International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 05 - Issue 07 || July 2019 || PP. 06-10

Research on Non-technical Loss Detection Method Based on Smart Grid

Yingying Zhao[1],Xiangrong Zu[2]

1 (School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)
2 (School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

Abstract: With the popularity of smart meters in the power grid, the use of distribution network and user-side measurement data to achieve efficient and accurate detection of NTL has received extensive attention from the academic community and the industry.In this paper, different machine learning algorithms are used to study the anomaly detection methods.

Keywords: electricity usage; non-technical loss; anomaly detection; data driven; smart grid

1. INTRODUCTION

From the researcher's point of view, the main NTL detection methods are mainly divided into three aspects: based on system state, data-driven and game-based. Based on the system state method, the core idea is to use the contradiction between the state estimation of the distribution network and the user measurement data to detect NTL. Based on the data-driven method, the core idea is to directly derive the shape and power distribution of the user's power curve. The characteristics of the game are based on the game theory. The core idea is to analyze the corresponding game equilibrium according to the behavior of the attacker and the detector, so as to detect the difference of the power distribution between different types of users.In this paper, we mainly analyze the data-driven anomaly detection method.

2. RESEARCH STATUS

The main problem existing in the anomaly detection method based on data-driven anomaly data is that the stolen samples have little or no existence of a given client, so that the sample data of the training is not balanced; the accuracy of the anomaly detection is not high.

2.1 Based on the problem of unbalanced training data

In [11], for the problem of data imbalance between the two types of customers obtained, this paper uses the SVM classifier to weight the sample ratio and calculate the sample ratio of each category (ie, the total number of classifier samples/single sample).) to adjust the weight.

For the problem of imbalance in the training data set, although weighting can be used to balance the sample, in most cases the theft sample has little or no presence for a given customer, for this zero attack situation. P. Jokae et al. [3] proposed a collection of data based on benign samples into malicious samples.

2.2 Based on the problem of low accuracy of abnormal detection

In [11] , we proposed an SVM-based approach that uses customer load profile information and other attributes to expose anomalous behaviors that are known to be highly correlated with NTL activity. With the implementation of this new fraud detection method, the TNBD detection hit rate will be 3 % increased to 60%.

P. Jokae et al. [3] proposed a consumption-based energy theft detector (CPBETD) that uses new techniques to overcome the problems associated with existing classification-based ETDS. In CPBETD, the total consumption per neighborhood is measured by the transformer meter and compared to the total usage reported by the smart meter. If a non-technical loss (NTL) is detected at this level, the customer in the zone with the exception mode will be selected as the suspicious user. For each customer, the user's historical data and synthetic attack data sets are used to train multiple types of support vector machines (SVMs). Then use the classifier to determine if the new sample is normal or malicious. This method makes the accuracy of anomaly detection reach 95%.

Q. Zhang et al. [4] proposed a machine learning-based approach called semi-supervised Gaussian mixture model (S2G2M2). It combines the user model of benign data generation with human prior knowledge to control the intensity of detection. Test indicators can be used to check user behavior at a frequency of 1 test/day without field inspection from the database side. In addition, S2G2M2 only needs historical benign data to identify whether current customer behavior is following normal mode.

Zheng, Zibin et al. [5] proposed a wide and deep convolutional neural network (CNN) model to learn power consumption data and identify power thieves. Our wide and deep CNN model consists of a wide component with a fully connected neural network layer and deep CNN components with multiple convolution layers, a pooled layer and a fully connected layer. In essence, wide components can learn global knowledge, while deep CNN components can capture the periodicity of power consumption data. The model integrates the advantages of wide components and deep CNN components to achieve good performance in power theft detection, which enables anomaly detection accuracy of 94%.

3. RESEARCH PROCESS

3.1 Data preprocessing

For the case of data loss and negative data, this paper uses the method of mean filling to process abnormal data. The missing phenomenon of data can be divided into two types: for the datamissing at a certain moment, we use the average value of the abnormal data point up and down timeas the abnormal point of the filling data;; the continuous time is missing, and the mean value of the dataat that time before and after day is used as the abnormal point of filling data.

After the exception point processing is completed, the malicious sample is constructed. A method for collecting malicious data samples based on benign samples using the literature [3]. There are mainly six malicious samples generated. The composition rules are as follows:

1.$h_1(x_t) = \alpha x_t$,α=random(0.1,0.8)(3-1)

2.$h_2(x_t) = \beta_t x_t$(3-2)

$$\beta_t = \begin{cases} 0 & start - time < t < end - time \\ 1 & else \end{cases}$$

start – time = random(1,23-minOFFTime)

duration = random(minOFFTime,24)

end-time = start-time+duration

3.$h_3(x_t) = \gamma_t x_t$, $\gamma_t = random(0.1,0.8)$(3-3)

4.$h_4(x_t) = \gamma_t mean(x), \gamma_t = random(0.1,0.8)$(3-4)

5.$h_5(x_t) = mean(x)$(3-5)

6.$h_6(x_t) = x_{24-t}$(3-6)

Where $h_1(.)$ all samples multiplies the same randomly selected value; $h_2(.)$sets its measured value to zero for a period of time; $h_3(.)$all samples multiplies a different random number; $h_4(.)$ send daily average multiplied by a random number; $h_5(.)$sendsdaily average; $h_6(.)$reverses the transmitted data, it can send low load when high price and send high load when low price number.

3.2 Method of choice

The seven classification models are mainly used to analyze the abnormality of the constructed data set.Namely: support vector machine (SVM), decision tree, logistic regression (LR), random forest (RF), gradient boosting algorithm ,XGBoost Algorithm, and lightGBM algorithm.

Theevaluation method is AUC and execution time.AUC is a performance indicator that measures the pros and cons of classification. It is defined as the area enclosed by the ROC curve and the coordinate axis. Obviously, the value of this area will not be greater than 1. Since the ROC curve is generally above the line y=x, the AUC value ranges between 0.5 and 1.The criteria for judging the classifier from AUC are as follows:

- AUC = 1, is the perfect classifier.
- AUC = [0.85, 0.95], works well
- AUC = [0.7, 0.85], the effect is average
- AUC = [0.5, 0.7], the effect is lower, but it is very good for predicting stocks.
- AUC = 0.5, followed by machine guessing (eg, lost copper), the model has no predictive value.
- AUC < 0.5, which is worse than random guessing; but as long as it is always anti-predictive, it is better than random guessing.

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Data

The data set was tested for one year's load data of a corporate user in the United States in 2012. The sampling interval was one hour (p=24). In order to detect the advantages and disadvantages of the classifier, we chose different machine learning methods for experiments.

4.2 Experimental results

The experimental summary diagram is shown in Table 4-1. The statistics are mainly from the AUC and the execution time. Fig. 1~Fig.7 show the experimental results of each category.

Tabel 4.1 Experimental results

Algorithm	AUC	execution time (s)
Decision Tree	0.89	0.025
SVM	0.694	1.87
LR	0.905	5.83
RF	0.979	0.04
Gradient Boosting	0.986	1.94
Xgboost	0.984	1.91
Lightgbm	0.988	2.15s

As can be seen from the table, in terms of execution speed, the decision tree and RF are executed faster; in terms of performance, Lightgbm and Gradient Boosting perform better.

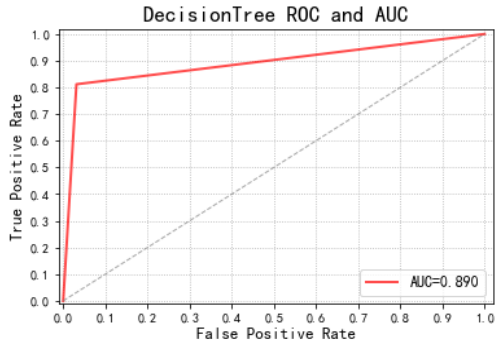

FIG.1 The result of Decision Tree

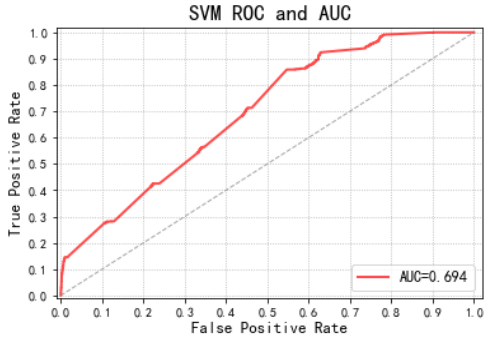

FIG.2 The result of SVM

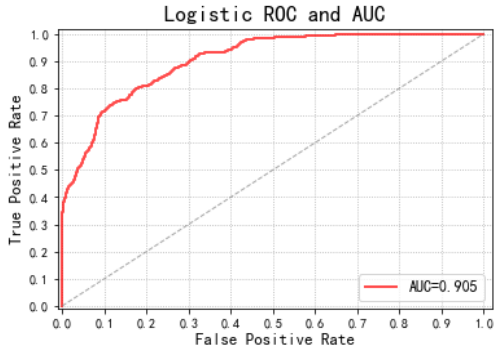

FIG.3 The result of LR

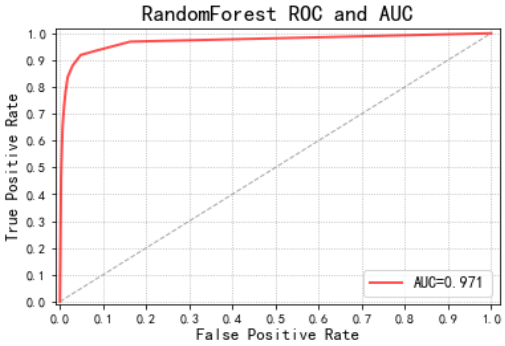

FIG.4 The result of RF

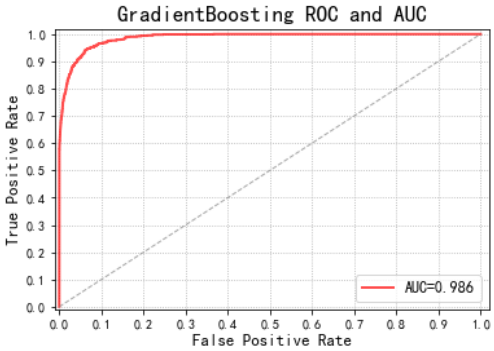

FIG.5 The result of Gradient Boosting

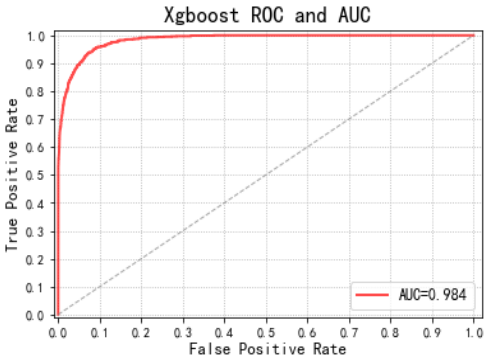

FIG.6The result of Xgboost

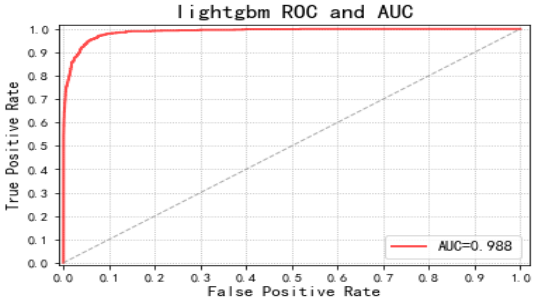

FIG.7 The result of lighgbm

5. CONCLUSION

In this paper, we mainly study the detection methods of non-technical losses in smart grids, and use different machine learning methods to conduct comparative experiments. By comparison, we found that the decision tree and RF execution time are similar, but the performance of RF is better than the decision tree; Gradient Boosting, Xgboost, lightgbm, and svm are similar in execution speed, but in comparison, Lightgbm has better performance and AUC reaches 0.988. Therefore, if we pay more attention to the accuracy rate, we can choose Lightgbm; if we pay more attention to the execution speed, we can choose RF.

The next step is to increase the data set. Comparing the performance comparison of different machine learning algorithms in the backgroundof big data

REFERENCES

[1]. JIANG, LU Rongxing, WANGYe, et al. Energy-theft detection issues for advanced metering infrastructure in smart grid[J].Tsinghua Since and Technology,2014,19(2):105-120.

[2]. FRAGKIOUDAKI A,CRUZ-ROMERO P,GOMEZEXPOSITO A, et al. Detection of non-technical losses in smart distribution networks: a review[M].New York, USA: Springer International Publishing,2016:43-54.

[3]. P. Jokar, N. Arianpoo and V. C. M. Leung, "Electricity Theft Detection in AMI Using Customers' Consumption Patterns," in IEEE Transactions on Smart Grid, vol. 7, no. 1, pp. 216-226, Jan. 2016.

[4]. Q. Zhang, M. Zhang, T. Chen, J. Fan, Z. Yang and G. Li, "Electricity Theft Detection Using Generative Models," 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), Volos, 2018, pp. 270-274.

[5]. Zheng, Zibin,;Yang, Yatao; Niu, Xiangdong; Dai, Hong-Ning; Zhou, Yuren;" Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids" in IEEE Transactions on Industrial Informatics, v 14, n 4, p 1606-1615, April 2018

[6]. McLaughlin, S.; Holbert, B.; Zonouz, S.; Berthier, R. "AMIDS: a multi-sensor energy theft detection framework for advanced metering infrastructures" in 2012 IEEE Third International Conference on Smart Grid Communications (Smart Grid Comm), p 354-9, 2012

[7]. CHEN Wei. Research on preventing electricity-stolen technology of intelligent walt-hour meter [D]. Beijing: North China Electirc Power Univeristy, 2015.

[8]. ZHOU Li, ZHAO Lujun, Gao Weiguo. Application of sparse coding in detection for abnormal electricity consumption behaviors [J].Power System Technology, 2015, 39(11): 3182-3188.

[9]. XU Gang, TAN Yuanpeng, Dai Tenghui. Sparse random forest based abnormal behavior pattern detection of electric power user side[J].Power System Technology,2017,41(6):1964-1973

[10]. SHEN Haitao, QIN Jingya, CHEN Hao , etal. Anomaly detection and category of electical and category of electrical utilization data [J].Power & Energy, 2016, 37(1): 17-22.

[11]. NAGI J,YAP K S,TIONG S K, etal. Nontechnical loss detection for metered customers in power utility using support vector machines [J].IEEE Transaction on Power Delivery, 2010, 25(2): 1162-1171.

[12]. JANETZKO H,STOFFEL F , MITTELSTAEDT S A. Anomaly detection for visual analytics of power consumption data[J].Computers & Graphics,2014,38(1):27-37

[13]. LIU Xiufeng, NIELSEN P S. Regression- based online anomaly detection for smart grid data[EB/OL].[2017-06-18].https://arxiv.org/pdf/1606.05781.pdf