4C-02

スパース・モデリングを応用した マハラノビス・タグチ法による異常検知

大久保 豪人[†]　　永田 靖[†]
早稲田大学　創造理工学部　経営システム工学科[†]

1. はじめに

マハラノビス・タグチ（MT）法[1]はタグチメソッドにおける代表的な手法であり，実用的な異常検知手法として，我が国の製造業を中心に広く普及している．特に近年では，センサー等から取得されたデータをもとに設備機器の状態監視へ適用された事例も報告されている．しかし，一般にセンサーから取得されたデータは，ノイズを伴い観測されるため，学習データから本質的な相関構造を推定することが難しい．

そこで本報告では，ガウシアン・グラフィカル・モデリング（GGM）に基づく正則化を応用することで，学習データから本質的な相関構造を精度よく推定することを目指す．具体的には，MT 法の解析目的に合わせて GGM に基づく正則化を定式化し，新たな解析プロセスとして提案する．また，数値実験を通して提案プロセスの有用性を示す．なお，本報告の内容は著者らの研究成果[2]に基づく．

2. MT 法の概要

MT 法では，均質な母集団を形成する群のことを単位空間と呼ぶ．MT 法とは，判定対象となる個体がこの単位空間に属するか否かをマハラノビス距離（MD）を用いて判定する手法である．

まず，単位空間に属する n 個の個体の各々が p 次のベクトル形式 $\boldsymbol{x}_i$ $(i = 1,2, ..., n)$で無作為に抽出されたとする． なお，個体番号を i とする．また，以降の数式中の小文字はベクトル，大文字は単一変量あるいは行列を表す．

次に，各個体に「各変数の単位空間に属する n 個のデータの平均を引き，標準偏差で割る」という標準化を行う．このとき，標準化後の各個体の p 次のベクトルを $\boldsymbol{u}_i$ $(i = 1,2, ..., n)$で表す．

個体 i の MD は，単位空間に属する n 個のデータから求めた相関係数行列を $\mathbf{R}$ として

$$MD_i^{\,2} = \boldsymbol{u}_i^{\,T}\mathbf{R}^{-1}\boldsymbol{u}_i \tag{1}$$

の平方根で定義される．なお，上付きの T は行列やベクトルの転置とする．

ここで，以降の議論では，(1)式を相関係数行列ではなく，共分散行列の逆行列（以降，精度行列と呼ぶ）を使用しているとみなして議論を進める．すなわち，母精度行列を $\mathbf{\Lambda}$，その最尤推定量を $\mathbf{\Lambda_{MLE}}$ として以下のように書き換える．

$$MD_i^{\,2} = \boldsymbol{u}_i^{\,T}\mathbf{\Lambda_{MLE}}\boldsymbol{u}_i \quad (i = 1, 2, \cdots, n) \tag{2}$$

以上の手順を，判定対象となる個体にも繰り返し実施して MD を算出する．ただし，各変数の平均および標準偏差，相関係数行列は，単位空間のデータから既に計算した値を用いる．そして判定対象となる個体の MD が，事前に定めた閾値よりも小さいならば単位空間に属し，大きいならば属さないと判断する．

3. GGM に基づく MT 法

従来の MT 法は GGM の観点から言えば，常に精度行列のフルモデルに基づいてパラメータを推定する方法とみなすことができる（以降，従来の MT 法を MT-FM 法と呼ぶ）．そのため，推定すべきパラメータ数は最大となり，学習データへの過剰適合リスクも最大となってしまう．

そこで，本章では， GGM に基づき精度行列のパラメータ数を削減するプロセスを MT 法に導入した MT-GGM 法を提案する．以下，MT-GGM 法で使用する正則化アルゴリズムとモデル評価規準について説明する．

正則化アルゴリズム．グラフィカル Lasso[3]を用いて精度行列の正則化を行う．なぜなら，変数間に線形関係が成立する場合でも安定的な結果が得られるからである．グラフィカル Lasso では次式に示す最適化問題を解くことで，縮約モデルを得る．すなわち，標本共分散行列を $\mathbf{V}$ とするとき，精度行列 $\mathbf{\Lambda_{glasso}}$ を次のように求める．

$$\mathbf{\Lambda}^{*}_{\mathbf{glasso}} = \underset{\mathbf{\Lambda_{glasso}}}{\arg\max}\, \ell\left(\mathbf{\Lambda_{glasso}};\mathbf{V}\right) - \rho\left\|\mathbf{\Lambda_{glasso}}\right\|_1 \tag{3}$$

$$\ell\left(\mathbf{\Lambda_{glasso}};\mathbf{V}\right) \equiv \log\left|\mathbf{\Lambda_{glasso}}\right| - tr\left[\mathbf{V}\mathbf{\Lambda_{glasso}}\right] \tag{4}$$

ただし，(3)式の正則化項の重み ρ は非負定数であり， |・|は行列式，tr[・]は行列のトレース，

Anomaly Detection with the Mahalanobis-Taguchi Method based on Sparse Modeling
†Masato OHKUBO・WASEDA University
†Yasushi NAGATA・WASEDA University

$\|\cdot\|_1$ は L_1 ノルム（行列の各要素の絶対値の和）を示す．また，(4)式は多変量正規分布の対数尤度関数である．

モデル評価規準．予測だけでなく，原因追究にも適したモデルを選択するため，BIC タイプの情報量規準を用いる．また，BIC タイプの規準の中でもグラフィカル Lasso のような罰則項付き最尤推定に対する規準である EBIC[4]を使用する．

4. 実データ解析

前章で提案した MT-GGM 法は，その正則化の特徴から，精度行列の非対角要素にゼロの値に近い要素が多ければ多いほど有効な過剰適合対策となる．本章では，そのような特徴をもつ実データを取り上げ，実際に解析を行う．

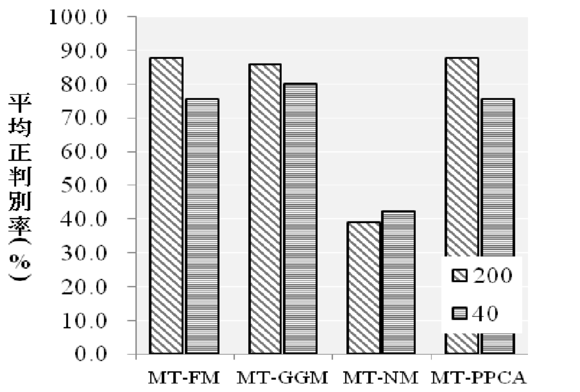
用いるデータ．UCI 機械学習レポジトリの Ionosphere データセットを用いる．変数は 34 次元であり，正常データは 225 個，異常データは 126 個である．また，全 34 変数中の 2 変数は正常データに属する個体から求めた標準偏差がゼロとなるため解析から除外する．

実験設定．用いるデータを次のような設定で 100 組用意する．単位空間は 225 個の正常な個体から，無作為に $N_0(N_0 = 40, 200)$個を抽出して定める．正常データは 225 個の全正常データ，異常データは 126 個の全異常データを用いる．

評価方法．評価方法は異常データの正判別率（以降，単に正判別率と呼ぶ）を用いる．正判別率は，0 から 100 までの値をとり，100 に近いほどよい．判別のための閾値は，正常データの誤判別率（以降，単に誤判別率と呼ぶ）が 1%となるように設定する．

比較手法．MT-FM 法および MT-GGM 法に加え，新たにナルモデルに基づく MT 法（以降，MT-NM 法と呼ぶ）および MT-PPCA 法を比較対象とする．ここで，MT-NM 法とは，MT-GGM 法において最小のパラメータ数となるモデルを常に選択する方法である．すなわち，(1)式の MD の算出の際，相関係数行列 **R** の代わりに単位行列 **I** を用いて計算を行う．また，MT-PPCA 法とは大久保・永田[5]が提案した PPCA モデルに基づく共分散行列の正則化の方法である．

実験結果・考察．図・1 に各手法による解析結果を比較したグラフを示す．グラフの縦軸は正判別率の平均であり，図中の棒は各手法の平均正判別率を，単位空間サンプル数の条件ごとに示したものである．図・1 より，サンプル数が大きい場合と小さい場合において最も安定的に高い性能を示しているのは MT-GGM 法であることがわかる．一方，MT-PPCA 法は MT-FM 法と同じ性能を示しており，正則化の効果は確認できない．実際，MT-PPCA 法ではパラメータの削減は行われていないため，本データは PPCA モデルで仮定するような特徴は有していないといえる．



図・1 各手法の性能比較（Ionosphere）

5. まとめと今後の課題

本報告ではタグチメソッドの手法である MT 法を取り上げ，ノイズを伴って観測されるデータへの適用方法について議論した．具体的には，GGM に基づく精度行列の正則化を MT 法の解析目的に合わせて定式化し，新たな解析プロセスとして提案した．また，数値実験を通して，新たな解析プロセスの有用性を示した．なお，当日の発表ではモンテカルロ・シミュレーションにより本実験の再現性を確認した結果も示す．

今後の課題として，観測変数がより高次元の場合においても，本報告で述べたような議論が成立するか検証することを挙げる．

参考文献

[1] Taguchi, G. and Jugulum, R. (2002): *The Mahalanobis-Taguchi Strategy: A Pattern Technology System*. John Wiley and Sons.

[2] 大久保豪人・永田靖: グラフィカル・モデリングに基づくマハラノビス・タグチ法，応用統計学(掲載予定).

[3] Friedman, J. , Hastie, T. and Tibshirani, R. (2008): Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, **9** (3), 432-441.

[4] Chen, J. and Chen, Z. (2008): Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, **95**(3), 759-771.

[5] 大久保豪人・永田靖(2015)：MT システムにおける小標本データの解析方法. 日本経営工学会論文誌, **66**(1), 30-38.