Automatic Baseline-Sample-Selection Scheme for Baseline Predictive Maintenance

Chun-Fang Chen, Yao-Sheng Hsieh, Fan-Tien Cheng, Hsien-Cheng Huang, and Saint-Chi Wang

Institute of Manufacturing Information and Systems
National Cheng Kung University
Tainan, Taiwan, R.O.C.

August 18, 2013
Contents

- Introduction
- Automatic Baseline-Sample-Selection Scheme
  - Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
  - Step 2: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
  - Step 3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
  - Step 4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)
- Illustrative Examples
- Conclusions and Future Work
Contents

- Introduction

- Automatic Baseline-Sample-Selection Scheme
  - Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
  - Step 2: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
  - Step 3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
  - Step 4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)

- Illustrative Examples

- Conclusions and Future Work
The baseline-predictive-maintenance (BPM) scheme based on virtual-metrology technology was proposed recently.

By applying the BPM scheme, fault diagnosis and prognosis can be accomplished and the requirement of massive historical failure data can also be released.

Due to the merit of not requiring historical failure data, automatic creation of a BPM model just after maintenance becomes possible. This makes the BPM scheme more applicable for on-line implementation.

The purpose of this paper is to develop an automatic baseline-sample-selection (ABSS) scheme to automatically prepare the modeling samples for creating the BPM model.
Introduction (2/4)

BPM Scheme

Baseline Models

Related Process Data

Z-Score

TD Baseline Model

ISIB Model

Z-Score

Target Device Data

DHI Module

BEI Module

FDC

PdM

RUL Predictive Module

RUL

TD Remaining Useful Life (RUL)

Normal

TD Abnormal

TD is OOC Due to Process Data Abnormal

Normal but Process Data Abnormal
Set throttle valve to be the target device (TD).

Introduction (3/4)

Red-rectangular blocks are un-healthy due to abrupt changes of the throttle values while their related process parameters do not have changes accordingly.

The health and quality check process is so tedious, which may prohibit the usage of the BPM scheme.
The C&H samples are selected manually, and several contradictory samples are embedded in the C&H samples.

Those contradictory samples should be discovered and deleted as well.
Contents

- Introduction
- Automatic Baseline-Sample-Selection Scheme
  - Step1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
  - Step2: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
  - Step3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
  - Step4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)
- Illustrative Examples
- Conclusions and Future Work
Introduction

Automatic Baseline-Sample-Selection Scheme

- Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
- Step 2: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
- Step 3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
- Step 4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)

Illustrative Examples

Conclusions and Future Work
Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples (1/5)

Healthy Samples Selection Scheme

- The necessity of including C&H samples in the baseline model is to enlarge the model-operating space.
- The flow of the HSS scheme is shown on left-hand side with four steps.

1. Collect enough historical $y_T$ samples together with their corresponding $X$.
2. Apply HMA filter to retain potential healthy samples with large deviation and delete un-healthy ones.
3. Exclude those healthy samples with large deviation and then apply DQIx & DQly algorithms to confirm the quality of the remaining healthy samples.
4. Combine those large-deviation potential healthy samples with the remaining good-quality healthy samples.
Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples (2/5)

1. Collect enough historical y_T samples together with their corresponding X.

2. Apply HMA filter to retain potential healthy samples with large deviation and delete un-healthy ones.

3. Exclude those healthy samples with large deviation and then apply DQ_lx & DQ_ly algorithms to confirm the quality of the remaining healthy samples.

4. Combine those large-deviation potential healthy samples with the remaining good-quality healthy samples.

- Apply the hollow-moving-average (HMA) filter to retain potential healthy samples with large deviation and delete un-healthy ones.

  - The large deviation of the potential healthy samples in TD datum is caused by the abnormal values of the related parameters.

  - The TD should be in healthy status.
Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples (3/5)

- Hollow-moving-average (HMA) filter:

The 2N samples are used to calculate HMA values ($\bar{y}_{T-i}$ & $\bar{x}_{j,i}$)

Target sample is used to calculate $i^{th}$ TD sample ($y_{T-i}$) and its corresponding $j^{th}$ process datum ($x_{j,i}$)

**Equation (1)**

$$\bar{y}_{T-i} = \begin{cases} y_{T-2} + y_{T-3} + y_{T-4} + y_{T-5} + y_{T-6} + y_{T-7}, & \text{if } i = 1 \\ y_{T-1} + y_{T-3} + y_{T-4} + y_{T-5} + y_{T-6} + y_{T-7}, & \text{if } i = 2 \\ y_{T-1} + y_{T-2} + y_{T-4} + y_{T-5} + y_{T-6} + y_{T-7}, & \text{if } i = 3 \\ y_{T-i-3} + y_{T-i-2} + y_{T-i-1} + y_{T-i+1} + y_{T-i+2} + y_{T-i+3}, & \text{if } 4 \leq i \leq n-3 \\ y_{T-n-6} + y_{T-n-5} + y_{T-n-4} + y_{T-n-3} + y_{T-n-1} + y_{T-n}, & \text{if } i = n-2 \\ y_{T-n-6} + y_{T-n-5} + y_{T-n-4} + y_{T-n-3} + y_{T-n-2} + y_{T-n}, & \text{if } i = n-1 \\ y_{T-n-6} + y_{T-n-5} + y_{T-n-4} + y_{T-n-3} + y_{T-n-2} + y_{T-n-1}, & \text{if } i = n \end{cases}$$

**Equation (2)**

$$\bar{x}_{j,i} = \begin{cases} x_{j,2} + x_{j,3} + x_{j,4} + x_{j,5} + x_{j,6} + x_{j,7}, & \text{if } i = 1 \\ x_{j,1} + x_{j,3} + x_{j,4} + x_{j,5} + x_{j,6} + x_{j,7}, & \text{if } i = 2 \\ x_{j,1} + x_{j,2} + x_{j,4} + x_{j,5} + x_{j,6} + x_{j,7}, & \text{if } i = 3 \\ x_{j,i-3} + x_{j,i-2} + x_{j,i-1} + x_{j,i+1} + x_{j,i+2} + x_{j,i+3}, & \text{if } 4 \leq i \leq n-3 \\ x_{j,n-6} + x_{j,n-5} + x_{j,n-4} + x_{j,n-3} + x_{j,n-1} + x_{j,n}, & \text{if } i = n-2 \\ x_{j,n-6} + x_{j,n-5} + x_{j,n-4} + x_{j,n-3} + x_{j,n-2} + x_{j,n}, & \text{if } i = n-1 \\ x_{j,n-6} + x_{j,n-5} + x_{j,n-4} + x_{j,n-3} + x_{j,n-2} + x_{j,n-1}, & \text{if } i = n \end{cases}$$

where $i = 1, 2, ..., n$, $j = 1, 2, ..., p$
Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples (4/5)

- The operating instructions of HMA filter:
  - Let $y_{ME} = \left| y_{T_i} - \bar{y}_{T_i} \right|$ and $x_{ME,j} = \left| x_{j,i} - \bar{x}_{j,i} \right|
  - Define the process Spec (the border of sickness) and HardSpec (the border of death) of the TD. Also, assign the X_spec of each $x_{j,i}$

```
HardSpec

\[ y \]

\[ x \]
```

where $i = 1, 2, \ldots, n$

$\quad j = 1, 2, \ldots, p$
Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples (5/5)

The operating instructions of HMA filter:

- Potential Healthy Sample due to \( y_{ME} > \text{Spec}, \) \( y_{T,i} < \text{HardSpec}, \) and \( x_{ME,j} > X_{\text{spec}} \)
- Abnormal due to \( y_{ME} < \text{Spec}, \) and \( x_{ME,j} < X_{\text{spec}} \)
- Abnormal due to \( y_{T,i} > \text{HardSpec} \)
- Abnormal due to \( y_{ME} < \text{Spec}, \) and \( x_{ME,j} > X_{\text{spec}} \)
Contents

- Introduction
- Automatic Baseline-Sample-Selection Scheme
  - Step 1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
  - Step 2: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
  - Step 3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
  - Step 4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)
- Illustrative Examples
- Conclusions and Future Work
Step 2: Applying the dynamic-moving-window (DMW) scheme to choose C&H samples

- The dynamic-moving-window (DMW) scheme is capable of adding a new sample into the model with a predefined window size and apply a clustering technology to do similarity clustering.

- Next, the number of samples in each cluster is checked. If the largest number of the samples is greater than the predefined threshold, then the oldest sample in the cluster of the largest population is deleted.

- Therefore, the DMW scheme is applied by the ABSS scheme to choose the C&H samples with a predefined number of sample size.
Introduction

Automatic Baseline-Sample-Selection Scheme

- **Step1**: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
- **Step2**: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
- **Step3**: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
- **Step4**: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)

Illustrative Examples

Conclusions and Future Work
Step3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples

- Fresh Samples are essential and fundamental for establishing the baseline model.
- Enough fresh and healthy samples newly-generated after maintenance should be collected on-line.
- After finishing collection, those newly-generated samples are added to the C&H historical samples.

Total historical & healthy data include $y_T$ and its corresponding $X$ (maybe Thousands of samples)

Running DMW to choose C&H Samples (Ex: 30 samples)

PM

Fresh Samples after Maintenance (Ex: 10 samples)

TD baseline modeling set (Ex: 30+10 samples)
Contents

- Introduction
- Automatic Baseline-Sample-Selection Scheme
  - Step1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
  - Step2: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
  - Step3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
  - Step4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)
- Illustrative Examples
- Conclusions and Future Work
Step 4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI) (1/2)

Start

4.1 Run Residual Analysis to obtain the \( t_i \) of each modeling sample

4.2 Run BPM Scheme on the 1\(^{st} \) and 2\(^{nd} \) available samples just after maintenance and calculate their BEI values

4.3 Both BEI values \( \geq 0.95 \)

4.4 Delete the modeling sample that has Max \( |t_i| \)

4.5 Reserve the remaining samples for modeling

End

Studentized deleted residual:

\[
t_i = \frac{e_i \sqrt{m - p - 1}}{\sqrt{SS_E \cdot (1 - h_i) - e_i^2}}
\]

- \( m \) is the number of modeling samples
- \( p \) is the total number of process parameters
- \( SS_E \) is the sum of squared errors
- \( h_i \) is the hat matrix diagonal element for observation \( i \)
Step 4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI) (2/2)

- The baseline error index (BEI) module converts $y_E (= |y_T - \hat{y}_B|)$ into the BEI index.
- The threshold of BEI for checking a sample being contradictory or not is assigned to be 0.95.

![Diagram showing the process of Step 4](image)

### Configurations of SPC Control Charts of BEI
(Converting $y_E$ into BEI)

- **Healthy**: $y_E = 0$ (BEI = 1)
- **Sick**: $y_E = 0.3$ (BEI = 0.7)
- **Dead**: $y_E = 1$ (BEI = 0)
- **Hard Spec**: $y_T - \bar{y}$
- **Spec**: $0.7$

---

**4.1 Run Residual Analysis to obtain the ($t_i$) of each modeling sample**

**4.2 Run BPM Scheme on the 1st and 2nd available samples just after maintenance and calculate their BEI values**

**4.3 Both BEI values $\geq 0.95$**

**4.4 Delete the modeling sample that has Max $|t_i|$**

**4.5 Reserve the remaining samples for modeling**

**End**
Contents

- Introduction
- Automatic Baseline-Sample-Selection Scheme
  - Step1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
  - Step2: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
  - Step3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
  - Step4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)
- Illustrative Examples
- Conclusions and Future Work
Example 1

Illustration of the Effect of Deleting Contradictory Samples with Residual Analysis and baseline-error-index (BEI)
The C&H samples are selected manually, and several contradictory samples are embedded in the C&H samples.
### Example 1: Testing results of deleting contradictory samples with residual analysis and baseline-error-index (BEI)

<table>
<thead>
<tr>
<th>No. of Round</th>
<th>Deletion of Contradictory Samples</th>
<th>BEI Values of the 1(^{st}) and 2(^{nd}) Test Samples</th>
<th>(t_i) value of deleted sample</th>
<th>Amount of Samples of Erroneous Sick-state Triggering Ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(^{st}) Round</td>
<td>0</td>
<td>0.909/0.913</td>
<td>-2.854</td>
<td>44</td>
</tr>
<tr>
<td>2(^{nd}) Round</td>
<td>1</td>
<td>0.910/0.914</td>
<td>-2.087</td>
<td>43</td>
</tr>
<tr>
<td>3(^{rd}) Round</td>
<td>2</td>
<td>0.890/0.891</td>
<td>-2.602</td>
<td>51</td>
</tr>
<tr>
<td>4(^{th}) Round</td>
<td>3</td>
<td>0.942/0.947</td>
<td>-3.483</td>
<td>23</td>
</tr>
<tr>
<td>5(^{th}) Round</td>
<td>4</td>
<td>0.958/0.962</td>
<td>--</td>
<td>2</td>
</tr>
</tbody>
</table>

- In the 5\(^{th}\) round, due to the fact that both of the BEI values (0.958/0.962) are greater than 0.95, the flow of deleting contradictory samples is accomplished.
- Four contradictory samples are deleted totally and the amount of samples of erroneous sick-state triggering ahead is reduced from 44 to 2.
**Example 1:**

BPM results with deletion of four contradictory samples

Remedying the problem that the initial C&H samples have the contradictory samples.
Example 2

BPM results with the Complete ABSS Scheme
### Example 2: Testing results of deleting contradictory samples with residual analysis, BEI and HSS

<table>
<thead>
<tr>
<th>No. of Round</th>
<th>Deletion of Contradictory Samples</th>
<th>BEI Values of the 1st and 2nd Test Samples</th>
<th>$t_i$ value of deleted sample</th>
<th>Amount of Samples of Erroneous Sick-state Triggering Ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Round</td>
<td>0</td>
<td>0.945/0.950</td>
<td>-3.545</td>
<td>22</td>
</tr>
<tr>
<td>2nd Round</td>
<td>1</td>
<td>0.957/0.962</td>
<td>--</td>
<td>2</td>
</tr>
</tbody>
</table>
Example 2:

BPM results with the complete ABSS scheme

The C&H samples are chosen by the HSS scheme, and the operating range of the throttle values is large.

Remedying the problem that health & quality check process is so tedious.
Contents

- Introduction
- Automatic Baseline-Sample-Selection Scheme
  - Step1: Applying the healthy-samples-selection (HSS) scheme to delete un-healthy samples
  - Step2: Applying the dynamic-moving-window (DMW) scheme to choose concise-and-healthy (C&H) samples
  - Step3: Collecting enough fresh samples after maintenance and combining those C&H and fresh samples
  - Step4: Deleting contradictory samples with residual analysis and baseline-error-index (BEI)
- Illustrative Examples
- Conclusions and Future Work
Conclusions

- An ABSS scheme for selecting the C&H samples and deleting the contradictory samples automatically is proposed in this paper.

- Two illustrative examples are presented to demonstrate the effect of deleting contradictory samples and the BPM results with the preliminary ABSS scheme.
Future Work

- This preliminary ABSS scheme is merely suitable for treating large amount of historical raw samples. If historical samples are not available or not easy to collect, then the C&H samples cannot be obtained. As such, only the fresh samples after maintenance are available.

- How can the BPM modeling samples be prepared when only the fresh samples after maintenance are available is the future work.
Thank you for your listening!

Q&A