Incorporating dimensional analysis and artificial neural networks into achieving process modeling in IC manufacturing

Kun-Lin Hsieh*
Department of Information Management, National Taitung University, 684, Chung Hua Road, Sector 1, Taitung, Taiwan, ROC

Received 09 August 2007; revised 09 May 2008

This study presents applications of dimensional analysis (DA) and artificial neural networks (ANNs) for constructing model in many applications in chemistry, physics and mechanics. An illustrative example for spin coating of thin films in IC manufacturing is employed to demonstrate rationality and effectiveness of proposed approach.

Keywords: Artificial neural networks, Dimensional analysis, Dimensionless terms, Function approximation

Introduction

Functional relationship may exist among inputs and outputs for a system\(^1\). There are lots of methodologies of model construction\(^1\) (direct experiment solution, test run, experiment of design, analytic solution, numerical solution, similitude modeling, and analogy). If relationships among variables are ambiguous or complicated, or number of variables is large, the model is difficult to be constructed. Hence, dimensional analysis (DA) has been used into system’s model construction\(^2\) - \(^7\). Primary advantage of DA is data independence, which is similar as principle component analysis\(^8\) (PCA). Model derived from DA will have characteristics of energy transformation, and relationship among input variables and output variables can be obtained. However, capability of data independence will be viewed as a difficulty for real application. Generally, regression analysis (RA) will be then used to make the model construction with assumption of linear relationship\(^9\). However, non-linear relationship frequently exists in real world and linear assumption made by RA will limit the application of model construction. Among approaches addressing function approximation, artificial neural networks\(^10\) - \(^12\) (ANNs) had been used well and applied into estimation, prediction and so on.

Artificial Neural Networks (ANNs)

Conventional supervised learning neural models include perceptron, backpropagation neural network (BPNN), learning vector quantization (LVQ), and counter propagation network (CPN); BPNN model\(^10\) - \(^13\) is frequently used and, therefore, selected herein.

The equation\(^12\) - \(^13\) utilized to adjust weights following the presentation of an input/output pair for output layer \(k\) is

\[
\Delta W_{kj} = \eta \delta_k O_j
\]

where \(\Delta W_{kj}\) = change to be made in weight from \(j\)-th to \(k\)-th unit following presentation of an input pattern, \(\delta_k\) = error signal for unit \(k\) after presentation of an input pattern, \(O_j\) = \(j\)-th element of output pattern produced by presentation of an input pattern, \(\eta\) = learning rate that governs how fast network will encode a set of input/output patterns.

BPNN for changing weights following presentation of input/output pair for hidden layer \(j\) is

\[
\Delta W_{ji} = \eta \delta_j O_i
\]

where \(\Delta W_{ji}\) = change to be made in weight from \(j\)-th to \(i\)-th unit following presentation of an input pattern, \(\delta_j\) = error signal for unit \(j\) after presentation of an input pattern, \(O_i\) = \(i\)-th element of output pattern produced by presentation of an input pattern, \(\eta\) = learning rate that governs how fast network will encode a set of input/output patterns.
In training this type of network, a signal input pattern is presented and network adjusts set of weights in all connecting links such that desired output is obtained at output node (Fig. 1). On accomplishing the adjustment, next pair of input and output target value is presented and network learns that association.

In this study, DA and ANNs are applied into developing an integrated approach for achieving functional approximation in IC manufacturing.

**Proposed Approach for Model Construction**

Advantage of DA is that meaningful multipliers of dimensionless terms can be understood by practitioners from the viewpoint of engineering practice\(^9,14\). Besides, it can reach the purpose of dimension reduction under the case with many variables. Although, constructed model will have better explanation capability after performing DA, the corresponding problems still be met (how the parameters can be estimated? How to estimate such parameters simply?). Hence, BPNN has been applied in the proposed model (Fig. 2).

**Model Construction**

To construct proposed model, clearly define the problem. Determine related variables about the problem. From related theories and engineering experiences, practitioners can determine the variable to be studied to their chosen problem.

Initialize weights between layers, \( \Delta W_{kj} \) and \( \Delta W_{ji} \);

\[ \text{for } m=1 \text{ to output layer iteration number or error criterion } \]

Decrease adjusting rate for \( \Delta W_{kj} \) and \( \Delta W_{ji} \) gradually;

\[ \text{for } n=1 \text{ to training samples size } \]

- calculate output for each hidden node;
- calculate output for each output nodes;
- accumulate difference between actual and target outputs;
- calculate modified gradient for \( \Delta W_{kj} \);
- calculate modified gradient for \( \Delta W_{ji} \);
- modify \( \Delta W_{kj} \);
- modify \( \Delta W_{ji} \);

**End**

Fig. 1—Training process for backpropagation neural network

Define the basic dimension. According to chosen variables, CGS unit and dimension can be decided. Then, basic dimension can be obtained. Construct dimension matrix according to basic dimension and chosen variables. Elements in such matrix will denote power for variables with respect to basic dimensions. Count of dimensionless multiplier terms will be computed as “number of variables including dependent and independent variables – the rank of matrix A”. Then structure dimension matrix (Fig. 3).

To form dimensionless multiplier term, list all dimensionless multiplier according to different combination of unit matrix, and possible dimensionless multiplier terms can be determined. Transfer form and make the form of \( \pi \) is the same as the known important dimensionless term \( \pi \). Necessary transformation will be done to make \( \pi \) terms obtained be the same as known dimensionless multiplier terms in practice\(^{16,17}\) according to the problem domain.

Before forming training set, objective function must be detected by engineers to check if it needs to modify. Then, according to the form, practitioners need to collect the data. Training dataset include about 90% data size.

---

![Flowchart diagram of proposed approach](image-url)
The response variable will be output signal of such model and other dimensionless multiplier terms will be the input variables. And practitioners can use them to train BPNN.

According to related parameters’ settings (learning rate, momentum, and epochs), practitioners should repeat BPNN model construction with different architectures. Practitioners also can set stopping criterions to determine stopping point (error less than a cutoff value or it arrives a predefined training cycles). Suitable evaluation index can be chosen to compare the result for different architectures. As for BPNN, root mean square of error (RMSE), distance value or correlation value will be frequently employed into making final decisions.

An Illustrative Example
To demonstrate procedure for proposed approach, an example owing to the spin coating of thin films in IC industry has been considered as per steps in Fig. 2.

Step 1. Clearly Define the Problem
During IC manufacturing processes, process of spin coating before photolithography is an important process. Thickness and uniformity will be two quality characteristics being considered in this process. Thicker coating can provide a better viscosity and pinholes and erosion will be reduced. However, thinner coating can provide a higher resolution. Hence, thickness of spin coating will be the problem to study.

Step 2. Determine Related Variables About The Problem
From related theories and engineering experiences, several variables can be chosen to have the effect on thickness of spin coating [density ($\rho$), thermal conductivity coefficient ($k$), latent heat ($\lambda$), specific heat ($C_p$), rotation speed ($\Omega$), viscosity ($\mu$) and diffusion coefficient ($D$)]. The partial data will be given in Table 1.

Step 3. Define the Basic Dimension
After reviewing related variables and response variable (CGS unit and dimension will be listed in Table 2), basic dimensions of mass (M), length (L), time (T) and temperature ($\theta$) will be defined.

Step 4. Construct Dimensional Matrix
Then, a dimension matrix (denoted as $A$) can be constructed as per structure of dimension matrix (Fig. 4).

where $\text{Rank}(A) = 4$, hence, count of dimensionless multipliers will be $8-4 = 4$
<table>
<thead>
<tr>
<th>Physical dimension</th>
<th>CGS unit</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thickness (H)</td>
<td>cm</td>
<td>L</td>
</tr>
<tr>
<td>density (ρ)</td>
<td>g/cm³</td>
<td>M/L³</td>
</tr>
<tr>
<td>thermal conductivity coefficient (k)</td>
<td>cal/cm/sec</td>
<td>ML/T</td>
</tr>
<tr>
<td>latent heat (λ)</td>
<td>cal/g</td>
<td>L²/T²</td>
</tr>
<tr>
<td>specific heat (Cp)</td>
<td>cal/g</td>
<td>L²/T²</td>
</tr>
<tr>
<td>rotation speed (Ω)</td>
<td>hz</td>
<td>1/T</td>
</tr>
<tr>
<td>viscosity (µ)</td>
<td>g/cm/sec</td>
<td>M/LT</td>
</tr>
<tr>
<td>diffusion coefficient (D)</td>
<td>cm²/sec</td>
<td>L²/T</td>
</tr>
</tbody>
</table>

Step 5. Perform Necessary Computation Of Matrix

Let $K = \begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ \vdots \\ K_n \end{bmatrix}$

Then, after necessary computation $A \times K = 0$, one can obtain dimensional matrix (Fig. 5). Left part of matrix can be viewed as a unit matrix, and right part as a residual matrix.

Step 6. Form Dimensionless Multiplier Term

Due to different combination of unit matrix, different dimensionless multipliers are obtained (Table 3).

One can obtain, $\pi_1 = \frac{L\sqrt{\Omega}}{\sqrt{D}}$, $\pi_2 = \frac{\rho D}{\mu}$, $\pi_3 = \frac{K}{\rho C_p \mu}$, $\pi_4 = \frac{\lambda}{\Omega D}$.

From related studies$^{2,3,5,7}$, one can find the term $(\mu/\rho D)$ to be Schmidt Number (dimensionless multiplier in mass transfer), term $(C_p \mu/k)$ to be Prandtl Number (dimensionless multiplier in heat transfer), and term $(k/\rho C_p D)$ to be Lewis Number (Schmidt Number/Prandtl Number). Next, necessary transformations are made to let chosen dimensionless term to be the same as the known dimensionless terms.

Step 7. Transfer Form and Make Form of $\bar{A}$ Is Same as Known Important Dimensionless Term

Transfer $\pi_i$ to the known dimensionless multipliers.

Hence, $\frac{H\sqrt{\Omega}}{\sqrt{D}}$, $\frac{\mu}{\rho D}$, $\frac{C_p \mu}{K}$, $\frac{\lambda}{\Omega D}$ will be chosen to be final dimensionless multiplier terms in this case. The equation of thickness of spin coating can be expressed as

$$H\frac{\sqrt{\Omega}}{\sqrt{D}} = f$$

$$(\mu \frac{\rho}{D} \frac{C_p \mu}{K} \frac{\lambda}{\Omega D}) = \alpha_1 (\mu \rho D) \alpha_2 (\frac{C_p \mu}{K}) \alpha_3 (\frac{\lambda}{\Omega D}) \alpha_4 \ldots (1)$$
Step 8. Form Training Data Set

Before forming training set, take Eq. (1) into another form for the purpose of prediction for thickness. Transferred equation will be denoted in Eq. (2)

\[ H = \alpha_1 \left( \frac{\mu}{\rho D} \right) \alpha_2 \left( \frac{C_p \mu}{K} \right) \alpha_3 \left( \frac{\lambda}{\Omega D} \right) \alpha_4 \left( \frac{\sqrt{D}}{\sqrt{\Omega}} \right) \]

\[ = f \left( \frac{\mu}{\rho D}, \frac{C_p \mu}{K}, \frac{\lambda}{\Omega D}, \frac{\sqrt{D}}{\sqrt{\Omega}} \right) \]

Each term in right side of Eq. (2) will be set as one processing element for ANNs. There are 24 data, and about 20 data are taken randomly as training dataset. Rest will be set as testing data.

Step 9. Training of Neural Network and Make Judgment to Neural Model

Suitable neural network is constructed via try and error, and evaluation index is set as RMSE of training and testing\[^1\]. For simplifying operation of construction, use professional software (Neural Profession Plus II) to perform necessary ANNs model construction. Herein, BPNN will be chosen to construct the model. Thickness
will be set as the response variable \( \frac{\mu}{\rho D^*} \), \( \frac{C_p \mu}{K} \), \( \frac{\lambda}{\Omega D} \) and \( \frac{\sqrt{D}}{\sqrt{\Omega}} \) will be set as the independent variable (number of processing element in input layer of BPNN will be set as 4). Optimum BPNN architecture with DA will be 4-5-1 due to compromising RMSE values of training/testing phase (Table 4).

Besides, original data structure is directly applied to construct BPNN model. Herein, thickness will be set as response variable and density (\( \rho \)), thermal conductivity coefficient (k), latent heat (\( \lambda \)), specific heat (\( C_p \)), rotation speed (\( \Omega \)), viscosity (\( \mu \)) and diffusion coefficient (D) will be set as independent variable (number of processing element in input layer of BPNN will be set as 7). Optimum BPNN architecture without DA will be 7-8-1 due to compromising RMSE values of training/testing phase (Table 5).

Reviewing the comparison, model with DA (Table 4) is found better model due to RMSE value (0.024146) being less than that (Table 5) without DA (0.035928). Herein, a cutoff value for RMSE value is set to be 0.05. According to the result obtained from above steps, BPNN model is obtained.

To verify effectiveness of proposed approach, evaluation indexes were applied to analyze it. The first index is degree of similarity, which can represent correlation between predicted data set and original data set; and it can verify degree of effectiveness for proposed model. Herein, one can find out that correlation for prediction with DA is larger than that without DA (0.993 > 0.989). Second index is distance, which can denote difference effect between predicted value and actual value. Herein, total distance is directly used to denote difference effect. And, distance value for prediction with DA is less than that without DA (0.0628 > 0.0799). From such comparison result, one can find out that the model construction with DA is better than that without DA.

### Concluding Remarks

A new model construction incorporating DA and ANNs is proposed. Effectiveness of proposed approach has been demonstrated via an illustrative example owing to spin coating of thin films in IC industry. Such procedure can be viewed as a rational reference in the future for similar cases that physical dimensions exist. Modeling procedure incorporating DA and ANNs has been found as a feasible and rational approach. Proposed model can be used since performing the pilot run or pre-test run before mass production. Engineers can initially obtain useful information about process modeling, and necessary analysis may be processed after such analysis. It can be regarded as an improvement approach for manufacturing industry.

### References


---

**Table 4.** Comparison table of RMSE for different architectures with DA

<table>
<thead>
<tr>
<th>Architecture</th>
<th>RMSE of training</th>
<th>RMSE of testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-2-1</td>
<td>0.0835</td>
<td>0.1025</td>
</tr>
<tr>
<td>4-3-1</td>
<td>0.0629</td>
<td>0.0742</td>
</tr>
<tr>
<td>4-4-1</td>
<td>0.0444</td>
<td>0.0563</td>
</tr>
<tr>
<td>4-5-1*</td>
<td>0.0243</td>
<td>0.0387</td>
</tr>
<tr>
<td>4-6-1</td>
<td>0.0208</td>
<td>0.0469</td>
</tr>
<tr>
<td>4-7-1</td>
<td>0.0184</td>
<td>0.0518</td>
</tr>
</tbody>
</table>

*denote optimum choice

**Table 5.** Comparison table of RMSE for different architectures without DA

<table>
<thead>
<tr>
<th>Architecture</th>
<th>RMSE of training</th>
<th>RMSE of testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-3-1</td>
<td>0.09125</td>
<td>0.15224</td>
</tr>
<tr>
<td>7-4-1</td>
<td>0.07386</td>
<td>0.13367</td>
</tr>
<tr>
<td>7-5-1</td>
<td>0.06675</td>
<td>0.11263</td>
</tr>
<tr>
<td>7-6-1</td>
<td>0.05988</td>
<td>0.08753</td>
</tr>
<tr>
<td>7-7-1</td>
<td>0.05324</td>
<td>0.06428</td>
</tr>
<tr>
<td>7-8-1*</td>
<td>0.04672</td>
<td>0.05336</td>
</tr>
<tr>
<td>7-9-1</td>
<td>0.03275</td>
<td>0.05855</td>
</tr>
<tr>
<td>7-10-1</td>
<td>0.03028</td>
<td>0.06469</td>
</tr>
</tbody>
</table>

*denote optimum choice


12 NeuralWare Inc, NeuralWorks Professional II/Plus and NeuralWorks Explorer (NeuralWare, Inc: Penn Center West) 1990.


