Intelligent Trend Indices and Recursive Modelling in Prognostics

Esko Juuso
Control Engineering Laboratory, Department of Process and Environmental Engineering, P.O.Box 4300, FI-90014 University of Oulu, Finland
Phone: +358-8-5532463
E-mail: esko.juuso@oulu.fi

Sulo Lahdelma
Mechatronics and Machine Diagnostics Laboratory, Department of Mechanical Engineering, P.O.Box 4200, FI-90014 University of Oulu, Finland
E-mail: sulo.lahdelma@oulu.fi

Abstract

Machines seldom break down suddenly without warning. Therefore, it is crucial to monitor the vibration levels as a function of time. Regression analysis and fuzzy reasoning with triangular episodic representations, which are based on the estimated first and second derivatives of a feature, have been used in trend analysis. Trend indices are calculated from the scaled values by using the means obtained for a short and a long time period. The severity of the situation can be evaluated with a deviation index, which combines the current level, the trend index and the change of the trend index. This index obtains its highest absolute values, when there is considerable difference between the variable and the reference point and this difference continues to grow with increasing speed. The scaling functions are updated in recursive modelling, which is triggered by a fast increase of the indices. Exponential models are used in the real scale. Vibration measurements from two paper machines are analysed in this paper. Periodic measurements from a press roll in the felt washer provided useful data for detecting the gradual increase of the resin problem at an early stage. Operating conditions had a major effect on the resonance of the press section.

1. Introduction

Temporal reasoning is a very valuable tool for diagnosing and controlling slow processes. Manual process supervision relies heavily on monitoring visually of the characteristic shapes of changes in process variables, especially their trends. Although the human eye can readily detect such patterns, it is a difficult problem for control system software. \(^{(1, 2)}\) In these cases, the appropriate time window is not always recognised. Fault diagnosis is a special area of detecting operating conditions.

The trend extraction method of Cheung and Stephanopoulos is called qualitative scaling \(^{(3)}\). It generates a hierarchical multiscale structure of trend descriptions over different time scales from process data. The first versions of intelligent trend indices were presented in \(^{(3)}\). A new scaling approach for finding the meanings of features was introduced in \(^{(3)}\). This approach uses the generalised norms and moments presented in \(^{(5)}\).
This paper deals with a new intelligent trend index approach based on variable meanings obtained by means of data analysis. The methodology is tested with measurements taken from two paper machines.

2. Data analysis

The mathematical expectation, expected value, or briefly the expectation, of a random variable is a very important concept in probability and statistics. The moments can be defined about some central value, e.g. the mean, median and mode. The variance \( \text{Var}(X) = \sigma^2 \) can be represented by

\[
\sigma^2 = M^2 = \mathbb{E}[(X - E(X))^2] = E(X^2) - [E(X)]^2,
\]

(1)

The positive square root of the variance, \( \sigma_X \), or briefly \( \sigma \), which is called the standard deviation, is used more often since it has the same dimension as the variable. Dimensionless features can be obtained by normalising the moments, for example by standard deviation \( \sigma_X \):

\[
\gamma_k = \frac{E[(X - E(X))^k]}{\sigma_X^k}.
\]

(2)

The feature \( \gamma_3 \) is called the coefficient of skewness, or briefly skewness, and the feature \( \gamma_4 \) as the coefficient of kurtosis. The skewness is a measure of asymmetry: \( \gamma_3 = 0 \) for a symmetric distribution. If \( \gamma_3 > 0 \), the skewness is called positive skewness and the distribution has a long tail to the right, and vice versa if \( \gamma_3 < 0 \).

A norm defined by

\[
\left\| M^p \right\|_p = (\mathbb{E} M^p)^{1/p} = \left( \frac{1}{N} \sum_{i=1}^{N} M_{i}^{p} \right)^{1/p},
\]

(3)

where \( \alpha \in \mathbb{R} \) is the order of derivation, the order of the moment \( p \in \mathbb{R} \) is non-zero. The signal is measured continuously, and the analysis is based on consecutive equally sized samples. Duration of each sample is called sample time, denoted \( \tau \). The number of signal values \( N = \tau N_s \) where \( N_s \) is the number of signal values which are taken in a second. This norm, which has the same dimensions as the signal \( x^{(\alpha)} \), is defined in such a way that \( -\infty < p < \infty, p \not= 0 \), i.e. the definition includes the \( l_p \) norms defined for \( 1 < p < \infty \). This norm introduced in (5) is a Hölder mean, also known as the power mean. The norm values increase with increasing order, i.e. for the generalised norms holds

\[
(\mathbb{E} M^p)^{1/p} \leq (\mathbb{E} M^q)^{1/q},
\]

if \( p < q \). The increase is monotonous if all the signals are not equal.

The normalised moments (2) were in (4) generalised by replacing the expectation value with the norm (3) as the central value:

\[
\gamma_k = \frac{\mathbb{E}\left[\left(X^{(\alpha)} - \left\| M^p \right\|_p\right)^k\right]}{\sigma_X^k},
\]

(5)
where $\sigma_x$ is calculated about the origin, and $k$ is a positive integer.

The basic idea of the linguistic equation (LE) methodology is the nonlinear scaling developed to extract the meanings of variables from measurement signals. The scaling function scales the real values of variables to the range of $[-2, +2]$, which combines two monotonously increasing functions: one for the values between -2 and 0, and one for the values between 0 and 2\(^6\). The functions consist of two second-order polynomials, which are defined by the points

\[
\{ (\min(x_j), -2), ((c_j)_j, -1), (c_j, 0), ((c_h)_j, 1), (\max(x_j), 2) \}.
\]

The feasible range can be represented by a trapezoidal membership function defined by the core $[ (c_j)_j, (c_h)_j ]$ and the support $[ \min(x_j), \max(x_j) ]$. The scaling functions are monotonously increasing if the coefficients, $a_j^-$ and $a_j^+$, are both in the range $\left[\frac{1}{3}, 3\right]$. Corrections are done by changing the borders of the core area, the borders of the support area or the centre point. Additional constraints for derivatives can also be taken into account. The coefficients of the polynomials can be represented by

\[
a_j^- = \frac{1}{2}(1 - a_j) \Delta c_j^-,
\]

\[
b_j^- = \frac{1}{2}(3 - a_j) \Delta c_j^-,
\]

\[
a_j^+ = \frac{1}{2}(a_j^+ - 1) \Delta c_j^+,
\]

\[
b_j^+ = \frac{1}{2}(3 - a_j^+) \Delta c_j^+,
\]

where $\Delta c_j^- = c_j - (c_j)_j$ and $\Delta c_j^+ = (c_h)_j - c_j$. Membership definitions may contain linear parts if some coefficients $a_j^-$ or $a_j^+$ equal to one.

The best way to tune the system is to first define the working point $c_j$ and the core, then the ratios $a_j^-$ and $a_j^+$ from the range $\left[\frac{1}{3}, 3\right]$, and finally to calculate the support. The scaling functions of each variable are configured with five parameters, including the centre point $c_j$ and three consistent sets:
The upper and the lower parts of the scaling functions can be convex or concave, independent of each other. Simplified functions can also be used, e.g. a linear membership definition requires two and an asymmetrical linear definition three parameters. Additional constraints can be taken into account for derivatives, e.g. locally linear function results if continuous derivative is chosen in the centre point.

In the analysis of the corner points, the value range of $x_j$ is divided into two parts by the central tendency value $c_j$, and then the core area, $[(c_l)_j, (c_h)_j]$, is limited by the central tendency values of the lower and upper part. Generalised skewness $\gamma_j$ defined by (5) is used in estimating the central tendency value and the core area. Derivation is not used, i.e. $\alpha = 0$. The central tendency value is chosen by the point where the skewness changes from positive to negative, i.e. $\gamma_j = 0$. Then the data set is divided into two parts: a lower part and an upper part. The same analysis is done for these two data sets. The estimates of the corner points, $(c_l)_j$ and $(c_h)_j$, are the points where the skewness goes to zero. The iteration is performed with generalised norms. Then the ratios $\alpha_j^-$ and $\alpha_j^+$ are restricted to the range \([\frac{1}{3}, 3]\) moving the corner points $(c_l)_j$ and/or $(c_h)_j$ or the upper and lower limits $\min(x_j)$ and/or $\max(x_j)$. The linearity requirement in the working point $c_j$ is taken into account, if possible.

3. Trend analysis

Typical reasoning systems have three components: a language to represent the trends, a technique to identify the trends, and mapping from trends to operational conditions (7). The fundamental elements are modelled geometrically as triangles to describe local temporal patterns in data (Figure 1). The elements are defined by the signs of the first and second derivative, respectively. These elements, which are also known as triangular episodic representations (8), have their origin in qualitative reasoning and simulation (9, 10). An interval-halving scheme to facilitate the automatic extraction of temporal features from sensor data in terms of the trend language of primitives was presented in (11).

Detecting changes is important for data compression, process control and fault diagnosis. Linear regression combined with fuzzy reasoning can be used to detect significant changes up or down in process variables (12). The episodes shown in Figure 1 provide features for more detailed analysis. Trend extraction methods are based on polynomials, linear segments, wavelets and B-splines, and also neural networks are used. Similarities between trends are analysed, e.g. with sequence matching, decision trees, pattern recognition and hidden Markov models, see (13).
Figure 1. Triangular episodic representation: seven basic types of episodes used for interval description, each episode type is denoted by a letter from the set \{A, B, C, D, E, F, G\} \(^2\).

For any variable \( x_j \), a trend index \( I_j^T(k) \) is calculated from the scaled values \( X_j \) with a linguistic equation

\[
I_j^T(k) = w_j \left[ \frac{1}{n_s + 1} \sum_{i=k-n_s}^{k} X_j(k) - \frac{1}{n_L + 1} \sum_{i=k-n_L}^{k} X_j(k) \right],
\]

which is based on the means obtained for a short and a long time period, defined by delays \( n_s \) and \( n_L \), respectively. The weight \( w_j \) is variable specific. The index value is in the linguistic range \([-2, 2]\) representing the strength of both decrease and increase of the variable \( x_j \) \(^3\).

The derivative of the index \( I_j^T(k) \), denoted as \( \Delta I_j^T(k) \), is used for analysing triangular episodic representations (Figure 1). An increase is detected if the trend index exceed a threshold \( I_j^T(k) > \varepsilon_i^+ \), and correspondingly, \( I_j^T(k) < \varepsilon_i^- \) for a decrease (Figure 2). These trends are linear if the derivative is close to zero: \(-\varepsilon_2^- < \Delta I_j^T(k) < \varepsilon_2^+\). Concave upward monotonic increase (D) and concave downward monotonic decrease (B) are dangerous situations. Concave downward monotonic increase (A) and concave upward monotonic decrease (C) mean that an unfavourable trend is stopping.

The trend detection logic is similar to a typical PI type linguistic equation (LE) controller \(^4\). The system could be implemented as a fuzzy set system by generating the
membership functions \(^{(6)}\). In the present system, fuzzy logic is only used to explain the progress of the analysis.

\[
\Delta I_j^T(k) = \left( k_{ITj} - k_{ITj+1} \right) \frac{1}{3} \left( X_j(k) + I_j^T(k) + \Delta I_j^T(k) \right).
\]

This index has its highest absolute values, when the difference to the set point is very large and is getting still larger with a fast increasing speed \(^{(3)}\). This can be understood as an additional dimension in Figure 2.

Severity of the situation can be evaluated by a deviation index

\[
I_j^D(k) = \frac{1}{3} \left( X_j(k) + I_j^T(k) + \Delta I_j^T(k) \right).
\]

Condition and stress indices provide information about the severity of condition problem and stress level, respectively. The severity levels are consistent with the vibration severity criteria. Performance measures can be handled in the same way, e.g. improvements of overall equipment effectiveness (OEE) can be based on following classification: \([0.25, 0.75)\) slight, \([0.75, 1.25)\) good, \([1.25, 1.75)\) very good, and \(\geq 1.75\) excellent improvement \(^{(15)}\).

Models can be specific to the triangular episodes shown in Figures 1 and 2. In condition monitoring, the concave upward monotonic increase requires special attention. New phenomena are activated when the system moves from linear increase to this area. The scaling functions should be defined for the whole operating area. However, a wide range of measurements is only available after failure. Initial estimates are needed at the beginning for the corner points \(^{(6)}\). Parameters for the lower part \([-2, 0]\) are updated when a change to episode D is detected with the trend analysis based on the initial estimates of the scaling functions. Parameters for the upper part \([0, +2]\) can be updated recursively.
4. Applications

Trends are analysed from rms velocity, \( v_{rms} = x^{(1)}_{rms} \), measurements of two paper machines: the resonance of the press section and resin problems of a press roll in the felt washer \(^{(16)}\). In both cases the \( v_{rms} \) measurements are scaled with the nonlinear scaling method described above, and the trend index (9) is calculated using appropriate short and long time periods.

4.1 Press section

Measurements are quite infrequent in this case. The scaled values increase from -2 to 2, and the reduced machine speed is seen as decreasing values (Figure 3(a)). The trend index \( I_j^T(k) \) is zero at the beginning, the increase corresponding to \( I_j^T(k) \approx 1 \), and finally the index reaches the value 1.5 (Figure 3(b)). The short and long periods are two and four measurements, respectively. The weight \( w_j = 1.5 \).

![Graphs showing trend analysis for a press section.](image)

**Figure 3.** Intelligent trend analysis for a press section.

The derivative \( \Delta I_j^T(k) \) is around zero at the beginning and then increases fast, \( \Delta I_j^T(k) \approx 1 \) (Figure 3(c)). For the fast increase \( \Delta I_j^T(k) \) reaches value 1.5. The fast increase stops, but starts again before the end of the period. The deviation index \( I_j^D(k) \)
is around or below zero a long time at the beginning (Figure 3(d)). Higher values after Day 26 give a warning, and the highest values at the end of the period can be regarded as alarms. The trend analysis is delayed since the measurements are infrequent.

The episode analysis starts from normal and moves to a concave upward increase, then returns to normal and again starts to speed up at the end. The roll was changed when the concave upward area was reached. The change is very strong.

![Figure 4. Trend episodes for a press section.](image)

4.2 Felt washer

The scaled values increase from -2 to 2: at the beginning the increase is slow but becomes faster at the end of the period (Figure 5(a)). The trend index $I_j^T(k)$ is zero at the beginning, the slight increase corresponds to $I_j^T(k) \approx 0.5$, and finally the index reaches the value $1.5$ (Figure 5(b)). The short and long periods are three and nine measurements, respectively. The weight $w_j = 1.5$.

The derivative $\Delta I_j^T(k)$ is around zero at the beginning and then increases slightly, $\Delta I_j^T(k) \approx 0.5$ (Figure 5(c)). For the fast increase $\Delta I_j^T(k)$ reaches value $1.5$. The fast increase stops, but starts again before the end of the period. The deviation index $I_j^D(k)$ is around or below zero a long time at the beginning (Figure 5(d)). Higher values after Day 84 give a warning, and the highest values at the end of the period can be considered as alarms.
Figure 5. Intelligent trend analysis for a press section.

The episode analysis starts from normal, gives the first warnings about the increase of $I_T^f(k)$, then slight linear increase, which first turns into a concave upward increase and then returns to quite strong linear increase that again starts to speed up at the end (Figure 6). The roll was changed just before the concave upward area would have been reached.

Figure 6. Trend episodes for a press section.
5. Conclusions

The new trend analysis methods provide logical results for the problems in the press section, although the measurements were very infrequent. More values can be used in the analysis of the roll in the felt washer, since there are much more measurements. The roll was changed at the right time. All parameters were obtained by means of data analysis. The indices and the episodes provide clear indications when the predictive models should be changed or updated.

The nonlinear scaling methodology based on generalised norms and skewness provides good results for scaling functions. Intelligent trend indices are suitable for real applications and can be tuned by means of data analysis. Some estimates of the limits are needed at the beginning, but later recursive tuning can be used for fine tuning the scaling functions.

References