

Research on LSSVM based on analytic selection and simulate anneal arithmetic

Yang Dan^{1, a}, Zang Lie^{2, b}, Gan Lu^{3, c}

¹College of computer science and technology, Nanjing university of aeronautics and astronautics, Nanjing, 211106, China

²College of computer science and technology, Nanjing university of aeronautics and astronautics, Nanjing, 211106, China

³College of computer science and technology, Nanjing university of aeronautics and astronautics, Nanjing, 211106, China

^aemail: miemieduan@163.com, ^bemail: zangliwen@nuaa.edu.cn, ^cemail: 1321710271@qq.com

Keywords: Software Reliability Model; Least Square Support Vector Machines (LSSVM); Analytic Selection (AS); Simulate Anneal Arithmetic (SA)

Abstract. In view of the existing situation of artificial parameter selection in the fitting process which using the LSSVM software reliability model, the combination method of analytic selection (AS) and simulate anneal arithmetic (SA) is proposed for parameter selection, then obtain the improved software reliability model based on ASSA-LSSVM. The ASSA-LSSVM model considers the optimization of the penalty parameters C and kernel parameter σ in the model, make the model more realistic, and reduce the time cost of artificial selection-parameter. Make use of the published three groups of software failure data to verify this proposed model, the result shows that, compared with the former LSSVM model, the ASSA -LSSVM model can achieve better fitting results.

Introduction

The existing Software Reliability Models [1] (SRMs) are mostly based on accessing to plenty of software failure data and then use these data to achieve the accurate prediction of software reliability. The most of the research works for SRMs are focused on the Nonhomogeneous Poisson Process model, the renewal model and the time series model [2]. Some of the others can be used in software reliability analysis include the model based on fault tree, the model based on the number of error, the model based on neural network, the model based on test coverage and the model based on the structure of system, etc. [3]. These models are subjected to the strict assumptions restrictions, and easy to generate a lot of overhead time and space. In addition, due to the cost of the required time and money of locating and repairing the bugs of the large software increased, the user eager to get higher efficiency and easier used software reliability model[4] [5].

Support Vector Machines (SVM) [6] is a machine learning method based on statistical learning theory, it's simply structured, versatility, and solving the problem with global optimality [7]. And Least Squares Support Vector Machines (LSSVM) is the improvement version of general SVM [8], software reliability model which based on LSSVM, with the advantage of no assumptions, strong applicability and pretty generalization.

In this paper, we compared and analysed the existing software reliability models, and the analyses is the software reliability model of LSSVM based on simulated annealing algorithm. At the same time, based on this model, use the analysis selection to optimize penalty parameter C and nuclear parameter σ in the software reliability model based on SA-LSSVM, software reliability model based on ASSA-LSSVM is constructed. Through the improvement of the model, which make the optimized parameters in the model, and complete the structure of the optimized model, reducing the influence of artificial factors effectively, and also reduces the effort of construct the model. Finally, simulation results of the proposed method verify the feasibility and effectiveness.

The existing software reliability models

Presently, the most common software reliability models are sowing seeds model, curve fitting model and reliability growth model, the Markov process model and Non-Homogeneous Poisson Process(NHPP)model [9]. Among them, the Nonhomogeneous Poisson Process model [10] is important one in software reliability growth model, and the research of NHPP are increased now, it also is a widely used model. NHPP model provides an analysis framework used to describe a software failure phenomenon during the test. According to the different assumptions, the mean value function expression of the model will differ in partial. According to the assumptions of NHPP, $N(t)$ obeys the Poisson distribution of averages $m(t)$, i.e:

$$\Pr\{N(t) = k\} = \frac{[m(t)]^k}{k!} e^{-m(t)} \quad k = 0, 1, 2, \dots \quad (1)$$

According to the definition, the mean function $m(t)$ is derived by the software failure density function, i.e:

$$m(t) = \int_0^t \lambda(s) ds \quad (2)$$

The available software reliability expression is:

$$R(t) = e^{-m(t)} = e^{-\int_0^t \lambda(s) ds} \quad (3)$$

In this paper we choose three kinds of NHPP software reliability model as comparison model, and compared with the proposed software reliability model based on ASSA - LSSVM in this paper.

Model I: Goel and Okumoto put forward NHPP model as early as in 1979, which was G-O model [11], this model regards the failure detection as a process of NHPP. Its mean value function is $m(t) = a(1 - \exp(-bt))$, parameters $a(t) = a$, $b(t) = b$.

Model II: Yamada of S-Shaped model [12] can be seen as G-O model which a failure rate increases at the beginning and the failure rate decreases at the end. In this model, the examination process of the software failure data is an s-shaped curve, which manifests the tester is not familiar with the software at the beginning of the test and the failure data is difficult to find in the late. Its mean value function is $m(t) = a(1 - (1 + bt)\exp(-bt))$, parameters $a(t) = a$, $b(t) = b^2 t / (1 + bt)$.

Model III: Inflected S-Shaped model [13] is the improvement of the type of G-O, this model assumes that the failure of the software is independent of each other, and then observed the software reliability growth curve is S type. Its mean value function is $m(t) = a(1 - \exp(-bt)) / (1 - \beta \exp(-bt))$, parameters $a(t) = a$, $b(t) = b / (1 - \beta \exp(-bt))$.

There are still some parts can be perfected in NHPP model when it is applied to practical problems, such as the target value of the model calculation is not enough to reflect the change features of software reliability, there are many constraints in use and the model itself will be more complicated if need to estimate more parameters, and estimate of the parameter error is bigger when the sample data is small and so on.

Least squares support vector regression machine

Assume that the training set consists of a sample points

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \quad (4)$$

Where $x_i \in R^n$ is input vector, $y_i \in R$ which is corresponding to the output.

The original optimization problem of Least squares support vector machine as follows:

$$\begin{aligned} \min_{\omega, b, e} J(\omega, b, e) &= \frac{1}{2} \omega^T \omega + \frac{C}{2} \sum_{k=1}^l e_k^2 \\ \text{s.t. } y_k &= \omega^T \phi(x_k) + b + e_k, \quad k = 1, \dots, l \end{aligned} \quad (5)$$

Where ω is the weight vector, C is the regularization parameter, e_k is the error variables, $\phi(\cdot)$ is the nonlinear mapping from input space to the high dimensional feature space, b is a partial quantity.

By Lagrangian dual method, we can get a quadratic optimization problem solution and the estimated output value is:

$$m(x) = \sum_{i=1}^l \alpha_i K(x_i, x) + b \quad (6)$$

$K(x_i, x)$ is the kernel Function, it is an important part of the support vector machine, the common used kernel functions are: linear kernel function, polynomial kernel function, Gaussian kernel function (Radial Basis Function, RBF), Sigmoid kernel function and composite kernel function. The Gaussian kernel function is a universal, widely applied kernel function, so this paper will research the failure model based on the Gaussian kernel function, i.e:

$$m(x) = \sum_{i=1}^l \alpha_i e^{(-\|x_i - x\|^2 / \sigma^2)} + b \quad (7)$$

SA-LSSVM software reliability model

We can see from formula (7), the optimization problem of LSSVM model based on RBF contains its optimal solution of the parameters. Usually we use the common parameter optimization algorithm such as: simulated annealing algorithm, genetic algorithm, particle swarm optimization algorithm (PSO) and differential evolution algorithm to the optimal solution of these parameters.

Simulated annealing algorithm [14] is a kind of widely used random search algorithm, it come from the extension of local search algorithm, and usually used in a large search space to search for the optimal solution of the proposition. In order to reduce the influence of the model's performance by the uncertainty of parameters C and σ , using simulated annealing algorithm for automatic optimization of these two parameters, in order to obtain the SA-LSSVM model for software reliability. Select Error Sum of Squares (SSE) as the objective function of the algorithm.

Setting the algorithm operating conditions:

% Setting the parameter of the cooling schedule

MarkovLength = 100; % the length of Markov chain

DecayScale = 0.95; % attenuation parameter

StepFactor = 0.02; % the factor of step length

Temperature = 10000; % initial temperature

How = 1e-6; % tolerance, the termination conditions of the algorithm

ASSA-LSSVM software reliability model

In order to improve the accuracy of the regression or fitting of software reliability model based on SA-LSSVM, improve the convergence speed, we think we can narrow the scope of the parameter value and gain better optimization effect along the direction to optimize at the same time in the search process of SA. In order to achieve this effect, we introduced analytic selection on specific search optimization problem which used SA.

Many related work about support vector machine are referred to close correlation of the optimal parameter and statistical characteristic of data selection. Mattera came up with that a "good" parameter C can be selected from the range of output values [15]:

$$C = \max(|\bar{y} + 3\sigma_y|, |\bar{y} - 3\sigma_y|) \quad (8)$$

Where \bar{y} is the mean value of the output value, σ_y is the standard deviation of the output value.

Same for the nuclear parameter values, the selection of it must reflect the variability of input vector. Considering the single variable input, simply σ can be set to [16]:

$$\sigma \sim (0.1 - 0.5) \times rang(x) \quad (9)$$

Where $rang(x) = |\max(x) - \min(x)|$. Formulas (8)-(9) are analytic selection, which is composed of the parameter combination of penalty parameter C and kernel parameter σ .

Using analytic selection to guide the simulated annealing algorithm optimization search, the

selected parameter value is used to adjust the cooling process of simulated annealing algorithm. In each cooling step we embed a simple operation named drifting anneal to modify the cooling process. The software reliability model based on ASSA-LAAVM includes the following main steps, flow diagram is shown in figure 1:

Table 1 The fitting results comparison of different models

name	SSE	SSE	SSE
	(NTDS)	(TCSDP)	(ATTSTP)
Model I	131	104870	27.37
Model II	174	1544	31.5
Model III	187	30015	28.97
SA-LSSVM model	67	556.6	22.6
ASSA-LSSVM model	53	486.9	14.1

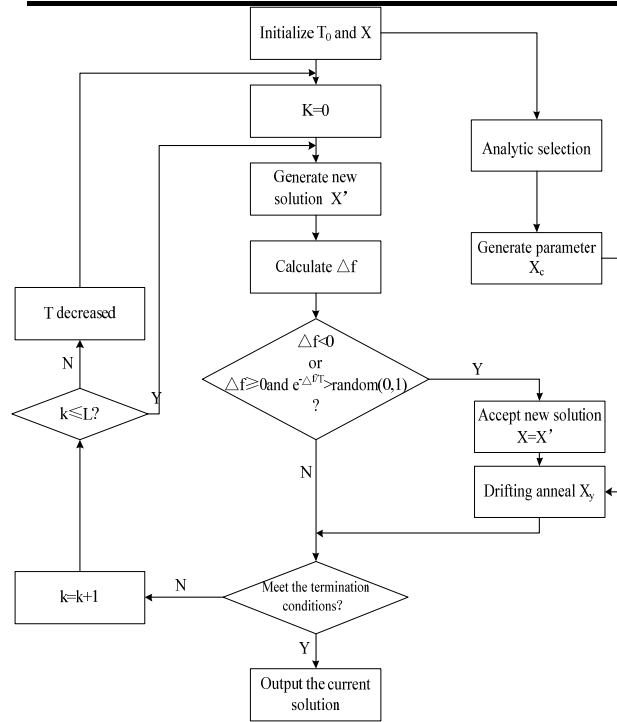


Fig.1. Flow diagram of ASSA-LSSVM

(1) Representation: The solution state x is directly expressed as parameters combination $x = [c, \sigma]$.

(2) The AS method: Through the application of formulas (8)-(9) in input vector, analytic selection method determines the value of the parameter vector x_c .

(3) Initialization: Initial temperature T_0 (sufficiently large), the initial solution state x is the iterative starting point of algorithm, the number of iterations for each T value is L .

(4) Generate new solution: The objective function from the current solution creates a new solution which locates in the solution space.

(5) Objective function: Computing the difference of the objective function, that is $\Delta f = f(x') - f(x)$, where $f(x)$ is the objective function SSE.

(6) Metropolis principles: Determine whether the new solution x' is accepted. If $\Delta f < 0$, then accepted x' as a new current solution and used x' as the initial point of the next simulated annealing; otherwise computing new acceptable probability $P(\Delta f) = e^{-\Delta f / KT}$, creating the pseudo random

numbers r on uniform distribution on $[0, 1]$. If $P(\Delta f) > r$, then accepted x' as a new current solution and used x' as the initial point of the next simulated annealing; otherwise giving up the new solution still take the original point as the initial point of the next simulation.

(7) Drifting anneal: newly generated solution is based on the AS reference vector, after the drifting annealing:

$$X_y = X' + 1/AS_factor \cdot (X_c - X') \quad (10)$$

Where AS_factor is a parameter which controls the influence of modification: the greater the value of AS_factor , the smaller influence on the simulated annealing algorithm.

(8) Termination conditions: If meeting the termination conditions, output the current solution as the optimal solution, end the program.

Software reliability model based on ASSA-LSSVM has the problem in the process of construction: excepting obtain the optimal solution of the structured calculation based on least squares support vector machine, there's still needing to optimize some other factors which affect the quality of data fitting, such as parameter C and σ (take RBF kernel function as an example). to make the whole structure of the model optimization. And the selection of the parameters C and σ in the model are artificial, and the good or bad of values rely on the expert experience heavily, laboriously. In this paper, we improve the SA-LSSVM software reliability model by analytic selection method, and make the parameters combined with characteristics of statistical data and prior knowledge in the process of automatic optimization to obtain the improved model.

Experiment study

This paper selects NHHP model, software reliability model based on SA-LSSVM and the improved software reliability model based on ASSA-LSSVM to contrast and analyze. In order to verify the validity of the model, and verify the stability of the evaluating and predictive ability, we use three groups of classic data collection of NTDS, TCSDP and ATTS as reference object for the verification of the model, to verify this model.

By the SSE parameters in table 1 we can see, the ASSA-LSSVM model is optimal in five models, and its error sum of squares is smaller compared to the SA-LSSVM model before improved, and it shows that it can be more accurately for fitting of failure data.

Experimental results of the data set of NTDS, TCSDP, ATTSTP are shown in figure 2, 4, 5 respectively, and the data set of NTDS, ATTSTP are complete data sets. It can be seen from the figure 3, 6 that ASSA-LSSVM model is better in the later fitting and closer to the actual data values. Data set TCSDP is an incomplete data set, and it can be seen from the figure 4 that ASSA-LSSVM model's improvement of fitting effect is small. It can be seen from the experimental results that using the combination method with analytic selection and simulated annealing algorithm, the failure model that constructed by the optimized LSSVM can further improve the accuracy of parameters C and σ , which provide a more optimization analysis model for software reliability assessment.

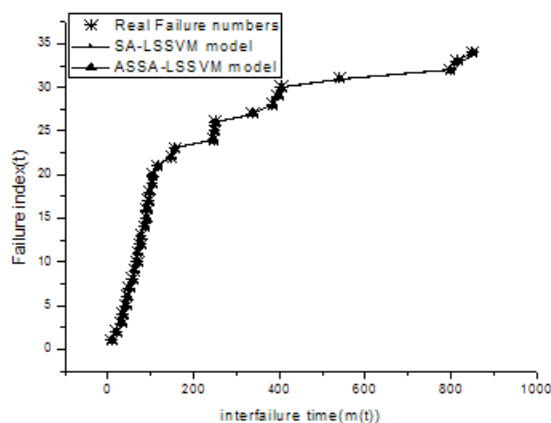


Fig.2.

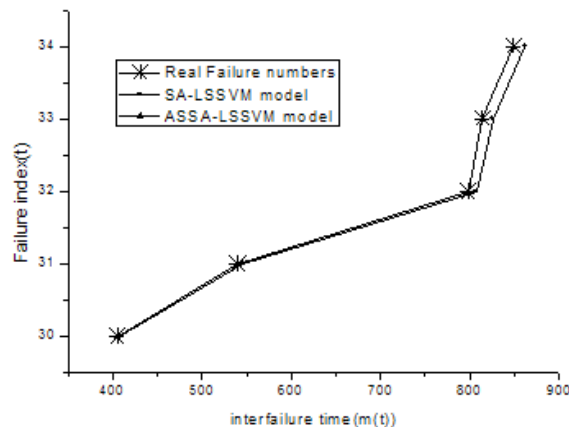


Fig.3.

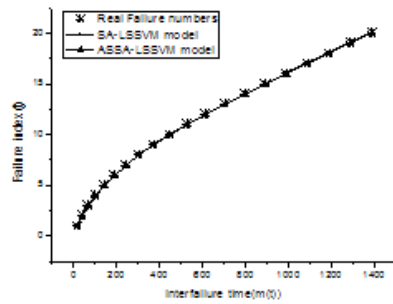


Fig.4.

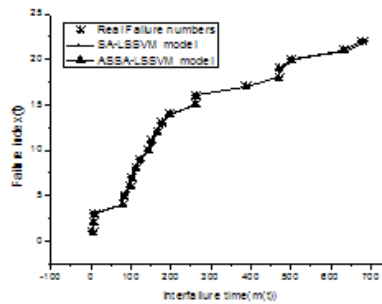


Fig.5.

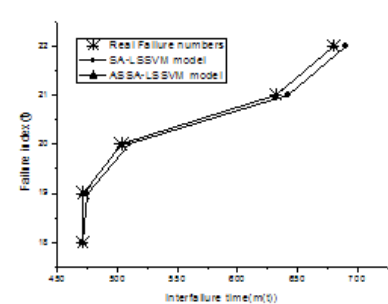


Fig.6.

Conclusion

In this paper, we introduced some existing software reliability models, analyzed emphatically the software reliability model based on LSSVM. And on the basis of simulated annealing algorithm to realize the automatic optimization, we combine with the analytic selection method to improve the software reliability model based on SA-LSSVM. Aimed at combining statistical characteristic data and prior knowledge to completed the structure of SVM, at the same time giving the optimal values of the parameters that do not constitute a support vector regression machine's optimal solution, so as to avoid the influence of artificial factors, also reducing the workload of model, to make the software reliability model based on ASSA-LSSVM more conformed to the actual fitting. This paper uses three representative groups of the software failure data sets to conduct simulation experiment with the software reliability model based on ASSA-LSSVM, the experimental results show that the improved model on the failure data set can obtain good fitting effect, so that they can get more accurate evaluate and forecast the software reliability.

References

- [1] Pham H. System software reliability [M]. Springer, 2007.
- [2] Bing Chao. The Classification of Software Reliability Models and Failure Analysis of Software Reliability Based on Support Vector Machines [D][D]. , 2010.
- [3] Chao B, Xu R Z. Software Failure Model Based on Optimized RLSSVM[J]. Computer Engineering, 2011, 5: 030.
- [4] Ling-zhang Wang. Software reliability method. Beijing, China machine press, 2012.
- [5] Ren-zuo Xu. Software reliability engineering. Beijing, Tsinghua university press, 2007.
- [6] Xiao-fang Yuan, Yao-nan Wang. Selection of SVM Parameters Using Chaos Optimization Algorithms [J][J]. Control and Decision, 2006, 21(1): 111-113.
- [7] Li L M, Wen G R, Wang S. Parameters selection of support vector regression based on genetic algorithm[J]. Computer Engineering and Applications, 2008, 44(7): 23-26.
- [8] Zhang X, Zhao Y, Wang S. Reliability prediction of engine systems using least square support vector machine[C]//2011 International Conference on Electronics, Communications and Control (ICECC). 2011.
- [9] Chin-Yu Huang, Michael R.Lyu.A Unified Scheme of Some Nonhomogeneous Poisson Process Models for Software Reliability Estimation [J]. IEEE Transaction on software engineering, 2003, 29(3):261~269
- [10]XIE J Y, AN J X, ZHU J H. NHPP Software Reliability Growth Model Considering Imperfect Debugging [J][J]. Journal of Software, 2010, 5: 009.
- [11]Goel A L, Okumoto K. Time-dependent error-detection rate model for software reliability and

other performance measures[J]. IEEE transactions on Reliability, 1979, 3: 206-211.

[12]Xie M. Software reliability modelling[M]. World Scientific, 1991.

[13]Lyu M R. Handbook of software reliability engineering[M]. CA: IEEE computer society press, 1996.

[14]Shou-Heng T U O. A Global Memetic Optimization Algorithm for Solving High-Dimensional Problems Based on Differential Evolution and Simulate Anneal [J][J]. Computer Systems & Applications, 2012, 2: 022.

[15]V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Netw. 17 (2004) 113–126.

[16]Zhao W, Tao T, Zio E. System reliability prediction by support vector regression with analytic selection and genetic algorithm parameters selection[J]. Applied Soft Computing, 2015, 30: 792-802.