Identification of Power Quality Disturbance Sources Using Gradient Boosting Decision Tree

Cheng Pan
Institute of Automation, Chinese Academy of Sciences
University of Chinese Academy of Sciences
Beijing 100190, P. R. China
pancheng2014@ia.ac.cn

Jie Tan*
Institute of Automation, Chinese Academy of Sciences
Chinese Academy of Sciences
Beijing 100190, P. R. China
tan.jie@tom.com

Dandan Feng
State Key Laboratory of Advanced Power Transmission Technology
Global Energy Interconnection Research Institute Co. Ltd
Beijing 102209, P. R. China
nobody.fd@163.com

***Abstract*—This paper proposed a new method based on statistical feature extraction and gradient boosting decision tree (GBDT) to recognize the power quality disturbance sources. Statistical calculation is adopted to extract the features of power quality disturbance sources, which has the advantage of small calculation. GBDT is proposed to apply in the recognition of power quality disturbance sources. First, according to the inherent characteristics of high-speed railway, ordinary railway, wind farm and photovoltaic plant, the proposed method uses statistical calculations to extract features which are the input of GBDT. Then, GBDT is applied to classify the power quality disturbance sources. Experiment results show that the proposed method can classify power quality disturbance sources accurately. Compared with other classification methods, GBDT has better recognition performance.**

***Index Terms*—gradient boosting decision tree, power quality, identification, disturbance sources**

I. Introduction

With the diversification and large-scale development of power quality disturbance sources, power quality problems such as harmonics, negative sequence and flicker caused by various disturbance sources, such as electrified railways, new energy generation, DC converter stations and non-linear load, are becoming more and more serious and complex. Identifying the disturbance sources of power quality is the basis and prerequisite for the management and improvement of power quality.

Generally, based on power quality monitoring data, we use statistical analysis methods, signal processing techniques, and data mining methods to study the identification of power quality disturbance sources. Power quality disturbance sources identification mainly includes two stages, namely feature extraction and classification. In the feature extraction stage, wavelet transform [1], S transform [2], Hilbert-Huang transform [3], statistical calculation [4] are commonly used. In the classification stage, approaches such as artificial neural networks (ANN) [5], fuzzy classification [6], expert system [7] and probabilistic neural network (PNN) [8] are among the methods used for classification. Classification of disturbances based on K-L transform and support vector machine (SVM) is presented in [9]. In [8], S transform and PNN was used to identify the power quality disturbance. The authors used S transform to extract features, and then classified the disturbance with PNN optimized by genetic algorithm (GA). Besides, In [10], It used wavelet multi-resolution analysis technology to extract disturbance energy features, and then combines the optimized wavelet neural network to identify the power quality disturbance. Reference [11] proposed an algorithm based on Stockwell's transform and ANN-based classifier and a rule-based decision tree for the recognition of single stage and multiple power quality disturbances.

At present, domestic and foreign scholars main focus on how to identify power quality types such as voltage swell, voltage sag, etc. However, we have less research on the sources of power quality disturbance. And with the improvement of power grid automation, it is urgent to identify the source of power quality disturbance, and then assist in the operation and maintenance decision of the power grid.

In this paper, we combine the statistical features of power quality monitoring data with GBDT to identify power quality disturbance sources. GBDT is an boosting method based on regression tree. It is considered as one of the best performance methods in statistical learning. Based on the power quality monitoring data of various kinds of disturbance sources, we select the key measurement indexes such as three-phase active power, negative sequence unbalanced current and other key measurement indicators according to prior knowledge, and calculate their statistical features, and then use them as the input of the classifier to identify high-speed railway, ordinary railway, wind farm, photovoltaic plant. The experimental results show that the proposed method can effectively and accurately distinguish the source of power quality disturbance.

II. Gradient Boosting Decision Tree

Boosting Tree is an boosting method based on classification tree or regression tree, which is considered as one of the best performance methods in statistical learning. It uses the additive model and forward stagewise algorithm to realize the optimization process of the algorithm. When the loss function is a square loss and exponential loss function, the optimization of each step is relatively simple. However, for a general loss function, it is often not easy to optimization each step. To solve this problem, Frediman [12] proposes a gradient boosting decision tree (GBDT), which uses the negative gradient of loss function in the current stage as the approximate value of residual to fit a regression tree. The negative gradient of loss function is given as

$$-\left[\frac{\partial L(y, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)} \tag{1}$$

which x_i is the i-th sample, $f_{m-1}(x)$ is the model obtained from $m-1$ stage. GBDT uses regression tree as the basic classifier and can be used to solve regression problems, binary classificaton problems, and multi-classification problems. In this paper, we use GBDT to identify power quality disturbance sources.

In the multi-classification problems, the loss function of GBDT is

$$L(\{y_k, F_k(x)\}_1^K) = -\sum_{k=1}^{K} y_k \log p_k(x) \tag{2}$$

where $y_k = 1(class = k) \in \{0, 1\}$, K is the total number of class, and $p_k(x) = P(y_k = 1|x)$. With the idea of softmax, $p_k(x)$ is written as,

$$p_k(x) = exp(F_k(x)) / \sum_{l=1}^{k} exp(F_l(x)) \tag{3}$$

where $F_k(x)$ is the function value of sample x at the current stage. Since GBDT fits a regression tree using the negative gradient values of the current stage, we substitute (3) into (2) and according to (1), take first derivatives one has

$$\begin{aligned}\hat{y}_{ik} &= -\left[\frac{\partial L(\{y_{il}, F_l(x_i)\}_{l=1}^K)}{\partial F_k(x_i)}\right]_{\{F_l(x)=f_{l,m-1}(x)\}_1^K} \\ &= y_{ik} - p_{k,m-1}(x_i)\end{aligned} \tag{4}$$

where m is the the number of iterations, $p_{k,m-1}(x_i)$ can be obtained according to $F_{k,m-1}(x)$ and (4). In each iteration, we use K-trees to fit the the negative gradient of loss function in the current stage for K-class respectively. Each tree has J-terminal nodes corresponding to the region $\{R_{jkm}\}$ in feature space. According to the properties of regression trees, we can calculate the output value γ_{jkm} on the region $\{R_{jkm}\}$. It is given as,

$$\gamma_{jkm} = \frac{K-1}{K} \frac{\sum_{x_i \in R_{jkm}} \hat{y}_{ik}}{\sum_{x_i \in R_{jkm}} |\hat{y}_{ik}|(1 - |\hat{y}_{ik}|)} \tag{5}$$

So we can get the function estimation F in an additive form:

$$F_{k,m}(x) = F_{k,m-1}(x) + \sum_{j=1}^{J} \gamma_{jkm} \mathbf{1}(x \in R_{jkm}) \tag{6}$$

where $F_{k,m-1}(x)$ is the function estimation at the $m-1$-th stage, $\sum_{j=1}^{J} \gamma_{jkm} \mathbf{1}(x \in R_{jkm})$ is the function estimation of residuals at the $m-1$-th stage, J is the number of leaf nodes in the regression tree.

After model training, we can use the final estimates $\{F_{kM}(x)\}_1^2$ to classification

$$\hat{k}(x) = \arg \min_{1 \le k \le K} \sum_{k'=1}^{K} c(k, k') p_{k'M}(x) \tag{7}$$

where $c(k, k')$ is the loss associated with predicting the k-th class when the truth is k'.

III. Recognition Methods

The proposed power quality disturbance sources recognition method is shown in Fig. 1. It can be divided into two stages. The first stage is feature extraction. We calculate the statistical characteristics of the daily measurement data, and use it as the features of the power quality disturbance sources. The second stage is classification that involves determination of power quality disturbance sources by GBDT. In this paper, we will identify four types of disturbance sources, which are high-speed railway, ordinary railway, wind farm and photovoltaic plant.

A. Feature Extraction Stage

Feature extraction is the key to pattern recognition which directly affects the performance of intelligent systems. The power quality disturbance sources monitoring data consist of active power, reactive power, harmonic current(2-50 times), three-phase unbalance and other measurement indicators. According to the prior knowledge, we separately extract the statitical features of three-phase active power, negative sequence unbalanced current, three-phase third harmonic and three-phase fifth harmonic. Fig. 2 shows the daily active power timing diagram of high-speed railway, ordinary railway, wind farm and photovoltaic plant. Photovoltaic plants are affected by sunlight, have typical time characteristics and are susceptible to cloud effects. Wind farms have random fluctuations. The high-speed railway and the ordinary railway are impulse loads, in which the high-speed railway has a power of zero in the early morning hours and its active power value is high. It can also be observed that the four types of disturbance sources have significant differences in mathematical expectations, so we will use it for classification. At the same time, compared with other feature extraction methods, expectation feature extraction has lower computation. The mathematical expectation feature is given by

$$\bar{x} = \frac{1}{N} \sum_{k=0}^{N-1} x_k \tag{8}$$

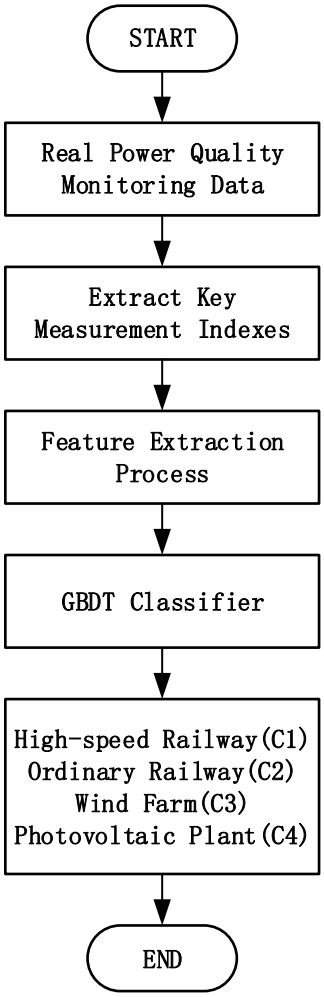

Fig. 1. The proposed power quality disturbance sources recognition system

where x_k represents the value of the measurement index at time k, N represents the numbers of data which obtained from monitor in one day.

Typical values of the four types of disturbance sources on the 10-dimensional measurement index are given in Table I. It can be seen that different power quality disturbance sources have obvious differences on some measurement indexes.

B. Recognition Stage

After the feature extraction process, we will get the statistical features of power quality disturbance sources. In the Recognition stage, we use statistical features as input to GBDT. The principle of GBDT is given in Section II. The output of GBDT is the types of power quality of disturbance sources, where C1 represents high-speed railway, C2 represents ordinary railway, C3 represents wind farm and C4 represents photovoltaic plant.

IV. Experimental results

We use power quality monitoring data from a certain region of China and divide it into training dataset and test dataset.

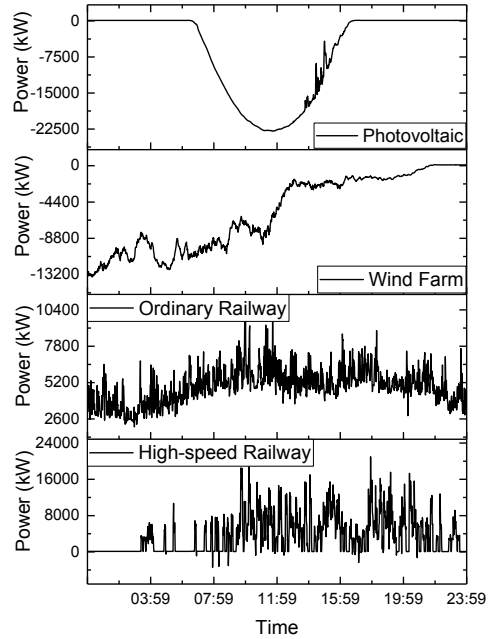

Fig. 2. active power of Four types of disturbance sources

TABLE I
Typical features of disturbance sources

	High-speed Railway	*Ordinary Railway*	*Wind Farm*	*Photovoltaic*
active power-A	3486	4893	-5655	-1280
active power-B	2119	6389	-5694	-1280
active power-C	1183	5117	-5689	-1288
Negative Sequence Unbalance	42.75	18.03	0.49	2.27
three harmonic current-A	0.54	1.16	0.64	0.16
three harmonic current-B	0.48	1.77	0.83	0.17
three harmonic current-C	0.37	1.53	0.49	0.07
five harmonic current-A	0.64	0.60	0.77	0.37
five harmonic current-B	0.44	1.22	0.76	0.32
five harmonic current-C	0.26	0.88	0.88	0.31

The training dataset is used to train the classifier, and the test dataset tests the classifier performance. The classification results of the proposed recognition system are shown in Table II. The diagonal elements represent correct classification results, and the remaining elements are misclassified results. The classification results show that GBDT can recognize high-speed railway(C1), ordinary railway(C2) and photovoltaic plant(C4) completely. For wind farm(C3), only one sample is misclassified. And the overall success rate is 98.68%.

Classification results obtained by different classifiers are

TABLE II
CLASSIFICATION RESULT BASED ON THE PROPOSED METHOD

True class	C1	C2	C3	C4	Accuracy(%)
C1	19	0	0	0	100
C2	0	19	0	0	100
C3	0	0	18	1	94.73
C4	0	0	0	19	100
Overall success rate (%): 98.68					

given in Table III. We train the classifiers SVM, KNN and GBDT using the same training dataset. These classifiers are then tested using the same batch of test dataset, and the overall classification accuracy of SVM, KNN and GBDT is 94.74%, 96.05% and 98.68%, respectively. From the Table III, we can see that all three classifiers can identify high-speed railway(C1) and ordinary railway (C2), SVM and KNN have misclassification for wind farms(C3) and photovoltaic plants (C4), GBDT only has misclassification for wind farms(C3). So, it is clear that GBDT gives the best classification results for this case.

TABLE III
CLASSIFICATION RESULTS OBTAINED BY DIFFERENT CLASSIFIERS

	Accuracy (%)				Overall
	C1	C2	C3	C4	Accuracy (%)
SVM (Kernal:Linear)	100	100	94.73	84.21	94.74
KNN ($K = 3$)	100	100	94.73	89.47	96.05
GBDT	100	100	94.73	100	98.68

V. CONCLUSIONS

In this work, the method of statistical feature extraction combined with GBDT, to classify various power quality disturbance sources, is proposed. Compared with feature extraction methods such as wavelet transform and HHT transform, the statistical feature extraction used in this paper is simpler and more effective. GBDT has high classification accuracy and can effectively identify power quality disturbance sources. Compared with the SVM and KNN classification methods, the GBDT classification has better recognition performance. Experimental results show that the proposed method can effectively identify power quality disturbance sources.

ACKNOWLEDGMENT

This work was supported by the National Nature Science Foundation of China under Grant U1701262 and the Intelligent Manufacturing New Model Application Project of the Ministry of Industry and Information Technology of the People's Republic of China .

REFERENCES

[1] Murat Uyar, Selcuk Yildirim, Muhsin Tunay Gencoglu, "An effective wavelet-based feature extraction method for classification of power quality disturbance signals," Electric Power Systems Research, vol. 78, pp. 1747-1755, 2008.

[2] P.K. Dash, B.K. Panigrahi, G. Panda, "Power quality analysis using S-transform," IEEE Transactions on Power Delivery, vol. 18, pp. 406-411, April 2003.

[3] TY Li, Y Zhao, N Li, G Feng, HH Gao, "New method for power quality detection based on HHT," proceedings of the Csee, vol. 17, pp. 52-56, 2005.

[4] H.Dehghani, B.Vahidi, R.A.Naghizadeh, S.H.Hosseinian, "Power quality disturbance classification using a statistical and wavelet-based Hidden Markov Model with DempsterShafer algorithm," International Journal of Electrical Power & Energy Systems, vol. 47, pp. 368-377, May 2013.

[5] D. Srinivasan, W.S. Ng, A.C. Liew, "Neural-network-based signature recognition for harmonic source identification," IEEE Transactions on Power Delivery, vol. 21, pp. 398-405, Jan 2006.

[6] Birendra Biswal, P. K. Dash, B. K. Panigrahi, "Power Quality Disturbance Classification Using Fuzzy C-Means Algorithm and Adaptive Particle Swarm Optimization," IEEE Transactions on Industrial Electronics, vol. 56, pp. 212-220, July 2008.

[7] Mamun Bin Ibne Reaz, Florence Choong, Mohd Shahiman Sulaiman, Faisal Mohd-Yasin, Masaru Kamada, "Expert System for Power Quality Disturbance Classifier." IEEE Transactions on Power Delivery, vol. 22, pp. 1979-1988, July 2007.

[8] Dr. Sivanthi, Thiagarajar, "Research on disturbances detection and classification for power quality," Journal of The Institution of Engineers (India), vol. 96(3), pp. 283-295, 2015.

[9] LUO Dian-sheng, HE Hong-ying, YAO Jian-gang, "Research of classification method of short duration power quality disturbance," Computer Engineering & Applications, vol. 44, pp. 238-241, 2008.

[10] Han Fuchun, Sun Jie, LingHu jinjun, Hao caiyun, "Power quality disturbances classification based on improved PSO-WNN," Electrical Automation, vol. 33, pp. 59-60+85, 2011.

[11] Raj Kumar, Bhim Singh, D. T. Shahani, Ambrish Chandra, Kamal Al-Haddad, "Recognition of Power-Quality Disturbances Using S-Transform-Based ANN Classifier and Rule-Based Decision Tree," IEEE Transactions on Industry Applications, vol. 51, pp. 1249-1258, 2014.

[12] Jerome H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," The Annals of Statistics, vol. 29, pp. 1189-1232, 2001.