature extraction of the instance

The data volume of the pixelated V-I track characteristics depends on the resolution of the track image. Since the mathematical relationship between the data volume and the resolution is a square relationship, the increase in the resolution will cause an exponential increase in the data volume. Of course, the recognition accuracy will also be improved. In order to achieve a good balance between the accuracy and the amount of data, this section uses the sample data in the instance to perform image experiments on the pixelated V-I track characteristics at different resolutions. It is found from experiments that at a resolution of 50, the V-I track has been sufficiently reflected and has many details. But in fact, it is not necessary to restore such sufficient image features in details, because when the resolution reaches a certain value, the increase in resolution cannot effectively improve the accuracy. Therefore, by viewing the V-I track images at various resolutions, it is found that when the resolution is 32, the V-I track can be reflected to a sufficient degree and the details are not much disappeared, and the accuracy rate has reached a high degree. Therefore, the pixelated V-I track characteristics with a resolution of 32 are used in the load identification algorithm of this instance.

When extracting the current harmonic characteristics from the sample data, it is found that the current only has obvious values at the odd-numbered harmonics, and when the order is higher, the harmonic values of different samples become very small, which is of little use value. Therefore, the first 11 odd-numbered harmonics of the current are used as the current harmonic characteristics in the load identification algorithm of this instance.

4.3 Results of the instance

From the samples of each type of load, 10% of the sample data is selected to form the test data set, and there are 174 sample data in the test data set. The training data set consists of 2464 samples, which is expanded by the synthetic minority oversampling technique for unbalanced data sets.

The test data set is input to the classification algorithm trained by the training data set for load identification, and the load identification accuracy rate of the load identification algorithm in this paper is 0.845. The instance result shows that the non-intrusive load identification algorithm based on deep learning and the compound feature proposed in this paper has a good load identification effect.

4.4 Analysis of the algorithm effect

In order to prove the improvement of the load identification effect of the algorithm in this paper, a general neural network is used to separately perform load identification on the two features selected in this paper,

and the accuracy of the load identification using pixelated V-I track characteristics is 0.782, and the accuracy of using current harmonic characteristics is 0.701.

Comparing the experimental results, it is found that the accuracy of the load identification of the compound feature selected in this paper through the two-layer neural network constructed in this paper is significantly higher than that of using a single feature to identify the load through a general neural network. Experimental results prove that the non-intrusive load identification algorithm based on deep learning and the compound feature proposed in this paper effectively improves the performance of load identification.

5 Conclusion

Aiming at the problem that the V-I trajectory cannot reflect the magnitude of the current amplitude and the insufficient effect of current harmonics on multi-state load identification, this paper proposes a non-intrusive load identification algorithm based on deep learning and a compound feature.

This paper uses neural networks in deep learning algorithms to combine pixelated V-I track characteristics and current harmonic characteristics to form a compound feature. By comparing the accuracy of load identification using the compound feature and single features, it proves that the use of the compound feature can achieve the complementarity of the pixelated V-I track characteristics and the current harmonic characteristics, thereby improving the load identification ability. Research in follow-up work can be the feasibility of using high-frequency transient characteristics for load identification and the optimization of the load identification algorithm.

Acknowledgments

Thanks to colleagues for their help with the work related to the paper, as well as the tutor's guidance and support for the research.

References

1. Yu, Y., Liu, Y. (2015) Challenging issues of smart grid. *Automation of Electric Power Systems*, 39(2): 1–5.
2. Leng, X., Chen, G., Jiang, Y., Zhang, J., Cao, Y. (2018) Data specification and processing in big-data analysis system for monitoring and operation of smart grid. *Automation of Electric Power Systems*, 42(19): 175–182.
3. Yang, T., Zhai, F., Zhao, Y., Pen, H. (2019) Explanation and prospect of ubiquitous electric power Internet of Things. *Automation of Electric Power Systems*, 43(13): 9–20.
4. Wang, S., Sun, Z., Liu, Z. (2015) Co-scheduling strategy of home energy for smart power utilization.

- Automation of Electric Power Systems, 39(17): 108–113.
5. Tian, Z., Xu, Q., Li, R., Zhang, J., Zhao, S. (2021) A non-intrusive load identification decision-making method based on time characteristics. <http://kns.cnki.net/kcms/detail/23.1202.TH.20200803.1424.016.html>.
 6. Liu, H., Shi, S., Xu, X., Zhou, D., Min, R., Hu, W. (2019) A non-intrusive load identification method with associated RNN model. *Power system protection and control*, 47(13): 162–170.
 7. Zhang Z., Chen, H.N., Luo, X.W. (2018) A supervised event-based non-intrusive load monitoring for non-linear appliances. *Sustainability*, 10(4): 1001.
 8. Gao, J.K., Kara, E.C., Giri, S., Berges, M. (2015) A feasibility study of automated plug-load identification from high-frequency measurements. In: 2015 IEEE Global Conference on Signal and Information Processing. Orlando. pp. 220-224.
 9. Develder, C, Dhaene, T., Deschrijver, D., Gao, J., Berges, M. (2017) Handling imbalance in an extended PLAID. In: 5th IFIP Conference on Sustainable Internet and ICT for Sustainability. Funchal. pp. 1-5.
 10. Dhaene, T, Deschrijver, D., Develder, C., Berges, M. (2018) VI-based appliance classification using aggregated power consumption data. In: 2018 IEEE International Conference on Smart Computing. Taormina. pp. 179-186.
 11. Srinivasan, D., Ng, W.S., Liew, A.C. (2018) Neural-network-based signature recognition for harmonic source identification. *IEEE Transactions on Power Delivery*, 21(1): 398–405.