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

早稲田大学　大久保豪人

1. はじめに

マハラノビス・タグチ（MT）法（Taguchi and Jugulum (2002)）はタグチメソッドの代表的手法の一つであり，異常検知やパターン認識のための基本的な多変量解析法として普及している．特に近年では，Internet of Things（IoT）関連技術の進展に伴って，センサーから取得されたデータをもとに設備機器の状況監視を行う目的で MT 法が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って観測されるため，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，本研究では，ガウシアン・グラフィカル（GG）モデリングの応用により，センサー・データのようなノイズを伴って観測されるデータから本質的な相関構造を推定するプロセスを MT 法に導入することを提案する．なお，本講演の内容は大久保・永田 (2017)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 法では，この単位空間からの離れ具合をマハラノビス距離によって定量化する．そして，マハラノビス距離の大きさが事前に定めた閾値を超えるか否かによって，判定対象となる個体が単位空間に属するか判定する．

いま p 次元変数 $\boldsymbol{x}$ が母集団で観測され，母集団からの大きさ n の無作為標本を $\boldsymbol{x}_i$ $(i = 1,2, ..., n)$とする．また，$\boldsymbol{x}$ の母平均ベクトルおよび母共分散行列を $\boldsymbol{\mu}$ および $\boldsymbol{\Sigma}$ とする（以降，母数 $\boldsymbol{\theta}$ の推定量を $\hat{\boldsymbol{\theta}}$ と記す）．このとき，標本マハラノビス距離 $D(\boldsymbol{x}_i)$ $(i = 1,2, ..., n)$を次のように定義する．

$$D^2\left(\boldsymbol{x}_i\right)=\left(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}\right)\hat{\Sigma}^{-1}\left(\boldsymbol{x}_i-\hat{\boldsymbol{\mu}}\right)^T \tag{1}$$

現行 MT 法では $\boldsymbol{\mu}$ および $\boldsymbol{\Sigma}$ の推定量には単位空間に属する n 個の個体から求めた標本平均ベクトルおよび標本共分散行列を用いる．また，判定対象となる新たな個体に対して標本マハラノビス距離を計算する際も，単位空間から既に計算した母数の推定量を用いる．

3. 提案プロセス

いま単位空間の真の分布が多変量正規分布であることが想定できるものとする．加えて，母共分散行列の逆行列（以降，「母精度行列」と呼ぶ）の非対角要素の大半がゼロの値をとるとの事前情報がある場合，統計モデルとして GG モデルを想定できる．そして，GG モデルの仮定のもと，本質的な相関構造を推定する方法論として GG モデリングがある．一方，MT 法においてもマハラノビス距離の算出時，精度行列を推定する必要がある．このことから，GG モデリングの応用により，本質的な相関構造を捉えた精度行列の推定，延いては異常検出性能の向上が実現できるといえる．

GG モデリングのアルゴリズムには様々な種類が存在するものの，本研究では Friedman et al. (2008)のグラフィカル Lasso を使用する．グラフィカル lasso では多変量正規分布の尤度関数に L_1 正則化項を加えた次の最適化問題を解く．そして，その最適解を GG モデルのもとでの母精度行列 $\boldsymbol{\Omega}_{\mathbf{GG}}$ の推定量とするアルゴリズムである．

$$\max_{\boldsymbol{\Omega}_{\mathrm{GG}}}\left\{\ell\left(\boldsymbol{\Omega}_{\mathbf{GG}};\mathbf{S}\right)-\lambda\left\|\boldsymbol{\Omega}_{\mathbf{GG}}\right\|_1\right\} \tag{2}$$

$$\ell\left(\boldsymbol{\Omega}_{\mathbf{GG}};\mathbf{S}\right)\equiv\ln\left|\boldsymbol{\Omega}_{\mathbf{GG}}\right|-\mathrm{tr}\left(\mathbf{S}\boldsymbol{\Omega}_{\mathbf{GG}}\right) \tag{3}$$

ここで，$|\cdot|$は行列式，tr(・)は行列のトレース，$\|\cdot\|_1$ は行列の L_1 ノルム（行列の全要素の絶対値の和）を示す．また，$\mathbf{S}$ は標本共分散行列，λ は正則化項の重み（非負の定数）である．また，(3)式は多変量正規分布の対数尤度関数に対応している．

参考文献

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

大久保豪人, 永田靖. (2017): 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, **46**(1), 13-26.

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