D_020

マハラノビスタグチ法を用いた監視サーバの異常発見・診断の検討

Fault Detection and Diagnosis with Mahalanobis Taguchi Strategy

酢山 明弘† 森 紘一郎† 折原 良平†
Akihiro Suyama Kouichirou Mori Ryohei Orihara

1．まえがき

社会インフラシステムの監視は一般に異常データの数が非常に少なく，データマイニング的な扱いが難しい．また，監視システム構成の違いによってデータの傾向が全く異なっており、さらに、外乱や経年劣化によって複雑かつ不規則にデータの傾向が変化していくため、予め入手したデータに対して静的に分析することができない．したがって、オンライン上で分析モデルを動的に適応させ，高速に分析を行える枠組みが必要となる。

本研究では，品質管理的なアプローチの１つであるマハラノビスタグチ法へ対象とする問題を適切に応用することで WMI, SNMP データから，異常予測や診断を可能とする分析法を提案する．

2．関連研究

2.1 マハラノビスタグチ法

マハラノビスタグチ法（MT 法）[1]は，単位空間が集団の端にあり，異常になると単位空間からの距離が大きくなって異常に対する検出を早期に行う方法である．MT 法では，式(1)で与えられるマハラノビス平方距離 D^2 を単位空間からの外れ度合（＝異常度合）として与えることが特徴である．

$$D^2 = \frac{1}{k}\mathbf{u}\mathbf{R}^{-1}\mathbf{u}^T \quad \mathbf{u} = \left(\frac{x_1-\mu_1}{\sigma_1},\ldots,\frac{x_k-\mu_k}{\sigma_k}\right) \cdots (1)$$

ここに、R^{-1} は単位空間の相関行列の逆行列，k は変量の数（MT 法では項目数と呼ぶ）であり，異常がない場合のマハラノビス平方距離の平均はほぼ１となる．

MT 法におけるマハラノビス平方距離では，一般的なマハラノビス距離を項目数で割っている．一般的なマハラノビス距離は自由度 k のカイ二乗分布に従うことが知られていることから，MT 法におけるマハラノビス平方距離のカイ二乗分布 生存関数は図１のように示すことができる．

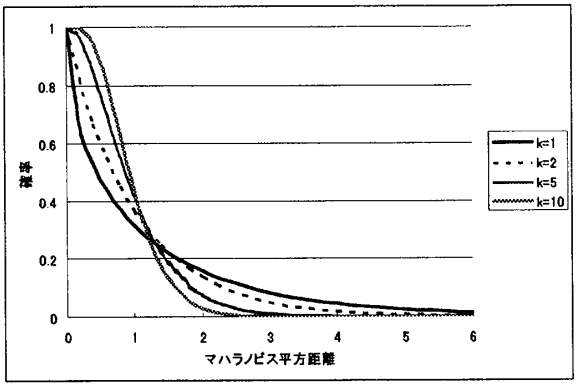

図1 MT 法におけるマハラノビス距離の生存関数

†(株)東芝 研究開発センター

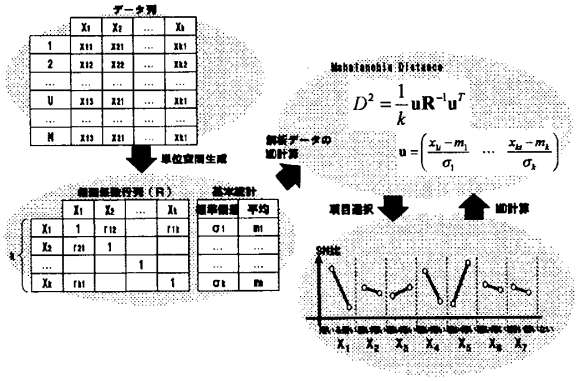

図2 マハラノビスタグチ法

図１は，X 軸がマハラノビス平方距離，Y 軸が確率であり，グラフ曲線はマハラノビス平方距離が x 以上の値となる確率を意味している．すなわち，項目数によらず 4～5 を超える確率がほぼ０であることから，一意に異常判定基準を設定することが可能である．

次に，MT 法の基本処理手順を図２に示し，以下に主要処理概要を記す．

1. 項目を決定する
2. 単位空間を求める（相関行列の逆行列，各項目の平均，標準偏差）
3. 各データのマハラノビス平方距離の計算
4. 要因効果図を用いた項目選択を行い３へ．

MT 法と呼ばれる所以はステップ 4 における直交表と要因効果図を用いた項目選択法にある．項目選択は精度が高い異常検出や計算コストの低減を実現する効果がある．単位空間を正常時の空間とした場合，距離が異常になったとき，その原因を探求する異常診断を次のように実施する．まず，すべての項目について２水準

第１水準：その項目を用いる
第２水準：その項目を用いない

にとり，２水準系の直交表 L^n にわりつけてＳＮ比を求める．続いて，各項目について用いた場合と用いない場合のＳＮ比の平均差を求める．用いた場合の利得の差が大きさで重要度を決定し，重要度が高い項目群を原因項目と推定する．ここでは，X1, X4 の利得差が多く，異常の原因であると推論される．

3．提案手法

3.1 対象問題のモデル化

本研究では，監視サーバの健全性保証を対象問題とし，監視サーバ，ネットワークで取得可能な WMI，SNMP データを入力として扱う．MT 法では対象問題となるデータの単位空間が集団の端になるようにマッピングする必要がある．そこで我々は，監視サーバの異常がハードウェ

アの経年的な劣化，利用者の操作ミス，ＯＳパッチによるアプリケーションの誤動作，アプリケーションの内在的なバグによって引き起こされると考え，このような状況が起こりにくい運用開始直後に獲得できるデータを単位空間とした．このように設定すると，正常状態におけるマハラノビス平方距離の時系列変化は，式(2)の線形モデルに近似することができる．

$$Y=\beta\cdot t+\gamma \quad (\beta\cong 0)\cdots(2)$$

このモデルを利用すると，MT 法によって異常検出と異常検出した結果に対する診断は容易に行うことが可能ある．以下本論文では，MT 法の枠組みでは行うことができない異常予測と予測結果に対する診断に関して説明する．

3.1 異常予測

異常予測は，何らかのトリガによって当初計算されていた余命よりも短い時間で故障となるような場合に，その変化をいち早く捉えて警告することで，未然に故障を防ぐことを目的に行う．

本研究における異常予測は，ウインドウ付きモデル系列候補生成とＭＤＬ原理[2][3]に基づく最適モデル系列選択によって常時モデルを更新し，時刻（ｔ－１）と時刻ｔとの間のモデルに変化が生じた場合に余寿命を再計算して警告を与えるものであり，以下の点において特徴がある．

1. ウインドウ幅を単位空間生成するまでに必要とした時間の定数倍で与え，開始時刻～現時刻を分割する組合せのパターンを候補とする．
2. 区間Ａのすべての部分組合せで作成したモデルが，区間Ａで作成したモデルと一致した場合，以降，区間Ａの部分候補は考慮しない．
3. それぞれの候補の各ウインドウで，最適な線形モデルを生成する．

特徴１と２は，生成する候補を減らすことで計算コストを減少させることを目的とする．最適モデル系列の選択は，モデルの確からしさが等しい２つのモデル系列があった場合，簡単に記述できるモデル系列を優先するので，予めこれらの特徴で切り捨てることが有効である．特徴３に関しては，計算コストを減少させることを目的として，高次のモデルへの近似は行わず線形モデルのみ生成する．ただし，特徴１と３に関しては実データに合わせて設計する必要があるので，現在は推測段階である．

モデル系列候補の中から最適なモデル系列を一つ選択する方法として，過剰なモデル変化を抑制するため，自由パラメータの数が多くなることによるペナルティが大きい MDL 原理を利用している．本研究では最適な線形モデル（モデルと実測値の誤差εが平均０，分散σの正規分布に従う）に近似することを仮定しているため，モデル選択の情報量は次式となる．

$$MDL=\frac{N}{2}\log\left(2\pi\sigma^2\right)+\frac{1}{2\sigma^2}\sum_{i=1}^{N}\varepsilon_i^2+\frac{m}{2}\log N\cdots(3)$$

ここに，σ は分散，ε_i は i 番目のデータとモデルとの誤差，m は自由パラメータの数，N はデータ数である．

3.2 異常予測に対する診断

明らかに異常状態であるデータの診断は，ＭＴ法のＳＮ比の計算により推論することが可能であるが，予測に対する診断では必ずしも正しい結果が得られるとは限らない．何故ならば，正常状態とほとんど変わらない予測点のデータにおいては，その利得差においてもほとんど差が生じないからである．そこで，その項目が最初から利得差が大きいのか，それとも異常になったことによって大きくなったのかを分析するために以下の方法を用いた．

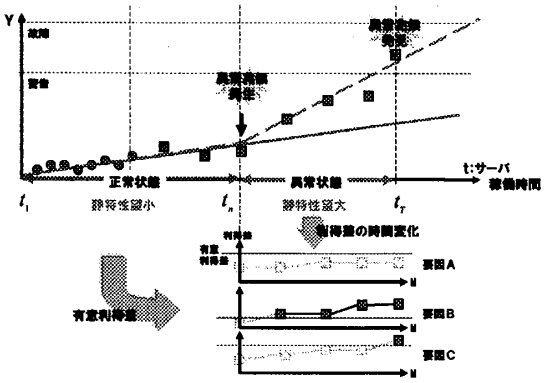

図3 異常予測に対する診断

モデルが変化する前を正常モデル，変化後を異常モデルとする．正常モデルでは各項目における望小特性の利得差平均 $\overline{Gd_i^L}=1/n\sum_{t=t_n}^{t_n} Gd^L(t)\ \ (i=1..k)$，異常モデルではそれぞれのデータにおける各項目の望大特性の利得差 $Gd_i^H(t)(i=1..k,t=t_n..t_T)$ を求め，$t\geq t_n$ において式(4)を計算し，閾値を超えた段階で異常原因項目と判断させる．

$$F_i(t)=\frac{Gd_i^H(t)\times\overline{Gd_i^L}}{Gd_i^H(t)+\overline{Gd_i^L}}\cdots(4)$$

これにより，単に項目 B が異常原因であるというだけでなく，ある時刻において項目 B が原因で異常が発生し始め，さらに項目 C に派生したとの時系列的な原因の出力が可能となる．

4．むすび

本論文では，社会インフラシステムなど異常がほとんど発生しないデータを対象とした異常分析の方法として，マハラノビスタグチ(MT)法を用いた提案手法を紹介した．特に，MT 法の枠組みでは行うことができない予測と予測結果に対する診断方法に関して詳細に述べた．

時系列データを扱う先行研究の成果として日本電気(株)の TrendLiner™ が知られている．これら先行研究に対しての本研究の特徴は，予測結果に対する診断にあるといえる．

今後，本提案手法を実データで検証するとともに，洗練化を行う予定である．

参考文献

- [1] 田口玄一，“研究開発の戦略～華麗なるタグチメソッドの真髄～”，日本規格協会（2005）．
- [2] 下平英寿，伊藤秀一，久保川達也，竹内啓，“モデル選択　予測・検定・推定の交差点”，岩波書店（2004）．
- [3] Jorma Rissanen, “Modeling By Shortest Data Description”, *Automatica*, Vol.14, pp465-471(1978).