Classification of the Degradation of Soft Sensor Models and Discussion on Adaptive Models

Hiromasa Kaneko¹, Kimito Funatsu*¹

¹ Department of Chemical System Engineering, Graduate School of Engineering, The University of Tokyo, Hongo 7-3-1 Bunkyo-ku, Tokyo 113-8656, Japan
* e-mail: funatsu@chemsys.t.u-tokyo.ac.jp

Abstract – Soft sensors are used widely to estimate a process variable which is difficult to measure online. One of the crucial difficulties of soft sensors is that predictive accuracy drops due to changes of state of chemical plants. It is called as the degradation of soft sensor models. In this study, we attempted to classify this degradation of models in terms of changes in an explanatory variable and an objective variable, and the rapidity of the changes. Moreover, we discussed characteristics of adaptive soft sensor models, based on the classification results. By analyzing simulated data sets, we could obtain knowledge and information on appropriate adaptive models for each type of the degradation.

Keyword: Process control, Soft sensor, Degradation, Adaptive model, Predictive ability
1. Introduction

In chemical plants, soft sensors have been widely used to estimate process variables that are difficult to measure online [1, 2]. An inferential model is constructed between those variables that are easy to measure online and those that are not, and an objective variable, \( y \), is then estimated using that model. Through the use of soft sensors, the values of \( y \) can be estimated with a high degree of accuracy.

Their use, however, involves some practical difficulties. One crucial difficulty is that their predictive accuracy decreases due to changes in the state of chemical plants, catalyzing performance loss, sensor and process drift, and the like. This is called as the degradation of soft sensor models. If the degradation is not solved, it is difficult to identify reasons of abnormal situations. On the site of plants, when a prediction error of \( y \) is above a threshold, it is recognized as an abnormal situation. There is no effective method to judge whether the reason of the abnormal situation is the trouble of \( y \)-analyzer or the degradation of a soft sensor model under present circumstances.

It is therefore strongly desired to solve the degradation. In order to reduce the degradation of a soft sensor model, the model is reconstructed with newest data. A moving window (MW) model [3, 4] and a recursive model [5] are categorized as a sequentially updating type and a distance-based just-in-time (JIT) model [6] and a correlation-based JIT model [7] are categorized as a JIT type. For example, a MW model is constructed with data that are measured most recently and a distance-based JIT model is constructed with data whose distances to prediction data are smaller than those of other data. Many excellent results have been reported based on the use of MW models and JIT models.

Meanwhile, problems of reconstructing a model such as the incorporation of abnormal data with training data and an increase of maintenance costs were discussed, and then, a model based on the time difference of \( y \) and that of explanatory variables, \( X \), is proposed [8, 9]. This model is referred to as a time difference (TD) model. The effects of deterioration with age such as the drift and gradual changes in the state of a plant can be accounted for by using a TD model without reconstruction of the model.

Kaneko et al. analyzed data obtained from the operation of a distillation column and compared a MW model and a TD model [8]. The results suggested that the predictive ability of each adaptive model changed depending on states of the plant. For instance, the TD model could predict \( y \) with high accuracy soon after and before process variations and the MW model had high predictive ability in process variations. We can say that the understanding of the state of a plant and the use of an appropriate model for each state are needed for highly accurate prediction.

In this study, therefore, we classify the degradation of soft sensor models in terms of the types of changes in \( X \) and \( y \) and the rapidity of the changes, and then, discuss adaptive models such as MW, JIT and TD models, which is based on the classification results in order to construct and select models that can handle each degradation. Finally, the results of the discussion are confirmed by using simulation data sets.

2. Research Methodology

2.1. Classification of the Degradation

Fig. 1 shows basic concepts of the degradation of a linear soft sensor model constructed between an explanatory variable, \( x \), and \( y \). Fig. 1(a) and (b) represent shifts of \( y \)-values and \( x \)-values, respectively. These are corresponding to sensor and process drift, scale deposition on a pipe, a change of operating condition such as the amount of raw material, and so on. The slope does not change between training data and new data, but values of a \( y \)-variable or an \( x \)-variable shift.

Fig. 1(c) represents a change of the slope of \( x \) and \( y \). This is corresponding to catalyzing performance loss, a change of operating condition such as a concentration in raw material, and so on. Of course, shifts of \( y \)-values and \( x \)-values, and a change of the slope will occur simultaneously.

When we focus on the rate of the degradation, each shift or change happens gradually, rapidly, or instantly. For example, catalyzing performance loss, process and sensor drift, a change of external temperature, and scale deposition on a pipe occur gradually; a sharp change in raw material occurs rapidly; and correction of drift, regular repair of a plant, and a stoppage of a pipe occur instantly.
2.2. Characteristics of Adaptive Models

A TD model can adapt shifts of both y-values and x-values because it achieves the same effect as a bias update. Even when the shifts happen gradually, rapidly, and instantly, a TD model can follow the shifts appropriately. On the one hand, it is difficult for a MW model to adapt a rapid or instantaneous shift because old data before the shift remains in training data.

Meanwhile, a TD model cannot adapt a change of the slope [9]. A MW model should be used to follow a gradual change of the slope by adding new data to training data. However, it is difficult for a MW model to handle a rapid or instant change because the model is affected by old data before the change.

In case of a JIT model, which is constructed with data sets close to test data in the space of \( x \), appropriate selection of data sets will be performed if a shift of \( x \)-values happens. However, besides that, data sets after a shift of \( y \)-values or a change of the slope cannot be selected because there is no change in the space of \( x \) as shown in Fig. 1(a) and (c).

3. Results and Discussion

In order to test predictive ability of each adaptive model when each type of the degradation occurs, we used simulated data sets. The number of \( X \)-variables was set as two. First, \( X \) of uniform pseudorandom numbers whose range was from 0 to 10 was prepared. Then, \( y \) was set as follows:

\[
y = Xb + UOD + N(0, 0.1)
\]  

where \( b \) means the magnitude of contribution of \( X \) to \( y \); \( UOD \) means unobserved disturbances; and \( N(0, 0.1) \) is random numbers from normal distribution given a standard deviation of 0.1 and a mean of 0. Shifts of \( y \)-values and \( x \)-values, and changes of the slope (Fig. 1) can be represented by changing \( UOD, X, \) and \( b \), respectively. Additionally, we can consider gradual, rapid, or instant shifts and changes. Lastly, \( N(0, 0.1) \) was added also to \( X \).

The number of data was 200. The first 100 data was used for training and the next 100 data was the test data. We used a PLS method [10] as a regression approach and set the number of data for constructing a MW model and a JIT model as 20. In this paper, not a correlation based JIT model [7] but a Euclidian distance-based JIT model was used because there is little correlation between \( X \)-variables.

3.1. Shift of \( y \)-values

First, we explain the prediction results where three types of \( UOD \) were used as follows:

- For \( UOD = 0.01t \)
  \[
  UOD = 3\sin(0.04t)
  \]  
  
- For \( UOD = \begin{cases} 
0 & (1 \leq t \leq 20, 101 \leq t \leq 120) \\
5 & (21 \leq t \leq 100, 121 \leq t \leq 200)
\end{cases}
\]

where \( t \) means time. Eqs. (2), (3), and (4) represent gradual, rapid, and instant changes of \( UOD \), respectively.

Table 1 shows the \( r_{pred}^2 \) values of each model for each \( UOD \). The \( r_{pred}^2 \) is the determination coefficient, \( r^2 \), for test data. As we discussed, the TD models had high predictive ability for all types of \( UOD \). Though the MW model could follow the gradual change of \( UOD \), the \( r_{pred}^2 \) values got low when \( UOD \) changed rapidly and instantly. The \( r_{pred}^2 \) value of the JIT model was the lowest in all cases, indicating appropriate selection of training data could not be performed.

Fig. 2 shows the relationships between \( y \) and predicted \( y \) of test data for each model when Eq. (4) was used as \( UOD \). The plot of Fig. 2(c) shows a much tighter cluster of predicted values along the diagonal for the TD model though the MW model also had high predictive accuracy. There seem two outliers in Fig. 2(c), which are corresponding to the times when the value of \( UOD \) changes from 0 to 5 and from 5 to 0. Practically, these changes mean, for example, the instant shift of \( y \)-values just after drift correlation. The occurrence of this event can be noticed in advance, and therefore, there is no problem with the outliers in practice. In addition, the same outliers are included also in training data. The accuracy of the TD model can be improved by eliminating the outliers.

The distribution of the plot in Fig. 2 (b) is divided in two. This indicates that the training data for predicting test data could not be selected appropriately in the construction of the JIT models when \( UOD \) was 0 or 5. We concluded that TD models should be used for these kinds of the degradation of a soft sensor model.

Table 1. \( r_{pred}^2 \) values of each model for each \( UOD \)

<table>
<thead>
<tr>
<th>UOD</th>
<th>MW model</th>
<th>JIT model</th>
<th>TD model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2)</td>
<td>0.999</td>
<td>0.965</td>
<td>1.000</td>
</tr>
<tr>
<td>(3)</td>
<td>0.782</td>
<td>0.747</td>
<td>0.998</td>
</tr>
<tr>
<td>(4)</td>
<td>0.816</td>
<td>0.781</td>
<td>0.977</td>
</tr>
</tbody>
</table>

3.2. Change of the Slope

Second, we set the magnitude of contribution of the first explanatory variable to \( y \), \( b_1 \), as 1 and that of the second explanatory variable to \( y \), \( b_2 \), as follows:

\[
b_1 = 3\sin(0.01\pi t) + 1
\]

\[
b_2 = 3\sin(0.02\pi t) + 1
\]
Eqs. (5) and (6) represent gradual and rapid changes of \( b_2 \), respectively. Then, we compared the prediction results of the MW, JIT, and TD models.

The \( r_{\text{pred}}^2 \) values and the root mean squared error (RMSE) values of test data of each model for each \( b_2 \) are shown in Table 2 and Table 3, respectively. In both cases where Eqs. (5) and (6) were used, the MW models had the highest predictive accuracy in three adaptive models as discussed in 2.2. However, the MW models could not adapt the latest relationship between \( X \) and \( y \) because the number of data for the construction of the MW models was set as 20, and then, the MW models were constructed with data where the relationship between \( X \) and \( y \) was old. Hence, the \( r_{\text{pred}}^2 \) values of the MW models were not so high.

In addition, as shown in Table 3, the RMSE values of the MW models increased as the rate of the change of the slope increased, reflecting that it is difficult for a MW model to handle a rapid or instant change of the slope because the model is affected by old data as mentioned in 2.2. But, the MW models could predict test data more accurately than the JIT and TD models did.

\[
\begin{align*}
\text{Table 2. } r_{\text{pred}}^2 \text{ values of each model for each } b_2 \\
\text{Eq. for } b_2 & \quad \text{MW model} & \quad \text{JIT model} & \quad \text{TD model} \\
(5) & 0.546 & -5.932 & -6.077 \\
(6) & 0.574 & 0.009 & 0.377 \\
\end{align*}
\]

\[
\begin{align*}
\text{Table 3. RMSE values of each model for each } b_2 \\
\text{Eq. for } b_2 & \quad \text{MW model} & \quad \text{JIT model} & \quad \text{TD model} \\
(5) & 3.85 & 15.05 & 15.21 \\
(6) & 7.32 & 11.17 & 8.85 \\
\end{align*}
\]

3.3 Shift of \( x \)-values and Change of the Slope

Lastly, we changed the first \( X \)-variable, \( x_1 \), to \( x_{1, \text{new}} \) below:

\[
x_{1, \text{new}} = x_1 + 0.01t 
\]

where \( x_{1, \text{new}} \) means the shifted values of \( x_1 \). Eqs. (7) and (8) represent gradual and instant changes of \( x_1 \). We also set \( b_2 \) as (5) or as follows:

\[
b_2 = \begin{cases} 1 & (1 \leq t \leq 20, 101 \leq t \leq 120) \\ 2 & (21 \leq t \leq 100, 121 \leq t \leq 200) \end{cases}
\]

In cases where both \( x_{1, \text{new}} \) and \( b_2 \) change gradually, the TD model could not follow the change of the slope and the appropriate JIT model could not be constructed as shown on the low values of \( r_{\text{pred}}^2 \), -5.676 and -6.071, in Table 4. The MW model had the highest predictive accuracy in three adaptive models though the \( r_{\text{pred}}^2 \) value was not so high. The MW model could adapt the change...
of $b_2$ given in Eq. (5) to an extent. However, when $x_{1,\text{new}}$ shifted instantly as given in Eq. (8), the $r_{\text{pred}}^2$ value was 0.213 and very low, reflecting that the MW model could hardly follow the instant shift of $x$-values.

When the slope of $x$ and $y$ changed instantly, i.e. Eq. (9) was used as $b_2$, the $r_{\text{pred}}^2$ value of the JIT model was very high if $x$-values also changed instantly. Fig. 3 shows the relationships between $y$ and predicted $y$ of test data for each model when Eq. (8) was used as $x_{1,\text{new}}$ and Eq. (9) was used as $b_2$. On the one hand, many data are far from the diagonal in the plots of Fig. 3 (a) and (c); on the other hand, the plot of Fig. 3(b) shows a much tighter cluster of predicted values along the diagonal for the JIT model. But, if $x$-values changed gradually, the predictive accuracy of the MW, JIT, and TD models was low totally. We can say that the data selection of a JIT model succeeds and a predictive JIT model can be constructed if the change of the slope and the shift of $x$-values occur simultaneously.

![Fig. 3. Relationships between $y$ and predicted $y$ of test data for each model when Eq. (8) was used as $x_{1,\text{new}}$ and Eq. (9) was used as $b_2$.](image)

Table 4. $r_{\text{pred}}^2$ values of each model for each $x_{1,\text{new}}$ and each $b_2$

<table>
<thead>
<tr>
<th>$x_{1,\text{new}}$</th>
<th>Eq. for $b_2$</th>
<th>MW model</th>
<th>JIT model</th>
<th>TD model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(7)</td>
<td>-</td>
<td>0.999</td>
<td>0.967</td>
<td>1.000</td>
</tr>
<tr>
<td>(8)</td>
<td>-</td>
<td>0.264</td>
<td>0.911</td>
<td>0.696</td>
</tr>
<tr>
<td>(7)</td>
<td>(5)</td>
<td>0.539</td>
<td>-5.676</td>
<td>-6.071</td>
</tr>
<tr>
<td>(8)</td>
<td>(5)</td>
<td>0.213</td>
<td>-6.777</td>
<td>-0.880</td>
</tr>
<tr>
<td>(7)</td>
<td>(9)</td>
<td>0.527</td>
<td>0.498</td>
<td>0.576</td>
</tr>
<tr>
<td>(8)</td>
<td>(9)</td>
<td>0.807</td>
<td>0.990</td>
<td>0.897</td>
</tr>
</tbody>
</table>

4. Conclusions

In this paper, we categorized the degradation of a soft sensor model and discussed characteristics of MW, JIT, and TD models, based on the classification results. Then, we confirmed the discussion results with simulated data sets. The results suggested that there exist appropriate models for each type of the degradation. When the shift of $y$-values or $x$-values occurs, a TD model is suitable. But, we should give attention to data just after the instant shift. A MW model is appropriate for the gradual change of the slope of $x$ and $y$.

When the slope of $x$ and $y$ changes instantly, a JIT model should be used if the shift of $x$-values also happens and the amount of training data that are close to new data is enough, but if not, a MW model will have higher predictive ability than a JIT model and a TD model do. However, the predictive accuracy of a MW model is not sufficient, and therefore, other methods are needed for predictive soft sensors.

We plan to analyze the real industrial data [8] and discuss adequate adaptive models for each state of the plant. By using the knowledge we obtained in this study, the predictive accuracy of soft sensors will improve, and then, chemical plants can be operated stably.

References


pc015-5