ＭＴ（マハラノビスタグチ）システム

１．　ＭＴシステムとは

ＭＴシステムとは、多次元の情報で構成されたデータベースに対して、一つの尺度を導入し、パターン認識をする情報処理の技術である。

この多次元空間の中に、普通・一様・中間の集団で単位空間を定義し、尺度を作り、単位空間に属さない対象について、単位空間と異なる集団かどうかとか、この先どうなるかを診断・予測する。

近年注目され、事例も急増しているが、初期の方法であるＭＴ法では多重共線性や分散ゼロについての問題があり、その対応によりＭＴＡ法・ＴＳ法・Ｔ（１）（２）（３）法と多くの対応策が提案されてきた。

今回は最初に提案されたＭＴ法と最新のＴ（１）法を中心に解説する。

２．　ＭＴ法

1970 年代に田口先生から提案があり、逓信病院の兼高先生との協同研究がスタートしたが、80 年代になって、ようやく成果がでるようになった。

インドの統計学者Ｐ．Ｃ．マハラノビスが提唱したマハラノビスの汎距離を使用する方法で、最初にＭＴシステムとして提案されたものである。

表．１のような単位空間データをｋ個集める。単位空間データは普通・一様・中間なもので十分な数があることが重要である。単位空間データの数ｋは項目の数ｎより大きい必要がある（ｋ＞ｎ）。

表．１の各項目の平均値をm、標準偏差をσとする。

表．２のように、表．１と同じ項目で、単位空間に属さない信号データを１個集める。信号データの数１は１以上であればよい。

信号データは一つひとつパターンが異なるが、単位空間内では同一パターンとし、その中に、パターン内距離としてマハラノビスによる複合距離（マハラノビスの汎距離）を求める。

表．１　単位空間データ

	項目１	項目２	・・・	項目ｎ
1	X_{11}	X_{21}	・・・	X_{n1}
2	X_{12}	X_{22}	・・・	X_{n2}
・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・
k	X_{1k}	X_{2k}	・・・	X_{nk}
平均値	m_1	m_2	・・・	m_n

表．２　信号空間データ

	項目１	項目２	・・・	項目ｎ
1	Y_{11}	Y_{21}	・・・	Y_{n1}
2	Y_{12}	Y_{22}	・・・	Y_{n2}
・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・
l	Y_{1l}	Y_{2l}	・・・	Y_{nl}

標準偏差　　σ_1　　σ_2　・・・　　σ_n

ＭＴ法ではデータの規準化を行う。単位空間データでの各項目の平均値mと標準偏差σにより

$$x_{ij} = \frac{X_{ij} - m_i}{\sigma_i}$$

$$y_{ij} = \frac{Y_{ij} - m_i}{\sigma_i}$$

とする。

単位空間データを用い、各項目間の相関係数ｒを求め、相関行列Ｒを作成する。

$$r_{ij} = \frac{S(x_i, x_j)}{\sqrt{S(x_i, x_i)\,S(x_j, x_j)}}$$

ここで

$$S(x_i, x_i) = \Sigma x_{ik}^2 / k$$

$$S(x_i, x_j) = \Sigma x_{ik} x_{jk} / k$$

とする。

相関行列Ｒは

$$R = \begin{pmatrix} r_{11} & r_{21} & \cdots & r_{n1} \\ r_{12} & r_{22} & \cdots & r_{n2} \\ r_{13} & r_{23} & \cdots & r_{n3} \\ \cdot & \cdot & \cdots & \\ \cdot & \cdot & \cdots & \\ \cdot & \cdot & \cdots & \\ r_{1n} & r_{2n} & \cdots & r_{nn} \end{pmatrix}$$

である。

相関行列の逆行列Ａを求める。

$$A = R^{-1} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ a_{13} & a_{23} & \cdots & a_{n3} \\ \cdot & \cdot & \cdots & \\ \cdot & \cdot & \cdots & \\ \cdot & \cdot & \cdots & \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}$$

逆行列Ａが求まらない場合があり、これを多重共線性という。

多重共線性については後で詳細を述べる。

逆行列Ａと規準化した単位空間データよりマハラノビスの距離を次式で計算する。

$$D^2 = (x_{1i} \cdots x_{ni}) \begin{pmatrix} a_{1i} \cdots a_{ni} \\ \cdots \\ a_{1n} \cdots a_{nn} \end{pmatrix} \begin{pmatrix} x_{1i} \\ \cdot \\ x_{ni} \end{pmatrix}$$

単位空間でのサンプルに対して２乗距離を求めたことになる。
単位空間のマハラノビス距離の平均値は１となり、１とならない場合は逆行列が正常に求まっていないので結果が信用できない。
信号データについても単位空間で求めた逆行列Ａと規準化したデータを用い、信号データのサンプルに対するマハラノビス距離を計算する。

$$D^2 = (y_{1i} \cdots y_{ni}) \begin{pmatrix} a_{1i} \cdots a_{ni} \\ \cdots \\ a_{1n} \cdots a_{nn} \end{pmatrix} \begin{pmatrix} y_{1i} \\ \cdot \\ y_{ni} \end{pmatrix}$$

信号データで求めたマハラノビス距離により、信号データのサンプルーつひとつが単位空間からどれくらい離れているかをみて、単位空間に属するものか否かを判別する。
信号データで真値がわかる場合は真値を信号Ｍにし、マハラノビス距離の平方根Ｄを特性値としてＳＮ比を計算する。
計測項目の中に無駄なものがあると、予測や判別の精度が落ちるので、計測項目を２水準系の直交表にわりつけ、水準１をその計測項目を用いる、水準２をその計測項目を用いないとして、マハラノビス距離を計算する。
このとき、直交表はＬ１２、Ｌ２０等の４×素数のものが望ましい。
直交表の組み合わせ数だけマハラノビス空間をつくり、ＳＮ比を計算して、利得を求めて、計測項目の取捨選択を行う。
信号データで真値がわからない場合は望大特性のＳＮ比を計算し、直交表を利用して、計測項目の取捨選択を行う。

適用事例に健康診断や火災報知機のもの、外観検査技術の研究等に成果が報告されている。
何をしたいのか目的を明確にし、目的に合った計測項目を選び、普通・均質・正常・一様な計測項目で単位空間とするデータベースを定義することからＭＴ法をスタートさせる。
例えば、偽札の判別をする場合、単位空間は本物の札のデータとし、いろいろな状態の本物の札（新品、使い古し、汚れたもの、折り目の付いたもの等）を用意する。
偽札には多くの種類・方式があり、いくら偽札の研究をしても意味がなく、ひとつの偽札に対応できても、すぐに新しい種類・方式の偽札が作られ、対応にはきりがない。
本物の札を研究し、単位空間以外のサンプル（信号データ）に対するマハラノビス距離を計算し、その距離で本物か偽札かを判別するのがＭＴ法である。

ＭＴ法には多くの問題点があり、取り扱うデータの吟味が必要である。

①単位空間の項目で標準偏差がゼロ（その項目のデータが全て同一）の場合、規準化をしようとするとゼロ除算となり、計算できない。
②多重共線性により計算が進まない、計算結果が信用できない。
多重共線性とは逆行列が求まらない状態をいい、下記が原因である。
１）相関係数の絶対値が１または１に近い
２）自由度以上の項目がある
３）サンプル数が項目数より少ない

これらの問題点をクリアするため、色々な対応策が提案され、現在では、ＭＴ法以外に、ＭＴＡ法、ＴＳ法、Ｔ（１）法（両側Ｔ法)、Ｔ（２）法（片側Ｔ法)、Ｔ（３）法（ＲＴ法）が存在する。

３．　Ｔ（１）法（両側Ｔ法）

ＭＴ法では普通・正常等の単位空間はゼロ点近傍で、単位空間に属さない信号データでは正の距離となる。
ＭＴ法の単位空間は端にあったが、Ｔ（１）法は単位空間が真ん中にあり、正負の符号が付いた距離を計算する予測を主体とする方法である。
Ｔ（１）法では計測項目以外に必ず真値が必要で、真値とは予測したい項目をいう。
企業の売上予測、販売数の予測、不動産価格の予測などでは、真中付近に単位空間をとるのがよく、売上高・販売数・価格等が真値で、Ｔ（１）法での予測が使える。
昨年に比べて今年の売上の伸びが普通（例えば±１％）の企業を集めて単位空間として、売上が５％増加した企業、３％減少した企業などは信号とする。

表．３のように単位空間データを収集する。
この時、真値については真ん中で均質なものとする必要がある。

表．３　単位空間データ

	項目１	項目２	・・・	項目ｎ	真値
1	X_{11}	X_{21}	・・・	X_{n1}	XM_1
2	X_{12}	X_{22}	・・・	X_{n2}	XM_2
・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・
k	X_{1k}	X_{2k}	・・・	X_{nk}	XM_k
平均値	m_1	m_2	・・・	m_n	ｘｍ

単位空間データの規準化を項目毎の平均値、真値の平均値を引いて行う。

$x_{ij} = X_{ij} - m_i$

$xm_i = XM_i - xm$

単位空間にとは異なるデータ群で信号データを収集する。

信号データにも真値（予測したい項目）が必ず必要である。

表．４　信号データ

	項目１	項目２	・・・	項目ｎ	真値
1	Y_{11}	Y_{21}	・・・	Y_{n1}	YM_1
2	Y_{12}	Y_{22}	・・・	Y_{n2}	YM_2
・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・	・ ・ ・
l	Y_{1l}	Y_{2l}	・・・	Y_{nl}	YM_l

信号データを単位空間の項目毎平均値、真値の平均値を引き規準化する。

$y_{ij} = Y_{ij} - m_i$

$ym_i = YM_i - xm$

規準化された信号空間の真値と項目１とでＳＮ比η（真数）と比例定数βを求める。

$\eta = (S_\beta - Ve) / r / Ve$　　$S\beta > Ve$の時

$\eta = 0$　　$S\beta \leqq Ve$の時

$\beta = L / r$

ここで

$S_T = \Sigma y_{1j}^2$

$r = \Sigma y_{mj}^2$

$L = \Sigma y_{1j} ym_j$

$S_\beta = L^2 / r$

$Se = S_T - S_\beta$

$Ve = Se / (l - 1)$

以下、項目２～ｎについても同様にＳＮ比と比例定数を求める。

この時、ｉ番目の項目のみで真値を予測すると、$y = \beta M$より

$ym_{ij} = y_{ij} / \beta_i$

となる。

ｊ番目のサンプルについて各項目のみの予測値を用い、項目毎ηの加重平均を総合予

測値em_jとする。

$$em_j = \frac{\eta_1 y_{1j}/\beta_1 + \eta_2 y_{2j}/\beta_2 + \cdots + \eta_n y_{nj}/\beta_n}{\eta_1 + \eta_2 + \cdots + \eta_n}$$

総合予測値は規準化されているので、真値の予測値は

$$EM_j = em_j + xm$$

となる。

ここで予測の精度を表す総合ＳＮ比を求める。

信号データの真値を信号に、予測値を特性値としてゼロ点比例式のＳＮ比を

$$\eta = 10\log\frac{(S_\beta - Ve)/r}{Ve}$$

とする。

ここで

$$S_T = \Sigma EM_j^2$$

$$r = \Sigma YM_j^2$$

$$L = YM_1 \cdot EM_1 + YM_2 \cdot EM_2 + \cdots + YM_l \cdot EM_l$$

$$S_\beta = L^2/r$$

$$Se = S_T - S_\beta$$

$$Ve = Se/(l-1)$$

である。

総合ＳＮ比は大きければ大きい程よい。

計測項目の中に無駄なものがあると、予測の精度が落ちるので、計測項目を２水準系の直交表にわりつけ、水準１をその計測項目を用いる、水準２をその計測項目を用いないとして、総合ＳＮ比を計算する。

このとき、直交表はＬ１２、Ｌ２０等の４×素数のものが望ましい。

直交表の組み合わせ数だけ総合ＳＮ比を計算して、利得を求めて、計測項目の取捨選択を行う。

４．その他のＭＴシステム

ＭＴシステムでは分散ゼロ対策、多重共線性対策のため、色々な提案がされ、現在に至っている。

以下、簡単に解説する。

（１）ＭＴＡ法

ＭＴ法からの変更点のみを述べる。

①基準化を単位空間の各項目の平均値を引くことで行う。

②相関行列でなく分散・共分散行列から余因子行列を用い、直接、逆行列を求める。

以下、ＭＴ法と同様で、単位空間は端にある。

行列演算が繰り返し行われるので、計算に時間が掛かる。

（２）ＴＳ法

予測を主体としたもので、正負の符号付き予測ができる。

真値が必ずあること。

項目間にグラム・シュミットの直交展開を採用した。

直交展開は逐次計算であるので、計測項目が多いと時間が掛かる。

項目の順序が結果に影響する。

順序の決め方には真値との相関係数の絶対値が大きい順、項目を主成分分析にかけ、因子付加量の大きい順等がある。

Ｔ（１）法同様、単位空間は真ん中にある。

（３）Ｔ（２）法（片側Ｔ法）

Ｔ（１）法の単位空間が真ん中であるのに対し、Ｔ（２）法では端にある。

真値が必ずあり、ランク値でもよい。

各項目毎の予測値を２乗の形で行い、項目毎ηの加重平均で２乗距離を予測する。

（４）Ｔ（３）法（ＲＴ法）

画像処理を念頭に置いたもので、真値がない。

単位空間の各項目の平均値ｍと単位空間での各サンプル毎の各項目値に対し、ゼロ点比例式のηとβを求める。

$$y_1=\beta \quad y_2=1/\sqrt{\eta}$$

としてＭＴＡ法により２乗距離を求める。

同様に単位空間の各項目の平均値ｍと信号データでの各サンプル毎の各項目値に対し、ゼロ点比例式のηとβを求める。

$$y_1=\beta \quad y_2=1/\sqrt{\eta}$$

としてＭＴＡ法により２乗距離を求め、単位空間に属するか否かを判別する。