

博士論文概要

論文題目

統計的モデリングに基づく
マハラノビス・タグチ・システム
The Mahalanobis-Taguchi system
based on statistical modeling

申請者

大久保	豪人
Masato	OHKUBO

経営システム工学専攻 統計科学研究

2017年11月

タグチメソッドの代表的な方法論の一つであるマハラノビス・タグチ (MT) システムは我が国の製造業を中心に広く普及しており、品質管理、医療等の様々な分野において重要な役割を担っている。例えば、品質管理の分野では、設備機器の状態監視等に応用されている。また、医療の分野では、健康診断等に応用された事例がある。しかしながら、MT システムを現行プロセスに則って実問題に適用する場合、実務的な需要に対応した適切な分析ができるとは限らない。そこで本研究では、実問題を適切に分析するため、統計的モデリングの枠組みに則った MT システムの新たな解析プロセスを提案する。

ある個体が正常な個体か異常な個体かを判定する問題を考える。ただし、正常な個体は同一のパターンであることが想定できる一方で、異常な個体には様々なパターンが存在するものとする。言い換えれば、正常な個体は群をなすと仮定できる一方で、異常な個体は群をなすとは仮定できない。このとき、元の問題を「ある個体が正常な群に属するか否かを判定する問題」として再定義する。このような問題を“異常検知問題”と呼ぶ。異常検知問題として定式化することで、未知のパターンが発生した場合でも、異常な個体が検知可能となる。

田口玄一博士は異常検知のための実用的な方法論として、MT システムを提唱した。MT システムはホテリングの多変量管理図に独自のアイデアを加えて発展させた方法論であるといえる。MT システムとホテリングの多変量管理図は、マハラノビス距離とその派生距離に基づく異常検知方式である点で共通している。一方、両者を比較すると、MT システムは母集団の分布に対して全く仮定をおいていない等、一見するとホテリングの多変量管理図よりも汎用性が高い解析プロセスが提案されているといえる。

しかしながら、実問題に対して MT システムを適用する場合、必ずしも適切な分析ができるとは限らない。言い換えれば、MT システムを使用する場合でも、一般的な多変量解析法と同様に、解析者は分析の目的や対象となるデータに合わせた適切な解析プロセスについて慎重に検討する必要がある。さらに、現行プロセスは統計的モデリングの観点から言えば、(1)モデル設定、(2)母数の推定、(3)モデル評価の各ステップに問題があるといえる。ただし、変数選択が統計的モデリングの一例であることに注意する。

第 1 に「(1) モデル設定」では、変数の組合せに対応した統計モデルを設定しているため、未知パターンの異常が検知できなくなる危険性がある。一般に異常検知問題では、未知パターンの異常検知も目的となるため、変数選択は保守的に行った方がよい。すなわち、事前には異常検知に役立つか不明である変数も未知パターンの異常検知には役立つ可能性がある。そのため、技術的に異常検知に役立たないことが明らかな場合を除いて変数選択は実行しない方がよいといえる。

第 2 に「(2) 母数の推定」についてはデータの特徴を考慮せず常に同じ推定法を使用することが問題であるといえる。現行プロセスでは標本マハラノビス距離

の計算に標本平均ベクトルおよび標本共分散行列を用いている．ここで，母集団分布に対して厳密な仮定をおいていないことに注意すると，現行プロセスは常に経験分布に基づくモーメント法によってパラメータを推定しているといえる．

第 3 に「(3) モデル評価」については解析の目的によらず常に予測力を規準としてモデル評価を実行していることが問題であるといえる．現行プロセスでは信号対雑音 (SN) 比に基づく変数選択を実行することで，予測精度が向上するような変数の組合せを選び出している．

そこで本研究では，統計的モデリングの枠組みに基づく MT システムの新たな解析プロセスを提案する．具体的には，統計的モデリングを通して統計モデルを推定後，その統計モデルのパラメータの推定量を用いて標本マハラノビス距離を計算する．このようなプロセスの導入により，解析目的に合致した適切な統計モデルの評価およびデータがもつ特徴を考慮した統計モデルのパラメータ推定が実行できる．そして，様々な実問題に対して解析目的やデータの特徴を考慮した適切な分析が実現できるようになるといえる．

本研究では，MT システムを小標本データ，高次元データ，汚染データに適用する場合を想定した統計的モデリングの枠組みに基づく新たな解析プロセスを提案する．そして，提案プロセスの有用性を実データ解析および数理的な検討結果を踏まえたモンテカルロ・シミュレーションを通して確認する．第 1 に小標本データの場合，提案プロセスにより異常検知性能が向上できるだけでなく，異常原因の特定に有益な知見を獲得できることを示す．第 2 に高次元データの場合，提案プロセスにより高次元データの特徴を活かした異常検知を実現できることを示す．第 3 に汚染データの場合，提案プロセスによりミスラベルのデータが大量に混入する状況でさえも安定的な異常検出を実現できることを示す．

第 1 の小標本データを対象とした場合の提案プロセスの概要は次の通りである．一般に小標本データ解析では，学習データへの過剰適合の問題が発生するため，推定すべきパラメータの削減が必須となる．加えて，異常検知問題では，未知パターンの異常検知性能の維持も重要となる．そこで本研究では，未知パターンの異常検知性能を考慮した小標本データ解析プロセスを提案する．具体的には，確率的主成分分析モデルあるいはガウシアン・グラフィカル・モデルをモデル族に設定したうえで，その母数の推定とモデル評価を実行する．このとき，異常検知問題では観測変数間に多重共線性が頻発するため，母数の推定法にはその問題を回避できる方法を導入する．またモデル評価には，異常原因の追究のための知見獲得の観点重視したベイズ情報量規準 (BIC) タイプの評価規準を使用する．

第 2 の高次元データを対象とした場合の提案プロセスの概要は次の通りである．高次元データ解析では，サンプル数 n よりも変数の次元 p が遥に多いデータ ($p \gg n$) を取り扱う必要がある．しかしながら， $p \gg n$ の状況では，そもそも現行プロセスのままでは解析不能となるため，新しい解析プロセスや異常検知方式自体の

改良が必要となる．そこで本研究では， $p \gg n$ の状況で計算可能な 2 種類の距離を取り上げたうえで，その距離を用いた新たな解析プロセスを提案する．具体的には，高次元データにおいて慣例的に想定される母共分散行列に対する固有値モデルをモデル族に設定したうえで，その母数の推定とモデル評価を実行する．このとき， $p \gg n$ の状況でも精度よく推定するため，スパース主成分分析に基づく推定法を使用する．またモデル評価には BIC タイプの評価規準を使用する．

第 3 の汚染データを対象とした場合の提案プロセスの概要は次の通りである．汚染データ解析では，十分なサンプル数が確保されている場合でさえ，現行プロセスは良好な異常検知性能を発揮できない可能性がある．単位空間に混入したミスラベルのデータの影響を受け，母数の推定が困難となるためである．そこで本研究では，統計的モデリングのステップの中でも母数の推定に焦点をあて，ミスラベル・データの混入に対してロバストな推定法の導入効果について検証する．そして実務的な動向を考慮して，ミスラベルのデータが大量に混入する場合でも安定的な異常検出を実現するための解析プロセスを提案する．具体的には，多変量正規分布をモデル族に設定したうえで，ガンマ・ダイバージェンスに基づくロバスト推定法を母数の推定に使用する．

近年，情報通信技術の進展に伴って，様々な産業分野においてセンサーやスマートデバイスから膨大なデータが取得・蓄積されている．特に製造業では，このように蓄積された履歴データを活用し，設備機器の異常検出を実現しようという動きがある．しかしながら，このような実問題に対して現行の MT システムを適用する場合，適切な分析が実行できる対象が非常に限定されてしまう．一方，提案プロセスでは様々な実問題に対して適切な分析ができるため，適用対象は格段に広がり，実用的な異常検知システムの実現につながるといえる．

今後の課題として，リアルタイム異常検知に向けた解析プロセスの改良を挙げる．MT システムをリアルタイム異常検知に使用する場合，単位空間データの更新等に対する適切な対応が必要となる．その際，ベイズ更新等が活用できるため，ベイズ的なアプローチが有用であると予想される．しかしながら，ベイズ的なアプローチを前提とする場合，本研究で用いたモデル選択よりも階層ベイズモデリングの実行を考えた方がよいといえる．

本論文は 7 つの章から構成される．第 1 章では，本研究の背景と目的を述べる．第 2 章では，MT システムについて概説する．また，ホテリングの多変量管理図との比較を通して MT システムの概念について説明する．第 3 章では，統計的モデリングの枠組みに基づく MT システムの新たな解析プロセスの概要を示す．続く第 4 章，第 5 章，第 6 章では，小標本データ，高次元データ，汚染データの各々に対して MT システムを適用する場合を想定した新たな解析プロセスを提案する．そして，提案プロセスの有用性を実データ解析およびモンテカルロ・シミュレーションを通して評価する．最後に第 7 章では，結論を述べる．

早稲田大学 博士（工学） 学位申請 研究業績書

氏名 大久保 豪人 印

(2017年 11月 現在)

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
論文	(学術誌原著論文)
○	[1] 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, Vol.46, No.1, pp.13-26, 2017.7. <u>大久保豪人</u> , 永田靖.
○	[2] 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, Vol.61, No.1, pp.30-38, 2015.4. <u>大久保豪人</u> , 永田靖.
○	[3] タグチの RT 法における同一次元でない連続量データへの適用方法. 品質, Vol.42, No.2, pp.86-102, 2012.4. <u>大久保豪人</u> , 永田靖.
講演	(国際会議)
○	[1] Anomaly detection in contaminated unit space using the Mahalanobis-Taguchi system. ANQ Congress 2017, Kathmandu, 2017.9. <u>Masato OHKUBO</u> and Yasushi NAGATA.
○	[2] Anomaly detection in high-dimensional data with the Mahalanobis-Taguchi system. 20th QMOD conference on Quality and Service Sciences ICQSS, Copenhagen / Elsinore, Denmark and Helsingborg, Sweden, 2017.8. <u>Masato OHKUBO</u> and Yasushi NAGATA.
	[3] Applying Graphical Modeling to the Mahalanobis-Taguchi Method. ANQ Congress 2016, Vladivostok, 2016.9. <u>Masato OHKUBO</u> and Yasushi NAGATA.
	(国内会議)
	[1] MT システムによる高次元データ解析. 日本品質管理学会研究発表会（関西支部）発表要旨集 Vol.115, pp.13-16, 大阪, 2017.9. <u>大久保豪人</u> , 永田靖.
	[2] 単位空間の汚染にロバストなマハラノビス・タグチ法. 統計関連学会連合大会講演報告集, pp.355, 愛知, 2017.9. <u>大久保豪人</u> , 永田靖.
	[3] γ ダイバージェンスに基づく MT 法. 日本品質管理学会研究発表会研究発表要旨集 Vol.113, pp.69-72, 東京, 2017.5. <u>大久保豪人</u> , 永田靖.

早稲田大学 博士（工学） 学位申請 研究業績書

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
	<p>[4] 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. 情報処理学会第 79 回全国大会講演論文集(2), pp.51-52, 愛知, 2017.3. <u>大久保豪人</u>, 永田靖.</p> <p>[5] グラフィカル・モデリングに基づく MT 法. 日本品質管理学会年次大会講演・研究発表要旨集 Vol.46, pp.165-168, 東京, 2016.11. <u>大久保豪人</u>, 永田靖.</p> <p>[6] MT システムにおける小標本データの解析方法. 日本品質管理学会研究発表会研究発表要旨集 Vol.101, pp.43-46, 東京, 2013.5. <u>大久保豪人</u>, 永田靖.</p> <p>[7] 次元圧縮を用いた MT システムにおける判定方法. 日本品質管理学会年次大会講演・研究発表要旨集 Vol.42, pp.53-56, 石川, 2012.10. <u>大久保豪人</u>, 永田靖.</p> <p>[8] タグチの RT 法におけるアンサンブル学習の導入. 日本品質管理学会研究発表会研究発表要旨集 Vol.98, pp.217-220, 東京, 2012.5. <u>大久保豪人</u>, 永田靖.</p> <p>[9] 同一次元でない連続量データのためのタグチの RT 法の改良手法. 日本品質管理学会研究発表会研究発表要旨集 Vol.95, pp.153-156, 東京, 2011.5. <u>大久保豪人</u>, 永田靖.</p>
その他	<p>(講演)</p> <p>[1] スパース・モデリングに基づくマハラノビス・タグチ・システム. 第 11 回日本統計学会春季集会, 企画セッション 4: テクノメトリックスー品質改善 を实践するための数理統計ー, 講演 3, 東京, 2017.3. <u>大久保豪人</u>.</p>