

Abnormal electricity detection with hybrid deep neural network model

Jie Liu¹, Xiang Cao^{2*}, Diangang Wang¹, Kejia Pan¹, Cheng Zhang¹, and Xin Wang¹

¹State Grid Sichuan information & communication company 610041, China

²College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 200090, China

Abstract. This paper tackles a new challenge in abnormal electricity detection: how to promptly detect stealing electricity behavior by a large-scale data from power users. Proposed scheme firstly forms power consumption gradient model by extracting daily trend indicators of electricity consumption, which can exactly reflect the short-term power consumption trend for each user. Furthermore, we design the line-losing model by analyzing the difference between power supplying and actual power consumption. Finally, a hybrid deep neural network detection model is built by combining with the power consumption gradient model and the line-losing model, which can quickly pin down to the abnormal electricity users. Comprehensive experiments are implemented by large-scale user samples from the State Grid Corporation and Tensorflow framework. Extensive results show that comparing with the state-of-the-arts, proposed scheme has a superior detection performance, and therefore is believed to be able to give a better guidance to abnormal electricity detection.

1 Introduction

With the development of smart grid, abnormal electricity, including the leakage, stealing behavior, has brought great loss for State Grid Corporation [1]. Detecting abnormal electricity behavior has become very popular. As one of effective solution methods, it is very important to deep mine the massive historical data of power companies and then achieves the leakage, stealing behavior and other abnormal patterns of detection [2].

Since daily electricity information recorded in the smart meter directly reflects the user's electricity consumption behavior, it is possible to find the difference between the normal and the abnormal users. Some methods have been developed. For example, in [3], the scheme can not directly calculate or measure the NTL, it needs to be estimated from other preliminary results, i.e. subtracting the technical loss from the total loss. A new automatic feature analysis method using wavelet technology is proposed in [4], which combined with multiple classifiers to identify fraud distribution networks. With the sophisticated electricity stealing technology development, the successful detection rate of abnormal electricity consu ditional detection methods are unsuitable for the abnormal electricity detection, because it greatly increases the cost of testing due to always combining manual field

* Corresponding author: caoxiang99@126.com

surveys, or temporarily increase the calibration of electrical energy metering equipment. Another scheme finds the characteristic that the power consumption from stealing users always has a downward trend and then compares the correlations among the consumption curves[8][9]. The abnormal users can finally be identified by the calculated Pearson coefficients. In [10][11], Nagi et al employ support vector machine (SVM) to implement a serial of experiments for NTL. Unfortunately, this method still has an inferior performance.

Facing these aforementioned problems, ones may ask that: how to promptly detect stealing electricity behavior by a large-scale data from power users? In this work, we try to answer this question with the following contributions:

- We propose a novel detection model by employing multiple deep neural networks, which is termed as hybrid deep neural network model. We emphasize that proposed scheme can perform a quick detection for abnormal electrical user.
- We design power consumption gradient model and the line-losing model, and then combine them to construct a hybrid deep neural network detection model, which can exactly reflect the short-term power consumption trend for each user and quickly pin down to the abnormal electricity users.
- Comprehensive experiments with a large-scale data from State Grid Corporation are tested. We built a general detection framework by using Tensorflow to implement a serial of experiments. The experimental results demonstrate that proposed scheme significantly improves the detectability, and shows superior performance over the existing schemes.

The rest of this paper is organized as the follows. We present proposed hybrid depth neural network model in Section 2. Subsequently, comprehensive experiments are performed to evaluate the performance of proposed scheme. The corresponding experimental results and discussions are presented in Section 3. Finally, Section 4 concludes the paper.

2 Hybrid detection model based on deep neural network

In this section, we try to build a hybrid detection model to effectively reflect whether the user has the leakage electric activity. Proposed model firstly introduces the identification process for user with the stolen electricity. Furthermore, we show the data extraction process, which is considered as the source of power consumption gradient model and the line-losing model. Finally, combining these two mentioned feature models, an efficient hybrid detection model is formed.

2.1 Identification process

Abnormal electrical behavior detection can be divided into five steps: data extraction, data exploration, preprocessing data, modeling diagnosis, the results of feedback and other steps as shown in figure 1.

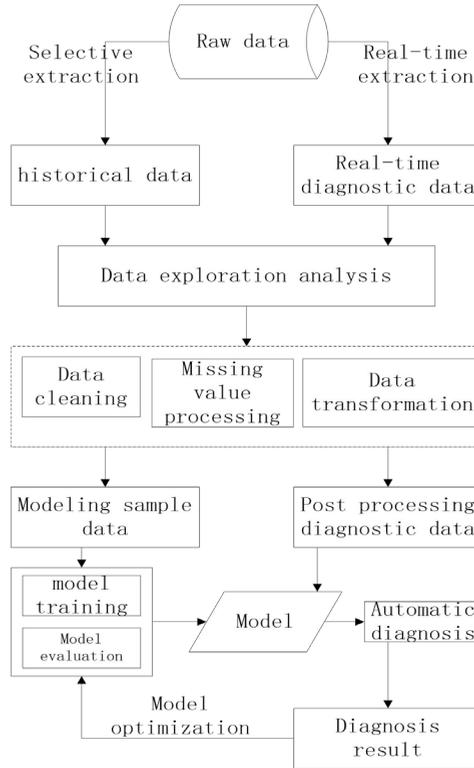


Fig. 1. The framework of detecting abnormal electricity behavior.

2.2 Data extraction

Since the raw data relating to abnormal electricity consumption contains the following aspects: electricity consumption information, line-losing information, alarm information and user data. Notably, we analyze the data and the corresponding signs of whether there exists electricity leakage every day from 20800 users in five years, including different types of normal and abnormal electricity users. Based on the overall analysis for all the users, there is not any non-resident electricity leakage (including government, hospitals, schools, etc.). Thus, we believe that the detection for non-resident electricity data is unnecessary.

2.3 Evaluation model

Although the collected raw electricity consumption data and alarm data can reflect some abnormal electricity consumption behaviors, it is obviously insufficient to summarize the vast majority of abnormal electricity consumption behaviors and need to undergo a transformation to generate new more efficient indicator definitions as follows.

Model 1: Power consumption gradient model

We define power trend decline indicator T

$$T = \sum_{n=i-t_1}^{i+t_2} D(n) \tag{1}$$

where, D (i) is the count of days the statistical trend is declining, that is

$$D(i) = \begin{cases} 1, & k_i < k_{i-1} \\ 0, & k_i \geq k_{i-1} \end{cases} \quad (2)$$

k_i is the statistics for 5 days before and after a total of 11 days, the power consumption of the slope change

According to model 1, we can use the days before and after as the statistical window to observe the downward trend during this period, and use the slope of the linear fitting of power as the measure. If the slope declines over time, then the user's abnormal electricity is very possible.

Model 2: Line-losing model

Defining the line-losing indicator $E(i)$.

$$E(i) = \begin{cases} 1, & \frac{V_i^1 - V_i^2}{V_i^2} > 1\% \\ 0, & \frac{V_i^1 - V_i^2}{V_i^2} \leq 1\% \end{cases} \quad (3)$$

where, V_i^1 is the average of line-loss rates for the after 5 days, V_i^2 is the average of line-loss rates for the previous 5 days.

According to **model 2**, we can use it to detect abnormal electricity behavior occurs because the relevant line-losing rate will be significantly rised .On the other hand, it can also detect the occurrence of abnormal electrical behavior through the calculation of the growth rate of line loss rate.

Model 3: Alarm model

Defining alarm index A_i :

$$A_i = \sum_{i=0}^n u_i(a) \quad (4)$$

A_i is to count the number of occurrences of these alarms within a certain period of time. Three types of user power anomaly feature indicators have been given, the next section to introduce how to construct a hybrid model.

2.4 Hybrid depth neural network model

Based on TensorFlow, a hybrid deep neural network model is constructed by using training samples, and the test set is predicted by using the generated neural network model. Figure 2 is the proposed depth neural network model:

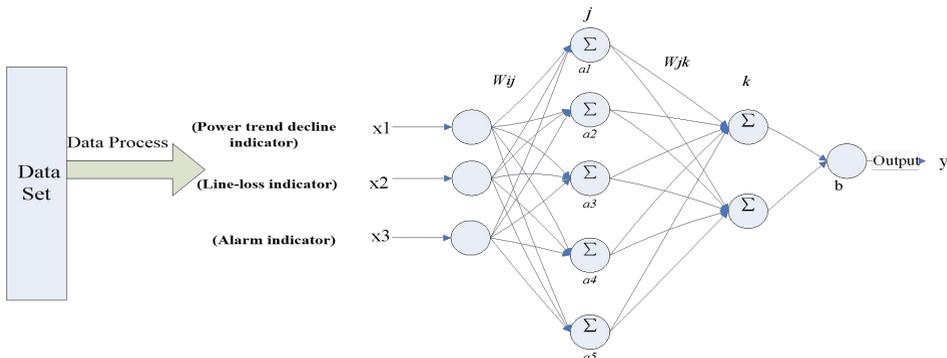


Fig. 2. Hybrid depth neural network model.

3 Experimental results

3.1 Construction of expert samples

For all abnormal and normal users that may have the electric leakage behavior, we firstly select raw data of electricity load, line loss data and protection alarm for data cleaning, and then perform conversion according to the abnormal electricity behavior evaluation index to obtain 29100 sets of sample data. Finally, the composition of expert sample library is shown in Table 1.

Table 1. The data of expert samples.

Time	User Id	T	Ei	Ai	Classify
2016/09/06	9900***154	4	1	1	1
2016/09/20	9900***431	4	0	4	1
2016/09/14	9900***154	9	0	0	0
2016/09/17	9900***050	3	1	0	0

3.2 Experimental setup

Preparation of expert samples completed, a total of 29100 samples were obtained. Attribute variable A contains three attributes: power trend down indicator, line loss indicators, and alarm indicators. The decision attribute Y is whether the user has an identifier of abnormal electricity usage. 20% of the experts' samples are randomly selected as the test samples and the remaining 80% of the training samples. Since the total sample size is 29,100, the training set size is 23,200 and the test set size is 5900.

Statistics were made on the prediction results, the confusion matrix of the model prediction results was obtained, and the results were visualized by Matplotlib. The training set's forecast result is shown in Figure 3, and the test set's forecast result is shown in Figure 4.

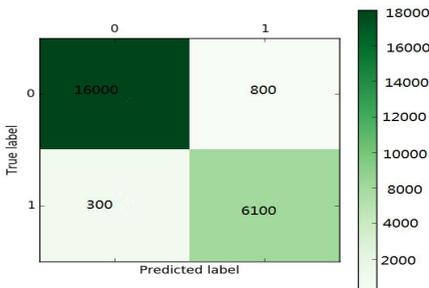


Fig. 3. Confusion matrix of training set prediction results.

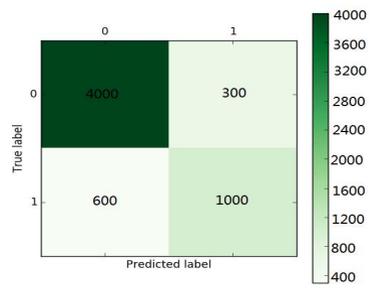


Fig. 4. Confusion matrix of test set forecast results.

According to the confusion matrix of training set prediction results, we can see that the classification accuracy rate.

$$Accuracy = \frac{TP + TN}{P + N} = \frac{6100 + 16000}{6100 + 800 + 16000 + 300} = 95.3\% \tag{5}$$

The Classification accuracy:

$$Precision = \frac{TP}{TP + FP} = \frac{6100}{6100 + 800} = 88.4\% \tag{6}$$

The Recall rate:

$$Recall = \frac{TP}{TP + FN} = \frac{6100}{6100 + 300} = 95.3\% \tag{7}$$

The normal user is misjudged as a stolen electricity user accounts for 4.8%, stolen electricity user is misjudged as normal user accounts for 4.7%. From the results of the test set, it can be found that the prediction model is reliable and the probability of over-fitting is low.

3.3 Performance comparison

In this paper, the user data of a certain province in the north is used to process the data that is given to the model through the method of the second section of this paper. Respectively with SVM , Random Forest for comparison. As shown in Figure 5.

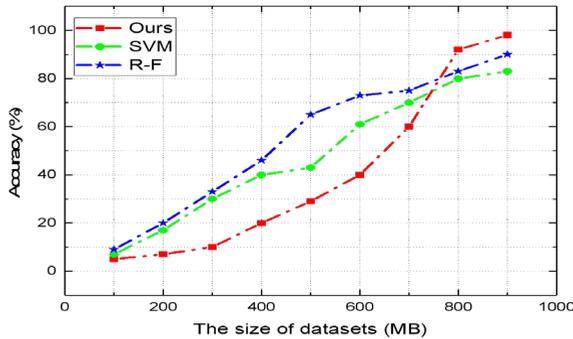


Fig. 5. Model comparison.

Through the above findings, when the quantity set is relatively small, the SVM and random forest model are better than the hybrid DNN model proposed in this paper, but when the data set size is higher than a certain data set size, the hybrid DNN model is obviously better than SVM and Random Forest model.

The final test results using the hybrid depth neural network in this paper are shown in Table 2.

Table 2. Abnormal electricity scene audit results table.

User ID	User Name	Abnormal electricity start date	Result
7110***608	xx electronics Co., Ltd.	2016.6.2	True
9900***537	xx Investment Co., Ltd.	2016.8.20	True
8210***409	xx town xx economic union	2016.8.10	True
8910***571	xx toy factory Co., Ltd.	2016.2.23	False
6710***757	xx economic union	2016.2.21	False

4 Conclusion

The main work of this paper is to establish the identification process of abnormal electricity user, to extract the features of abnormal electricity detection based on the original data, and to establish three technical indicators (electric power trend decline indicator, line loss indicator, Alarm index), and build an expert sample. Based on the supervised hybrid depth neural network, the abnormal electric power detection can detect the abnormal electric

behavior quickly after a long time of training, resulting in a higher accuracy and help the relevant departments to narrow the scope of the investigation. A hybrid deep neural network is characterized by long training time, but after obtaining the model, online analysis and forecasting time is very short, with strong online analytical capabilities.

By comparing with SVM and random forest, the hybrid depth neural network model is more accurate than SVM and random forest when the training dataset is large. Although the training model takes a long time, the training process mainly involves a large number of iterative behaviors, so the Spark platform can take advantage of the memory operation characteristics to optimize the operation.

Acknowledgement

This research was financially supported by National Natural Science Foundation of China under Grants (No. 61472236, 61672337, 61602295, and 61562020).

References

1. Dincer I. Energy and GDP analysis of OECD countries. *Energy conversion and management*, (1997), 38(7): 685-696
2. Smith T B. Electricity theft: a comparative analysis. *Energy Policy*, (2004), 32(18): 2067-2076
3. Nizar A H, Dong Z Y, Zhao J H, et al. A data mining based NTL analysis method. *Power Engineering Society General Meeting, 2007. IEEE. IEEE*, (2007): 1-8
4. Jiang R, Tagaris H, Lachs A, et al. Wavelet based feature extraction and multiple classifiers for electricity fraud detection. *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES. IEEE*, (2002), 3: 2251-2256
5. Leal A G, Boldt M. A big data analytics design patterns to select customers for electricity theft inspection. *Transmission & Distribution Conference and Exposition-Latin America (PES T&D-LA), 2016 IEEE PES. IEEE*, (2016): 1-6
6. Smith T B. Electricity theft: a comparative analysis. *Energy Policy*, (2004), 32(18): 2067-2076
7. Yap K S, Tiong S K, Nagi J, et al. Comparison of supervised learning techniques for non-technical loss detection in power utility. *International Review on Computers and Software*, (2012), 7(2): 626-636
8. Monedero I, Biscarri F, León C, et al. Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees. *International Journal of Electrical Power & Energy Systems*, (2012), 34(1): 90-98
9. Gontijo E M, Delaiba A C, Mazina E, et al. Fraud identification in electricity company customers using decision tree. *Systems, Man and Cybernetics, 2004 IEEE International Conference on. IEEE*, (2004), 4: 3730-3734
10. Nagi J, Mohammad A M, Yap K S, et al. Non-technical loss analysis for detection of electricity theft using support vector machines. *Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International. IEEE*, (2008): 907-912
11. Nagi J, Yap K S, Tiong S K, et al. Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system. *IEEE Transactions on power delivery*, (2011), 26(2): 1284-1285.