Fault Diagnosis for Power System Transmission Line Based on PCA and SVMs

Yuanjun Guo, Kang Li, Xueqin Liu, Queen's University, Belfast

***Abstract*—This paper presents the application of a fault detection method based on the principal component analysis (PCA) and support vector machine (SVM) in power system transmission lines. Consider that the data maybe huge with a number of strongly correlated variables, a method that incorporates both the principal component analysis (PCA) and support vector machine (SVM) is proposed. This algorithm has two stages. The first stage involves the use of the PCA to reduce the dimensionality as well as to find violating point of the signals according to the confidential limit. The second stage is to use pattern recognition method to distinguish the phase of the faulty situation. The proposed scheme is able to solve the problems encountered in traditional magnitude and frequency based methods. The benefits of this improvement are demonstrated.**

***Index Terms*— Fault Diagnosis, Principal Component Analysis (PCA), Support Vector Machine (SVM).**

I. Introduction

IN power systems, the transmission line is the vital link between the electricity power production and usage. To find the accurate fault location in the transmission line based on the measurement of the currents and voltages is of great importance. Since the modern power system is well equipped with advanced measurement and protection instruments, huge amount of data has been collected with greater dimensionality and quantity. Applications of statistical monitoring techniques could be useful in extracting and interpreting process information from massive data sets in order to discriminate between power system normal or faulty states. At the same time, pattern recognition techniques could also be used to distinguish which phase of the power system is faulty.

Widely uses of control charts are based on the correlations between all the process variables. One approach that has proved particularly powerful is the use of Principal Component Analysis (PCA). PCA divides data information into the significant patterns, such as linear tendencies or directions in model subspace, and the uncertainties, such as noises or outliers located in residual subspace [1].

For the use of PCA technique to fault detection and identification, Venkat has summarized a comprehensive and systematic review about the process diagnosis [2]. Recently, R. Salat utilized fault diagnosis method in power system using multi-class least square support vector machines classifiers and proved the effectiveness [3].

Two major problems exist in the detection of the transmission lines. One problem is that there are a lot of data collected from online monitoring systems, which are multivariate and correlated. The other one is that there are various types of faults in transmission lines. For real-time fault diagnosis in transmission lines, the two problems should be considered simultaneously. This is a problem to balance the real time implementation and the accuracy.

In this paper, a fault diagnosis approach based on PCA and SVM is proposed to tackle the problem. In the proposed approach, a feature extraction algorithm based on PCA is used to reduce the dimensionality. The PCA monitoring scheme and the SVMs have been utilized to pinpoint the fault inception point and implement fault recognition, respectively. The experimental results show that the proposed approach is capable of detecting and recognizing the faults effectively.

II. Methodology

PCA is an optimal dimensionality reduction technique in terms of capturing the variance of the data, and it accounts for correlations among variables [4]. On this basis, more new methods have been developed, such as Dynamic PCA (DPCA), Independent Component Analysis (ICA), Nonlinear PCA [5].

Fisher Discriminant Analysis, FDA is a dimensionality reduction technique developed and studied within the pattern classification community [6].

Partial Least Square is a data decomposition method for maximizing covariance between predictor and predicted variables. Principal curves are smooth one-dimensional curves that pass through the middle of a p-dimensional data set, providing a nonlinear summary of the data.

My research methods will focus on the nonlinear extensions of PCR, which rely on PCA and the Radial Basis Function (RBF) networks. RBF networks based on the Fast Recursive Algorithm (FRA) [7] shows the usefulness of detecting and isolating faults in engineering system.

III. Implementation

The method was applied to the fault diagnosis in the power system transmission line. The simulation was carried out using SimPowerSystems package in Matlab, which described

This work was financially supported by RCUK including Engineering and Physical Sciences Research Council (EPSRC) under grant EP/G 042594/1, and partly supported by the China Scholarship Council, and National Natural Science Foundation of China under Grants 61271347 and 61273040.

a three-phase, 60Hz, 735kV power system transmitting power from a power plant through 600 km transmission line. The transmission line was split into two 300km lines connected between three buses. To simulate all the situations, different type of faults were considered by different combinations of source impedances. For each condition, the fault simulation was carried out by changing the parameters in the fault breaker and the data was collected respectively.

Two data sets were both recorded from the simulated transmission line corresponding to fault free and faulty conditions. The fault free data set described normal power transmission behavior and included 4000 samples. The faulty conditions consisted seven groups of data represented different type of faults as described previously, and were used for testing the performance of the proposed PCA and SVM method.

The result is shown in Fig.1, the sample point of 689 can be considered as the violating fault inception point as it violated both the T^2 and Q control limits. In the section before point 689, which is also referred to the pre-fault signal, the calculated results of T^2 and Q statistics are both under the control limit. The section after the inception point is named post-fault signal, with high values of statistic results. Then the unusual event can be detected by these high values of statistics. An effective set of multiple variety control charts is therefore a T^2 chart on the five dominant orthogonal PCs plus a SPE or Q chart. Fig.1 also indicates that these statistics respond sharply to the abrupt changes in the voltages and currents caused by transmission line faults.

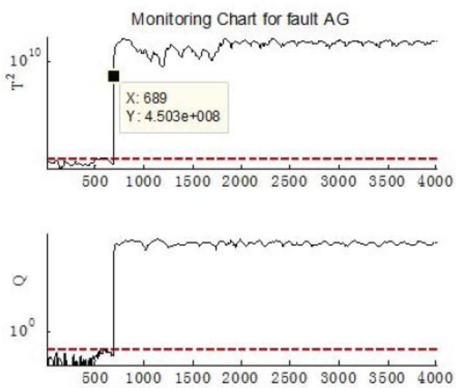

Fig.1. Monitoring chart for faulty signals

For a verification of the PCA model working under noisy condition, the model based on the noisy signals was identified next. In order to provide a fair comparison, the same five PCs were retained in the model. Noise was generated based on the random normal distribution, with 5% to 30% of the signal amplitude as the fluctuation range at the interval of 5%.

Fig.2 presents the entire situation under the 10% noisy condition. This indicates that the value of T^2 decrease with the increasing amplitude of the added noise. Some spikes appeared when the noise amplitude was bigger than 20% of the signal amplitude. In general, the error caused by different range of random noises were under 1% which was acceptable. These have led to a conclusion that the statistics control limit work well in the charts under the noise condition less than 30% except some fluctuations added.

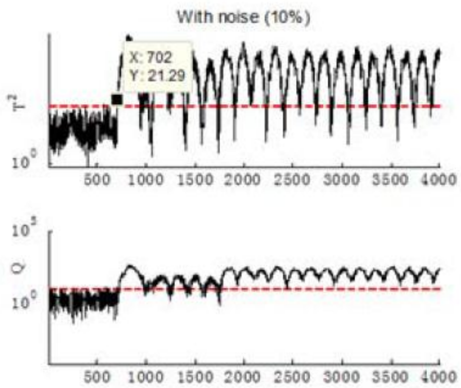

Fig.2 Monitoring chart for noisy faulty signals

IV. CONCLUSION AND FUTURE WORK

This methods is capable of capturing the relationship between the recorded variables from the data, and providing confidential limit charts for the violate fault points. It also helps to extract the features of the faulty signal under different faulty situations, which are used as the inputs of the SVMs to classify these faults correctly.

The work however has also raised some important issues with respect to the implementation of the combination method: (1) the monitoring statistics based on linear PCA are not sufficient to identify the nonlinear relations of the process variables, therefore, it is necessary to develop nonlinear and dynamic extensions for the nonlinear and dynamic system; (2) parameters of SVMs in this paper is not optimized or selected to achieve precise classifications. So the work in the near future is to solve these problems and to improve the PCA-SVM methods.

REFERENCES

[1] JF MacGregor and T. Kourti. Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3):403–414, 1995.
[2] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S.N. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & Chemical Engineering*, 27(3) : 293–311, 2003.
[3] R. Salat and S. Osowski. Accurate fault location in the power transmission line using support vector machine approach. *Power Systems, IEEE Transactions on*, 19(2):979–986, 2004.
[4] I.T. Jollie and MyiLibrary. Principal component analysis, volume 2. *Wiley Online Library*, 2002.
[5] D.J.H. Wilson, G.W. Irwin, and G. Lightbody. Rbf principal manifolds for process monitoring. *Neural Networks, IEEE Transactions on*, 10(6):1424{1434, 1999.
[6] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification and scene analysis 2^{nd} ed.* 1995.
[7] K. Li, J.X. Peng, and G.W. Irwin. A fast nonlinear model identification method. *Automatic Control, IEEE Transactions on*, 50(8):1211{1216, 2005.