University of Alberta

Support Vector Machines for Regression and Its Application for Prediction of Machine Degradation based on Vibration Signals

by

Hui Lin 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science.

Department of Mechanical Engineering

Edmonton, Alberta

Spring 2006

Your file Votre référence
ISBN: 0-494-13845-9
Our file Notre référence
ISBN: 0-494-13845-9

Abstract

A condition based predictive maintenance system will have the following functions: analyzing collected vibration signals and identifying the components that have deteriorated significantly, predicting the degradation of these components, and making an appropriate maintenance plan for minimizing total equipment operation cost. In this thesis, Support Vector Machines (SVM) regression is used for prediction of machine's degradation and is studied in depth. The selection of SVM model parameters is investigated based on current problems that have arisen in the industrial application of SVM regression. A new rule is proposed for selection of the error zone value, one of the SVM model parameters. The proposed rule is also compared with CMa's method and the results show that applying the new rule, SVM regression can provide better prediction than CMa's method.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Dr. Ming J Zuo, an excellent professor and supervisor, for giving me the opportunity to study in the Reliability Group, and for providing me with patient guidance, continual support and valuable suggestions throughout my research.

I wish to thank Dr. XianFeng Fan for his thoughtful discussions and kind help.

I wish to express my gratitude to all members of the Reliability Research Lab group: Wei Li, Siyan Wu, Yao Wang, and Zhigang Tian, whose knowledge and ability have helped me a great deal.

I wish to dedicate my thesis to my parents for their encouragement and trust throughout my education.

Above all I wish to thank my husband, Yougang, and my daughter, Cenfeng, for their understanding and support along my lengthy and wandering path through my graduate studies.

Table of Contents

List of Tables

List of Figures

List of Abbreviation

KKT ..... Karush Kuhn Tucker Conditions

SVM..... Support Vector Machines

SRM..... Structural Risk Minimization

ERM ..... Empirical Risk Minimization

SLT ..... Statistical Learning Theory

QP ..... Quadratic Programming

poly ..... Polynomial Kernel

rbf..... Radial Basis Function Kernel

polyrbf ..... Mixed Kernel which combined polynomial and radial basis function kernels

mse..... Mean Square Error

CHAPTER 1

INTRODUCTION

1.1 Motivation

The prediction of Machine degradation is very useful in industries. With precise prediction, catastrophic failure and a decrease in product quality can be mitigated. Moreover, accurate prediction is helpful to engineers attempting to make reasonable inspection and maintenance plans. Degradation prediction regarding mechanical equipment relies heavily on mechanical vibration signals. Mechanical vibration is sensitive to noise and outliers, two features which complicate degradation prediction making it difficult to achieve a high degree of accuracy.

Since the creation of the SVM theory by Vapnik in 1995 [29] at the AT&T Bell Laboratories, there have been breakthroughs in the application of SVM to classification. Results in pattern recognition have been promising. On the other hand, the application of SVM to regression has only recently come under development. Some researchers have used SVM for time series pre-

dictions, such as the forecasting of financial markets [2, 14], the estimation of power loads [33], and the prediction of travel time [31]. Support vector machines have been applied to the mechanical industry mostly with regard to fault diagnosis, which uses SVM classification instead of SVM regression. Using support vector regression to predict the degradation of mechanical equipment has been developed by Jiang [12]. For his thesis, Jiang applied SVM regression to equipment deteriorating prediction based on vibration signals. He also compared SVM and time series models in terms of prediction accuracy. His experiments showed that SVM regression makes better predictions than are possible with time series models [12]. In studies of condition trend prediction Lin *et al* [16] applied Support Vector Machines regression to machines condition trend prediction based on vibration signals collected from machines. He concluded that the SVM regression does an excellent job of predicting the condition trend of mechanical equipment and will be used increasingly for that purpose.

As we can see, SVM has great potential and has performed well in many fields; this is because SVM is theoretically based on the structural risk minimization (SRM) principle. The SRM principle is better able to generalize and is therefore superior to the empirical risk minimization (ERM) principle used in time series models and neural networks [4]. Because of the SRM principle, the global optimal value of SVM regression model can be obtained, whereas other prediction models can obtain only a local optimal value [29]. Thus, the SVM regression model has great theoretical advantages over other

prediction tools.

The SVM regression model does, however, face a big challenge. This is because there are three factors that control the prediction performance of the SVM regression model. They are the error zone term, regulation factor used for capacity control, and the type of kernels and kernel parameters [22]. Improper selection of the above three factors could result in bad prediction results. As reported in the literature, parameter selection and kernel type selection are complicated because the performance of the SVM regression model depends on all three factors. So far, despite the work that has been done, no single rule or method has been judged superior for selecting the suitable parameters and kernel types. Under the circumstances, the SVM regression model requires expert interface and is not user friendly. The application of the SVM regression model to real industry is under development, therefore, finding some rule or method for selecting a suitable model parameter that will improve the prediction accuracy of SVM regression is a top priority.

1.2 Research Objective

The goal of this thesis is to investigate how these factors affect the performance of SVM regression, and what rule can be followed to select the three factors so as to obtain the most reasonable prediction results. With this goal in mind, one learning process will be proposed that focuses on the selection of a suitable error zone value, kernel type and kernel parameter. Several

benchmark data sets will be adopted to show how SVM regression performance fluctuates as both kernel type and SVM parameters vary. The focus will largely be on selection of the error zone value because its flexibility is the source of the undependable predictions produced by the SVM regression model. In contrast, the selection of kernel type and kernel parameter is relatively limited although some research work would be worthwhile. As well, the application of SVM regression to the prediction of machine degradation based on vibration signals will be studied, because this is a very important step in making possible condition based maintenance in the mechanical industry.

1.3 Organization of Thesis

The thesis is organized as follows. Chapter 2 provides a literature review on SVM principle, theory and its application in different fields. It also introduces the three most important factors affecting SVM regression and their impact on the prediction performance of the SVM regression model. Chapter 3 proposes one learning process for selecting suitable error zone value, kernel type and kernel parameter. Three data sets are used to test SVM regression's performance following the proposed learning process. One conclusion on optimal error zone value is initially brought forward based on the analysis of the experiments results. Chapter 4 adopts the Golden Section Search Optimization method to further prove the conclusion proposed in Chapter 3. Chapter 5 then compares this thesis' conclusion with CMa's method. The

result demonstrates the superiority of this thesis' conclusion. In the next chapter, the SVM regression model with optimal error zone value is used to predict the vibration trend of machines based on their vibration signals. The resulting prediction shows that it is successful. The final chapter gives the summary and conclusion of this thesis and suggests possible areas for future research.

Chapter 2

Literature Review on SVM Regression Theory and Its Application

2.1 The Basic Principles of Support Vector Machines Regression

The Support Vector Machines regression model is developed based on the Structural Risk Minimization (SRM) principle [29]. SRM came from the statistical learning theory (SLT) developed by Vapnik in the 1960's. SRM provides a solid foundation for SVM, making it superior to other learning machines. The SRM principle defines a trade-off between the quality of the approximation of the given learning data and the complexity of the approximation function [9]. In SVM, the quality of the approximation of the given learning data represents the learning ability of the machines. Complexity is a term that describes how many parameters are in the potential regression function. If the complexity of a learning machine is too high, the learning ability of this machine may be good, but its prediction ability may not. The

learning machine will over-fit the data. On the other hand, when the learning machine has too little complexity, it may not have good learning ability. The learning machine will under-fit the data. The SRM principle is concerned with finding the right balance between the learning ability and prediction ability of the learning machines.

In support vector regression, structural risk consists of two parts: the empirical risk of the function, and a complexity term. The expression is as follows:

$$R_{reg} = CR_{emp}[f] + \frac{1}{2}\|\omega\|^2 = C\sum_{i=1}^{l}|y_i - f(x_i)|_\varepsilon + \frac{1}{2}\|\omega\|^2, \qquad (2.1)$$

where $x_i \in R^n$, $y_i \in R$, $\omega \in R^n$, $R_{emp}[f]$ is the empirical risk which is defined as the training error; $\sum_{i=1}^{l}|y_i - f(x_i)|_\varepsilon$ is defined as loss function, which is the sum of the error between the predicted data, $f(x_i)$, and the observed data, y_i; $\|\omega\|^2$ represents the complexity of the learning machines, which is defined as how many parameters there are in the learning machines; l denotes the sample size, and C is called the regularization constant, which is used to control the trade off between the complexity and the error term.

Structural risk minimization determines the parameter ω by minimizing the equation (2.1). We can see that applying structural risk minimization provides a good balance between the complexity of the machine and the training error.

2.2 Linear Regression Using Support Vector Machines

Suppose we have a data set $\{(x_1, y_1)\}, ..., (x_l, y_l)\}$, $\mathbf{x}_i \in \mathrm{R}^n$, $y \in R$. SVM is applied in regression by introducing a class of loss functions. Figure 2.1 illustrates four loss functions.

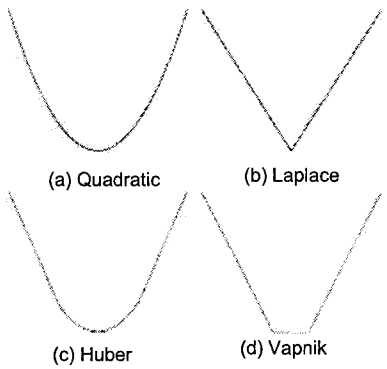

Figure 2.1: Loss Functions

The loss function in Figure 2.1 (a) corresponds to the traditional least square error criterion. The loss function in Figure 2.1 (b) is a Laplacian loss function that is less sensitive to outliers. Huber proposed the so-called robust loss function shown in Figure 2.1 (c), which has better property when the underlying distribution of the data is unknown [9]. However, these three loss functions do not have the sparseness property, which will be explained later in the section on support vectors. Vapnik proposed the loss function in Figure 2.1 (d), which is called the ε-loss function [29]. The ε-loss function is an approximation to Huber's loss function but make it possible to obtain a

sparse set of support vectors.

The ε-loss function is:

$$|y - f(\mathbf{x})|_{\varepsilon} = \max\{0, |y - f(\mathbf{x})| - \varepsilon\}. \tag{2.2}$$

In ε-SV regression, our goal is to find a function, $f(\mathbf{x})$, such that all data points are at most ε deviation from this function. Errors less than ε are ignored, but those greater than this are penalized. Figure 2.2 shows the ε-tube and the slack variables, ξ and ξ^*.

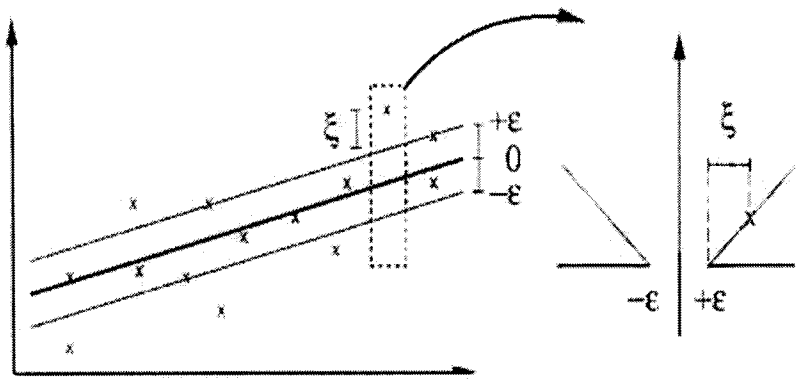

Figure 2.2: The ε-loss function

Based on the structural risk minimization principle and the ε-loss function, we obtain the following regression model:

$$\min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{2.3}$$

$$\text{Subject to:} \quad \begin{array}{l} y_i - \langle \omega, \mathbf{x}_i \rangle - b \leq \varepsilon + \xi_i \\ \langle \omega, \mathbf{x}_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \end{array} \quad \text{and } \xi_i, \xi_i^* \geq 0,$$

where the term $C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$ represents the error term. The constant C determines the trade off between the complexity and the empirical risk.

Now, the regression problem becomes a quadratic optimization problem that can be dealt with by using Lagrange theory. The key idea is to construct a Lagrange function from both the objective function and the corresponding constraints by introducing a set of multipliers. This Lagrange function is:

$$\begin{aligned} L = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) - \sum_{i=1}^{l} \alpha_i (\varepsilon + \xi_i - y_i + \langle \omega, \mathbf{x_i} \rangle + b) \\ - \sum_{i=1}^{l} \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle \omega, \mathbf{x_i} \rangle - b) - \sum_{i=1}^{l} (\eta_i \xi_i + \eta_i^* \xi_i^*), \end{aligned} \tag{2.4}$$

where the multipliers $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$.

A very important property for Lagrange is that at the saddle point, the partial derivatives of L with respect to the primal variables $(\omega, b, \xi_i, \xi_i^*)$ should be zero.

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0, \tag{2.5}$$

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \mathbf{x}_i = 0, \tag{2.6}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0, \tag{2.7}$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0. \tag{2.8}$$

Substituting (2.5), (2.6), (2.7), (2.8) into (2.4), yields the dual optimization problem [29].

$$\begin{aligned} max \quad W(\alpha, \alpha^*) &= -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\left\langle \mathbf{x_i}, \mathbf{x_j} \right\rangle \\ &- \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i\left(\alpha_i - \alpha_i^*\right) \end{aligned} \tag{2.9}$$

$$\text{Subject to: } \begin{array}{l} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C]. \end{array}$$

In deriving (2.9), the dual variables η_i, η_i^* are eliminated through the condition (2.7) and (2.8).

From the condition (2.6), parameter ω is obtained.

$$\omega = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\mathbf{x}_i. \tag{2.10}$$

Therefore, the following regression model is obtained:

$$f(x) = \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left\langle \mathbf{x}_i, \mathbf{x} \right\rangle + b. \tag{2.11}$$

Equation (2.10) is also called the Support Vector expansion since ω can be completely described as a linear combination of the training data $\mathbf{x_i}$. The vector $\mathbf{x}$ represents the testing points.

Now, let us compute the parameter b in equation (2.11). Karush-Kuhn-Tuker (KKT) condition [29] states that at the optimal solution, the products between multipliers and inequality constraints equal zero. This means:

$$\begin{aligned} \alpha_i \left(\varepsilon + \xi_i - y_i + \langle \omega, \mathbf{x}_i \rangle + b\right) = 0 \\ \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle \omega, \mathbf{x}_i \rangle - b) = 0 \end{aligned}, \tag{2.12}$$

$$\begin{aligned} & (C - \alpha_i)\, \xi_i = 0 \\ \textit{and} \quad & (C - \alpha_i^*)\, \xi_i^* = 0. \end{aligned} \tag{2.13}$$

From equations (2.12) and (2.13), it is apparent that [29]:

1. Only the samples $(\mathbf{x_i}, \mathrm{y_i})$ corresponding to $\alpha_i = C$ and $\alpha_i^* = C$ that lie outside the ε tube around $f(\mathbf{x})$ are used to construct the SVM model.

2. If $\alpha_i, \alpha_i^* \in (0, C)$, then $\xi_i, \xi_i^* = 0$ and therefore the second factor of equation (2.12) equal zero. b is obtained from Equation (2.12)

$$b = y_i - \langle \omega, \mathbf{x}_i \rangle - \varepsilon, \quad for \quad \alpha_i \in (0, C). \tag{2.14}$$

$$b = y_i - \langle \omega, \mathbf{x}_i \rangle + \varepsilon, \quad for \quad \alpha_i^* \in (0, C). \tag{2.15}$$

A very important property of SVR is the sparsity of the SVR expansion

which is expressed by Equation (2.10). From Equations (2.12) and (2.13), only if $|f(\mathbf{x}_i) - y_i| \geq \varepsilon$ Lagrange multipliers may be nonzero. In other words, for samples inside the ε-tube, α_i, α_i^* would be zero. For $|f(\mathbf{x}_i) - y_i| < \varepsilon$, the second factor in Equation (2.12) is nonzero, hence α_i, α_i^* have to be zero so that the KKT conditions is satisfied. Therefore, in the expansion Equation (2.10), $\mathbf{x}_i$ refer only to the sample data with nonzero coefficients, i.e. sample data outside the ε-tube. The sample data $\mathbf{x}_i$ used in the expansion Equation (2.10) are called support vectors [29]. Thus, the parameters of the regression model depend only on the support vectors, which represent the sparseness property of support vector machines.

2.3 Nonlinear Regression Using Support Vector Machines

SVM was first developed as linear learning machines; however, most problems encountered in real life are nonlinear. This requires learning machines be able to map nonlinear data in the input space into linear data in a higher dimension space [29]. This higher dimension space is called the feature space. In this way, the capacity (sometimes, called generality) of the support vector machines is increased; however, as the number of the dimensions grows the dimensionality of the feature space can become computationally difficult to handle. Furthermore, the learning machine's prediction performance will degrade as the number of input dimensions increases [23]. This phenomenon is called the curse of dimensionality. Kernel method is the key to solving this problem.

The idea of the kernel function enables operations to be performed in input space rather than in high dimensional feature space [22]. Now, let's re-exam the support vector machines in linear form. The equation is:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \langle \mathbf{x}_i, \mathbf{x} \rangle + b, \tag{2.16}$$

where $\mathbf{x}_i$ represents the support vectors.

If the input data $(\mathbf{x}, y)$ are nonlinear, the specific mapping would be made from the input space into high dimensional feature space. Assume that the mapping function is $\phi(\mathbf{x})$; being the case, the SVM regression model can be expressed in nonlinear case. Its model is:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle + b. \tag{2.17}$$

We can see that the inner product $\langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle$ needs to be evaluated in the feature space. This is computationally difficult. In order to solve this problem, support vector machines developed a theory based upon Reproducing Kernel Hilbert Space (RKHS) [29].

An inner product in feature space has an equivalent kernel in input space.

$$K(\mathbf{x}, \mathbf{x}_i) = \langle \phi(\mathbf{x}), \phi(\mathbf{x}_i) \rangle . \tag{2.18}$$

Hence, the inner product does not need to be evaluated in feature space.

The nonlinear SVM regression model can be written as:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{x_i}) + b. \tag{2.19}$$

Now, the key problem becomes finding the kernel function. The conditions for a kernel are: $K(\mathbf{x}, \mathbf{x}_i)$ is an admissible kernel function if it produces a kernel matrix that is symmetric and positively semi-definite [29]. This condition can be used to build various kinds of kernels so that we do not need to consider feature space in explicit form. Through using the kernel function, the inner product in feature space is determined directly as a function of the input data in input space [22]. Furthermore, the SVM regression model uses only the support vector in input space. Thus, using the kernel function, only the inner products between support vectors and the input data have to be calculated. The regression model (2.11) can be written as follows when the kernel is used in the SVM model.

$$\begin{aligned} \max W(\alpha, \alpha^*) &= -\frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) \\ &- \varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i (\alpha_i - \alpha_i^*) \end{aligned} \tag{2.20}$$

$$\text{Subject to:} \quad \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0$$
$$\alpha_i, \alpha_i^* \in [0, C].$$

2.4 Implementation of Support Vector Machines Regression

The optimization problem in equation (2.20) can be solved by using standard optimization routines that define the quadratic programming (QP) problem in terms of the Hessian matrix and separate matrixes for the equality and inequality constraint [17].

The Hessian matrix, $\mathbf{H}$, is now a $(2l \times 2l)$ matrix such that

$$\mathbf{H} = \begin{pmatrix} K(\mathbf{x}_i, \mathbf{x}_j) & -K(\mathbf{x}_i, \mathbf{x}_j) \\ -K(\mathbf{x}_i, \mathbf{x}_j) & K(\mathbf{x}_i, \mathbf{x}_j) \end{pmatrix}. \tag{2.21}$$

Assume $\beta = \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix}$. The SVM for regression in matrix formation then becomes:

$$\min \quad \frac{1}{2}\beta^T \mathbf{H}\beta + \begin{bmatrix} \varepsilon \mathbf{1}_l - \mathbf{y} \\ \varepsilon \mathbf{1}_l + \mathbf{y} \end{bmatrix}^T \beta \tag{2.22}$$

$$\text{Subject to: } \begin{bmatrix} \mathbf{1}_l \\ -\mathbf{1}_l \end{bmatrix}^T \beta = 0$$

$$\mathbf{0}_{2l} \leq \beta \leq \frac{C}{l}\mathbf{1}_{2l},$$

where $\mathbf{1}_l$ presents $[1_1, ..., 1_l]^T$. The SVM model is then a quadratic optimization problem with equality and inequality constraints. The solution can be obtained by solving this QP optimization model.

2.5 Three Factors in SVM Regression Models

There are three factors that affect the prediction performance of SVM regression. They are the error term, ε; the regulation factor (also called "additional capacity control"), C; and the type of kernels and the kernel parameter. The selection of model parameters and kernel type is critical to optimizing SVM performance.

2.5.1 The Error Zone Value

The error zone value, ε, controls the width of the ε-insensitive zone. In SVM regression, ε is used to control the noise on the performance of the model. The value of ε affects the number of support vectors that are used to construct the SVM regression model [29]. If the error zone value chosen is larger, fewer support vectors are used to construct the SVM model and the prediction calculation time is shorter. The prediction result may be under-fitting. If the error zone value chosen is smaller, more support vectors are used to construct the SVM model and a longer calculating time is needed. The prediction result may be over-fitting. Therefore, proper selection of the error zone value has a significant effect on SVM regression performance.

2.5.2 The Regulation Factor

Parameter C is the regulation factor of a SVM regression model which determines the trade off between model complexity (flatness) and the degree

to which deviations larger than ε are tolerated in equation (2.2). A large C value assigns higher penalties to errors, thus the regression minimizes error with a lower generalization while a smaller C value, which assigns lower penalties to errors, results in a higher generalization ability [28].

2.5.3 The Kernel Function and the Kernel Parameter

2.5.3.1 Kernel Functions

The type of kernel function used in SVM regression for prediction has significant impact on SVM performance. Normally, which kernel is to be used depends on the properties of the data set [21]. In this chapter, several commonly used kernels are introduced. The SVM model currently includes several kernel functions such as: polynomial function, radial basis function, exponential radial basis function, multi-layer perception, fourier series, splines, B-splines, and additive kernels. Among these kernels, the polynomial kernel and the radial basis function are the most popular since most of prediction work can be solved by either of the two [22]. Moreover, Maria [17] made a mixture of these two kernels. It appears that this new mixed kernel possesses the advantages of both the polynomial and Gaussian radial basis kernels. These three kernels are discussed below.

1. Polynomial Kernel ("poly")

$$K(x, x_i) = \langle x, x_i \rangle^d \quad . \tag{2.23}$$

This type of kernel, classified as a global kernel, gives the support vector model good extrapolation properties [17]. The extrapolation property is the ability to estimate a value for a variable outside a known range from values within a known range by assuming that the estimated value follows logically from the known values. For this reason, the polynomial kernel should be the first choice for regression; however, the parameter of the polynomial kernel is degree d, which could influence the extrapolation property. When d increases, the interpolation ability of the polynomial will increase and the extrapolation ability will decrease [17].

2. Radial Basis Function Kernel ("rbf")

$$K(x, x^*) = \exp\left\{-\frac{|x - x^*|^2}{2\sigma^2}\right\}. \tag{2.24}$$

The parameter is the width, σ, which depends on the scale of the input data [17]. This type of kernel, classified as a local kernel and it gives the support vector machines good interpolation ability [17]. The kernel function has the ability to detect local phenomena.

3. Mixed Kernel ("polyrbf")

$$K_{mix} = \rho K_{poly} + (1 - \rho) K_{rbf}. \tag{2.25}$$

Maria [17] made a mixed kernel which combined polynomial and radial

basis kernels. In this way, the new kernel has both interpolation and extrapolation properties. ρ is the mixing coefficient. Normally, it falls into the range $[0.5 : 0.95]$.

2.5.3.2 Kernel Parameters

The selection of the kernels and kern el parameters can directly affect prediction quality. Two commonly used kernels are the polynomial kernel and the radial basis function. The parameters of these kernels are the degree of the polynomial kernel and the width of the radial basis function. The kernel parameters should be carefully chosen as they implicitly define the structure of the high dimensional feature space, $\phi(x)$, and thus control the complexity of the final solution. Selection of a particular kernel type and the kernel parameter is usually based on application domain knowledge and may reflect distribution of the input value of the training data [29]. No criteria are set regrading how to choose the kernel and the kernel parameter for SVM. It is well known that SVM generalization performance (estimation accuracy) depends on a good setting of the hyper-parameters C, ε, kernel type and the kernel parameter. Model parameters selection is further complicated by the fact that SVM model complexity depends on all three factors [21]. How to determine these three factors is always critical in SVM applications. It is the focus throughout this thesis.

2.6 Applications of the SVM Regression Model

2.6.1 SVM Software Development

There are several software packages available for SVM regression and classification; these include Stefan Ruping's mySVM for Windows and Unix, Chih-Jen Lin's Looms, and Gunn's Matlab Support Vector Machines Toolboxes [39]. Among these, Gunn's SVM Matlab toolbox [9] is the most widely used because it is user friendly. Gunn used two benchmark data sets (Sinc and titanium) to show how SVM regression performs when using this software. In his experiments, the "rbf" kernel parameter is set at from 0.1 to 1. The "poly" kernel parameter is set at from 1 to 5. Gunn presented and compared the prediction results but he did not explain how to determine the parameters.

2.6.2 SVM Regression for Time Series Prediction

Muller et al. applied SVM for time series prediction [18]. In his paper, Muller compared SVM prediction with radial basis function networks. Two benchmark time series: Mackey Glass and Santa Fe Competition were evaluated by applying these two prediction methods. In both cases, SVM performed better still, Muller was unable to explain how to determine the three factors that work best for the SVM model.

2.6.3 SVM Regression in the Mechanical Industry

Jiang and Zuo applied SVM to the mechanical industry [13]. In their paper, they compared SVM prediction with the time series model. Vibration data was used to represent equipment deterioration levels and the "spline" kernel was chosen for SVM regression. The results showed that SVM regression predicted more accurately than the time series model does. The authors did not give reasons for having chosen the "spline" kernel but they did study the impact of the regulation factor on SVM prediction performance. The line search optimization method was used to find the optimal value for the regulation factor. Investigation indicated that SVM would predict best when the regulation factor is infinity but no study was made of how the other parameters, "error zone" and "the kernel parameter" influence SVM regression performance.

Lin *et al* [16] applied support vector machines to condition trend prediction of mechanical equipment based on vibration signals from machines. The one step and 24 steps ahead prediction methods were adopted by using Peak-Peak value of the vibration signal. The radial basis kernel ("rbf") function was adopted. Their experiments showed that the optimal kernel parameter should not be large if the prediction accuracy of the SVM regression model was a consideration. Lin also observed that a large kernel parameter would result in longer calculating time.

2.6.4 SVM Regression in Transportation

Wu *et al* [31] made prediction for travel time, which is one type of time series data using SVM regression. His experiments showed that SVM regression can significantly reduce mean square error and relative mean error significantly compared with traditional predictors such as time series, linear model, and neural networks. Wu demonstrated the feasibility of applying SVM regression in travel time prediction and proved that SVM regression was applicable and performed well for traffic data analysis. No investigation was conducted into the selection of SVM regression model factors.

Vanajakshi and Rilett [27] applied the SVM regression model to traffic speed prediction. Their paper presents a comparison between Neural Networks and the SVM regression for short-term prediction of traffic speed. The Neural Networks model used was a multi-layer feed forward Neural Networks and the SVM regression model used was a support vector regression with radial basis kernel function. The conclusion was that SVM regression performed better than Neural Networks when training data were less in quality and quantity. The paper didn't show how to determine the model parameters.

2.6.5 SVM Regression in the Chemical Industry

Maria applied SVM to the chemical industry [17]. In his thesis, he proposed a new kernel which combined two commonly used kernels, "poly" and "rbf". The new combined kernel is supposed to have both interpolation and

extrapolation properties. Some experiments were done to test this hypothesis. Results showed that the new combined kernel improved SVM regression prediction accuracy in some situations. In Maria's thesis, the "rbf" kernel parameter chosen ranges from 0.1 to 1 and "poly" kernel parameter chosen from 1 to 5. Maria did not give the reasons for these selections.

2.6.6 SVM Regression in Financial Forecasting

Stock market prediction is one example of a challenging financial time-series prediction. Theodore used the SVM regression model for IBM, Yahoo and America Online daily stock price predictions [26]. He compared two different QP algorithms of the SVM regression model, specifically, a primal dual interior point method and the standard QP algorithm. In addition, he compared the SVM regression model with other techniques such as backpropagation and RBF Networks. His conclusion was that the SVM regression was better at prediction than the other two techniques. The model parameter error zone value was simply fixed at zero throughout all the experiments he conducted, without any reasons being given.

Tay and Cao [25] examined the feasibility of SVM in forecasting financial time series, comparing SVM regression with a multi-layer back-propagation Neural Networks. They investigated the prediction variability in the SVM regression performance with respect to the SVM parameters. The conclusion they reached is that the SVM regression model provides a better generalization ability than does BP Networks because SVM regression model is based

on the structural risk minimization principle while BP Network is based on the empirical risk minimization principle. Tay and Cao also concluded that improper selection of model parameters can cause either over-fitting or under-fitting of the training data; because parameters have a significant impact on prediction performance, it is important to develop a structural way of selecting optimum parameters for SVM regression.

Wang and Wu [32] used the SVM regression model with an ε-insensitive loss function to forecast Shanghai stock composite index. BP Networks was used as a benchmark for comparison. Their conclusion was that the forecasting variation of SVM is smaller than that of BP, and the direction forecasting of SVM is more accurate.

Kim [14] applied SVM regression for financial time series forecasting. In his study, SVM regression and BP Networks are compared. SVM regression model outperformed BP Networks. Kim investigated the effect the value of SVM parameters had on his experiment, and concluded that the prediction performance of SVM regression model is sensitive to the value of these parameters. He concluded it is important to find the optimal values for SVM regression model parameters.

Chapter 3

Determination of SVM Regression Parameters

3.1 Proposed Learning Procedure for Selection of Parameters

In this chapter, a learning procedure based on SVM regression theory is given. The purpose of the proposed learning procedures is to find the optimal combination of error zone value, the kernel parameter and suitable kernel function since these three factors control SVM model prediction performance. Kernel function is chosen from among the "rbf", "poly", and"polyrbf" kernels as these three kernels have been theoretically proved able to represent most unknown functions. Moreover, these three kernels have been documented widely throughout the literature as providing excellent predictions.

The learning procedure involves four steps. Following this learning procedure, experiments are conducted and the results are presented numerically and in figures. The learning procedure is given below:

- Step 1: Choosing loss function: the ε-insensitive loss function is the first choice because it has sparse property over the other loss functions.

- Step 2: Selecting a kernel function from among the "rbf","poly" and "polyrbf" kernels.

- Step 3: Using the line search optimization method to find the optimal error zone, ε, and kernel parameter, σ. A program with two loops is compiled. One loop is used for optimizing ε, the other loop is used for optimizing σ.

  Parameter C is defined as infinite. This is based on Fan's recommendation. His thesis proved that SVM regression model obtained the best prediction results with a infinite regulation factor [12].

- Step 4: Adopting "mean square error" as a criterion for evaluating regression model performance. There are several statistical measures that are used to assess the overall performance of a regression model. These include error statistics like the mean square error and standard deviation, and traditional statistics like the residual analysis [7]. Among these, error statistics, "mean square error" is the most widely used in model regression assessment. It is defined as:

$$mse = \frac{\sum_{i=1}^{l} (y_i - \hat{y}_i)^2}{n}, \tag{3.1}$$

  where y is the observed data and $\hat{y}$ is the prediction data. The value

n is the size of the data set. Many works on SVM regression have used this error statistics as benchmark representing SVM regression performance. The mean square error is also used in this thesis since it makes easier to compare the prediction results from SVM regression with others obtained by other methods. The searching process has indicated that the optimal parameters can be found in terms of the smallest "mse". As the searching goes on, the "mse" values in consecutive iterations become very close. As a result, the "mse" ratio has been adopted as the stopping criterion. It is expressed by the following equation:

$$mse\ ratio = \frac{mse_n - mse_{n+1}}{mse_n}, \tag{3.2}$$

where mse_n represents the "mse" obtained in iteration n and mse_{n+1} represents the "mse" obtained in iteration $n+1$. Two prediction results are regarded as identical when the "mse" ratio between them is smaller than 0.01. In this situation, the experiment stops, and the corresponding parameters at the n^{th} iteration are regarded as the optimal parameters.

3.2 Data Analysis

Experiments based on the learning procedure are conducted and the results are presented numerically or in figures. Among the parameters, the model error zone value and kernel parameter are studied in detail. One program

with two loops is compiled to find the optimal value for the two parameters in terms of the smallest "mse".

The line search optimization method is adopted in every loop. A very simple version of this method is used. The starting point and the step size are fixed in every cycle. Later this thesis refers to this version of the line search optimization method as the "Fixed Step Size" method. Three data sets are used in the experiments.

3.2.1 The Sinc Data Set

The first data set is the Sinc data set. It comes from the Sinc function:

$$Sinc(x) = \frac{\sin(\pi x)}{\pi x}; \qquad x \in [-10, 10].$$

The "Sinc" function, also called "sample function", arises frequently in signal processing and the theory of Fourier transform. It is widely used to test the performance of regression models. The step size of x is 0.4. There are, in total, 51 data points in the Sinc data set. The accuracy of the input target value $[Sinc(x)]$ for this data set is 10^{-5}.

A set of prediction experiments are conducted using this data set. In the experiments, the "Radial basis function" kernel is chosen based on the prediction accuracy. The "Fixing Step Size" optimization method is used in order to find the optimal error zone value and kernel parameter simultaneously. The program used in this set of experiments includes two loops. One loop is used to find the optimal error zone value. The other loop is used to

find optimal kernel parameter. Based on the reports in the literature, the kernel parameter is usually in the range from 1 to 10. The error zone value could be chosen from a wide range. The ranges searched first are: [1, 10] with a step size 1, [0.1, 2] with a step size 0.1, and [0.01, 0.2] with a step size 0.01. The values 1, 0.1 and 0.01 are found to be the error zone values corresponding to the smallest "mse" in the corresponding ranges. Unfortunately, the prediction results at these three values are very bad because the prediction line is almost flat, whereas the plot of the Sinc data set is a curve. Analyzing the reasons for this, SVM regression theory can be of some assistance. In the SVM regression model the error zone value is used to control prediction noise. It seems that the error zone value given here is too large to play its role.

Since 0.01 is the smallest error zone value searched so far, the decision was made to search for the optimal error zone value in a smaller range, [0, 0.02] with a step size 0.001. The prediction results are illustrated in Tables 3.1 and 3.2 (see pages 41 and 42). The smallest "mse" (0.0034) was obtained when ε equalled 0.001 and σ equalled 3. The experiment also offered us two important insight. First, the "mse" value is relatively large when the error zone value is zero. Thus, we can't simply fix the optimal error zone value at zero. Second, the smaller the error zone value, the smaller the "mse" value. As a result, the search work for the optimal error zone value continued.

The range in which to search for the optimal error zone value was refined. Around the value 0.001, the new searching range was [0, 0.002], and

the step size was 10^{-4}. The result is shown in Tables 3.3 and 3.4 (see pages 43 and 44). The smallest "mse" (4.9325×10^{-4}) was obtained when ε equalled 10^{-4} and σ equalled 3. This prediction result is obviously better than the last one. According to Equation (3.2), the "mse" ratio was $(0.0034 - 4.9325 \times 10^{-4})/0.0034$=0.85. The search continued because the ratio was still pretty large.

Since the smallest "mse" was obtained at the error zone value of 10^{-4} in the previous iteration, the new searching range was refined to $[0, 2.0\times10^{-4}]$, with a the step size of 10^{-5}. Tables 3.5 and 3.6 (see pages 45 and 46) illustrate the results that the smallest "mse" (2.7088×10^{-4}) was obtained when ε equalled 10^{-5}, and σ equalled 3. The ratio between the "mse" obtained in this iteration and the previous iteration was $(4.9325 \times 10^{-4} - 2.7088 \times 10^{-4})/(4.9325 \times 10^{-4})$ =0.45, still greater than the required precision of 0.01. The prediction obviously improved during the searching process. The search continued.

The new searching range was refined to $[0, 2.0 \times 10^{-5}]$ with a step size of 10^{-6}. The result is illustrated in Tables 3.7 and 3.8 (see pages 47 and 48). When ε equalled 10^{-6}, σ equalled 3, the smallest "mse" (2.5684×10^{-4}) was obtained. The ratio between this "mse" value and the last "mse" value was $(2.7088 \times 10^{-4} - 2.5684 \times 10^{-4})/(2.7088 \times 10^{-4})$=0.05, still greater than 0.01.

It is apparent that the "mse" ratio becomes smaller as the error zone value becomes smaller. Based on this observation, the search for the optimal error zone value continued because the stopping criterion had not been reached

yet.

The new searching range was refined to $[0, 2 \times 10^{-6}]$ with a step size of 10^{-7}. The prediction result is presented in Tables 3.9 and 3.10 (see pages 49 and 50). The smallest "mse" 2.5552×10^{-4} was obtained when ε equalled 10^{-7} and σ equalled 3. The ratio between the two smallest "mse" values was $(2.5684 \times 10^{-4} - 2.5552 \times 10^{-4})/(2.5684 \times 10^{-4}) = 0.0051$ according to Equation (3.2). Since this is smaller than 0.01, the prediction results between these two searching cycles are very close. According to the stopping criterion, the predicting results of $\varepsilon = 10^{-7}$ and $\varepsilon = 10^{-6}$ are considered virtually identical, the search stopped here. Figure 3.1 shows the prediction result when the error zone value is 10^{-6} and 10^{-7} respectively. Clearly the prediction curves have almost merged together. Because there can be no significant improvement in prediction accuracy there is no need to try smaller error zone values for this data.

As has already been stated, the smaller the error zone value and the more support vectors there are, the longer the computation time is. Here, experiment has shown that there is no significant different between the prediction accuracy obtained with an error zone value of 10^{-6} and 10^{-7}. However, the predicting time when the error zone value is 10^{-7} is longer than that when the error zone value is 10^{-6}. Considering both the predicting time and prediction accuracy, fixing the error zone value at 10^{-6} is reasonable. We call this value the optimal error zone value for this data set. We also found that the error zone value is closely related to the accuracy of the input target

value. The accuracy of the input target value of the Sinc data set is 10^{-5}. The optimal error zone value is equal to the one-tenth of the accuracy of the input target value.

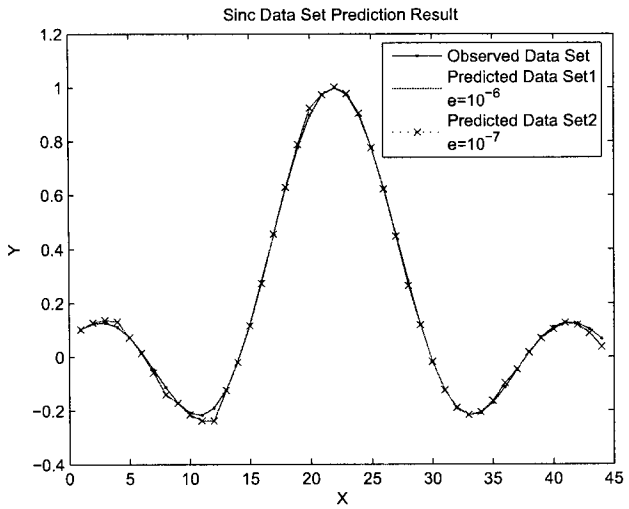

Figure 3.1: Sinc Data Set Prediction Result

3.2.2 The Sunspot Data Set

The Sunspot data set, the record of solar magnetic activity, has been studied by other researchers as a non-stationary time series, using various kinds of prediction models such as the ARMA model and the Neural Networks model [40]. There are 459 data points in the Sunspot data set. The accuracy of the input target value is 1. In the experiment, the "polyrbf" kernel was chosen because neither the "poly" kernel nor the "rbf" kernel alone could give

good prediction results. The experiment shows that only when the kernel parameter is fixed at 1, does the prediction program run normally. Otherwise, the predicting program breaks down. Thus only one loop is needed for the program to find its optimal error zone value. In searching for the optimal error zone value, the range [10, 100] with step size 10 was tried first. The prediction results were not good because the distance between the predicted curve and the observed curve was very great. This means that the error zone value given here must be too large to control the prediction noise. We decided to start the search for the optimal error zone value in the range [0, 11], with a step size of 1. The results are illustrated in Table 3.11 (see page 51). When the ε equalled 1, the smallest "mse" (144.48) was obtained. From the prediction result, we found the "mse" value was relatively large when the error zone value was zero. The optimal error zone value can not simply be fixed at zero. We also noticed that for this data set, the smaller the error zone value that was given, the smaller the "mse" value we could obtain. Hence, the search for the error zone value continued.

A new searching range was refined around 1, it was [0, 2], with a step size of 0.1. Table 3.12 (see page 51) illustrates the results. When ε equalled 0.1, the smallest "mse" (142.03) was obtained. This prediction result was obviously better than the previous one as can be seen by comparing the smallest "mse" values. The ratio between the two "mse" values is $(144.48 - 142.03)/144.48 = 0.017$ according to Equation (3.2). Based on experience ob-

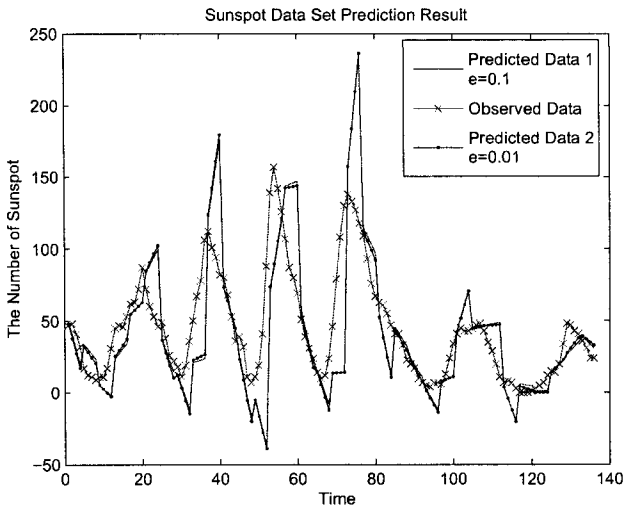

Figure 3.2: Sunspot Data Set Predicting Result

tained from the Sinc data set experiment, the smaller the error zone value searched, the smaller the "mse" ratio. The search continued.

The next searching range needed to be around 0.1. The new refined searching range was [0, 0.2]. The starting point of the error zone was set at zero, and the step size of the error zone was set at 0.01. Table 3.13 (see page 51) illustrates this experiment's results. When ε equalled 0.01, the smallest "mse" (141.79) was obtained. We can see that there is litter difference between the smallest "mse" values when the error zone is equal to 0.1 and 0.01 respectively. The ratio between these two "mse" values is $(142.03 - 141.79)/142.03 = 0.002$, smaller than the required precision of 0.01. The two predicting results can be regarded as the same, so the search stopped

here. Figure 3.2 shows the prediction results when the error zone values are 0.1 and 0.01 respectively. We can see that the predicted curves at 0.1 and 0.01 have almost merged together. Under this circumstances, there is no need to continue the search for the optimal error zone value. With the increase of searching iterations, the "mse" ratio will continue to decrease, so there will be no significant improvement of the prediction accuracy. On the other hand, the predicting time when the error zone value is 0.01 is longer than that when the error zone value is 0.1. Based on this observation, we may say that the optimal error zone value should be 0.1 in terms of both prediction accuracy and calculating time. We know the accuracy of the input target value is 1. For this data set, the optimal error zone value is equal to one-tenth of the input target value.

3.2.3 The Vibration Data "Rmsv"

The third data set is processed one-dimensional vibration data which is called the vertical root mean square (Rmsv) data [38]. It is from a gearbox life-time assessment experiment conducted in the Reliability Research Lab at the University of Alberta. We used the Rmsv data from points 750~986 as the input data. The accuracy of the input target value of the Rmsv data set is 10^{-4}.

Similar to the last two data sets, the "poly" kernel gave the best prediction result among the three kernels used in the experiments. The kernel

parameter was fixed at 1 since the range of "poly"kernel parameter is very limited. In most cases the "poly" kernel parameter value offers only two choices, either 1 or 2. Otherwise, the Hessian matrix of the SVM quadratic optimization programming (QP) will not be positively definite and the implementation of the SVM quadratic optimization programming will break down. In this situation, a prediction program with one loop is used to find the optimal error zone value. Since the accuracy of the input target value of the Rmsv data set is very small, and based on experience from other data sets, the small value range of [0, 1] with a step size 0.1 was searched first. At the value zero, the smallest "mse" was obtained. We found that the prediction results were not good - the distance between the predicted curve and observed curve was very large; and "mse" value was large too. As a result, the optimal error zone value was searched in the range [0, 0.020] with a step size of 0.001. The prediction result is illustrated in Table 3.14 (see page 52). The result shows that at ε=0, "mse" $S_{mse} = 1.0893 \times 10^{-7}$. Based on the experience obtained from the last two data sets, the optimal error zone value should not be zero. This result implies that the error zone value chosen from this range was too large and had lost the ability to control the prediction noise. In this case, an error zone 0 controls the prediction accuracy, but this prediction result was not what we want and so the search for the optimal error zone value continued.

A new searching range was refined to [0, 0.001] with a step size of 10^{-4}. Table 3.15 (see page 52) illustrates the prediction results. This result shows

that when the error zone value was 10^{-4}, the smallest "mse" (5.4658×10^{-8}) was obtained. This prediction result was obviously better than the previous one, but still the searching process continued.

Since the new search range should be around 10^{-4}, it was refined to [0, 2×10^{-4}], with a step size of 10^{-5}. Table 3.16 (see page 53) illustrates the experiment results. When ε equalled 10^{-5}, the smallest "mse" (4.8233×10^{-9}) was obtained. The ratio between this "mse" value and the previous one was: $(5.4658 \times 10^{-8} - 4.8233 \times 10^{-9})/(5.4658 \times 10^{-8}) = 0.90$ according to Equation (3.2). The search continued and it is clear that the "mse" value becomes smaller as the error zone value does.

A new searching range was refined to [0, 2.0×10^{-5}], with a step size of 10^{-6}. Table 3.17 (see page 53) illustrates the prediction results. When the error zone value is equal to 9.0×10^{-6}, the smallest "mse" (4.8086×10^{-9}) is obtained. The ratio between this "mse" value and the previous one was: $(4.8233 \times 10^{-9} - 4.8086 \times 10^{-9})/(4.8233 \times 10^{-9})$ =0.003. It is smaller than the required precision of 0.01. This means that there is no significant difference between the two "mse" values in terms of prediction accuracy. The prediction results when the error zone is 10^{-5} and 9.0×10^{-6} respectively are considered identical. The experiment for finding the optimal error zone value is stopped because the "mse" ratio will decrease as the search goes on. As a result, the prediction results would not show significant improvement if the search continued. Figure 3.3 shows the prediction results at the error zone values of 10^{-5} and 9.0×10^{-6} respectively. It can be clearly seen that the

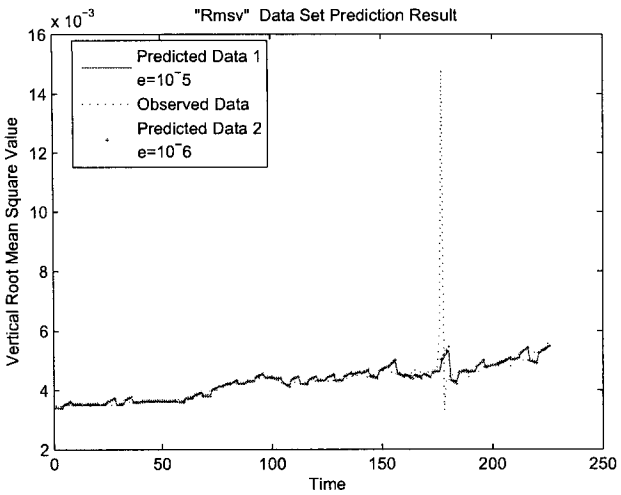

Figure 3.3: "Rmsv" Data Set Prediction Result

predicted curves merged together when the error zone values are 10^{-5} and 9.0×10^{-6}. However, the predicting time changes as different error zone values are taken into the SVM regression model. The smaller the error zone value, the longer the predicting time. Taking the predicting time into account, 10^{-5} is considered to be the optimal error zone value for this data set. As well since the accuracy of the input target value is 10^{-4}, the optimal error zone value equals to one-tenth of the accuracy of input target value.

3.3 Discussion of the Error Zone Value

From the above experiments, we can see that the optimal error zone, ε, is closely related to the accuracy of the input target value. In SVM theory,

the model parameter, ε, is called the error zone value; it is used to control the noise between the predicted data and the real data [30]. For every data point in the data set, the predicted noise (also called predicted errors) can be presented in the following formula:

$$\hat{\delta}_i = \hat{y}_i - y_i.$$

The average noise of the whole data set is:

$$\hat{\delta} = \frac{1}{n}(\delta_1 + \delta_2 + ... + \delta_n).$$

In the SVM regression model, the error zone value depends on the average noise, $\hat{\delta}$ [15]. That is to say the error zone value is closely related to the input target value, y_i, and the predicting target value, $\hat{y}_i$. Through literature review and our experiments, we have found that the predicted performance of SVM regression is more accurate than other predictors and hence is superior to them. This means that the predicted target value, $\hat{y}_i$, is very close to the input target value, y_i. It is possible that the noise between these two values is related to the accuracy of the input target value, which is the only information available before a prediction is made. In accordance with SVM regression theory, the experiments show that the error zone value depends on the accuracy of the input target value. The rule of thumb summarized here is: the optimal error zone value is equal to one-tenth of the accuracy of input target value.

Table 3.1: Sinc Data Set Prediction Result 1-1: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	1	2	3	4	5
0.000	0.0832	0.0786	0.0881	0.0948	0.0989
0.001	0.0363	0.0059	**0.0034**	0.0056	0.0075
0.002	0.0401	0.0071	0.0036	0.0048	0.0061
0.003	0.0434	0.0086	0.0041	0.0046	0.0055
0.004	0.0466	0.103	0.0052	0.0052	0.0057
0.005	0.0500	0.0121	0.0066	0.0061	0.0064
0.006	0.0532	0.0142	0.0082	0.0074	0.0075
0.007	0.0539	0.0164	0.0101	0.0090	0.0089
0.008	0.0546	0.0185	0.0123	0.0110	0.0107
0.009	0.0552	0.0204	0.0147	0.0132	0.0129
0.010	0.0557	0.0224	0.0163	0.0151	0.0149
$\varepsilon \setminus \sigma$	1	2	3	4	5
0.011	0.0564	0.0243	0.0181	0.0166	0.0163
0.012	0.0570	0.0260	0.0203	0.0185	0.0180
0.013	0.0576	0.0280	0.0227	0.0208	0.0201
0.014	0.0582	0.0289	0.0250	0.0232	0.0225
0.015	0.0588	0.0294	0.0268	0.0254	0.0249
0.016	0.0595	0.0296	0.0266	0.0269	0.0271
0.017	0.0601	0.0298	0.0264	0.0265	0.0269
0.018	0.0608	0.0300	0.0262	0.0262	0.0265
0.019	0.0615	0.0303	0.0260	0.0259	0.0261
0.020	0.0621	0.0305	0.0259	0.0256	0.0257

Table 3.2: Sinc Data Set Prediction Result 1-2: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	6	7	8	9	10
0.000	0.1014	0.1031	0.1042	0.1050	0.1056
0.001	0.0091	0.0100	0.0104	0.0106	0.0107
0.002	0.0070	0.0077	0.0081	0.0085	0.0087
0.003	0.0063	0.0068	0.0072	0.0074	0.0076
0.004	0.0062	0.0066	0.0069	0.0071	0.0072
0.005	0.0068	0.0071	0.0073	0.0074	0.0075
0.006	0.0077	0.0079	0.0081	0.0082	0.0083
0.007	0.0090	0.0092	0.0093	0.0093	0.0094
0.008	0.0107	0.0108	0.0109	0.0109	0.0109
0.009	0.0128	0.0128	0.0128	0.0129	0.0129
0.010	0.0148	0.0149	0.0149	0.0150	0.0151
$\varepsilon \setminus \sigma$	6	7	8	9	10
0.011	0.0162	0.0162	0.0162	0.0162	0.0163
0.012	0.0179	0.0178	0.0178	0.0178	0.0178
0.013	0.0198	0.0197	0.0197	0.0197	0.0196
0.014	0.0222	0.0220	0.0220	0.0219	0.0219
0.015	0.0247	0.0246	0.0245	0.0244	0.0244
0.016	0.0269	0.0268	0.0267	0.0267	0.0266
0.017	0.0272	0.0274	0.0276	0.0277	0.0278
0.018	0.0268	0.0269	0.0271	0.0272	0.0273
0.019	0.0263	0.0265	0.0267	0.0267	0.0268
0.020	0.0259	0.0260	0.0260	0.0263	0.0263

Table 3.3: Sinc Data Set Prediction Result 2-1: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	1	2	3	4	5
0.0000	8.3204e-002	7.8549e-002	8.8080e-002	9.4801e-002	9.8892e-002
1.0e-004	3.2716e-002	2.1343e-003	**4.9325e-004**	7.7552e-004	1.1004e-003
2.0e-004	3.3160e-002	2.6280e-003	8.4367e-004	1.1191e-003	1.3808e-003
3.0e-004	3.3611e-002	3.1606e-003	1.3538e-003	1.6890e-003	1.9101e-003
4.0e-004	3.4026e-002	3.6391e-003	1.9098e-003	2.4628e-003	2.6849e-003
5.0e-004	3.4387e-002	4.1454e-003	2.5529e-003	3.4189e-003	3.6622e-003
6.0e-004	3.4753e-002	4.6324e-003	2.9980e-003	4.5338e-003	4.8397e-003
7.0e-004	3.5125e-002	4.8861e-003	3.4423e-003	4.9702e-003	6.1452e-003
8.0e-004	3.5501e-002	5.0854e-003	3.4034e-003	5.3621e-003	7.1939e-003
9.0e-004	3.5882e-002	5.4583e-003	3.3315e-003	5.7652e-003	7.4519e-003
1.0e-003	3.6268e-002	5.8554e-003	3.4321e-003	5.5695e-003	7.5222e-003
$\varepsilon \setminus \sigma$	1	2	3	4	5
1.1e-003	3.6659e-002	6.0040e-003	3.5246e-003	5.4628e-003	7.3347e-003
1.2e-003	3.7049e-002	6.1201e-003	3.5071e-003	5.3293e-003	7.0971e-003
1.3e-003	3.7438e-002	6.2385e-003	3.4940e-003	5.2318e-003	6.8778e-003
1.4e-003	3.7852e-002	6.3591e-003	3.4853e-003	5.1395e-003	6.7373e-003
1.5e-003	3.8230e-003	6.4819e-003	3.4809e-003	5.0524e-003	6.6027e-003
1.6e-003	3.8603e-002	6.6070e-003	3.4955e-003	4.9707e-003	6.4739e-009
1.7e-003	3.8979e-002	6.7343e-003	3.5178e-003	4.8943e-003	6.4033e-003
1.8e-003	3.9359e-002	6.8639e-003	3.5437e-003	4.8380e-003	6.2335e-003
1.9e-003	3.9742e-003	6.9958e-003	3.5731e-003	4.7971e-003	6.1023e-003
2.0e-003	4.0129e-002	7.1298e-003	3.6062e-003	4.7605e-003	6.0615e-003

Table 3.4: Sinc Data Set Prediction Result 2-2: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	6	7	8	9	10
0.0000	1.0144e-001	1.0310e-001	1.0422e-001	1.0501e-001	1.0558e-001
1.0e-004	1.3419e-003	1.5191e-003	1.6614e-003	1.7973e-003	1.9560e-003
2.0e-004	1.5817e-003	1.7377e-003	1.8773e-003	2.0311e-003	2.2349e-003
3.0e-004	2.0802e-003	2.2204e-003	2.3563e-003	2.5208e-003	2.7507e-003
4.0e-004	2.8523e-003	2.9875e-003	3.1232e-003	3.2933e-003	3.5348e-003
5.0e-004	3.8354e-003	3.9734e-003	4.1110e-003	4.2827e-003	4.5252e-003
6.0e-004	5.0300e-003	5.1752e-003	5.3140e-003	5.4814e-003	5.7126e-003
7.0e-004	6.4328e-003	6.5894e-003	6.7288e-003	6.8861e-003	7.0940e-003
8.0e-004	7.5855e-003	7.8488e-003	8.0597e-003	8.2423e-003	8.4522e-003
9.0e-004	8.6930e-003	8.9839e-003	9.1964e-003	9.3772e-003	9.5660e-003
1.0e-003	9.0805e-003	9.9704e-002	1.0410e-002	1.0574e-002	1.0725e-002
$\varepsilon \setminus \sigma$	6	7	8	9	10
1.1e-003	8.7124e-003	9.8059e-003	1.0580e-002	1.1142e-002	1.1558e-002
1.2e-003	8.2995e-003	9.3379e-003	1.0076e-002	1.0611e-002	1.1009e-002
1.3e-003	8.0815e-003	8.8605e-003	9.5668e-003	1.0080e-002	1.0462e-002
1.4e-003	7.8693e-003	8.6500e-003	9.2039e-003	9.5742e-003	9.9356e-003
1.5e-003	7.7059e-003	8.4683e-003	9.0043e-003	9.3911e-003	9.6776e-003
1.6e-003	7.5486e-003	8.2928e-003	8.8166e-003	9.1948e-003	9.5023e-003
1.7e-003	7.3974e-003	8.1236e-003	8.6352e-003	9.0049e-003	9.2789e-003
1.8e-003	7.2522e-003	7.9605e-003	8.4601e-003	8.8214e-003	9.0893e-003
1.9e-003	7.1132e-003	7.8037e-003	8.2914e-003	8.6443e-003	8.9062e-003
2.0e-003	7.0118e-003	7.6755e-003	8.1447e-003	8.4845e-003	8.7367e-003

Table 3.5: Sinc Data Set Prediction Result 3-1: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	1	2	3	4	5
0.00000	8.3204e-002	7.8549e-002	8.8080e-002	9.4801e-002	9.8892e-002
1.0e-005	3.2294e-002	1.6414e-003	**2.7088e-004**	6.5881e-004	1.0592e-003
2.0e-005	3.2341e-002	1.6920e-003	2.8822e-004	6.06277e-004	1.0011e-003
3.0e-005	3.2389e-002	1.7436e-003	3.0741e-004	6.6898e-004	1.0511e-003
4.0e-005	3.2437e-002	1.7962e-003	3.2843e-004	6.7748e-004	1.0507e-003
5.0e-005	3.2485e-002	1.8499e-003	3.5130e-004	6.8816e-004	1.0528e-003
6.0e-005	3.2533e-002	1.9047e-003	3.7601e-004	7.0113e-004	1.0574e-003
7.0e-005	3.2581e-002	1.9605e-003	4.0256e-004	7.1635e-004	1.0645e-003
8.0e-005	3.2628e-002	2.0174e-003	4.3095e-004	7.3382e-004	1.0740e-003
9.0e-005	3.2672e-002	2.0753e-003	4.6118e-004	7.5354e-004	1.0860e-003
1.0e-004	3.2716e-002	2.1343e-003	4.9325e-004	7.7552e-004	1.1004e-003
$\varepsilon \setminus \sigma$	1	2	3	4	5
1.1e-004	3.2762e-002	2.1943e-003	5.2716e-004	709975e-004	1.1173e-003
1.2e-004	3.2804e-002	2.2554e-003	5.6292e-004	8.2622e-004	1.1367e-003
1.3e-004	3.2849e-002	2.3108e-003	6.0051e-004	8.5495e-004	1.1586e-003
1.4e-004	3.2893e-002	2.3603e-003	6.3995e-004	8.8593e-004	1.1829e-003
1.5e-004	3.2937e-002	2.4081e-003	6.8123e-004	9.1916e-004	1.2097e-003
1.6e-004	3.2982e-002	2.4499e-003	7.1556e-004	9.5464e-004	1.2390e-003
1.7e-004	3.3026e-002	2.4881e-003	7.4530e-004	9.9238e-004	1.2708e-002
1.8e-004	3.3071e-002	2.5341e-003	707657e-004	1.0324e-003	1.3050e-003
1.9e-004	3.3115e-002	2.5808e-003	8.0936e-004	1.0746e-003	1.3416e-003
2.0e-004	3.3160e-002	2.6280e-003	8.4367e-004	1.1191e-003	1.3808e-003

Table 3.6: Sinc Data Set Prediction Result 3-2: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	6	7	8	9	10
0.00000	1.0144e-001	1.0310e-001	1.0422e-001	1.0501e-001	1.0558e-001
1.0e-005	1.3473e-003	1.5480e-003	1.6921e-003	1.8056e-003	1.9182e-003
2.0e-005	1.3363e-003	1.5343e-003	1.6782e-003	1.7944e-003	1.9132e-003
3.0e-005	1.3280e-003	1.5231e-003	1.6669e-003	1.7858e-003	1.9106e-003
4.0e-005	1.3222e-003	1.5146e-003	1.6582e-003	1.7798e-003	1.9105e-003
5.0e-005	1.3190e-003	1.5088e-003	1.6522e-003	1.7763e-003	1.9127e-003
6.0e-005	1.3184e-003	1.5056e-003	1.6487e-003	1.7754e-003	1.9173e-003
7.0e-005	1.3204e-003	1.5050e-003	1.6480e-003	1.7770e-003	1.9243e-003
8.0e-005	1.350e-003	1.5071e-003	1.6498e-003	1.7812e-003	1.9338e-003
9.0e-005	1.3322e-003	1.5118e-003	1.6543e-003	1.7879e-003	1.9456e-003
1.0e-004	1.3419e-003	1.5191e-003	1.6614e-003	1.7973e-003	1.9600e-003
$\varepsilon \setminus \sigma$	6	7	8	9	10
1.1e-004	1.3542e-003	1.5291e-003	1.6711e-003	1.8091e-003	1.9766e-003
1.2e-004	1.3692e-003	1.5417e-003	1.6835e-003	1.8236e-003	1.9956e-003
1.3e-004	1.3869e-003	1.5570e-003	1.6985e-003	1.8405e-003	2.0171e-003
1.4e-004	1.4068e-003	1.5749e-003	1.7162e-003	1.8601e-003	2.0410e-003
1.5e-004	1.4295e-003	1.5954e-003	1.7364e-003	1.8822e-003	2.0673e-003
1.6e-004	1.4547e-003	1.6186e-003	1.7593e-003	1.9069e-003	2.0960e-003
1.7e-004	1.4826e-003	1.6444e-003	1.7849e-003	1.9341e-003	2.1271e-003
1.8e-004	1.5131e-003	1.6728e-003	1.8130e-003	1.9639e-003	2.1606e-003
1.9e-004	1.5461e-003	1.7039e-003	1.8438e-003	1.9962e-003	2.1966e-003
2.0e-004	1.5817e-003	1.7377e-003	1.8773e-003	2.0311e-003	2.2349e-003

Table 3.7: Sinc Data Set Prediction Result 4-1: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	1	2	3	4	5
0.000000	8.3204e-002	7.8549e-002	8.8080e-002	9.4801e-002	9.8892e-002
1.0e-006	3.2250e-002	1.5968e-003	**2.5684e-004**	6.5716e-004	1.0661e-003
2.0e-006	3.2255e-002	1.6017e-003	2.5839e-004	6.5726e-004	1.0652e-003
3.0e-006	3.2260e-002	1.6066e-003	2.583e-004	6.5737e-004	1.0644e-003
4.0e-006	3.2265e-002	1.6115e-003	2.6136e-004	6.5751e-004	1.0636e-003
5.0e-006	3.2270e-003	1.6165e-003	2.6290e-004	6.5767e-004	1.0628e-003
6.0e-006	3.2275e-003	1.6215e-003	2.6446e-004	6.5785e-004	1.0620e-003
7.0e-006	3.2279e-003	1.6264e-003	2.6603e-004	6.5806e-004	1.0613e-003
8.0e-006	3.2289e-003	1.6364e-003	2.6763e-004	6.5828e-003	1.0606e-003
9.0e-006	3.2289e-003	1.6364e-003	2.6924e-004	6.5853e-004	1.0599e-003
1.0e-005	3.2293e-003	1.6414e-003	2.7088e-004	6.5881e-004	1.0592e-003
$\varepsilon \setminus \sigma$	1	2	3	4	5
1.1e-005	3.2298e-003	1.6464e-003	2.7253e-004	6.5910e-004	1.0586e-003
1.2e-005	3.2303e-003	1.6514e-003	2.7420e-004	6.5942e-004	1.0579e-003
1.3e-005	3.2308e-003	1.6564e-003	2.7589e-004	6.5976e-004	1.0574e-003
1.4e-005	3.2313e-003	1.6615e-003	2.7760e-004	6.6012e-004	1.0579e-003
1.5e-005	3.2317e-003	1.6665e-003	2.7932e-004	6.6054e-004	1.0562e-003
1.6e-005	3.2322e-003	1.6716e-003	2.106e-004	6.6091e-004	1.0557e-003
1.7e-005	3.2327e-003	1.6767e-003	2.8283e-004	6.6134e-004	1.0552e-003
1.8e-005	3.2332e-003	1.6818e-003	2.8461e-004	6.6180e-004	1.0548e-003
1.9e-005	3.2337e-003	1.6868e-003	2.8640e-004	6.6227e-004	1.05432e-003
2.0e-005	3.2341e-003	1.6919e-003	2.8822e-004	6.6277e-004	1.0539e-003

Table 3.8: Sinc Data Set Prediction Result 4-2: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	6	7	8	9	10
0.000000	1.0144e-001	1.0310e-001	1.0422e-001	1.0501e-001	1.0558e-001
1.0e-006	1.3593e-003	1.5627e-003	1.7069e-003	1.8178e-003	1.9248e-003
2.0e-006	1.3579e-003	1.5609e-003	1.7052e-003	1.8163e-003	1.9240e-003
3.0e-006	1.3565e-003	1.5592e-003	1.7034e-003	1.8149e-003	1.9232e-003
4.0e-006	1.3551e-003	1.5575e-003	1.7017e-003	1.8135e-003	1.9223e-003
5.0e-006	1.3537e-003	1.5559e-003	1.7001e-003	1.8121e-003	1.9216e-003
6.0e-006	1.3524e-003	1.5543e-003	1.6984e-003	1.8107e-003	1.9209e-003
7.0e-006	1.3510e-003	1.5527e-003	1.6968e-003	1.8094e-003	1.9202e-003
8.0e-006	1.3498e-003	1.5511e-003	1.6952e-003	1.8081e-003	1.9195e-003
9.0e-006	1.3485e-003	1.5495e-003	1.6937e-003	1.8068e-003	1.9189e-003
1.0e-005	1.3473e-003	1.5480e-003	1.6921e-003	1.8056e-003	1.9182e-003
$\varepsilon \setminus \sigma$	6	7	8	9	10
1.1e-005	1.3460e-003	1.5465e-003	1.6906e-003	1.8043e-003	1.9176e-003
1.2e-005	1.3449e-003	1.5451e-003	1.6891e-003	1.8031e-003	1.9170e-003
1.3e-005	1.3437e-003	1.5436e-003	1.6877e-003	1.8020e-003	1.9165e-003
1.4e-005	1.3426e-003	1.5422e-003	1.6862e-003	1.8008e-003	1.9159e-003
1.5e-005	1.3415e-003	1.5408e-003	1.6848e-003	1.8000e-003	1.9154e-003
1.6e-005	1.3404e-003	1.5394e-003	1.6834e-003	1.7986e-003	1.9149e-003
1.7e-005	1.3393e-003	1.5381e-003	1.6821e-003	1.7975e-003	1.9145e-003
1.8e-005	1.3383e-003	1.5368e-003	1.6808e-003	1.7964e-003	1.9140e-003
1.9e-005	1.3373e-003	1.5355e-003	1.6795e-003	1.7954e-003	1.9136e-003
2.0e-005	1.3363e-003	1.5343e-003	1.6782e-003	1.7944e-003	1.9132e-003

Table 3.9: Sinc Data Set Prediction Result 5-1: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	1	2	3	4	5
0	8.3204e-002	7.8549e-002	8.8080e-003	9.4801e-003	9.8892e-002
1.0e-007	3.2246e-002	1.5924e-003	**2.5552e-004**	6.5710e-003	1.0669e-003
2.0e-007	3.2247e-002	1.5929e-003	2.5567e-004	6.5711e-003	1.0008e-003
3.0e-007	3.2267e-002	1.6014e-003	2.5698e-004	7.2200e-003	1.0668e-003
4.0e-007	3.2248e-002	1.5938e-003	2.5596e-004	6.5712e-003	1.0666e-003
5.0e-007	3.2248e-002	1.5943e-003	2.5611e-004	6.5713e-004	1.0665e-003
6.0e-007	3.2248e-002	1.5948e-003	2.5625e-003	6.5713e-004	1.0664e-003
7.0e-007	3.2249e-002	1.5953e-003	2.5640e-004	6.5714e-004	1.0663e-003
8.0e-007	3.2250e-002	1.5958e-003	2.5655e-004	6.5715e-004	1.0663e-003
9.0e-007	3.2250e-002	1.5963e-003	2.5670e-004	6.5716e-004	1.0662e-003
1.0e-006	3.2251e-002	1.5968e-003	2.5684e-004	6.5716e-004	1.0661e-003
$\varepsilon \setminus \sigma$	1	2	3	4	5
1.1e-006	3.2251e-002	1.5973e-003	2.5699e-004	6.5717e-004	1.0660e-003
1.2e-006	3.225e-002	1.598e-003	2.5714e-004	6.5718e-004	1.0659e-003
1.3e-006	3.225e-002	1.5982e-003	2.5729e-004	6.5719e-004	1.0658e-003
1.4e-006	3.225e-002	1.5987e-003	2.5744e-004	6.5720e-004	1.0657e-003
1.5e-006	3.225e-002	1.5992e-003	2.5758e-004	6.5720e-004	1.0011e-003
1.6e-006	3.2314e-002	1.6016e-003	2.7110e-004	7.2211e-004	1.0656e-003
1.7e-006	3.225e-002	1.6002e-003	2.5788e-004	6.5723e-004	1.0655e-003
1.8e-006	3.225e-002	1.6007e-003	2.5803e-004	6.5724e-004	1.0654e-003
1.9e-006	3.225e-002	1.6011e-003	2.5818e-004	6.5725e-004	1.0653e-003
2.0e-006	3.225e-002	1.6017e-003	2.583e-004	6.5726e-004	1.0652e-003

Table 3.10: Sinc Data Set Prediction Result 5-2: Corresponding "mse" with ε and σ Varying

$\varepsilon \setminus \sigma$	6	7	8	9	10
0	1.0144e-001	1.0310e-001	1.0442e-001	1.0501e-001	1.0558e-001
1.0e-007	1.3606e-003	1.5642e-003	1.7085e-003	1.8191e-003	1.9256e-003
2.0e-007	1.3605e-003	1.5641e-003	1.7083e-003	1.8191e-003	1.9255e-003
3.0e-007	1.3603e-003	1.5639e-003	1.7082e-003	1.8188e-003	1.9254e-003
4.0e-007	1.3602e-003	1.5637e-003	1.7080e-003	1.8187e-003	1.9253e-003
5.0e-007	1.3600e-003	1.5635e-003	1.7078e-003	1.8185e-003	1.9252e-003
6.0e-007	1.3599e-003	1.5634e-003	1.7076e-003	1.8184e-003	1.9251e-003
7.0e-007	1.3598e-003	1.5632e-003	1.7074e-003	1.8182e-003	1.9250e-003
8.0e-007	1.3596e-003	1.5630e-003	1.7073e-003	1.8181e-003	1.9250e-003
9.0e-007	1.3595e-003	1.5628e-003	1.7071e-003	1.8179e-003	1.9249e-003
1.0e-006	1.3593e-003	1.5627e-003	1.7069e-003	1.8178e-003	1.9248e-003
$\varepsilon \setminus \sigma$	6	7	8	9	10
1.1e-006	1.3592e-003	1.5625e-003	1.7067e-003	1.8177e-003	1.9247e-003
1.2e-006	1.3590e-003	1.5623e-003	1.7065e-003	1.8175e-003	1.9246e-003
1.3e-006	1.3589e-003	1.5621e-003	1.7064e-003	1.8174e-003	1.9245e-003
1.4e-006	1.3587e-003	1.5620e-003	1.7062e-003	1.8172e-003	1.9245e-003
1.5e-006	1.3586e-003	1.5618e-003	1.7060e-003	1.8171e-003	1.9244e-003
1.6e-006	1.3584e-003	1.5616e-003	1.7059e-003	1.8169e-003	1.9243e-003
1.7e-006	1.3583e-003	1.5614e-003	1.7057e-003	1.8168e-003	1.9242e-003
1.8e-006	1.3582e-003	1.5613e-003	1.7055e-003	1.8166e-003	1.9241e-003
1.9e-006	1.3580e-003	1.5611e-003	1.7053e-003	1.8165e-003	1.9241e-003
2.0e-006	1.3579e-003	1.5609e-003	1.7052e-003	1.8163e-003	1.9240e-003

Table 3.11: Sunspot Data Set Prediction Result 1: Corresponding "mse" with ε Varying

ε	0	1	2	3	4	5
mse	202.47	**144.48**	147.79	150.67	152.92	155.35
ε	6	7	8	9	10	11
mse	157.81	160.40	163.50	166.85	170.41	174.19

Table 3.12: Sunspot Data Set Prediction Result 2: Corresponding "mse" with ε Varying

ε	0.0	0.1	0.2	0.3	0.4	0.5	0.6
mse	202.47	**142.03**	142.27	142.52	142.78	143.05	143.31
ε	0.7	0.8	0.9	1.0	1.1	1.2	1.3
mse	143.59	143.87	144.17	144.48	144.78	145.09	145.42
ε	1.4	1.5	1.6	1.7	1.8	1.9	2.0
mse	145.77	146.11	146.46	146.81	147.14	147.46	147.79

Table 3.13: Sunspot Data Set Prediction Result 3: Corresponding "mse" with ε Varying

ε	0.00	0.01	0.02	0.03	0.04	0.05	0.06
mse	202.47	**141.79**	141.82	141.85	141.87	141.90	141.93
ε	0.07	0.08	0.09	0.10	0.11	0.12	0.13
mse	141.96	141.98	142.01	142.03	142.05	142.08	142.10
ε	0.14	0.15	0.16	0.17	0.18	0.19	0.20
mse	142.13	142.15	142.17	142.20	142.23	142.25	142.27

Table 3.14: Rmsv Data Set Prediction Result 1: Corresponding "mse" with ε Varying

ε	0.000	0.001	0.002	0.003	0.004
mse=1.0e-007*	**1.0893**	6.3837	20.5906	55.6539	92.8517
ε	0.005	0.006	0.007	0.008	0.009
mse=1.0e-007*	109.908	119.489	125.047	128.509	130.901
ε	0.010	0.011	0.012	0.013	0.014
mse=1.0e-007*	132.546	133.730	134.615	135.294	135.829
ε	0.015	0.016	0.017	0.018	0.019
mse=1.0e-007*	136.257	136.606	136.895	137.136	37.341
ε	0.020				
mse=1.0e-007*	137.515				

Table 3.15: Rmsv Data Set Prediction Result 2: Corresponding "mse" with ε Varying

ε=1.0e-004*	0	1	2	3	4	5
mse=1.0e-008*	10.8933	**5.4658**	11.639	16.937	21.751	27.671
ε=1.0e-004*	6	7	8	9	10	11
mse=1.0e-008*	32.270	40.049	49.672	55.247	63.837	75.460
ε=1.0e-004*	12	13	14	15	16	17
mse=1.0e-008*	90.098	107.654	128.254	142.760	148.620	157.918
ε=1.0e-004*	18	19	20			
mse=1.0e-008*	170.057	186.557	205.906			

Table 3.16: Rmsv Data Set Prediction Result 3: Corresponding "mse" with ε Varying

ε=1.0e-005*	0	1	2	3	4	5	6
mse=1.0e-009*	108.93	**4.8233**	5.409	6.866	10.122	15.383	19.874
ε=1.0e-005*	7	8	9	10	11	12	13
mse=1.0e-009*	27.14	35.74	44.78	54.658	63.31	70.92	78.33
ε=1.0e-005*	14	15	16	17	18	19	20
mse=1.0e-009*	85.69	92.65	97.53	102.53	107.03	111.52	116.39

Table 3.17: Rmsv Data Set Prediction Result 4: Corresponding "mse"with ε Varying

ε=1.0e-006*	0	1.0	2.0	3.0	4.0	5.0	6.0
mse=1.0e-009*	108.93	5.0162	4.9780	4.9225	4.8968	4.8435	4.8341
ε=1.0e-006*	7.0	8.0	9.0	10	11	12	13
mse=1.0e-009*	4.8112	4.8105	**4.8086**	4.8233	4.8465	4.8791	
ε=1.0e-006*	14	15	16	17	18	19	20
mse=1.0e-009*	4.9204	4.9670	5.0152	5.0771	5.1608	5.2227	5.3185

Chapter 4

Using The Golden Section Search Optimization Routine For Verifying the Optimal Error Zone Value

4.1 Introduction to the Golden Section Search Method

The golden section search method is a one-dimensional optimization search method which determines a minimizer of a unimodal function over a closed interval [3]. The solution to the golden section search routine is a local minimizer in the whole data range, that is a global minimizer which remains within two given points as long as the function is unimodal. The golden section search method is defined as the follows:

Let L be a closed interval $L = [a, b]$. The function $f(x)$ for $x \in L$ is an unimodal function with a minimizer if and only if there exists a point $x^* \in L$ such that $f(x)$ is monotonically decreasing on the left side of $x^*$and monotonically increasing on the right side of x^*. After each iteration, the

remaining interval length is reduced to 61.8% of the starting interval length. We define τ as 0.618; Ω is the length of the final interval. Thus, the number of iterations needed is $N = \log\left(\Omega/L\right) / \log\tau$. The golden section search algorithm is:

$L = b - a,$

$x1 = b - \tau L,\ x2 = a + \tau L,$

For $i = 1$ to N by 1

If $f(x1) > f(x2)$, $a = x1,\ L = b - a,\ x1 = x2,\ x2 = a + \tau L,$

Else $b = x2,\ L = b - a,\ x2 = x1,\ x1 = b - \tau L,$

End

$x^* = (b - a)/2,\ f(x^*).$

4.2 Motivation for Using the Golden Section Search Method

The rule of thumb on the optimal error zone value was derived based on the experiment results in Chapter 3. It is found to be one-tenth of the accuracy of the input target value. In this chapter, the optimal error zone value is searched only because the optimal kernel parameter and the suitable kernel have been found out for every data set in Chapter 3. The experiments described in Chapter 3 present the relationship between the "mse" value and the error zone value. It is:

1. The smaller the error zone value, the smaller the "mse" value.

2. The "mse" will get smaller very slowly after the error zone value reaches

a very small value and approaches to zero.

3. When the error zone value is given as zero, the "mse" value is relatively large.

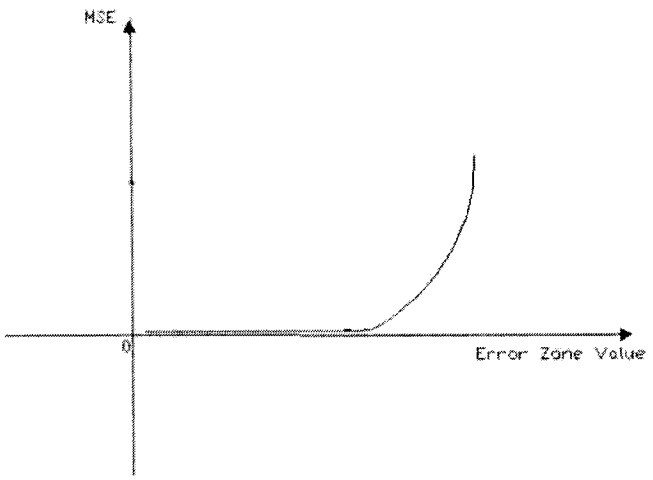

Figure 4.1: MSE vs. Error Zone Values

The relationship between "mse" value and error zone value are presented in Figure 4.1. Based on the above observation, we believe that for every data set, there is a optimal error zone value at which the "mse" function will monotonically increase on the right side of the optimal point and monotonically decrease on the left side of the optimal point. The objective function "mse" with a variable error zone value is regarded as a unimodal function. In this situation, the golden section search optimization method is used to find the smallest "mse" value and corresponding optimal error zone value. Nor-

mally, the golden section search optimization method is a local optimization method, which means that the result obtained by this method is not necessarily a global minimizer; however, the optimal error zone value has been figured out for every data set in Chapter 3. Thus, the optimal error zone value searched for here using the golden section search optimization method would be global. The stopping criterion, on the basis of which the searching process is halted, is defined below.

When the "mse" ratios [see Equation (3.2)] for five consecutive golden section search iterations are all less than 10^{-5}, the searching process will stop.

4.3 Case Studies

4.3.1 The Sinc Data Set

The first data set is the Sinc data set. There are 51 data points in this data set. The accuracy of the input target value of this data set is 10^{-5}. From the experiments carried out in Chapter 3, we have confirmed that the optimal error zone value is in the range of [0, 1]. Thus, the optimal error zone value is searched for in this range by using the golden section search optimization method.

The middle point of the interval at iteration n ($n = 1, 2, 3, ...$) of the error zone values and the corresponding "mse" values are shown in Table 4.1 (see page 71). With the number of the iterations increasing, the error zone value

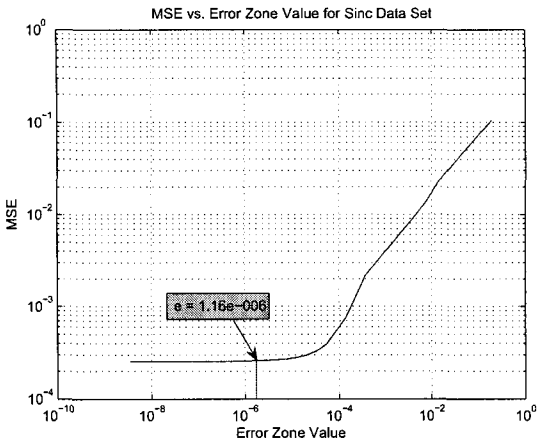

Figure 4.2: MSE vs. Error Zone Values for the Sinc Data Set

and "mse" are both decreasing. The searching process is stopped at iteration 43. This is because the "mse" ratio between 38 and 39 is larger than 10^{-5}, while the "mse" ratios for the five consecutive iterations, from iteration 39 to iteration 43, are all less than 10^{-5}, which is the criterion for stopping. The process is presented below.

- The "mse" ratio between iterations 38 and 39 is: $(2.55383 \times 10^{-4} - 2.55380 \times 10^{-4})/(2.55383 \times 10^{-4}) = 1.17 \times 10^{-5}$
- The "mse" ratio between iterations 39 and 40 is: $(2.55380 \times 10^{-4} - 2.55379 \times 10^{-4})/(2.55379 \times 10^{-4}) = 3.916 \times 10^{-6}$
- The "mse" ratio between iterations 40 and 41 is: $(2.55379 \times 10^{-4} - 2.55378 \times 10^{-4})/(2.55379 \times 10^{-4}) = 3.916 \times 10^{-6}$

- The "mse" ratio between iterations 41 and 42 is: $(2.55378 \times 10^{-4} - 2.55377 \times 10^{-4})/(2.55378 \times 10^{-4}) = 3.916 \times 10^{-6}$
- The "mse" ratio between iterations 42 and 43 is: $(2.55377 \times 10^{-4} - 2.55376 \times 10^{-4})/(2.55377 \times 10^{-4}) = 3.916 \times 10^{-6}$

So far, the searching process has satisfied the stopping criterion, which takes into consideration the fact that the "mse" values don't change even though the error zone value is decreasing. There is no need to search for a smaller "mse". The searching process stops here. The experiment result is plotted in Figure 4.2. It presents the relation between the "mse" value and the error zone value. We find that the error zone value, 1.16×10^{-6} is a critical point in Figure 4.2. The "mse" curve remains almost flat once the error zone value is smaller than it, whereas, the "mse" curve has a obvious slope once the error zone value is larger than it. This means that the "mse" changes significantly when the error zone value is larger than 1.16×10^{-6}, and the "mse" value shows no significant change when the error zone value is smaller than 1.16×10^{-6}. In addition, we know the predicting time is also an important factor in the assessment of prediction. We have mentioned before that the smaller the error zone value, the longer the predicting time since smaller error zone values result in more support vectors, and more support vectors need more calculating time. The experiments regarding the relation between the predicting time and the error zone value are conducted, some error zone values in Table 4.1 are chosen and the calculating time for each

error zone value is recorded. The results are presented in Table 4.2 (see page 71). The predicting time has varied in a range from 29s to 56s while the error zone value has changed in the range of $[3.1942 \times 10^{-10}, 1]$. The smaller the error zone value, the longer the calculating time.

Summarizing the results described in Table 4.1, Table 4.2, and Figure 4.2, we may say that fixing the error zone value at 1.16×10^{-6} was a good choice for this data set. It can guarantee not only the predicting accuracy but also the short predicting time. It should be noted, however, that we got this error zone value from the plot; it is a rough value instead of an exact one. In this situation, it would be a wise choice to use a round value which is smaller, the one closest to 1.16×10^{-6}. From Figure 4.2, it is clear that 10^{-6} is the value we need. The reasons are presented below:

1. 10^{-6} is a round, smaller value and the one closest to 1.16×10^{-6}. The experiment shows that the predicting time between error zone values 1.16×10^{-6} and 10^{-6} are almost the same. Thus, choosing 10^{-6} as the error zone value not only guarantees the prediction accuracy but also a short predicting time.

2. Numerically, the "mse" ratio between the error zone 10^{-6} and the smallest error zone value (3.1942×10^{-10}) is: $(2.56844 \times 10^{-4} - 2.55376 \times 10^{-4})/(2.56844 \times 10^{-4}) = 0.0057$. It is smaller than 0.01. In Chapter 3, we stipulated that the two prediction results are regarded as identical if the "mse" ratio between them is less than 0.01. For this reason, we

can say that the prediction results between the error zone values 10^{-6} and 3.1942×10^{-10} are identical. On the other hand, the predicting time for these two values are 29.1s and 56s respectively, which are very difference.

3. Fixing the error zone value at 10^{-6} is very simple. It is equal to one-tenth of the input target value. It can be obtained directly from the input target value before the prediction is performed.

From the above analysis, we can say that fixing 10^{-6} as the error zone value is a reasonable choice. It can guarantee both high prediction accuracy and short predicting time. The rule obtained in Chapter 3 is approved further.

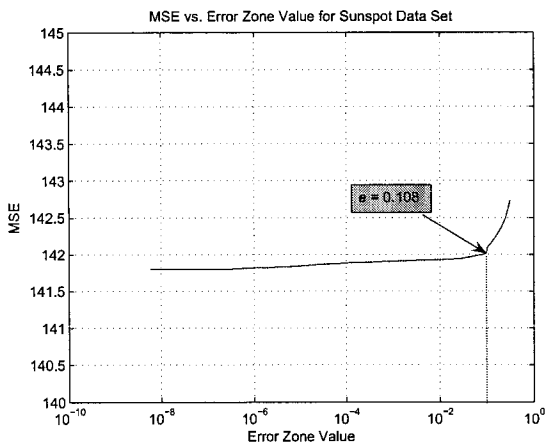

Figure 4.3: MSE vs. Error Zone Values for the Sunspot Data Set

4.3.2 The Sunspot Data Set

The second data set is the Sunspot data set. There are 459 data points in this set. The accuracy of the input target value of this data set is 1. The experiment process is the same as that for the Sinc data set. The experiments conducted in Chapter 3 show that the range of the optimal error zone value is within [0, 1]. The golden section search optimization method is used to find the optimal error zone value in this range for the desired prediction accuracy. Similar to the last data set, the middle point of the interval at the n_{th} iteration, $(n = 1, 2, 3, ...)$ of the error zone values and corresponding "mse" values are presented in Table 4.3 (see page 72). With the increase in iteration number, the error zone value decreases. The search is stopped at iteration 17. This is because the "mse" ratio for iterations 12 and 13 is bigger than 10^{-5}, while the "mse" ratios for the five consecutive iterations from 13 to 17 are all smaller than 10^{-5}, the defined stopping criterion. The results are presented below:

- The "mse" ratio between iterations 12 and 13 is:
  $(141.766 - 141.7645)/141.766 = 1.06 \times 10^{-5}$
- The "mse" ratio between iterations 13 and 14 is:
  $(141.7645 - 141.76375)/141.7645 = 5.29 \times 10^{-6}$
- The "mse" ratio between iterations 14 and 15 is:
  $(141.76375 - 141.76370)/141.76375 = 3.53 \times 10^{-7}$

- The "mse" ratio between iterations 15 and 16 is:
  $(141.76370 - 141.76298)/141.76370 = 5.079 \times 10^{-6}$
- The "mse" ratio between iterations 16 and 17 is:
  $(141.76298 - 141.76279)/141.76298 = 1.34 \times 10^{-6}$

The above "mse" ratios from iterations 13 to 17 are all smaller than 10^{-5}, the defined stopping criterion. The "mse" curve has become very flat so the "mse" remains virtually unchanged even though the error zone value is decreasing. The searching process stops here. The error zone value is divided from 1 to 8.6604×10^{-5}. Figure 4.3 presents the trend between the "mse" and error zone values. The error zone value 0.108 is a critical point, for the "mse" curve is relatively flat at error zone values smaller than 0.108, but has a obvious slop at error zone values larger than 0.108. This means that the "mse" value changes significantly when the error zone value is larger than 0.108, but remains virtually unchanged when the error zone value is smaller than 0.108. Normally, the prediction accuracy and predicting time are two factors considered crucial to prediction performance. The experiment regarding the relation between predicting time and error zone value is conducted, in which several error zone values in Table 4.3 are chosen and the calculating time for each error zone value is recorded. Table 4.4 (see page 72) and Figure 4.4 present the result - the smaller the error zone value, the longer the predicting time. Based on this observation, it would appear that fixing the error zone value at 0.108 is a good choice. The reasons are the same as those given

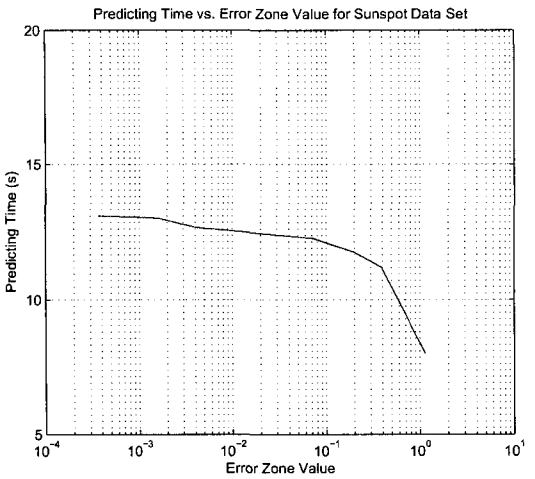

Figure 4.4: Predicting Time vs. Error Zone Values for the Sunspot Data Set

for the last data set analyzed above: the ability to obtain a high prediction accuracy and, as well, a short predicting time. Note, however, that we obtained this value from the figure; it is a rough value, not an exact one. In this case, using the value 0.1 to replace the value 0.108 would be a reasonable choice. This is because 0.1 is smaller than 0.108, so it can guarantee prediction accuracy without adversely affecting time, which remains almost the same since the two values are so close. As well, 0.1 is equal to one-tenth of the accuracy of the input target value. We can obtain this value directly from the input target value. The "mse" ratio 0.1 and the smallest error zone value of 8.6604×10^{-5} is calculated. At $(142.03 - 141.76279)/142.03 = 0.002$, it is smaller than 0.01, the defined stopping criterion given in Chapter 3. According to this stopping criterion, the prediction results between the error

zone value 0.1 and the smallest error zone value (8.6604×10^{-5}) is identical, while the predicting times are 14s and 31s respectively. Thus, fixing the error zone value at 0.1 can guarantee both the prediction accuracy and short predicting time. The rule obtained from Chapter 3 has been supported once again.

4.3.3 The Vibration Data Set "Root Mean Square"

The vibration data vertical root mean square (Rmsv) is the processed vibration signals collected from the Gearbox Lifetime Assessment Experiment carried out in the Reliability Research Lab at the University of Alberta. It represents the vibration trend of the machines in a degradation state. There are 237 data points in this data set. The accuracy of the input target value of this data set is 10^{-4}. Experiments are conducted that are similar to those conducted on the last two data sets. The searching rang for the error zone value is [0, 0.01] based on the experiments in Chapter 3. The middle point of the interval at the n_{th} iteration, $(n = 1, 2, 3, ...)$ of the error zone values and corresponding "mse" values are presented in Table 4.5 (see page 73). As the number of the iteration increases, the error zone value decreases. The smaller the error zone value, the smaller the "mse" value. The smallest error zone value searched is 2.6910×10^{-7}. The searching process stopped at iteration 29. This is because even though the "mse" ratio between iterations 24 and 25 is $((4.8255 \times 10^{-9}) - (4.812541 \times 10^{-9}))/(4.8255 \times 10^{-9}) = 2.685 \times 10^{-3}$, which is larger than 0.01, the "mse" values from 25 to 29 are all equal to

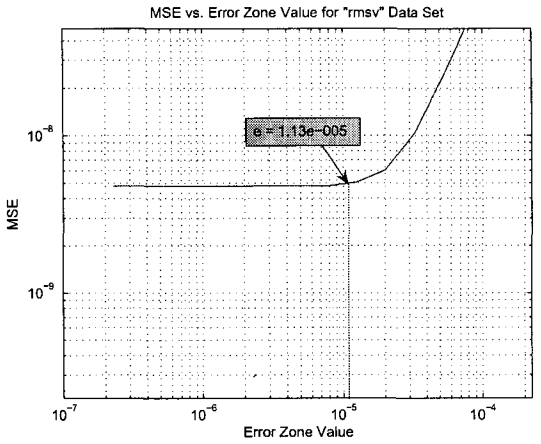

Figure 4.5: MSE vs. Error Zone Value for the Rmsv Data Set

$4.812584136851701 \times 10^{-9}$. As a result, the "mse" ratios between the consecutive iterations are all less than 10^{-5}, the defined stopping criterion. Thus, the searching process has stopped here. The trend between "mse" values and error zone values is plotted in Figure 4.5. From this figure, we find that the error zone value of 1.13×10^{-5} is a sensitive point. The "mse" curve has a very flat slope when the error zone value is smaller than 1.13×10^{-5}, whereas, it has a relatively steep slope when the error zone value is larger than 1.13×10^{-5}. This means that the prediction accuracy changes obviously when the error zone value is larger than 1.13×10^{-5}, but shows no significant change when the error zone value is smaller.

Moreover, the predicting time is still an important factor that should be considered here. As well, the relation between time and error zone value

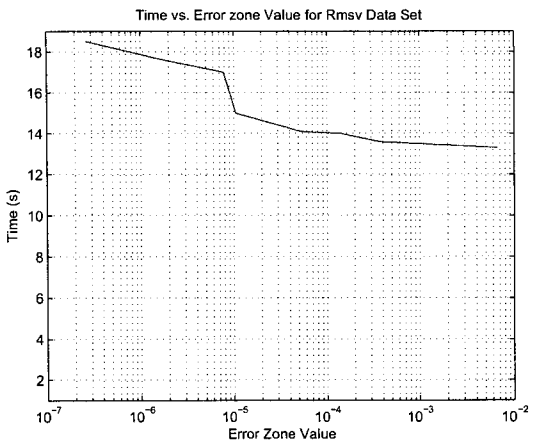

Figure 4.6: Time vs. Error Zone Value for the Rmsv Data Set

is studied. Several error zone values are chosen and the calculating time for each error zone value is recorded. The results are illustrated in Table 4.6 (see page 73) and Figure 4.6. We find that the predicting time changes following the change of error zone value. The smaller the error zone value, the longer the predicting time. Based on the above analysis, it seems that fixing the error zone value at 1.13×10^{-5} is a good choice, considering both prediction accuracy and predicting time. Note, however, that we got this value roughly from the plot; It is not an accurate value. Using the same methods as the last two data sets, the round value is chosen that is smaller and closest to the value 1.13×10^{-5}. It would be a wise choice to use the value 10^{-5} to replace the value 1.13×10^{-5}. The "mse" ratio between 10^{-5} and the smallest error zone value (2.6910×10^{-7}) is calculated: (4.8233×10^{-9} –

$4.812541 \times 10^{-9})/(4.8233 \times 10^{-9}) = 0.0023$. This is smaller than 0.01, the stopping criterion defined in Chapter 3. According to this stopping criterion, the predicting results of the error zone values of 10^{-5} and 2.6910×10^{-7} can be considered identical, while the predicting times between these two values are 15.2s and 21s respectively. Thus, fixing the value at 10^{-5} can guarantee both high prediction accuracy and a short predicting time. Moreover, we noticed that the error zone value of 10^{-5} is one-tenth of the accuracy of the input target value of the "Rmsv" data set. It is much easier to obtain this value from the input target value. The rule formulated in Chapter 3 has received further support.

4.4 Discussion and Conclusion

In the experiments conducted in Chapters 3 and 4, three data sets were used to test the prediction performance of the SVM regression model. First, the optimal error zone value and kernel parameter were found simultaneously according to the procedures proposed in Chapter 3. The range of the optimal error zone value was found for every data set. Roughly, the conclusion reached in Chapter 3 was: the optimal error zone value is equal to one-tenth of accuracy of the input target value. In Chapter 4, the error zone value was found only because the optimal kernel parameter and type had been figured out for every data set in Chapter 3. Because the "mse" function with variable error zone value is considered as a unimodal function, the golden search optimization method was adopted for a more in-depth search on the optimal

error zone value. The rule on the optimal error zone value is further proved in Chapter 4 considering both the prediction accuracy and predicting time. The benefits of the rule on the SVM regression for prediction include:

1. In SVM prediction performance, the complication of prediction calculations can be reduced greatly if the error zone value, one of the important parameters, can be fixed before the prediction is processed.

2. The proper selection of the error zone value would guarantee not only higher prediction accuracy with SVM regression but also shorter predicting time.

3. Fixing the optimal error zone value according to the accuracy of the input target value, the only information that can be obtained before prediction, can guarantee high prediction accuracy will be obtained without any predicting time being wasted. But this doesn't mean that the SVM regression model can't obtain the same results at other error zone values. If the error zone value closed to the optimal error zone value is carried into the SVM regression model, the prediction accuracy and predicting time may be almost the same as those obtained when the optimal error zone value is used. This doesn't mean that the optimal error zone value is useless; when a new data set is collected and there is no prior knowledge about the properties of this data set, this rule provides a basis for the selection of the error zone value. Without this, the wrong selection of error zone value could lead to SVM prediction

performance failure. Such failure would leave SVM users confused and dubious regarding SVM regression's ability to predict.

Table 4.1: Mean Square Error according to the Error Zone Values for the Sinc Data Set

ε	0.3090	0.191	0.118	0.07296
mse	0.11712	0.10592	0.07108	0.04942
ε	0.04509	0.02787	0.017224	0.006580
mse	0.03662	0.02872	0.02605	0.01263
ε	0.004067	0.002513	0.0015536	9.6015e-004
mse	0.006614	0.00424	0.00358	0.00351
ε	5.9349e-004	3.6678e-004	2.2671e-004	1.4011e-004
mse	3.4569e-003	2.2394e-003	1.2423e-003	7.5517e-004
ε	8.6604e-005	5.3527e-005	3.3083e-005	2.0445e-005
mse	5.1694e-004	3.9216e-004	3.3040e-004	2.9811e-004
ε	1.2637e-005	7.810e-006	4.8276e-006	2.9834e-006
mse	2.8040e-004	2.7031e-004	2.6440e-004	2.6088e-004
ε	1.8441e-006	1.1397e-006	7.04461e-007	4.3536e-007
mse	2.5875e-004	2.5745e-004	2.5665e-004	2.56164e-004
ε	2.6910e-007	1.6630e-007	1.0280e-007	6.3529e-008
mse	2.55862e-004	2.55676e-004	2.5562e-004	2.5549e-004
ε	3.9269e-008	2.4268e-008	1.5000e-008	9.2704e-009
mse	2.5545e-004	2.5542e-004	2.5540e-004	2.5539e-004
ε	5.7302e-009	3.5412e-009	2.1890e-009	1.3528e-009
mse	2.5539e-004	2.55383e-004	2.55380e-004	2.55379e-004
ε	8.3618e-010	5.1676e-010	3.1942e-010	
mse	2.55378e-004	2.55377e-004	2.55376e-004	

Table 4.2: Predicting Time according to the Error Zone Values for the Sinc Data Set

ε	0.309	0.07296	0.0066	0.04509	0.00155
Time(s)	17	19	20	21	22
ε	1.4e-004	3.3e-005	1.16e-006	1.0e-006	2.82e-009
Time(s)	24	27	29	29.1	56

Table 4.3: Mean Square Error according to the Error Zone Values for the Sunspot Data Set

ε	0.3094	0.1914	0.1181	0.07296
mse	142.73	142.36	142.14	142.01
ε	0.04509	0.02787	0.017224	0.006580
mse	141.92	141.86	141.82	141.79
ε	0.004067	0.002513	0.0015536	9.6015e-004
mse	141.78	141.77	141.768	141.766
ε	5.9349e-004	3.6678e-004	2.2671e-004	1.4011e-004
mse	141.7645	141.76375	141.76370	141.76298
ε	8.6604e-005			
mse	141.76279			

Table 4.4: Predicting Time vs. Error Zone Values for the Sunspot Data Set

ε	1.14	0.108	0.1	0.0730	0.0041	0.0016	0.00037	4.8276e-006
Time (s)	8	14	14.1	17.26	19.67	22.01	25.10	31

Table 4.5: Mean Square Error according to the Error Zone Value for the Rmsv Data Set

ε	0.3094	0.1914	0.1181	0.07296
mse	1.3917e-005	1.3916e-005	1.3914e-005	1.3908e-005
ε	0.04509	0.02787	0.017224	0.006580
mse	1.3894e-005	1.3859e-005	1.3771e-005	1.2889e-005
ε	0.004067	0.002513	0.0015536	9.6015e-004
mse	1.1025e-006	6.0868e-006	1.9023e-006	8.8021e-007
ε	5.9349e-004	3.6678e-004	2.2671e-004	1.4011e-004
mse	4.3370e-007	2.4626e-007	1.6053e-007	1.0409e-007
ε	8.6604e-005	5.3527e-005	3.3083e-005	2.0445e-005
mse	6.0961e-008	2.3970e-008	1.0525e-009	6.0369e-009
ε	1.2637e-005	7.810e-006	4.8276e-006	2.9834e-006
mse	5.0633e-009	4.8255e-009	4.8324e-009	4.8255e-009
ε	1.8441e-006	1.1397e-006	7.0446e-007	4.3536e-007
mse	4.812541e-009	4.812541e-009	4.812541e-009	4.812541e-007
ε	2.6910e-007			
mse	4.812541e-009			

Table 4.6: Predicting Time vs. Error Zone Value for the Rmsv Data Set

ε	0.0066	9.6e-004	1.4e-004	5.4e-005	1.2637e-005
Time(s)	13.3	13.5	14	14.1	15
ε	1.0e-005	1.13e-005	1.8e-006	7.04e-007	1.028e-007
Time(s)	15.2	15.3	17.6	19	21

Chapter 5

Comparison between CMa's Method and My Method on the Optimal Error Zone Value

5.1 Introduction to CMa's method

Cherkassky and Ma [6] published a paper which describes a simple yet practical analytical approach to select the error zone value, one of the SVM regression parameters, directly from the training data. This method is simply called CMa's method. The error zone value obtained using this method is called the calculated error zone value. The approach involves basing the selection of the error zone value on the noise in the training data, and the (known) number of training samples. The assumption is that the error zone value is proportional to the input noise level. They assumed that the larger data set needs smaller error zone value. CMa's method relies heavily on the "knn" method [6], which makes predictions for the training data first, then calculates the noise level. Here, "noise" is defined as the error between the

observed data and the predicted data. The error zone value is thus calculated according to the formula defined by the authors, which is:

$$\varepsilon = 3\sigma\sqrt{\frac{\ln n}{n}}, \tag{5.1}$$

where, σ is the variance of the noise and n is the sample size.
Thus, CMa's method can be realized in the following steps:

1. The "Knn" regression method is used to make predictions, and $\hat{y}$ is obtained.
2. The formula $\hat{\sigma}^2 = \frac{n^{1/5k}}{n^{1/5_{k-1}}} * \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is used to calculate the noise variance.
3. The ε value is fixed according to Equation (5.1).

5.2 Case Studies

In this section, three data sets, the Sinc data set, the Sunspot data set and the vibration data set "Rmsv" are used to provide comparison results.

5.2.1 The Sinc Data Set

The first data set is the Sinc data set, with a size of 51. The error zone value is calculated first using CMa's method. The results are presented below:

- When the first half of the data set, the first 25 data, are used as training data, the error zone value is 0.0055.

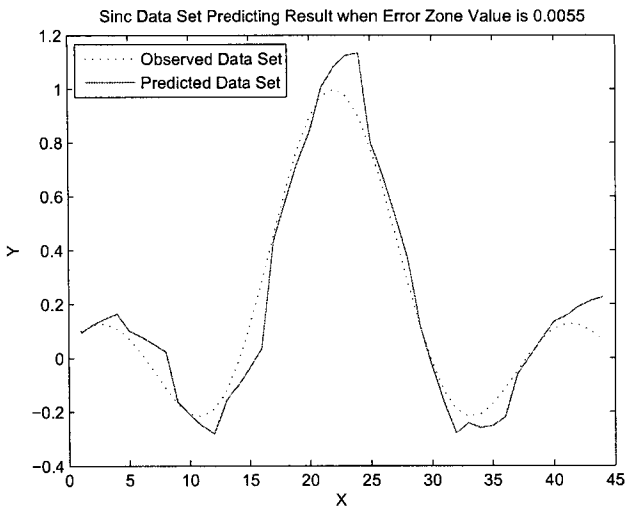

Figure 5.1: The Sinc Data Set Prediction Result when the Error Zone Value is 0.0055

- When the first 30 data are used as training data, the error zone value is 0.0048.
- When the first 40 data are used as the training data, the error zone value is 0.0039.
- When the whole data set is used as the training data, the error zone value is 0.0034.

We can see that for this data set the error zone value decreases as the training data size increases. When the largest and the smallest calculated error zone values are inserted into the SVM regression model, the following results are obtained:

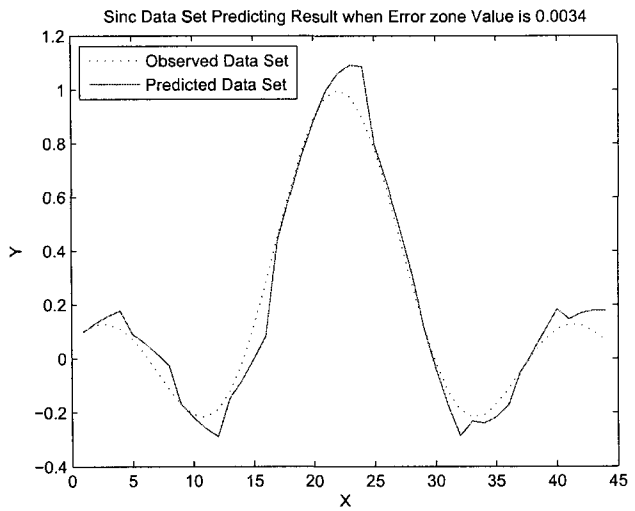

Figure 5.2: The Sinc Data Set Prediction Result when the Error Zone Value is 0.0034

- When the error zone value is 0.0055, "mse" is 0.007564, the result is as illustrated in Figure 5.1.
- When the error zone value is 0.0034, "mse" is 0.00452679, the result is as shown in Figure 5.2.
- When the optimal error zone value, 10^{-6}, is inserted into the SVM regression model for the Sinc data set, "mse" is 0.00025684.

Comparing the above results, we can see that, for the Sinc data set, interpolating the above two calculated error zone values into the SVM regression model doesn't produce the same good prediction result as that obtained using the optimal error zone value. The "mse" ratio as the error zone value

is 0.0055 and 10^{-6} is 0.96, while, the "mse" ratio as the error zone value is 0.0034 and 10^{-6} is 0.94. It shows that the prediction accuracy has significant difference between two calculated error zone value and the optimal error zone value. Moreover, it is not reasonable to take the whole data set as training data.

5.2.2 The Sunspot Data Set

The second data set is the Sunspot data set, with a data size of 459. The above procedure is repeated again. The following results are obtained:

- When the training data size is 70, the error zone value is 1.14, "mse" is 144.85.
- When the training data size is 100, the error zone value is 0.76, "mse" is 143.73.
- When the training data size is 200, (almost half the whole data set), the error zone value is 0.39, "mse" is 142.70.
- Using the whole data set of 459 as the training data, the error zone value is 0.19966, "mse" is 142.25.
- When the error zone value adopts the optimal error zone value of 0.1, "mse" is 142.03.

The above experiments results demonstrate that the larger the training data size, the smaller the error zone value. When the training data size is

70 or 100, the calculated error zone values are larger than the optimal error zone value, and the "mse" obtained by using the calculated error zone value is much larger than that obtained by using the optimal error zone value. The "mse" ratios are calculated as below:

- The "mse" ratio as error zone values are 1.14 and 0.1 is: $(144.85 - 142.03)/144.85 = 0.019$, larger than 0.01.
- The "mse" ratio as error zone values are 0.76 and 0.1 is: $143.73 - 142.03)/143.73 = 0.012$, larger than 0.01.

When half of the data points are used as training data, the calculated error zone value closes with the optimal error zone value, and the "mse" is not significantly different from that obtained by using the optimal error zone value (the "mse" ratio is less than 0.01). When the full data set is used as training data, the calculated error zone value much closely approximates the optimal error zone value, and the "mse" value is almost the same as that obtained by using the optimal error zone value. It is not reasonable, however, to use the whole data set as training data.

5.2.3 The Vibration Data Set "Rmsv"

The third data set is the "Rmsv" data set, with a data size of 236. The above procedure is repeated again and the following results are obtained:

- When the training data size is 50, the error zone value is 1.0522e-005, the "mse" is 4.8315×10^{-9}.

- When the training data size is 100, the error zone value is 6.5947e-006, the "mse" is 4.8257×10^{-9}.
- When the training data size is 150, the error zone value is 5.1822e-006, the "mse" is 4.8104×10^{-9}.
- When the training data size is 200, the error zone value is 7.7695e-005, the "mse" is 3.0781×10^{-8}.
- When the training data size is 236, the error zone value is 6.2729e-005, the "mse" is 2.6752×10^{-8}.
- The optimal error zone value based on my rule is 0.00001, the "mse" is 4.8283×10^{-9}.

From the above results, we can see that the calculated error zone value is different with different training data sizes, however, this result breaks the rule formulated on the basis of the above two data sets - the bigger the data size, the smaller the error zone value. In contrast, for this data set, larger data sizes produced larger error zone values and result in larger "mse" value. The "mse" ratios between every calculated error zone value and the optimal error zone value are calculated. The results show that the "mse" ratios are less than 0.01 as the training data size is 50, 100, and 150 respectively, whereas, the "mse" ratios are more than 0.80 as the training data size is 200 and 236 (full data set) respectively. The experiments showed that CMa's method is not stable.

5.3 Discussion and Conclusion

In this chapter, CMa's method is compared with the rule of thumb presented in Chapter 3. Three data sets are used, and the comparisons are analyzed. For the Sinc data set, we can see that the error zone value obtained using CMa's method is larger than the optimal error zone value obtained using my method even though the whole data set is used as training data. As a result, the "mse" obtained by using the calculated error zone value is larger that obtained by using the optimal error zone value. Thus, the prediction accuracy obtained using calculated error zone value from CMa's method is inferior to that obtained using the optimal error zone value from my method. For the Sunspot data set, the calculated error zone value is larger than the optimal error zone value, when the training data size is smaller. As a result, the prediction accuracy is inferior to that obtained using the optimal error zone value. On the other hand, the calculated error zone value is very close to the optimal error zone value when the training data size is equal to or larger than half of the whole data set. As a result, the same prediction accuracy can be obtained by using either method. It is not reasonable, however, that the training data size is very larger. Thus, if CMa's method is used for this data set, the problem arises of how to fix the training data size to obtain a suitable calculated error zone value. For the vibration Rmsv data set, the error zone value calculated using CMa's method is very close to the optimal error zone value from my method even though the training size is smaller.

As a result, the "mse" value is not significantly different for either method, and the same prediction accuracy is possible. It should be noted, however, that the result from the Rmsv vibration data set breaks the rule set by the other two data sets. The rule is that the larger the training data size, the smaller the error zone value. According to CMa's paper, their method is based on the concept that the error zone value chosen is related to the data size; this is because they believe larger data sets require smaller error zone values in order to obtain high prediction accuracy. The above experiment showed that this is not always true. From the above discussion, we can conclude that the calculated error zone value is, for some data sets, very close to or even the same as, the optimal error zone value. This means that CMa's method can obtain the same prediction accuracy as my method for some data sets. On the other hand, the calculated error zone value is, for other data sets, larger than the optimal error zone value, and results in lower prediction accuracy. Moreover, CMa's method requires more complex calculations. Since the training data size needs to be determined before prediction, my method is superior to CMa's.

Chapter 6

Application of SVM Regression For Vibration Trend Forecasting

6.1 Vibration Data from the Gearbox Lifetime Assessment Experiment

Using the rule of thumb developed earlier, we applied the SVM regression model with an optimal error zone value to the prediction of vibration trends for rotating machines. A gearbox lifetime assessment experiment was carried out in order to obtain real vibration signals. Some kinds of vibration data sets were obtained after the vibration signals were processed. The vibration data sets were used as indicators reflecting the health status of the pieces of equipment. There are several indicators that demonstrate the trend of deterioration for different aspects of the deteriorating equipment. The commonly used indicators are amplitude of vibration signal, peak-peak value, root mean square value, and kurtosis value. Among these indicators, the

peak-peak value is chosen as the deterioration trend indicator in our experiment. This is because it is stable and can best present the vibration trend among these indicators. The peak-to-peak value is defined as:

$$peak\_peak = \max[x_i] - \min[x_i]. \tag{6.1}$$

We can see that the peak-peak value measures the difference between maximum amplitude and minimum amplitude. It reflects the total fluctuation of the vibration trend [8, 37].

6.2 SVM with Optimal Error Zone Value for Vibration Trend Prediction

A peak-to-peak value data set is chosen for vibration trend prediction by using the SVM regression model. According to the rule, the error zone value can be fixed using the accuracy of input target value. For this data set, the accuracy of the target value is 10^{-4}; the optimal error zone value should be equal to one-tenth of the accuracy of the input target value. On this basis, the optimal ε should be 10^{-5}. Thus, the prediction program becomes simple because one loop normally used to find the optimal error zone value can be omitted. Through experiments, we found that the "poly" was the best choice among the three kernels in terms of prediction accuracy. The kernel parameter for "poly"is fixed at 1. The second loop used to find the optimal kernel parameter is omitted as well. The prediction program thus become

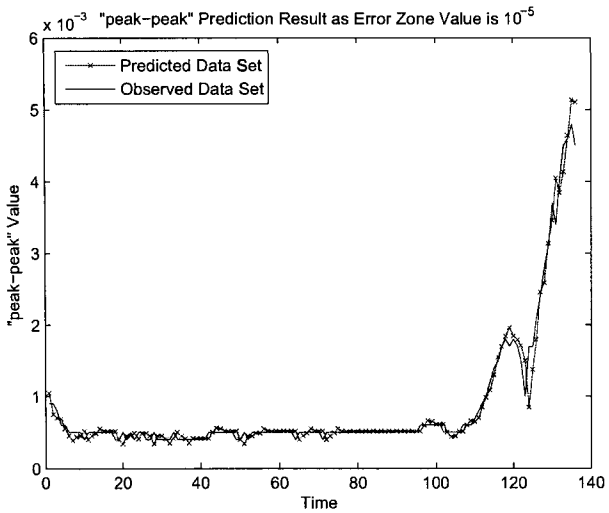

Figure 6.1: The "peak-peak" Data Set Prediction Result 1

very simple, and the predicting time is reduced greatly. It only takes for 54 seconds. On the other hand, it spent 4 minutes and 20 seconds for the prediction without using the rule on the error zone value. The prediction results with error zone value is presented in Figure 6.1. We can see that the SVM regression model used for predicting the degradation status of mechanical equipment can not only offer a total trend but also some details on the degradation status of the equipment. This prediction result is compared with other prediction results as different error zone values is used. The prediction results as error zone value at 10^{-3}, 10^{-4}, and 10^{-6} respectively are plotted in Figures 6.2, 6.3, and 6.4 (see pages 87 and 88). Even virtually, it is clear that the prediction results as error zone value at 10^{-3} and 10^{-4} are worse

than that as error zone value is 10^{-5}; the prediction results when error zone values are 10^{-5} and 10^{-6} are almost the same.

6.3 Discussion and Summary

In this chapter, The SVM regression model has been used to predict vibration trends. The peak-peak vibration data was chosen because it could best present the total vibration trend. The proposed rule on the optimal error zone value has been adopted. It makes the work of prediction much easier and the predicting time is reduced greatly. This is because one loop used to search for the error zone value can be reduced. The error zone value can be set according to the accuracy of the input target value. In this way, the high prediction accuracy is guaranteed and predicting time is reduced greatly.

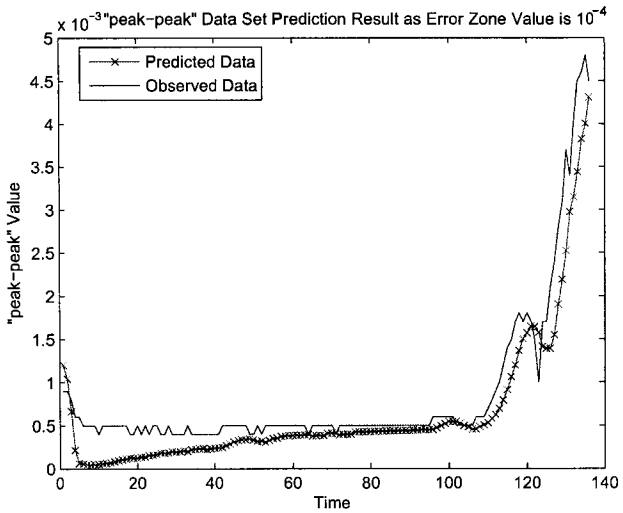

Figure 6.2: The "peak-peak" Data Set Prediction Result 2

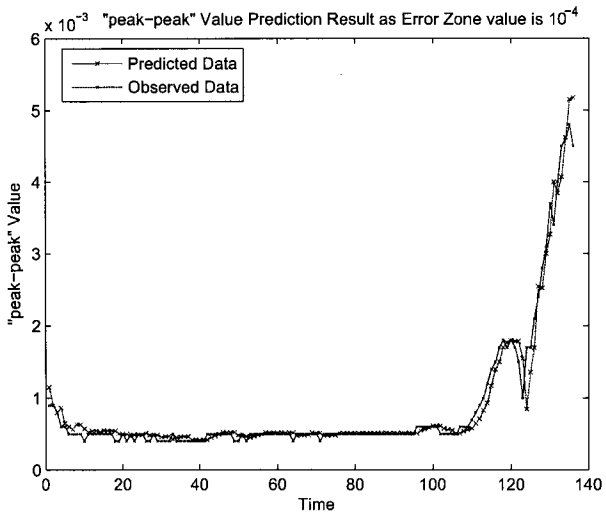

Figure 6.3: The "peak-peak" Data Set Prediction Result 3

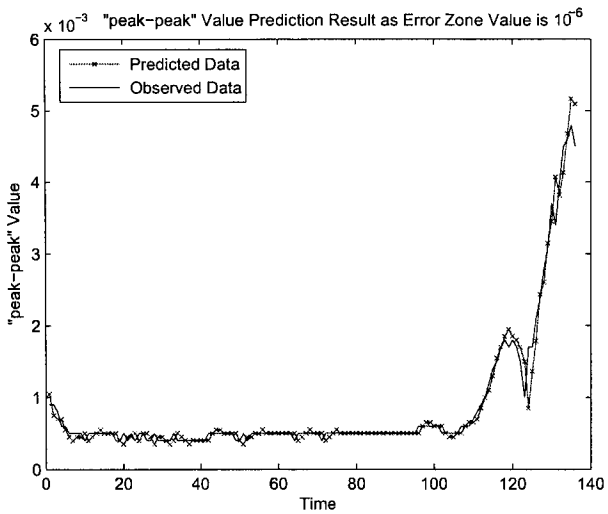

Figure 6.4: The "peak-peak" Data Set Prediction Result 4

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this thesis, the SVM regression model applied for prediction is studied in detail. Two benchmark data sets from the literature and two vibration data sets from the "gearbox lifetime assessment experiment" conducted in the Reliability Lab in University of Alberta are used to investigate the impact of SVM parameters on SVM prediction performance. The results of these are analyzed and the following conclusions are reached.

- The SVM model's parameter, error zone value, can be fixed according to the accuracy of the input target value. The high prediction accuracy can be obtained without time being waste by using the optimal error zone value, which is equal to one-tenth of the accuracy of the input target value.

- The above conclusion is compared with CMa's method and found to

be superior.

- The SVM regression model with optimal error zone value can be used in prediction of mechanical vibration trends. Successful use of the rule on optimal error zone value simplifies the forecasting of machines' degradation status, and greatly reduces predicting time.

7.2 Future Work

Thus, the focus has been on determining how to select one of the SVM regression model parameter, error zone value. Some conclusions have been proposed, however, the prediction performance of the SVM regression model also depends on the other two factors and the rule of selection of the other two, needs to be investigated further. Through the experiment, we found that the value of the regulation factor greatly affects the predicting time. Some research could focus on this point. Moreover, the SVM regression model with optimal error zone value is used in the mechanical engineering field for prediction of machines' degradation status based on vibration signals. In my view, other modern predictors should be applied and tried out in predicting the lifetime of mechanical equipment. As mentioned in the literature review, several papers have compared SVM regression and Neural Networks in some fields. The same work can be done for the prediction of the degradation status of mechanical equipment.

BIBLIOGRAPHY

[1] L. J. Cao, *Support Vector Machines Experts for Time Series Forecasting.* Neurocomputing 51: 321-339 (2003)

[2] L. J. Cao and F. H. Tay, *Support Vector Machines With Adaptive Parameters in Financial Time Series Forecasting.* IEEE TRANSTRACTIONS ON NEURAL NETWORKS, VOL.14, NO.6, NOVEMBER 2003.

[3] E. Chong and S. Zak,*An introduction to optimization.* 2nd Edition, John Wiley and Sons, New York, 2002.

[4] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Mehods.* Cambridge University Press, 2000.

[5] S. M. Clarke and J. H. Griebsch, *Analysis of Support Vector Regression for Approxmation of Complex Engineering Analyses.* Proceedings of the ASME Design Engineering Technical Conference. Vol:2, 535 543.

[6] V. Cherkassky and Y. Q. Ma, *Practical Selection of SVM parameters and Noise Estimation for SVM Regression.* Neural Networks 17 (2004) 113-125.

[7] C. Douglas and A. Elizabeth, *Introduction to linear regression analysis.* New York; Wiley c1982.

[8] R. S. Figliola and D. E. Beasley, *Theory and Design for mechanical Measurements.* Third Edition, John Wiley and Sons, 2000.

[9] S. R. Gunn, *Support Vector Machines for classification and regression.* Technical Report: Imagine Speech and Intelligent systems research group, University of Southampton, 1997.

[10] I. Huseyin, *Short term forecasting with support vector machines and application to stock price prediction.* Intelligent Engineering System Through Artificial Neural Networks Vol.13 737-742, 2003.

[11] Z. Huang, H. Chen and C. Hsu, *Credit rating analysis with support vector machines and neural networks: a market comparative study.* Decision Support System 37 (2004) 543-558.

[12] F. Jiang, *Prediction of Machines' Degradation Based on Vibration Analysis.* Master's Thesis, University of Alberta, 2002.

[13] F. Jiang and M. J. Zuo, *Forecasting of Machine Degradation Based on Vibration Trends and Support VectorMachines.* CSME, Kingston, Canada, 2002

[14] K. Kim, *Financial time series forecasting using support vector machines.* Neurocomputing 55 (2003) 307-319.

[15] V. Kecman, *Learning and soft computing : support vector machines, neural networks, and fuzzy logic models.* Cambridge, MA. MIT Press, 2001.

[16] L. J. Lin, Z. Zhang and Z. He, *Research on Condition trend Prediction of Mechanical Equipment based on Support Vector Machines.* Journal of Xi'an Jiaotong University Vol.38 No3 (2004) 230-238.

[17] J. E. Maria, *Development of Robust Inferential Sensor: Industrial application of support vector machines for regression.* PHD Thesis, School Technische Universiteit Eindhoven (The Netherlands), 2002.

[18] K. R. Muller, *Predicting Time Series with Support Vector Machines.* In *Advances in Kernel Methods- Support Vector Learning.* Cambridge, MA. MIT Press, 1999.

[19] S. Mukherjee, E. Osuna and F. Girosi, *Nonlinear Prediction of Chaotic Time Series Using Support Vector Machines.* Proceedings of IEEE NNSP'97, Amelia Island, Fl, 24-26 Sep., 1997.

[20] S. S. Rao, *Mechanical Vibation.* Fourth Edition, Pentice Hall; March 31, 2003, pp:789.

[21] A. J. Smola, *Regression Estimation with Support Vector Learning Machines.* Master's Thesis, Physik Department, Technische University Munchen, 1996.

[22] A. J. Smola and B. Scholkof, *A tutorial on support vector regression.* NC2-TR-1998-030, NEUROCOLT2 Technical Report Series, 1998.

[23] B. Scholkopf and S. Mika, *Input Space Versus Feature Space in Kernel-Based Methods.* IEEE Transaction on Neural Network, 10(5), 1999.

[24] Z. Sun, L. Gao and Y. Sun, *Using Support Vector Machines for Mining Regression Class in Large Data Sets.* Proceedings of IEEE TENCON'02, 2002.

[25] F. H. Tay and L. J. Cao *Application of support vector machines in financial time series forecasting.* Omega 29 (2001) 309-317.

[26] T. B. Trafalis and H. Ince, *Support Vector Machines for Regression and Application to Financial Forecasting.* IJCNN 2000: proceedings of the IEEE-INNS International Joint Conference on neural Networks, Como, Italy, 24-27 July 2000/Sponsored by IEEE Neural Network Council, International Neural Network. Vol 6, 348 353.

[27] L. Vanajakshi and L. R. Rilet, *A Comparison of the Performance of Artificial Neural Networks and Support Vector machines for the Prediction of Traffic Speed.* 2004 IEEE Intelligent Vechicles Symposium, University of Parma, Parma, Italy, 194-199.

[28] V. Vapnik, *An Overview of Statistical Learning Theory.* IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL.10, NO.5, SEPTEMBER 1999.

[29] V. Vapnik, *The nature of statistical learning theory.* Spring, New York, 1995.

[30] V. Vapnik, S. E. Golowich and A.Smola, *Support Vector Machines for Function Approximation, Regression Estimation, and Signal Processing.* In: Advances in Neural Information Systems. Vol. 9. MIT Press, Cambridge

[31] C. Wu, J. Ho and D. T. Lee, *Travel Time Prediction with Support Vector Regression.* Institute of Information Science, Academia Sinica, Taipei, Taiwan, 2002.

[32] D. Wang and W. Wu, *Application of support Vector Machines Regression in Prediction Shanghai Stock Composite Index.* Wuhan University of Natural Sciences Vol.8 No.4 2003 1126-1130.

[33] S. Zhu and X. Wang, *A SVM Approach to Ship Power Load Forecasting based on RBF Kernel.* Proceedings of the 5^{th}World Congress on Intelligent Control and Automation, June 15-19, Hangzhou, P.R.China, 2004.

[34] Z. Sun, L. Gao and Y. Sun, *Using Support Vector Machines for Mining Regression Class in Large Data Sets.* Proceedings of IEEE TENCON'02, 2002.

[35] *Http://svm.first.gmd.de.*

[36] *Http://www-personal.buseco.monash.edu.au/ hyndman/TSDL/.*

[37] *Http://www.atis.org/tg2k/peaktopeak.value.html*

[38] *Http://mathworld.wolfram.com/root-mean-square.html*

[39] *Http://www.support-vector.net/software.html*

[40] *http://www-personal.buseco.monash.edu.au/ hyndman/TSDL/*