An Application of Support Vector Machines for Fault Diagnosis

Pham Minh Hai * and Tu Minh Phuong **
* Dept of Automation in mining and petrolium, University of mining, Hanoi, Vietnam
email: pmhai80@yahoo.com
** Faculty of information technology, Posts & Telecom. Institute of Technology, Hanoi, Vietnam.
Tel. 84-4-8545604 email:phuongtm@fpt.com.vn

Abstract:
Fault diagnosis is one of the most studied problems in process engineering. Recently, great research interest has been devoted to approaches that use classification methods to detect faults. This paper presents an application of a newly developed classification method - support vector machines - for fault diagnosis in an industrial case. A real set of operation data of a motor pump was used to train and test the support vector machines. The experiment results show that the support vector machines give higher correct detection rate of faults in comparison to rule-based diagnostics. In addition, the studied method can work with fewer training instances, what is important for online diagnostics.

Keywords: support vector machines, fault diagnosis.

1. INTRODUCTION

Fault detection and diagnosis has received considerable attention from the industry and academia because of the economic and safety impact of this problem. Prompt detection of faults is essential for maintaining efficiency and quality in industrial processes. In recent years, the increasing availability of computing environments and the progress in problem solving paradigms have facilitated the development and application of a wide variety of computer-aided methods for on-line fault diagnosis. These methods differ from each other by the type of knowledge they are based on and how this knowledge is used.

Fault diagnosis is the identification or localization of the cause of fault operation. It involves determining which of the possible causes of faulty behavior are consistent with the observed behavior. Automated fault diagnosis relies entirely on sensor and so may not be able to identify the nature of the fault unambiguously, although it may be able to eliminate some of the possible causes. There are two basic ways to approach the analytical fault diagnosis problem: The model-based approach and the data-based approach [10]. In the model-based approach, the engineer has access to a model of the system whose behavior is being monitored. Most application of this approach have dealt with linear systems because they can be easily described and studied. The data-based approach bypasses the obtaining of a mathematical model and deals directly with the data. This approach passes a process called training, after that they use learned knowledge to identify fault situations.

In the past, the model-based approach was dominant. This approach requires explicit modeling of the process, filtering of measured data and estimation of the state variables. The output of the model is compared to the measurements from the process, the generated residuals or errors are then used for making decision about the state of the system. Faults are detected with a quantitative model by using measured values of the control signals and some of the sensor signals as inputs to the model. The remaining sensor signals are then compared with the predictions of the model, as shown in Figure 1. Significant differences indicate the presence of a fault somewhere in the part of the system treated by the model.

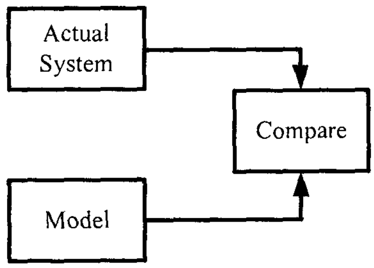

Fig. 1. The model-based fault detection scheme

In the last ten years, data-based approaches that do not assume any form of model information have received much of research interest. These include cased-based reasoning, rule-based systems [12], and knowledge-based expert systems among others. Classification algorithms have been extensively researched for fault diagnosis. There are some useful methods applied to fault diagnosis problems, for example, artificial neural network (ANN) [7,10,11], fuzzy logic (FL), neuro-fuzzy (NF) [8].

Case-based reasoning is a technique for solving new problems by adapting solution that have been used to solve old problems. The most similar cases must be retrieved in the current case base. The case base is an experience-based repository of currently known cases for a focused problem domain. Thus, a case is a documented experience, placed in the context of the domain, and it should comprise a description of the problem, the solution used to solve that problem, and the outcome describing the changes that the solution produced.

Rule-based systems are used to solve classification problems. The rules, which may be Boolean or fuzzy, can either obtained from experts and then checked for consistency and completeness or can be generated from simulation and/or real data using rule learning algorithms.

In this paper, we study an application of support vector machines for fault diagnosis in an industrial case study. The support vector machine has been chosen due to their good performance in a number of applications. We compare the performance of SVM and a rule-based method in a real operation data set of a motor pump.

The rest of the paper is structured as follows: section 2 briefly reviews the SVM method, in section 3, the application of SVM in motor pump fault diagnosis problem is described, and in section 4, we draw our conclusions.

2. SUPPORT VECTOR MACHINES

Support vector machine is a new machine learning method. The SVM can be used to learn polynomial, radial basis function (RBF) and multi-layer perceptron (MLP) classifiers. SVM has rooted from the statistical learning theory in the 1960s. However, since 1995, the algorithms have used for SVM to start emerging with greater availability of computing power, paving the way for numerous practical applications [1,2,5].

SVM is an approximate implementation of the method of structural risk minimization. This introduction principle is based on the fact that the error rate of a learning machine on test data, which is bound by the sum of the training-error rate and a term that depends on the Vapnik-Chervonenkis (VC) dimension. An SVM for classification attempts to find a hyperplane that maximizes the margin between positive and negative examples, while minimizing training set misclassifications. SVM are based on strong mathematical theory, and their observed empirical performance is extremely good. However, training an SVM requires the solution of a quadratic program in as many variables as there are data points in the training set. SVM were developed to solve the classification problem, but recently they have been extended to the domain of regression problems.

An SVM has two distinct features. Firstly, it is often associated to the physical meaning of the data so it is easy to interpret (in contrast, ANN does not possess any physical meaning). Secondly, it requires only small amount of training samples. In other researches, the results also show that SVM has higher or equal success rate than ANN [3,7].

Here we focus on SVM for two-class classification, class I and class II. This can easily be extended to $k-$class classification by constructing k two-class classifiers. Kernel functions are then introduced in order to construct non-linear decision surfaces.

Let n-dimensional input $\mathbf{x}_i$ $(i=1,..,M)$ belong to class I or class II and associated labels be $y_i=1$ for class I and $y_i=-1$ for class II. For linearly separable data, we can determine a hyperplane $f(\mathbf{x})$ that separates the data. For a separating hyperplane $f(\mathbf{x}) =0$, if the input $\mathbf{x}$ belongs to positive class, and $f(\mathbf{x})<0$, if $\mathbf{x}$ belongs to the negative class.

$$f(x) = w.x + b = \sum_{j=1}^{n} w_j x_j + b \qquad (1)$$

$$y_i f(x_i) = y_i (w.x_i + b) \geq 0, for \quad i = 1,..,M \qquad (2)$$

where w is an n-dimensional vector and b is a scalar.

The weighting vector **w** defines the direction of the separating hyperplane $f(\mathbf{x})$ and b (bias) defines the hyperplane's distance from the origin.

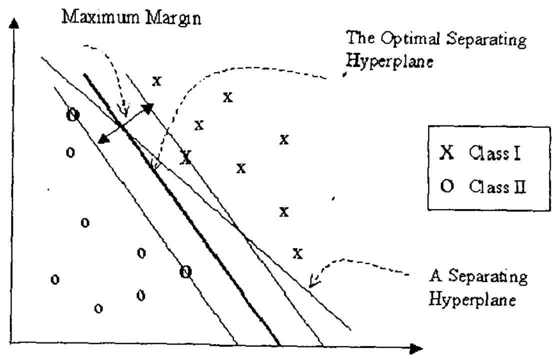

Fig. 2. Optimal hyperplane

The separating hyperplane that has the maximum distance between the hyperplane and the nearest data, i.e. the maximum margin, is called the optimal separating hyperplane. The generalisation ability is maximized with the optimal hyperplane. An example of optimal separating hyperplane of two datasets is presented in Figure 2. From the geometry, the optimal hyperplane can be obtained by solving the following convex quadratic optimisation problem:

$$\text{minimize } \frac{1}{2}\|w\|^2 \qquad (3)$$

supject to $y_i(w.x_i + b) \geq 1$

If the number of attributes of data examples is large, we can considerably simplify calculations by converting the problem with Kuhn-Tucker conditions into the equivalent Lagrange dual problem. Lagrange function for (3) is:

$$L(w,b,\alpha) = \frac{1}{2}(w.w) - \sum \alpha_i \left[y_i((w_i.x_i) + b) - 1 \right] \qquad (4)$$

where $\alpha = (\alpha_1,...,\alpha_M)$ is the Lagrange multiplier. The dual problem is:

$$\text{maximize } L(w,b,\alpha) \qquad (5)$$

supject to $\alpha_i \geq 0, \ i = 1,...,M$

By differentiating (4) with respect to **w** and **b** and imposing stationarity, we get:

$$\frac{\partial L}{\partial w}(w,b,\alpha) = w - \sum_{i=1}^{M} y_i \alpha_i x_i = 0$$
$$\frac{\partial L}{\partial b}(w,b,\alpha) = \sum_{i=1}^{M} y_i \alpha_i = 0 \qquad (6)$$

From (4), (5) and (6) we get the dual representation of the optimisation problem:

maximize $W(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,k=0}^{M} \alpha_i \alpha_k y_i y_k x_i . x_k$

$$\text{supject to } \sum_{i=1}^{M} y_i \alpha_i = 0,\ \alpha_i \geq 0,\ i = 1,...M \qquad (7)$$

The number of variables of the dual problem is the number of training data.

Let us assume that optimal solution for the dual problem is α^* and b^*. According to the Karush-Kuhn-Tucker theorem, the equality condition in (2) holds for the training input-output pair ($\mathbf{x}_i$,y_i) only if the associated α_i^* is not 0. In this case the training example $\mathbf{x}_i$ is a support vector. Solving (7) for $\alpha = (\alpha_1),...,\alpha_M$, we can obtain the support vectors for classes I and II. Then the optimal separating hyperplane is placed at the equal distances from the support vectors for classes I and II, and b^* is given by:

$$b^* = -\frac{1}{2} \sum_{k=1}^{M} y_k \alpha_k^* (s_1 x_k + s_2 x_k)$$

where **s**1 and **s**2 are respectively, arbitrary support vectors for class I and class II. In Figure 2, support vectors are bolded. Notice, that support vectors are such training samples that are on the margin of two datasets. The optimal separating hyperplane would be the same, if only support vectors had been used as training data. So far we have assumed that the training data is linearly separable. In the case where the training data cannot be linearly separated, we introduce non-negative slack variables ξ_i to (2), and add to the objective function given by (5), the sum of the slack variables multiplied by the parameter *C*. This corresponds to adding the upper bound *C* to α. In both cases, the decision functions are the same and are given by:

$$f(x) = \sum_{i=1}^{M} \alpha_i^* y_i x_i . x + b^*$$

Then unknown data example x is classified as follows:

$$x \in \begin{cases} Class\ I\ if\ \ f(x) > 0 \\ Class\ II\ otherwise \end{cases}$$

SVM is a non-linear kernel-based classifier, which maps the data to be classified, *X*, onto a space, where the data can be linearly classified. The space is called a feature space, *F*. This is depicted in Fig.3.. . Now using the non-linear vector function $\Phi(x) = (\Phi_1(x),...,\Phi_l(x))$ that maps the *n*-dimensional input vector **x** into the *l*-dimensional feature space, the linear decision function in dual form is given by:

$$f(x) = \sum_{i=1}^{M} \alpha_i y_i \Phi(x_i).\Phi(x) \qquad (8)$$

Notice that in (8) as well in the optimisation problem (7), the data occur only in inner products. In SVM, the actual mapping function, Φ, is not necessary to be known, but the classes optimally separating hyperplane is possible to calculate with inner products of the original data samples. If it is possible to find this kind of procedure to calculate inner products of feature space in original data space, it is called a kernel, $K(x,z) = \Phi(x).\Phi(z)$. Then the learning in the feature space does not require evaluating Φ or even knowing it, because all the original samples are handled only with Gram matrices $G = ((x_i . x_j))^M{}_{i,j=1}$. Using a Kernel function, the decision function will be:

$$f(x) = sign\left(\sum_{\sup port\ vector} \alpha_i^* y_i K(x_i, x) \right)$$

and the unknown data example is classified in a following way:

$$x \in \begin{cases} Class\ I\ if\ \ f(x) = +1 \\ Class\ II\ if\ \ f(x) = -1 \end{cases}$$

There are a number of Kernel functions mentioned in the literature [6,8,9]. A concrete selection depends on a concrete problem and must be defined by experiments. However, all kernels do not correspond inner products in some feature space *F*. With a so called Mercer's theorem it is possible to find out, whether a kernel *K* depicts an inner product in that space where Φ is mapped.

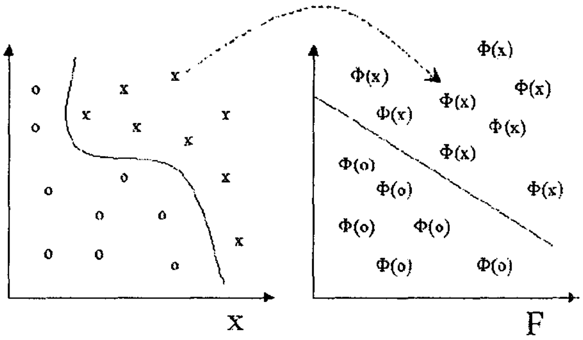

Fig. 3. Mapping Φ from the data space X to the feature space F

3. EXPERIMENTS AND RESULS

In this section, we will present an application of SVM in fault diagnosis of a motor pump described in [4]. The results are then compared with the results from the original paper, which used a rule learning based algorithm to detect faults. The motor pump was chosen due to the popularity of such rotating apparatuses in pratice and, more important, due to the availability of training data.

3.1. Preliminaries

The motor pump studied belongs to a large class of apparatuses that range from small motor pumps to very large turbo. All the considered apparatuses share the common feature of possessing a rotating shaft to which various rotors are connected. If the motion has to be transmitted between separate shafts, transmission parts such as rigid or elastic joints, belt drivers and gear trains will be used. The scheme of a typical motor pump is reported in Figure 4.

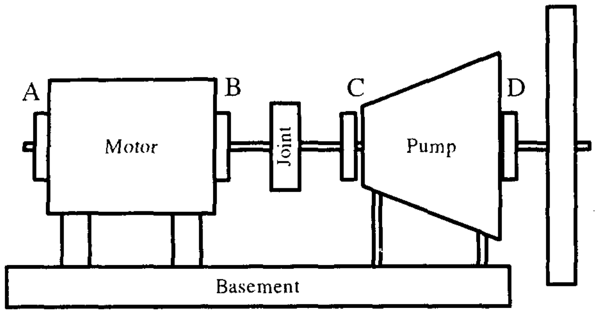

Fig. 4. Scheme of a horizontal centrifugal motor pump. The supports of the motor are labeled A and B, those of the pump C and D. The motor and the pump shafts are connected through an elastic joint. Both the motor and the pump are anchored to the ground by means of elastic support. Rolling bearings are located inside the four supports.

When a machine possesses rotating elements, several unavoidable vibratory motions are included in its parts, these vibrations occur also during the correct machine operation and are not dangerous as long as their amplitude remains limited. When some fault occurs in the machine, new anomalous vibrations will appear, beside other manifestations. The aim of fault daignosis is to locate faults through an analysis of these vibrations. The diagnosis process basically performs a Fourier analysis of the vibratory motions taken in prespecified and labeled points, precisely on the supports of the machine components. By means of a special analyzer, the technician obtains, for each support, the amplitude and velocity of the global vibration along the vertical, horizontal, and axial direction. Furthermore, the same data can be taken for each of the harmonic components of the vibrations.

Mechanalysis has a strong mathematical foundation in vibration theory, and hence, the relationships between anomalous frequencies and faults could be predicted. To keep all the electromechanical operate in the control, the aim to solve fault diagnosis problem is necessary. This creates a taxonomic structure in the set of diagnoses, which must be taken into account both in learning and in classification.

3.2. The data set

For testing SVM in fault diagnosis of the motor pump, we have used the data set called "Mechanical analysis data" [4]. The data were collected from a number of sensors mounted on the motor pump. The data set has 209 instances, each with seven attibutes.

The seven attributes are:

1. class - classification (1..6, the same for components of one example)
2. # - component number (integer)
3. sup - support in the machine where measure was taken (1..4)
4. cpm - frequency of the measure (integer)
5. mis - measure (real)
6. misr - earlier measure (real)
7. omega - rpm of the machine (integer, the same for components of one example)

Table 1 shows the six conditions tested (6 classes), with the 6^{th} is the normal operating condition.

Since SVM can classify instances into only two classes we have combined several SVM for the six-class classification problem as in our case. Hence, we used five SVMs to classify all conditions. One SVM is used to detect faults while others are use to identify the faults. This design is necessary because SVM can deal with only two classes and it helps to quickly detect the faults.

Table 1. Operating conditions of motor pump

Conditions (Classes)	Description
1.	Problems in the joint
2.	Faulty Bearings
3.	Mechanical Loosening
4.	Basement Distortion
5.	Unbalance
6.	Normal operating conditions

Then, combinations of the classes are possible with the following groupings:

7. Shaft misalignment (includes class 1 and class 4)
8. Problems in the pump (includes classes 2, 3 and 5)
9. Problems in the motor (includes classes 2, 3 and 5)
10. Problems in the machine (includes all basic classes except class 6)

Figure 5 shows the diagnosis of the motor pump fault.

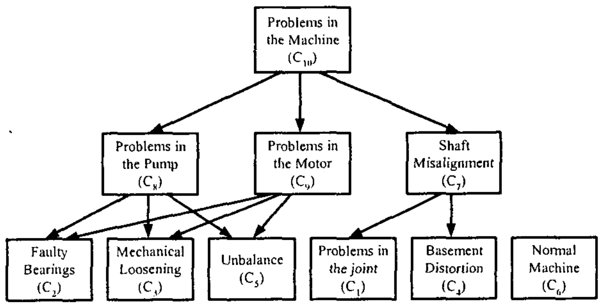

Fig. 5. Diagnosis taxonomy of the motor pump faults. The leaves correspond to single faults ($C_1 \div C_6$). The internal nodes ($C_7 \div C_9$) correspond to more generic fault locations, whereas the root C_{10} corresponds to the case in which a fault does occur, but it is impossible to locate it.

For using SVM in pratice, choosing an appropriate kernel function is important. The kernel function reflected the geometric relationship between the input vector and the support vector, or the similarities of the features of the fault. We have tried several kernel functions (such as polynomial, radial basis function and sigmoid). The experiments show that the polynomial kernel function gives the best performance both of correct prediction rate and time to converge to SVM.

To calculate the success rate of the method, we used popular *n-fold cross validation*. The data set is randomly divided into n subsets. In each test, n-1 subsets are used to train SVM and the remain subset is used to test the learned SVM. The success rate is then calculated as the average of n tests. We used ten folds in our experiments. In addition, we counted an instance as correctly classified if and only if the instance is classified correctly to a specific class.

In the original paper, the authors used a rule-based system called ENIGMA for fault diagnosis. Table 2 shows the rate of correct fault detection using SVM in comparison with the detection rate of ENIGMA as reported in [4]. Although the success rate of 75% is not as high as reported in some other papers, it is a level higher than that of ENIGMA. The low success rate can be explained by the specificities of the data set we used.

Table 2. Testing results

	SVM	ENIGMA
Success rate (%)	75.33	70,49

4. CONCLUSION

In this paper, we have used a new classification technique, support vector machines, for fault diagnosis of a motor pump. The experiment results with a real data set show the better perfomance of SVM in comparison with a rule-based classification system. These results are promising because they demonstrate the potential of using SVM in fault detection and diagnosis problems. In future research we intend to do more comparative study of SVM with other classification methods such as ANN, HMM-based.

REFERENCES

[1] R. Burbidge, B. Buxton, "An introduction to Support vector machines for data mining", Gower Street, UK, 2000.

[2] C. J. C. Burges, "A tutorial on Support vector machines for pattern recognition" Data mining and knowledge discovery 2, pp. 121-167, 1998.

[3] M. Ge, R. Du, G. Zhang and Y. Xu, "Fault diagnosis using Support vector machines with an application in sheet metal stamping operations", Mechanical systems and signal processing, vol. 18, pp. 143-159, 2004.

[4] A. Giordana, L. Saitta, F. Bergadano, F. Brancadori and D. D. Marchi, "ENIGMA: A System that learns diagnosis knowledge" IEEE transactions on knowledge and data engineering, vol. 5, no. 1., pp.15-28,1993.S.

[5] R. Gunn, "Support vector machines for classification and regression", Technical Report, University of Southampton Press, 1998.

[6] C.-W. Hsu, C.-C. Chang and C.-J. Lin, "A practical guide to Support vector classification", National Taiwan University Press, 2002.

[7] B. Li, M.-Y. Chow, Y. Tipsuwan and J. C. Hung, "Neural-network-base motor rolling bearing fault diagnosis" IEEE transactions on industrial electronics, vol. 47,no. 5, pp. 1060-1069, 2000.

[8] V. Palade, R. J. patton, F. J. Uppal, J. Quevedo and S. Daley, "Fault diagnosis of an industrial gas turbine using neuro-fuzzy methods" IFAC Press, 2002.

[9] W. G. Philip, M. H. paul and G. M. Laurence, "Fault diagnosis for industrial printer using case-based reasoning" Engng Applic. Artif. Intell, vol. 9, no. 2, pp. 163-173, 1996.

[10] B. Samanta, "Gear fault using Artificial neural networks and Support vector machines with genetic algorithms", Mechanical systems and signal processing, vol. 18, pp. 625-644, 2004.

[11] B. Samanta and K. R. Al-Balushi, "Artificial neural network based fault diagnosis of rolling element bearings using time-domain features", Mechanical systems and signal processing, vol. 17, no. 2, pp. 317-328, 2003.

[12] V. Venkatasubramanian, R. Rengaswamy and S. N. Kavuri, "A review of process fault diagnosis. Part II: qualitative models and search strategies" Computer and chemical engineering, vol. 27, pp. 313-326, 2003.