Title: SYSTEM MONITORING

Abstract: A method of novelty detection for system monitoring utilises multimodal multivariate extreme value theory to set a novelty threshold based on the probability distribution of extreme values drawn from a model of the system. Extreme values are defined as the least probable of a set of values based on the system model.
SYSTEM MONITORING

The present invention relates to the field of systems monitoring and in particular to the automated, continuous analysis of the condition of a system.

Systems monitoring is applicable to fields as diverse as the monitoring of machines, or the monitoring of human patient's vital signs in the medical field, and typically such monitoring is conducted by measuring the state of the system using a plurality of sensors each measuring some different parameter or variable of the system. To assist in the interpretation of the multiple signals acquired from complex systems, developments over the last few decades have led to automated analysis of the signals with a view to issuing an alarm to a human user or operator if the state of the system departs from normality. A basic and traditional approach to this has been to apply a threshold to each of the individual sensor signals, with the alarm being triggered if any, or a combination of, these single-channel thresholds is breached.

However, it is often difficult to set such thresholds automatically at a point which on the one hand provides a sufficiently safe margin by alarming reliably when the system departs from normality, but on the other hand does not generate too many false alarms, which leads to alarms being ignored. Further, such single-channel thresholds do not allow for situations where the system is in an abnormal state as indicated by an abnormal combination of signals from the sensors even though each individual signal is within its individual single-channel threshold.

Consequently more recently techniques have been developed which assess the state of a system relative to a model of normal system condition, with a view to classifying data from the sensors as normal or abnormal with respect to the model.

Such novelty detection, or 1-class classification, is particularly well-suited to problems in which a large quantity of examples of normal behaviour exist, such that a model of normality may be constructed, but where examples of abnormal behaviour are rare, such that a traditional multi-class approach cannot be taken. Novelty detection is therefore useful in the analysis of data from safety-critical systems such as jet engines, manufacturing processes, or power-generation facilities, which spend the majority of their operational life in a normal state, and which exhibit few, if any, failure conditions. It is also applicable, though, in the medical field, where human vital signs are treated in the same way as data acquired from mechanical systems.
As indicated above, novelty detection is performed with respect to a model of normality for the system. Such a model can typically be produced by taking a set of measurements of the system while it is assumed to be in a normal state (these measurements then being known as the training set) and fitting some analytical function to the data. For example, for multivariate and multimodal data the function could be a Gaussian Mixture Model (GMM), Parzen Window Estimator, or other mixture of kernel functions. In this context, multivariate means that there are a plurality of variables – for example each variable corresponds to a measurement obtained from a single sensor or some single parameter of the system and multimodal means that the function has more than one mode (i.e. more than one local maximum in the probability distribution function that describes the distribution of values in the training set). The model of normality can therefore be represented as a probability density function $p(x)$ (the GMM or other function fitted to the training set) over a multidimensional space with each dimension corresponding to an individual variable or parameter of the system.

Having constructed such a model of normality one approach to novelty detection is simply to set a novelty threshold on the probability density function such that a data point $x$ is classified as abnormal if the probability density $p(x)$ is less than the threshold. Such thresholds are simply set so that the separation between normal and any abnormal data is maximised on a large validation data set, containing examples of both normal and abnormal data labelled by system domain experts. A similar alternative approach is to consider the cumulative probability function $P(x)$ associated with the probability distribution: that is to find the probability mass $P$ obtained by integrating the probability density function $p(x)$ up to the novelty threshold and to set the threshold at that probability density which results in the desired integral value $P$ (for example so that 99% of the data is classified normal with respect to the threshold). This allows a probabilistic interpretation, namely: if one were to draw a single sample from the model, it would be expected to lie outside the novelty threshold with a probability $1-P$. For example, if the threshold were set such that $P$ is 0.99, so that 99% of single samples could be expected to be classified normal, then $1-P$ is 0.01, and 1% of single samples would expected to be classified abnormal with respect to that threshold. However, these approaches encounter the problem that although the probabilistic interpretation is valid for consideration of a single sample taken from the model, if multiple samples are taken from the model, as
occurs in the continuous monitoring of real-life systems, the probability that the
novelty threshold will be exceeded increases, and is no longer given by 1-\(P\). Thus
while the technique above is valid for applications where one is comparing a single
measurement to a model of normality (for example comparing a single mammogram
to a model constructed using "normal" mammogram data) it is not valid for
applications where systems are being continually monitored with sensor
measurements being sampled on a continual basis generating a continual stream of
readings.

Because abnormal states of a system will generally be associated with extreme
values of the variables being measured, interest has developed in using extreme value
theory in the monitoring of systems. Extreme value theory is a branch of statistics
concerned with modelling the distribution of very large or very small values (extrema)
with respect to the probability distribution function describing the location of the
normal data. Extreme value theory allows the examination of the probability
distribution of extrema in data sets drawn from a particular distribution. For example
Figure 1 of the accompanying drawings illustrates a Gaussian distribution labelled
\(p(x)\) of one dimensional data \(x\) (i.e. a univariate unimodal distribution) in the solid
line with corresponding extreme value distributions (EVD) labelled \(p^*(x)\) for data sets
having different numbers of samples \(m=10, 100, 1000\). Thus the extreme value
distribution gives the probability of each value of \(x\) appearing as an extremum in a set
of \(m\) data points drawn randomly from the Gaussian distribution. The shape of the
extreme value distribution can be understood by considering that points which are at
the centre of the Gaussian distribution are very unlikely to appear as extrema of a data
set, whereas points far from the centre (the mode) of the Gaussian are quite likely to
be extrema if they appear in the data set, but they are not likely to appear very often.
Thus as illustrated the form of the EVD is that it takes low values at the centre and
edge of the Gaussian with a peak between those two areas. The particular shape of
the curve for a Gaussian distribution of data is a Gumbel distribution.

Figure 1 also illustrates the problem mentioned above of setting a threshold
(labelled \(x_K\)) on a particular data value. Although it can be seen that for data sets with
small values (e.g. \(m=10\)) the peak of the EVD is below the threshold \(x_K\), which means
that the most probable extreme value of such data sets (which, it should be recalled,
are data from a system in its normal condition), are below the threshold, as the size of
the data set increases the peak of the EVD moves to the right, above the threshold, so
that for data sets of 100 or 1000 samples the most likely extreme values are beyond the threshold $x_K$. This means that even though the system is normal, an extremum of a large data set is quite likely to trigger a false alarm by exceeding the threshold, and the situation gets worse as more readings are taken (i.e. as $m$ increases).

Because of these problems, extreme value theory has been proposed for novelty detection in the engineering, health and finance fields. By examining the extreme value distribution it is possible to use it to classify data points as normal or abnormal. It is possible, for example, to set a threshold on the extreme value distribution, for example at 0.99 of the integrated Gumbel probability distribution, which can be interpreted as meaning that out of a set of actual measurements on the system, if the extremum of those measurements is outside the threshold, this has less than a 1% chance of being an extremum of a normal data set. Consequently, that measurement can be classified as abnormal. Obviously the threshold can be set as desired.

However, existing work has been limited to unimodal univariate data for example as illustrated in Figure 1 and, as mentioned above, for complex systems data is likely to be multivariate and may also be multimodal.

Figure 2 illustrates a bivariate Gaussian distribution (the centre peak) together with its corresponding extreme value distribution (the surrounding torus). Although one might expect that the novelty detection techniques used in univariate extreme value theory could straightforwardly be extended to two dimensions as illustrated in Figure 2, by using the radius from the mode as the univariate variable, in fact as the dimensionality of the data set increases, classical extreme value theory tends to introduce increasing error in its estimates of the EVD. In particular the extreme value distribution obtained analytically from the distribution of the data departs ever more significantly with increasing dimensionality from the actual extreme value distribution obtained experimentally by looking at the extrema in actual data sets. One reason for this is that as the dimensionality of the data increases (i.e., as more sensor channels are included in the monitoring system) more and more of the data space is effectively empty – in other words the probability of obtaining most of the values in the data set is zero or close to zero, and the zero or low probability density areas increase in size as dimensionality increases. Thus if one considers the distribution of probability density values in the data set, as dimensionality increases the most likely probability density value in the data space approaches zero. Figure 3 illustrates this for
multivariate Gaussian distributions of increasing numbers of dimensions. Figure 3 illustrates normalised histograms (integrating to unity) of \( p(x) \) values for \( N = 10^6 \) samples drawn from multivariate standard Gaussian distributions of dimensionality \( Q = 1, 2, 3, 4, 7, 12 \). As can be seen in the single dimension univariate case of Figure 3a the most likely probability density value obtained is 0.4. With a two dimensional data set of Figure 3b probability density values from 0 to 0.15 are all broadly similarly likely. However in Figure 3c for a three dimensional data set the very small probability density values close to zero have become much more likely than higher probability density values, with this effect increasing through Figures 3d-f as dimensionality increases. For a 12 dimensional data set as shown in Figure 3f, very low probability density values are much more likely than higher ones and the probability mass is highly clustered towards zero (i.e., if data are generated according to the model, they are more likely to be of a very low probability density \( p(x) \) as model dimensionality increases). This leads to classical extreme value theory incorrectly estimating the parameters of the extreme value distribution, and hence it cannot be used to determine where the extent of normality should lie to a set probability \( P \). This makes it inappropriate to use the extreme value distribution obtained from classical extreme value theory to classify accurately data points as normal or abnormal.

It should also be noted that the data in Figure 2 is unimodal – i.e. there is a single peak in the probability distribution. The extension of extreme value theory to multimodal, for example bimodal, data is also problematic. Figure 4 illustrates a bimodal generative probability density function (the dashed line) representing a model of normal data in a training data set, with the extreme value distribution predicted by existing methods (solid line). The bimodal distribution in Figure 4 is a mixture of two Gaussian distributions and so the extreme value distribution is a Gumbel type distribution around each of the Gaussian modes or kernels. These extreme value distributions obtained by existing classical methods are generated on the assumption that the closest Gaussian kernel dominates the distribution of extreme values and thus the other kernel can be ignored. Also illustrated in Figure 4, though, by the circles is a histogram for \( N = 10^6 \) experimentally-obtained extrema of data sets each including 100 data points. It can be seen that the fit between the experimentally obtained data (circles) and the predicted extreme value distribution (solid line) is poor.
In summary, therefore, although existing classical extreme value theory appears to offer the prospect of meaningful probabilistic interpretations of the thresholds for use in novelty detection, particularly when considering that the abnormal states of systems are likely to be reflected by extrema in the data, the extension of current techniques from univariate unimodal data to multivariate and/or multimodal data is inappropriate and has not been successful.

The present invention provides a way of extending extreme value theory to multimodal multivariate data to allow novelty detection on such data. It is based on a new way of understanding the extreme value distribution and a new way of defining extrema in the data set.

Normally an extreme value of the data set is defined to be that which is either a minimum or maximum of the set in terms of absolute signal magnitude. For example in novelty detection, when considering the extrema of unimodal distributions as illustrated in Figures 1 and 2, the extrema are at the minimum or maximum distance from the single mode of the distribution. However for multimodal data there is no single mode from which distance may be defined. For example in Figure 4 data midway between the two modes is clearly extremely unlikely, because this region has very low probability density with respect to the model, and thus represents an abnormal state for the system. However such data is not at an extreme value of x in terms of absolute magnitude, and so classical extreme value theory would not class data falling within this improbably region as being abnormal.

According to the present invention the extreme value is redefined in terms of probability given that the goal for novelty detection is to identify improbable events with respect to the normal state of the system, rather than events of extreme absolute magnitude. Thus in accordance with the present invention, for novelty detection, the most extreme of a set of m samples \(x = x_1, x_2 \ldots x_m\) distribution according to the model \(p(x)\) is that which is most improbable with respect to the probability distribution of the model.

This definition provides the mechanism required to determine the extent of data space that is considered normal: thus if m normal data distributed according to the distribution of the model of normality \(p(x)\) are observed, the Extreme Value Distribution (EVD) \(p'(x)\) describes where the least probable of those m normal data will lie. Therefore the EVD can be used to set a novelty threshold and to perform
novelty detection when multiple data from the same system are observed, as occurs with continuous monitoring of real-life systems.

As a result of defining the extrema in terms of the probability the inventors have found that the EVD can be defined as a function of the probability distribution used to model the distribution of normal data in the training set, i.e. $p^*(x) = g(p(x))$. While this new function $g()$ cannot be stated explicitly, we can accurately determine its values using a numerical scheme, and so accurately determine the EVD which describes where we expect the most improbable data to lie under normal conditions.

In accordance with the invention, and using the definition of extrema above, once one has a model of a system (for example by using a training data set and fitting a model to it to estimate the probability density function of the normal data in the training set), an extreme value distribution can be obtained not in terms of the values of the individual variables $x$ making up the data, but in terms of the probability density values of the extrema. Thus from the model artificial data sets of $m$ data points are generated, the data point with the lowest probability density value is taken as the extremum and its probability density value noted. This is repeated a large number $N$ times (such as $N = 10^6$) to generate $N$ extrema and the distribution of probability density values of these extrema is then constructed. A function can then be fitted to this distribution of the probability density values of the extrema. It is possible then to define a threshold in probability space (rather than in the original multivariate space of the data acquired from the system) so that if the probability density value of a point defined as an extremum is below the threshold then one can say that it is unlikely (to a probability defined by the threshold) to appear as an extremum of a data set from a system in a normal state. The threshold can be set, of course, to give the desired chance, e.g. 1% or 0.1% or so on, that an extremum having less than that the threshold probability density value comes from a normal data set.

Thus with the invention the threshold is set with respect to extreme values of data sets and is set in probability space, not in the data space (i.e. not in relation to actual data values themselves). This avoids the problems associated with finding appropriate data values for thresholds and the problems (described earlier) associated with increasing chances of exceeding the threshold as more and more data is obtained. It also allows the threshold to be set for multimodal and multivariate data because the definition of the extrema in terms of probability avoids the difficulties of finding extrema in a multimode distribution and the setting of the threshold in probability
space avoids the problem of the probability mass of extreme values clustering towards zero as dimensionality increases.

An embodiment of the invention provides a further enhancement which improves the ability to analyse the distribution of probability density values and set the threshold appropriately. This enhancement is to apply a transform to the probability density value of each extremum \( p(x) \) which tends to spread out the distribution of lower probability values, and enables an extreme value distribution to be easily estimated. The transform is preferably of a logarithmic form. Having set the threshold on the transformed probability density values, the threshold can either itself be converted back into a threshold on the probability density function, for use in a novelty detection system, or in the novelty detection system the probability of data points defined as extrema can be subject to the same transform for comparison with the threshold.

In more detail, therefore, the present invention provides a method of setting a normality threshold for use in a system monitor to allow automatic classification of measured states of the system as normal or abnormal, comprising:

constructing a statistical model of the normal states of the system giving the probability density value for each measured state when the system is normal;

defining as an extreme value the least probable of a set of measurements of the state of the system, the set comprising a predetermined number \( m \) of measurements;

finding from the model the distribution of the probability density values of the so-defined extreme values;

setting as a threshold a probability density value on said distribution which divides said distribution in predetermined proportions whereby extreme values having a probability density value higher than said threshold are classified as normal and extreme values having a probability density value lower than said threshold are classified as abnormal.

The statistical model can be obtained by fitting a suitable analytic function to a set of training data. As discussed above the invention is applicable to models which are multimodal and/or multivariate, though it is equally applicable to unimodal univariate models.

Preferably the distribution of probability density values of extreme values for the models is found numerically by generating from the model a number \( N \) (for example \( 10^5 \)) sets of \( m \) (for example 100) values representing measurements of the
state of the system, taking the extreme value from each set (i.e. the value with the lowest probability density value determined with respect to the model) and then constructing a histogram of the probability density values of the so-obtained $N$ extrema.

The threshold can be set directly on the distribution of the probability density values of the extrema (i.e. on the histogram) in particular at a particular value (e.g. 0.99) of the cumulative probability density. Such a threshold then has the meaning that an extreme value with a probability density value outside the threshold has less than a predetermined (1% if the threshold is set at 0.99) chance of being an extremum from a normal data set.

As mentioned above, with increasing dimensionality the distribution of probability density values of extrema tends to cluster heavily close to zero and so it can be difficult to set the threshold accurately. Consequently the probability density values of the extrema are preferably transformed to spread out the low probability density values.

It is possible to define two or more thresholds representing increasing chances of abnormality, for example to give the possibility of “amber” and “red” alarms. For example this may be achieved by setting a second threshold on said distribution for classifying extreme values into three classes (e.g., “normal”, “warning”, and “critical alert”).

The invention also extends to a system monitor which utilises thresholds set in the manner above. The monitor stores the model and is adapted to acquire a number $m$ of measurements of the state of the system, to compare each measurement to the model to find the least probable of the $m$ measurements (that thus being defined as an extremum of that data set), and then comparing the least probable of the $m$ measurements to the threshold to classify the system state as normal or abnormal.

As mentioned above the comparison can be made either on the probability density value of the extremum or after transforming that value using the transform mentioned above.

The system monitor does not analyse individual measurements one by one, but instead collects sets (or “windows”) of $m$ measurements and looks at the extremum within that set. In a system monitor which is acquiring measurements of the system state continually (for example receiving continual inputs from various sensors on the system), it preferably uses a rolling window of $m$ successive measurements. It may,
therefore be that one particular value persists as an extremum until it falls out of the rolling window.

Preferably measurements from the system when it is classified as being in a normal state are stored for addition to a new training set of data which can be used to retrain the model. In this way the model can be closely tuned to the system being monitored, and can change over time following the gradual "normal" deterioration of system condition, in the case of a mechanical system.

The invention is applicable to the monitoring of complex systems in the engineering and medical field. For example it is applicable to jet engine vibration data and to human vital sign health monitoring.

It will be appreciated that the invention is preferably embodied in software executable on a computer system and thus the invention extends to an executable software application embodying the methods, a data carrier storing the executable application and to a computer system programmed to execute the methods.

The invention will be further described by way of example with reference to the accompanying drawings in which:-

Figure 1 illustrates a Gaussian PDF $p(x)$ of data $x$ together with the corresponding Extreme Value Distribution (EVD) $p_\alpha(x)$;

Figure 2 illustrates a bivariate Gaussian distribution and corresponding EVD;

Figure 3(a)-(f) illustrate how the distribution of probability density values varies with dimensionality of a data set;

Figure 4 illustrates a bimodal probability density function with classically predicted EVD and experimentally obtained EVD;

Figure 5 illustrates probability contours on a trimodal, bivariate probability distribution together with corresponding (experimentally obtained) EVD contours;

Figure 6 is a flow diagram explaining how a novelty threshold is set in accordance with one embodiment of the invention;

Figure 7 is a flow diagram explaining a system monitor in accordance with an embodiment of the invention;

Figure 8(a) illustrates a normalised histogram of probability values for $N = 10^6$ extrema generated from the trimodal distribution of Figure 5 and Figure 8(b) illustrates in grey the histogram of psi-transformed values together with a Gumbel distribution fitted to it and a threshold set at a cumulative density of 0.99 (dashed line);
Figure 9 illustrates a psi-transformed EVD for trivariate, trimodal data and a histogram and fitted Gumbel for a 6-dimensional mixture of 15 Gaussian kernels, together with a novelty threshold dashed at $P=0.99$; and

Figures 10 (a) and (b) illustrate the performance of an embodiment of the invention on two sets of gas turbine engine data.

Figure 6 explains how a threshold is set in accordance with one embodiment of the invention. First it is necessary to have a model of the data representing measurements of the system state when the system is normal. This is normally obtained by making many measurements on a system when it is known to be in a normal state and using those measurements as a training data set. For example in the case of a jet engine the jet engine would be run for several hours and the measurements obtained from the various sensors, such as vibration sensors, stored and used as a training data set if the engine run has been normal. In the case of a jet engine the measurements can be the vibration amplitude at different frequencies corresponding to the different shafts in the engine. In the case of a human health monitor the different measurements could be the blood pressure, heart rate, oxygen saturation, breathing rate, body temperature, etc. Typically such measurements are produced on a continual basis and so at each sampling time point the value of each of the parameters being measured is used as one component of a multivariate vector defining a point in the data space. The data space has a number of dimensions corresponding to the number of different parameters being measured.

Once the training data set has been obtained, in step 601 a model of the data is constructed by fitting an analytic function, such as a Gaussian mixture model, to the set of training data using one of the well-known methods such as the Expectation-Maximisation algorithm. In accordance with this embodiment of the invention the model is then used to find the distribution of extreme values in a numerical manner. Thus in step 602 a set of $m$ (e.g. 100) data points are drawn from the model (these 100 points thus forming an artificial representation of data from the system) and in step 603 the data point from that set having the lowest probability density value (determined with respect to the model describing the distribution of normal data) is defined as the extremum for that set. Figure 6 illustrates on the right the steps applied to a simply univariate bimodal distribution in which is can be seen that of the five illustrated synthetically generated points $x_1$, $x_2$, $x_3$, $x_4$, $x_m$ it is $x_3$ which has the lowest probability density value and so would be selected as the extremum for that data set.
In step 604 it is checked whether the required number of extrema have been obtained and, if not, steps 602 and 603 are repeated until \( N \) (e.g. \( 10^6 \)) extrema have been obtained. For each of the extrema the probability density value \( p(x) \) is noted and to the right of step 604 a typical distribution of such probability density values (i.e. a histogram of the probability density values) is illustrated. As discussed above the majority of the probability density values are close to zero resulting in the histogram being skewed towards zero.

Therefore in this embodiment in step 605 a psi transform is applied to the probability density value \( p(x) \) of each of the \( N \) extrema and a distribution of these psi values is then constructed. In this embodiment the transform is defined as follows:

\[
\Psi[p(x)] = \begin{cases} 
-2 \ln p(x) - d \ln 2\pi \left( x \right)^{d/2} & \text{if } p(x) < (2\pi)^{-d/2} \\
0 & \text{if } p(x) \geq (2\pi)^{-d/2}
\end{cases}
\]

where \( d \) is the dimensionality of the model, and \( \ln \) is the natural logarithm function.

It can be noted that if the model were a unimodal Gaussian distribution then this transform would replicate the Gumbel distribution which is the correct EVD for a Gaussian distribution. Thus the psi transform maps the multidimensional probability density values \( p(x) \) into a single dimensional space where the EVD takes the form of the Gumbel distribution. To the right of steps 605 and 606 the form of the distribution of psi values is shown (grey) together with a Gumbel function fitted to it (black line) as indicated in step 606.

In order to set a threshold for use in novelty detection the cumulative density function \( P^c \) (which is univariate in psi-space) is equated to some probability mass, e.g. 0.99. This corresponds to setting the threshold such that the integral of the fitted Gumbel distribution from minus infinity up to the threshold equals the desired value, e.g. 0.99. The psi value of this threshold can then be converted back to a probability density value \( p(x) \) by reversing the psi transform above.

The meaning of this threshold is then that if, in an actual data set of \( m \) data is observed from the system, the probability density value of whichever data point is the extremum of that set is lower than the threshold, then that data point has less than the threshold chance of being an extremum from a normal data set. In such a
case, where the extremum is below the threshold, that extremum can therefore be classified as representing an abnormal state of the system.

Figure 8 illustrates actual histograms of the extreme values for the trimodal probability distribution of Figure 5. Figure 8(a) illustrates the distribution of probability density values of the $10^6$ extrema generated from Figure 5 with $m = 100$. Each of these $10^6$ probability density values is then psi transformed and the distribution of psi values is shown in Figure 8(b) with the corresponding (Maximum Likelihood Estimate) fitted Gumbel distribution shown in black and a novelty threshold at $P^e = -0.99$ as a dashed line. In the example of Figure 8(b) the threshold is at a psi value of about 4.6 corresponding to a probability density value of $p(x) = 0.0004$. This is clearly a very low probability density value and in a heavily-populated part of the highly skewed histogram of Figure 8(a), showing the value of setting the threshold on the psi distribution. The peak psi value of about 3.3 corresponds to a probability density value of $p(x) = 0.0007$, again very close to the left-hand end of the skewed histogram of Figure 8(a).

The method above thus can be regarded as defining a probability contour in the data space (e.g. of Figure 5) which describes where the most extreme of $m$ normal samples generated from the model will lie, to some probability (e.g. 0.99). Figure 5 illustrates in the darker lines such contours corresponding to the threshold set at increasing probability levels. Figure 9 illustrates another example using the method for a trivariate, trimodal distribution, where contours on the probability density function $p(x)$ are shown which correspond to thresholds set in the psi-space at increasing probabilities $P^e(x)$. Figure 9(b) illustrates the histogram of psi transformed extrema (shown in grey) from a different, 6-dimensional model and an MLE Gumbel fitted to these transformed extrema (shown in black). It should be noted that although the novelty threshold is set at a value which corresponds to a contour on the original probability density function which models the distribution of normal data, it is not heuristic. The threshold is set such that generating $m$ samples will result in an extremum exceeding the threshold with a probability of $1 - P^e = 1 - 0.99 = 0.01$; that is, the novelty threshold has a valid probabilistic interpretation provided by extreme value theory, which existing work cannot provide. This allows the false-positive alarm rate to be quantified, which existing methods cannot (because their estimates of the EVD
probabilities are incorrect for multivariate and/or multimodal data). Further, the
method is portable in that it can be applied to models of different systems (e.g.
different engines, different types of engines, different patients) and to models with
differing dimensionality and modality. In each case the threshold set at $P = 0.99$
on the psi distribution will result in the threshold being set at a different
probability density value $p(x)$ on the original data model. Further, as new normal
data is added to the training set and the model retrained, the position of the
contour on the model will change even thought the threshold in the psi space is
unchanged at $P = 0.99$.

Figure 7 illustrates how the threshold set above can be used in a system
monitor. The system monitor will typically receive input from multiple system
sensors illustrated as $s_1$ to $s_i$. Thus in this case the data space is \textit{i}-dimensional. In
step 702 the sensor readings are assembled into a time-series of state
measurements $x = (s_1, s_2, \ldots, s_i)$ each having as components a measurement from
each sensor. In step 703 a rolling window is defined in the time-series containing
$m$ (e.g. 100) consecutive state measurements $x$. Then in step 704 each of the $m$
measurements is compared to the model to read off its probability density value
$p(x)$ and in step 705 the one with the lowest probability is selected as being the
extremum for that set.

There are then two possibilities. One possibility is that the probability
density value $p(x)$ of the extremum selected in step 605 is transformed using the
psi transform and then compared to the threshold defined in psi space as illustrated
in steps 706b and 707b. Alternatively if the threshold in psi space has been
transformed back into a probability density value $p_{\psi}(x)$ corresponding to the
threshold, then in step 706a the probability density value of the extremum from
step 705 can simply be compared to the threshold in probability space in step
706a. Based on whether the probability density value or psi value of the
extremum exceeds the threshold the system state is classified as being normal or
abnormal in steps 707a or b.

Depending on the application an alarm can be generated if the system state
is classified as abnormal, or it may be preferable to wait to determine if the
abnormal state persists for a predetermined number of samples or a predetermined
time, or whether a predetermined number of abnormal classifications occur within
a certain time.
In step 708 the rolling window is moved (e.g. by one sample along in time) and steps 704 to 707 are repeated.

Figures 10 (a) and (b) illustrate the advantages of the invention in an application to gas turbine engine monitoring. In each figure, which relate to different engine runs, the first section of data (which plots a fused “health score” indicating the condition of the engine against time) was used as a training set. The subsequent data shown as a lighter line is the data before traditional techniques detected a fault, and the part of the data shown by a heavy line at the end of each run is the data after traditional techniques detected it. The threshold set according to the multivariate EVT method above is illustrated as a horizontal dotted line and it can be seen that in both cases the “health score” crossed the multivariate-EVT based threshold (and thus would create an alert) long before the traditional techniques detected the fault.

The method can also be applied to manufacturing by, for example, monitoring manufacturing process and looking for abnormality in the parameters acquired from that manufacturing process. An example would be a manufacturing process involving drilling or machining where vibration and acoustics together with the power consumed by the machine can be monitored in which case the system looks for abnormal combinations of vibration acoustics and power consumption which could be indicative of a fault in the manufacturing process.
CLAIMS

1. A method of setting a normality threshold for use in a system monitor to allow automatic classification of measured states of the system as normal or abnormal, comprising:
   constructing a statistical model of the normal states of the system giving the probability density value for each measured state when the system is normal;
   defining as an extreme value the least probable of a set of measurements of the state of the system, the set comprising a predetermined number \( m \) of measurements;
   finding from the model the distribution of the probability density values of the so-defined extreme values;
   setting as a threshold a probability density value on said distribution which divides said distribution in predetermined proportions whereby extreme values having a probability density value higher than said threshold are classified as normal and extreme values having a probability density value lower than said threshold are classified as abnormal.

2. A method according to claim 1 wherein the statistical model is multimodal.

3. A method according to claim 1 or 2 wherein the statistical model is multivariate, each of said measurements comprising a plurality of component variables each corresponding to the output of a sensor on the system.

4. A method according to claim 1, 2 or 3 wherein the distribution of the probability density values of the extreme values is obtained by generating from the model a number \( N \) of sets of \( m \) values representing measurements of the state of the system, taking the extreme value from each set, and constructing the distribution of the probability density values of the so-obtained \( N \) extreme values.

5. A method according to any one of the preceding claims wherein the step of setting the threshold as a probability density value on said distribution comprises transforming the probability density value of each of the extreme values and constructing the distribution with the transformed values, wherein said transform expands the low probability part of the distribution.
6. A method according to claim 5 wherein the transform is logarithmic.

7. A method according to claim 5 or 6 wherein the distribution of transformed values approximates an extreme value distribution.

8. A method according to any one of the preceding claims, further comprising the step of setting a second threshold on said distribution for classifying extreme values into three classes.

9. A system monitor for monitoring the state of a system by reference to a threshold set in accordance with the method of any one of the preceding claims, the monitor storing said model and being adapted to acquire a number $m$ of measurements of the state of the system, to compare each measurement to the model to find the least probable of the $m$ measurements, and to compare the least probable of the $m$ measurements to the threshold to classify the system state as normal or abnormal.

10. A system monitor according to claim 9, wherein the comparison with the threshold is performed after transforming the least probable measurement using the transform of claim 5, 6 or 7.

11. A system monitor according to claim 9 or 10 adapted to acquire measurements of said system state continually and to use as said $m$ measurements a rolling window of $m$ successive measurements.

12. A system monitor according to claim 9, 10 or 11 further adapted to store measurements of the system state classified as normal for use in retraining the statistical model.
Fig. 1.
Fig. 6.

Construct model by fitting function, e.g.
GMM, to set of training data

Draw a set of \( m \) (e.g. 100) data
points from the model

Of the \( m \) points select the one with
the lowest probability \( p(x) \) and define
it as the extreme value for that set

\( N \) (e.g. \( 10^6 \))

extreme values

Yes

No

Apply psi transform to the probability
value \( p_\psi(x) \) of each of the \( N \) extreme values
and construct distribution of psi values

Fit Gumbel curve to psi distribution

Define threshold \( \psi_{th} \) on distribution
of psi values such that threshold is
at, e.g. 0.99, of the cumulative density

Optionally convert psi threshold value back
to a probability value \( p_{th} \) using reverse
psi transform
Fig. 7.

1. Receive input from multiple sensors on system
2. Assemble into time series of state measurements $x$ each having as components a measurement value from each sensor $x = (s_1, s_2, s_3, ..., s_i)$
3. Define a rolling window containing $m$ consecutive state measurements $x$
4. Compare the $m$ measurements to the model to read off the probability $p(x)$ of each measurement $x$
5. Choose the one with the lowest probability and note its probability
6. Compare its probability to the threshold $p_{th}$ of step 608
   - If $p < p_{th}$, system state normal, so store measurements for addition to training set
   - If $p > p_{th}$, convert its probability using the psi transform and compare to the threshold of step 607
     - If $\Psi < \psi_{th}$, system state abnormal, so consider alarming
     - If $\Psi > \psi_{th}$, system state normal, so store measurements for addition to training set
7. Move rolling window along to analyze next set of $m$ state measurements
Fig. 8(a)

Fig. 8(b)
**Fig. 10(a)**

Engine 1, performance model

**Fig. 10(b)**

Engine 2, performance model
**INTERNATIONAL SEARCH REPORT**

**A. CLASSIFICATION OF SUBJECT MATTER**

**INV.** G05B 23/02  
**ADD.**

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)
G 05 B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic database consulted during the international search (name of database and, where practical, search terms used)
EPO-Internal

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>page 1, left-hand column, paragraph 1 - right-hand column, paragraph 4 page 4, left-hand column, paragraph 1 - right-hand column, paragraph 2</td>
<td>2, 3, 5-7, 10</td>
</tr>
</tbody>
</table>

[X] Further special categories of cited documents:

- [ ] Special categories of cited documents:
  - [ ] document defining the general state of the art which is not considered to be of particular relevance
  - [ ] earlier document but published on or after the international filing date
  - [ ] document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
  - [ ] document referring to an oral disclosure, use, exhibition or other means
  - [ ] document published prior to the international filing date but later than the priority date claimed

- [ ] "*" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- [ ] "*" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- [ ] "*" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- [ ] "*" document member of the same patent family

Date of the actual completion of the international search: 17 November 2010

Date of mailing of the international search report: 24/11/2010

Name and mailing address of the ISA/
European Patent Office, P. B. 5018 Patentlaan 2  
NL - 2330 HN Rijswijk  
Tel. (+31-70) 540-2040, Fax: (+31-70) 340-3016

Authorized officer: Kelperis, K

Form PCT/ISA/210 (second sheet) (April 2005)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>CHAN ET AL: &quot;Point and interval estimation for extreme-value regression model under Type-II censoring&quot;, COMPUTATIONAL STATISTICS AND DATA ANALYSIS, NORTH-HOLLAND, AMSTERDAM, NL, vol. 52, no. 8, 5 February 2008 (2008-02-05), pages 4040-4058, XP022575632, ISSN: 0167-9473 page 4041, line 1 - line 3</td>
<td>5-7,10</td>
</tr>
</tbody>
</table>