Diagnosis and Analysis of Uncertain Information based on Probability-box

LIU Yan Jun[a], WANG Qing Xin[b], DING Jia Man*[c]

Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming Yunnan 650500, China

[a]454104763@qq.com, [b]yunnan403@sina.com, [c]tjom2008@163.com

Keywords: uncertain; fault information; probability box; support vector machine; information fusion

Abstract. The fault information always contains many uncertain, such as the missing information, malfunction and failure to operation. In this paper, we use a new method based on probability-box theory and support vector machine to solve the uncertain problems and improve the diagnosis ability. The first, structural the p-boxes based on the data of fault record. The second, using the fusion rules to get the different probability boxes fusion, then extract the feature from the p-boxes. The final, get the diagnosis result based on the support vector machine. This paper shows that the probability box has high diagnostic rate by comparing with the traditional method.

Introduction

It is crucial to locate the fault equipment quickly and shorten the fault processing time when the power grid fault [1][2], so researching a new diagnose method to fuse different data source is very meaningful.

Probability-box (P-box) is used to characterize uncertainty and uncertainty about the uncertainty, often used in risk analysis or quantitative uncertainty modeling [3]. SVM (support vector machine, SVM) is based on statistical learning theory of VC dimension theory and structural risk minimization principle based, according to the limited sample information between the complexity and the study ability of the model for the best compromise. In order to obtain the best generalization ability. SVM has demonstrated many unique advantages in solving small sample, nonlinear and high dimensional pattern recognition [4].

This paper proposes fault diagnosis method based on probability-box and support vector machine. The method uses multi-source information construct p-boxes, then fusion information and fault feature extraction. It can solve the small sample problem of uncertainty, improve the diagnosis rate. Finally, using the SVM method of fault classification and get the diagnosis results. The experimental results show that this method has high accuracy and efficiency of diagnosis.

Basic concepts

Probability box. Suppose $\bar{F}$ and $\underline{F}$ are non-decreasing functions from the real line R into $[0,1]$ and $\underline{F}(x) \le \bar{F}(x)$ for all $x \in R$. Let $[\bar{F}, \underline{F}]$ denote the set of all non-decreasing functions F from the reals into [0, 1] such that $\underline{F}(x) \le F(x) \le \bar{F}(x)$. When the functions $\bar{F}$ and $\underline{F}$ circumscribe an imprecisely known probability distribution, we call $[\bar{F}, \underline{F}]$, specified by the pair of functions, a “probability box” or “p-box” for that distribution[7]. From a lower probability measure* P for a random variable X, one can compute upper and lower bounds on distribution functions using.

Upper bound:

$$\bar{F}_X(x) = 1 - \underline{P}(X > x) \tag{1}$$

Lower bound:

$$\underline{F}_X(x) = \underline{P}(X \leq x) \tag{2}$$

Support Vector Machine. A complete, can handle linear and nonlinear, and can tolerate noise and outliers SVM described as follows.

$$\min \frac{1}{2} \| w^2 \| + C \sum_{i=1}^{n} x_i \tag{3}$$

$$\text{s.t.} \begin{array}{l} y_i(w^T x_i + b) \geq 1 - x_i, i = 1, 2....., n \\ x_i \geq 0, i = 1, 2....., n \end{array} \tag{4}$$

Here, w the (not necessarily normalized) normal vector to the hyper-plane; the parameter C controls the relative weighting between the twin goals of making the large (which we saw earlier makes the margin small) and of ensuring that most examples have functional margin at least 1; non-negative slack variables, ξ_i, which measure the degree of misclassification of the data x_i;b is the classification threshold.

SVM diagnosis based on probability-box theory

Construction of the p-boxes and fusion. The most important uncertain quantities characterized are estimated by modeling. The modeling approach to be employed in any particular case is determined by the modeler who understands something about the underlying physics or engineering. There are many ways to decompose a quantity in question into other quantities. In practice, modeling often consists of several or even many of these steps, which are sequentially applied to build up the desired estimate out of the sub-problems. These operations are convolution, transformations, enveloping, intersection, mixture, composition and de-convolution.

The steps in getting the p-boxes are shown as following:

1. Divide the data into several groups according to the sampling frequency.
2. Get the DSS as given format and discretizing of the DSS.
3. Construct the p-boxes.

Through different methods or under different conditions are obtained by different probability boxes, different probability boxes to be integrated can be relatively complete, consistent information. There several strategies for combing different p-boxes, such as null aggregation, intersection, envelope, Dempster’s rule and its modificatios and so on.

Feature extraction after fusion can increase the diagnostic accuracy. The cumulative uncertainty measures methods are to get some single scalars or intervals from the p-box’s upper and lower bounds.

SVM diagnosis based on p-boxes. With the increase of sampling frequency, the probability of overlap box is decreased with the uncertain region, but there are still some overlap, inevitable. Frequent sampling will increase the computational cost. In order to solve this problem, the support vector machine pattern recognition method, solve the problem of pattern recognition probability box. The process of SVM pattern recognition is shown as following [6]:

1. Establish the training database and testing database.
2. Data normalization.
3. Select the kernel function and the parameters of SVM.
4. Train the SVM by using the training database to get the trained Pattern.
5. Enter the testing data, using the trained Pattern to obtain the classification results.

Examples

The experimental environment as follows:

a. Voltage level is 220KV;
b. The length of the line is 300KM;
c. Sampling frequency is 2000HZ.

The voltage, current single index for fault diagnosis, compared with the fusion results, the diagnosis results are shown in table 1.

Table.1 The diagnosis results of different index contrast

index	fault		Non fault	
	accuracy (%)	time (ms)	Accuracy (%)	time (ms)
U	81.31	655	83.44	652
I	87.96	647	86.59	644
Fusion	94.37	643	97.12	639

As shown on table 1, the probability box after fusion experimental results more accurate, more efficient, reflect the use of probability box method for effective feature extraction, and the necessity of fusion. The probability box has some advantages in dealing with small samples, the uncertainty problem, can improve the accuracy of diagnosis.

Conclusions

This paper uses probability box theory for feature extraction, fusion of multi fault information, make full use of the advantages of the probability box dealing with uncertain problems. The SVM method has great advantage in fault diagnosis. Extracting different types of cumulative uncertainty measures from p-boxes can establish the SVM Features Database. The analysis result shows that the combination of p-box and SVM can achieve a high diagnosis rate.

The next target is fault distance measurement with more information.

Acknowledgements

This work was financially supported by the National Science Foundation of China (51467007), and the Application Basic Research Plan in Yunnan Province of China (2013FZ020). The corresponding author of this paper is Ding Jiaman.

References

[1] GAO Zhenxing, GUO Chuangxin, PENG Jianwei et al. Study of a fault diagnosis approach for power grid with information fusion based on multi-data resources[J]. Power System Protection and Control. 2013, 39(6):17-23.

[2] ZHANG Xu, WEI Juan, ZHAO Dongmei, et al. Research Course and Prospects of Power Grid Fault Diagnosis[J]. Power System Technology. 2013, 37(10): 2746-2753.

[3] ZHANG Xu, WEI Juan, ZHAO Dongmei, et al. Research Course and Prospects of Power Grid Fault Diagnosis[J]. Power System Technology. 2013, 37(10): 2746-2753.

[4] FANG Biao. Fault diagnosis model of transformer based on Karhunen Loeve feature extracting and suppose vector machine[J]. Sichuan Electric Power Technology, 2012, 35(6): 58-61.

[5] Fulvio Tonon. Using random set theory to propagate epistemic uncertainty through a mechanical system[J]. Reliability Engineering and System Safety. 85(2004) 169-181.

[6]DING Jia-man, DU Yi, WANG Qing-xin, et al. Fault Diagnosis of Power Transformer based on Probability-box Theory[C].

[7] SU Guangning, ZHANG Peichao, HU Yan, et al. A Novel Fault Diagnosis Method Based on Information Fusion of Multi-resources for Power Grids[J]. Automation of Electric System, 2012, 36(1): 61-64.

[8] GUO Chuangxin, PENG Mingwei, LIU Yi. Novel Approach for Fault Diagnosis of the Power Grid With Information Fusion of Multi-data Resoutces[J]. Proceedings of the CSEE, 2009 ，29(31) ：1-7 .

[9] XIA Keqing, CHEN Genjun, LI Li, et al. Real-time Analysis of Power Grid Fault Based on Multi-source Information Fusion and Its Application[J]. Automatic of Electric Power System, 2013, 37(24): 81-88.

[10] ZHANG Xiaoyuan, ZHANG Xinping, SU Baoping. Vibrant fault diagnosis for hydro-turbine generating unit using minmax kernel K-means clustering algorithm[J]. Power System Protection and Control, 2015, 43(5): 28-34.