

# Research on Emergency Judgement of Low Voltage Area Based on Classification Technique

Xiaoqiang Zhong

State Grid Fujian Electric Power Co, Ltd, Fuzhou 350003, China

xqzhong @163.com

**Keywords:** low voltage, emergency, classification technique, area.

**Abstract.** As an important evaluation index of power quality, voltage is the basic condition for guaranteeing power supply service. With the development of industry and the improvement of the living standards, the demand of electricity increases. Low voltage has become a major problem affecting the work, life and entertainment of the people. There are many causes for the low voltage. According to the sudden degree of maintenance and overhaul, such causes can be divided into sudden ones and non-sudden ones. It is of great significance to the daily maintenance of the power grid if the sudden cause can be quickly determined and the maintenance personnel can be sent for site maintenance. In this paper, the classification technique of machine learning is used to analyze the voltage related data in the power grid, and then a method for determining the cause of low voltage in distribution network areas is put forward.

## 1. Introduction

Strengthening the power demand-side management is an important channel and effective way to build a resource-conserving society and promote the sustainable development of economy and society. Research on low voltage is a common problem of power system anomalies. The cause of the low voltage can be summarized as follows <sup>[1]</sup>: (1) large power supply radius; (2) small sectional area of wires; (3) aging of equipment; (4) line overload; (5) unreasonable switch position of tap changer contact in the distribution transformer is; (6) unbalanced low voltage three-phase load; (7) capacity assignment of reactive power compensation; (8) over capacity by customers. According to the sudden degree of maintenance and overhaul, the causes can be divided into two categories. One is the sudden cause. Normally, the cost of solving such problems is small. For example, if it's due to the aging equipment of a line, it's necessary to replace the equipment. The other is the non-sudden cause. The cost of solving such problems is greater. For example, due to the large power supply radius, it is necessary to re-define the radius of power supply area. And if it's caused by the small sectional area of the wire, it is necessary to replace the wire.

At present, the main research directions of low voltage are: (1) genetic analysis and remediation research <sup>[2-5]</sup>. For example, paper <sup>[2]</sup> further analyzed the internal relationship between the line load moment and the cross-section of wire, breaking the bottleneck of the traditional method which only calculated the same type of wire. Also it normalized the overhead lines of different diameters to determine the optimal segment of replacing the wire diameter. (2) Low voltage prediction. For example, study <sup>[6]</sup> collected nine characteristic parameters reflecting the voltage quality of low-voltage distribution network, and trained a low-voltage court prediction model based on self-organizing competitive neural networks to achieve the automatic clustering of three types of zones with normal low voltage, low-voltage risk, as well as severe low-voltage risk.

As for maintenance and overhaul of the power grid, it is of great significance to the daily maintenance and overhaul of the power grid if the sudden cause can be quickly determined and the maintenance personnel can be sent for site maintenance.

## 2. Technology Background

### 2.1 Decision Tree [7].

The decision tree is an instance-based inductive learning algorithm. It deduces the classification rules of the decision tree representation from a set of unordered and irregular tuples. It uses the top-down recursive way to compare the attribute values within the decision tree's internal nodes, and splits from the nodes in branches according to different attribute values. The leaf node is the class to learn classification. A path from the root to the leaf node corresponds to a conjunction rule, and the entire decision tree corresponds to a set of disjunctive expression rules.

### 2.2 Naive Bayes [8].

Let each data sample describe  $n$  attribute values with an  $n$ -dimensional eigenvector, i.e.  $X = \{x_1, x_2, \dots, x_n\}$ . Suppose there are  $m$  classes, denoted by  $C_1, C_2, \dots, C_m$ , respectively. Set an unknown data sample  $X$  (i.e., no class label), if the naive Bayesian classification assigns the unknown sample  $X$  to class  $C_i$ , it must be

According to the Bayes theorem, since  $P(X)$  is constant for all classes, the maximized posterior probability  $P(C_i | X)$  can be transformed into the maximized prior probabilities  $P(X | C_i) P(C_i)$ . If the training data set has many attributes and tuples, the cost of calculating  $P(X | C_i)$  may be very large. To this end, it is usually assumed that the values of each attribute are independent of each other, so that the prior probabilities  $P(x_1 | C_i), P(x_2 | C_i), \dots, P(x_n | C_i)$  can be obtained from the training data set.

According to this method, for a sample  $X$  of unknown class, we can first calculate the probability  $(P(X | C_i) P(C_i))$  of  $X$  belonging to each class  $C_i$ , and then select the category with the highest probability as its category.

### 2.3 SVM [9].

Support vector machine (SVM) is to improve the generalization ability of the learning machine by seeking the minimum of structural risk, so as to realize the minimization of empirical risk and confidence interval, thereby achieving the purpose of obtaining good statistical law in the case of less statistical sample size. In popular terms, it is a kind of second class classification model. Its basic model is defined as the largest linear classifier in the feature space, that is, the learning strategy of support vector machine is the maximized interval, which can be eventually transformed into solving a convex second programming problem.

## 3. Model introduction

### 3.1 Data Format.

The data related to the situation of the power grid in reactive courts are: A-phase voltage, B-phase voltage, C-phase voltage, A-phase current, B-phase current, C-phase current, frequency, total active power, A-phase active power, B-phase active power, C-phase active power, total reactive power, A-phase reactive power, B-phase reactive power, C-phase reactive power, total power factor, A-phase power factor, B-phase power factor, C-phase power factor, positive active power, positive reactive power and so on.

The data of the nature of the reaction zone are as follows: capacity (rated capacity of distribution transformer), power supply in the area, total quantity of electricity of the same day, total number of users in the area, electricity consumption by residents for living of the same day, electricity consumption for non-resident lighting of the same day, electricity consumption for agricultural production of the same day, commercial electricity consumption on the same day, electricity consumption for non-general industry on the same day and so on.

Because the low voltage is not anomaly at a single point, it is on a daily basis to determine whether low voltage anomalies occur in the area. Therefore, all of the above features and voltages are combined in the following chart, where each row (except "sudden or not") represents the different characteristics of the same date, and each column (except "sudden or not") represents the same characteristics of the different dates. The "sudden or not" value is 0 (not sudden) or 1 (sudden).

Table 1. Format List of data

Feature_1	Feature_2	Feature_n-1	Feature_n	sudden or not
T11	T12	T1(n-1)	T1n	V1
T21	T22	T2(n-1)	T2n	V2
T(k-1)1	T(k-1)2	T(k-1) (n-1)	T(k-1)n	Vk-1
Tk1	Tk2	Tk(n-1)	Tkn	Vk

**3.2 Choosing Useful Feature.**

There are many related data of voltage in the area. If the classified counting is used, it will be time-consuming, yet the efficiency is low. Therefore, it is necessary to pre-process the data to determine the characteristic data which is strongly correlated with the voltage in the area. The steps are as follows:

- R1 select feature 1 to do regression analysis on V1, go to R2;
- R2 whether the feature is selected, go to R5 if “yes”, otherwise go to R3
- R3 introduce a new feature, carry out regression analysis, test the partial regression of each characteristic, and go to R3;
- R4 discard indistinctive features, then go to R2;
- R5 selection finished.

**3.3 Setting up Classifier.**

Randomly select 80% of the filtered data as a training set, and the remaining 20% as a test set. Apply the decision tree algorithm, the naive Bayesian algorithm and the support vector machine algorithm. Build a classifier according to the procedures in the following figure.

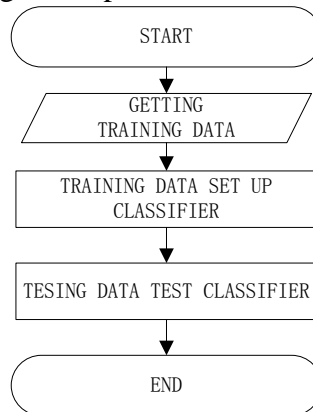


Fig. 1 Flow chart of building classifier

**3.4 Setting up Strong Classifier (SC).**

Three classifiers created in 2.3 are optimized according to the method shown in Figure 2 to construct a strong classifier. Among them, classifiers 1, 2 and 3 represent the classifiers constructed by decision tree algorithm, naive Bayesian algorithm and support vector machine algorithm respectively.

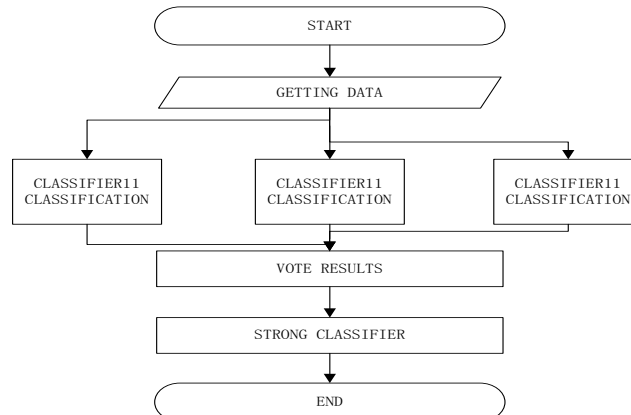


Fig. 2 Flow chart of building strong classifier

#### 4. Practice

If you follow the “checklist” your paper will conform to the requirements of the publisher and facilitate a problem-free publication process.

The low-voltage maintenance records of a region in Fujian province include low-voltage causes. We divide them into sudden ones and non-sudden ones, conduct analysis on 1000 low-voltage records in Fujian in April, 2016, and pre-process the data according to the format shown in Table 1. After the selection of useful features, the processing results of the classifier are shown in Table 2.

Table 2. Comparison chart of classifier

Algorithm	Decision Tree	Naïve Bayes	SVM	SC
Accuracy	57.3%	54.5%	58.1%	62.4%

From the above results, we cannot analyze whether the low voltage is due to sudden causes. When we return to the low-voltage maintenance records, it is found that there is repeated or wrong judgment of the fault reasons in the record. For example, for the same area, the low voltage situation, as well as fault diagnosis occurred many times, that is, the cause of the trouble was not determined in one troubleshooting, but the record was increased, which played a negative role for data learning. After the data is re-cleaned, the results are shown in Table 3; the contrast between before and after the calibration is shown in Figure 3:

Table 3. Comparison chart of classifier

Algorithm	Decision Tree	Naïve Bayes	SVM	SC
Accuracy	69.1%	75.4%	77.8%	82.3%

It can be seen from Table 3 and Figure 3 that, the data after the re-cleaning are characterized by good regularity, the judgment results are remarkably improved, and strong classifier is obviously better than any single one of the three algorithms.



Fig. 3 Comparison chart of data cleaning

#### 5. Summary

In this paper, aiming at the problem of low voltage prevalent in the network system, the causes and the research status of low voltage are introduced first. And as for the need of sudden maintenance, it analyzes the low-voltage data with the classification technique, which has greater probability to determine the cause of the failure, and provide decision support for maintenance work. With the accumulation of data and the standardization of maintenance records, the data quality will be greatly improved, and the accuracy of the method proposed in this paper will be higher.

In addition, the proposed method to determine the sudden failure is dynamic. In other words, with the adding of the data and the change of the results, the generated classifier will change accordingly. At the same time, when the result is inaccurate, the correction of the data will also affect the classifier.

## References

- [1]. Barros J, Perez E. Measurement and analysis of voltage events in a low-voltage distribution network [C]// Electrotechnical Conference, 2004. Melecon 2004. Proceedings of the, IEEE Mediterranean. IEEE Xplore, 2004:1083-1086 Vol.3.
- [2]. Yang G, He T. Circuit loss analysis and application development of three-phase unbalanced load of low voltage distribution network [C]// China International Conference on Electricity Distribution. IEEE, 2011:1-4.
- [3]. Feng N I, Jian-Ming Y U, Pan Z M. Low-Voltage Distribution Network Theoretical Energy Loss Calculation Based on Real-Time Unbalance in Three Phases [J]. Journal of Xian University of Technology, 2008.
- [4]. Zhang T, Zhang B. Power supply reliability evaluation for consumers in low voltage distribution network based on probability statistics [J]. Power System Technology, 2004, 28 (17):81-84.
- [5]. Stewart R A, Bennett C J, Lu J W. Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system [J]. Energy, 2014, 67 (4):200-212.
- [6]. Pan X, Tang A, Yang H, et al. Reliability Evaluation and Prediction of Low Voltage Distribution Network [C]// International Conference on Electrical, Computer Engineering and Electronics. 2015.
- [7]. Quinlan J R. Induction on decision tree [J]. Machine Learning, 1986, 1 (1):81-106.
- [8]. Rish I. An empirical study of the naive Bayes classifier [J]. Journal of Universal Computer Science, 2001, 1 (2):127.
- [9]. Joachims T. Making large-scale SVM learning practical [J]. Technical Reports, 1999, 8 (3):499-526.