Predicting Time to Failure using Support Vector Regression

Yuan Fuqing
Division of Operation and Maintenance Engineering,
Luleå University of Technology, SE-971 87 Luleå, Sweden
yuan.fuqing@ltu.se

Uday Kumar
Division of Operation and Maintenance Engineering,
Luleå University of Technology, SE-971 87 Luleå, Sweden
uday.kumar@ltu.se

ABSTRACT

Support Vector Machine (SVM) is a new but prospective technique which has been used in pattern recognition, data mining, etc. Taking the advantage of Kernel function, maximum margin and Lanrangian optimization method, SVM has high application potential in reliability data analysis. This paper introduces the principle and some concepts of SVM. One extension of regular SVM named Support Vector Regression (SVR) is discussed. SVR is dedicated to solve continuous problem. This paper uses SVR to predict reliability for repairable system. Taking an equipment from Swedish railway industry as a case, it is shown that the SVR can predict (Time to Failure) TTF accurately and its prediction performance can outperform Artificial Neural Network (ANN).

Keywords

Support Vector Machine; Support Vector Regression; Kernel Function; Crossings and Switches; Time to Failure.

1. INTRODUCTION

As a new trend of statistical learning theory [7,8], Support Vector Machine (SVM) is gaining popularity and lots applications of it have been found in state of art. The principle of SVM is shown in Figure 1. Provided a set of empirical data (x_i,y_i), x_i, y_i denotes the input and output value(response variable) respectively. SVM will adjust its internal parameters to fit the input data. SVM tries to minimize the margin between real output y_i and predicted $\underline{y}_i$. Furthermore, take an example of binary classification and suppose the real output and the predicted output is identical, i.e. margin error is zero. SVM then further employs another tactic named regularization to maximize distance of the two collections, where each collection leads to one output class. Section 2.3 presents an example to demonstrate it.

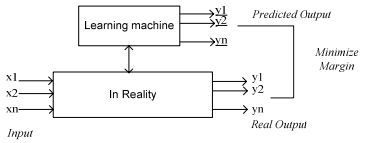

Figure 1. SVM Leaning Process

This paper first describes the basic theory of SVM. Later on the paper presents an application of Support Vector Regression (SVR) to predict Time To Failure (TTF) for Crossings and Switches (C&S) from railway industry.

2. Basic of SVM

2.1 Principle of SVM

SVM was originally developed as classifier. Regression analysis and principle component analysis can be regarded as special case of classifiers. A simple illustration is introduced here to demonstrate the principle of SVM. As shown in Figure 2, there are two kinds of dots (black and white). Suppose one line existing to separate them. Obviously the line should locate between these two groups, as the solid line in Figure 2. The line is called separator or **Hyperplane**. SVM tries to find the optimal line which maximizes the distance between these two groups of dots. Usually, Lagrangian optimization approach is used to find the optimal line.

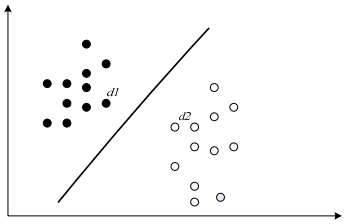

Figure 2. Separator for dots

Support Vector: From geometry's view, each dot can be represented by an n-dimension vector, for example, in Figure 2, each dot can be represented by a 2-dimension vector. d_1, d_2 are nearest and locate in the edge of their respective group. When the solid line in Figure 2 is selected as their separators, only d_1, d_2 take effect. d1, d2 is the so-called support vectors.

Learning Machine (Machine): Imitating the learning process of human being, SVM is capable of learning from given empirical data. During leaning, SVM extracted features from these empirical data. From statistical learning's point of view, the process is automatically completed. So it is called learning machine.

2.2 Architecture of SVM

Figure 3 illustrates the architecture of SVM. SVM obtains its hyperplane (i.e. separator) from a given training data set. Lagrangian optimization method is used to obtain optimal solution during this procedure. After training, the SVM is built up, i.e. the hyperspace is found. After that, given an input data set, SVM can generate its predicted output value by the hyperplane, e.g. identify the group the dot belonging to.

ISBN 978-91-7439-120-6

Kernel function comprises a methodology to measure the distance between two vectors. Kernel Function can transform the input vector. For example, transform the lower dimensions input vector to higher dimensions. In kernel function, inner dot product is the most simple and common measurement to measure distance.

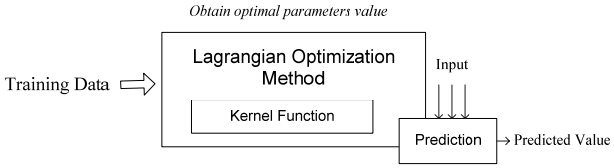

Figure 3. Architecture of SVM

2.3 A Simple Linear SVM

In order to illustrate SVM, a simple SVM named linear SVM classifier is presented in this section. The objective of this SVM is to classify the dark dots from white dots as shown in Figure 4. Obviously, any line locating between the nearest dots (d_1,d_2) between these two classes can separate these two classes. Sharing a same normal line W_1, among L_1, L_2, L_3, the most reasonable separator line should be L_2. The distances from d_1 to L_2 and from d_2 to L_2 are equal.

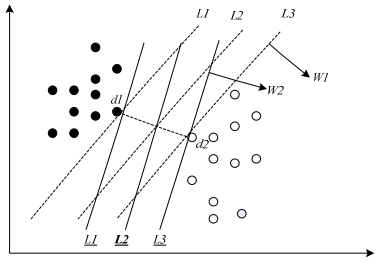

Figure 4. Linear SVM

Other than these, there are some other lines can separate these two classes with another normal line, e.g. L_1, L_2, L_3. In order to separate the two classes as far as possible, the line with normal line perpendicular to line d_1d_2 is optimal, i.e. W_2 is selected as the optimal line's normal line. So, the Line L_2 is the best (optimal) separator. Formulate the problem into a constrained optimization problem.

$$\min_{w\in h, b\in R} \frac{1}{2}\|w\|^2 \qquad (1)$$
$$s.t. \quad y_i(\langle w, x_i\rangle + b) \geq 1, \quad i = 1,2,....,m.$$

Each dots in Figures 4 has been transformed to a constraints. SVM uses Lagrangian Relaxation Method to obtain optimal solution. By introducing Lagrange multipliers, Formula (1) is rewritten to:

$$\begin{aligned} max \quad & W(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle \\ s.t. \quad & \alpha_i \geq 0,\ i = 1,2,....,m \\ & \sum_{i=1}^{m}\alpha_i y_i = 0 \end{aligned} \qquad (2)$$

Nevertheless, in practice, the separator may not exist, i.e. there are no any line can separate these dots. Then error tolerance should be introduced. We loose these constraints in Formula (1) by introducing slack variables. The corresponding Lagrangian Formulation is thereby written to:

$$\begin{aligned} & \max \sum_{i=1}^{m}\alpha_i - 1/2\sum_{ij}^{m}\alpha_i\alpha_j y_i y_j K(x_i, x_j) \\ & s.t. 0 \leq \alpha_i \leq C \\ & \sum_{i}^{m}\alpha_i y_i = 0 \end{aligned} \qquad (3)$$

The optimal hyperplane can be obtained from above formulations. Given new x_i, the following formulation is used to decide which class the dot belong to.

$$f(x_i) = \mathrm{sgn}(\sum_{j=1}^{m}\alpha_j y_j K(x_i, x_j) + b) \qquad (4)$$

2.4 Support Vector Regression

Support vector regression estimates continuous function given definite training data sets. Analogous to support vector classifier, SVR uses soft margin to separate features which denote the characteristics of the desired function $f(x)$. Usually ε -insensitive loss function (See Scholkopf and Smola 2002) is used to balance the accuracy of approximation and computation complexity[1,2]. A constrained optimization problem is constructed to approximate the desired function as follows:

$$\begin{aligned} & \min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \\ & s.t. \ \ f(x_i) - y_i \leq \varepsilon + \xi_i \\ & \quad y_i - f(x_i) \leq \varepsilon + \xi_i^* \\ & \quad \xi_i \geq 0, \xi_i^* \geq 0.\ i = 1,2,3,....m. \end{aligned} \qquad (5)$$

The above formula is called primal problem. By introducing Lagrangian multipliers, the corresponding dual problem of Formula (5) is:

$$\begin{aligned} & max \ \ W(\alpha, a^*) = \\ & -\varepsilon\sum_{i=1}^{m}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{m}(\alpha_i^* - \alpha_i)y_i \\ & -\frac{1}{2}\sum_{j=1}^{m}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\langle x_i, x_j\rangle \\ & s.t. \quad \alpha_i \geq 0,\ \alpha_j \geq 0; \\ & \quad \alpha_i^* \leq C, \alpha_j^* \leq C; \\ & \quad \sum_{i=1,j}^{m}(\alpha_i^* - \alpha_i) = 0,\ i = 1,2,....,m \end{aligned} \qquad (6)$$

Substitute $\langle x_i,x_j\rangle$ in above Formula(6) with $K\langle x_i,x_j\rangle$, which is called kernel function. The desired function is approximated as follows[5,6]:

$$f(x) = \sum_{i=1}^{m}(\alpha_i^* - \alpha_i)K(x_i, x_j) + b \qquad (7)$$

3. A Case Study

3.1 Switches and Crossings

The railway infrastructure of Sweden has in total 17,000 km of railway and about 12,000 switches and crossings. Crossings and

ISBN 978-91-7439-120-6

Switches (C&S) is a mechanical installations enabling railway trains to be guided from one track to another at a railway junction [4],and allowing slower trains to be overtaken. Functions of an S&C can be summarized as:

- Carry load
- Be part of the track signaling circuit.
- Act as flange protection.
- Move the switch blade to enable one of two or more alternative ways.
- To enable train to move from sidings and re-enter main track.

Figure 5. A kind of Switches and Crossings

The life of S&C is approximate 40 years. As a result of analysis for cause of delay time, the C&S related failures contribute 14% of total causes of train delay time. "That means roughly 15 minutes delay time per S&C/year, assuming on C&S in main track per 2km and 50 trains per day". Furthermore, the C&S cost covers at least 13% of the total maintenance cost. Consequently, C&S plays an important role in railway industry both from functional and financial point of view [4]. Analyzing the reliability of C&S is hence necessary.

3.2 Data Collection

The data has been collected from Swedish Banverket's asset register system BIS and failure reporting system 0felia. Database BIS is collecting all features concerning C&Ss, such as *track section*, *C&S type*, *Put in place year* and so on. 0felia is collecting data covers *Failure reported data and time*, *time for maintenance*, *failure symptom* and so on.

In order to demonstrate the proposed methodology, we select a C&S in Section 111 as example, which locates between Kiruna and Riksgränsen in Northern Sweden where failures are more frequent due to severe natural weather. The data concerning *asset No.1* is collected. This asset is put into service in 2005. We collected the data from 2005 to 2007. Part of these data is tabulated as follows.

Table 1. Failure Report for S&C

Report ID	*Failure Declaration Time*	….
FR00324115	2005-09-15 06:05	…
FR00325402	2005-09-24 08:52	
FR00326198	2005-09-30 14:19	
FR00326495	2005-10-03 17:22	
FR00327341	2005-10-10 06:48	…
…	…	…
FR00332014	2005-11-14 20:43	
FR00347162	2006-02-18 15:18	
FR00347301	2006-02-19 23:31	
FR00349796	2006-03-05 15:27	
FR00350165	2006-03-06 23:02	
…	...	

3.3 Data transformation

In order to facilitate failure analysis, the calendar failure time in Table 1 is transformed into accumulate time to failure. It is tabulated in Table 2.

Table 2. Transformed Time to Failure Data

No.	*Failure Time*	*Time To Failure*
1	2005-09-24 08:52	218,7863889
2	2005-09-30 14:19	149,4422222
3	2005-10-03 17:22	75,06027778
4	2005-10-10 06:48	157,4258333
5	2005-10-13 21:18	86,50583333
6	2005-10-23 14:57	233,6477778
7	2005-11-05 14:43	311,7713889
8	2005-11-14 08:10	209,4413889
9	2005-11-14 20:43	12,55361111
10	2005-12-27 06:51	1018,124722
	…	

3.4 Predict Time to Failure

Use the approach described in Section 2.4. Among the 41 data sets tabulated in Table 2. Take last 4 data to validate the model. There are 37data sets therefore remaining. In order to obtain the optimal parameter for SVR for this problem, last 4 out of the remaining 37 data sets are used to supervise the parameter tuning. Using these 4 out of 37 data sets, the desired optimal parameter is parameter with a minimum total error between predicted *TTFs* and real *TTFs*.

After the optimal parameter obtained, we take the whole 40 data sets as training data to train the SVR again. Then use the trained SVR to predict *TTFs* for the remaining 4 data sets.

Table 3 n-step lagged input array and its corresponding output

x(input)					y(output)
x_1	x_2	….	x_{m-1}	x_m	x_{m+1}
x_2	x_3	….	x_{n-2}	x_{n-1}	x_{m+1}
…	…	….	…	…	…
x_{n-m}	x_{n-m+1}	….	x_{n-2}	x_{n-1}	x_n

SVR approach also uses n-lagged time series as input to train and predict the *TTF*. The input of the SVR is shown in Table 3. After trying of several n-lagged time series (change the step size n), the

ISBN 978-91-7439-120-6

best performance comes out with 2-lagged time series. The consequently predicted *TTFs* are tabulated in Table 4, along with the predicted result yielded from the other techniques: Artificial Neural Network (ANN)[3].

Table 4. Comparison of real TTF with predicted TTF

No.	Real TTF	ANN	SVM
1	1211,4	1200,69	1204,30
2	1212,0	1197,60	1212,30
3	1220,54	1210,06	1222,23
4	1223,49	1212,30	1224,56

In order to facilitate the comparison among different techniques, the data in Table 4 are plot in the Figure 6. It is shown that the SVR outperform ANN.

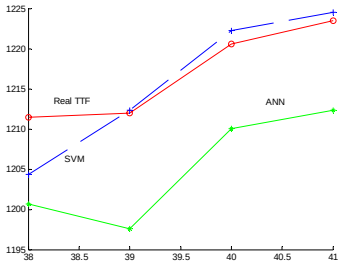

Figure 6 Predicted *TTFs* and real *TTFs*

Measures *TE*, *MTE* and *NRMSE* are used to compare their performance. They are tabulated in Table 5.

Table 5. Comparison of real TTF with predicted TTF

	TOTAL ERROR	SSE	MSE	NRMSE
ANN	36.6347	558.4198	139.6049	0.0097
SVR	4.1131	55.4053	13.8513	0.0031

According to all the three measures, the total error of SVR is less than the corresponding ANN. Consequently, the proposed SVR can obviously outperform ANN for this case.

4. Conclusion

By comparing the results predicted both from ANN and SVR, it is evident that SVR can outperform ANN. The result experimentally proves the regularization part of SVM is effective.

5. References

[1] Fuquing Y., Kumar U., Claudio M. Rocco S. and Misra K. B. (2009) Complex System Reliability Evaluation using Support Vector Machine, Proceedings SMRLO10, Israel.

[2] Gunn, S.R. (1998) Support Vector Machines for Classification and Regression. Technical report, School of Electronics and Computer Science, University of Southhampton.

[3] Harvey,R.L. (1994)Neural network principles. Englewood Cliffs: Prentice Hall.

[4] Nissen,A. (2009) Development of Life Cycle Cost Model and Analyses for Railway Switches and Crossings. Doctoral Thesis: Lulea University of Technology.

[5] Smola, A.J., Muller, K.R. (1998). General cost functions for support vector regression. *Proceedings of the 8th International Conference on Artificial Neural Networks.*

[6] Scholkopf. B. and Smola A.j. (2002) Learning with kernels. The MIT Press: London.

[7] V.Vapnik, (1995)The Nature of Statistical Learning Theory. New York, Springer.

[8] V.Vapnik, (1998) Statistical Learning Theory. New York, John Wiley & Sons, Inc.

ISBN 978-91-7439-120-6