

# データマイニングによる異常検知技術

鈴木 英明, 内山 宏樹, 湯田 晋也

新興国を中心とした経済がけん引役となっている現在、産業成長は国内から海外へとその軸足が向けられている。グローバル市場において各産業分野では技術の伝承と業務の効率化の両方が求められている。国内で培った技術や製品を海外に移転するためには、人材育成と同時に、業務の IT 化を進める必要がある。製品の製造だけでなく、運用を含めた効率化が要求され、製品のライフサイクルにわたる利用者へのサポートが重要である。本稿では、保守業務の IT 化に寄与する異常検知技術について、センサのデータマイニング活用を中心に議論する。

キーワード：データマイニング, 機械学習, 異常検知

## 1. まえがき

新興国を中心とした世界経済の伸びに伴い、わが国の製造業における海外生産高比率および売上高比率は 2011 年で 35% 前後になっており、今後さらに伸びることが予想される [1]。とくに、電力・水道や鉄道などの社会インフラ産業については、機器の納入が中心となっているものの、3 割以上が中国やインドといった新興国への進出に関心をもっている。一方、国内においてはベテランの大量退職がピークを過ぎ、若い世代への技術移転が十分進まないうちに、ベテラン技術者が海外に移って新興国の技術者を育てるという状況が起きている。社会インフラ産業は、製品設備そのものの提供だけでなく、運用や保守といった活用段階を含めたサポートが必要となるため、技術的な難しさもっている。すなわち、運用する人材の育成が重要である。

運用や保守などの人的リソースに依存する事業においては、手順書に従った実践とそのための教育プログラムの充実が必須である。一方で、新興国においてはとくに、優秀な人材の確保は難しく、インフラ設備の需要に人材育成が追いつかないという状況が増加しつつある。この状況を少しでも緩和するためには、システム化・自動化を進めるとともに、人が介在するオペレーションを極小化していく取り組みが必要である。

本稿では、設備監視における自動化あるいはその支援システムとして注目されている状態基準保守とそれ

を実現するための機械学習技術について説明し、従来手法との違いや課題などについて例を交えながら述べる。

## 2. 保守の分類

保守の方法は大きく分けて、事後保守と予防保守の二つがある。JIS では保守 (保全) の方式を JISZ8115-2000 で図 1 のように定めている [2]。

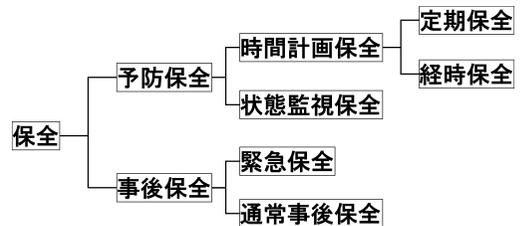


図 1 JIS による保全方式の分類

事後保守 (Breakdown Maintenance) は、対象となる設備あるいは機器が機能停止となる故障後に設備や機器の修理を実施するものである。家電製品などはこの保守方式が多い。予防保守は、対象機器が故障する前に点検や部品交換などの整備を実施し、故障そのものの発生を回避するものである。重要インフラにおける設備や停止による影響が大きい機器などではこちらが一般的である。直感的には、事後保守は故障してから保守を実施するため部品などの寿命を使い切るという意味で直接コストが少なく済み、予防保守は対象機器を止めずにすむというメリットがある。しかしながら、事後保守では、機器が故障する際に原因となった部位だけでなくその影響を受けて他の部位も故障するといったことが起こってしまう可能性がある。また、

すずき ひであき, ゆだ しんや

(株) 日立製作所 日立研究所

〒 319-1292 茨城県日立市大みか町 7-1-1

うちやま ひろき

(株) 日立製作所 横浜研究所

〒 244-0817 神奈川県横浜市戸塚区吉田町 292 番地

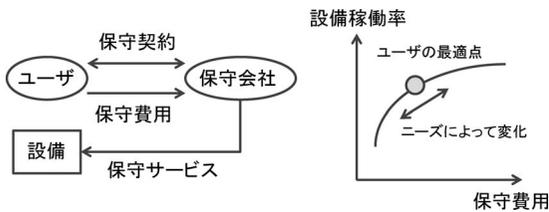


図2 利用者のニーズに合わせた保守サービス

予防保守の場合でも、保守作業中は機器を停止する必要があります。したがって、図2に示すような利用者のニーズに合わせた保守契約の形態を対象機器ごとに決めておくのが一般的である [3]。

機器が故障停止している間は、所望の機能を利用することができないため、例えば電力設備であれば電力が供給されないといった問題が発生する。電力が供給されないことで生じる影響は2次、3次といった損害につながることもある。電力が供給されないことで冷凍設備が使用できず、中のものをダメにしてしまったというような場合である。インフラ設備においては、このような間接的な影響の範囲が大きいという特徴がある。

上述のように故障の影響が大きいインフラ設備は、停止しないように予防保守の措置が取られることが多い。予防保守では定期保守（時間計画保全）が一般的である。これは定期点検義務などの法律の制約による部分もある。時間を基準とした保守では、時間が機器の劣化状態を反映しているとは言えないことも多い。その理由として、時間基準そのものが設計由来である一方で、機器の使い方など時間に依存しない要素が組み込まれていないという点がある。利用者、利用環境、稼働状態などさまざまな条件の下に利用されるにもかかわらず、異なる機器が同じ時間という条件で保守されるという点が難しい状況を作っている。そこで、機器の状態に応じて保守対応を柔軟に変更するという状態監視保全という考え方が生まれ、近年注目されるようになってきている。

### 3. 状態基準保守

前述の状態監視保全は、より一般的には状態基準保守 (CBM, Condition Based Maintenance) と呼ばれる。インフラ設備では監視制御 (SCADA, Supervisory Control And Data Acquisition) システムという形で専任のオペレータが設備の稼働中は状態監視と必要に応じた操作を行うのが一般的である。状態を参照するという意味では状態基準保守も同様であるが、目的が

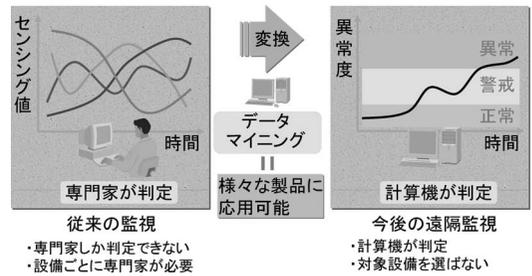


図3 データマイニングによる設備監視

設備や機器の制御ではなく、保守の最適化にある点が異なる。監視制御においても、設備に組み込まれたロジックにしたがって機器の状態異常がアラームによって報知され、オペレータが適切な対応を実施する。しかし、機器の発報するアラームは機器の保護を目的とした警告や機器の故障そのものを知らせるものであり、その段階においてはすでに機器の異常がかなり進んでいることが考えられる。状態基準保守を実現するためには、機器がアラームを発報する前の段階でその状態の「変化」を検知する必要があるが、「変化」をどのように定義あるいは設計すればよいかという新たな課題が生じる。

この課題を計算機の力と数学を用いて解析的に解決しようとするものが、いわゆるデータマイニングである (図3)。データマイニングは検索エンジンを初め、さまざまな分野に应用されているが、最近では状態基準保守に活用しようとする動きが活発になってきている。

### 4. 機械学習

データマイニングという言葉は、多量のデータの中から意味のある、あるいは有用な知見を導き出し、実務に活用するという意味で用いられることが多く、最近ではビッグデータ活用という言い方をされる機会が増えてきた。

このデータマイニングを支える技術のことを、データ処理あるいは多変量解析の分野では、機械学習と呼ぶ。機械学習は、人が行っているパターン認識や経験則を導き出す活動を模して、計算機に訓練データを与え、さまざまなアルゴリズムを用いて規則性を学習させることを指す。機械学習の手法は、PCA (Principal Component Analysis, 主成分分析) [4], クラスタリング [5][6], SVM (Support Vector Machine) [7] など、さまざまな手法が提案されており、その用途も言語処理、画像認識、数値予測、データ分類など多岐にわたる。応用としては、マーケティング、医療、金融、産

業応用などほとんどの分野で活用されるようになって  
いる。

産業分野では、従来から用いられている閾値判定や  
回帰分析などがまだ一般的であるが、これは異常判定  
のロジックが設計時に組み込まれることが多いこと  
による。つまり、物理的に説明ができることが前提で、設  
計時に取得可能なデータも実験室あるいは実証試験時  
の限られた範囲で判断する必要がある。このような設  
計時にモデル化された判定ロジックでは、想定外の異  
常を検知することができないという課題がある。機械  
学習を活用した状態基準保守では、実際に運用中の機  
器から収集したデータに基づいて判定することが前提  
であるため、むしろ現場の状況に合わせた異常判定が  
有効な場面も出てくる。

## 5. 機械学習による異常検知

機械学習による異常検知は、機器のセンサが取得し  
た観測値を活用する。対象により取得できるセンサ数、  
サンプリング間隔などは異なるが、時系列の連続した  
データであることが前提となる。

機械学習は、訓練データとなる正常動作時の稼働デ  
ータを学習させる学習段階と、判定データとなる異常動  
作時の稼働データを診断させる検知段階に分かれる  
(図4)。時系列のセンシングデータは各時刻における  
センシングデータの値を1組のベクトルとして扱う。  
機械学習では、この各時刻のデータを多次元空間上の  
1点として扱い、学習アルゴリズムによって、その空  
間上での分布を学習する [8][9][10][11][12]。

シンプルな利用方法としては、設備が正常に動作し  
ている期間の稼働データを機械学習させ、あとは日々の  
稼働データを入力すれば、異常が検知できる。図5に  
示すように機械学習による異常検知の本質は、学習さ  
せる正常データと診断させるデータとがどれだけ“似  
ているか”あるいは“似ていないか”を適切に数値化で  
きるかという点にある。データが似ていなければ、そ  
の「ずれ」が異常として検知されたことになる。異常  
検知はコンテキストによっては変化点検知と呼ばれる  
こともある。図6に示すように変化点検知の場合は、

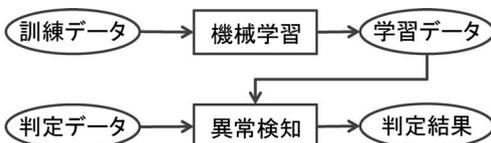


図4 機械学習による異常検知

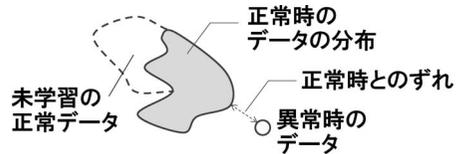


図5 正常データと異常データのずれ

訓練データを時々刻々更新していくという点異なる。

このようにデータさえあれば、設備の詳細な知識が  
なくとも異常を知ることができるように思えるが、実  
際には容易ではない。自然現象とは異なり、設備には  
必ず「制御」が伴い、制御ロジックに応じた挙動をす  
る。すなわち、いかに設備の振る舞いを学習によって  
模擬できるかが重要である。そのために、設備のあら  
ゆる動作状態を訓練データとして取得する必要があり、  
同時に、さまざまな周囲の環境条件も含めた設備の挙  
動を機械学習によって習得する必要がある。もし訓練  
データが不十分であれば、欠損している部分のデータ  
領域での挙動を、正常稼働にもかかわらず、誤って異  
常と検知する可能性がある。最近では訓練データと診  
断時の条件の差を吸収しようとする研究も進んでいる  
が [13]、基本的には網羅的なデータ収集が必要であり、  
この点が設計時にモデル化された判定ロジックとの違  
いになる。

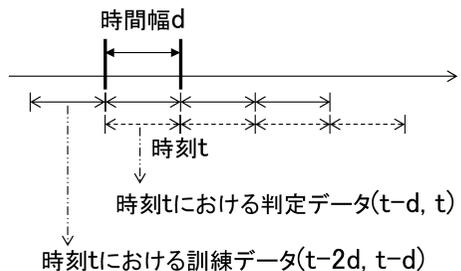


図6 変化点検知におけるデータ選択

## 6. 異常検知の実験例

図7はある機器の実稼働データに基づいて、模擬異  
常データを生成し、機械学習による学習と異常検知の  
実験を行ったものである。訓練データと同じような振  
る舞いをするデータについては、「ずれ」は観測され  
ず、模擬異常を作りこんだ部分については「ずれ」が観  
測されている。以下この「ずれ」のことを異常度と呼  
ぶ。ここで注目すべきはセンサ値がステップ状に変化  
するような状況にも対応している点である。訓練デー

タに含まれていない正常データの範囲では異常度は大きくならず、模擬異常データの部分で大きく振れている。この方法では面倒なセンサごとの閾値設定や機器の制御モードごとの切り分け判定は必要がない。

しかしながら、機械学習アルゴリズムに与える訓練データが不適切であると、正しくない判定結果が出力されるようになる。図8では訓練データが不十分な

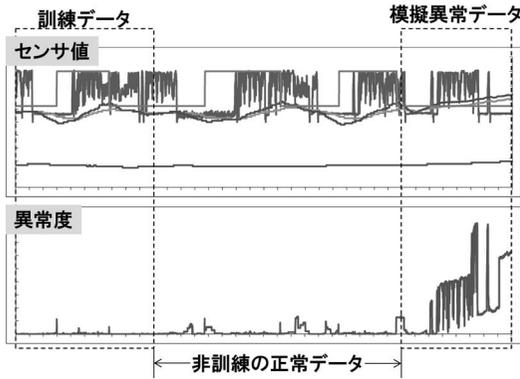


図7 模擬異常データとその検知

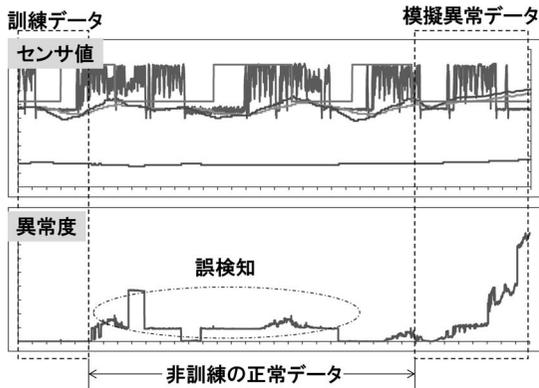


図8 誤検知

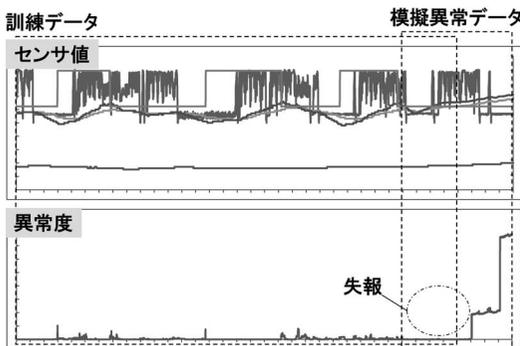


図9 失報

め、正常であるにもかかわらず異常と検知される誤検知の例を示している。検知された理由は、与えられた訓練データに該当するセンサデータの組合せが存在しなかったことを示している。また、図9に示す例は、与えた訓練データの中に検知すべき異常データが含まれており、異常な状態も正常と学習したため、異常を正しく検知できていない例である。このように、訓練データの適切な選定は、検知性能を左右する重要なポイントである。FDAやLFDA[14]、SVMでは、正常であるか異常であるかといったラベル情報付きの訓練データが準備できれば、最適な境界条件を導き出してくれるが、設備機器の異常検知においては、必ずしも異常データが初めから準備できるわけではない。したがって、正常時のデータしか存在しないという条件を前提とすることが必要であると考えられる。

訓練データの選択ミスは、前述の変化点検知のような場合は特に問題となる。変化点検知では、訓練データが時々刻々と更新されていくため、訓練データとして与える範囲が少なすぎると誤検知が発生し、多すぎると失報の可能性を増やしてしまう。つまり、指定する時間幅の最適化が課題となる。徐々に進行するような劣化の場合には指定する時間幅との関係で検知できないということが起こりうる。

## 7. 異常事象の推定

前節までに示した異常検知技術により、機器の状態が正常か異常かを判定することが可能となったものの、さらなる保守業務の最適化を図るためには、どのような異常が発生しており、この先どのような故障につながるのかといった情報を事前に知る必要がある。このような事前情報により、異常の内容に従った点検整備や万が一の故障時に必要な交換部品の準備などを事前に行うことができると考えられる。

図10は異常事象の推定において利用する異常寄与度の概念を表している。訓練データも判定データも多変量ベクトルであり、異常検知では前述のようにその

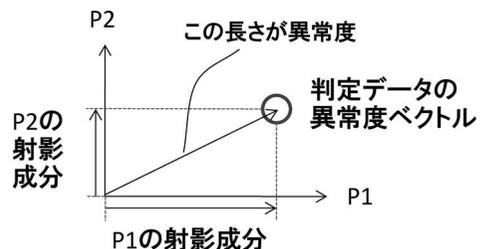


図10 異常寄与度の算定

ベクトルの差を異常として識別する。これらのベクトル間の差分ベクトルを異常度ベクトルとしたときに異常度ベクトルの長さが異常度に相当する。ここで、異常度ベクトルの各センサ値を表す軸への射影成分を考え、これを異常寄与度と定義する。異常度が正常データとの「ずれ」を表しているとすれば、異常寄与度は「ずれ方」を表していると言える。

この異常寄与度（ベクトル）について、異常発生時のデータだけについて切り出してサンプリングし、そのサンプルデータについて異常寄与度の算出と、機器の点検結果から得られた実際の異常事象とをセットでデータベース化することを考える。この異常事象のデータベースを構成することができれば、過去の異常事象から現在あるいは今後起こる異常事象を推定できる可能性がある。図 11 に示すように、異常事象の判定は、判定データの異常寄与度ベクトルと過去の異常事象ごとの異常寄与度ベクトルとの内積を計算し、それらが最小となる異常事象を推定異常事象として出力する。

また、図 12 に示すように、保守員の実施した作業結果を逐次異常寄与度と共に記録してデータベース化し、随時アップデートすることも可能である。このよ

$$\text{推定異常事象ID} = \text{argmin}(\theta_1, \theta_2, \dots, \theta_N)$$

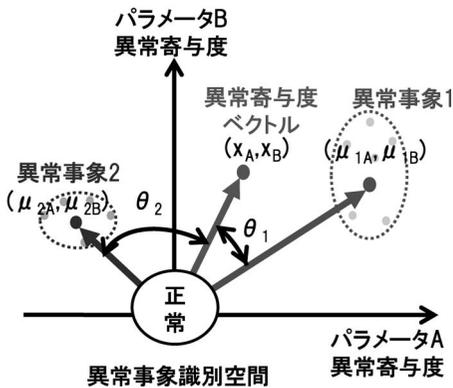


図 11 異常寄与度の算定

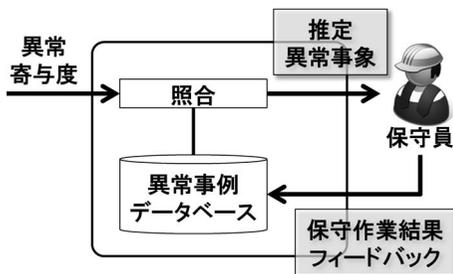


図 12 異常事例データベース

うなデータベースを活用することにより、データベースに格納されている情報の精度を高め、次に起こりうる異常事象を推定するという循環を構成することが可能になる。これにより、過去の保守作業による経験が知識化される可能性が得られる。

## 8. 今後の展開

保守業務の IT 化のためには、本稿で述べたような技術をパッケージ化し、広く利用したいと考えるユーザーや関係者に提供していく必要がある。そのため、SaaS(Software as a Service) のような形態で誰もが Web サービスとして利用可能になるような仕組みが最適であるとわれわれは考えている。われわれはこのようなデータを付加価値の高い知識へと処理する情報サービスとして KaaS(Knowledge as a Service)[15] を提案している。KaaS の 1 つの実現例として、本稿で述べたような保守業務のための情報サービスが考えられる。図 13 に示すように、対象機器のセンシングデータをクラウドサーバに集約し、異常検知のようなデータ処理による付加価値を情報サービスとして提供することで、ユーザーが容易に保守診断サービスを受けたり、保守員どうしが互いの保守経験をクラウド経由で共有したりすることが可能になる。

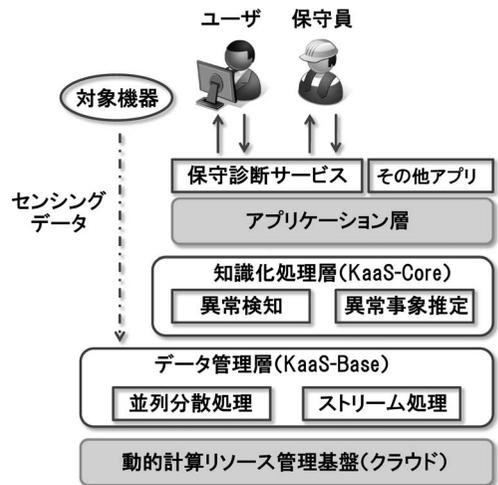


図 13 保守診断サービス

このような情報サービスを行うメリットは、ユーザーと提供者の両方に発生する。ユーザーは大幅な初期投資をすることなく、一定の利用料を支払うことで最新の機能を利用できる。クラウド上に構築されているので、診断対象の設備が増えてもその増加量に応じてサーバ

容量を増やせばよいための、投資効率が良くなる。提供者はユーザから預かる大量のデータに基づいてアルゴリズムの改良や知識ライブラリの充実を図ることができるとともに、ユーザからのニーズを吸い上げて新たな知識処理サービスを展開することができる。

今後は、保守診断サービスのための技術の充実を図るとともに、関連する他のデータや情報サービスのための技術開発についても取り組んでいきたい。

#### 参考文献

- [1] 国際協力銀行, 「わが国製造業企業の海外事業展開に関する調査報告」 <http://www.jbic.go.jp/ja/about/press/2011/1202-01/besshi1.pdf>, 2011.
- [2] 日本規格協会, [http://www.jsa.or.jp/stdz/edu/pdf/b2/2\\_11.pdf](http://www.jsa.or.jp/stdz/edu/pdf/b2/2_11.pdf), 2007.
- [3] 「標準サービスレベルマネジメント」, 2003, オーム社.
- [4] Jolliffe, I. T., *Principal Component Analysis* (Second ed.). Springer, 2002.
- [5] MacQueen, J. B., Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Symposium on Math, Statistics, and Probability Berkeley, CA: University of California Press.* pp. 281–297, 1967.
- [6] T. Kohonen, “Self-Organizing Maps,” Springer, 1996.
- [7] V. Vapnik, “The Nature of Statistical Learning Theory,” Springer, 1995.
- [8] 乾 稔, 矢入健久他, 「次元削減の再構成誤差を用いた異常検知手法の比較」, JSAI2009.
- [9] 藤巻遼平, 「異常検出サポートベクトルマシン」, FIT2008.
- [10] 井手 剛, 「近傍保存原理による異常検知」, IBIS2007.
- [11] 河原吉伸, 矢入健久他, 「部分空間法に基づく変化点検知アルゴリズム」, 人工知能学会論文誌, Vol. 23, No. 2, pp. 76–85, 2008.
- [12] 杉山 将, 「密度比に基づく機械学習の新たなアプローチ」, 統計数理, Vol. 58, No. 2, pp. 141–155, 2010.
- [13] 杉山 将, 「共変量シフト下での教師付き学習」, 日本神経回路学会誌, Vol. 13, No. 3, pp. 111–118, 2006.
- [14] M. Sugiyama, “Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis,” *JMLR*, Vol. 8, pp. 1027–1061, 2007.
- [15] 植田良一, 佐藤嘉則他, 「社会インフラの革新に貢献する知識化サービス基盤 KaaS」, 日立評論, 2010.