2012 International Conference on Solid State Devices and Materials Science

Wick Sintered Temperature Forecasting Based on Support Vector Machines with Simulated Annealing

Zeng Dehui*a,c, Liu Yuan*a, Jiang Liano*a, Li Liao*a, Xu Gang*a,b,*

* College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, 518060
*b Key Laboratory of Advanced Optical Precision Manufacturing Technology of Guangdong Higher Education Institutes, Shenzhen University, Shenzhen, 518060, China
*c School of mechanical and automotive engineering, South China University of Technology, Guangzhou, 510640

Abstract

The law of sintering temperature’ changing during resistance furnace sintering is the very important technology information. Support vector machines (SVMs) have been successfully employed to solve nonlinear regression and time series problems. In order to improve time efficiency of prediction, a new sintered furnace temperature law prediction model and method based on SVM in this paper. Moreover, simulated annealing (SA) algorithms were employed to choose the hyperparameters of a SVM model. A comparison of the performance between SVM optimized by Particle Swarm Optimization (SVM-PSO) and SVM-SA is carried out. Experiments results demonstrate that SVM-SA can achieve better accuracy and generalization than the SVM-PSO. Consequently, the SVM-SA model provides a promising alternative for sintered furnace temperature law.

© 2012 Published by Elsevier B.V. Selection and/or peer-review under responsibility of Garry Lee
Open access under CC BY-NC-ND license.
Keywords: support vector regression, simulated annealing, sintered, temperature forecasting

1. Introduction

With the development of microelectronic technique, power consumption and heat dissipation have become the bottleneck that hinders microelectronic technique from further development [1]. Efficient cooling is important to the successful operation of modern electronic devices. The amount of heat generated in these electronic devices must be dissipated in order to maintain the operating temperature limit. One of the most promising candidates for the electronic components is heat pipes, which offer a better alternative to liquid cooling provided they can handle the heat flux and power requirements.

* Corresponding author.
E-mail address: xgrobot@126.com
Micro heat pipes (MHPs) have many good characteristics, such as high heat conductivity, good isothermal temperature, variability of heat flux, reversibility of heat flow direction, thermodiode, heat switch performance, temperature steadiness and adaptability to environment\textsuperscript{[2–4]}, and greater heat transfer efficiency that is several hundredfold or even thousand fold of that of copper rod. As MHP’s heat transfer capability, to a large extent, depends on the capillary structure of its wick. Sintered powder metal wicks offer several advantages over other wick structures. An emerging advantage of the sintered powder wick is its ability to handle high heat fluxes with usually low thermal resistance. Since sintered powder wicks are generally more than 50% porous, there is, accordingly, a large surface area for evaporation. Another advantage of a heat pipe with a sintered powder wick, especially for the cooling system of electronic devices, is that it can work in any orientation. Since a sintered powder wick is integral with the heat pipe envelope, and there is enough fluid charge just to saturate the wick, the heat pipe can be subjected to freeze/thaw cycles without degradation in performance. The above attributes make the sintered powder wick the optimal structure for many thermal management solutions.

Manufacturing of MHPs’ sintered wicks has rigorous demands for quality, purity and cell size of a single copper powder, sintering temperature and sintering uniformity. The whole apparatus formed was then placed on the steadier and sintered at some certain sintering temperatures in RXL–12–11 resistance furnace for some durations of time. The appropriate sintering technological parameters were determined, which could help to solve the technical problem of sintering a thin layer of copper powders in inner wall of slender copper tube with a larger porosity, so as to provide some technical data for manufacturing MHPs with sintered wicks. The law of sintering temperature changing during resistance furnace sintering is the very important technology information. If adopting the traditional trial and error methods, scores of, even hundreds long time of tests must be done in order to find the optimum technology parameters because technology parameters vary so sharply that the data are very difficult to be collected.

In order to overcome the shortcoming, support vector machines (SVMs) proposed by Vapnik\textsuperscript{[5]} has been receiving increased attention with remarkable results. With the introduction of \( \varepsilon \)-insensitive loss function, SVM has been extended to solve nonlinear regression estimation problems, such as new techniques known as support vector regression (SVR), which have been shown to exhibit excellent performance. Thus, SVR had been successfully employed to solve forecasting problems in many fields. Such as traffic flow time series forecasting, engineering and software field forecasting, atmospheric science forecasting, and so on. Meanwhile, SVR model also had been successfully applied to forecast electric load. The practical results indicated that poor forecasting accuracy is suffered from the lack of knowledge of the selection of the three parameters (\( \sigma, C, \) and \( \varepsilon \)) in a SVR model.

Due to lacking of structured ways in determining three free parameters in SVR, in this investigation, SVR model, whose control parameters were optimized by the Simulated Annealing algorithm (SA), was developed to predict the sintered furnace temperature law. Subsequently, SVR model was built using sintered furnace temperature historic data as the training samples.

The remainder of this paper is organized as follows. In Sections 2 and 3, we explain the methodology. Section 4 gives experimental results. Finally, Section 5 concludes the paper.

2. Regression arithmetic of support vector machine (SVR)

SVR is to map the input data \( x \) into a higher dimensional feature space through a nonlinear mapping \( F \) and then a linear regression problem is obtained and solved in this feature space. Here, given a set of data \( T = \{(x_1, y_1), \cdots, (x_l, y_l), \cdots, (x_l, y_l)\} \), where \( x_i \in \mathbb{R}^n \) is an arbitrary data point and \( y_i \in \mathbb{R} \) denotes the corresponding output value, the SVM regression function can be formulated as

\[
y = \omega \cdot \varphi(x) + b
\]  
(1)
where $\varphi(\cdot)$ denotes the non-linear mapping function, $\omega$ denotes the weight vector and $b$ denotes the bias term. $\varepsilon$-SVR is to find a function $y$ that has at most $\varepsilon$ deviation from the actually obtained targets $d_i$ for all the training data and at the same time is as flat as possible. We do not care about errors as long as they are less than $\varepsilon$, but will not accept any deviation larger than this. Flatness in this case means to reduce the model complexity, so that we can write this problem as a convex optimization problem:

$$\min R(y) = C \frac{1}{N} \sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2} \|\omega\|^2$$  \hspace{1cm} (2)

It is known that the regression estimation function is the one that minimizes (2) with the following $\varepsilon$-insensitive loss function, defined as:

$$L_\varepsilon(d_i, y_i) = \begin{cases} 
|d_i - y_i| - \varepsilon, & |d_i - y_i| > \varepsilon \\
0, & \text{otherwise} 
\end{cases}$$  \hspace{1cm} (3)

where both $C$ and $\varepsilon$ are user-determined parameters; $d_i$ denotes the actual value at period $i$; $y_i$ represents the forecasting value at period $i$. Additionally, the first term in (2) denotes the empirical error. The second term in (2) represents the function flatness. The $C$ is used as the trade-off between the empirical risk and the model flatness. Sometimes, this may not be the case, or we also may want to allow for some errors. Two positive slack variables $\xi_i$ and $\xi_i^*$ are introduced to represent the distance from actual values to the corresponding boundary values of the $\varepsilon$-tube. Then, Eq.(2) is transformed into the following constrained form:

$$\min S = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$  \hspace{1cm} (4)

$$\text{s.t. } \begin{cases} 
|d_i - \omega \cdot \varphi(x_i)| - \varepsilon + \xi_i, & \omega \cdot \varphi(x_i) + b - d_i \leq \varepsilon + \xi_i^*, \\
\xi_i, \xi_i^* \geq 0, \end{cases}$$

Introduction Lagrange multipliers $\alpha_i, \alpha_i^*, \eta_i$ and $\eta_i^*$, we can write the corresponding primal Lagrangian as:

$$L(\omega, \xi_i, \xi_i^*) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) - \sum_{i=1}^{n} \alpha_i (\varepsilon + \xi_i + d_i - \omega \cdot \varphi(x_i) - b)$$

$$- \sum_{i=1}^{n} \alpha_i^* (\varepsilon + \xi_i^* + \omega \cdot \varphi(x_i) + b - d_i) - \sum_{i=1}^{n} (\eta_i \xi_i + \eta_i^* \xi_i^*)$$  \hspace{1cm} (5)

Finally, by applying Karush-Kuhn-Tucker (KKT) conditions for regression, the optimization formulation can be transformed into a dual problem, get:

$$\max Q(\alpha_i, \alpha_i^*) = \sum_{i=1}^{n} y_i (\alpha_i^* - \alpha_i) - \varepsilon \sum_{i=1}^{n} (\alpha_i^* + \alpha_i) - \frac{1}{2} \sum_{i=1}^{n} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i, x_j)$$

$$\text{s.t. } \begin{cases} 
\sum_{i=1}^{n} (\alpha_i^* - \alpha_i) = 0, & 0 \leq \alpha_i, \alpha_i^* \leq C, i=1,2,\cdots,n 
\end{cases}$$  \hspace{1cm} (6)
where $\alpha_i, \alpha_i^*$ are the so-called Lagrangian multipliers. They satisfy the equalities $\alpha_i \cdot \alpha_i^* = 0$.

The Lagrange multipliers $\alpha_i$ and $\alpha_i^*$, are calculated and an optimal weight vector of the regression hyperplane is expressed as:

$$w^* = \sum_{i=1}^{nsv} (\alpha_i - \alpha_i^*) K(x_i, x)$$

(7)

Hence, the regression function is expressed as

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

(8)

Where the constant $b$ is written as

$$b = \frac{1}{2} \sum_{i=1}^{nsv} (\alpha_i - \alpha_i^*) (K(x_r, x_i) + K(x_s, x_i))$$

(9)

Notably, a number of coefficients ($\alpha_i - \alpha_i^*$) are non-zero values and the corresponding training data points have approximation error equal to or larger than zero. They are called support vectors. $x_r$ and $x_s$ in Eq.(9) are support vectors. $nsv$ denotes the number of the total support vectors.

The term $K(x_i, x_j)$ in Eq.(8) is defined as kernel function, where the value of kernel function equals the inner product of two vectors $x_i$ and $x_j$ in the feature space $\varphi(x_i)$ and $\varphi(x_j)$, meaning that $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$. Any function that satisfies Mercer's condition can be used as the kernel function. In SVM, radial basis function (RBF) $K(x_i, x_j) = \exp(-\|x_i-x_j\|^2/2\gamma^2)$, polynomial basis function $K(x_i, x_j) = ((x_i \cdot x_j) + b)^d$ ($b \geq 0$, $d$ is the natural number), and sigmoid function $K(x_i, x_j) = \tanh(k(x_i \cdot x_j) + v)$ are typical examples of kernel function. Only one variable needs to be determined in the radial basis function. In addition, SVM constructed by radial basis function has excellent nonlinear classification ability. Thus, in this work, RBF is used in the SVM.

The selection of the three positive parameters, $\gamma$, $\varepsilon$ and $C$ of a SVM model is important to the accuracy of the forecasting. However, structural methods for confirming the selection of parameters efficiently are lacking. Therefore, a simulated annealing algorithm (SA) is used in the proposed SVM model to optimize the parameter selection.

3. SA for Parameter Tuning of SVM

The original intention of Simulated Annealing (SA) is similarity between annealing procedure of solid material in physics and general combination optimization problem and its application field is extended from the original combination optimizing problem to continuous space optimizing problem. Under a given original temperature the simulated annealing algorithm could randomly search the globally optimal solution of the object function in the solution space with the gradually-descending annealing temperature. Even at the locally optimal solution its probability jumping property will make optimizing procedure escape from this point and finally go to the global optimal solution.

Based on the study of Boltzmann and Metropolis et al [6], Kirkpatrick claimed that the Metropolis approach is conducted for each temperature on the annealing schedule until thermal equilibrium is reached [7]. Additionally, a prerequisite for applying the SA algorithm is that a given set of the multiple variables defines a unique system for which the objective function can be calculated. The SA algorithm in our investigation is described as follows:
Step 1 (Initialization). Set upper bounds of the three SVM positive parameters, $\delta$, $\epsilon$ and $C$. Then, generate and feed the initial values of the three parameters into the SVM model. The absolute value of the forecasting error is defined as the system state ($E$). Here, the initial state ($E_0$) is obtained.

Step 2 (Provisional state). Make a random move to change the existing system state to a provisional state. Another set of the three positive parameters are generated in this stage.

Step 3 (Acceptance tests). The following equation is employed to determine the acceptance or rejection of the provisional state:

$$
\begin{cases}
\text{Accept the provisional state} & \text{if } E(S_{\text{new}}) > E(S_{\text{old}}), \text{and } p < P(E(S_{\text{new}}), 0 \leq p < 1) \\
\text{Accept the provisional state} & \text{if } E(S_{\text{new}}) \leq E(S_{\text{old}}) \\
\text{Reject the provisional state} & \text{otherwise}
\end{cases}
$$

(10)

In Eq.10, $E(s)$ denotes the energy of state $s$, $p$ is a random number to determine the acceptance of the provisional state. If the provisional state is accepted, then set the provisional state as the current state.

Step 4 (Incumbent solutions). If the provisional state is not accepted, return to Step 2. Furthermore, if the current state is not superior to the system state, then repeat step 2 and 3 until the current state is superior to the system state and, finally, set the current state as the new system state. Previous studies indicated that maximum number of loops ($N_{sa}$) is $100d$ to avoid infinitely repeated loops, where $d$ denotes the problem dimension. In this study, the three parameters ($\delta$, $\epsilon$ and $C$) are used to determine the system states, therefore, $N_{sa}$ is set to 300.

Step 5 (Temperature reduction). After the new system state is obtained, reduce the temperature. The new temperature reduction is obtained by the following equation

$$
\text{New temperature} = \text{(Current temperature)} \times \rho, \text{ where } 0 < \rho < 1
$$

(11)

Where $\rho$ is set at 0.9 in this work. If the pre-determined temperature is reached, then stop the algorithm, and the latest state is an approximate optimal solution. Otherwise, go to step 2.

The following mean square deviation index, shown as Eq.11, used as the criterion for identifying suitable parameters for use in the SVM-SA algorithm.

$$
I = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2
$$

(12)

Where $N$ is the number of samples in the training set, $\hat{y}_i$ and $y_i$ represents the training output value and the expectation output value of SVM, respectively.

The architecture of proposed SVM forecasting model can be illustrated in Fig. 1. The SA algorithm is used to seek a better combination of the three parameters in the SVMs.

**Fig.1 Architecture of SVM-SA forecasting model**

### 4. Wick Sintered temperature forecasting model based on SVR-SA

#### 4.1. Problem formulation
According to the changes of time series of the temperature, the current temperature is certainly linked with that of several hours ago. Thus the previous hydrogen-producing reactor temperature sequence data can be used to predict the future reactor temperature. Assuming $T_i(t)$ is the temperature value of sintered temperature in $t$ moment, $T_i(t-1)$ is the temperature value in $t-1$ moment. The temperature value of sintered temperature of current and previous $m$ time period can be used to forecast the reactor temperature value in the future time period. Let $T_i(t)$, $T_i(t-1)$, ..., $T_i(t-m)$ be the samples input vector in $t$ moment as $x_i$, $T_i(t+1)$ be the samples output value $y_i$.

Chosen sintered temperature value samples as initial training set: $S = \{s_i | s_i = (x_i, y_i), i=1,2,\ldots,n\}$. According to SVR algorithm, the initial prediction regression function is obtained. Then the prediction value for sintered furnace temperature $\hat{y}_i$ is:

$$\hat{y}_i = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

4.2. Data Pre-processing

The samples are divided into two data sets: the training data sets and the testing data sets. Subject to random factors (e.g. transmission errors, etc.), it can not be avoided to lose data accuracy such as data errors and data loss, so the primary data preprocessing need to be implemented to correction errors. The training data is smoothly processed to eliminate singular value, and the experimental data (including training data and testing data) is normalized, which can improve generalization capability of SVR. If the data is unequal interval series, unequal interval series are transformed into equal interval series to train SVR.

In order to avoid the influence of difference between factors, the parameters of input and output are normalized as Eq.(19), the learning data are scaled into $[0, 1]$.

$$\tilde{y}_i = \frac{2(y_i - y_{\min})}{y_{\max} - y_{\min}} - 1$$

4.3. Sintered Temperature using SVR-SA Model

While using SVR-SA model for sintered furnace temperature law forecasting, in the training stage, firstly the hyper-parameters $C$, $\varepsilon$ and $\sigma$ of SVM model are optimized by SA, Then, the optimal parameters are utilized to train SVM model. The testing data sets are used to examine the accuracy of the forecasting model.

Cross-validation is a popular technique for estimating generalization performance. In order to better evaluate the performance of the proposed approach, we also use the $k$-folds cross validation to chosen the appropriate parameters ($C$, $\varepsilon$ and $\sigma$). In this example, we use $k = 5$ for the number of folds. The optimal parameters of SVR searched by SA are $C = 8.048$, $\varepsilon = 0.9813$, $\sigma = 0.7541$. Using the parameters determined by SA and several groups of parameters selected randomly to establish reactor temperature prediction models based on SVR-SA, the tuning results are given in Table 1. It can be seen from Table 1, the model established by the optimal parameters has the best prediction effect. Therefore, the practical prediction model of sintered furnace temperature based on SVR-SA is established.
### Tab. 1 SVR-SA optimal forecasting parameters

| No. | Hyper-parameters | NMSE of testing |
|-----|------------------|-----------------|
|     | $\sigma$ | $C$ | $\varepsilon$ |                     |
| 1   | 0.6415  | 6.125 | 0.5798 | 0.2561 |
| 2   | 0.4126  | 3.196 | 0.4521 | 0.3721 |
| 3   | 0.5642  | 7.103 | 0.4213 | 0.3345 |
| 4   | 0.7541  | 8.048 | 0.9813 | 0.2416 |
| 5   | 0.6958  | 9.242 | 0.9214 | 0.2945 |

From Fig. 2, it can be observed that some parts of the fitness curve of SVM based on SA obviously vibrate but the fitness curve of SVM based on PSO does not, which shows that the SVM based on SA accepts worse solutions sometimes and indicates its better ability to escape from a local optimum. Moreover, SVM based on SA got the better results than SVM based on PSO in the end, owing to the increased diversity in the particles by SA algorithm.

![Fig. 2. Prediction comparison using SVM-SA and SVM-PSO](image)

### 5. Conclusions

This study applied SVR to the forecasting fields of sintered furnace temperature time series. To build stable and reliable forecasting models, the parameters of SVR must be specified carefully. Generally, SA can be used to select suitable parameters to forecast sintered furnace temperature, which avoids over-fitting or under-fitting of the SVM model occurring because of the improper determining of these parameters. This study proposed a SA algorithm for dynamically optimizing the three parameters of the support vector regression. Experimental results demonstrate the feasibility of successfully applying this novel hybrid SVR-SV model to the complex forecasting problem. The proposed model has a better ability to escape from the local optimum and a better predicting ability than SVM-PSO. Additionally, there are a large number of research directions that can be considered as useful extensions of this research.
Acknowledgments

This work is partly supported by the National Natural Science Foundation of China (50905119); the Natural Science Foundation of Guangdong Province (941806001001985); China Postdoctoral Science Foundation special fund (201104354); Shenzhen Science & Technology Program (SYB200907090130A / JSA200903190981A) and the SZU R/D Fund (2010004);

References

[1] VASILIEV L L. Micro and miniature heat pipes: Electronic component coolers. Applied Thermal Engineering, 2006, 28(4): 266–273.
[2] TAN B K, WONG T N, OOI K T. Analytical effective length study of a flat plate heat pipe using point source approach [J]. Applied Thermal Engineering, 2005, 25(14/15): 2272–2284.
[3] TANG Yong, CHI Yong, CHEN Jing-chuang. Experimental study of oil-filled high-speed spin forming micro-groove fin-inside tubes [J]. International Journal of Machine Tools and Manufacture, 2007, 47(7/8): 1059–1068.
[4] ZHANG Guang-liang, ZHANG Shi-hong, LI Bing, ZHANG Hai-qu. Analysis on folding defects of inner grooved copper tubes during ball spin forming. Journal of Materials Processing Technology, 2007, 184(1/3): 393–400.
[5] V. Vapnik, Statistical Learning Theory. New York: Wiley, 1998.
[6] Cercignani C. The Boltzmann equation and its applications. Berlin: Springer-Verlag; 1988.
[7] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by Simulated Annealing Science, Vol.220, (1983): 671–680.