

令和 4 年 6 月 9 日現在

機関番号：32665

研究種目：基盤研究(C)（一般）

研究期間：2019～2021

課題番号：19K04891

研究課題名（和文）品質工学の判別尺度による多変量データの識別性の検討と品質管理への応用

研究課題名（英文）Examination of discrimination of multivariate data by quality engineering discrimination scale and application to quality control

研究代表者

矢野 耕也（YANO, Koya）

日本大学・生産工学部・教授

研究者番号：30514153

交付決定額（研究期間全体）：（直接経費） 600,000円

研究成果の概要（和文）：近年のビッグデータの利用や、画像とか音声情報といった大量のデータを用いた識別、判別、分類、予測が行われているが、従来の統計的な解析方法では、データの中の項目情報より多くサンプルデータが必要となる。そのことから、解析に必要なデータ数が揃わないとか、項目情報が多すぎて解析が対応できないという問題が生じることが多く、さらにパターン認識では効率性や正確性が求められるにも関わらず、項目情報の多さとサンプルデータ数のアンバランスにより適切な解析が出来ないことが多い。そこで品質工学の転写機能を利用し、多数の項目情報を2項目に縮約し、2種類の統計量で解析をすることで、的確な識別や判別、分類を行うことができた。

研究成果の学術的意義や社会的意義

大量のデータの解析には従来は多変量解析という手法が適用されてきたが、データ数や情報量に制約がつくケースが多かった。ここでは品質工学で用いる転写という概念を適用し、多数の情報を全て2に圧縮し、精度を落とさずに2つの統計量で1つの尺度にまとめることを可能とし、識別、判別、分類等を可能とした。対象分野は無数にあるが、一例として医学データや品質管理上の成分分析で実施を行い、効率の良い的確な類型分類を可能とした。数万以上のデータも2変量に圧縮し、最終的に1つの尺度で議論が可能のために、数値管理も容易になることが期待できる。

研究成果の概要（英文）：In recent years, big data has been used, and identification, discrimination, classification, and prediction have been performed using a large amount of data such as images and audio information. However, the conventional statistical analysis method requires more sample data than the item information in the data. Therefore, there are often problems that the number of data required for analysis is not available or that there is too much item information to handle the analysis. In addition, although efficiency and accuracy are required in pattern recognition, appropriate analysis is often not possible due to the large amount of item information and the imbalance in the number of sample data. Therefore, by using the transcription function of quality engineering, reducing a large number of item information to two items and analyzing with two types of statistics, we were able to perform accurate identification, discrimination, and classification.

研究分野：品質工学

キーワード：品質工学 多変量解析 パターン認識 品質管理

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

(1) データサイエンスの進化に伴い、大量のサンプリングデータ数や高密度の多次元データが用いられるケースが増加しつつあり、大量のデータ(多変量データ)を効率的に処理し、一元化を可能とするような判別尺度が重要になってきていた。

(2) 上記(1)に係るデータは情報間の相互作用も強く、統計科学的な処理が容易でない分野であり、多数の情報から特徴を的確に分類し、明確に識別可能とする尺度が要求されていた。またデータ量すなわち情報量が増えるに従い、加速度的に処理が困難になるのも克服すべき課題であった。

(3) そこで(1)、(2)を踏まえ、品質工学で用いる方法を多変量データの分析に適用し、特に判別、分類、異質性、同等性といったパターン認識や識別に目的を特化し、実データを用いて検証を行った。

### 2. 研究の目的

(1) データサイエンスの進化に伴い、大量のサンプリングデータ数や高密度の多次元データが用いられることが多くなり、それらの複雑さは増すばかりであるが、極力単純化してシンプルに識別することも重要であり、そのための効率的な判別尺度が重要になっている。

(2) 品質工学では、通信工学で用いられる識別精度の尺度である SN 比や、多次元データの総合的尺度である Mahalanobis の距離という古典的な判別分析を中心とした多次元・多変量データに対する尺度があるが、いずれも通常用いられる評価の方法ではなく、予測、判別、分類、異質性、同等性といった、パターン認識や識別を強化するような尺度、解析方法を検討することを目的とし、現実のデータに適用をすることで適用可能性の現実性の検討を行った。

(3) 第 1 の目的として、解析手法そのものの原理を検討し、背景にある転写(コピー)技術を利用し、ある種の統計量に縮約することの妥当性を検討する。

(4) 第 2 の目的として、本研究課題では実際の医学データ、および医薬品成分の品質管理技術への適用を行い、解析方法の現実的な適用可能性の検討を行う。

### 3. 研究の方法

(1) 全般に医療データを中心に利用をすることを念頭としたため、2019 年から 2020 年にかけて、東京慈恵会医科大学狛江第 3 付属病院消化器・肝臓内科(以下慈恵医大第 3 病院とする)において、生化学的データを得るものとし、倫理審査を通過するための手続きを行い、また日本大学生産工学部での倫理審査も 2020 年 1 月に終了し、2020 年 4 月以降は実データの取得に進めることが出来た。ただし取得制約が多いため、生化学的データについては上記慈恵医大第 3 病院から得たものを使用した。

(2) 次に技術的・数理工学的な方法として、品質工学で用いられている手法であるパターン認識や、識別のための SN 比として、(1)パターン認識の評価に適した手法といわれる RT 法について、転写性技術との対応関係、(2)基準データが  $n=1$  の場合の転写性の SN 比、を検討した。

(3) 3 章(1)の転写性とはパターンの写し取り(転写技術)に由来するもので、品質工学では RT 法という名称で使われているが、原理的な面での解明が不十分であった。今回の検討で、情報数(項目数) $k$  がいくつあっても制限なく取り扱うことを可能とし、また  $k$  を 2 項目に縮約し、余因子行列により 1 次元の尺度に圧縮することができ、転写性という概念と情報圧縮の関連性が明らかになった。転写とは 1 つの基準データに対し、もう 1 つの対象データの対応性が等倍であれば 1:1 になることを理想としており、倍率が 2 倍であれば 1:2 であり、これを感度 という尺度で表す。等倍の場合は  $=1.00$  で、2 倍の場合は感度  $=2.00$  となる。さらに対応性感度の他に、 $k$  個のデータの打点のばらつきの尺度でとらえ、これに SN 比や標準偏差を用いる。このようにして情報量(項目数) $k$  がいくつあっても、感度 と、ばらつきの尺度である SN 比や標準偏差の 2 種類にまとめることで、最小量の 2 種類の統計量(特徴データ)に縮約をするものである。

(4) また 3 章(2)の  $n=1$  の転写性の SN 比は、3 章(1)の転写技術の利用になるが、教師学習信号である基準データ数が 1 個で済むという簡便さがあり、技術的には 3 章(1)で触れた RT 法と同様の転写技術に基づいているが、それほど検討例が多くないため、3 章(1)と同様に適用可能性を実証した。

### 4. 研究成果

(1) 当初は医学データ(生化学的データ)を中心に計画を据えてきたが、医学データに限らず、医薬品成分の均一性管理という品質管理上のデータにも適用して、当初の目的を達成することができた。

(2) 近年は医学データや品質管理データに限らず、サンプリング機能の向上と相俟って、いわゆるビッグデータという大量データの処理の必要性に迫られていたが、これらは通常の変数解析の原則である項目数  $k$  個 < データ数  $n$  個の原則を圧迫する要因であった。仮に 1 万項目あった場合、サンプルデータを 1 万データ以上集める必要があるとしても現実的でない場合がある。そこで、品質工学で用いられている転写性という技術を適用し、項目数がいくつあっても 2 項目に縮約し、データ数  $n$  を極力少なくしながらも、識別性が下がらない方法を検討した。

(3) 3 章(1)についてであるが、複数の成分が混合された医薬品成分を高速液体クロマトグラフィーという装置で成分含量分析を行い、複数の成分として 7 指標を得て、 $k=7$  とした。またサンプルデータとして  $n=82$  を得た。このデータではそのまま通常分析も可能であるが、 $k=7$  を転写性の原理を利用し、 $k=2$  まで縮約することで、識別性にどの程度影響が及ぶかを検討した(図 1~図 4 では明示のために  $k=4$  で説明)。

(4) 図 1 は、模式的に  $k=4$  の場合の医薬品成分の含有波形を示し、図 2 は含有成分にばらつきがある場合の含有波形を示している。これを波形のピーク高さの対応性すなわち転写性で表したものが図 3 と図 4 である。図 3 は、図 1 のようなばらつきがほとんどない場合の波形の転写性を示しており、図 4 は図 2 のように、含有成分にばらつきがある場合の転写性を示している。図 3 のばらつきがない場合の波形の転写性は、ほぼ  $45^\circ$  の角度で直線性をもつ線形式が得られるが、図 4 のように成分含有量にばらつきがある場合は打点位置がずれてしまい、 $45^\circ$  の傾きとはならないばらつきのある状態となり、転写の状態が良好でないことが明白となる。

(5) と SN 比ないし標準偏差の 2 種類を用いて 2 行 2 列の行列式を用いるため、逆行列の解析や余因子行列の解析が必要となることから、最小データ数  $n$  は 4 以上になるが、非常に少ないデータ数でも解析が可能となっている。

(6) 以上のように、図 1 や図 2 のような視覚的な判断に依存するもの、また情報量(項目数)が複数に渡るものが、図 3 や図 4 のような  $x$  軸と  $y$  軸の 2 軸表現が可能となるのが転写性の原理であり、その打点の線形性でシンプルな評価を可能とするものである。

(7) その結果、転写技術を応用した 3 章の(1)RT 法や 3 章の(2) $n=1$  の転写性の SN 比を用いることで、項目数  $k$  の制限を受けることなく、分析・適用が可能であることを示すことができた。

(8) 結果として、大量の情報の処理をシンプルに処理する道筋を与えることができ、効率化、簡略化の方法として汎用性があることを示すことができた。

(9) 3 章(2)で示した  $n=1$  の転写性の SN 比を用いた場合も同様で、この場合は行列式を用いないために、基準データ 1 個に対して対象データが 1 個でも比較ができ、よりシンプルなパターン認識やパターンマッチングを可能としている。

#### < 引用文献 >

矢野耕也、高橋和仁、MT システムのパターン認識への適用、品質工学、第 30 巻、第 2 号、2022、24-30

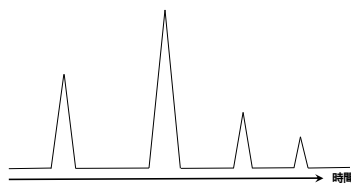


図 1 基準波形模式図(4 ピークの例)

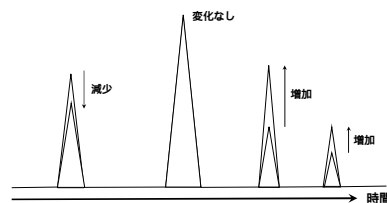


図 2 ピーク高さにばらつきがある場合の模式的波形の重複例

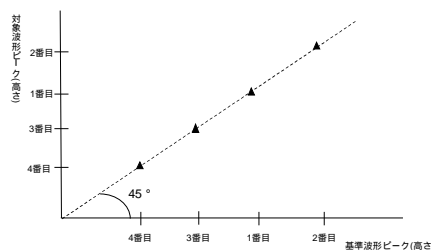


図 3 基準と対象が一致している場合の波形情報のプロット

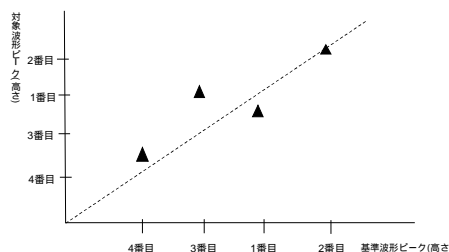


図 4 基準と対象の間に差(ばらつき)がある場合の波形情報のプロット

## 5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 7件/うち国際共著 0件/うちオープンアクセス 7件）

1. 著者名 中島尚登、矢野耕也	4. 巻 68
2. 論文標題 大学病院本院群の類型分類と機能性評価係数 に影響を及ぼす要因について	5. 発行年 2021年
3. 雑誌名 厚生指標	6. 最初と最後の頁 36-42
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 中島尚登、矢野耕也	4. 巻 75
2. 論文標題 経済協力開発機構方式を用いた都道府県別平均寿命・余命に影響を及ぼす要因の検討	5. 発行年 2020年
3. 雑誌名 日本衛生学雑誌	6. 最初と最後の頁 1-18
掲載論文のDOI（デジタルオブジェクト識別子） 10.1265/jjh.20001	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 中島尚登、矢野耕也	4. 巻 67
2. 論文標題 在院日数の短縮に影響を及ぼす主要診断軍分類と診療行為について	5. 発行年 2020年
3. 雑誌名 厚生指標	6. 最初と最後の頁 15-21
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 矢野耕也、高橋和仁	4. 巻 30
2. 論文標題 MTシステムのパターン認識への適用	5. 発行年 2022年
3. 雑誌名 品質工学	6. 最初と最後の頁 24-30
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 中島尚登、矢野耕也	4. 巻 74
2. 論文標題 クラスター分析を用いた都道府県別医療制度パフォーマンスの地域差の検討	5. 発行年 2019年
3. 雑誌名 日本衛生学雑誌	6. 最初と最後の頁 1-21
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 中島尚登、矢野耕也	4. 巻 66
2. 論文標題 経済協力開発機構方式を用いた都道府県別医療制度パフォーマンスの地域差の検討	5. 発行年 2019年
3. 雑誌名 厚生学	6. 最初と最後の頁 8-20
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 中島尚登、矢野耕也	4. 巻 28
2. 論文標題 MTシステムによるOECD35か国の医療制度のパフォーマンスの評価 - 単位空間メンバーが1つの場合の評価法 -	5. 発行年 2020年
3. 雑誌名 品質工学	6. 最初と最後の頁 14-25
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 矢野耕也	4. 巻 36
2. 論文標題 多成分情報の比較による品質評価法 ~ 品質工学の方法を用いた波形パターンの簡易的解析方法 ~	5. 発行年 2020年
3. 雑誌名 PHARM TECH JAPAN	6. 最初と最後の頁 123-134
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 中島尚登、矢野耕也、湯川豊一、松平浩、伊藤周二
2. 発表標題 B型肝炎ウイルスによる消化器がん死亡率の憎悪について
3. 学会等名 第117回日本内科学会講演会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	中島 尚登  (NAKAJIMA HISATO)  (90207788)	東京慈恵会医科大学・医学部・准教授   (32651)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------