

Application of Fuzzy Support Vector Machines to Partial Discharge Pattern Recognition

Fu Lihua, Lei Wanzhong, Zhang Xiaomei
Henan Institute of Engineering
Zhengzhou 451191

Abstract-In the field of pattern recognition, Support Vector Machines (SVMs) has great advantage than other traditional methods due to its simple structure, strong ability of generalization and good performance in recognizing. The feature of discharge is extracted using the 2D pattern chart and the Fuzzy Support Vector Machines (FSVMs) based on the affinity among samples is used to recognize the discharge models in this paper. Experimental results show that FSVMs has the better robust and classification performance than the SVMs.

Keywords-Partial Discharge (PD), Pattern Recognition, Fuzzy Support Vector Machines

I. INTRODUCTION

Partial Discharge is one of the main reason for the electrical aging of electric installations, so it becomes a most direct characterization of insulation failure. By on-line Partial Discharge monitoring, we can find partial fault of insulations timely and forecast the hidden accident. It is especially important to extract features and recognize discharge patterns because the level of electrical aging gets in touch with the discharge models closely. At present, neural network and fuzzy recognition are applied most widely in the pattern recognition of Partial Discharge [1],[2]. In this paper, fuzzy theory and SVM are used to recognize discharge models.

The rest of this paper is organized as follows. Section 2 gives a brief review of SVMs and FSVMs. The experimental results and discussions are presented in Section 4. Finally the conclusion is given in Section 4.

II. SVMs AND FSVMs

A. SVMs

The theory of SVM is a powerful methodology for solving pattern classification and regression estimation problems [3], [4]. While traditional statistical theory keeps to empirical risk minimization, SVM satisfies structural risk minimization based on statistical learning theory, who outperform conventional classifiers especially when the number of training data is small and there is no overlap between classes.

Given a set of labeled training data

$L = \{(x_i, y_i), i = 1, 2, \dots, l\}$, each training point $x_i \in R^n$ belongs to either of two classes and is given a label $y_i \in \{-1, 1\}$ for x_i . If these data are linearly separable, we wish to find the hyper plane

$$\omega \cdot x_i + b = 0 \quad (1)$$

Where ω is an l -dimensional vector, b is a scalar, and

$$y_i (\omega \cdot x_i + b) \geq 1 \quad i = 1, 2, \dots, l \quad (2)$$

The distance between the separating hyper plane and the training datum nearest to the hyper plane is called the margin. The hyper plane with the maximum margin is called the optimal separating hyper plane.

The data that satisfy the equality in (1) are called support vectors.

Then we can convert the optimal hyper plane problem into the following quadratic program (QP) problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & y_i (\omega \cdot x_i + b) \geq 1 \quad i = 1, 2, \dots, l \end{aligned} \quad (3)$$

If these data are linearly nonseparable, a set of non-negative slack variables ξ_i is introduced such that the following conditions are satisfied.

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} \quad & y_i (\omega \cdot x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad i = 1, 2, \dots, l \end{aligned} \quad (4)$$

Where C is a predefined positive constant, a smaller C imposes a less penalty on empirical errors. Tuning this parameter can make balance between margin maximization and classification violation.

If these data are non-linearly separable, the solution is to embed feature space R^n in higher-dimensional feature space R^N using a non-linearly function $\phi(x)$, and by working with linear classification in that space. The vector $\phi(x_i)$ in the feature space corresponds to vector x_i in the original space. To solve the QP problem, one needs to compute the scalar products of the form $\phi(x_i) \cdot \phi(x_j)$ and do not have to know the shape of $\phi(x_i)$. It is therefore convenient to introduce the kernel

function $K(x_i \cdot x_j) = \phi(x_i) \cdot \phi(x_j)$. By using the Lagrange multiplier method and kernel method, one can construct the QP problem as follows:

$$\begin{aligned} \min & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^l \alpha_i \\ & \sum_{i=1}^l \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \quad i = 1, 2, \dots, l \end{aligned} \quad (5)$$

Some commonly used kernel functions are polynomial, sigmoid and Gaussian function.

B. FSVMs

In many real word problems, there are cases that some training points are corrupted by noise. Moreover, some points in the training data are misplaced on the wrong side by accident. These points are all outliers; thus do not completely belong to one class, but with different memberships in the two classes. In this case, the traditional SVMs training algorithm will make the decision boundary to severely deviate from the optimal hyper plane because SVMs is very sensitive to outliers [5]. Solving this problem, FSVMs technique has been proposed [6]. FSVMs is developed on the theory of SVMs. In FSVMs, each sample is given a fuzzy membership which denotes the attitude of the corresponding point toward one class. The membership represents how important is the sample to the decision surface. The bigger the fuzzy membership, the corresponding point is treated more important; thus, different input points can make different contributions to the learning of decision surface.

Suppose the set of training data is

$$\{(x_i, y_i, \mu(x_i)), i = 1, 2, \dots, l\}$$

The QP is described as follows:

$$\begin{aligned} \min & \psi(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \mu(x_i) \xi_i \\ \text{s.t.} & y_i (\omega \cdot \phi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad i = 1, 2, \dots, l \end{aligned} \quad (6)$$

$\mu(x_i)$ is a membership function satisfying

$\eta \leq \mu(x_i) \leq 1$ with a sufficiently small positive constant η . Obfuscation the penalty parameter and using different penalty parameter, $C \cdot \mu(x_i)$ denotes the importance of x_i in FSVMs. The bigger the $C \cdot \mu(x_i)$, the smaller probability of x_i is classified wrong, the shorter distance between the separating hyper plane and samples. If x_i is outliers, $\mu(x_i)$ is very small, so is to $C \cdot \mu(x_i)$, so the contribution of this sample to FSVMs is decreased drastically, then the influence of outliers to the training is decreased accordingly. The FSVMs can eliminate influence

of outliers and the decisive action of the normal support vectors to the separating hyper plane is not weakened at the same time. According to above analysis, that choosing a proper fuzzy membership function is quite important to solve classification problem with FSVMs.

C. Membership Functions for FSVMs

At present, the membership functions building method mainly bases on the distance between the sample and its corresponding class center which measures the size of membership function [7]. There includes the membership function based on linear distance and the membership function based on S-function [8]. In these two methods, each sample is treated in a same way. That lead to the effective samples and the outliers is not divided accurately; i.e., they can't reflect the uncertainty of samples.

The membership function based on affinity is chosen in this paper [9]. In this method, the fuzzy membership is dependent on not only the relation between a sample and its corresponding class center, but also the relation between a sample and those non-corresponding samples. Compared with other FSVMs algorithms, this method can effectively distinguish between the valid samples and the outliers.

III. APPLICATION OF FUZZY SUPPORT VECTOR MACHINES TO PARTIAL DISCHARGE PATTERN RECOGNITION

At present, on the study of Partial Discharge pattern recognition, there is difficult in obtaining the test data on site, so, the research should be processed based on the laboratory result of nondestructive Partial Discharge under the condition of laboratory. The training data in this paper come from the standard discharge models samples data of Tsinghua university laboratory [12]. There are four models: tiny air-gap discharge, interior air-gap discharge, endings surface discharge and slots discharge. The features in this paper are acquired from computing the positive half period and negative half period subsidiary spectrogram of two-dimension spectrogram above four models.

A. Skewness

Skewness describes the measure of skewness that the subsidiary spectrogram distribution compares to the normal distribution. $S_k = 0$ denotes the shape of subsidiary spectrogram is lateral symmetry; $S_k > 0$ indicates the shape of subsidiary spectrogram deviate left and $S_k < 0$

denotes the shape of subsidiary spectrogram deviate right.

$$S_k = \sum_{i=1}^W p_i (\phi_i - \mu)^3 / \sigma^3 \quad (7)$$

Where W denotes the quantities of phase window in half period, ϕ_i denotes the phase of the phase window whose order is number i . p_i , μ and σ denote event probability, average value and variance in the number i

phase window when ϕ_i is stochastic variable, whose computation method as:

$$p_i = y_i / \sum_{i=1}^W y_i \quad (8)$$

Where y_i is y-axis of spectrogram which denotes apparent discharge magnitude or discharge repetition frequency.

$$\mu = \frac{\sum_{i=1}^W p_i \phi_i}{\sum_{i=1}^W p_i} \quad (9)$$

$$\sigma = \sqrt{\sum_{i=1}^W p_i (\phi_i - \mu)^2} \quad (10)$$

B. Kurtosis

Kurtosis describes the precipitous level that the subsidiary spectrogram contour compares to the normal distribution contour as

$$K_u = \frac{\sum_{i=1}^W (\phi_i - \mu)^4 \cdot p_i}{\sigma^4} - 3 \quad (11)$$

C. Cross-relation Factor (CC)

Cross-relation Factor can be expressed as

$$CC = \frac{\sum_{i=1}^W q_i^+ q_i^- - (\sum_{i=1}^W q_i^+ \sum_{i=1}^W q_i^-) / W}{\sqrt{[\sum_{i=1}^W (q_i^+)^2 - (\sum_{i=1}^W q_i^+)^2 / W][\sum_{i=1}^W (q_i^-)^2 - (\sum_{i=1}^W q_i^-)^2 / W]}} \quad (12)$$

Where q_i indicates the average discharge magnitude in the phase window the number i in which “+” and “-” denote positive half period and negative half period of $\phi - q$ spectrogram. W denotes the quantity of phase window which half period of spectrogram is divided equally. CC close to 1 means positive half period and negative half period of spectrogram contour is similar extremely; CC close to 0 means spectrogram contour is different very. According to above features, the feature vector of Partial Discharge is

$$V = (S_{kq}^+, K_{uq}^+, S_{kq}^-, K_{uq}^-, S_{kn}^+, K_{un}^+, S_{kn}^-, K_{un}^-, CC) \quad (13)$$

Partial Discharge pattern recognition is a multi-classification problem, traditional SVM originally formulated for two-classification, so there developed some methods to solve this problem. One is to construct several two-class classifiers, and then to assemble a multi-class classifier, such as one -against-one, one-against-all etc [10]. M-ary is used to solve the multi-classification problem is this paper[11]. Structuring $\lceil \log_2 4 \rceil = 2$ SVM subsidiary classifiers, samples data of interior air-gap discharge and endings slots discharge are labeled positive all and samples data of tiny air-gap discharge and endings surface discharge are labeled negative all to the first classifier; samples data of interior air-gap discharge and endings surface discharge

slots are labeled positive all and samples data of tiny air-gap discharge and slots discharge endings are labeled negative all second classifier. According the results of the two classifiers, the models of the testing samples can be obtained.

The whole data set includes 160 samples in which every sample dimension is 9 and 4 models. In order to overcome lack of testing samples, 5% white noise is added to these 160 testing samples factitiously. That another 160 modified testing samples are obtained.

The normal software package LibSVM is used in the experiment. All samples are divided into two part 50% used to train and 50% used to test. There chooses 40 sets of data regarded as training set to learn for SVMs first and the remaining 40 sets of data are regarded as testing set. There has not a uniform method how to select the kernel function and its parameter, RBF kernel is used in this paper. The experimental results are shown in Table I. Recognizing the PD models with the traditional SVMs is used in this paper too. For verifying the anti-noise ability of SVMs, uniformly distributed noises superimposed on the training samples and the testing samples remained unchanged, the experiment is done once more. The experimental results are shown in Table II.

As we see, using the FSVMs, the recognition ratio and the anti-noise performance all are better than the SVMs.

IV. SUMMARIES

FSVMs is used to recognize the discharge models in this paper. Experimental results show that this method gives good performance on reducing the effects of outliers and significantly improves the classification accuracy and robustness.

REFERENCES:

- [1] M.M.A. Salama and R. Bartnikas. Fuzzy Logic Applied to PD Pattern Classification [J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2000.1,pp 118-123.
- [2] R. Candela, G.Mirelli, R. Schifani. PD Recognition by Means of Statistical and Fractal Parameters and a Neutral Network [J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2000.1,pp 87-94.
- [3] Cortes C, Vapnik V N. Support vctor networks [J]. Ma-chines Learning, 1995, 20.3,pp273-297.
- [4] Vapnik V N. Statistical learning theory[M].New York:John Wiley, 1998.
- [5] S. Abe . Pattern Classification: Neuro-fuzzy Methods and Their Comparison. Springer-Verlag, London, UK, 2001.
- [6] Lin C F, Wan Sh D. Fuzzy support vector machines [J]. IEEE Transactions on Neural Networks, 2002, 13(2) pp: 464-471.
- [7] Xiufeng Jiang , Zhang Yi , Jian Cheng Lv. Fuzzy SVM with a new fuzzy membership function[J]Neural Comput & Applic[J]. 2006. 15, pp :268-276.
- [8] Bian Zhao-qi Zhang Xue-gong. Pattern Recognition [M]. Beijing: Tsinghua University Press, 2000: 135-136.
- [9] ZHANG Xiang, XIAO Xiao-ling, Xu Guang-you. Determination and Analysis of Fuzzy Membership for SVM [J] Journal of Image and Graphics2006.8, pp: 1188-1192.

[10] Hsueh-W, LING-J. A comparison of methods for multi-class support vector machines [J].IEEE Trans on Neural Networks, 2002.2, pp:415-425.

[11] SEBALDDJ, BUCHLEWJA. Support vector machines and the multiple hypothesis test problem [J]. IEEE Trans on Signal Processing, 2001, 49(11): 2865-2872

[12] Gao kai. The study of partial discharge models for generator [M]. Tsinghua university doctor dissertation, 2002

TABLE I. RECOGNITION RESULTS

Discharge model	Testing samples	Training samples	Recognition ratio (%)	
			SVMs	FSVMs
Tiny air-gap discharge	40	40	95	100
Interior air-gap discharge	40	40	92.5	97.5
Endings surface discharge	40	40	92.5	95
Slots discharge	40	40	95	100

TABLE II. RECOGNITION RESULTS COMPARISON

Discharge model	Noise ratio (%)	Recognition ratio (%)					
		9.7		19.5		30.4	
		SVMs	FSVMs	SVMs	FSVMs	SVMs	FSVMs
Tiny air-gap discharge		92.5	100	90.5	97.5	87.5	92.5
Interior air-gap discharge		90	95	87.5	92.5	82.5	90
Endings surface discharge		87.5	92.5	82.5	87.5	77.5	82.5
Slots discharge		92.5	97.5	90	92.5	82.5	90