



HU000033477T2

(19) **HU****MAGYARORSZÁG****Szellemi Tulajdon Nemzeti Hivatala**(11) Lajstromszám: **E 033 477**(13) **T2**

EURÓPAI SZABADALOM SZÖVEGÉNEK FORDÍTÁSA

(21) Magyar ügyszám: **E 03 781728**(22) A bejelentés napja: **2003. 11. 03.**(51) Int. Cl.: **G05B 17/02** (2006.01)**G05B 23/02** (2006.01)**G05B 13/02** (2006.01)

(96) Az európai bejelentés bejelentési száma:

EP 20030781728

(97) Az európai bejelentés közzétételi adatai:

EP 1579288 A2 **2004. 05. 21.**

(86) A nemzetközi (PCT) bejelentési szám:

PCT/US 03/035001

(97) Az európai szabadalom megadásának meghirdetési adatai:

EP 1579288 B1 **2017. 03. 15.**

(87) A nemzetközi közzétételi szám:

WO 04042531

(30) Elsőbbségi adatok:

423476 P**2002. 11. 04.****US**

(73) Jogosult(ak):

**GE Intelligent Platforms, Inc., Charlottesville,
VA 22911-8300 (US)**

(72) Feltaláló(k):

**WEGERICH, Stephan, W., Glendale Heights, IL 60139
(US)****XU, Xaio, Naperville, IL 60540 (US)**

(74) Képviselő:

SBGK Szabadalmi Ügyvivői Iroda, Budapest

(54)

Rendszerállapot-figyelés visszatérő helyi tanulógép segítségével

Az európai szabadalom ellen, megadásának az Európai Szabadalmi Közlönyben való meghirdetésétől számított kilenc hónapon belül, felszólalást lehet benyújtani az Európai Szabadalmi Hivatalnál. (Európai Szabadalmi Egyezmény 99. cikk(1))

A fordítást a szabadalmas az 1995. évi XXXIII. törvény 84/H. §-a szerint nyújtotta be. A fordítás tartalmi helyességét a Szellemi Tulajdon Nemzeti Hivatala nem vizsgálta.

(19)



(11)

EP 1 579 288 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:
15.03.2017 Bulletin 2017/11

(51) Int Cl.:
G05B 17/02 (2006.01) G05B 23/02 (2006.01)
G05B 13/02 (2006.01)

(21) Application number: **03781728.5**

(86) International application number:
PCT/US2003/035001

(22) Date of filing: **03.11.2003**

(87) International publication number:
WO 2004/042531 (21.05.2004 Gazette 2004/21)

(54) **SYSTEM STATE MONITORING USING RECURRENT LOCAL LEARNING MACHINE**

SYSTEMZUSTANDSÜBERWACHUNG DURCH EINE ITERATIVE LOKALELERNMASCHINE

SURVEILLANCE D'ETAT D'UN SYSTEME AU MOYEN D'UNE MACHINE A APPRENTISSAGE
LOCAL RECURRENT

(84) Designated Contracting States:
**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IT LI LU MC NL PT RO SE SI SK TR**

(30) Priority: **04.11.2002 US 423476 P**

(43) Date of publication of application:
28.09.2005 Bulletin 2005/39

(73) Proprietor: **GE Intelligent Platforms, Inc.**
Charlottesville, VA 22911-8300 (US)

(72) Inventors:
• **WEGERICH, Stephan, W.**
Glendale Heights, IL 60139 (US)
• **XU, Xaio**
Naperville, IL 60540 (US)

(74) Representative: **Illingworth-Law, William**
Illingworth
GPO Europe
GE International Inc.
The Ark
201 Talgarth Road
Hammersmith
London W6 8BJ (GB)

(56) References cited:
WO-A1-02/086726 WO-A2-02/35299
WO-A2-02/057856 US-A1- 2002 128 731
US-A1- 2003 060 808 US-A1- 2004 019 406
US-A1- 2005 021 187 US-A1- 2005 027 400

EP 1 579 288 B1

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

DescriptionFIELD OF THE INVENTION

[0001] The present invention relates to a method and system for modeling a process, piece of equipment or complex interrelated system. More particularly, it relates to equipment condition and health monitoring and process performance monitoring for early fault and deviation warning, based on recurrent non-parametric modeling and state estimation using exemplary data.

[0002] US 2002/0128731 A1 discloses an improved empirical model-based surveillance or control system for monitoring or controlling a process or machine, which provides identification of transitions between operational states. Empirical model-based estimates generated in response to receiving actual operational parameters are compared using a global similarity operator to the actual parameters to indicate whether the process or machine is operating in a stable state, or is in transition from one state to another.

[0003] WO-A-02/35299 discloses a method for estimating and reducing uncertainties in process measurements. A reference matrix contains valid measurements characterizing operation of a multivariate process. Modeling parameters of the reference matrix are derived. The final model parameters, balanced with respect to measuring and modeling uncertainties, are applied to model a new set of measurements. If the new set has no faults then all modeled values and modeling uncertainties can be used to control the process. If the new set has only one fault then the modeled value and modeling uncertainty of the faulted measurement plus the measured values and measuring uncertainties of the unfaulted measurements can be used to control the process while repair procedures are implemented for the identified fault. If the new set has more than one fault then the process should be shut down, and repair procedures should be implemented for all identified faults.

SUMMARY OF THE INVENTION

[0004] According to the present invention, there is provided a method of system state monitoring as set out in claim 1.

[0005] The present invention also provides an apparatus for system state monitoring as set out in claim 20.

[0006] Optional features are set out in the dependent claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007]

FIG. 1 is process flowchart for equipment health monitoring using the model of an embodiment of the invention;

FIG. 2 shows a diagram for windowed adaptation in a model according to an embodiment of the invention; and

FIG. 3 shows a block diagram of a system according to an embodiment of the invention for monitoring equipment health.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0008] In the following, there is described an empirical, non-parametric multivariate modeling method and apparatus for state modeling of a complex system such as equipment, processes or the like, and provides equipment health monitoring, process performance optimization, and state categorization. In a machine, process or other complex system that can be characterized by data from sensors or other measurements, the modeling method comprises first acquiring reference data observations from the sensors or measurements representative of the machine, process or system, and then computing the model from a combination of the representative data with a current observation from the same sensors or measurements. The model is recomputed with each new observation of the modeled system. The output of the model is an estimate of at least one sensor, measurement or other classification or qualification parameter that characterizes the state of the modeled system.

[0009] Accordingly, for equipment health monitoring, the model provides estimates for one or more sensors on the equipment, which can be compared to the actually measured values of those sensors to detect a deviation indicative of an incipient failure mode. Alternatively, the model can estimate a performance parameter that can be used to optimize a process, by indicating how that performance parameter changes with controllable changes in inputs to the process. The estimate provided by the model can even be a logical or qualitative output designating the state of the modeled system, as in a quality control application or a disease classification medical application.

[0010] Advantageously, the modeling method employs similarity-based modeling, wherein the model estimate is com-

prised of a weighted composite of the most similar observations in the reference data to the current observation. The model employs matrix regularization to control against ill-conditioned outputs, e. g. , estimates that blow up to enormous or unrealistic values, which are useless in the applications of the model. For applications in which the size of the reference data is large, or the sampling time of observations (and thus the need for estimates from the model) is fast, the current observation can be indexed into a subset domain or fuzzy subset of the reference data using a comparison of the current observation with a reference vector, for quicker computation of the estimate.

[0011] The described apparatus comprises a memory for storing the reference data; an input means such as a networked data bus or analog-to-digital converter connected directly to sensors, for receiving current observations; a processing unit for computing the model estimate responsive to the receipt of the current observation; and output means such as a graphic user interface for reporting the results of the modeling. The described apparatus further comprises a software module for outputting the model estimates to other software modules for taking action based on the estimates.

[0012] The modeling method of an embodiment the present invention can be used in equipment condition monitoring where the model estimates sensor readings in response to current readings, and the estimates and actual readings are compared to detect and diagnose any equipment health issues. The modeling method can also be extended for use in classification of a system characterized by observed variables or features, where the output of the model can be an estimate of a parameter used for classifying. Generally, an embodiment of the invention will be described with respect to equipment health monitoring.

[0013] A reference data set of observations from sensors or other variables of the modeled system comprises sufficient numbers of observations to characterize the modeled system through all of the dynamics of that system that are anticipated for purposes of the modeling. For example, in the case of monitoring a gas combustion turbine for equipment health and detection of incipient failures, it may be sufficient to obtain 500 to 10,000 observations from a set of 20-80 temperature, flow and pressure sensors on the turbine, throughout the operational range of the turbine, and throughout environmental changes (seasons) if the turbine is located outside. As another example of equipment health monitoring, 10-20 sensors on a jet engine can be used to obtain 50-100 observations of take-off or cruise-mode operation to provide adequate modeling. In the event that all such data is not available up front (for example, seasonally affected operation), the reference set can be augmented with current observations.

[0014] Observations may comprise both real-world sensor data and other types of measurements. Such measurements can include statistical data, such as network traffic statistics; demographic information; or biological cell counts, to name a few. Qualitative measurements can also be used, such as sampled opinions, subjective ratings, etc. All that is required of the input types used is that they are related in some fashion through the physics, mechanics, or dynamics of the system being modeled (or are suspected to be so), and in aggregate represent "states" the modeled system may take on.

[0015] With reference to FIG. 1, in an embodiment of the invention, the reference set of observations is formed into a matrix, designated H for purposes hereof, in a step 102 typically with each column of the matrix representing an observation, and each row representing values from a single sensor or measurement. The ordering of the columns (i. e. , observations) in the matrix is not important, and there is no element of causality or time progression inherent in the modeling method. The ordering of the rows is also not important, only that the rows are maintained in their correspondence to sensors throughout the modeling process, and readings from only one sensor appear on a given row. This step 102 occurs as part of the setup of the modeling system, and is not necessarily repeated during online operation.

[0016] After assembling a sufficiently characterizing set H of reference data observations for the modeled system, modeling can be carried out. Modeling results in the generation of estimates in response to acquiring or inputting a real-time or current or test observation, as shown in step 107, which estimates can be estimates of sensors or non-sensor parameters of the modeled system, or estimates of classifications or qualifications distinctive of the state of the system. These estimates can be used for a variety of useful modeling purposes as described below.

[0017] The generation of estimates according to the modeling method of an embodiment comprises two major steps after acquiring the input in 107. In the first step 118, the current observation is compared to the reference data H to determine a subset of reference observations from H having a particular relationship or affinity with the current observation, with which to constitute a smaller matrix, designated D for purposes hereof. In the second step 121, the D matrix is used to compute an estimate of at least one output parameter characteristic of the modeled system based on the current observation. Accordingly, it may be understood that the model estimate Y_{est} is a function of the current input observation Y_{in} and the current matrix D, derived from H:

$$\bar{Y}_{est} = \bar{D} \bullet \bar{W} \quad (1)$$

$$\bar{W} = \frac{\hat{W}}{\left(\sum_{j=1}^N \hat{W}(j) \right)} \quad (2)$$

$$\hat{W} = (\bar{D}^T \otimes \bar{D})^{-1} \bullet (\bar{D}^T \otimes \bar{Y}_{in}) \quad (3)$$

$$\bar{D} = F(\bar{H}, \bar{Y}_{in}) \quad (4)$$

where the vector $\bar{Y}_{est}$ of estimated values for the sensors is equal to the contributions from each of the snapshots of contemporaneous sensor values arranged to comprise matrix $\bar{D}$. These contributions are determined by weight vector $\bar{W}$. The multiplication operation is the standard matrix/vector multiplication operator, or inner product. The similarity operator is symbolized in Equation 3, above, as the circle with the "X" disposed therein. Both the similarity operation of Equation 3 and the determination F of membership comprising $\bar{D}$ from $\bar{H}$ and the input observation $\bar{Y}_{in}$ are discussed below.

[0018] As stated above, the symbol $\otimes$ represents the "similarity" operator, and could potentially be chosen from a variety of operators. In the context of the embodiment, this symbol should not to be confused with the normal meaning of designation of $\otimes$, which is something else. In other words, for purposes of the present embodiment the meaning of $\otimes$ is that of a "similarity" operation.

[0019] The similarity operator, $\otimes$, works much as regular matrix multiplication operations, on a row-to-column basis, and results in a matrix having as many rows as the first operand and as many columns as the second operand. The similarity operation yields a scalar value for each combination of a row from the first operand and column from the second operand. One similarity operation that has been described above involves taking the ratio of corresponding elements of a row vector from the first operand and a column vector of the second operand, and inverting ratios greater than one, and averaging all the ratios, which for normalized and positive elements always yields a row/column similarity value between zero (very different) and one (identical). Hence, if the values are identical, the similarity is equal to one, and if the values are grossly unequal, the similarity approaches zero.

[0020] Another example of a similarity operator that can be used determines an elemental similarity between two corresponding elements of two observation vectors or snapshots, by subtracting from one a quantity with the absolute difference of the two elements in the numerator, and the expected range for the elements in the denominator. The expected range can be determined, for example, by the difference of the maximum and minimum values for that element to be found across all the data of the reference library $\bar{H}$. The vector similarity is then determined by averaging the elemental similarities.

[0021] In yet another similarity operator that can be used in an embodiment of the present invention, the vector similarity of two observation vectors is equal to the inverse of the quantity of one plus the magnitude Euclidean distance between the two vectors in n -dimensional space, where n is the number of elements in each observation, that is, the number of sensors being observed. Thus, the similarity reaches a highest value of one when the vectors are identical and are separated by zero distance, and diminishes as the vectors become increasingly distant (different).

[0022] Other similarity operators are known or may become known to those skilled in the art, and can be employed in the embodiments of the present invention as described herein. The recitation of the above operators is exemplary and not meant to limit the scope of the invention. In general, the following guidelines help to define a similarity operator for use in an embodiment of the invention as in equation 3 above and elsewhere described herein, but are not meant to limit the scope of the invention:

1. Similarity is a scalar range, bounded at each end.
2. The similarity of two identical inputs is the value of one of the bounded ends.
3. The absolute value of the similarity increases as the two inputs approach being identical.

[0023] Accordingly, for example, an effective similarity operator for use in an embodiment the present invention can generate a similarity of ten (10) when the inputs are identical, and a similarity that diminishes toward zero as the inputs become more different. Alternatively, a bias or translation can be used, so that the similarity is 12 for identical inputs, and diminishes toward 2 as the inputs become more different. Further, a scaling can be used, so that the similarity is 100 for identical inputs, and diminishes toward zero with increasing difference. Moreover, the scaling factor can also be a negative number, so that the similarity for identical inputs is -100 and approaches zero from the negative side with

increasing difference of the inputs. The similarity can be rendered for the elements of two vectors being compared, and summed or otherwise statistically combined to yield an overall vector-to-vector similarity, or the similarity operator can operate on the vectors themselves (as in Euclidean distance).

[0024] Significantly, an embodiment of the present invention can be used for monitoring variables in an autoassociative mode or an inferential mode. In the autoassociative mode, model estimates are made of variables that also comprise inputs to the model. In the inferential mode, model estimates are made of variables that are not present in the input to the model. In the inferential mode, Equation 1 above becomes:

$$\bar{Y}_{est} = \bar{D}_{out} \bullet \bar{W} \quad (5)$$

$$\underline{\hat{W}} = (\bar{D}_{in}^T \otimes \bar{D}_{in})^{-1} \bullet (\bar{D}_{in}^T \otimes \bar{Y}_{in}) \quad (6)$$

where the D matrix has been separated into two matrices D_{in} , and D_{out} , according to which rows are inputs and which rows are outputs, but column (observation) correspondence is maintained.

[0025] A key aspect of one embodiment is that D is determined recurrently with each new input observation, from the superset of available learned observations H characterizing the dynamic behavior of the modeled system. In doing so, sufficiently relevant exemplars or learned observations are used to characterize the modeled behavior in the neighborhood of the current observation, but the model avoids both undue overfitting as well as impractical computational time. The determination of membership in D according to an embodiment is accomplished by relating the current input observation to observations in H, and when there is a sufficient relationship, that learned observation from H is included in D, otherwise it is not included in D for purposes of processing the current input observation.

[0026] According to one embodiment of the invention, the input observation is compared to exemplars in H using the similarity operation to render a similarity score for each such comparison, called "global similarity" for purposes hereof. If the resulting global similarity is above a certain threshold, or is one of the x highest such global similarities across all exemplars in H, the exemplar or learned observation is included in D. For a similarity operator rendering similarity scores between zero (different) and one (identical), a typical threshold may be 0.90 or above, by way of example. However, the choice of threshold will depend on the nature of the application, and especially on the number of exemplars in the set H. In the event that the highest x similarities are used to determine membership in D, it is not uncommon to use somewhere in the range of 5 to 50 exemplars in D, even when selecting from a set H that may have an enormous number of exemplars, such as 100,000. A hybrid of threshold and count can be used to determine membership of D, for example by using a threshold for inclusion, but requiring that D contain no less than 5 exemplars and no more than 25.

[0027] Importantly, not all elements of the observations need be used for determining global similarity. Certain variables or sensors may be deemed more dominant in the physics of the monitored system, and may be the basis for determining membership of D, by performing the global similarity calculation only on a subvector comprised the those elements from each of the current observation and each learned observation. By way of example, in an inferential model, in which the input observation has ten (10) sensor values, and the output of the model is an estimate for five (5) additional sensor values not among the inputs, the global similarity may be computed using a subvector of the input vector and the learned observations comprising only the 1st, 2nd, 5th, and 7th sensor values, even though the estimate of the 5 outputs will be performed using all 10 inputs. Selection of which input sensors to rely on in determining global similarity for constituting membership in the D matrix can be made using domain knowledge, or can be determined from the least root mean square error between actual values and estimates produced by the model when tested against a set of test data (not part of the set H) characterizing normal system behavior, among other methods.

[0028] In an alternative to the use of the global similarity, membership in D can be determined by examining one or more variables at an elemental level, and including exemplars from H that have elemental values fitting a range or fitting some other criterion for one or more elements. For example, in the fanciful 10-input, 5-output model mentioned above, D might be comprised by exemplars from H with the 5 closest values for the 1st sensor to the same sensor value for the current observation, the 5 closest value for the 2nd sensor, the 5 closest for the 5th sensor, and the 5 closest for the 7th sensor, such that D has at most 20 vectors from H (though possibly less if some repeat). Note that this is different from the global similarity in that a learned observation may be included in D solely because it has a closely matching value on an *m*th sensor, irrespective of the rest of its sensor values.

[0029] In a preferred embodiment, the examination at an elemental level for membership in D can be performed on variables that do not in fact comprise inputs to the model, but are nonetheless sensor values or measurements available from the system with each observation of the other sensors in the model. A particularly important circumstance when this can be useful is with ambient condition variables, such as ambient air temperature, or ambient barometric pressure.

Such ambient variables - while not necessarily serving as inputs to any given model - may be proxies for overall conditions that impact the interrelationships of the other sensor values that are in the model. Consequently, the use of ambient variables for determining membership in D of exemplars selected from H can be a good way of providing a D matrix with relevant exemplars to seasonal variation. For example, in an application for monitoring the health of the engine of a locomotive, a variety of engine parameters (e.g., fuel flow, exhaust gas temperature, turbo pressure, etc.) may be used to model the behavior of the engine, and ambient temperature may be used as an ambient variable for selecting observations from H for D. The ambient temperature is a proxy for the weather conditions that affect how all the other parameters may interrelate at any given temperature. H ideally contains historic data of normal performance of the engine, for all temperature ranges, from below freezing in winter, to sweltering temperatures of a desert summer. Exemplars from H (coming from all across this temperature range) may be selected for a particular D matrix if the ambient temperature of the exemplar is one of the x closest values to the ambient temperature of the current input reading. Note that in computing the model estimates per equations 1-4 above, ambient temperature would not necessarily be an input or an output.

[0030] In a preferred embodiment, a hybrid of the ambient variable data selection and one of either global similarity or elemental test for inclusion, is used in combination. Thus, for example, ambient temperature may be used to select from a superset of H having 100,000 learned observations covering temperatures from well below freezing to over 100 degrees Fahrenheit, a subset of 4000 observations to comprise an intermediate set H', which 4000 observations are those within +/- 5 degrees from the current ambient temperature. This intermediate subset H' can then be used without alteration for several hours worth of input data (during which ambient temperature has not shifted significantly), to repeatedly generate a D matrix of, say, 30 vectors selected from the 4000 by means of global similarity for each input observation. In this way, the current observation can be closely modeled based on the performance characteristics of the system at that moment, within the framework of a set of data selected to match the ambient conditions. This cuts down on computational time (avoiding processing all 100,000 observations in H), avoids overfitting, and provides high fidelity modeling tuned to the conditions in which the monitored equipment is encountered.

[0031] Yet another way of determining membership in D involves a modified use of global similarity, for improving the computational speed of this step. Accordingly, a reference vector, which may be one of the exemplars in H, is first compared to all the learned observations to generate a global similarity for each comparison. This can be done before on-line monitoring is commenced, and need be done only once, up front. Then, during monitoring the current observation is compared to that reference vector using global similarity, instead of comparing the current observation to all learned observations in H. The resulting global similarity score is then compared to the pre-calculated global similarities of the reference vector vis-à-vis the learned observations in H, and the closest n scores indicate the learned observations to include in D; or alternatively, those global similarities within certain limits around the global similarity of the current observation, indicate the learned observations to include in D.

[0032] According to yet another way to determine membership in D, the reference set of learned observations in H are grouped using a clustering method into a finite number of clusters. In real-time, the current observation is then analyzed to determine which cluster it belongs to, and the learned observations in that cluster are then drawn from to constitute the D matrix. All of the learned observations in the cluster can be included, or a sampled subset of them can be included in order to keep the size of D manageable if the cluster contains too many vectors. The subset can be sampled randomly, or can be sampled from using a "characterized" sampling method as disclosed later herein.

[0033] To select the clusters for the clustering algorithm, seed vectors can be selected from H. A vector becomes a seed for a cluster based on containing a maximum or minimum value for a sensor across all the values of that sensor in H. One clustering technique that can be used is fuzzy C means clustering, which was derived from Hard C-Means (HCM). Accordingly, vectors in H can have partial membership in more than one cluster. Fuzzy C-Means (FCM) clustering minimizes the objective function:

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m d^2(x_k, v_i) \quad (7)$$

where $X = (x_1, x_2, \dots, x_n)$ is n data sample vectors (the learned observations in H), U is a partition of X in c part, $V = (v_1, v_2, \dots, v_c)$ are cluster centers in R^V (seeded as mentioned above from actual observations in H), u_{ik} is referred to as the grade of membership of x_k to the cluster i , in this case the member of u_{ik} is 0 or 1, and $d^2(x_k, v_i)$ is an inner product induced norm on R^V :

$$d(x_k, v_i) = \sqrt{(x_k - v_i)^T (x_k - v_i)} \quad (8)$$

The problem is to determine the appropriate membership u_{ik} , which is done through iterative determination to convergence of:

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^n u_{ij}^m} \quad (9)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (10)$$

where c is the number of clusters. The u_{ik} are randomly selected initially subject to the constraints:

$$0 \leq u_{ik} \leq 1, \quad \text{for } 1 \leq i \leq c, 1 \leq k \leq n \quad (11)$$

$$0 < \sum_{k=1}^n u_{ik} < n, \quad \text{for } 1 \leq i \leq c \quad (12)$$

$$\sum_{i=1}^c u_{ik} = 1, \quad \text{for } 1 \leq k \leq n \quad (13)$$

During monitoring, the input observation is compared using global similarity, Euclidean distance, or the like, to the cluster centers v_i , to determine which cluster the input observation is most related to. The D matrix is then constituted from the identified cluster. A cluster is determined to be those vectors in H that have a fuzzy membership u_{ik} that is above a certain threshold, typically 0.70 (but dependent on the application and availability of data in H). Thus, a particular observation in H could belong to more than one cluster. The cluster in H matching the input observation can be used in its entirety for D, or can be selected from to comprise D, using any of the other methods described herein. Fuzzy c-means clustering can thus be used to reduce the number of vectors in H that need to be analyzed with some other method for inclusion in D, such as global similarity, as a computational savings.

[0034] An additional important aspect of one embodiment is adaptation of the model. Especially for equipment health monitoring (but also for other applications) the issue of keeping the model tuned with slow and acceptable changes to the underlying modeled system is critical for practical use. When monitoring equipment, graceful aging is assumed, and should not become a source of health alerts. Therefore the model must adapt through time to gradual aging and settling of the monitored equipment, and not generate results that suggest an actionable fault is being detected.

[0035] Adaptation can be accomplished in an embodiment in a number of ways. According to a first way, called for purposes herein "out-of-range" adaptation, certain of the monitored variables of the system are considered drivers or independent variables, and when they take on values outside of the ranges heretofore seen in the set H of exemplars, then the current observation is not alerted on, but rather is added to the set H, either by addition or by substitution. In this way, when a driver variable goes to a new high or a new low, the model incorporates the observation as part of normal modeled behavior, rather than generating an estimate that in all likelihood is different from the current observation. The drawback of this out-of-range adaptation is two-fold: (1) not all variables can be considered drivers and thus cause out-of-range adaptation and thus there is an application-specific art to determining which variables to use; and (2) there exists the possibility that an out-of-range event is in fact initial evidence of an incipient fault, and the model may now not as easily identify the fault. With regard to the first concern, ambient variables can usually make good candidates for out-of-range adaptation as a rule. For the second concern, a preferred embodiment of the invention does not permit n successive out-of-range adaptations, where n is typically in the range of 2 and up, depending

on the sampling rate of the data acquisition.

[0036] Usually, out-of-range adaptation is additive to the H matrix, rather than replacing exemplars in H. According to another kind of adaptation that can be employed in parallel with out-of-range adaptation, vectors are added to H that occur in a window of observations delayed by some offset from the current observation, and these additions replace the oldest exemplars in H. Thus H is a first-in, first-out stack, and is eventually turned over entirely with updated observations, thus tracking the graceful aging of the monitored equipment. The offset is required so that observations aren't learned that include developing faults, and the choice of delay size is largely a function of the application, the data sampling rate, and the nature of expected failures and how they manifest themselves.

[0037] Turning to FIG. 2, this method of moving window adaptation can be better understood in view of a timeline of sequential current observations being monitored. Monitoring begins at time step 210. A reference library H of learned observations 213 has been assembled from prior normal operation of the monitored equipment. The current real-time observation 220 is being monitored presently. A window of past observations 225 is drawn from to provide updated exemplars to reference library H 213, which may or may not employ a replacement scheme by which older exemplars are deleted from the library 213. The window 225 moves forward with the timeline 206, at some delay separation 230 from the current observation 220. If faults are detected in observations that enter the leading edge of window 225, there are two alternatives for avoiding adapting into the developing fault. First, the faulted observations themselves can be flagged to not be adapted on. Second, windowed adaptation can be turned off until the fault is resolved. Upon resolution of the fault, the window would be reinitiated starting with "normal" data beyond (in time) the fault resolution event.

[0038] The observations in window 225 can be sampled for addition to library 213, or can all be added. Methods for sampling a subset of observations to add to library 213 include random sampling; periodic sampling; and sampling, in which the set of observations in window 225 is mined for those observations that characterize the dynamics of operation throughout the window. For example, one way is to pick those observations which contain a highest value or a lowest value for any one of the sensors in the observations throughout the window, optionally augmented with observations having sensor values that cover the sensor range (as seen throughout the window) at equally spaced values (e. g., for a temperature range of 50-100 degrees, picking vectors at 60,70, 80 and 90, as well as the extremes of 50 and 100).

[0039] Turning to FIG. 3, the use of the modeling of an embodiment of the present invention is described in the context of a complete apparatus for performing equipment health monitoring. An H reference library 304 is stored in memory, typically permanent disk drive read/write memory, and comprises learned observations characterizing the anticipated operational dynamics of the monitored equipment in normal, non-faulted operation. Data acquired or supplied from sensors or other measurement systems on the equipment are provided for active monitoring to a set of preconditioning modules 307, including data cleaning, feature extraction and complex signal decomposition. Data cleaning includes filtering for spikes, smoothing with filters or splines, and other techniques known in the art. Feature extraction can include spectral feature extraction, and translation of analog data values into classes or other numeric symbols, as is known in the art. For sensors such as acoustics and vibration, complex signal decomposition is a form of feature extraction in which pseudo-sensors are provided from the spectral features of these complex signals, and can be FFT components as signals, or subbands.

[0040] The preconditioned data is then supplied to the D selector module 311 and the estimate generator 315. The D selector module 311 employs the techniques mentioned above to compare the (preconditioned) current observation to the exemplars in the library 304, to select a subset to comprise the D matrix. The estimate generator uses the D matrix and the current observation to generate an estimate for sensors describing equipment health according to Equations 1 through 4 above. Estimates are provided along with the current observation to a statistical testing module 320 which is described below. The purpose of the statistical testing module is to test the estimate in contrast to the actual current readings to detect incipience of faults in the equipment. The estimated sensor values or parameters are compared using a decision technique to the actual sensor values or parameters that were received from the monitored process or machine. Such a comparison has the purpose of providing an indication of a discrepancy between the actual values and the expected values that characterize the operational state of the process or machine. Such discrepancies are indicators of sensor failure, incipient process upset, drift from optimal process targets, incipient mechanical failure, and so on.

[0041] The estimates and current readings are also available to a diagnostics module 324, as are the results of the statistical testing module. The diagnostics module 324 can comprise a rules-based processor for detecting patterns of behavior characteristic of particular known failure modes, by mapping combinations of residuals, statistical test alerts, raw values and features of raw values to these known failure modes. This is described in greater detail below.

[0042] The results of both the statistical testing module 320 and the diagnostics module 324 are made available to a user interface module 330, which in a preferred embodiment is a web-based graphical interface which can be remotely located, and which displays both failure messages and confidence levels generated by the diagnostics module 324, and charts of residuals, statistical testing alerts, and raw values. Diagnostic results and statistical test results can also be made available through a software interface 335 to downstream software that may use the information, e.g., for scheduling maintenance actions and the like. The software interface 335 in a preferred embodiment comprises a messaging service that can either be polled or pushes information to subscribing systems, such as .NET services.

[0043] An adaptation module 340 employs the out-of-range adaptation and the windowed adaptation described above to update the library 304 as frequently as with every new current observation.

[0044] The statistical testing module can employ a number of tests for determining an alert condition on the current observation or sequence of recent observations. One test that can be used is a simple threshold on the residual, which is the difference between the estimate of a sensor value and the actual sensor value (or actual preconditioned sensor value) from the current observation. Alerts can be set for exceeding both a positive and/or a negative threshold on such a residual. The thresholds can be fixed (e.g., +/- 10 degrees) or can be set as a multiple of the standard deviation on a moving window of the past n residuals, or the like.

[0045] Another test or decision technique that can be employed is called a sequential probability ratio test (SPRT), and is described in the aforementioned U.S. Patent No. 5,764,509 to Gross et al. Broadly, for a sequence of estimates for a particular sensor, the test is capable of determining with preselected missed and false alarm rates whether the estimates and actuals are statistically the same or different, that is, belong to the same or to two different Gaussian distributions.

[0046] The SPRT type of test is based on the maximum likelihood ratio. The test sequentially samples a process at discrete times until it is capable of deciding between two alternatives: $H_0: \mu=0$; and $H_1: \mu=M$. In other words, is the sequence of sampled values indicative of a distribution around zero, or indicative of a distribution around some non-zero value? It has been demonstrated that the following approach provides an optimal decision method (the average sample size is less than a comparable fixed sample test). A test statistic, Ψ_t , is computed from the following formula:

$$\Psi_t = \sum_{i=1+j}^t \ln \left[\frac{f_{H1}(y_i)}{f_{H0}(y_i)} \right] \quad (14)$$

where $\ln()$ is the natural logarithm, $f_{Hs}()$ is the probability density function of the observed value of the random variable Y_i under the hypothesis H_s and j is the time point of the last decision.

[0047] In deciding between two alternative hypotheses, without knowing the true state of the signal under surveillance, it is possible to make an error (incorrect hypothesis decision). Two types of errors are possible. Rejecting H_0 when it is true (type I error) or accepting H_0 when it is false (type II error). Preferably these errors are controlled at some arbitrary minimum value, if possible. So, the probability of a false alarm or making a type I error is termed α , and the probability of missing an alarm or making a type II error is termed β . The well-known Wald's Approximation defines a lower bound, L , below which one accepts H_0 and an upper bound, U above which one rejects H_0 .

$$U = \ln \left[\frac{1-\beta}{\alpha} \right] \quad (15)$$

$$L = \ln \left[\frac{\beta}{1-\alpha} \right] \quad (16)$$

[0048] Decision Rule: if $\Psi_t \leq L$, then ACCEPT H_0 ; else if $\Psi_t \geq U$, then REJECT H_0 ; otherwise, continue sampling.

[0049] To implement this procedure, this distribution of the process must be known. This is not a problem in general, because some a priori information about the system exists. For most purposes, the multivariate Gaussian distribution is satisfactory, and the SPRT test can be simplified by assuming a Gaussian probability distribution p :

$$p = \frac{1}{\sigma \sqrt{2\pi}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2} \right]} \quad (17)$$

Then, the test statistic a typical sequential test deciding between zero-mean hypothesis H_0 and a positive mean hypothesis H_1 is:

$$\Psi_{t+1} = \Psi_t + \frac{M}{\sigma^2} \left(y_t - \frac{M}{2} \right) \quad (18)$$

where M is the hypothesized mean (typically set at a standard deviation away from zero, as given by the variance), σ is the variance of the training residual data and y_t is the input value being tested. Then the decision can be made at any observation t+1 in the sequence according to:

1. If $\Psi_{t+1} \leq \ln(\beta/(1-\alpha))$, then accept hypothesis H_0 as true;
2. If $\Psi_{t+1} \geq \ln((1-\beta)/\alpha)$, then accept hypothesis H_1 as true; and
3. If $\ln(\beta/(1-\alpha)) < \Psi_{t+1} < \ln((1-\beta)/\alpha)$, then make no decision and continue sampling.

The SPRT test can run against the residual for each monitored parameter, and can be tested against a positive biased mean, a negative biased mean, and against other statistical moments, such as the variance in the residual.

[0050] Other statistical decision techniques can be used in place of SPRT to determine whether the remotely monitored process or machine is operating in an abnormal way that indicates an incipient fault. According to another technique, the estimated sensor data and the actual sensor data can be compared using the similarity operator to obtain a vector similarity. If the vector similarity falls below a selected threshold, an alert can be indicated and action taken to notify an interested party as mentioned above that an abnormal condition has been monitored.

[0051] It should be appreciated that a wide range of changes and modifications may be made to the embodiments of the invention as described herein. Thus, it is intended that the foregoing detailed description be regarded as illustrative rather than limiting.

Claims

1. A method of system state monitoring comprising:

providing a reference set of data (H) comprising a plurality of learned observations from sensors of a modeled system that characterize the dynamic behaviour of the modeled system, wherein the reference set of data (H) is in the form of a matrix, with each column of the matrix representing an observation and each row representing values from a single sensor;

providing a current observation regarding the modeled system;

comparing the current observation to the reference set of data (H) using a similarity operator to render a similarity score for each learned observation in the reference set of data (H); and

if the similarity score for a learned observation is above a threshold or is one of a predetermined number of the highest similarity scores across all learned observation, including that learned observations in a subset of data (D) of the reference set of data (H);

computing a model of the system based on the current observation and the current subset of data (D) derived from the reference set of data (H), wherein computing the model comprises generating a model estimate comprising a weighted composite of the subset of data (D).

providing a series of subsequent current observations regarding the modeled system;

re-computing the model by re-determining the subset of data (D) for each new current observation; and

detecting incipience of faults in the system by testing the model estimate in contrast to the current observation.

2. The method of claim 1, wherein providing a reference set of data (H) further comprises receiving the observations for a plurality of different times.

3. The method of claim 1, wherein providing a current observation regarding the modeled system further comprises monitoring the system using a plurality of sensors.

4. The method of claim 1, wherein:

providing a reference set of data further comprises receiving information as corresponds to a first plurality of information sources; and

providing a current observation further comprises receiving information as corresponds to a second plurality of information sources.

5. The method of claim 4, wherein the second plurality of information sources are at least partially the same as the first plurality of information sources.

6. The method of claim 5, wherein the second plurality of information sources are at least partially the same as the first plurality of information sources but not fully inclusive of all of the first plurality of information sources.

7. The method of claim 1, wherein comparing the current observation to the reference set of data (H) to determine a subset of data (D), further comprises determining similarity as a function, at least in part, by:

defining similarity as a scalar range, bounded at each end thereof;
defining a level of similarity for two identical inputs as comprising a value that corresponds to one of the ends of the scalar range; or
providing for an absolute value of a similarity value to increase as two inputs approach being identical.

8. The method of claim 1, wherein comparing the current observation to the reference set of data (H) to determine a subset of data (D) comprises calculating the similarity between reference observations of the reference set of data and the current observation, wherein not all of the elements in the observations being compared are used to determine the similarity.

9. The method of claim 1, wherein providing a reference set of data (H) further comprises at least one of:

receiving observations regarding non-sensor measurements related to the modeled system, the non-sensor measurements comprising at least one of statistical data, demographic data network traffic statistics, biological cell counts or qualitative measurements; and
receiving information regarding at least one ambient condition as corresponds to the given system.

10. The method of claim 1, further comprising comparing the current observation to the reference set of data (H) to determine a subset of data (D) comprises selecting learned observations from the reference set of data (H) as a function, at least in part, on the basis of at least one variable that is not an input or an output of the model.

11. The method of claim 10, wherein the at least one variable comprises an ambient condition variable.

12. The method of claim 10, comparing the current observation to the reference set of data (H) to determine a subset of data (D) comprises using the at least one variable to eliminate a portion of the reference data from inclusion in the subset and using a predetermined level of similarity to select reference data for inclusion in the subset from what is left after elimination by the at least one variable.

13. The method of claim 1, further comprising modifying the reference set of data (H).

14. The method of claim 13, wherein modifying the reference set of data further comprises:

identifying at least one monitored system variable as an identified variable, including identifying at least one ambient variable;
determining when the identified variable exceeds a range of values for the identified variable as are presently included in the reference set of data; and
modifying the reference set of data as a function, at least in part, of the identified variable that exceeds the range of values.

15. The method of claim 14, wherein modifying the reference set of data as a function, at least in part, of the identified variable that exceeds the range of values further comprises adding additional data to the reference set of data.

16. The method of claim 14, wherein modifying the reference set of data as a function, at least in part, of the identified variable that exceeds the range of values further comprises substituting new data for existing data in the reference set of data.

17. The method of claim 14, and further comprising:

when the identified variable exceeds a range of values for the identified variable as are presently included in

the reference set of data, determining whether the reference set of data should nevertheless not be modified.

18. The method of claim 17, wherein determining whether the reference set of data should nevertheless not be modified further comprises determining whether the reference set of data has already been modified a predetermined number of times.

19. The method of claim 17, wherein determining whether the reference set of data should nevertheless not be modified further comprises determining whether the given system is likely exhibiting a fault.

20. An apparatus for system state monitoring comprising:

first means for providing a reference set of data (H) comprising a plurality of learned observations from sensors of a modeled system that characterize the dynamic behaviour of the modeled system, wherein the reference set of data (H) is in the form of a matrix, with each column of the matrix representing an observation and each row representing values from a single sensor;

second means for providing a current observation regarding the modeled system;

third means for comparing the current observation to the reference set of data (H) using a similarity operator to render a similarity score for each learned observation in the reference set of data (H), and, if the similarity score for a learned observation is above a threshold or is one of a predetermined number of the highest similarity scores across all learned observations, including that learned observation in a subset of data (D) of the reference set of data (H);

fourth means for computing a model of the system based on the current observation and the current subset of data (D) derived from the reference set of data (H), wherein computing the model comprises generating a model estimate comprising a weighted composite of the subset of data (D);

fifth means for providing a series of subsequent current observations regarding the modeled system;

sixth means for recomputing the model by re-determining the subset of data (D) for each new current observation; and

seventh means for detecting incipience of faults in the system by testing the model estimate in contrast to the current observation

21. The apparatus of claim 20, wherein the first means for providing a reference set of data (H) comprises means for receiving the observations for a plurality of different times.

22. The apparatus of claim 20, wherein the second means for providing a current observation in relation to the modeled system comprises means for monitoring the system using a plurality of sensors.

23. The apparatus of claim 20, wherein:

the first means for providing a reference set of data further comprises means for receiving information as corresponds to a first plurality of information sources; and

the second means for providing a current observation comprises means for receiving information as corresponds to a second plurality of information sources.

24. The apparatus of claim 23, wherein the second plurality of information sources are at least partially the same as the first plurality of information sources.

25. The apparatus of claim 24, wherein the second plurality of information sources are at least partially the same as the first plurality of information sources but not fully inclusive of all of the first plurality of information sources.

26. The apparatus of claim 20, wherein the third means for comparing the current observation to the reference set of data (H) to determine a subset of data (D) comprises means for determining similarity as a function, at least in part, by:

defining similarity as a scalar range, bounded at each end thereof;

defining a level of similarity for two identical inputs as comprising a value that corresponds to one of the ends of the scalar range; or

providing for an absolute value of a similarity value to increase as two inputs approach being identical.

27. The apparatus of claim 20, wherein the third means for comparing the current observation to the reference set of

data (H) to determine a subset of data (D) comprises means for calculating the similarity between reference observations of the reference set of data and the current observation, wherein not all of the elements in the observations being compared are used to determine the similarity.

5 **28.** The apparatus of claim 20, wherein the first means for providing a reference set of data (H) further comprises at least one of:

10 means for receiving observations regarding non-sensor measurements related to the modeled system, the non-sensor measurements comprising at least one of statistical data, demographic data network traffic statistics, biological cell counts or qualitative measurements;
 and
 means receiving information regarding at least one ambient condition as corresponds to the given system.

15 **29.** The apparatus of claim 20, wherein the third means for comparing the current observation to the reference set of data (H) to determine a subset of data (D) comprises means for selecting learned observations from the reference set of data (H) as a function, at least in part, on the basis of at least one variable that is not an input or an output of the model.

20 **30.** The apparatus of claim 29, wherein the at least one variable comprises an ambient condition variable.

25 **31.** The apparatus of claim 29, wherein the means for selecting learned observations from the reference set of data (H) as a function, at least in part, on the basis of at least one variable that does not comprise an observation of the reference set of data (H) and the current observation comprises means for using the at least one variable to eliminate a portion of the reference data from inclusion in the subset and using a predetermined level of similarity to select reference data for inclusion in the subset from what is left after elimination by the at least one variable.

30 **32.** The apparatus of claim 20, further comprising means for modifying the reference set of data (H).

35 **33.** The apparatus of claim 32, wherein the means for modifying the reference set of data further comprises:

 means for identifying at least one monitored system variable as an identified variable, including identifying at least one ambient variable;
 means for determining when the identified variable exceeds a range of values for the identified variable as are presently included in the reference set of data; and
 means for modifying the reference set of data as a function, at least in part, of the identified variable that exceeds the range of values.

40 **34.** The apparatus of claim 33, wherein the means for modifying the reference set of data as a function, at least in part, of the identified variable that exceeds the range of values further comprises means for adding additional data to the reference set of data.

45 **35.** The apparatus of claim 33, wherein the means for modifying the reference set of data as a function, at least in part, of the identified variable that exceeds the range of values further comprises means for substituting new data for existing data in the reference set of data.

50 **36.** The apparatus of claim 33, further comprising:

 means for determining whether the reference set of data should nevertheless not be modified, when the identified variable exceeds a range of values for the identified variable as are presently included in the reference set of data.

55 **37.** The apparatus of claim 36, wherein comprises means for determining whether the reference set of data has already been modified a predetermined number of times.

38. The apparatus of claim 36, wherein the means for determining whether the reference set of data should nevertheless not be modified further comprises means for determining whether the given system is likely exhibiting a fault.

Patentansprüche

1. Verfahren zur Systemzustandsüberwachung, umfassend:

5 Bereitstellen eines Referenzdatensatzes (H), der mehrere gelernte Beobachtungen von Sensoren eines modellierten Systems umfasst, die das dynamische Verhalten des modellierten Systems kennzeichnen, wobei der Referenzdatensatz (H) die Form einer Matrix aufweist, wo jede Spalte der Matrix für eine Beobachtung steht und jede Reihe für Werte eines einzelnen Sensors steht;
 Bereitstellen einer aktuellen Beobachtung bezüglich des modellierten Systems;
 10 Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H) unter Verwendung eines Ähnlichkeitsoperators, um eine Ähnlichkeitsbewertung für jede gelernte Beobachtung im Referenzdatensatz (H) zu erbringen; und
 falls die Ähnlichkeitsbewertung für eine gelernte Beobachtung über einem Schwellenwert liegt oder eine von einer vorbestimmten Zahl der höchsten Ähnlichkeitsbewertungen über die gesamte gelernte Beobachtung ist,
 15 Aufnehmen der gelernten Beobachtungen in einem Datenuntersatz (D) des Referenzdatensatzes (H);
 Berechnen eines Modells des Systems, basierend auf der aktuellen Beobachtung und dem aktuellen Datenuntersatz (D), der vom Referenzdatensatz (H) abgeleitet ist, wobei das Berechnen des Modells ein Erstellen einer Modellschätzung umfasst, die eine gewichtete Zusammensetzung des Datenuntersatzes (D) umfasst;
 Bereitstellen einer Serie nachfolgender aktueller Beobachtungen bezüglich des modellierten Systems;
 20 Neuberechnen des Modells durch Neuermitteln des Datenuntersatzes (D) für jede neue aktuelle Beobachtung; und
 Erfassen eines Anfangs von Fehlern im System durch Testen der Modellschätzung in Kontrast zur aktuellen Beobachtung.

25 2. Verfahren nach Anspruch 1, wobei das Bereitstellen eines Referenzdatensatzes (H) des Weiteren ein Empfangen der Beobachtungen für mehrere verschiedene Zeiten umfasst.

3. Verfahren nach Anspruch 1, wobei das Bereitstellen einer aktuellen Beobachtung bezüglich des modellierten Systems des Weiteren ein Überwachen des Systems unter Verwendung mehrerer Sensoren umfasst.

30 4. Verfahren nach Anspruch 1, wobei:
 das Bereitstellen eines Referenzdatensatzes des Weiteren ein Empfangen von Informationen, wie es einer ersten Vielzahl von Informationsquellen entspricht, umfasst; und
 35 das Bereitstellen einer aktuellen Beobachtung des Weiteren ein Empfangen von Informationen, wie es einer zweiten Vielzahl von Informationsquellen entspricht, umfasst.

40 5. Verfahren nach Anspruch 4, wobei die zweite Vielzahl von Informationsquellen zumindest teilweise dieselbe ist wie die erste Vielzahl von Informationsquellen.

6. Verfahren nach Anspruch 5, wobei die zweite Vielzahl von Informationsquellen zumindest teilweise dieselbe ist wie die erste Vielzahl von Informationsquellen, aber nicht vollständig alle der ersten Vielzahl von Informationsquellen enthält.

45 7. Verfahren nach Anspruch 1, wobei das Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H), um einen Datenuntersatz (D) zu ermitteln, des Weiteren ein Ermitteln einer Ähnlichkeit als eine Funktion umfasst, durch, zumindest teilweise:

50 Definieren der Ähnlichkeit als eine Skalarspanne, die an jedem Ende davon begrenzt ist;
 Definieren eines Ähnlichkeitsgrads für zwei identische Eingaben als einen Wert umfassend, der einem der Enden der Skalarspanne entspricht; oder
 Bereitstellen einer Zunahme für einen absoluten Wert eines Ähnlichkeitswertes, wenn zwei Eingaben sich annähern, ident zu sein.

55 8. Verfahren nach Anspruch 1, wobei das Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H), um einen Datenuntersatz (D) zu ermitteln, ein Berechnen der Ähnlichkeit zwischen Referenzbeobachtungen des Referenzdatensatzes und der aktuellen Beobachtung umfasst, wobei nicht alle der Elemente in den Beobachtungen,

die verglichen sind, verwendet werden, um die Ähnlichkeit zu ermitteln.

9. Verfahren nach Anspruch 1, wobei das Bereitstellen eines Referenzdatensatzes (H) des Weiteren zumindest eines umfasst von:

Empfangen von Beobachtungen bezüglich Nicht-Sensormessungen, die dem modellierten System zugehörig sind, wobei die Nicht-Sensormessungen zumindest eines von statistischen Daten, demographischen Datennetzwerksverkehrsstatistiken, biologischen Zellzählungen oder qualitativen Messungen umfassen; und Empfangen von Informationen bezüglich zumindest einer Umgebungsbedingung entsprechend dem gegebenen System.

10. Verfahren nach Anspruch 1, des Weiteren umfassend, dass das Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H), um einen Datenuntersatz (D) zu ermitteln, ein Auswählen gelernter Beobachtungen aus dem Referenzdatensatz (H) als eine Funktion, zumindest teilweise, auf Basis zumindest einer Variable, die keine Eingabe oder Ausgabe des Modells ist, umfasst.

11. Verfahren nach Anspruch 10, wobei die zumindest eine Variable eine Umgebungsbedingungsvariable umfasst.

12. Verfahren nach Anspruch 10, wobei das Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H), um einen Datenuntersatz (D) zu ermitteln, ein Verwenden der zumindest einen Variable, um einen Abschnitt der Referenzdaten vom Einbezug in den Untersatz auszuschließen, und ein Verwenden eines vorbestimmten Ähnlichkeitsgrads zum Auswählen von Referenzdaten zum Einbezug in den Untersatz dessen, was nach dem Ausschluss durch die zumindest eine Variable übrig ist, umfasst.

13. Verfahren nach Anspruch 1, des Weiteren umfassend ein Modifizieren des Referenzdatensatzes (H).

14. Verfahren nach Anspruch 13, wobei das Modifizieren des Referenzdatensatzes des Weiteren umfasst:

Identifizieren zumindest einer überwachten Systemvariable als eine identifizierte Variable, enthaltend ein Identifizieren zumindest einer Umgebungsvariable;
Ermitteln, wann die identifizierte Variable eine Spanne von Werten für die identifizierte Variable, wie sie momentan im Referenzdatensatz enthalten sind, überschreitet; und
Modifizieren des Referenzdatensatzes als eine Funktion, zumindest teilweise, der identifizierten Variable, die die Spanne an Werten überschreitet.

15. Verfahren nach Anspruch 14, wobei das Modifizieren des Referenzdatensatzes als eine Funktion, zumindest teilweise, der identifizierten Variable, die die Spanne an Werten überschreitet, des Weiteren ein Hinzufügen zusätzlicher Daten zum Referenzdatensatz umfasst.

16. Verfahren nach Anspruch 14, wobei das Modifizieren des Referenzdatensatzes als eine Funktion, zumindest teilweise, der identifizierten Variable, die die Spanne an Werten überschreitet, des Weiteren ein Einwechseln neuer Daten für bestehende Daten im Referenzdatensatz umfasst.

17. Verfahren nach Anspruch 14 und des Weiteren umfassend:

wenn die identifizierte Variable eine Spanne an Werten für die identifizierte Variable, wie sie momentan im Referenzdatensatz enthalten sind, überschreitet, Ermitteln, ob der Referenzdatensatz nichtsdestotrotz nicht modifiziert werden sollte.

18. Verfahren nach Anspruch 17, wobei das Ermitteln, ob der Referenzdatensatz nichtsdestotrotz nicht modifiziert werden sollte, des Weiteren ein Ermitteln umfasst, ob der Referenzdatensatz bereits eine vorbestimmte Anzahl an Modifizierungen aufweist.

19. Verfahren nach Anspruch 17, wobei das Ermitteln, ob der Referenzdatensatz nichtsdestotrotz nicht modifiziert werden sollte, des Weiteren ein Ermitteln umfasst, ob das gegebene System wahrscheinlich einen Fehler aufweist.

20. Vorrichtung zur Systemzustandsüberwachung, umfassend:

erste Mittel zum Bereitstellen eines Referenzdatensatzes (H), der mehrere gelernte Beobachtungen von Sensoren eines modellierten Systems umfasst, die das dynamische Verhalten des modellierten Systems kennzeichnen, wobei der Referenzdatensatz (H) die Form einer Matrix aufweist, wo jede Spalte der Matrix für eine Beobachtung steht und jede Reihe für Werte eines einzelnen Sensors steht;

zweite Mittel zum Bereitstellen einer aktuellen Beobachtung bezüglich des modellierten Systems;

dritte Mittel zum Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H) unter Verwendung eines Ähnlichkeitsoperators, um eine Ähnlichkeitsbewertung für jede gelernte Beobachtung im Referenzdatensatz (H) zu erbringen und falls die Ähnlichkeitsbewertung für eine gelernte Beobachtung über einem Schwellenwert liegt oder eine von einer vorbestimmten Zahl der höchsten Ähnlichkeitsbewertungen über die gesamte gelernte Beobachtung ist, Aufnehmen der gelernten Beobachtungen in einem Datenuntersatz (D) des Referenzdatensatzes (H);

vierte Mittel zum Berechnen eines Modells des Systems, basierend auf der aktuellen Beobachtung und dem aktuellen Datenuntersatz (D), der vom Referenzdatensatz (H) abgeleitet ist, wobei das Berechnen des Modells ein Erstellen einer Modellschätzung umfasst, die eine gewichtete Zusammensetzung des Datenuntersatzes (D) umfasst;

fünfte Mittel zum Bereitstellen einer Serie nachfolgender aktueller Beobachtungen bezüglich des modellierten Systems;

sechste Mittel zum Neuberechnen des Modells durch Neuermitteln des Datenuntersatzes (D) für jede neue aktuelle Beobachtung; und

siebte Mittel zum Erfassen eines Anfangs von Fehlern im System durch Testen der Modellschätzung in Kontrast zur aktuellen Beobachtung.

21. Vorrichtung nach Anspruch 20, wobei die ersten Mittel zum Bereitstellen eines Referenzdatensatzes (H) des Weiteren Mittel zum Empfangen der Beobachtungen für mehrere verschiedene Zeitpunkte umfassen.

22. Vorrichtung nach Anspruch 20, wobei die zweiten Mittel zum Bereitstellen einer aktuellen Beobachtung bezüglich des modellierten Systems des Weiteren Mittel zum Überwachen des Systems unter Verwendung mehrerer Sensoren umfassen.

23. Vorrichtung nach Anspruch 20, wobei:

die ersten Mittel zum Bereitstellen eines Referenzdatensatzes des Weiteren Mittel zum Empfangen von Informationen, wie es einer ersten Vielzahl von Informationsquellen entspricht, umfassen; und

die Mittel zum Bereitstellen einer aktuellen Beobachtung des Weiteren Mittel zum Empfangen von Informationen, wie es einer zweiten Vielzahl von Informationsquellen entspricht, umfassen.

24. Vorrichtung nach Anspruch 23, wobei die zweite Vielzahl von Informationsquellen zumindest teilweise dieselbe ist wie die erste Vielzahl von Informationsquellen.

25. Vorrichtung nach Anspruch 24, wobei die zweite Vielzahl von Informationsquellen zumindest teilweise dieselbe ist wie die erste Vielzahl von Informationsquellen, aber nicht vollständig alle der ersten Vielzahl von Informationsquellen enthält.

26. Vorrichtung nach Anspruch 20, wobei die dritten Mittel zum Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H), um einen Datenuntersatz (D) zu ermitteln, des Weiteren ein Ermitteln einer Ähnlichkeit als eine Funktion umfasst, durch, zumindest teilweise:

Definieren der Ähnlichkeit als eine Skalarspanne, die an jedem Ende davon begrenzt ist;

Definieren eines Ähnlichkeitsgrads für zwei identische Eingaben als einen Wert umfassend, der einem der Enden der Skalarspanne entspricht; oder

Bereitstellen einer Zunahme für einen absoluten Wert eines Ähnlichkeitswertes, wenn zwei Eingaben sich annähern, ident zu sein.

27. Vorrichtung nach Anspruch 20, wobei die dritten Mittel zum Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H), um einen Datenuntersatz (D) zu ermitteln, Mittel zum Berechnen der Ähnlichkeit zwischen Referenzbeobachtungen des Referenzdatensatzes und der aktuellen Beobachtung umfassen, wobei nicht alle der Elemente in den Beobachtungen, die verglichen sind, verwendet werden um die Ähnlichkeit zu ermitteln.

28. Vorrichtung nach Anspruch 20, wobei die Mittel zum Bereitstellen eines Referenzdatensatzes (H) des Weiteren zumindest eines umfassen von:

Mitteln zum Empfangen von Beobachtungen bezüglich Nicht-Sensormessungen, die dem modellierten System zugehörig sind, wobei die Nicht-Sensormessungen zumindest eines von statistischen Daten, demographischen Datennetzwerksverkehrsstatistiken, biologischen Zellzählungen oder qualitativen Messungen umfassen; und Mitteln zum Empfangen von Informationen bezüglich zumindest einer Umgebungsbedingung entsprechend dem gegebenen System.

29. Vorrichtung nach Anspruch 20, wobei die dritten Mittel zum Vergleichen der aktuellen Beobachtung mit dem Referenzdatensatz (H), um einen Datenuntersatz (D) zu ermitteln, Mittel zum Auswählen gelernter Beobachtungen aus dem Referenzdatensatz (H) als eine Funktion, zumindest teilweise, auf Basis zumindest einer Variable, die keine Eingabe oder Ausgabe des Modells ist, umfassen.

30. Vorrichtung nach Anspruch 29, wobei die zumindest eine Variable eine Umgebungsbedingungsvariable umfasst.

31. Vorrichtung nach Anspruch 29, wobei die Mittel zum Auswählen gelernter Beobachtungen aus dem Referenzdatensatz (H) als eine Funktion, zumindest teilweise, auf der Basis zumindest einer Variable, die keine Beobachtung des Referenzdatensatzes (H) umfasst, und der aktuellen Beobachtung, Mittel zum Verwenden der zumindest einen Variable umfassen, um einen Abschnitt der Referenzdaten vom Einbezug in den Untersatz auszuschließen und Verwenden eines vorbestimmten Ähnlichkeitsgrades zum Auswählen von Referenzdaten zum Einbezug in den Untersatz dessen, was nach dem Ausschluss durch die zumindest eine Variable übrig ist.

32. Vorrichtung nach Anspruch 20, des Weiteren umfassend Mittel zum Modifizieren des Referenzdatensatzes (H).

33. Verfahren nach Anspruch 32, wobei die Mittel zum Modifizieren des Referenzdatensatzes des Weiteren umfassen:

Mittel zum Identifizieren zumindest einer überwachten Systemvariable als eine identifizierte Variable, enthaltend ein Identifizieren zumindest einer Umgebungsvariable;
Mittel zum Ermitteln, wann die identifizierte Variable eine Spanne von Werten für die identifizierte Variable, wie sie momentan im Referenzdatensatz enthalten sind, überschreitet; und
Mittel zum Modifizieren des Referenzdatensatzes als eine Funktion, zumindest teilweise, der identifizierten Variable, die die Spanne an Werten überschreitet.

34. Vorrichtung nach Anspruch 33, wobei die Mittel zum Modifizieren des Referenzdatensatzes als eine Funktion, zumindest teilweise, der identifizierten Variable, die die Spanne an Werten überschreitet, des Weiteren Mittel zum Hinzufügen zusätzlicher Daten zum Referenzdatensatz umfassen.

35. Vorrichtung nach Anspruch 33, wobei die Mittel zum Modifizieren des Referenzdatensatzes als eine Funktion, zumindest teilweise, der identifizierten Variable, die die Spanne an Werten überschreitet, des Weiteren Mittel zum Einwechseln neuer Daten für bestehende Daten im Referenzdatensatz umfassen.

36. Vorrichtung nach Anspruch 33, des Weiteren umfassend:

wenn die identifizierte Variable eine Spanne an Werten für die identifizierte Variable, wie sie momentan im Referenzdatensatz enthalten sind, überschreitet, Ermitteln, ob der Referenzdatensatz nichtsdestotrotz nicht modifiziert werden sollte.

37. Vorrichtung nach Anspruch 36, wobei sie Mittel zum Ermitteln umfasst, ob der Referenzdatensatz bereits eine vorbestimmte Anzahl an Modifizierungen aufweist.

38. Vorrichtung nach Anspruch 36, die Mittel zum Ermitteln, ob der Referenzdatensatz nichtsdestotrotz nicht modifiziert werden sollte, des Weiteren Mittel zum Ermitteln umfassen, ob das gegebene System wahrscheinlich einen Fehler aufweist.

Revendications**1.** Procédé de surveillance d'état d'un système comprenant :

la fourniture d'un ensemble de données de référence (H) comprenant une pluralité d'observations apprises de capteurs d'un système modélisé qui caractérisent le comportement dynamique du système modélisé, dans lequel l'ensemble de données de référence (H) se présente sous la forme d'une matrice, chaque colonne de la matrice représentant une observation et chaque rangée représentant des valeurs issues d'un capteur individuel ;

la fourniture d'une observation courante concernant le système modélisé ;

la comparaison de l'observation courante à l'ensemble de données de référence (H) en utilisant un opérateur de similarité pour rendre un score de similarité pour chaque observation apprise dans l'ensemble de données de référence (H) ; et

si le score de similarité pour une observation apprise se situe au-dessus d'un seuil ou est l'un d'un nombre prédéterminé des scores de similarité les plus élevés sur toute l'observation apprise, comprenant cette observation apprise dans un sous-ensemble de données (D) de l'ensemble de données de référence (H) ;

le calcul d'un modèle du système sur la base de l'observation courante et du sous-ensemble de données courant (D) tiré de l'ensemble de données de référence (H), dans lequel le calcul du modèle comprend la génération d'une estimation de modèle comprenant un composite pondéré du sous-ensemble de données (D), la fourniture d'une série d'observations courantes ultérieures concernant le système modélisé ;

le re-calcul du modèle par une re-détermination du sous-ensemble de données (D) pour chaque nouvelle observation courante ; et

la détection de l'apparition de défauts dans le système en testant l'estimation de modèle au regard de l'observation courante.

2. Procédé selon la revendication 1, dans lequel la fourniture d'un ensemble de données de référence (H) comprend en outre la réception des observations une pluralité de fois différentes.**3.** Procédé selon la revendication 1, dans lequel la fourniture d'une observation courante concernant le système modélisé comprend en outre la surveillance du système en utilisant une pluralité de capteurs.**4.** Procédé selon la revendication 1, dans lequel :

la fourniture d'un ensemble de données de référence comprend en outre la réception d'informations qui correspondent à une première pluralité de sources d'information ; et

la fourniture d'une observation courante comprend en outre la réception d'informations qui correspondent à une seconde pluralité de sources d'informations.

5. Procédé selon la revendication 4, dans lequel la seconde pluralité de sources d'information est au moins en partie la même que la première pluralité de sources d'informations.**6.** Procédé selon la revendication 5, dans lequel la seconde pluralité de sources d'information est au moins en partie la même que la première pluralité de sources d'information, mais n'inclut pas totalement la première pluralité de sources d'information.**7.** Procédé selon la revendication 1, dans lequel la comparaison de l'observation courante à l'ensemble de données de référence (H) pour déterminer un sous-ensemble de données (D) comprend en outre la détermination de la similarité en tant que fonction au moins en partie :

en définissant la similarité comme une plage scalaire délimitée à chacune de ses extrémités ;

en définissant un niveau de similarité pour deux entrées identiques comme comprenant une valeur qui correspond à l'une des extrémités de la plage scalaire ;

ou

la condition pour qu'une valeur absolue d'une valeur de similarité augmente lorsque deux entrées se rapprochent étant identique.

8. Procédé selon la revendication 1, dans lequel la comparaison de l'observation courante à l'ensemble de données de référence (H) pour déterminer un sous-ensemble de données (D) comprend le calcul de la similarité entre des

observations de référence de l'ensemble de données de référence et l'observation courante, dans lequel tous les éléments des observations en comparaison n'étant pas utilisés pour déterminer la similarité.

- 5 9. Procédé selon la revendication 1, dans lequel la fourniture d'un ensemble de données de référence (H) comprend en outre au moins l'une ou l'autre des étapes suivantes :

la réception d'observations concernant des mesures ne venant pas des capteurs en rapport avec le système modélisé, les mesures ne venant pas des capteurs comprenant au moins l'une parmi les données statistiques, les données démographiques, les statistiques de trafic de réseaux, les comptages de cellules biologiques ou les mesures quantitatives ; et
10 la réception d'informations concernant au moins une condition ambiante qui correspond au système donné.

- 15 10. Procédé selon la revendication 1, comprenant en outre la comparaison de l'observation courante à l'ensemble de données de référence (H) pour déterminer un sous-ensemble de données (D) comprend la sélection d'observations apprises de l'ensemble de données de référence (H) en tant que fonction, au moins en partie sur la base d'au moins une variable qui n'est pas une entrée ou une sortie du modèle.

- 20 11. Procédé selon la revendication 10, dans lequel la au moins une variable comprend une variable de condition ambiante.

- 25 12. Procédé selon la revendication 10, dans lequel la comparaison de l'observation courante à l'ensemble de données de référence (H) pour déterminer un sous-ensemble de données (D) comprend l'utilisation de la au moins une variable pour éliminer une partie des données de référence d'une inclusion dans le sous-ensemble et l'utilisation d'un niveau prédéterminé de similarité pour sélectionner des données de référence pour une inclusion dans le sous-ensemble à partir de ce qui reste après élimination par la au moins une variable.

- 30 13. Procédé selon la revendication 1, comprenant en outre la modification de l'ensemble de données de référence (H).

- 35 14. Procédé selon la revendication 13, dans lequel la modification de l'ensemble de données de référence comprend en outre :

l'identification d'au moins une variable du système surveillé en tant que variable identifiée, comprenant l'identification d'au moins une variable ambiante ;
la détermination du moment où la variable identifiée dépasse une plage de valeurs pour la variable identifiée telles qu'elles sont comprises actuellement dans l'ensemble de données de référence ; et
la modification de l'ensemble de données de référence en tant que fonction, au moins en partie, de la variable identifiée qui dépasse la plage de valeurs.

- 40 15. Procédé selon la revendication 14, dans lequel la modification de l'ensemble de données de référence en fonction, au moins en partie, de la variable identifiée qui dépasse la plage de valeurs comprend en outre l'addition de données supplémentaires à l'ensemble de données de référence.

- 45 16. Procédé selon la revendication 14, dans lequel la modification de l'ensemble de données de référence en fonction, au moins en partie, de la variable identifiée qui dépasse la plage de valeurs comprend en outre la substitution de nouvelles données pour les données existantes de l'ensemble de données de référence.

- 50 17. Procédé selon la revendication 14 et comprenant en outre :

lorsque la variable identifiée dépasse une plage de valeurs pour la variable identifiée telles qu'elles sont actuellement comprises dans l'ensemble de données de référence, la détermination du fait que l'ensemble de données de référence ne doit néanmoins pas être modifiée.

- 55 18. Procédé selon la revendication 17, dans lequel la détermination du fait que l'ensemble de données de référence ne doit néanmoins pas être modifiée encore comprend la détermination du fait que l'ensemble de données de référence a déjà été modifié un nombre prédéterminé de fois.

19. Procédé selon la revendication 17, dans lequel la détermination du fait que l'ensemble de données de référence ne doit pas néanmoins pas être modifié comprend encore la détermination du fait que le système donné est sus-

ceptible de présenter un défaut.

20. Appareil de surveillance de l'état d'un système comprenant :

- 5 un premier moyen pour fournir un ensemble de données de référence (H) comprenant une pluralité d'observations apprises venant de capteurs d'un système modélisé qui caractérisent le comportement dynamique du système modélisé, dans lequel l'ensemble de données de référence (H) se présente sous la forme d'une matrice, chaque colonne de la matrice représentant une observation et chaque rangée représentant des valeurs venant d'un capteur individuel ;
- 10 un deuxième moyen pour fournir une observation courante concernant le système modélisé ;
un troisième moyen pour comparer l'observation courante à l'ensemble de données de référence (H) en utilisant un opérateur de similarité afin d'obtenir un score de similarité pour chaque observation apprise de l'ensemble de données de référence (H) et, si le score de similarité pour une observation apprise se situe au-dessus d'un seuil ou est l'un d'un nombre prédéterminé des scores de similarité les plus élevés d'entre toutes les observations apprises, notamment l'observation apprise dans un sous-ensemble de données (D) de l'ensemble de données de référence (H) ;
- 15 un quatrième moyen pour calculer un modèle du système sur la base de l'observation courante et du sous-ensemble courant de données (D) tirées de l'ensemble de données de référence (H), dans lequel le calcul du modèle comprend la génération d'une estimation de modèle comprenant un composite pondéré du sous-ensemble de données (D) ;
- 20 un cinquième moyen pour fournir une série d'observations courantes ultérieures concernant le système modélisé ;
un sixième moyen pour recalculer le modèle en re-déterminant le sous-ensemble de données (D) pour chaque nouvelle observation courante ; et
- 25 un septième moyen pour détecter l'apparition de défauts dans le système en testant l'estimation de modèle au regard de l'observation courante.

21. Appareil selon la revendication 20, dans lequel le premier moyen pour fournir un ensemble de données de référence (H) comprend un moyen pour recevoir les observations une pluralité de différentes fois.

22. Appareil selon la revendication 20, dans lequel le deuxième moyen pour fournir une observation courante en rapport avec le système modélisé comprend un moyen pour surveiller le système en utilisant une pluralité de capteurs.

23. Appareil selon la revendication 20, dans lequel :

- le premier moyen pour fournir un ensemble de données de référence comprend en outre un moyen pour recevoir des informations qui correspondent à une première pluralité de sources d'information ; et
- le deuxième moyen pour fournir une observation courante comprend un moyen pour recevoir des informations qui correspondent à une deuxième pluralité de sources d'information.

24. Appareil selon la revendication 23, dans lequel la deuxième pluralité de sources d'information est au moins en partie la même que la première pluralité de sources d'information.

25. Appareil selon la revendication 24, dans lequel la deuxième pluralité de sources d'information est au moins en partie la même que la première pluralité de sources d'information, mais ne comprend pas la totalité de toutes la première pluralité de sources d'information.

26. Appareil selon la revendication 20, dans lequel le troisième moyen pour comparer l'observation courante à l'ensemble de données de référence (H) pour déterminer un sous-ensemble de données (D) comprend un moyen pour déterminer la similarité en tant que fonction, au moins en partie :

- en définissant une similarité comme une plage scalaire délimitée à chacune de ses extrémités ;
- en définissant un niveau de similarité pour deux entrées identiques en ce qu'elles comprennent une valeur qui correspond à l'une des extrémités de la plage scalaire ;
- ou
- à condition qu'une valeur absolue d'une valeur de similarité qui augmente lorsque deux entrées se rapprochent soit identique.

27. Appareil selon la revendication 20, dans lequel le troisième moyen pour comparer l'observation courante à l'ensemble de données de référence (H) afin de déterminer un sous-ensemble de données (D) comprend un moyen pour calculer la similarité entre des observations de référence de l'ensemble de données de référence et l'observation courante, dans lequel tous les éléments des observations comparées ne sont pas utilisés pour déterminer la similarité.

28. Appareil selon la revendication 20, dans lequel le premier moyen pour fournir un ensemble de données de référence (H) comprend en outre au moins l'un des suivants :

un moyen pour recevoir des observations concernant des mesures ne venant pas de capteurs se rapportant au système modélisé, les mesures ne venant pas de capteurs comprenant au moins l'une ou l'autre de données statistiques, de données démographiques, de statistiques de trafic de réseau, de comptage de cellules biologiques ou de mesures qualitatives ; et
un moyen recevant des informations concernant au moins une condition ambiante qui correspond au système donné.

29. Appareil selon la revendication 20, dans lequel le troisième moyen pour comparer l'observation courante à l'ensemble de données de référence (H) afin de déterminer un sous-ensemble de données (D) comprend un moyen pour sélectionner des observations apprises provenant d'un ensemble de données de référence (H) en tant que fonction au moins en partie, sur la base d'au moins une variable qui n'est pas une entrée ou une sortie du modèle.

30. Appareil selon la revendication 29, dans lequel la au moins une variable comprend une variable de condition ambiante.

31. Appareil selon la revendication 29, dans lequel le moyen de sélection d'observations apprises provenant de l'ensemble de données de référence (H) en tant que fonction, au moins en partie, sur la base d'au moins une variable qui ne comprend pas une observation de l'ensemble de données de référence (H) et l'observation courante comprend un moyen pour utiliser la au moins une variable afin d'éliminer une partie des données de référence d'une inclusion dans le sous-ensemble et utiliser un niveau prédéterminé de similarité pour sélectionner des données de référence pour inclusion dans le sous-ensemble à partir de ce qui est resté après élimination par la au moins une variable.

32. Appareil selon la revendication 20, comprenant en outre un moyen pour modifier l'ensemble de données de référence (H).

33. Appareil selon la revendication 32, dans lequel le moyen pour modifier l'ensemble de données de référence comprend en outre :

un moyen pour identifier au moins une variable du système surveillé en tant que variable identifiée, notamment pour identifier au moins une variable ambiante ;
un moyen pour déterminer à quel moment la variable identifiée dépasse une plage de valeurs pour la variable identifiée qui sont actuellement comprises dans l'ensemble de données de référence ; et
un moyen pour modifier l'ensemble de données de référence en fonction, au moins en partie, de la variable identifiée qui dépasse la plage de valeurs.

34. Appareil selon la revendication 33, dans lequel le moyen de modification de l'ensemble de données de référence en fonction, au moins en partie, de la variable identifiée qui dépasse la plage de valeurs comprend en outre un moyen pour ajouter des données supplémentaires à l'ensemble de données de référence.

35. Appareil selon la revendication 33, dans lequel le moyen de modification de l'ensemble de données de référence en fonction, au moins en partie, de la variable identifiée qui dépasse la plage de valeurs comprend en outre un moyen pour substituer de nouvelles données aux données existantes dans l'ensemble de données de référence.

36. Appareil selon la revendication 33, comprenant en outre :

un moyen pour déterminer si l'ensemble de données de référence ne doit néanmoins pas être modifié lorsque la variable identifiée dépasse une plage de valeurs pour la variable identifiée telles qu'elles sont comprises actuellement dans l'ensemble de données de référence.

37. Appareil selon la revendication 36, qui comprend un moyen pour déterminer si l'ensemble de données de référence

a déjà été modifié un nombre prédéterminé de fois.

- 38.** Appareil selon la revendication 36, dans lequel le moyen pour déterminer si l'ensemble de données de référence ne doit néanmoins pas être modifié comprend en outre un moyen pour déterminer si le système donné est susceptible de présenter un défaut.

5

10

15

20

25

30

35

40

45

50

55

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 20020128731 A1 [0002]
- WO 0235299 A [0003]
- US 5764509 A, Gross [0045]

RENDSZERÁLLAPOT-FIGYELEÉS VISSZATÉRŐ HELYI TANULÓGÉP SEGÍTSÉGÉVEL

Szabadalmi igénypontok

1. Eljárás rendszerállapot-figyelésre, amely a következőket tartalmazza:

hivatkozási adathalmaz (H) biztosítása, amely modellezett rendszer olyan érzékelőtől való tanult megfigyelések sokaságát tartalmazza, amelyek jellemzik a modellezett rendszer dinamikus viselkedését, ahol a hivatkozási adathalmaz (H) mátrix formájában van, a mátrix egyes oszlopaival egy-egy megfigyelést ábrázolva, az egyes sorok pedig egy-egy önálló érzékelőtől való értékeket ábrázolnak;

a modellezett rendszerre vonatkozó aktuális megfigyelés biztosítása;

az aktuális megfigyelés összehasonlítása a hivatkozási adathalmazzal (H) hasonlósági műveleti jel segítségével, hogy hasonlósági pontot adjon a hivatkozási adathalmazban (H) lévő minden egyes tanult megfigyelésért; és

a hasonlósági pontszám egy tanult megfigyelésre egy küszöbérték fölött van-e, vagy az összes tanult megfigyelésre a legmagasabb hasonlósági pontszám előre meghatározott számainak egyike, beleértve a hivatkozási adathalmaz (H) részhalmazában (D) lévő tanult megfigyeléseket;

a rendszer modelljének kiszámítása az aktuális megfigyelés és a hivatkozási adathalmazból (H) származtatott aktuális részhalmaz (D) alapján, ahol a modell kiszámítása az adatok részhalmazának (D) súlyozott összetételét tartalmazó modellbecslés előállítását tartalmazza,

a modellezett rendszerre vonatkozó egymást követő aktuális megfigyelések sorozatának biztosítása;

a modell újraszámítása az adatok részhalmazának (D) minden egyes megfigyelésre való újbóli meghatározásával; és

a rendszerben lévő meghibásodások kezdetének jelzése az aktuális megfigyeléssel szembeni modellbecslés vizsgálatával.

2. Az 1. igénypont szerinti eljárás, ahol a hivatkozási adathalmaz (H) biztosítása továbbá tartalmazza különböző időpontokhoz való megfigyelések vételét.
3. Az 1. igénypont szerinti eljárás, ahol a modellezett rendszerre vonatkozó aktuális megfigyelés biztosítása továbbá tartalmazza a rendszernek érzékelők sokaságával való figyelését.
4. Az 1. igénypont szerinti eljárás, ahol:

hivatkozási adathalmaz biztosítása továbbá tartalmazza információ vételét, ahogy megfelel az információforrások első sokaságának; és

aktuális megfigyelés biztosítása továbbá tartalmazza információ vételét, ahogy megfelel az információforrások második sokaságának.
5. A 4. igénypont szerinti eljárás, ahol az információforrások második sokasága legalább részben ugyanaz, mint az információforrások első sokasága.
6. Az 5. igénypont szerinti eljárás, ahol az információforrások második sokasága legalább részben ugyanaz, mint az információforrások első sokasága, de nem teljesen beleértve az információforrások összes első sokaságát.
7. Az 1. igénypont szerinti eljárás, ahol az aktuális megfigyelés összehasonlítása a hivatkozási adathalmazzal (H) az adatok részhalmazának (D) meghatározására továbbá tartalmazza a hasonlóság függvényként való meghatározását, legalább részben, a következőkkel:

a hasonlóság skaláris tartományként való meghatározása, mindkét végén határolva;

a hasonlóság szintjének meghatározása két azonos bemenetre, amely tartalmaz egy értéket, amely megfelel a skaláris tartomány egyik végének; vagy

növelendő hasonlósági érték abszolút értékének biztosítása, amint a két bemenetes megközelítés azonos.

8. Az 1. igénypont szerinti eljárás, ahol az aktuális megfigyelés összehasonlítása a hivatkozási adathalmazzal (H) adatok részhalmazának (D) meghatározásához tartalmazza a hivatkozási adathalmaz hivatkozási megfigyelései és az aktuális megfigyelés közötti hasonlóság kiszámítását, ahol nem az éppen összehasonlított megfigyelésekben lévő összes elemet használják a hasonlóság meghatározásához.
9. Az 1. igénypont szerinti eljárás, ahol egy hivatkozási adathalmaz (H) biztosítása továbbá legalább a következők egyikét tartalmazza:

a modellezett rendszerre vonatkozó nem érzékelős méréseket érintő megfigyelések vétele, a nem érzékelős mérések legalább egyet tartalmaznak a statisztikai adatok, a demográfiai adatok hálózati forgalmi statisztikája, a biológiai sejtszámok vagy a minőségi mérések közül; és

Az adott rendszerhez tartozó legalább egy környezeti állapotra vonatkozó információ vétele.
10. Az 1. igénypont szerinti eljárás, amely továbbá tartalmazza az aktuális megfigyelésnek a hivatkozási adathalmazzal (H) való összehasonlítását adatok részhalmazának (D) meghatározására, tartalmazza tanult megfigyeléseknek a hivatkozási adathalmazból (H) való kiválasztását, mint funkciót, legalább részben, legalább egy változó alapján, amely nem a modell bemenete vagy kimenete.
11. A 10. igénypont szerinti eljárás, ahol a legalább egy változó tartalmaz egy környezeti feltétel változót.
12. A 10. igénypont szerinti eljárás, az aktuális megfigyelés összehasonlítása a hivatkozási adathalmazzal (H) adatok részhalmazának (D) meghatározására, tartalmazza a legalább egy változó használatát a hivatkozási adatok egy részének a részhalmazba foglalásból való kihagyására és egy előre meghatározott hasonlósági szint használatát a hivatkozási adatoknak a részhalmazba foglalásra való kiválasztásához abból, ami marad a legalább egy változó általi kihagyás után.

13. Az 1. igénypont szerinti eljárás, amely továbbá tartalmazza a hivatkozási adathalmaz (H) módosítását.

14. A 13. igénypont szerinti eljárás, ahol a hivatkozási adathalmaz módosítása továbbá a következőket tartalmazza:

legalább egy figyelt rendszerváltozó azonosítása azonosított változóként, beleértve legalább egy környezeti változó azonosítását;

annak meghatározása, hogy az azonosított változó mikor lép túl egy értéktartományt az azonosított változóra, ahogy az értékek jelenleg bele vannak foglalva a hivatkozási adathalmazba; és

a hivatkozási adathalmaz legalább részbeni módosítása az értéktartományt túllépő azonosított változó függvényeként.

15. A 14. igénypont szerinti eljárás, ahol a hivatkozási adathalmaz legalább részbeni módosítása az értéktartományt túllépő azonosított változó függvényeként továbbá tartalmazza további adatoknak a hivatkozási adathalmazhoz való hozzáadását.

16. A 14. igénypont szerinti eljárás, ahol a hivatkozási adathalmaz legalább részbeni módosítása az értéktartományt túllépő azonosított változó függvényeként továbbá tartalmazza új adatok helyettesítését a hivatkozási adathalmazban meglévő adatok számára.

17. A 14. igénypont szerinti eljárás, amely továbbá a következőket tartalmazza:

amikor az azonosított változó túllép egy értéktartományt az azonosított változóra, ahogy az értékek jelenleg bele vannak foglalva a hivatkozási adathalmazba, annak meghatározása, hogy a hivatkozási adathalmazt mindazonáltal kell-e módosítani.

18. A 17. igénypont szerinti eljárás, ahol annak meghatározása, hogy a hivatkozási adathalmazt mindazonáltal nem kell módosítani, továbbá tartalmazza annak meghatározását, hogy a hivatkozási adathalmazt már előre meghatározott számú alkalommal módosították.

19. A 17. igénypont szerinti eljárás, ahol annak meghatározása, hogy a hivatkozási adathalmazt mindazonáltal nem kell módosítani, továbbá tartalmazza annak meghatározását, hogy az adott rendszer valószínűleg hibát mutat.

20. Berendezés rendszerállapot-figyelésre, amely a következőket tartalmazza:

első eszköz hivatkozási adathalmaz (H) biztosítására, amely egy modellezett rendszer olyan érzékelőitől tanult megfigyelések sokaságát tartalmazza, amelyek jellemzik a modellezett rendszer dinamikus viselkedését, ahol a hivatkozási adathalmaz (H) mátrix formájában van, a mátrix egyes oszlopaival egy-egy megfigyelést ábrázolva, illetve mindegyik sor egy önálló érzékelőből való értékeket ábrázol;

második eszköz a modellezett rendszerre vonatkozó aktuális megfigyelés biztosítására;

harmadik eszköz az aktuális megfigyelésnek a hivatkozási adathalmazzal (H) való összehasonlítására, hasonlósági műveleti jelet használva, hogy hasonlósági pontszámot adjon a hivatkozási adathalmazban (H) lévő minden egyes tanult megfigyelésért, és ha a hasonlósági pontszám egy tanult megfigyelésre egy küszöbérték fölött van, vagy az összes tanult megfigyelésre a legmagasabb hasonlósági pontszám előre meghatározott számainak egyike, beleértve a hivatkozási adathalmaz (H) részhalmazában (D) lévő tanult megfigyelést.

negyedik eszköz a rendszer modelljének az aktuális megfigyelés és a hivatkozási adathalmazból (H) származtatott adatok aktuális részhalmaza (D) alapján való kiszámítására, ahol a modell kiszámítása tartalmazza egy modellbecslés előállítását, amely az adatok részhalmazának (D) súlyozott összetételét tartalmazza;

ötödik eszköz a modellezett rendszerre vonatkozó egymást követő aktuális megfigyelések sorozatának biztosítására;

hatodik eszköz a modellnek az egyes új megfigyelésekre az adatok részhalmazának (D) újbóli meghatározásával való újraszámítására; és

hetedik eszköz a rendszerben lévő meghibásodások kezdetének az aktuális megfigyeléssel szembeni modellbecslés vizsgálatával való jelzésére.

21. A 20. igénypont szerinti berendezés, ahol a hivatkozási adathalmaz (H) biztosítására való első eszköz különböző időpontokra vonatkozó megfigyelések vételére való eszközt tartalmaz.
22. A 20. igénypont szerinti berendezés, ahol a modellezett rendszerre vonatkozó aktuális megfigyelés biztosítására való második eszköz eszközolet tartalmaz a rendszernek érzékelők sokaságának segítségével való figyelésére.
23. A 20. igénypont szerinti berendezés, ahol:
 - a hivatkozási adathalmaz biztosítására való első eszköz továbbá információ vételére való eszközt tartalmaz, ahogy megfelel az információforrások első sokaságának; és
 - az aktuális megfigyelés biztosítására való második eszköz információ vételére való eszközt tartalmaz, ahogy megfelelő az információforrások második sokaságának.
24. A 23. igénypont szerinti berendezés, ahol az információforrások második sokasága legalább részben ugyanaz, mint az információforrások első sokasága.
25. A 24. igénypont szerinti berendezés, ahol az információforrások második sokasága legalább részben ugyanaz, mint az információforrások első forrása, de nem teljesen beleértve az információforrások összes első sokaságát.
26. A 20. igénypont szerinti berendezés, ahol a harmadik eszköz, amely az aktuális megfigyelésnek a hivatkozási adathalmazzal (H) való összehasonlítására, adatok részhalmazának (D) meghatározására szolgál, tartalmaz eszközt a hasonlóság, mint függvény legalább részbeni meghatározására, a következők által:
 - a hasonlóság skaláris tartományként való meghatározása, mindkét végén határolva;
 - hasonlósági szint meghatározása két azonos bemenetre, mint olyan értéket tartalmazó, amely megfelel a skalári tartomány egyik végének; vagy

hasonlósági érték abszolút értékének növelése, mivel a két bemenetes megközelítés azonos.

27. A 20. igénypont szerinti berendezés, ahol a harmadik eszköz az aktuális megfigyelésnek a hivatkozási adathalmazzal (H) való összehasonlítására adatok részhalmazának (D) meghatározásához tartalmaz eszközt a hivatkozási adathalmaz hivatkozási megfigyelései és az aktuális megfigyelés közötti hasonlóság kiszámítására, ahol nem az éppen összehasonlított megfigyelésekben lévő összes elemet használják a hasonlóság meghatározásához.

28. A 20. igénypont szerinti berendezés, ahol az első eszköz egy hivatkozási adathalmaz (H) biztosítására továbbá legalább a következők egyikét tartalmazza:

eszköz a modellezett rendszerre vonatkozó nem érzékelős méréseket érintő megfigyelések vételére, a nem érzékelős mérések legalább egyet tartalmaznak a statisztikai adatok, a demográfiai adatok hálózati forgalmi statisztikája, a biológiai sejtszámok vagy a minőségi mérések közül; és

eszköz adott rendszerhez tartozó legalább egy környezeti állapotra vonatkozó információ vételére.

29. A 20. igénypont szerinti berendezés, ahol a harmadik eszköz az aktuális megfigyelésnek a hivatkozási adathalmazzal (H) való összehasonlítására, adatok részhalmazának (D) meghatározására, tartalmazza tanult megfigyeléseknek a hivatkozási adathalmazból (H) való kiválasztását, mint funkciót, legalább részben, legalább egy változó alapján, amely nem a modell bemenete vagy kimenete.

30. A 29. igénypont szerinti berendezés, ahol a legalább egy változó tartalmaz egy környezeti feltétel változót.

31. A 29. igénypont szerinti eszköz, ahol a tanult megfigyeléseknek a hivatkozási adathalmazból (H) történő kiválasztására, mint funkcióra való eszköz, legalább részben, legalább egy olyan változó alapján, amely nem tartalmazza a hivatkozási adathalmaz (H) megfigyelését és az aktuális megfigyelést, tartalmaz eszközt, hogy a legalább egy válto-

zót arra használja, hogy kihagyja a hivatkozási adatok egy részét a részhalmazba való befoglalásból és előre meghatározott hasonlósági szintet használjon, hogy hivatkozási adatokat válasszon ki a részhalmazba való befoglalásra, amiből maradt a legalább egy változó általi kihagyás után.

32. A 20. igénypont szerinti berendezés, amely továbbá eszközt tartalmaz a hivatkozási adathalmaz (H) módosítására.
33. A 32. igénypont szerinti berendezés, ahol a hivatkozási adathalmaz módosítására való eszköz továbbá a következőket tartalmazza:

eszköz legalább egy figyelt rendszerváltozónak azonosított változóként való azonosítására, beleértve legalább egy környezeti változó azonosítását;

eszköz annak meghatározására, hogy az azonosított változó mikor lép túl egy értéktartományt az azonosított változóra, ahogy az értékek jelenleg bele vannak foglalva a hivatkozási adathalmazba; és

eszköz a hivatkozási adathalmaz legalább részbeni, az értéktartományt túllépő azonosított változó függvényeként való módosítására.

34. A 33. igénypont szerinti berendezés, ahol a hivatkozási adathalmaz legalább részbeni, az értéktartományt túllépő azonosított változó függvényeként történő módosítására való eszköz továbbá tartalmaz eszközt további adatoknak a hivatkozási adathalmazhoz való hozzáadására.
35. A 33. igénypont szerinti berendezés, ahol a hivatkozási adathalmaz legalább részbeni, az értéktartományt túllépő azonosított változó függvényeként történő módosítására való eszköz továbbá tartalmaz eszközt új adatoknak a hivatkozási adathalmazban meglévő adatokhoz való behelyettesítésére.
36. A 33. igénypont szerinti berendezés, amely továbbá a következőket tartalmazza:

eszköz annak meghatározására, hogy a hivatkozási adathalmazt mindazonáltal nem kell módosítani, amikor az azonosított változó túllép egy értéktartományt az

azonosított változóra, ahogy az értékek jelenleg bele vannak foglalva a hivatkozási adathalmazba.

37. A 36. igénypont szerinti berendezés, amely eszközt tartalmaz annak meghatározására, hogy a hivatkozási adathalmazt módosították-e egy előre meghatározott számú alkalommal.
38. A 36. igénypont szerinti berendezés, ahol az eszköz annak meghatározására, hogy a hivatkozási adathalmazt mindazonáltal nem kell módosítani, továbbá tartalmaz eszközt annak meghatározására, hogy az adott rendszer valószínűleg meghibásodást mutat.