

センサー・データによる状態監視技術

データの取得・蓄積環境の整備と、熟練エンジニアの高齢化と不足を背景にして、センサー・データを活用した状態把握と保守の半自動化に注目が集まっています。ものづくりを得意にしてきた日本においては、自動状態監視はいわゆるビッグデータ活用の、いわば本丸と言うべき重要な領域です。状態監視の中で中核となる技術は異常検知技術ですが、多くのセンサー・データを同時に監視して異常の兆候を見いだすには大変高度な解析技術が必要とします。そのため、各応用領域において多大な労力を費やして個別の工夫に基づく状態判定ルールを構築している状況だと思われます。

この解説では、状態判定ルールの自動構築に向け、多変量時系列データからの異常検知技術の動向を解説します。とりわけ、機械学習の分野で 2000 年代半ばに提案された技術的ブレイクスルーが、従来知られていた課題の多くを解決していることを説明します。

1. はじめに

日本はこれまで、緻密な共同作業を得意とする質の高い労働力に支えられて、自動車、鉄道、造船、電機、電子、建築、土木など、「もの」が関わる大変多くの分野で世界をリードする存在でした。しかしここ 10 年ほど、日本をめぐる状況は大幅に変わりつつあります。韓国、中国、台湾を中心とする新興国がものづくりの国際分業の主要な部分を担うようになり、日本は急速にその立ち位置を見失いつつあるように見えます。

そのような中で、ハードウェアに付帯した保守もしくは監視業務に、最近大きな注目が集まっています。これには少なくとも 2 つの理由があります。

第 1 に、保守業務の自動化を進めることでワランティに関わるコストを下げ、利益率を改善できる可能性があるという点です。これについては、**時間基準保全** (time-based maintenance) から**状態基準保全** (condition-based maintenance) への進化という言い方で最近よく知られるようになりました。すなわち、パーツ類の定期交換を基本とする保守から、センサーなどの客観的な情報を基に保守対象機器の状態を把握し、安全性が担保される限り「使い倒す」という戦略への進化です。

第 2 に、保守サービスそれ自体の事業化と、保守サービスに関連して取得したデータを用いた新ビジネス創出の期待が高まっているという点が挙げられます。例えば、カー・ナビゲーション・システムや、インターネットに接続された家電を情報源にした統合的な情報基盤の構想が、すでにいくつかの企業により提案されています。しかしながら、現状では、データをビジネスに変えるための決定版と言えるビジネス・モデルが確立してい

るわけではなく、生データを価値ある 2 次データに変換するためにどのようなデータ解析技術が使えるかについて、検討が行われているという段階だと思えます。

やや引いた視点で考えるならば、これは最近の流行語となっている「ビッグデータ」の日本的展開と言えるでしょう。ビッグデータの名の下に報じられる成功事例のほとんどは、いわゆる**トランザクション・データ** (売り上げデータ) に基づくもので、ものづくりに密接に関係するセンサー・データを対象にしたものは多くはありません。

しかし、例えば海事業界では、マーケティング的領域よりはむしろ、設計開発、製造、運用といった領域で、センサー・データを駆使した業務革新が可能です。ものづくりと密接に関連したセンサー・データの活用、とりわけ職人芸的ノウハウとデータ解析技術の結合は、日本におけるビッグデータ解析の本丸と言うべきであり、ぜひとも日本が世界を先導したい分野です。

本解説では、センサー・データに基づく保守監視業務で最も重要な技術的課題となっている異常検知技術について、最新技術の動向を概観します。

日本では、1980 年代に日本で成功を収め、その後世界中に広まった品質工学の手法に源を発する「MT 法 (Mahalanobis-Taguchi method)」[1] が試されることが多いように思われます。しかし、MT 法には実用上の課題が多く、それが本格的な状態監視業務の展開のボトルネックになってきた感は否めません。実はそれらの課題の多くは、品質工学とはほぼ独立に関係してきた統計的機械学習と呼ばれる分野の理論により解決されています。本解説の目的の一つは、そのような現代の発展を紹介することです。

2. 状態監視技術の現状

2-1 状態監視の4つの技術分野とその成熟度

それぞれの業種に合った状態監視技術を考える上で、状態監視技術の成熟度を俯瞰的に把握しておくことは大変重要です。図1にセンサー・データの活用に関する4つの技術分野と、それぞれの技術成熟度を模式的にまとめたものを示します。



図1. センサー・データ活用のための技術分野とその成熟度

センサー・データの活用に関する技術分野には、大きく分けて計測、通信、蓄積、解析の4つがあります。このうち、センシングそれ自体（計測）とデータを収集する手段（通信）については、今なお活発に技術開発が行われているにせよ、少なくとも何らかの選択肢が存在するという意味において、相対的に成熟度が高いと思われる。

それに比べて、蓄積（データベース）と解析の技術は、成熟度においてまだ低い段階にあると言わざるを得ません。中でも解析技術には課題が多く、センサー・データを元にした状態監視という分野に限っていても、後述する通り、手法が確立していると言えるのは、非常に理想的な状況に限られます。

現代のデータベースの主流をなす関係モデルという方式は、もともとトランザクション・データを対象に、その格納と高速検索を想定して開発されたため、数値データが時々刻々来るような状況への対応において困難があると認識されています。そのような問題意識の下、例えば「列格納方式（column-store）」と呼ばれる新しいデータ構造が盛んに研究されています [2]。

データベースの設計は、データに対してどのような解析を行うかという点と密接に関係しているため、特定のデータ解析技術、あるいはそれに基づくビジネス・モデルに最適化した新たなアーキテクチャーが提案される余地はまだあります。

2-2 センサー・データからの異常検知技術の課題

状態監視においては、センサー値から異常の発生を判断するという作業、すなわち、異常検知が本質的に重要です。

図2に示すように、一般的に異常検知というときに想定される状況は「単純な状況」だと思います。例えば、普段マイナス20度に保たれている冷凍庫の温度計が、あるときプラス5度になっていたら誰も異常を疑います。しかし産業上の状態監視にお

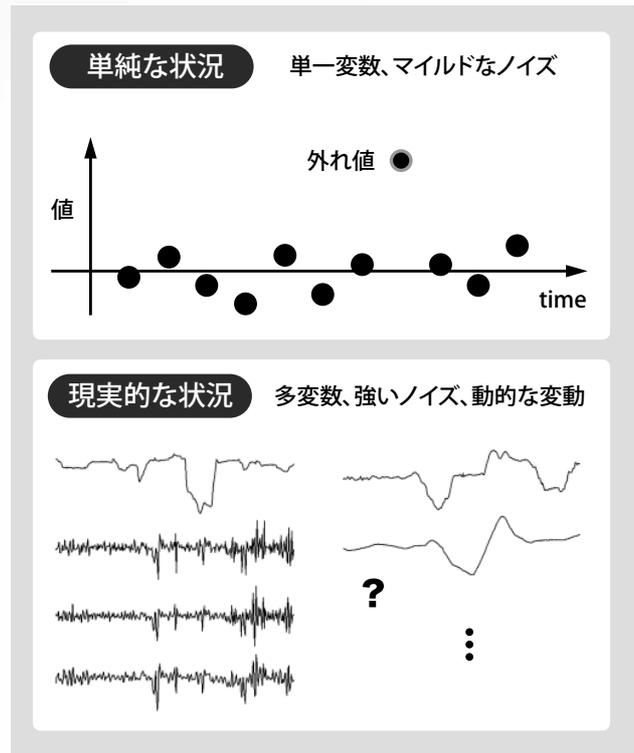


図2. センサー・データからの異常監視における典型的状況

いては、そういう単純な状況はむしろまれで、次のような課題が常に存在します。

1. 「その値だけを見ていれば完璧に異常が判別できる」というようなセンサーは通常存在せず、複数のセンサーを同時に監視しなければならない（例えば、車の健康度を表す単一のセンサーは存在せず、走行距離、オイルの量などを組み合わせて総合的に判断するしかない）。
2. 監視すべき量が多い。特に変数の組み合わせにおいて起こる異常（例えば、アクセルを踏んでも回転数が上がらない）については、すべての変数の組み合わせを監視しようとすると、監視対象が爆発的に増大する。
3. センサー値にはノイズがあり、また外乱による出力変動が避けられない。ノイズや外乱を異常と弁別するのは一般に難しい（例えば、夏にはエアコンを使うため発電機の出力が上がるが、出力上昇自体は異常ではない）。

図から分かる通り、センサー・データによる状態監視を行うということは、いわば多次元時系列のふるまいを一つのパターンとして捉え、そのパターンから異常を判定することに他なりません。これは一般に困難な課題であるため、長らく理想化された状況の下で異常検知の手法が開発されてきました。次章ではそのような代表的手法を紹介します。

3. ホテリング理論と MT 法

3-1 ホテリング統計量 (1 変数)

多変量解析における異常検知は、米国の経済学者 Harold Hotelling が創始した理論がその中核となっています。例えば、半導体業界はセンサー・データによる状態監視技術が最も進んだ業界の一つですが、そこではホテリング統計量と呼ばれる異常の指標が標準的に使われています。例えば 1 種類のセンサーしかない場合 (図 2 の「単純な状況」)、ホテリング統計量は次のように定義されます*1。

$$(1) \quad T_{M=1}^2 = (x - m)^2 S^{-1}$$

ここで、 x は異常かどうか確かめたい観測値、 m と S はそれぞれ、それ以前の (異常ではない) 観測値から計算された平均値と分散、センサーの数を M と表します (今は $M=1$)。

ホテリング統計量はしばしば (定数倍は別にして) マハラノビス距離とも呼ばれます。実際、 $x - m$ という量は、 m から x への距離に他ならず、距離という呼び方は納得できます。この距離は物理量の次元により異なるはずですので、標準偏差で割ることで無次元化したものを外れの度合いとして採用しています。

統計学においてホテリング統計量が関心を引いた理由は、観測値が正規分布に従うという仮定の下、ホテリング統計量の確率分布が厳密に導出できるからです。実際、式 (1) の平方根に当たる量は t 分布と呼ばれる分布に従い、それゆえ、「確率 5% 未満でしか起こらない値なら、異常と見なそう」というように、客観的に異常判定基準を置くことが原理的には可能です。

しかし実際には、話はさほど簡単ではありません。動的な変動が強いセンサー・データでは、 t 分布に基づいて異常判定基準を置いても誤警報が多く、使い物にならないことがほとんどです。正規分布に従うというもとの仮定が正しくないからです。

*1 : 厳密に言えば過去の観測値の数 N と、センサーの数 M に依存する係数が右辺にはかかりますが、 $N \gg M$ の状況では近似的に係数は 1 となります。

3-2 ホテリング統計量 (多変数)

先に述べたように、現実的な状況では、異なるセンサー値同士の組み合わせで起こる異常にも対処しなければなりません。そのため、式 (1) を、 $M > 1$ (1 つ以上のセンサー数) に拡張することを考えます。結果から示せば、多変数の場合のホテリング統計量は次のように定義されます*2。

$$(2) \quad T_M^2 = (\mathbf{x} - \mathbf{m})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m})$$

この場合、観測値およびその平均値は M 次元の (列) ベクトルとなります。また、式 (1) における分散は、 $M \times M$ 次元の共分散

行列 \mathbf{S} になり、 \mathbf{S}^{-1} というのはただの逆数ではなくて、逆行列となります。それぞれの定義は次の通りです。

$$(3) \quad \mathbf{m} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}^{(n)}, \quad \mathbf{S} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}^{(n)} - \mathbf{m})(\mathbf{x}^{(n)} - \mathbf{m})^T$$

ただし、 N は過去に観測した観測値の数、上付きの T はベクトルまたは行列の転置を表します。

式 (2) のホテリング統計量が、1 変数の場合の自然な拡張になっていることは式の形から明らかです。観測値 \mathbf{x} の外れ度合いは、この場合も単に平均値からのずれで測られ、共分散行列を使ってそれをある意味で規格化します。この場合、変数が複数あるため、 \mathbf{S} は単なる無次元化よりもより重要な役割を担います。

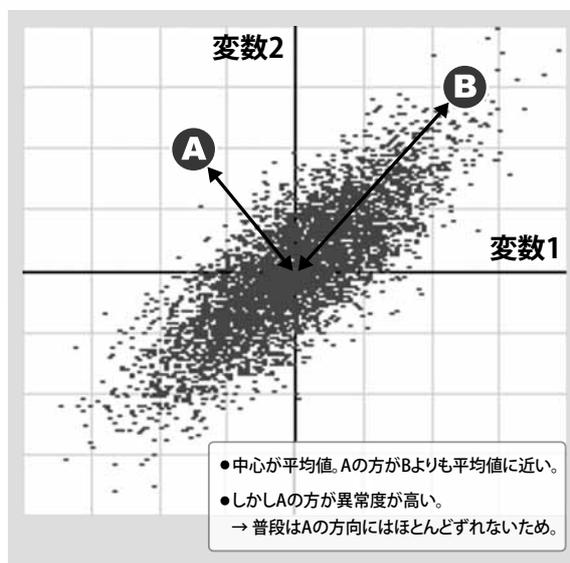


図 3. マハラノビス距離の性質の説明 (2 変数の場合)

図 3 に、 $M=2$ の場合を例示しました。小さな点の一つ一つが、時々刻々出てくるセンサーの観測値で、これらはいずれも監視対象機器が正常の場合に観測されたものと想定します。センサーは 2 つですので、観測値は 2 次元平面の 1 点として表されます。この例では、A の方向と B の方向でばらつき的大小が異なることがわかります。このような場合、B 方向では、A 方向に比べて多少の値の変動は「大目に見る」ことが合理的です。多変数のホテリング統計量における \mathbf{S}^{-1} の項は、このような変数ごとの相関関係とばらつきの異方性を正規化する働きを持っています。それにより、原理的には誤警報を減らす効果期待できます。

多変数のホテリング統計量の計算には、ベクトルと行列からスカラーを作る複雑な演算を必要とします。しかしやや驚くべきことに、統計学の理論によれば、式 (2) (の定数倍) が F 分布という既知の分布に従うことが証明できます [3]。それゆえ、少なくとも原理的には 1 変数の場合と同様に、客観的なパーセント値を用いて異常判別ができます。F 分布に従うことの証明には、標本共分散行列がどう分布に従うのか (そもそも「行列の分布」

とは何か)などの深遠な問いに答える必要があり、多変量解析の代表的な理論的成果になっています。

しかしその美しい理論的な結果とは裏腹に、ホテリング統計量による異常判別には、実用上の多くの課題があります。これは、MT法の課題と併せて次項でまとめて述べましょう。

*2:単純化のため付加係数を省略しています。また、不偏分散でなく、最尤推定値を使っています。

3-3 MT法の課題

状態監視はその発想において品質工学との親和性が高いため、ホテリング理論は、日本においては時折MT法(もしくはMTシステム(MTS))という呼び方をされます。MT法とは、ホテリング統計量に基づく外れ値検出手法と、分散分析などに基づく異常変数選択手法を組み合わせた手法です。MT法には様々なバリエーションがありますが、Taguchi-Jugulum [1]で述べられている内容を状態監視の文脈で解釈すると次の通りです。

- 正常データに基づき、式(3)でSとmを計算し、それらを式(2)に入れる。
- 式(2)において、 \mathbf{x} に観測値を代入する。値がある閾値を越えたものを異常と判定する。
- その際、式(2)から各変数を順番に取り除き、式(2)の値の減少を記録する。その値を各変数の異常度とする。

オリジナルのMT法では、2水準(変数を含める or 含めない)の割り付けに関する実験計画法の手法を用い、「SN比」と呼ばれる指標で変数の取捨選択の判断を行っています。SN比という量は、状態監視の文脈ではホテリング統計量を対数変換したものと定義されますが、その理論的根拠は必ずしも明確ではありません。

このほかにも、MT法は統計学的には直ちに正当化しにくい点をいくつか含んでおり、例えば、宮川 [4]は「MTSには理論的に未解決な要素がいくつかあり、それらがほとんど議論されないまま教条的に使われてきた」と注意を喚起しています。

実用上の観点からは、MT法の問題は次のようにまとめられます。

1. ホテリング統計量における逆行列 \mathbf{S}^{-1} の計算が数値的に不安定で、変数の数が100くらいで計算困難となる。
2. ホテリング統計量は、M次元系を単一の指標で表すため、変数が多くなるにつれ、少数の変数のみに生ずる異常がかき消される傾向がある。
3. 異常度の定義が、一定の平均値からのずれという形で与えられているため、値が動的に変化するシステムでは適用が困難である。

第1の点については、Taguchiらはデータ行列のQR分解に基づく計算法を提案しています [1]。それにより数値的には安定しますが、その代償として、元の変数を混合した人為的な変数を扱わざるを得なくなるという問題があります。第2の点については、手作業などで事前に変数をグルーピングしておかなければならず、センサー数が数百になるとその手間が作業の障害になりがちです。第3の点は、適切な前処理でトレンド成分を除去するなどである程度対応できますが、一般的には、MT法が妥当なのは、非常によく制御されて値がほぼ静的に一定値に留まるようなシステムのみとされています。

4. IBM Anomaly Analyzer for Correlational Data

上に述べた従来法の課題を解決するために必要な異常検知技術の要件をまとめると次の通りです。

- 変数の数が数千から数万の程度でも安定して異常度の計算が行えること。
- システムのモジュール構造(変数のグループ)を自動認識した上で個別の変数の異常度を算出できること。
- 観測値が動的に変化するようなシステムに対しても、異常検知が可能であること。

IBMの研究部門の開発した異常検知エンジンIBM Anomaly Analyzer for Correlational Data(以下、ANACONDA) [5]は、これらの条件をすべて満たす、現時点でおそらく唯一の異常検知手法です。上記に加えて、ANACONDAでは、変数間の隠された依存関係を発見しユーザーに提供することができます。単に結果をブラックボックス的に与えるのではなく、システム内部で起こっている現象に気づきを与えることで、より納得性の高い異常検知システムが構築可能です。

以下、MT法との比較を軸に、ANACONDAの概要を説明します。

4-1 計算手順の概要

MT法の異常検知の思想は、要するに「普通に考えてありえないくらい平均値からずれたら異常と見なす」ということです。図2で説明したとおり、このような考え方は単純な状況なら良いのですが、負荷変動が頻繁に起こるような動的なシステムにおいては適用困難です。平均値に当たる量もまた動的に変わるからです。

ANACONDAの異常検知の思想は、MT法とはまったく異なります。ANACONDAでは、各変数の個々の値の平均値からのずれよりも、変数同士の依存関係の変化に着目します。この発想の転換により、動的に値が変わるような状況でもある程度の状態

監視が可能になります。

ANACONDA のアルゴリズムは大きく分けてモデルの構築と、異常度の計算という 2 ステップに分かれます。図 4 に状態監視シナリオでの動作イメージを示します。

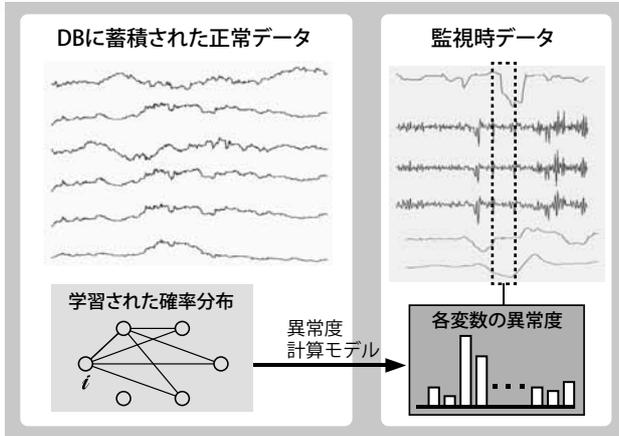


図 4. IBM Anomaly Analyzer for Correlational Data における動作イメージ

モデル構築時においては、正常と信じられる過去のデータから変数の依存関係のモデルを学習します。依存関係のモデリングには、グラフィカル・ガウシアン・モデル (GGM) というモデルを用います [5]。GGM の歴史は実は長く、GGM の学習には共分散構造選択 (covariance selection) という名前の古典的な手法が 1970 年代から知られていました。しかし MT 法と同様の数値計算的な困難を抱えており、GGM がセンサー・データの解析に使われることはほとんどありませんでした。

ところが 2006 年以降、状況が一変します。共分散構造選択を Lasso 回帰という特殊な回帰問題に帰着させる画期的なアルゴリズムが発見されたからです [6]。これにより数十から数百、さらには数万という数の変数を GGM で容易に扱う道が開けました。ANACONDA はまさにこのブレイクスルーの上に、統計的機械学習の観点で首尾一貫した理論を構築した初めての異常検知アルゴリズムです。ANACONDA では、次の最適化問題を解くことで GGM の学習を行います。

$$(4) \max_{\Lambda} (\ln \det \Lambda - \text{Tr}(\Lambda S) - \rho \|\Lambda\|_1)$$

ここで Λ は S の逆行列に対応する $M \times M$ 行列であり、GGM においては、グラフ構造と 1 対 1 の対応がつかます。詳しい説明は省きますが、重要なのがカッコ内第 3 項で、この項により、ノイズを除去しながら **スパースな** (つまり本質的な依存関係だけを与えるような) 依存関係のモデルを見出すことができます。

なお、変数間の類似度指標を適当に作るなどして、アドホックに依存関係のモデルを作ることは可能かもしれませんが、そのようなモデルでは通常、擬似相関の影響が排除できないなど、結果の解釈に深刻な問題が生じます。いわゆるグラフィカル・モデ

ル [7] として依存関係が統計的に正しく定義されている必要があります。

ANACONDA の中核となるアイデアは、上記で学習された GGM を用いて変数ごとに異常度スコアを計算することです。正常時のモデルが GGM における確率分布 $p(\mathbf{x})$ として得られた後、各変数について異常度を時々刻々計算していきます。 \mathbf{x} が現在観測されたデータだとすると、第 i 変数の異常度スコアは次のようになります。

$$(5) a_i(\mathbf{x}) \equiv -\ln p(x_i | x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_M)$$

この式は、第 i 番目のセンサーにとってみれば、いわば「他のセンサー仲間から期待されるふるまいをしているかどうかの度合い」を、条件付き確率の性質を利用して計算していることに対応しています。逆に言えば、ANACONDA は「センサー同士の仲間割れ」を見出す計算手法とも表現できます。

4-2 ANACONDA の特長

実用上の ANACONDA の革新的特長の一つは、システムに内在するモジュール構造を自動的に見出すことができるという点です。これは **スパース構造学習** の手法 (式 (4)) に多くを負っています。ANACONDA においては、センサーの数が非常に多くても、センサー同士の本質的な依存関係のみを抽出することが可能です。これは直感的には、変数同士を密接に絡み合う小グループに分けることと同じです。

数値計算の観点では、式 (4) は大変性質の良い最適化問題であり、非常に効率良く解くことができます [5]。逆行列計算に伴う数値的不安定性が MT 法の課題でしたが、そのような問題はここでは存在しません。

さらに ANACONDA では、個々のセンサーに対する異常度を、統計的に自然なやり方で定義できます。変数間の依存関係についても同様で、グラフィカル・モデルとして妥当なやり方でスパースに依存関係を推定し、MT 法などでは困難であった変数同士の関係についての有益な情報を提供します。論理的な首尾一貫性は、学問的な美しさという点のみならず、将来的な手法の拡張を可能にするという観点でも重要です。実際、IBM 東京基礎研究所のグループでは、ANACONDA にさまざまな拡張を加え、ビジネス上の多くの実問題を解決しています。

なお、ANACONDA は GGM をベースに異常度を計算していますが、これはデータの確率分布として正規分布を仮定していることを意味しません。GGM は単に依存関係を見るためのフィルタとして使われているだけであり、元のデータが正規分布に従わない時でも有効です。

