Modeling and Optimization of Ultrasonic Cleaning Process for Hard Disk Drive Arm Using Support Vector Regression

Presented in 1st Data Storage Technology Conference (DST-CON 2008)

*Pongsak Holimchayachotikul and Wimalin Laosiritaworn**

Abstract

The ultrasonic cleaning has been widely adopted in hard disk drive manufacturing industry due to its effectiveness in removing contamination from the final products, sub-assemblies and components. This process is suitable for delicate product with high complexity. The process efficiency depends on six factors including ultrasonic frequency, type of liquid medium, time, temperature, power, and finally a number parts in cleaning basket. These factors, if not carefully set, might result in failure in contamination removal indicated by high level of liquid particle count (LPC), damages of hard disk drive and ultimately shorten product's durability. The ultrasonic cleaning process setting is usually determined by experience of operators, which might not always result in optimum condition. Therefore, this study presents an application of the integration between Taguchi Method and Computational Intelligence (CI) techniques, increasingly applied for modeling and optimizing the performance of manufacturing industry, to identify the optimum setting of the cleaning process parameters focusing on quality improvement of the finished hard disk drive arm that can deliver clean surface of the product with no damage. The proposed method is as follows. Firstly, both Support Vector Regression (SVR) and Artificial Neural Network (ANN) were trained with experimental data to model LPC level of the cleaning process. The model with highest accuracy was selected to be a Suitable Computational Intelligence Model (SCIM). Then, a Grid search was opted to the SCIM to find the optimum settings. Data from real experiments of nigh hawk 1H hard disk drive arm were used to demonstrate the proposed method. The experimental results suggested that SVR technique is capable of high accuracy modeling and results in much smaller error and learning time in comparison with ANN.

Department of Industrial Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200 THAILAND

**corresponding author; e-mail: winmalin@hotmail.com*

Introduction

An ultrasonic cleaning plays a crucial role in the hard disk drive assembly process as cleanness is very important in hard disk drive production. Ultrasonic cleaning utilizes ultrasonic energy to extract particles from a part and puts them into aqueous suspension. Then, a liquid particle counter (LPC) is applied to measure the particle concentration in the extraction solution. Among the procedures of hard disk drive arm production line are contaminated, from soft and hard particle, the arm (Gouk , 1997). Due to the different materials and designs of the parts; it is technically difficult to achieve acceptable LPC level set by the customers. Optimizing a manufacturing process for the ultrasonic cleaning requires a thorough study of involving parameters relating to cleanness quality. The usage of design of experiments in this area was proven to be very beneficial. For example Gouk (Gouk, 1997) used classical design of experiment approach, namely full factorial design to study the optimum condition for ultrasonic cleaning.

Since there has been rapid growth in the hard disk drive industry driven by the advances technologies. Hard disk drive has to increase in its capacity yet the size needs to be smaller. This abrupt change has made the hard disk drive arm becomes smaller and much more complex, the traditional statistical analytical approaches, such as Response Surface Methodology (RSM) and Taguchi method, are not always suitable with this change according to the non-linear nature of the problem. Recently, CI techniques, namely support vector machine (SVM) and artificial neural network (ANN), have considered as novel alternatives on the ground of its capability to deal with imprecision, uncertainty and fault tolerance (Smola and Schölkopf, 2004; Nascimento et al., 2000). Moreover, integrations of CI with conventional approaches have been increasingly utilize to model and optimize process in various fields such as Superplastic forming process in aerospace industry (Sukthomya and Tannock , 2005), flux cored arc welding process (Holimchayachotikul et al.,2007) and insulin dose forecasting for type II diabetes patients (Holimchayachotikul et al.,2007). ANN approaches have suffered from problems such as local minimum and over fitting (Kuman , 2005). While ANN learning with back-propagation algorithm (BP) was designed by empirical risk minimization, SVM was developed by structural risk minimization. SVM, has been reported in the literature to be superior to ANN approaches in certain applications (Vapnik, 1999). SVM can be used for either classification or regression task.

This paper is organized as follows. In section 2, the fundamental concepts of SVR algorithm was briefly described. In section 3, the methodology of the proposed approach is discussed. Results and discussion of the proposed framework are mentioned and analyzed in section 4. Finally, conclusions are provided in section 5.

Support Vector Regression

The purpose of SVR is to find a function $f(x)$ that has at most ε deviation from the actual obtained target y_i for all training set, $\{x_k, y_k\}$, $x\in \Re^n$, $y\in \Re^n$ with k observations. The commencement of SVR is described by the linear function using the form $\langle w \bullet x \rangle + b$, then the nonlinear problem is transferred into a linear problem by a nonlinear map $\Phi(x)$from the low dimensional input space to a higher-dimensional feature space. At the same time, $f(x)$ is as flat as possible. SVR estimates function, taking the following form:

$$f(x)=\sum_{i=1}^{k}(\alpha_i-\alpha_i^*)\langle x_i \bullet x\rangle+b \qquad (1)$$

Let the nonlinear transformation function be $\Phi(\bullet)$, b is the "bias" term and the kernel functions are defined as

$$K(x_i x)=\langle \Phi(x_i)\bullet\Phi(x)\rangle \qquad (2)$$

Equation (1) implies that the dot product in high dimensional space is equivalent to a kernel function of the input space.

Methodology

A schematic diagram of the proposed procedure is shown in Figure 1. This comprises of the combination of ANN, SVR, and Taguchi method applied to find the optimum setting of the process parameters of ultrasonic cleaning for Night Hawk 1H hard disk drive arm. The study started from obtaining data from real experiments based on Taguchi design. Before starting CI learning process, Pareto ANOVA was conducted to identify significant factors. At the same time, Taguchi analysis was applied to find optimum condition from another study (Amornsisudja and Chattinnawat, 2008). Data files, consisting of input parameters (frequency, power, temperature, and time) and the corresponding outputs (LPC of particle, sized 0.3 and 0.6 μ/cm^3), are then used to train both SVR and ANN concurrently.

After training prediction accuracy of testing data, in terms of mean absolute percentage error (MAPE), was used to compare models performance. The model with higher accuracy was identified as a 'SCIM'. Finally, the Grid search method was adopted to the SCIM model (in this case was SVR) to find optimum process parameter settings.

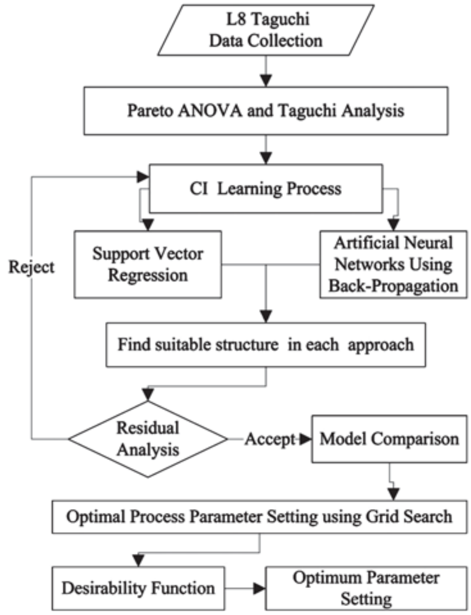

Figure 1. Schematic diagram of the proposed framework.

Results and Discussion

The experimental phase

The Taguchi's $L_8(2^7)$ was used with the four controllable factors of ultrasonic cleaning for Night Hawk 1H hard disk drive arm process, each at low and high level, as shown in Table 1. Experiments were carried out with three replications, as a result, 24 experimental trials, shown in Table 2, was obtained. Pareto ANOVA was used to identify significant factors. In this study the responses are the-smaller-the-better S/N ratio (Taguchi et al., 2005) and the average form of LPC values. From linear graph of, Factor T,F,P ,and Time were assign to columns 1,2,4 and 7 respectively. This permitted two-level interactions (TxF, FxP and TxP) to be assigned to columns 3, 5 and 6 respectively. According to the Pareto ANOVA chart in Figure 2, the important factors, FxP, F, T ,and Time and

T, FxP, Time ,and F for LPC sized 0.3 and 0.6 respectively, are selected from the left-hand side, which contribution of approximately 90% of the entire effect. But according to the hierarchy principle (Montgomery, 2005), the lower-order terms, such as P, should be included in the model since it is a composition of significant higher order FxP. In practice, the interaction terms (FxP TxP FxT) cannot be actually set; therefore, the CI models were developed based on four factors.

Table 1. Taguchi's $L_8(2^7)$ experimental levels

Factor	Factor description	level Low (-1)	Medium (0)	High (1)
T	Temperature (°C)	35	40	45
F	Frequency (kHz)	40	-	80
P	Power (% of 2500 W)	70	80	90
Time	Time (sec)	240	360	480

Table 2. Taguchi's $L_8(2^7)$ results.

Factor				LPC particle size 0.3 μ/cm^3		0.6 μ/cm^3	
T	F	P	Time	mean	S/N	mean	S/N
-1	-1	-1	-1	66353.58	-96.7147	5532.54	-75.1979
-1	-1	1	1	66783.75	-96.8560	6043.96	-76.0546
-1	1	-1	1	56696.67	-95.4739	4866.71	-74.2401
-1	1	1	-1	52466.75	-94.8319	4914.63	-74.4748
1	-1	-1	1	63543.38	-96.8341	6627.67	-77.4122
1	-1	1	-1	76683.63	-99.0530	7779.79	-79.5181
1	1	-1	-1	72184.21	-98.2043	7592.21	-79.1402
1	1	1	1	51110.79	-94.8356	5999.75	-76.5551

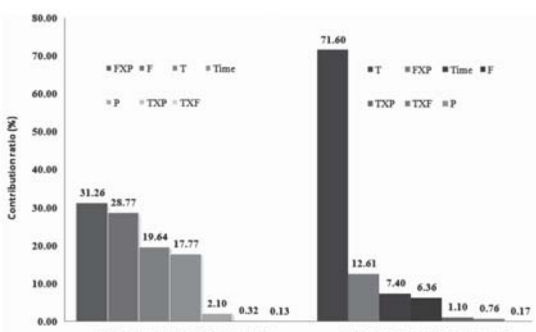

Figure 2. Pareto ANOVA charts for each response.

8 experiment trials were conducted at 3 replicates. From each replicate, 8 hard disk drive arms were collected from the washing basket, 4 from the upper position and 4 from lower positions in the basket. As a result, a total of 192 (8x3x8) experimental trials were available for CI learning process.

The learning process was carried out based on 5-fold cross validation with 50,000 iterations for each candidate model. Optimum architecture of SVR and ANN were obtained from Grid search, illustrated in sub-process of Figure 1. However, both models failed to accurately predict the process responses. Hence, output transformation by using the-smaller-the-better S/N ratio formula was used. A total of 24 S/N ratio values were calculated from the upper, lower and total positions from each condition. Training process was then carried out again with the transformed outputs.

Artificial Neural Networks Training and Testing

In this study, all computational experiments are performed on an Intel Core(TM) 2 Duo, 1.80 GHz CPU and 1.5 GB of memory. A total of 24 data were available for training. The first 80% of data (19 samples) were used for training, while the rest 20% (5 samples) were used in the testing process. Accuracy of the network was measured by mean square error. Exhaustive search was used to identify network parameters to achieve the best setting with minimum error. The search was conducted with 3,000 iterations and two retrain in each combination. Testing error fitness criterion was used. The smaller the error on the test set the better the accuracy of the network. Space search of two hidden layers started from 1 to 50 hidden nodes in each layer with search step one by one. The best topology obtained from the search was 4-8-11-1 network and 4-38-23-1 network for LPC sized 0.3 and 0.6 μ/cm^3 , respectively. The BP algorithm was used in training process. The networks were trained with sigmoid transfer function for hidden and output layers. The momentum factor of 1.75 and initial value of learning rate of 0.05 were chosen for fast convergence. Thus, the number of training iterations was 50,000 epochs and the initial values of weights and biases are random. Alyuda NeuroIntelligence 2.2 was the software for ANN learning process.

Support Vector Regression Training and Testing

The same data set used to train and test neural network were, again, used with SVR. This data set comprises of the inputs vector and the corresponding output vector. Many types of kernel functions can be applied, for instance linear, polynomial, radial basis function, and sigmoid function. The training processes are as follows; Firstly, input data were mapped from the input space into a high dimensional feature space using radial basic function kernel (RBF)(Anguita et al., 2000). The amplitude of RBF was controlled by the vital parameter γ. Exhaustive search was used to identify model parameters to achieve the best setting with minimum error. The search was conducted with 10,000 iterations to find the setting with lowest MSE. Space search of γ, ε and C started from 0.01 to 0.5 , 0.0001 to 0.01 and 1000 to 1,000,000 respectively. In this research, we found that γ of 0.25, ε =0.0001 and C=500,000 from the grid search results, provided the best predictive results for LPC sized 0.3 and 0.6 μ/cm^3. STATISTICA 7 was applied to run SVR training and testing process.

Model Perfomance Comparison

The results from SVR and ANN are shown in Figure 3. The performance of each model was compared by using MAPE of the overall data set, training data set and testing data set. According to the testing data set error, it was found that SVR has lower error than ANN. It can be clearly seen that SVR provides a better result and is more suitable in this case. As a result, it was used for further optimization. Figure 4 shows the comparison of the desired output of the S/N ratio compared with the results obtained from the three approaches. The testing data were cases 4, 15, 18, 20 and 22.

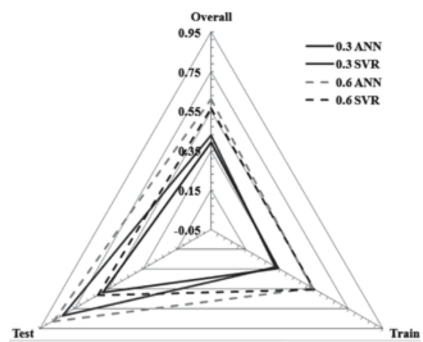

Figure 3. MAPE comparison of all approaches.

Process Optimization

In this process, SVR, the most accurate model identified in section 4.4, was used to find the optimized parameter settings using Grid search. The Grid search step size of each model for temperature, frequency, power, and time parameters were 0.05, 4, 0.05, and 1.5, respectively. The optimum cleaning conditions from two approaches and verification experiments are shown in Table 3. The finest condition for two levels of LPC responses was 37.75 ^{o}C and the maximum level of other factors, obtained from the desirability function of the-smaller-the better (Montgomery , 2005).

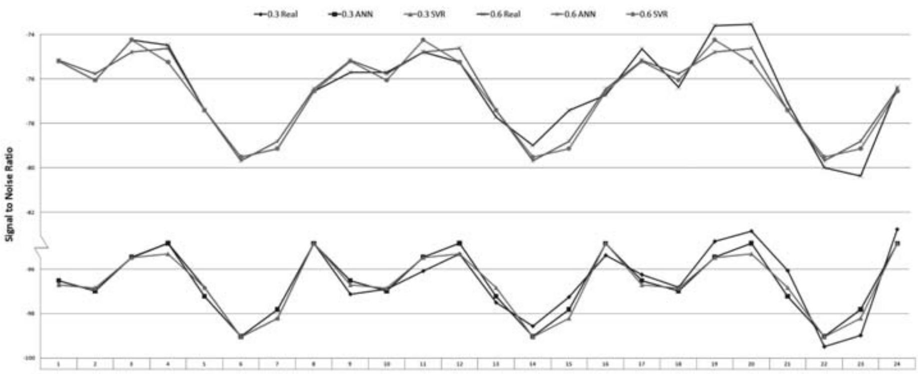

Figure 4. The comparison of desired output and the output from each model.

Table 3. The optimum conditions from SVR Models.

For LPC	T	F	P	Time	S/N	Particles	D
0.3	37.75	80	90	480	-94.6085	53755.76	0.9257
0.6	35	80	76	480	-74.1901	5122.77	0.8502

Conclusion

The cleaning process using ultrasonic is a crucial process in a hard disk drive arm production. Most engineers are familiar with the Taguchi method to optimize the process. Nevertheless, this approach has some concerning points. Firstly, it can only get the optimum solution within the specified level of control factors. Secondly, it is unable to find the optimum values when the specified parameters are continuous in nature on the ground of its address to discrete controllable factors. Finally, it can not effectively cope with interactions among parameters. This research proposed an integrated approach of a SVR trained with Taguchi data to solve the mentioned problems. This novel approach utilizes Taguchi method as a means of efficient data collection tool, as feature selection tool, through a small number of experiments. Taguchi's experimental relationships were then captured by SVR technique. Then a Grid search is applied to obtain the optimum parameter settings of the process. While Taguchi method can provide only one proper condition, at this point SVR model based on Grid search is superior to Taguchi method in terms of real practice, A practical example of SVR competitiveness, in terms of cost and time, is: Given a rival company achieve the same cleanness level of the arm in 360 seconds, SVR can be modeled to achieve optimum results similar to or even better than that of rival company's in that 360 seconds, which Taguchi method fails to comply.

SVR technique has potential advantages, as it has an easy-and-quick capability to explore a nonlinear multivariate relationship between parameter condition and response. Moreover, it utilizes small number of experimental data to construct the model, results in minimal process disruption. However, there are certain limitations of SVR. For instance, the quality of solution of SVR depends significantly on the variation and pattern of the learning data set and SVR can provide only one output vector. Despite of these limitations, this study points out that it can provide a robust and an accurate model in the complex processes, faced in the current advances of technology. Further potential research may be embraced by combining output vector from each SVR model using fuzzy desirability function to achieve optimum condition or apply other local search methods such as GA, ant, tabu search algorithm. Moreover, we can use the optimum condition from this proposed study to suggest feasible level determinations of factors in final process modeling and optimization.

Acknowledgement

The authors would like to acknowledge National Electronics and Computer Technology Center of Thailand (NECTEC) for financial support and Lanna Thai Electronic Components, Ltd (LTEC) for allowing experiments to be conducted on Nighthawk 1H product.

References

Amornsisudja, C., Chattinnawat, W, 2008. "Optimization of Parameters in Ultrasonic Cleaning Process for Hard Disk Drive Arm Using TaguchiDesign of Experiment," The 1st International Data Storage Technology Conference (DST-CON2008).

Anguita, D., Boni, A., and Pace, S. 2000. "Fast training of Support Vector Machines for regression," *Proc. of the IEEE-INNS- ENNS Infl. Joint Confon Neural Nehuorh*, vo1.5, pp. 210-214.

Montgomery DC., 2005.*Design and Analysis of Experiments*, 5th Ed, John Wiley& Sons, New York.

Nascimento, C.A.O., Giudici, R., Guardani, R., 2000. "Neural network based approach for optimization of industrial chemical processes," *Computers and Chemical Engineering*, 24 (9-10), pp. 2303-2314

Gouk, R., 1997. "Optimizing ultrasonic cleaning for disk drive components", *Precision Cleaning*, 5 (8), pp. 13-17,

Holimchayachotikul, P., Jintawiwat, R., Leksakul, K. 2007. "Support vector regression based insulin dose forecasting for type II diabetes patients" The 5th International Conference on Quality and Reliabilty pp.192-197.

Smola A.J. and Schoölkopf B., 2004. "A tutorial on support vector regression", *Statistics and Computing*, **14 (3)**, pp. 199-222.

Kuman, S., 2005.*Neural networks a classroom approach*, International Ed, Mc Graw Hill.

Sukthomya, W., Tannock, J., 2005. "The optimisation of neural network parameters using Taguchi's design of experiments approach: an application in manufacturing process modeling" *Neural Computing and Applications*, 14 (4), pp. 337-344.

Taguchi, G., Chowdhury, S. and Wu, **Y.**, 2005. ***Taguchi's quality engineering handbook***, Hoboken, N.J. : John Wiley & Sons.

Vapnik, V.N., 1999. "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, 10 (5), pp. 988-999.