International Journal of Innovative Computing, Information and Control
Volume **8**, Number **10(B)**, October **2012**

ICIC International ©2012 ISSN 1349-4198
pp. **7321–7328**

RECIPE GENERATION FROM SMALL SAMPLES: INCORPORATING WEIGHTED KERNEL REGRESSION WITH ARTIFICIAL SAMPLES

Mohd Ibrahim Shapiai[1], Zuwairie Ibrahim[2], Marzuki Khalid[1]
Lee Wen Jau[3], Soon-Chuan Ong[3] and Junzo Watada[4]

[1]Centre of Artificial Intelligent and Robotics
Universiti Teknologi Malaysia
Jalan Semarak, 54100, Kuala Lumpur, Malaysia
ibrahim@fke.utm.my; marzuki@utm.my

[2]Faculty of Electrical and Electronic Engineering
Universiti Malaysia Pahang, Kampus Pekan
26600 Pekan, Pahang, Malaysia
zuwairie@fke.utm.my

[3]Department ATTD Automation (APAC) Pathfinding
Intel Technology Sdn. Bhd.
Kulim, Malaysia
{ wen.jau.lee; soon.chuan.ong }@intel.com

[4]Department Graduate School of Information and Systems
Waseda University
2-7 Hibikino, Wakamatsu, Kita-Kyushu 808-0135, Japan
junzow@osb.att.ne.jp

Received March 2011; revised July 2011

Abstract. *The cost of the experimental setup during the assembly process development of a chipset, particularly the under-fill process, can often result in insufficient data samples. In INTEL Malaysia, for example, the historical chipset data from an under-fill process consist of only a few samples. As a result, existing machine learning algorithms for predictive modeling cannot be applied in this setting. Despite this challenge, the use of data-driven decisions remains critical for further optimization of this engineering process. In this study, a weighted kernel regression with artificial samples (WKRAS) is introduced to improve the predictive modeling in a setting with limited data samples. In the proposed framework, the original weighted kernel regression (WKR) is strengthened by incorporating artificial samples to fill the information gaps between available training samples. The artificial samples generation is based on the dependency measurement between every independent variable and dependent variable with subject to the calculated correlation coefficients. Even though only four samples are used during the training stage of the setup experiment, the proposed technique is able to provide an accurate prediction within the engineer's requirements as compared with other existing predictive modeling systems, including the WKR and the artificial neural networks with back-propagation algorithm (ANNBP).*

Keywords: Recipe generation, Predictive modeling, Weighted kernel regression, Small samples, Artificial samples

1. **Introduction.** Recipe generation provides the key references needed by engineers to set up a new experiment for a new product and plays an important role in determining the success of product development. Currently, the ingredients chosen for the recipe mainly depend on the engineer's knowledge. Optimizing the input parameters will facilitate the engineering decisions needed to fulfill certain requirements. As the assembly process for

chipsets is rapidly progressing towards smaller scales and greater complexity, the accuracy and efficiency requirements are more vital. For example, a semiconductor process flow requires hundreds of fabrication operations steps with a lead-time of a few months. In addition, device fabrication and manufacturing costs continue to escalate. In addition to the usual strategy of increasing the wafer size and shrinking devices to reduce the cost per transistor, automation and modeling are becoming more important. Fowler [1] revealed that the productivity improvement strategy of semiconductor manufacturing is based on operational improvement at the front-end of wafer fabrication; this strategy accounts for almost half of the total annual productivity improvement target.

The use of artificial intelligence techniques for process modeling during the downstream assembly and all the involved tests is expected to reduce the overall manufacturing cost. As artificial intelligence techniques have been successfully applied in various engineering applications [2], introducing intelligent modeling to the assembly process promises to accelerate the engineering decisions even at early stages when very few collected samples are available. Inherently, intelligent modeling can improve equipment and resource utilization. In general, the development of recipe generation for assembly processes has only limited samples. However, most of the current machines learning algorithms are hindered by the limited number of available samples. In other words, the performance of existing algorithms degrades because the sample size is insufficient [3].

In INTEL Malaysia, the under-fill process shown in Figure 1, which consists of six input parameters with a small and sparse data set, is considered. Those input parameters are die size (dimension of die), gap height, the number of bumps, dispense distance, dispense weight, and the output is the dispense tongue length. In practice, it is difficult to define the input-output relationship, and improperly determined input setting parameters frequently cause the yield to be 'excess epoxy', 'epoxy on die', or 'insufficient epoxy'. Notably, the experiment usually involves large samples, and it is rather expensive to determine the recipe that prevents the tongue generated during the under-fill process from touching the keep out zone (KOZ), as illustrated in Figure 2. Hence, it is important to develop a cost-effective method to arrive at the optimal setting.

The problem being solved can be categorized as of learning from small samples which has gained increasing attention in many fields, such as in assembly process for sparse prediction modeling [4,5], engine control modeling [6], medical problem [7], and pulp and paper industry [8]. In general, most of the existing techniques rely on the pre-data processing technique, utilizing bias data points, and artificial samples generation in solving the problem. Generating artificial samples is also known as a technique to incorporate prior knowledge in machine learning [9] which gives better generalization. However, only few works have been reported in literature particularly for regression problem [10,11].

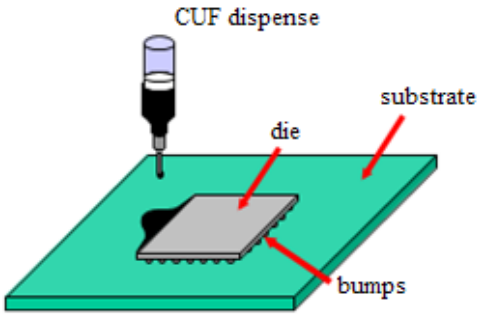

FIGURE 1. Illustration of an under-fill process in an assembly

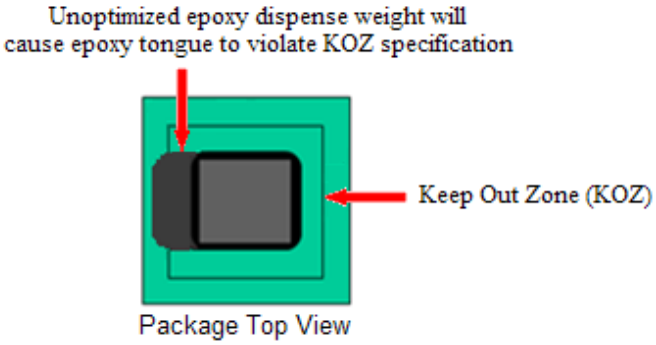

FIGURE 2. Illustration of an epoxy tongue that touches the keep out zone

Therefore, the main objective of this study is to propose a new technique for generating artificial samples by using the Weighted Kernel Regression (WKR) in solving the recipe generation problem. With this idea, correlation factor offers a feasible solution in generating artificial sample which based on the dependency coefficients between each independent variable and dependent variable. Also, the study aims to incorporate the generated artificial samples together with the available training samples to the WKR. By assuming that the generated artificial samples are relevant, with enough samples, the WKRAS will improve the prediction performance of the WKR. Previously, WKR has proved to solve small sample with good accuracy for theoretical functions [12] and application in semiconductor problem [5].

The remainder of this paper is organized as follows. A brief review of the WKR is given in Section 2. The proposed WKRAS is presented in Section 3. Section 4 includes the implementation of the proposed technique and the experimental results. Finally, the conclusions are provided in Section 5.

2. **Weighted Kernel Regression Review.** In this section, we first review the basic algorithm of the WKR. The concept of the WKR is introduced in the following. Given training samples, $\{x_i, y_i\}_{i=1}^{n}$, where n is the number of training samples, $x_i \in \Re^d$ is the input and $y_i \in \Re$ is the target output. WKR is the technique to regress the output space by mapping the input space $\Re^d$ to $\Re$. In general WKR is a modified Nadaraya-Watson kernel regression (NWKR) by expressing the weight based on the observed samples through a kernel function. The existing WKR relies on the Gaussian kernel function as given in Equation (1).

$$K(X, X_i) = \frac{1}{\sqrt{2\pi}} \exp \frac{\left(-\|X - X_i\|^2\right)}{h} \tag{1}$$

where h is the smoothing parameter. As in NWKR, the selection of smoothing parameter, h, is important to compromise between smoothness and fitness [13]. As in existing WKR, Equation (2) is employed to determine the value of h.

$$h = \sum_{i=1}^{n} \left(\|X_i\|^2 - \overline{\|X\|}^2\right)^2 \tag{2}$$

The kernel matrix $K = [K_{ij}]$, where $i = j = 1, \ldots, n$, with a generalised kernel matrix based on the Gaussian kernel, is given in Equation (3). The matrix K transforms the linear observed samples to non-linear problems by mapping the data into a higher dimensional

feature space.

$$K_{ij} = \begin{cases} \dfrac{\prod_{p=1}^{d} K\left(X_i^p, X_j^p\right)}{\sum_{l=1}^{n}\left[\prod_{p=1}^{d} K\left(X_{i\vee j}^p, X_j^p\right)\right]} & i \neq j \\ \dfrac{1}{\sum_{l=1}^{n}\left[\prod_{p=1}^{d} K\left(X_{i\vee j}^p, X_j^p\right)\right]} & i = j \end{cases} \tag{3}$$

In WKR, the most popular function for regression problems is used which to minimize the sum of squared error (SSE) to estimate the weight parameters, W.

$$\min f(W) \Leftrightarrow \min \|Kw - y\|^2 \tag{4}$$

Once the optimum weight is estimated, the model is ready to predict any unseen samples (test samples). The test samples can be predicted by using Equation (5)

$$\hat{y}\left(X, \hat{W}\right) = \frac{\sum_{i=1}^{n} \hat{w}_i \left(\prod_{p=1}^{d} K\left(X^p, X_i^p\right)\right)}{\sum_{i=1}^{n}\left(\prod_{p=1}^{d} K\left(X^p, X_i^p\right)\right)} \tag{5}$$

3. **The Proposed Weighted Kernel Regression with Artificial Samples.** The proposed technique relies on WKR to predict the unseen data. However, instead of relying on the limited samples, a considerable quantity of artificial samples is generated based on the correlation factor. The entire working procedure of the proposed technique is shown in Figure 3.

Step 1. Calculating the correlation coefficient for each independent variable against the output based on Equation (6).

$$r_{x_p y} = \frac{\operatorname{cov}(x_p, y)}{\sigma_{x_p}\sigma_y} = \frac{E\left[\left(x_p - \mu_{x_p}\right)(y - \mu_y)\right]}{\sigma_{x_p}\sigma_y} \tag{6}$$

where $r_{x_p y}$ is the correlation coefficient between the independent variable p, $p = 1, 2, \ldots, d$, and the dependent variable, y. x_p is n observed independent variables of p and μ_{x_p} is the corresponding mean value of x_p. y is n observed dependent variables and μ_y is the corresponding mean value of y. σ_{x_p} and σ_y are the standard deviations of x_p and y respectively. A highly correlated independent variable, $x_c = \{x_c^1, x_c^2, \ldots, x_c^n\}$, is defined as an independent variable with the highest correlation coefficient, which is subjected to a two-dimensional WKR in the next step.

Step 2. The highly correlated independent variable, $x_c = \{x_c^1, x_c^2, \ldots, x_c^n\}$, and the corresponding outputs $(Y_1, Y_2, \ldots, Y_n)$ are used as training samples for a two-dimensional WKR and the smoothing parameter, h, is set to be the variance of the original outputs $\{Y_1, Y_2, \ldots, Y_n\}$. With n samples of x_c, $n-1$ regression intervals, I_r, are defined where $r = 1, 2, \ldots, (n-1)$. The range of the entire interval is defined from the minimum value of x_c, $\min(x_c)$, and the maximum value of x_c, $\max(x_c)$. Within the entire interval, m new independent variables are generated as the test sample, X_s, where $s = 1, 2, \ldots, m$, based on Equation (7).

$$X_s = \min(x_c) + \left(\frac{s \times (\max(x_c) - \min(x_c))}{m}\right) \tag{7}$$

X_s is then used as the input for regressing m corresponding outputs, O_s, by applying Equation (5) based on the estimated weight. Because $S_s = (X_s, O_s)$, and $T_r \in S_s$, where

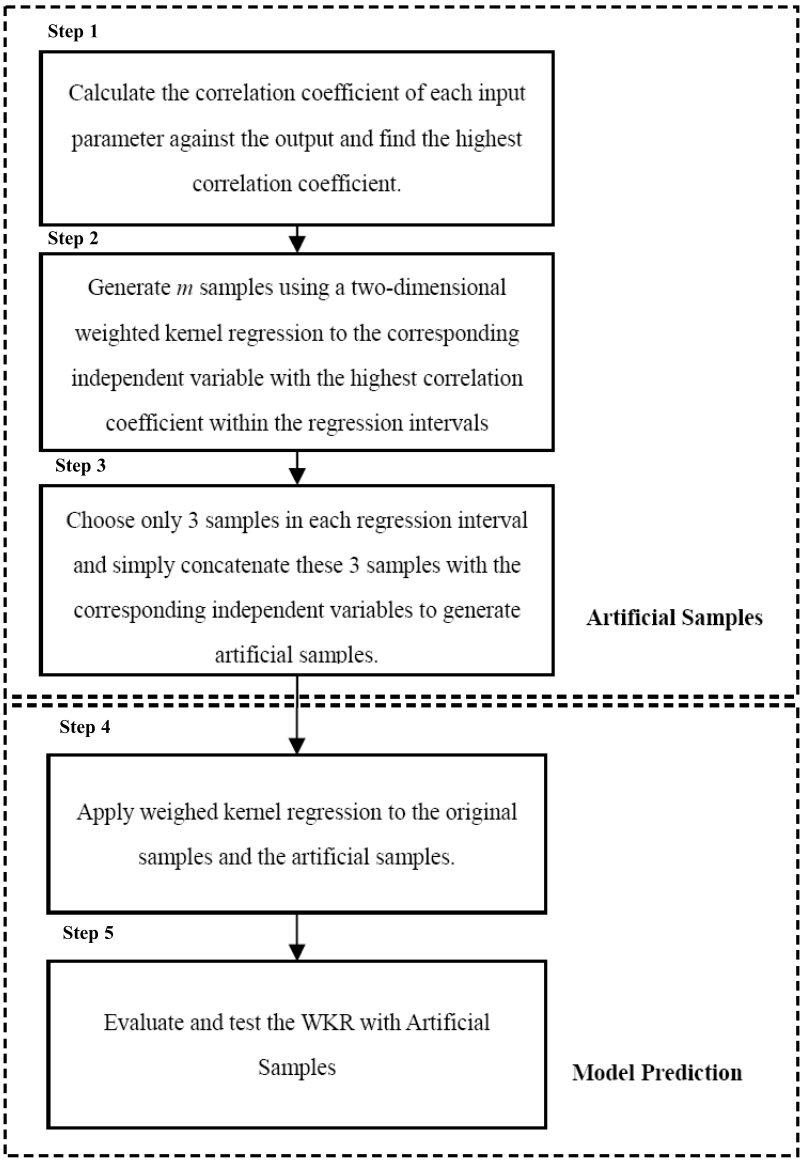

FIGURE 3. A series of steps for the WKR with artificial samples

T_r are generated samples in each regression interval and the chosen m value must satisfy T_r, there must be at least 3 samples ($T_r \geq 3$) in each regression interval condition.

Step 3. For each regression interval, only three samples, which are the minimum, median, and maximum values of T_r, are chosen; they correspond to $\{T_r^{\min}, T_r^{med}, T_r^{\max}\}$, respectively. Notably, in Step 1, among the d independent variables, only the highest correlated independent variable is chosen for artificial sample generation based on a two-dimensional WKR. The other $d-1$ independent variables that correspond to the particular $x_c = \{x_c^1, x_c^2, \ldots, x_c^n\}$ (only the first x_c^{n-1}'s are applicable due to the $n-1$ interval) are then simply concatenated to each of $T_r^{\min}$, T_r^{med} and $T_r^{\max}$ to generate artificial samples. This step is implemented in each regression interval. The total number of artificial samples is given by $\{AS_1, AS_2, \ldots, AS_k\}$, where $k = 3 \times (n-1)$.

Step 4. WKR is applied to the artificial samples obtained in Step 3 and the original samples. Again, h has to be calculated using Equation (2) for model prediction

Step 5. Evaluate and test the WKRAS using the test samples (unseen samples).

4. Experiments and Results.

4.1. **Experimental setup.** The historical DOE data set obtained from INTEL Malaysia, shown in Table 1, was employed in the experiment. The total number of available samples was ten, and four training samples from the first four rows were chosen because those training samples covered the minimum and the maximum range of the input and output values. This was a relevant assumption because the problem became an interpolation problem based on the observed samples.

Several existing techniques are considered in the conducted experiment including WKR [5] and ANNBP. The main purpose of employing other techniques is to highlight the capability of the proposed technique in term of the performance quality. Initially, all the parameter settings for each predictive modeling algorithm are predefined. The parameter settings are summarized in Table 2.

4.2. **Performance measure.** A simple but useful concept from [4] is used to evaluate the performance of the prediction based on the error of the acceptance rate, E, within the accuracy of the guard band, A, as shown in Equation (8).

$$E = \left| \frac{predict - actual}{predict} \right| \times 100\% \leq A \tag{8}$$

Table 1. The historical DOE data set: a and b are the dimension sizes; gh, the gap height; nb, the number of bumps; dd, the distance dispense; sw, the amount of epoxy; the output, the length of the tongue

a	b	gh	nb	dd	sw	output
14795.66	13475.28	3035.64	6782064	67870	61700	256305.3
17238.98	17238.98	3134.36	6782064	80210	49360	166709.3
6170	6170	3072.66	662658	49360	17276	114980.7
16671.34	16362.84	3356.48	6415566	74040	61700	250800.1
14795.66	13475.28	3035.64	6782064	67870	49360	237581.9
17238.98	17238.98	3134.36	6782064	80210	55530	243672.4
16671.34	16362.84	3356.48	6415566	74040	49360	215971.4
16671.34	16362.84	3356.48	6415566	67870	61700	246692
14795.66	13475.28	3035.64	6782064	67870	40722	199574.8
14795.66	13475.28	3035.64	6782064	67870	57998	251815.5

Table 2. Parameter settings for each of the predictive modeling algorithms

Technique	Parameter Settings
WKRAS	$\sigma = \sum_{i=1}^{n} \left(\|X_i\|^2 - \overline{\|X\|}^2 \right)^2$, iteration = 1000 (whichever is reached first), artificial samples generation
WKR [5]	$\sigma = \sum_{i=1}^{n} \left(\|X_i\|^2 - \overline{\|X\|}^2 \right)^2$, iteration = 1000 (whichever is reached first),
ANNBP	Input Layer (6 nodes), One Hidden Layer (15 nodes with sigmoid function), Output Layer (1 node with linear function), momentum rate = 0.9, learning rate = 0.7 and stopping criteria either training error MSE < 10e-6 or iteration = 1000 (whichever is reached first)

Within a specified acceptance rate, the coverage accuracy, as shown in Equation (9), is calculated to determine how many samples fulfill the setting guard band value.

$$C = \frac{total\ number\ of\ accepts}{total\ number\ of\ predictions} \times 100\% \tag{9}$$

4.3. **Results.** The presented results in Table 3 show the coverage accuracy of all employed techniques for three different guard band values. The WKRAS achieves a high accuracy even dealing with a very small and sparse dataset, as compared with the WKR and ANNBP. Also, the ANNBP tended to produce inconsistent predictions when dealing with small samples due to the non-deterministic nature of the ANNBP [14]. As a result, ANNBP had the worst prediction quality. Hence, we only reported the average coverage accuracy for ANNBP in this study.

TABLE 3. The coverage accuracy of all predictive modeling algorithms

Technique	Sample Size		Coverage Accuracy, C (%)		
	Train	Test	A = 8%	A = 12%	A = 15%
WKRAS	4	6	66.67	100	100
WKR [5]	4	6	50	83.33	83.33
ANNBP	4	6	50	50	50

TABLE 4. Calculated correlation coefficients of the intel dataset using Equation (6)

	c_1	c_2	c_3	c_4	c_5	c_6
Correlation Factor	0.6966	0.6208	0.3777	0.7839	0.5394	0.9352

The introduction of artificial samples based on the correlation factor reflected the particular input relationship against the output and successfully filled the information gaps of the original training samples. Theoretically, the introduction of these samples agreed with the nature of the dataset, in which the length of the tongue from the under-fill process was highly correlated with the amount of dispensed epoxy. In other words, the 'sw' input contributed substantially to the length of the tongue, as shown in Table 4. Explicitly, the calculated coverage accuracy also agrees with the assumption of the correlation factor in generating the artificial samples. However, if the generated artificial samples are not well represented the actual distribution of the training samples, the performance of the WKRAS may degrade and causes a large error.

The chosen guard band values provide an indicator for the engineer and facilitate the establishment of a new experiment for a new product at a certain confidence level. As a result, the experiment conducted here to model the under-fill process will allow the full use of resources and indirectly reduce costs by creating a recipe from the proposed technique.

5. **Conclusions.** Because of limited information, learning from small samples is extremely difficult, especially during the under-fill process of the assembly process. This study shows that the modified version of kernel regression with artificial samples, namely WKRAS, is superior to the existing technique, WKR and ANNBP. The WKRAS requires a training process that incorporates artificial samples during the learning stages to find the optimum weight before the model is ready to use for the recipe generation process development. The artificial sample generation successfully signifies the dependencies of

the highly correlated input parameter to the output and provides necessary information when there is no training sample available. In the future, a technique to systematically generate artificial samples will be investigated to increase the number of relevant samples and thereby improve the prediction of the model and avoid irrelevant artificial samples generation.

Acknowledgment. This work was financially supported by the Universiti Teknologi Malaysia (UTM) GUP Research Funds (Q.J130000.7109.02H20)

REFERENCES

[1] J. W. Fowler, Modeling and analysis of semiconductor manufacturing, *Dagstuhl Seminar*, 2002.
[2] M. Kurban and U. B. Filik, Next day load forecasting using artificial neural network models with autoregression and weighted frequency bin blocks, *International Journal of Innovative Computing, Information and Control*, vol.5, no.4, pp.889-898, 2009.
[3] M. I. Shapiai et al., A non-linear function approximation from small samples based on Nadaraya-Watson Kernel regression, *Proc. of the 2nd International Conference on Computational Intelligence, Communication Systems and Networks*, Liverpool, UK, pp.28-32, 2010.
[4] W. Lee and S. Ong, Learning from small data sets to improve assembly semiconductor manufacturing processes, *Proc. of the 2nd ICCAE*, Singapore, pp.50-54, 2010.
[5] M. I. Shapiai et al., Recipe generation from small samples by weighted kernel regression, *Proc. of International Conference on Modeling Simulation and Applied Optimization*, Kuala Lumpur, Malaysia, pp.1-4, 2011.
[6] G. Bloch et al., Support vector regression from simulation data and few experimental samples, *Information Sciences*, vol.178, pp.3813-3827, 2008.
[7] R. Andonie, Extreme data mining: Inference from small datasets, *International Journal of Computers, Communications and Control*, vol.5, pp.280-291, 2010.
[8] R. Lanouette et al., Process modeling with neural networks using small experimental datasets, *Computers and Chemical Engineering*, vol.23, pp.1167-1176, 1999.
[9] P. Niyogi et al., Incorporating prior information in machine learning by creating virtual examples, *Proc. of the IEEE*, vol.86, 1998.
[10] C. Huang and C. Moraga, A diffusion-neural-network for learning from small samples, *International Journal of Approximate Reasoning*, vol.35, pp.137-161, 2004.
[11] T. Tsai and D. Li, Approximate modeling for high order non-linear functions using small sample sets, *Expert Systems with Applications*, vol.34, pp.564-569, 2008.
[12] M. I. Shapiai et al., Enhanced Nadaraya-Watson Kernel regression: Surface approximation for extremely small samples, *Proc. of the 5th Asia Modelling Symposium*, Kuala Lumpur, Malaysia and Manila, Philippines, pp.7-12, 2011.
[13] J. Zhang et al., An improved kernel regression method based on Taylor expansion, *Applied Mathematics and Computation*, vol.193, pp.419-429, 2007.
[14] M. Graczyk et al., Nonparametric statistical analysis of machine learning algorithms for regression problems, *Knowledge-Based and Intelligent Information and Engineering Systems*, pp.111-120, 2010.