

RESEARCH

Open Access



Timing shift-based bi-residual network model for the detection of electricity stealing

Jie Lu, Jingfu Li^{*} , Wenjiang Feng, Yongqi Zou, Juntao Zhang and Yuan Li

^{*}Correspondence:
jingfuli@cqu.edu.cn
School of Microelectronics
and Communication
Engineering,
Chongqing University,
Chongqing 400044, China

Abstract

With the increasing number of electricity stealing users, the interests of countries are jeopardized and it brings economic burden to the government. However, due to the small-scale stealing and its random time coherence, it is difficult to find electricity stealing users. To solve this issue, we first generate the hybrid dataset composed of real electricity data and specific electricity stealing data. Then, we put forward the timing shift-based bi-residual network (TS-BiResNet) model. It learns the features of electricity consumption data on two aspects, i.e., shallow features and deep features, and meanwhile takes time factor into consideration. The simulation results show that TS-BiResNet model can detect electricity stealing behaviors that are small scaled and randomly coherent with time. Besides, its detection accuracy is superior to the benchmark schemes, i.e., long short-term memory (LSTM), gated recurrent unit (GRU), combined convolutional neural network and LSTM (CNN-LSTM) and Bi-ResNet.

Keywords: Electricity stealing, Detection model, Bi-ResNet, Timing shift

1 Introduction

Electricity is crucial to build the social economy, where its consumption and generation can affect the overall policy of one country [1]. In the past decade, thanks to the development of smart grid and intelligent meters, the power systems became more intellectualized and robust. Accordingly, the efficiency of energy utilization got rise and tactics of power scheduling was more reasonable. However, due to the increase in electricity consumption and the rise in electricity prices, the electricity charge brought considerable economic burden to the populace and enterprises [2]. Thereby, the problem of electricity stealing [3] became more severe which may lead to a great economic loss for one country. To solve this issue, amounts of research on effective detection methods have been studied and, among them, two types of ideas are popular, i.e., to monitor the physical characteristics of meters, or to analyze features of electricity data.

For the first type thought, the data of the smart meters are tampered illegally and the detection methods are supposed to find the abnormal data. According to types of electrical parameters, the electricity stealing methods can be divided into five categories [4], i.e., under-voltage stealing, under-current stealing, phase shift stealing, spread stealing, no-meter stealing. For the first three kinds of methods, the common

approach is to improve the structure of intelligent meters whose threshold value of each electrical parameter is set, e.g., [5, 6]. As for the latter two methods, the common approach is to monitor the changes of electricity consumptions with networks, e.g., [7, 8]. However, because the electricity data are varied and redundant, the mentioned approaches are hard to achieve synchronous transmission and efficient management in energy systems, and meanwhile, their ability of abnormal data detection is limited.

To handle these issues and improve the capability of detection, the second type thought attracts wide attention. The core idea of these methods is that the electricity data are collected in energy systems and its features are analyzed with various algorithms. In the beginning, the energy systems mainly calculate the line loss rate (LLR) [9] to determine whether users steal electricity. However, because of the massive data, this method brings heavy workload to energy systems, which makes detection efficiency become low. To handle this issue, some new techniques, e.g., grey model (GM) [10, 11] and neural network (NN) [12, 13], are adopted to predict LLR automatically, but the detection models are still limited by the downside of LLR. It means that only long-term electricity stealing can be identified, while the short-term electricity stealing cannot be detected. Subsequently, various studies concern more about the features of electricity data rather than LLR. Specifically, paper [14, 15] utilizes support vector machines (SVM) to seek for the hyperplanes which can distinguish ordinary users and electricity stealing users. Paper [16, 17] adopts long and short-term memory (LSTM) network to study the time characteristics of electricity data and achieves anti-electricity theft load. Paper [18, 19] leverages convolutional neural network (CNN) to deeply learn the features of electricity data, and the electricity stealing users can be identified accurately with these features. Although the mentioned methods can recognize electricity stealing users with kinds of features gotten from training process, when the total amount of stolen electricity is small or its change does not have strong regularity on time, the performance of these methods drops dramatically.

From the above description, we know that kinds of detection models have been proposed and the ability to detect electricity stealing has been improved. However, for the small scale of electricity stealing or the stealing behaviors is not strongly related to the time, the existing methods do not work well on the electricity stealing detection. Thus, in this paper, we put forward a timing shift-based bi-residual network (TS-BiResNet) to consider the features of electricity data on three aspects, i.e., shallow features, deep features and time factor. The main contribution of this paper is fourfold as follows:

- Owing to the fact that electricity stealing data are difficult to acquire, we first analyze the distribution of a real electricity dataset without any electricity stealing behaviors. Then, we generate specific electricity stealing data based on the distribution of a real electricity dataset and the principle of electricity stealing. After preprocessing, the obtained hybrid dataset, composed of real electricity data and specific electricity stealing data, can be used for the training and testing of our detection model.
- In order to consider the deep features of electricity stealing data, we redesign the residual network (ResNet) detection model for the hourly electricity data within

one week. The model leverages shortcut connections to preserve shallow features so that the issue about vanishing gradient can be avoided during deep learning process and this model is fit for the detection of small-scale electricity stealing.

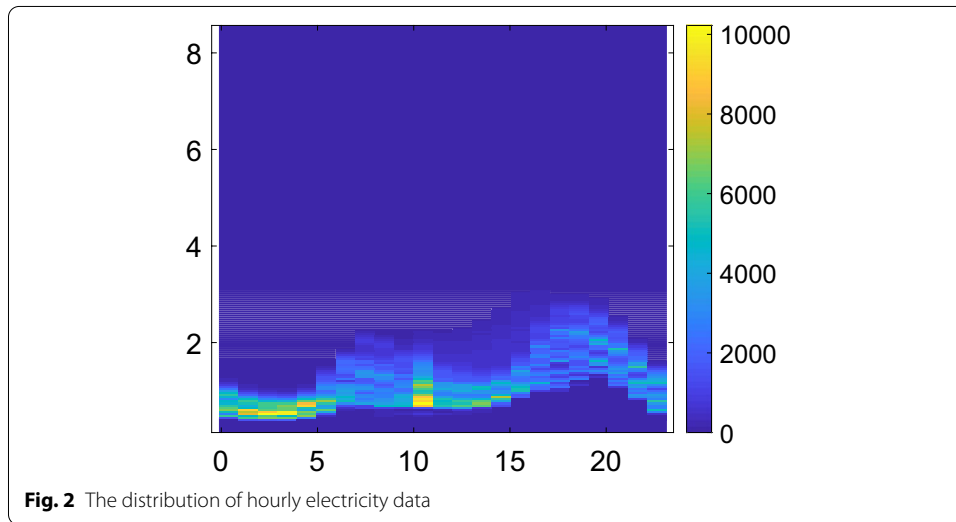
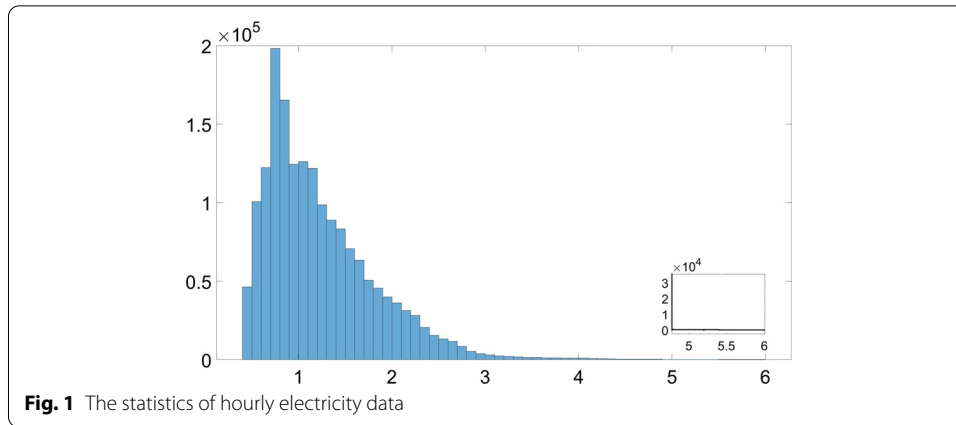
- To make the model applicable to the electricity stealing users with varying temporal correlation, we propose timing shift (TS) preprocessing and improve ResNet to a two-layer structure, i.e., Bi-ResNet. For the former, the electricity data in different periods can be selected by the same convolution kernel. From which, the time factor is no longer restricted by the time interval. For the latter, the convolution kernels of different sizes are used in the same layer, so that Bi-ResNet model can extract time features with different durations.
- To further analyze the performance of TS-BiResNet model, we perform numerical evaluations via simulations. The simulation results show that the accuracy of our proposed model is higher than that of existing work. Meanwhile, this model can also detect the electricity stealing users whose scale of stolen electricity is small or time periods of electricity stealing are not regular.

The rest of this paper is organized as follows. In Sect. 2, we first analyze the distribution of real electricity dataset based on hourly electricity data and time interval and then generate specific electricity stealing data. In Sect. 3, the Bi-ResNet model is put forward and its input data are preprocessed by the TS preprocessing in Sect. 4. The performance of our proposed model is verified by simulation results in Sect. 5. Finally, we draw conclusions in Sect. 6.

2 Problem formulation: electricity stealing data

It is known that electricity data are often confidential due to the state secrecy provisions [20], which means that the real data are hard to be obtained by the populace. For this case, we take the open-source dataset provided by National Renewable Energy Laboratory [21] as the original data \mathbf{S} , Denoted as $D1$. This dataset contains the hourly electricity data of 936 users, i.e., $\mathbf{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_{936}\}$. For each user $k, \forall k \in \{1, \dots, 936\}$, it records the electricity data within 1 year, i.e., $\mathbf{s}_k = \{s_{k,1}, \dots, s_{k,365}\}$, and for any day $d, \forall d \in \{1, \dots, 365\}$, there are 24 hourly electricity data within it, i.e., $\mathbf{s}_{k,d} = \{s_{d,1}, \dots, s_{d,24}\}$. However, in this dataset, no electricity stealing users are available there. To make the data suitable for the detection models, we first analyze the real electricity dataset and then provide a universal method to generate electricity stealing data. Thus, we randomly select the hourly electricity data of 400 users in half a year, i.e., $\mathbf{S}_{\text{selected}} \in \mathbb{C}^{168 \times 26 \times 400}$. From which, 168 is the quantity of hourly electricity data within 1 week, 26 denotes the amounts of weeks within half a year, and 400 presents the number of all selected users. For convenience, we abbreviate $\mathbf{S}_{\text{selected}}$ as \mathbf{S} . The statistical results of \mathbf{S} are shown as Figs. 1 and 2.

Figure 1 presents the statistics of hourly electricity data for all users within half a year. It indicates that, although the range of hourly electricity data is wide, i.e., from 0.4 to 6, its values are mainly concentrated between 0.4 and 2.3, especially around 1. It means that when the amount of stolen electricity account for a tiny proportion of the electricity consumption during period of time, e.g., 10% or 20%, the value of electricity detected by energy systems is still within the normal range and it does not slump rapidly. For this case, the stealing users are hard to be detected so that deep features are desired.



Besides, we further analyze the distribution of hourly electricity data on the basis of time period. Figure 2 shows that the electricity behaviors of all users are correlated with time. Specifically, in the wee hours, i.e., from 10:00 p.m. to 6:00 a.m., the distribution of electricity consumption is concentrated and its value is relatively small. For this case, the range of electricity stealing is limited and behaviors are easy to be detected, where the traditional detection methods and neural network-based methods can detect the electricity stealing users successfully. In the morning and at noon, i.e., from 6:00 a.m. to 4:00 p.m., the amount of electricity consumption is rising and the variance becomes large. For this case, electricity stealing is relatively easy to be detected but some traditional detection methods will fail. As for the last case, i.e., from 4:00 p.m. to 10:00 p.m., the electricity consumption of all users reaches its peak, and its distribution is the most scattered. At this moment, the electricity stealing is the most difficult to be found. Herein, we focus on the latter two cases, i.e., electricity stealing behavior is relatively difficult to be detected.

Based on above analysis, we randomly preprocess 150 users out of selected 400 users as the electricity stealing users. From which, the original data are divided into

two parts, i.e., $\mathbf{S} = \{\mathbf{S}_{\text{real}}, \mathbf{S}_{\text{stl}}\}$, where $\mathbf{S}_{\text{real}} \in \mathbb{C}^{168 \times 26 \times 250}$ is the real electricity data and $\mathbf{S}_{\text{stl}} \in \mathbb{C}^{168 \times 26 \times 150}$ is the specific electricity stealing data. According to the [4], the electricity stealing methods can be equivalent to the product of electricity consumption and a stealing weighting factor. Thus, we define the weighting factor as electricity stealing coefficient (ESC) and stipulate that the value of ESC ranges from 0.2 to 1, i.e., $\text{ESC} \in (0.2, 1)$. Meanwhile, we further define the duration of electricity stealing as electricity stealing period (ESP). Considering the possibility of stealing electricity shown in Fig. 2, the ESP is expressed as

$$\text{ESP} = r(\{6, 22\}), \quad (1)$$

where $r(\cdot)$ is the random function which returns an array. Herein, the electricity stealing function is designed as

$$f(s_{k,i}) = \begin{cases} \text{ESC} \times s_{k,i}, (182 - \lfloor i/7 \rfloor \times 7) \in \text{ESP} \\ s_{k,i}, (182 - \lfloor i/7 \rfloor \times 7) \notin \text{ESP} \end{cases}, \quad (2)$$

where $\lfloor \cdot \rfloor$ is the round down operation. Herein, we can get the hybrid dataset $\tilde{\mathbf{S}} = \{\mathbf{S}_{\text{real}}, \tilde{\mathbf{S}}_{\text{stl}}\}$ which is composed of real electricity data \mathbf{S}_{real} and specific electricity stealing data $\tilde{\mathbf{S}}_{\text{stl}}$, as shown in Table 1.

Notice that, electricity stealing users with different ESC come from the same users in original dataset. For each experiment, only ordinary users and electricity stealing users with one kind of ESC are selected. This operation can reflect how the ESC and ESP affect the accuracy of electricity stealing detection.

3 Method: Bi-ResNet model

After getting the hybrid dataset composed of real electricity data and specific electricity stealing data, various detection models can be adopted [9–19]. However, due to two features of $\tilde{\mathbf{S}}$, i.e., small scale of electricity stealing and random coherence with time, the performance of these models is not optimal and deep features are needed. In the process of seeking for deep features, the value of electricity consumption data is small and its change trend is usually gentle. For this case, if we take neural networks with multiple layers, the vanishing gradient [22] will affect the performance of detection models. Aiming at this issue, ResNet model [23] attracts wide attention. It adds some shortcut connections to preserve shallow features so that the problem of vanishing gradient can be solved. However, the performance of ResNet is still not optimal, because every time when the convolution

Table 1 Hybrid dataset composed of real electricity data and specific electricity stealing data

Type	Application	Size	Quantity	ESC	ESP
Ordinary	Training	168	5200	None	None
	Testing	168	1300	None	None
Stealing	Training	168	2600	0.2	$r(\{6, 22\})$
	Testing	168	1300	0.2	$r(\{6, 22\})$

	Training	168	2600	0.9	$r(\{6, 22\})$
	Testing	168	1300	0.9	$r(\{6, 22\})$

kernels extract features of electricity data, the obtained information is limited. Therefore, in this paper, we introduce feature-map attention [24] and then propose Bi-ResNet.

3.1 Unit block of Bi-ResNet

The unit block of Bi-ResNet consists of two kinds of basic components, i.e., convolution modules and shortcut connections, and two kinds of logical operations, i.e., copy and vector addition. Its structure is shown in Fig. 3. For convenient analysis, we further divide the unit block into three steps.

For the step 1, the obtained $\bar{\mathbf{S}}$ is first mapped from three dimension to two dimensions for the adaption of the input. For the user k on the d -th day, the input data $\bar{\mathbf{S}}_{k,d}$ are changed from the $\mathbf{s}_{k,d}$ as follows

$$\bar{\mathbf{S}}_{k,d} = \begin{bmatrix} s_{d,1} & s_{d,2} & \cdots & s_{d,167} & s_{d,168} \\ s_{d,2} & s_{d,3} & \cdots & s_{d,168} & s_{d,1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{d,167} & s_{d,168} & \cdots & s_{d,165} & s_{d,166} \\ s_{d,168} & s_{d,1} & \cdots & s_{d,166} & s_{d,167} \end{bmatrix} \quad (3)$$

When $\bar{\mathbf{S}}_{k,d}$ is put into the network, the copy function $\text{cp}(\cdot)$ is taken, which duplicates the input data into two same copies

$$\{\mathbf{X}_1, \mathbf{X}_2\} = \text{cp}(\bar{\mathbf{S}}_{k,d}). \quad (4)$$

Then, two convolution modules with different size of convolution kernels are taken. Specifically, each convolution module function $i, \forall i \in \{1, 2\}$ consists of three parts, i.e., convolution function $\text{conv}(\cdot)$, batch normalization $\text{BN}(\cdot)$ and rectified linear unit function $\text{relu}(\cdot)$. We use $H_{i,\text{size}}(\cdot)$ function to represent the i -th convolution model with $\text{size} \times \text{size}$ convolution kernel, thereby the i -th pseudo-output \mathbf{Y}_i is calculated by

$$\mathbf{Y}_i = H_{i,\text{size}}(\mathbf{X}_i). \quad (5)$$

Notice that, when two convolution kernels with different sizes are used for the same input data, the features can be extracted more thoroughly, that is essential for deep features.

For the step 2, we fuse the same feature under different scopes, i.e., convolution modules with 3×3 convolution kernel and 5×5 convolution kernel and then use one convolution module with 1×1 convolution kernel to extract the deep feature based on the fused feature. Specifically, we first duplicate the pseudo-output $\mathbf{Y}_i, i \in \{1, 2\}$ as

$$\{\mathbf{X}_{i \times 2 + 1}, \mathbf{X}_{(i+1) \times 2}\} = \text{cp}(\mathbf{Y}_i). \quad (6)$$

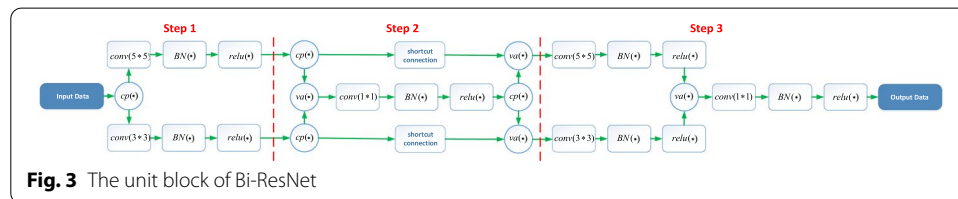


Fig. 3 The unit block of Bi-ResNet

After duplicating, $\mathbf{X}_{i \times 2+1}$ are connected by a shortcut connection that can maintain the shallow feature. As for $\mathbf{X}_{(i+1) \times 2}$, they are fused by the $\text{va}(\cdot)$ function as

$$\text{va}(\mathbf{X}_4, \mathbf{X}_6) = 1/2 \times \mathbf{X}_4 + 1/2 \times \mathbf{X}_6, \quad (7)$$

and the deep feature is extracted by the 3-rd convolution model with 1×1 convolution kernel. Accordingly, the 3-rd pseudo-output \mathbf{Y}_3 is calculated by

$$\mathbf{Y}_3 = H_{3,1}(\text{va}(\mathbf{X}_4, \mathbf{X}_6)). \quad (8)$$

To make the pseudo-output contain the deep feature and the shallow feature, we duplicate the \mathbf{Y}_3 as

$$\{\mathbf{X}_7, \mathbf{X}_8\} = \text{cp}(\mathbf{Y}_3), \quad (9)$$

and fuse them with the $\mathbf{X}_{i \times 2+1}, i \in \{1, 2\}$ which come from shortcut connections, calculated by

$$\text{va}(\mathbf{X}_{i \times 2+1}, \mathbf{X}_{i \times 3+1}) = 1/2 \times \mathbf{X}_{i \times 2+1} + 1/2 \times \mathbf{X}_{i \times 3+1}. \quad (10)$$

Herein, we get two set of excellent pseudo-output with hybrid features. One is the feature that maintains the shallow feature gotten by 3×3 convolution kernel and meanwhile contains more attentive and deep feature gotten by 5×5 convolution kernel. On the contrary, the other one maintains the shallow feature gotten by 5×5 convolution kernel and contains more attentive and deep feature gotten by 3×3 convolution kernel.

For the step 3, the features of pseudo-output $\text{va}(\mathbf{X}_{i \times 2+1}, \mathbf{X}_{i \times 3+1}), i \in \{1, 2\}$ are extracted by two convolution modules with 5×5 convolution kernel and 3×3 convolution kernel. The pseudo-output $\mathbf{Y}_{i+3}, i \in \{1, 2\}$ is calculated by

$$\mathbf{Y}_{i+3} = H_{i+3, \text{size}}(\text{va}(\mathbf{X}_{i \times 2+1}, \mathbf{X}_{i \times 3+1})). \quad (11)$$

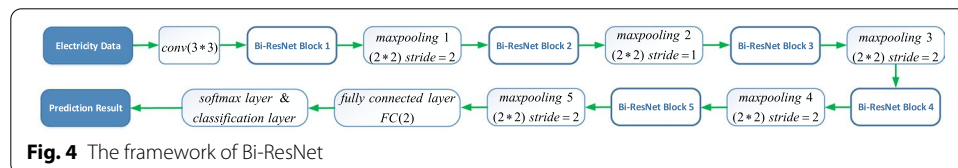
Then, the $\mathbf{Y}_{i+3}, i \in \{1, 2\}$ are merged again with $\text{va}(\cdot)$ function and 6-th convolution model with 1×1 convolution kernel, expressed as

$$\mathbf{Y}_6 = H_{6,1}(\text{va}(\mathbf{Y}_4, \mathbf{Y}_5)). \quad (12)$$

Notice that, \mathbf{Y}_6 is the final output of the Bi-ResNet's unit block and, compared with the input data $\bar{\mathbf{S}}_{k,d}$, its dimension is reduced by 6, i.e., from 168×168 to 162×162 .

3.2 Framework of Bi-ResNet

After a series of mixed and cross-learning, the final hybrid feature contains various deep features and shallow features with different levels of attention, thereby the vanishing gradient disappears and deep features are obtained. Herein, the framework of Bi-ResNet is developed, as shown in Fig. 4.



Considering that the input data of detection model are the hourly electricity data within one week, we design Bi-ResNet which contains the 3×3 convolution kernel, 5 Bi-ResNet blocks, 4 maxpooling layers with stride 2, 1 maxpooling layer with stride 1, and some traditional components of CNN [25]. Specifically, due to the fact that shallow features are important but relatively easy to extract, we first pick a convolution kernel with smaller size, i.e., 3×3 , and Bi-ResNet block is unnecessary for this step. Then, we use Bi-ResNet blocks and maxpooling layers to extract deep features. Notice that, when the input data are processed by the Bi-ResNet blocks, maxpooling layers with stride 2 and maxpooling layers with stride 1, the size of output data is reduced by 4, halved, and reduced by 1, respectively. Therefore, a hourly electricity data with the size of 168×168 are put into the Bi-ResNet and the output data with the size of 6×6 can be obtained after maxpooling 5. Subsequently, the type of hourly electricity data will be determined by the fully connected layer, and classification layer distinguishes whether the user steals electricity.

4 Method improvement: timing Shift Preprocessing

For the hourly electricity data, our proposed Bi-ReNet model can take both the deep features of electricity data and time factor into account, so the prediction results related to the electricity stealing users are accurate. However, the learning of time factor is not sufficient. Taking a 10×10 input data and 3×3 convolution kernel for instance, the convolution process is shown as Fig. 5.

Due to the fact that the size of convolution kernel is fixed, the width of each convolution is the same and the vertical interval will not change. This means that neural networks can only learn the features of time factor with respect to the same time period, e.g., 1, 2, 3 corresponding to 3 hours and time interval, e.g., 1, 2, 3 1, 10, 19 corresponding to per 9 h. To handle this issue, we propose timing shift algorithm and its specific process is shown in Table 2.

Notice that, the value of stride st is randomly selected from 0 to 23 due to the features of electricity stealing behaviors, i.e., the time period for electricity stealing will change

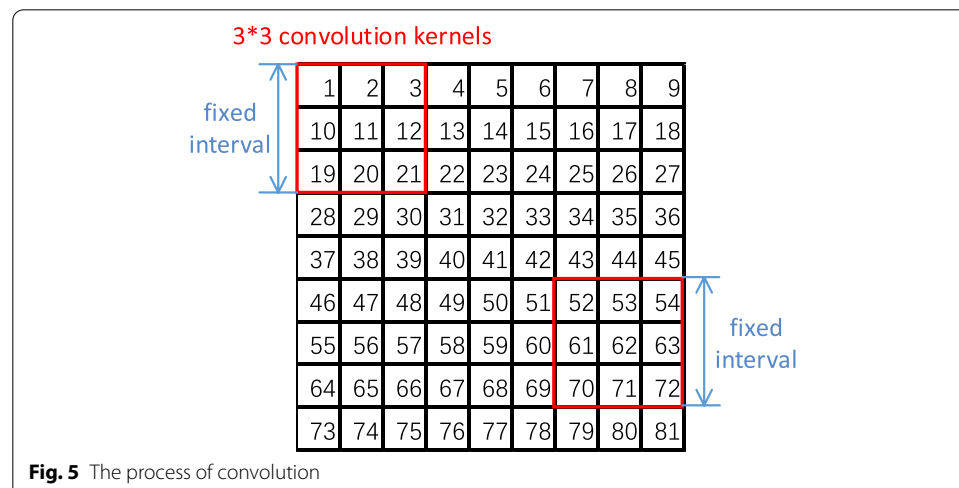
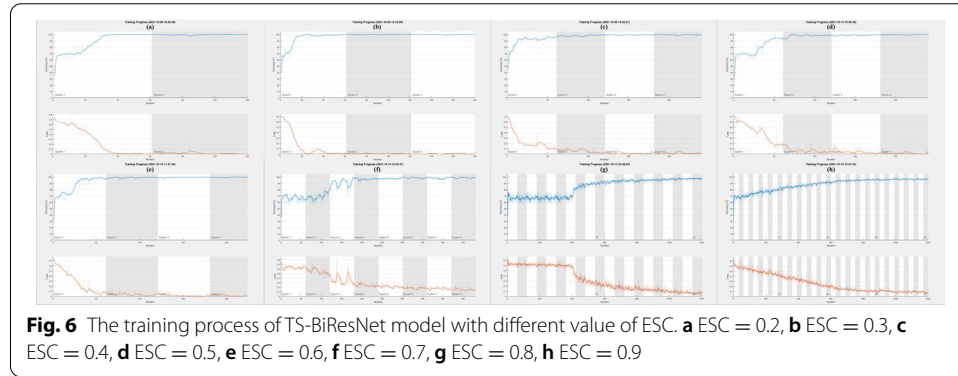


Table 2 Timing shift algorithm

Input: The segment hourly electricity data $\{s_{k,1}, \dots, s_{k,7}\}$ with size 1×168
Output: timing shift based input data $\bar{S}_{k,d}$ with size 168×168
Initialize $\bar{S}_{k,d}$, stride st , pseudo-data $X = \{s_{k,1}, \dots, s_{k,7}\}$
record $\bar{S}_{k,d}(1,:) = X$
for $i=2 : \text{size}(\bar{S}_{k,d})$
$st = \text{random}(0, 23)$
for $d=1 : 7$
$\bar{s}_{k,d} = [s_{k,d}(st : \text{size}(s_{k,d})) \quad s_{k,d}(1 : (st - 1))]$
$X = \{\bar{s}_{k,1}, \dots, \bar{s}_{k,7}\}$
end
record $\bar{S}_{k,d}(i,:) = X$
end



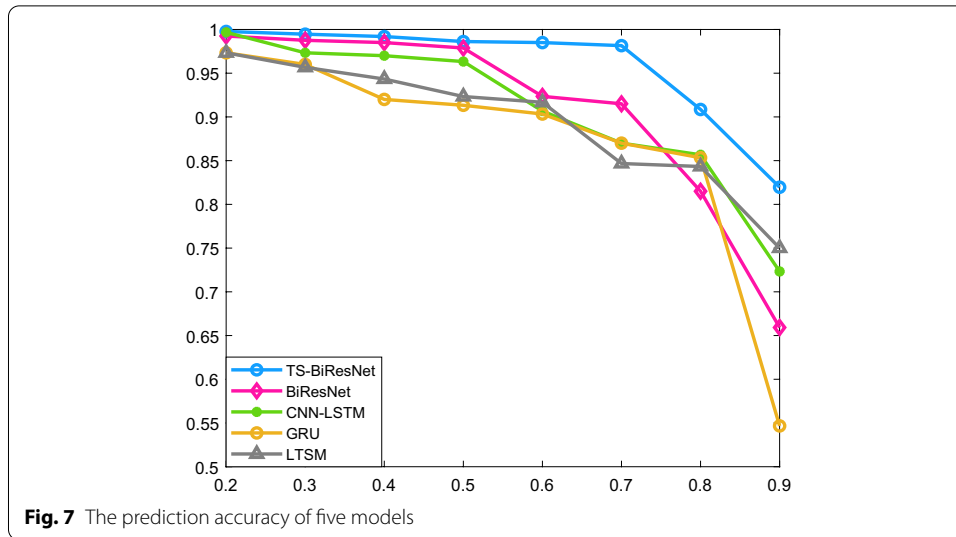
based on date. For this case, we should extract the features of different dates and time periods, so that the combination of different time periods is considered.

5 Results and discussion

In this section, we present numerical results on three aspects, i.e., the training process of TS-BiResNet model, the accuracy of detection, and the further discussion about generalization ability. The first part verifies that TS-BiResNet model is fit for the detection of electricity stealing and its training process is completed successfully. For the second part, it evaluates the performance of the proposed model by comparing it with four benchmark models, i.e., LSTM model [16], Bi-ResNet model [26], gated recurrent unit (GRU) model and combined convolutional neural network and LSTM (CNN-LSTM) model [27, 28]. The simulation results are presented as Figs. 6 and 7, respectively. Then, the issue about generalization ability is discussed and analyzed with two additional datasets, whose simulation results are presented in Table 3.

5.1 Training process

Figure 6 is the training process of the TS-BiResNet model with different values of ESC. We take the situation that ESC = 0.8 as an example to analyze the training process of the model. Since the curves, i.e., the accuracy and the loss function, share the

**Table 3** Comparison of prediction accuracy for three datasets

Dataset	ESC	TS-BiResNet (%)	Bi-ResNet (%)	LSTM (%)	GRU (%)	CNN-LSTM (%)
D1	0.6	98.50	92.35	91.67	90.33	90.67
	0.7	98.15	91.50	84.67	87.00	87.00
	0.8	90.85	81.50	84.33	85.33	85.67
	0.9	81.96	65.92	75.00	54.67	72.33
D2	0.6	83.08	82.67	NONE	NONE	73.50
	0.7	82.60	80.00	NONE	NONE	70.25
	0.8	78.27	73.25	NONE	NONE	58.67
	0.9	63.94	60.93	NONE	NONE	52.33
D3	0.6	95.85	93.67	91.00	92.67	87.33
	0.7	93.15	90.25	87.67	85.67	81.00
	0.8	90.65	81.27	68.93	75.33	66.33
	0.9	67.54	60.77	62.67	56.33	52.67

same variation trend, we only analyze the curve of accuracy. The figure reflects that shallow features and deep features of the hourly electricity data coexist and both of them can be learned by TS-BiResNet model. Specifically, the accuracy of detection soars into 65%, while the accuracy fluctuates around this value for several epochs. This is because we take 3×3 convolution kernel to find the shallow features but this kind of preliminary detection is ineffective due to the fact that the value of ESC is large, so that deep features are desired. After long training, the accuracy of detection increases sharply at the time integration = 450, reaching 80%. Then, its value rises gradually and reaches the peak around 98%. It indicates that deep features have great impact on the accuracy of detection and these features can be found by the TS-BiResNet model. Overall, the two curves are convergent, which means that the proposed model is suitable for detecting electricity stealing users in the dataset *D1*, and there is no problem about over-fitting and under-fitting during prediction process.

5.2 Prediction accuracy

Figure 7 is the performance comparison between the TS-BiResNet model and four benchmark models in terms of prediction accuracy. According to the mentioned analysis, we set the value of ESC ranges from 0.2 to 0.9. It verifies the TS-BiResNet model has capability to detect electricity users for extreme situations, i.e., the total amount of stolen electricity is small or its change does not have strong regularity on time. In general, the accuracy of all models is declined with the increase in ESC, this comes from the fact that the value of electricity stealing data becomes closer to the value of ordinary data, so that the features of electricity stealing users are difficult to be found. Specifically, when the value of ESC ranges from 0.2 to 0.5, the features of electricity stealing users are relatively obvious. Thereby, the accuracy of TS-BiResNet model, BiResNet model and CNN-LSTM model is very high, more than 98.0%, which exceeds the accuracy of other two models more than 5%. The reason for this phenomenon is that these three models learn features of electricity stealing while the other does not. Notice that, when the value of ESC reaches 0.5, the accuracy of BiResNet model and CNN-LSTM model decrease dramatically. This is because BiResNet model and CNN-LSTM need to extract the shallow features of electricity stealing for prediction, while the shallow features are difficult to be found and deep features are desired. When ESC ranges from 0.6 to 0.7, although the performance of BiResNet model deteriorates due to the inconspicuous shallow features, it is still superior to the performance of other four models. Besides, when ESC ranges from 0.7 to 0.8, the accuracy of TS-BiResNet model and BiResNet model experiences a great decline. This is because the shallow features disappear due to the distribution of electricity consumption data, so that only deep features and time factors can be used for the detection. It is worth noting that the performance of GRU model is greater than that of Bi-ResNet model when $ESC = 0.8$. The reason for this phenomenon is that GRU model can utilize the temporal correlation of data rather than make use of shallow features, so that its performance is close to the GRU model and LSTM model. At this time, the performance of CNN-LSTM model, GRU model and LSTM model is greater than that of Bi-ResNET but worse than BiResNet model. Last but not least, the TS-BiResNet model is able to detect electricity stealing users even when the value of ESC is 0.8 and 0.9, reaching 90.8% and 81.9% accuracy, respectively. It shows that the TS-ResNet can learn deep features and time factors for users with small scale of electricity stealing and varying temporal correlation, and its performance is better than that of four benchmark models. (When $ESC = 0.9$, the performance of GRU model deteriorates because of its model structure)

5.3 Further analysis

Based on the above analysis, we get the conclusion that our proposed TS-Bi-ResNet model can achieve detection for the users with small scale of electricity stealing and varying temporal correlation. However, it is worth noting that the convergence of model is a necessary condition for overfitting problem, and the generalization ability should be further proved. Thus, we further introduce two different datasets [29] and [30], denoted as $D2$ and $D3$, respectively.

In specific, $D2$ is a dataset consisting of the hourly electricity data of 32 users within half a year. Compared with the dataset $D1$, we can verify that whether our proposed

model is suitable for small dataset. This is because $D1$ and $D2$ have some similar characters, i.e., they are, respectively, taken from the residential users within a limited region, which means that the difference among the maximum value of each user is small. Besides, for each user, the habit with respect to the electricity consumption is the same as the others, which means that the distribution of electricity consumption is similar. However, the scale of $D1$ is 10 times larger than $D2$, so that $D2$ is a small dataset.

As for $D3$, it is a dataset containing the hourly electricity data of 913 companies within half a year. Compared with the dataset $D1$, we can analyze that whether our proposed model is fit for the dataset with various individual differences. This is because, although the data of companies are taken from the same region, the scale and habit of electricity consumption are varied, e.g., for company A , the scale of electricity consumption may be 10 times larger than company B , and meanwhile, company A 's peak electricity consumption is in the morning, while Company B 's peak electricity consumption is in the afternoon. In general, compared with $D1$, $D3$ is a dataset with large individual differences.

For these three types of datasets, we carried out simulations, respectively, and the simulation results are shown in Table 3. It indicates that, for the small dataset $D2$, the performance of all models decreases sharply. It means that recurrent neural network (RNN)-based prediction model and deep neural networks (DNN)-based prediction model are not fit for small sample classification or prediction, especially for the former. This is because LSTM and GRU require temporal depth and regularity, that is to say the scale of the data should be large and the data should be temporally correlated. In comparison, although the prediction effect of the TS-BiResNet and Bi-ResNet model is not good enough, when the value of ESC is low, e.g., 0.6 or 0.7, they can still achieve a certain accuracy of detection, around 80%. As for the $D3$ dataset with large individual differences, the performance of different models varies greatly. In specific, when the value of ESC is low, e.g., 0.6 or 0.7, the performance of the models is similar to the case in $D1$, except for the CNN-LSTM model. The reason why the performance of CNN-LSTM model drops dramatically is that it tries to get the correlation of the same moment in different weeks, rather than finding the relationship between the current moment and the next moment. However, it seems that there is no regulation can be found so that CNN-LSTM model does not perform well. Besides, when the value of ESC is 0.8, the performance of TS-BiResNet and Bi-ResNet model maintains good performance, reaching 90.65% and 81.27%, respectively. It proves that our proposed model has capability to detect electricity stealing users even if they have significant individual differences, but other models cannot. However, this advantage is not kept when ESC is 0.9, in other words, the detection of all models fails. This is because the individual difference is so huge that the models cannot determine whether it is individual difference or electricity stealing behaviors. In general, for large-scale datasets, our proposed model can effectively detect electricity stealing users, proving its generalization, while the performance of the detection regarding to small-scale datasets should be improved.

6 Conclusions

For electricity stealing detection, we first analyze the distribution of the real electricity dataset and behaviors of electricity stealing. Based on this, we generate the hybrid dataset composed of real electricity data and specific electricity stealing data. To detect electricity

stealing users, we design Bi-ResNet model for learning deep features of electricity data. Then, we propose the TS-Bi-ResNet model which takes time factor into consideration. Simulation results show that our proposed model can achieve detection for the users with small scale of electricity stealing and varying temporal correlation. Meanwhile, its detection accuracy is roundly superior to the four benchmark models. Its generalization ability is further discussed and simulated with two additional datasets.

Abbreviations

TS-BiResNetL: Timing shift-based bi-residual network; LSTM: Long short-term memory; LLR: Line loss rate; GM: Grey model; NN: Neural network; CNN: Convolutional neural network; SVM: Support vector machines; ResNet: Residual network; TS: Timing shift; ESC: Electricity stealing coefficient; ESP: Electricity stealing period.

Acknowledgements

Not applicable.

Authors' contributions

JL revised this paper, replied to the comments given by the reviewer, collected the data set, added new experiments and solved the problems in the simulations. JL put forward the concept of this paper, built and optimized the model, made the performance simulations, and wrote the paper. Other authors assisted in related work and revised the paper. All authors read and approved the final manuscript.

Authors' information

Jie Lu received her B.Eng. degree in electronic information engineering from the South-Central Minzu University. Now, she is a postgraduate student in Chongqing University. She is currently focusing on the research about next generation wireless communication, interference management and deep learning.

Jingfu Li received his B.Eng. degree in communication engineering from the Chongqing University of Posts and Telecommunications and M.Eng. degree in information and communication engineering from the Chongqing University. Now, he is a doctoral candidate in Chongqing University. He is currently focusing on the research about the wireless communication, interference management and machine learning. Accordingly, three SCI papers and two EI papers have been obtained.

Wenjiang Feng received the Ph.D. degree in electrical engineering from Chongqing University, Chongqing, China, in 2000. He is currently a Professor with the College of Microelectronics and Communication Engineering, Chongqing University. His research interests include all aspects of MIMO communication, including limited feedback techniques, antenna design, interference management and full-duplex communication, cognitive radio, special mobile communication systems, and emergency communication. He is a Peer Review Expert of the Natural Science Foundation of China and is a Senior Member of the China Institute of Communications. He also serves as an Editorial Board Member of Data Communication, China.

Yongqi Zou received his B.Eng. degree in electronic information engineering from the Shandong University Of Technology. Now, she is a master degree candidate in Chongqing University. She is currently focusing on the research about the deep learning in the field of energy.

Juntao Zhang received his B.Eng. degree in communication engineering from the Chongqing University. Now, he is a master candidate in Chongqing University. He is currently focusing on the research about the radio frequency fingerprinting of wireless devices and machine learning.

Yuan Li received his Bachelor of Science degree in electronic information science and technology from the Chong Qing University. Now, he is a postgraduate student in Chongqing University. He is currently focusing on the research about backscatter communication system and deep learning.

Funding

The work was supported by the Science and Technology Project of State Grid Chongqing Electric Power Company under Grant 2021 Yudian Technology No.8 and Chongqing Basic Science and Frontier Technology Research Project under Grant cstc2017jcyjBX0047.

Availability of data and materials

The original data can be found in <https://data.openai.org/submissions/153/>. Besides, if the original code is needed, please contact author Jingfu Li for data requests.

Declarations

Ethics approval and consent to participate

There is no experimental research that need to be approved by ethics committee. Also, there is no research related to human or animals.

Competing interests

The authors declare that they have no competing interests.

Received: 7 December 2021 Accepted: 23 March 2022

Published online: 08 April 2022

References

1. G. Memarzadeh, F. Keynia, Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electr. Power Syst. Res.* **192**, 106995 (2021)
2. F. Jamil, E. Ahmad, The relationship between electricity consumption, electricity prices and GDP in Pakistan. *Energy Policy* **38**(10), 6016–6025 (2010)
3. T.B. Smith, Electricity theft: a comparative analysis. *Energy Policy* **32**(18), 2067–2076 (2004)
4. C. Cheng, H. Zhang, Z. Jing, M. Chen, L. Yang, Study on the anti-electricity stealing based on outlier algorithm and the electricity information acquisition system. *Dianli Xitong Baohu yu Kongzhi Power Syst. Prot. Control* **43**(17), 69–74 (2015)
5. W. Xuewei, W. Decong, Application of electric energy meter transformer and anti-electricity loss technology. Application of electric energy meter transformer and anti-power loss technology (2000). (in China)
6. F. Wang, F. Yang, T. Liu, X. Hu, Measuring energy meter of three-phase electricity-stealing defense system, in *2011 6th IEEE Conference on Industrial Electronics and Applications*, pp. 11–15 (2011)
7. D. Zheng, W. Shuai, Research on measuring equipment of single-phase electricity-stealing with long-distance monitoring function, in *Power and Energy Engineering Conference, 2009. APPEEC 2009. Asia-Pacific* (2009)
8. M. Zhang, X. Liu, Y. Shang et al., Research on comprehensive diagnosis model of anti-stealing electricity based on big data technology. *Energy Rep.* **8**, 916–925 (2022). (2021 International Conference on New Energy and Power Engineering)
9. H. Liu, J. Yang, Y. Zhang et al., Monitoring of wastewater treatment processes using dynamic concurrent kernel partial least squares. *Process Saf. Environ. Prot.* **147**, 274–282 (2021)
10. F. Utaminigrum, S.J.A. Sarosa, C. Karim et al., The combination of gray level co-occurrence matrix and back propagation neural network for classifying stairs descent and floor. *ICT Express* **8**, 151–160 (2021)
11. N. Xu, Y. Dang, Y. Gong, Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in china. *Energy* **118**(JAN.1), 473–480 (2017)
12. T. Ahmad, R. Madonski, D. Zhang et al., Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: key developments, challenges, and future research opportunities in the context of smart grid paradigm. *Renew. Sustain. Energy Rev.* **160**, 112128 (2022)
13. L. Ke, G. Wenyan, S. Xiaoliu et al., Research on the forecast model of electricity power industry loan based on GA-BP neural network. *Energy Procedia* **14**, 1918–1924 (2012). (2011 2nd International Conference on Advances in Energy Engineering (ICAEE))
14. H. Hong, Y. Su, P. Zheng, N. Cheng, J. Zhang, A SVM-based detection method for electricity stealing behavior of charging pile. *Procedia Comput. Sci.* **183**(2), 295–302 (2021)
15. X. Yan, N.A. Chowdhury, Mid-term electricity market clearing price forecasting: a hybrid LSSVM and ARMAX approach. *Int. J. Electr. Power Energy Syst.* **53**, 20–26 (2013)
16. Y. Shen, P. Shao, G. Chen, X. Gu, J. Zhu, An identification method of anti-electricity theft load based on long and short-term memory network. *Procedia Comput. Sci.* **183**(8), 440–447 (2021)
17. Z. Wang, D. Jiang, F. Wang et al., A polymorphic heterogeneous security architecture for edge-enabled smart grids. *Sustain. Cities Soc.* **67**, 102661 (2021)
18. E.U. Haq, J. Huang, H. Xu et al., A hybrid approach based on deep learning and support vector machine for the detection of electricity theft in power grids. *Energy Rep.* **7**, 349–356 (2021). (2021 The 4th International Conference on Electrical Engineering and Green Energy)
19. Y. Shen, P. Shao, G. Chen et al., An identification method of anti-electricity theft load based on long and short-term memory network. *Procedia Comput. Sci.* **183**, 440–447 (2021). (Proceedings of the 10th International Conference of Information and Communication Technology)
20. C. L. Government, Law on the protection of state secrets of the People's Republic of China (september 5, 1988). *Chin. Law Govern.* **51**, 79–83 (1994)
21. O. E. D. I. (OEDI), Commercial and residential hourly load profiles for all tmy3 locations in the United States, <https://data.openet.org/submissions/153/>
22. Z. Hu, J. Zhang, Y. Ge, Handling vanishing gradient problem using artificial derivative. *IEEE Access* **10**(99), PP (2021)
23. S. Fan, Y. Wang, S. Cao et al., A deep residual neural network identification method for uneven dust accumulation on photovoltaic (PV) panels. *Energy* **239**, 122302 (2022)
24. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2016, pp. 770–778 (2016)
25. K. Gopalakrishnan, S.K. Khaitan, A. Choudhary et al., Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Constr. Build. Mater.* **157**, 322–330 (2017)
26. Z. Kou, Y. Fang, An improved residual network for electricity power meter error estimation. *Int. J. Pattern Recognit. Artif. Intell.* **33**(8), 1959024.1–1959024.19 (2019)
27. M. Ismail, M.F. Shaaban, M. Naidu, E. Serpedin, Deep learning detection of electricity theft cyber-attacks in renewable distributed generation. *IEEE Trans. Smart Grid* **11**(4), 3428–3437 (2020)
28. Y. He, G.J. Mendis, J. Wei, Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism. *IEEE Trans. Smart Grid* **8**(5), 2505–2516 (2017)
29. UMassTraceRepository, Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States, <https://data.openet.org/submissions/153>
30. O. E. D. I. (OEDI), The project of UMass Trace Repository supported in part by the National Science Foundation under Grants CNS-323597 and 0325868, <https://traces.cs.umass.edu/index.php/Smart/Tools>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.