

# 马田系统研究进展

常志朋<sup>†</sup>

- (1. 安徽工业大学 商学院, 安徽 马鞍山 243002;
2. 安徽工业大学 安徽创新驱动发展研究院, 安徽 马鞍山 243002)

**摘要:** 马田系统是一种产生于质量工程学领域中的模式识别方法,该方法能够在不平衡数据环境中进行异常值识别和特征选择,并且不需要任何数据分布假设,具有样本需求小、原理简单和易于操作等优点. 为更好地促进马田系统理论及应用研究,首先对马田系统的异常值识别和特征选择原理进行介绍;然后从马氏距离、信噪比、马氏空间、特征选择、阈值、数据环境和应用等7个方面,梳理马田系统的研究进展;最后,对马田系统研究进展进行总结,指出马田系统仍存在的问题以及未来可能的研究方向.

**关键词:** 马田系统; 马氏距离; 马氏空间; 信噪比; 正交表; 特征选择

**中图分类号:** O235      **文献标志码:** A

## Research progress of Mahalanobis-Taguchi system

CHANG Zhi-peng<sup>†</sup>

- (1. School of Business, Anhui University of Technology, Maanshan 243002, China; 2. Institute of Anhui's Innovation Driving and Development, Anhui University of Technology, Maanshan 243002, China)

**Abstract:** As a pattern recognition method generated in the quality engineering field, the Mahalanobis-Taguchi system (MTS) can carry out outlier recognition and feature selection in an imbalanced data environment without any assumption of data distribution. It has the advantages such as small samples, simple principle and easy operation. In order to better promote the theoretical and applied research of the MTS, this paper first introduces the basic principles of outlier recognition and feature selection of the MTS, and then reviews the research progress in terms of Mahalanobis distance, signal to noise ratio, Mahalanobis space, feature selection, threshold, data environment, and the application of the MTS. Finally, this paper summarizes the research progress and proposes a detailed analysis of the future possible research directions of the MTS.

**Keywords:** Mahalanobis-Taguchi system; Mahalanobis distance; Mahalanobis space; signal to noise ratio; orthogonal array; feature selection

## 0 引言

马田系统(Mahalanobis-Taguchi system, MTS)是日本著名质量工程学家田口玄一(Taguchi G)博士于20世纪90年代初提出的一种面向不平衡数据的模式识别方法,由马氏距离(Mahalanobis distance, MD)和田口方法(Taguchi's design of experiment, DOE)整合而来. 2000年, Taguchi等<sup>[1]</sup>在《New trends in multivariate diagnosis》中详细介绍了MTS的基本原理、使用方法和适用条件,标志着MTS正式产生. 同年,文献[2-3]将MTS引入我国. 2001年和2002年, Taguchi等又分别出版了两本专著<sup>[4-5]</sup>,对MTS进行了系统介绍和补充.

MTS的主要功能是面向不平衡数据进行异常值识别和特征选择,与其他经典模式识别方法相比,MTS主要具有以下6个特点: 1)不需要任何数据分布假设; 2)需要的样本量较小; 3)只有一个数据总体; 4)通过构建测量尺度进行识别; 5)能够测度异常样本的异常程度和异常方向; 6)原理简单,易于操作,便于在实践中应用. 本文将对MTS的研究进展进行全面梳理和总结,希望能更好地促进MTS的理论和应用研究.

## 1 MTS基本原理

本节将重点介绍MTS的异常值识别和特征选择原理,MTS的具体操作步骤可参见文献[1, 5].

收稿日期: 2019-01-03; 修回日期: 2019-03-12.

基金项目: 国家自然科学基金面上项目(71673001); 安徽省高校优秀青年人才支持计划重点项目(gxyqZD2017040).

责任编辑: 高会军.

<sup>†</sup>通讯作者. E-mail: changzp@126.com.

1.1 异常值识别原理

利用 MTS 进行异常值识别,首先要收集正常样本数据集,并将其定义为马氏空间 (Mahalanobis space, MS),也称基准空间;然后以 MS 为基点,利用马氏距离构建一个测量尺度,测度未知样本与 MS 的距离;最后通过设定阈值达到异常值识别的目的. MTS 异常值识别原理如图 1 所示.

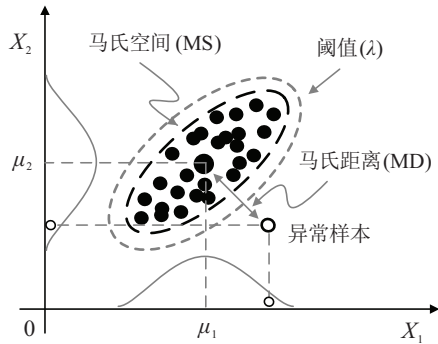


图 1 MTS 异常值识别原理

1.2 特征选择原理

为提高测量尺度的稳健性 (robustness),即降低 MD 的波动性,在利用 MTS 进行异常值识别之前,往往需要进行特征选择,将影响 MD 波动的因素去掉.

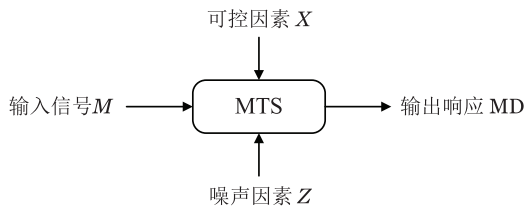


图 2 MTS 的输入和输出

如图 2,输出响应 MD 主要受两类因素影响:一类为可控因素  $X = \{X_1, X_2, \dots, X_p\}$ ;另一类为噪声因素  $Z = \{Z_1, Z_2, \dots, Z_d\}$ . 噪声因素对 MD 有影响,但很难控制,如环境因素(温度、湿度等)、诊断对象、操作人员等. 为便于讨论,可将 MD 与 X 和 Z 之间可能存在的函数关系写成如下形式:

$$MD = f(X, Z). \tag{1}$$

MD 在 X 和 Z 的共同影响下会产生一定的波动性,波动性越大,MD 的稳健性越差,MTS 的异常值识别能力也越差. MD 的波动性主要来自可控因素 X 的微小变差  $\Delta X$  和不可控的噪声因素 Z. 为使 MD 对  $\Delta X$  和 Z 不敏感,可以利用 X 与 Z 之间存在的交互效应,通过改变 X 中因素组合来减小 MD 的波动性. 下面用式(1)的一个特例来解释这一点,考虑如下模型<sup>[6]</sup>:

$$MD = \mu + \alpha X_1 + \beta Z_1 + \gamma X_2 Z_1 + \varepsilon. \tag{2}$$

其中:  $X_1$  和  $X_2$  为可控因素,  $Z_1$  为噪声因素.

在式(2)中,有一个显著的  $X_1$  主效应和一个显著的  $X_2$  与  $Z_1$  的交互作用,  $\varepsilon$  表示不包含在  $Z_1$  中的剩余变差. 如果对式(2)进一步整理,可得

$$MD = \mu + \alpha X_1 + (\beta + \gamma X_2) Z_1 + \varepsilon, \tag{3}$$

则可通过选择  $X_2$  的值使式(3)中  $Z_1$  的系数  $\beta + \gamma X_2$  的绝对值尽可能的小,从而减弱  $Z_1$  对 MD 的影响. 在进行数据分析时,可以选择使 MD 与  $Z_1$  之间的响应最平坦的可控因素组合来实现,如图 3 所示.

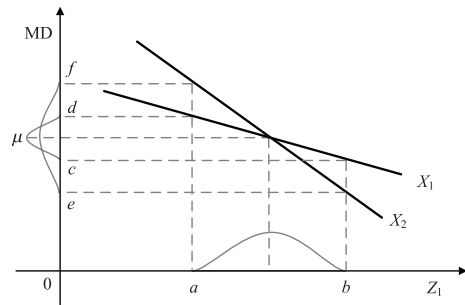


图 3 MTS 的特征选择原理

在图 3 中,可控因素  $X_2$  在  $[a, b]$  范围内可以使 MD 与  $Z_1$  之间的响应更为平坦,因此,选择可控因素  $X_2$  可以减少 MD 的波动性.

以上分析表明,选择使 MD 波动性最小的可控因素组合,可以达到挑选特征子集 (feature subset, FS) 的目的,但还需要一个衡量 FS 优劣的目标函数. 最优的 FS 应该满足两个条件:一是 FS 应使输出响应 MD 的波动性尽量小;二是 FS 应使输出响应 MD 尽量集中在信号 M 附近. 因此,可以根据如下思路构建衡量 FS 优劣的目标函数:

假设输出响应 MD 在可控因素 X 和噪声因素 Z 的共同作用下服从正态分布,即  $MD \sim N(\mu_{MD}, \sigma_{MD}^2)$ ,则输出响应 MD 的波动性可以用方差  $\sigma_{MD}^2$  描述,在信号 M 附近的集中程度可以用均值  $\mu_{MD}$  描述. 如果将  $\mu_{MD}^2$  定义为信号 (signal),  $\sigma_{MD}^2$  定义为噪声 (noise),则信号强度与噪声强度之比可以定义为信噪比 (signal to noise ratio, SNR),即

$$SNR = \frac{\mu_{MD}^2}{\sigma_{MD}^2}. \tag{4}$$

在实际计算中,通常模仿通信理论中的处理方法,将 SNR 值转化为分贝值 (dB), 即

$$SNR = 10 \log_{10} \frac{\mu_{MD}^2}{\sigma_{MD}^2}. \tag{5}$$

这样做的目的是使 SNR 值较接近于正态分布,且因素效应也大多具有线性可加性<sup>[7]</sup>. 在 MTS 中,信噪比主要用于<sup>[5]</sup>: 1) 确定有效特征变量; 2) 评估系统的性能; 3) 改善系统的性能. 信噪比对于优化系统变量和降低诊断成本而言非常重要,信噪比只利用异常

样本马氏距离来计算. 根据信号  $M$  的不同, 常用的信噪比主要有以下3种类型.

### 1) 望目特性信噪比.

对于望目特性 (nominal-the-best, NB) 信噪比, 希望  $\mu_{MD}$  值越接近目标值  $MD_0$  越好, 同时也希望  $\sigma_{MD}^2$  越小越好, 故可将望目特性信噪比定义为

$$SNR_{NB} = 10 \log_{10} \frac{\mu_{MD}^2}{\sigma_{MD}^2}. \quad (6)$$

由于方差  $\sigma_{MD}^2$  和均值  $\mu_{MD}^2$  的无偏估计分别为

$$\hat{\sigma}_{MD}^2 = V_e = \frac{1}{n} \sum_{k=1}^n (MD_k - \overline{MD})^2,$$

$$\hat{\mu}_{MD}^2 = \frac{1}{n} (S_m - V_e).$$

其中:  $\overline{MD} = \frac{1}{n} \sum_{k=1}^n MD_k$ ,  $S_m = \frac{1}{n} \left( \sum_{k=1}^n MD_k \right)^2$ ,  $k = 1, 2, \dots, n$  为样本数. 在实际计算中, 望目特性信噪比的公式为

$$SNR_{NB} = 10 \log_{10} \left( \frac{1}{n} \frac{S_m - V_e}{V_e} \right). \quad (7)$$

### 2) 望小特性信噪比.

对于望小特性 (smaller-the-better, SB) 信噪比, 希望输出响应  $MD$  越小越好, 因此不仅希望  $\mu_{MD}$  越小越好, 同时也希望  $\sigma_{MD}^2$  越小越好. 为使量纲一致, 可以要求  $\sigma_{MD}^2 + \mu_{MD}^2$  越小越好, 故可将望小特性信噪比定义为

$$SNR_{SB} = -10 \log_{10} (\sigma_{MD}^2 + \mu_{MD}^2).$$

由于  $\sigma_{MD}^2 + \mu_{MD}^2$  的无偏估计为

$$\hat{\sigma}_{MD}^2 + \hat{\mu}_{MD}^2 = E(MD^2) = \frac{1}{n} \sum_{k=1}^n MD_k^2,$$

在实际计算中, 望小特性信噪比的公式为

$$SNR_{SB} = -10 \log_{10} \left( \frac{1}{n} \sum_{k=1}^n MD_k^2 \right). \quad (8)$$

### 3) 望大特性信噪比.

对于望大特性 (larger-the-better, LB) 信噪比, 希望输出响应  $MD$  越大越好, 故一方面希望  $\mu_{MD}$  越大越好, 另一方面又希望  $\sigma_{MD}^2$  要小. 在实际计算中, 望大特性信噪比可以对望小特性信噪比中的马氏距离取倒数得到, 即

$$SNR_{LB} = -10 \log_{10} \left( \frac{1}{n} \sum_{k=1}^n \frac{1}{MD_k^2} \right). \quad (9)$$

下面根据目标函数  $SNR$  来挑选最优特征子集 (optimal feature subset,  $FS^{opt}$ ). 最优的挑选方法是穷举法 (exhaustive method, EM), 就是将所有可能的特征子集的  $SNR$  都计算出来再加以比较, 但计算量较大, 如  $|X| = p$ , 则所有可能的特征子集有  $2^p$  个. 因

此, 找到一种简单且易于操作的挑选方法是非常必要的. 需要指出的是, 任何非穷举法都不能保证所得结果最优. Taguchi 等<sup>[1]</sup> 采用 DOE 法挑选  $FS^{opt}$ , 该方法利用正交表 (orthogonal array, OA) 从特征变量集  $X$  中筛选  $FS$  ( $FS \subset X$ ). 正交表是一套科学安排试验条件的规格化表格, 具有均衡搭配和综合可比的特点, 能以少量试验获取较全面的信息. 在 MTS 中, 运用 2 水平正交表  $L_q(2^p)$  挑选特征变量, 其中  $q$  表示试验次数,  $p$  表示参与试验的特征变量个数, 水平“1”表示选择该变量, 水平“2”表示不选择该变量. 以  $|X| = 7$  为例, 假如进行全面试验需要计算  $2^7 = 128$  个  $FS$  的  $SNR$ , 如果采用  $L_8(2^7)$  正交表, 则只需根据正交表挑选 8 个  $FS$ , 然后计算其  $SNR$  即可.

选择合适类型的信噪比之后, 就可以通过优化如下目标函数来挑选最优的特征子集  $FS^{opt}$ :

$$\max_{FS \subset X} SNR_{FS} \xrightarrow{OA} FS^{opt}, \quad (10)$$

其中  $SNR_{FS}$  为利用  $FS$  计算得到的信噪比.

## 2 MTS 研究进展

本节将从马氏距离、信噪比、马氏空间、阈值、特征选择、数据环境拓展以及 MTS 应用等 7 个方面梳理 MTS 的研究进展.

### 2.1 马氏距离研究进展

与其他一些模式识别方法一样, MTS 也采用距离来测度样本之间的相似性以达到识别的目的, 但 MTS 并没有像其他一些模式识别方法采用欧氏距离, 而是采用马氏距离. 马氏距离是由印度统计学家 Mahalanobis<sup>[8]</sup> 提出的一种协方差距离, 可以消除特征变量之间的相关性以及量纲影响. 因此, 相较于欧氏距离, 马氏距离更适合进行异常值诊断. 在 MTS 中, 采用尺度化的平方马氏距离, 其计算公式为

$$MD_k^2 = \frac{1}{p} \mathbf{z}_k^T \mathbf{R}^{-1} \mathbf{z}_k. \quad (11)$$

其中:  $p$  为特征变量个数,  $\mathbf{z}_k$  为第  $k$  个样本的标准化向量,  $\mathbf{R}$  为样本标准化后总体的相关系数矩阵.

通过尺度化处理, 可以使平方马氏距离的期望值在 1 左右分布<sup>[5]</sup>. 为叙述方便, 以下将尺度化的平方马氏距离简称为马氏距离. 式 (11) 一般称为逆矩阵马氏距离, 常志朋等<sup>[9]</sup> 在此基础上, 提出了一种加权逆矩阵马氏距离, 该距离充分考虑了决策者的主观偏好, 其计算公式如下:

$$MD_k^2 = \frac{1}{p} \mathbf{z}_k^T \mathbf{W} \mathbf{R}^{-1} \mathbf{W} \mathbf{z}_k. \quad (12)$$

其中:  $\mathbf{W} = \text{diag}[w_1, w_2, \dots, w_p]$ ,  $0 \leq w_j \leq 1$  为第  $j$

个变量的权重,且  $\sum_{j=1}^p w_j = 1$ .

当特征变量存在多重共线性 (multicollinearity) 时,逆矩阵马氏距离很难计算或计算不准确. Taguchi 等<sup>[5]</sup> 提出通过施密特正交化 (Gram-Schmidt orthogonalization process, GSP) 的方法计算马氏距离,该方法不依赖  $\mathbf{R}^{-1}$ , 因而不受多重共线性的影响,并且利用该方法计算的马氏距离与利用逆矩阵法计算的马氏距离相等<sup>[5]</sup>,其计算公式为

$$MD_k^2 = \frac{1}{p} \sum_{j=1}^p \frac{u_{kj}^2}{\hat{s}_j^2}. \quad (13)$$

其中:  $u_{kj}$  为施密特正交化后第  $k$  个样本的第  $j$  个变量值,  $\hat{s}_j$  为施密特正交化后第  $j$  个变量的标准差.

Su 等<sup>[10]</sup> 相应地构建了加权施密特正交马氏距离,即

$$MD_k^2 = \frac{1}{p} \sum_{j=1}^p w_j \frac{u_{kj}^2}{\hat{s}_j^2}. \quad (14)$$

陶建波等<sup>[11]</sup> 提出利用岭估计 (ridge estimate) 处理多重共线性问题,构建岭马氏距离,其计算公式为

$$MD_k^2 = \frac{1}{p} \mathbf{z}_k^T (\mathbf{R} + \eta \mathbf{I})^{-1} \mathbf{z}_k. \quad (15)$$

其中:  $\eta \in [0, 1]$  为岭参数,  $\mathbf{I}$  为单位矩阵.

Taguchi 等<sup>[5]</sup> 认为,伴随矩阵与逆矩阵具有相同的性质,因而也可以利用相关矩阵  $\mathbf{R}$  的伴随矩阵  $\mathbf{R}_{adj}$  计算马氏距离,将其称为伴随矩阵马氏距离,具体计算公式为

$$MD_k^2 = \frac{1}{p} \mathbf{z}_k^T \mathbf{R}_{adj} \mathbf{z}_k. \quad (16)$$

常志朋等<sup>[12]</sup> 采用伪逆马氏距离处理多属性决策问题,即

$$MD_k^2 = \frac{1}{p} \mathbf{z}_k^T \mathbf{R}^+ \mathbf{z}_k, \quad (17)$$

其中  $\mathbf{R}^+$  为  $\mathbf{R}$  的伪逆.

Hwang 等<sup>[13]</sup> 针对特征变量有时呈现正偏度或负偏度分布,提出一种偏度 (skewness) 马氏距离,其计算公式为

$$SMD_k^2 = \frac{1}{p} \left( \sum_{j=1}^p \left( \frac{x_{kj} - \bar{x}_j}{s_j} \right)^3 \right) MD_k^2. \quad (18)$$

其中:  $x_{kj}$  为第  $k$  个样本的第  $j$  个变量值,  $\bar{x}_j$  为第  $j$  个变量的均值,  $s_j$  为第  $j$  个变量的标准差.

上述马氏距离的计算均是针对单点的实数数据. 常志朋等<sup>[14]</sup> 针对双点的区间数据,提出了一种区间马氏距离,其计算方法如下: 设  $[\mathbf{X}] = [\mathbf{x}_{kj}^L, \mathbf{x}_{kj}^U]_{n \times p}$  为区间样本矩阵,则第  $k$  个区间样本

$$[\mathbf{x}_k^L, \mathbf{x}_k^U] = [[x_{k1}^L, x_{k1}^U], [x_{k2}^L, x_{k2}^U], \dots, [x_{kp}^L, x_{kp}^U]]^T$$

的区间马氏距离为

$$[MD_k^L, MD_k^U] = \min/\max \left[ \sqrt{\frac{1}{p} (\mathbf{z}_k^L)^T \bar{\mathbf{R}}^{-1} \mathbf{z}_k^L}, \sqrt{\frac{1}{p} (\mathbf{z}_k^U)^T \bar{\mathbf{R}}^{-1} \mathbf{z}_k^U} \right]. \quad (19)$$

其中:  $\mathbf{z}_k^L$  和  $\mathbf{z}_k^U$  分别为  $\mathbf{x}_k^L$  和  $\mathbf{x}_k^U$  的标准化向量;  $z_{kj}^L$  和  $z_{kj}^U$  可利用下式计算:

$$z_{kj}^L = \frac{x_{kj}^L - \bar{x}_j}{s_j}, \quad z_{kj}^U = \frac{x_{kj}^U - \bar{x}_j}{s_j}.$$

$\bar{x}_j$  和  $s_j$  分别为第  $j$  个特征变量的均值和标准差,其计算公式如下:

$$\begin{aligned} \bar{x}_j &= \frac{1}{2n} \sum_{k=1}^n (x_{kj}^L + x_{kj}^U), \\ s_j^2 &= \frac{1}{3n} \sum_{k=1}^n [(x_{kj}^L)^2 + x_{kj}^L x_{kj}^U + (x_{kj}^U)^2] - \\ &\quad \frac{1}{4n^2} \left[ \sum_{k=1}^n (x_{kj}^L + x_{kj}^U) \right]^2. \end{aligned}$$

$\bar{\mathbf{R}}$  为相关系数矩阵,有

$$\bar{\mathbf{R}} = \begin{bmatrix} s_{ij} \\ s_i s_j \end{bmatrix}_{p \times p}, \quad i, j = 1, 2, \dots, p,$$

$s_{ij}$  为特征变量  $i$  与  $j$  之间的协方差,即

$$\begin{aligned} s_{ij}^2 &= \frac{1}{4n} \sum_{k=1}^n (x_{ki}^L + x_{ki}^U)(x_{kj}^L + x_{kj}^U) - \\ &\quad \frac{1}{4n^2} \left[ \sum_{k=1}^n (x_{ki}^L + x_{ki}^U) \right] \left[ \sum_{k=1}^n (x_{kj}^L + x_{kj}^U) \right]. \end{aligned}$$

综上,现有的研究成果主要使马氏距离具备以下几个功能: 1) 能够处理共线性问题; 2) 能够处理相关度很低的问题; 3) 能够判断异常样本的异常方向; 4) 能够考虑主观偏好; 5) 能够处理偏度数据; 6) 能够处理区间数据. 表 1 总结了不同类型马氏距离的功能及代表性文献.

表 1 不同类型马氏距离的功能及代表性文献

类型	公式	功能	代表性文献
逆矩阵马氏距离	(11)	2)	[5]
加权逆矩阵马氏距离	(12)	2)、4)	[9]
施密特正交马氏距离	(13)	1)、2)、3)	[5]
加权施密特正交马氏距离	(14)	1)、2)、3)、4)	[10]
岭马氏距离	(15)	1)、2)	[11]
伴随矩阵马氏距离	(16)	2)	[5]
伪逆马氏距离	(17)	1)、2)	[12]
偏度马氏距离	(18)	2)、5)	[13]
区间马氏距离	(19)	2)、6)	[14]

## 2.2 信噪比研究进展

传统信噪比的计算主要基于单点实数型马氏距离,常志朋等<sup>[14]</sup> 针对区间马氏距离,分别推导了望大特性、望小特性和望目特性 3 种区间信噪比 (interval

signal to noise ratio, ISNR),其计算公式分别如下:

$$ISNR_{LB} = -10 \log_{10} \left( \frac{1}{n} \sum_{k=1}^n \frac{1}{MD_k^U MD_k^L} \right), \quad (20)$$

$$ISNR_{SB} = -10 \log_{10} \left[ \frac{1}{3n} \sum_{k=1}^n ((MD_k^U)^2 + MD_k^U MD_k^L + MD_k^U MD_k^L + (MD_k^L)^2) \right], \quad (21)$$

$$ISNR_{NB} = -10 \log_{10} \left[ \frac{1}{3(n-1)} \sum_{k=1}^n ((MD_k^U)^2 + MD_k^U MD_k^L + (MD_k^L)^2) - \frac{1}{4n(n-1)} \left( \sum_{k=1}^n (MD_k^U + MD_k^L) \right)^2 \right]. \quad (22)$$

其中:  $n$  表示样本的个数,  $[MD_k^L, MD_k^U]$  表示第  $k$  个样本的区间马氏距离.

### 2.3 马氏空间构建方法研究进展

马氏空间(MS)一般依靠专业人员的领域内知识或丰富经验来界定,但存在一定的主观性和不确定性. 针对这种情况, Yang 等<sup>[15]</sup> 提出利用质量控制图构建马氏空间, 见图4.

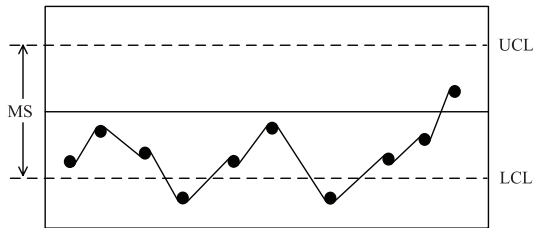


图4 质量控制图

在图4中,  $UCL = \bar{x}_{MD} + 3s_{MD}$  和  $LCL = \bar{x}_{MD} - 3s_{MD}$  分别为控制上限和下限, 其中  $\bar{x}_{MD}$  和  $s_{MD}$  分别为样本马氏距离的均值和标准差, 可以分别利用下式计算:

$$\bar{x}_{MD} = \frac{1}{n} \sum_{k=1}^n MD_k^2, \quad s_{MD} = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (MD_k^2 - \bar{x}_{MD})^2}.$$

Wang 等<sup>[16]</sup> 采用 Soylemezoglu 等<sup>[17]</sup> 的实例数据, 对 Yang 等<sup>[15]</sup> 的方法进行了有效性验证. Das 等<sup>[18]</sup> 提出利用  $k$ -means 聚类分析构建马氏空间, 并在此基础上构建了一种非监督 MTS 算法. Liparas 等<sup>[19-20]</sup> 分别采用两步聚类分析 (two-step cluster analysis) 和  $k$ -means 聚类分析构建马氏空间.

由于数据包络分析 (DEA) 模型可以将有效率的样本和无效率的样本分开, 叶芳羽等<sup>[21]</sup> 采用 DEA 中的 CCR 模型构建马氏空间, 该模型具体形式如下:

$$\begin{aligned} & \min \theta_o. \\ & \text{s.t.} \sum_{j=1}^n \xi_j x_{ij} \leq \theta_o x_{io}, \quad i = 1, 2, \dots, m; \\ & \sum_{j=1}^n \xi_j y_{rj} \geq y_{ro}, \quad r = 1, 2, \dots, s; \\ & \xi_j \geq 0, \quad j = 1, 2, \dots, n. \end{aligned} \quad (23)$$

其中:  $\theta_o$  为被评价决策单元的 DEA 效率值,  $\xi_j$  为产出和投入的权重系数,  $x_{ij}$  和  $y_{rj}$  分别为第  $j$  个决策单元的第  $i$  个输入指标和第  $r$  个输出指标,  $x_{io}$  和  $y_{ro}$  分别为被评价决策单元的第  $i$  个输入指标和第  $r$  个输出指标.

### 2.4 阈值确定方法研究进展

MTS 的另外一个重要任务是合理确定识别阈值  $\lambda$ . Taguchi 等<sup>[5]</sup> 根据“经济损失”与“功能波动”之间的函数关系, 提出利用质量损失函数法 (quality loss function, QLF) 确定阈值. 然而, QLF 法缺乏合理的统计解释, 主要依据领域内知识确定阈值, 存在一定的主观性<sup>[22-23]</sup>, 并且在实际应用中确定每个样本的相对成本或损失很困难. 针对这种情况, 一些替代 QLF 的阈值确定方法相继提出, 通过梳理可以发现, 这些方法主要从以下两个方面切入确定阈值:

1) 从 MTS 分类结果切入. 一些学者<sup>[24-26]</sup> 根据 MTS 分类结果, 采用两类错误法确定最优阈值  $\lambda$ . Liparas 等<sup>[19]</sup> 在两类错误法的基础上, 提出采用“受试者工作特征” (receiver operating characteristics, ROC) 曲线确定阈值  $\lambda$ , 但是该方法要通过不断地试探来确定最优阈值  $\lambda$ . Mahmoud 等<sup>[27]</sup> 根据 ROC 曲线构建了一种确定  $\lambda$  值的优化模型.

2) 从马氏距离的概率分布切入. 由于马氏距离的概率分布未知, Su 等<sup>[28]</sup> 利用切比雪夫定理 (Chebyshev's theorem) 提出了一种概率阈值计算方法, 其计算公式为

$$\lambda = \bar{x}_{MD} + \sqrt{\frac{1}{1 + \delta - \theta}} \cdot s_{MD}. \quad (24)$$

其中:  $\bar{x}_{MD}$  为正常样本马氏距离平均值,  $s_{MD}$  为正常样本马氏距离的标准差,  $\delta$  为一个比较小的参数,  $\theta$  表示在正常样本马氏距离中比异常样本中最小马氏距离还小的百分比.

牛俊磊等<sup>[29]</sup> 在此基础上根据样本误分率构建了一种优化模型来确定阈值  $\lambda$ , 优化模型为

$$\begin{aligned} & \min \alpha. \\ & \text{s.t.} \quad 1 - \beta \leq \frac{s_{md}^2}{(\bar{x}_{MD} + s_{MD}/\sqrt{\alpha} - \bar{x}_{md})}, \\ & \alpha = \beta. \end{aligned} \quad (25)$$

其中:  $\alpha$  为将正常样本误判为异常的概率上限,  $\beta$  为将异常样本误判为正常的概率上限,  $\bar{x}_{MD}$  和  $s_{MD}$  分别为正常样本马氏距离的均值和标准差,  $\bar{x}_{md}$  和  $s_{md}$  分别为异常样本马氏距离的均值和标准差.

Huang 等<sup>[30]</sup> 根据正常样本和异常样本重叠比率, 提出一种确定阈值方法, 其计算公式为

$$\lambda = \bar{x}_{MD} \cdot \sqrt{\frac{100}{100 + \alpha - \nu}} \cdot s_{MD}. \quad (26)$$

其中:  $\alpha$  为模型置信水平,  $\nu$  为正常样本马氏距离与异常样本马氏距离的重叠比率.

Kumar 等<sup>[31]</sup> 提出对马氏距离进行 Box-Cox 变换, 将其转换为正态分布, 具体转换公式如下:

$$MD_k^2(\gamma) = \begin{cases} \frac{1}{\gamma} ((MD_k^2)^\gamma - 1), & \gamma \neq 0; \\ \ln(MD_k^2), & \gamma = 0. \end{cases} \quad (27)$$

其中:  $MD_k^2$  为第  $k$  个样本的马氏距离;  $MD_k^2(\gamma)$  为  $MD_k^2$  的 Box-Cox 变换值,  $\gamma$  值由如下最大似然求得:

$$\max_{\gamma} f(\gamma) = -\frac{n}{2} \ln \left[ \frac{1}{n} \sum_{k=1}^n (MD_k^2(\gamma) - \bar{x}_{MD(\gamma)})^2 \right] + (\gamma - 1) \sum_{k=1}^n \ln(MD_k^2), \quad (28)$$

$\bar{x}_{MD(\gamma)}$  为经过 Box-Cox 变换后, 正常样本马氏距离的均值. 由于  $MD_k^2(\gamma)$  服从正态分布, 且  $MD_k^2(\gamma) > 0$ , 可以采用一元质量控制图确定阈值. 文献<sup>[31]</sup> 给出了两种阈值的计算公式, 即

$$\lambda = \bar{x}_{MD(\gamma)} + 2s_{MD(\gamma)}, \quad \lambda = \bar{x}_{MD(\gamma)} + 3s_{MD(\gamma)},$$

其中  $s_{MD(\gamma)}$  为经过 Box-Cox 变换后, 正常样本马氏距离的标准差.

Johnson 等<sup>[32]</sup> 指出马氏距离一般服从卡方分布, 因此未被尺度化的马氏距离服从自由度为  $p$  的卡方分布, 进而可知尺度化马氏距离具有如下性质:

$$MD^2 = \frac{1}{p} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \frac{1}{p} \chi^2(p). \quad (29)$$

其中:  $\boldsymbol{\mu}$  为总体均值向量,  $\boldsymbol{\Sigma}$  为总体协方差矩阵.

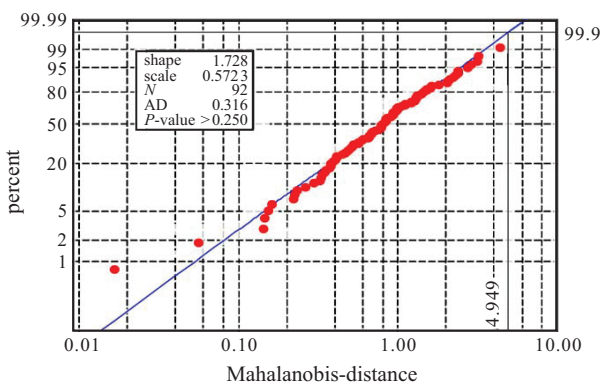


图5 正常样本马氏距离的伽马分布

Das 等<sup>[18,33]</sup> 根据卡方分布提出一种确定阈值方法, 其计算公式为

$$\lambda = \frac{1}{p} \chi^2(p). \quad (30)$$

由于卡方分布是更一般的伽马分布<sup>[32]</sup>, Rai 等<sup>[34]</sup> 提出通过对正常样本马氏距离进行伽马分布拟合来确定阈值, 如图5所示.

### 2.5 特征选择研究进展

MTS 采用正交表和信噪比进行特征选择, 但是 Woodall 等<sup>[22]</sup> 指出该方法存在如下缺陷: 1) 缺乏完备的统计理论基础; 2) 用正交表筛选出的特征子集, 其信噪比未必较大; 3) 信噪比较大的特征子集其分类效果未必良好. 因此建议使用搜索算法进行特征选择. Abraham 等<sup>[35]</sup> 进一步证实正交表和信噪比并不是最优的特征子集选择策略, 并质疑其是否为最优的特征选择方法. Jagulum 等<sup>[36]</sup> 指出, 如果一种特征选择方法能够构建更好的测量尺度, 则可将其应用于 MTS.

在上述背景下, 一些学者尝试采用其他方法进行特征选择, 如 Foster 等<sup>[37]</sup> 采用自适应一次一个因子设计 (adaptive one-factor-at-a-time, aOFAT) 法代替正交表, 即采用下式进行特征选择:

$$\max_{FSCX} SNR_{FS} \xrightarrow{\text{aOFAT}} FS^{\text{opt}}. \quad (31)$$

Foster 发现利用 aOFAT 筛选特征变量比正交表成本更低, 并且速度更快.

Ramlie 等<sup>[38]</sup> 利用蜂群算法 (bees algorithm, BA) 对下式优化求解得到  $FS^{\text{opt}}$ :

$$\max_{FSCX} SNR_{FS} \xrightarrow{\text{BA}} FS^{\text{opt}}. \quad (32)$$

Pal 等<sup>[39]</sup> 根据样本误分率和变量选择率构建了如下二进制多目标整数规划模型:

$$\begin{aligned} \min f(\mathbf{x}) &= \alpha \underbrace{\left( w_1 \frac{n_1^e}{n_1} + w_2 \frac{n_2^e}{n_2} \right)}_{f_1(\mathbf{x})} + \beta \frac{p_{\text{selected}}}{p}; \\ \text{s.t. } &\sum_{j=1}^p x_j \leq p, \\ &\sum_{j=1}^p x_j = p_{\text{selected}}, \\ &f_1(\mathbf{x}) \leq f_1^{\text{max}}. \end{aligned} \quad (33)$$

其中:  $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$  是一个  $p$  维向量, 如果变量  $x_j$  未被选择, 则  $x_j = 0$ , 如果变量  $x_j$  被选择, 则  $x_j = 1$ ;  $p_{\text{selected}}$  为被选择的变量数;  $n_1$  为正常样本数,  $n_2$  为异常样本数;  $n_1^e$  是正常样本被错分为异常样本的数量,  $n_2^e$  是异常样本被错分为正常样本的数

量;  $w_1 = c_1/(c_1 + c_2)$ ,  $w_2 = c_2/(c_1 + c_2)$ ,  $c_1$  和  $c_2$  分别为错分正常样本和异常样本的成本;  $f_1^{\max}$  为所有变量均被选择时的误分率。

对于该模型的求解, Pal 等<sup>[39]</sup> 和 Edgar 等<sup>[40]</sup> 分别尝试利用二进制粒子群优化 (binary particle swarm optimization, BPSO) 算法和二进制蚁群 (binary ant colony optimization, BACO) 算法对该模型进行求解. Edgar<sup>[41]</sup> 引入戈珀兹二进制粒子群优化 (Gompertz binary particle swarm optimization, GBPSO) 算法对式 (33) 展开求解, 并与 BPSO 和 BACO 两种算法进行比较, 结果表明 GBPSO 算法比 BPSO 算法和 BACO 算法更快, 因此建议使用 GBPSO 算法. Yadira 等<sup>[42]</sup> 采用粘性二进制粒子群优化 (sticky binary particle swarm optimization, SBPSO) 算法对该模型进行求解, 并与 BPSO 和 GBPSO 两种方法进行比较, 结果表明 3 种方法得到的  $FS^{opt}$  相同, 但是 SBPSO 收敛速度最快, Yadira 推荐使用 SBPSO 算法。

牛俊磊等<sup>[43]</sup> 在式 (33) 的基础上, 根据样本误分率、变量选择率和望大特性信噪比构建多目标优化模型, 并采用全方位优化 (omni-optimizer, OO) 算法对该模型进行求解. 通过 UCI 数据进行实例验证表明, 该多目标优化模型的特征选择和分类效果均优于传统 MTS. 另外, 牛俊磊等<sup>[29]</sup> 还针对不平衡数据环境, 采用  $g$ -means、 $F$ -value、降维效率和望大特性信噪比构建多目标优化模型, 同样采用全方位优化算法对模型求解. 通过 8 个 UCI 数据集进行实例验证, 结果表明该模型能够处理不平衡数据的异常值识别问题, 并且降维效果显著。

上述研究没有考虑过拟合和正则化问题. Iquebal 等<sup>[44]</sup> 基于粗糙集理论 (rough set theory, RST), 利用特征子集  $FS$  和样本类别集  $\omega$  之间的依赖度构建如下目标函数, 并采用 GA 对其优化得到  $FS^{opt}$ :

$$\max_{FS \subseteq X} \delta_{FS}(\omega) \xrightarrow{GA} FS^{opt}. \quad (34)$$

其中:  $\omega$  为样本类别集,  $\delta_{FS}(\omega)$  为  $FS$  对  $\omega$  的依赖度。

由于该方法充分考虑了过拟合和正则化问题, 其分类效果优于 PSO 法. Liparas 等<sup>[45]</sup> 根据最小正常样本马氏距离和异常样本马氏距离分布的重叠率, 即最大化图 6 中阴影面积, 构建如下目标函数:

$$\max_{FS \subseteq X} f(FS) = \frac{h}{n_2} \xrightarrow{GA} FS^{opt}. \quad (35)$$

其中:  $h$  为比最大正常样本马氏距离大的异常样本数,  $n_2$  为异常样本数。

对于该模型的求解, Liparas 等<sup>[45]</sup> 采用 GA 法对其优化求解得到  $FS^{opt}$ , 并分别与传统 MTS、朴素贝叶

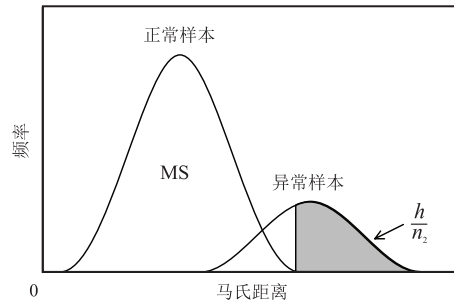


图 6 阴影面积分布

斯 (naive Bayes)、支持向量机 (support vector machine, SVM)、多层感知器 (multilayer perceptron, MP)、决策树 (decision tree, DT) 和随机森林 (random forest, RF) 等方法进行分类性能比较. 验证结果表明该方法在训练准确率、测试准确率、敏感度、AUC (area under ROC curve) 和相对敏感度等 5 类性能指标均优于其他方法。

对以上特征选择方法的目标函数、优化方法及代表性文献进行总结, 具体见表 2。

表 2 特征选择的目标函数、优化方法及代表性文献

目标函数	优化方法	代表性文献
信噪比	OA	[1]
	aOFAT	[37]
	BA	[38]
样本误分率、变量选择率	BPSO	[39]
	BACO	[40]
	GBPSO	[41]
	SBPSO	[42]
$g$ -means、 $F$ -value、降维效率、信噪比	OO	[29]
样本误分率、变量选择率、信噪比	OO	[43]
依赖度(粗糙集)	GA	[44]
重叠率	GA	[45]

## 2.6 数据环境拓展研究进展

传统 MTS 只能应用于低维、静态的实数环境中, 常志朋等<sup>[14]</sup> 将 MTS 数据环境拓展到低维、静态的区间数据环境. 但是随着互联网和大数据技术的飞速发展, 现代工业过程可以在线获得大量的高维动态数据. 为使 MTS 的数据环境拓展到高维动态数据环境中, 一些学者提出首先对高维动态数据进行特征提取, 然后再利用 MTS 进行特征选择或异常值诊断. 如 Lv 等<sup>[46]</sup> 利用多重分形去趋势分析 (multifractal detrended fluctuation analysis, MF-DFA) 算法从时间序列数据中提取非线性特征, 然后利用 MTS 进行诊断. Hu 等<sup>[47]</sup> 利用多重分形谱 (multifractal spectrum) 和广义分形维数 (generalized fractal dimensions) 提取振动信号数据的特征, 然后利用 MTS 进行特征选择。

Wang 等<sup>[48]</sup> 针对非线性、非平稳的振动信号数

据,首先利用经验模态分解 (empirical mode decomposition, EMD)把信号分解为若干个本征模态分量 (intrinsic mode functions, IMF),然后利用奇异值分解提取特征向量,最后利用 MTS 进行特征选择和诊断. 陈俊洵等<sup>[49]</sup>将该方法应用于滚动轴承的质量诊断. 陈俊洵等<sup>[50]</sup>利用 EMD 对振动信号数据提取 IMF,然后直接应用 MTS 进行特征选择. Chen 等<sup>[51]</sup>利用集成经验模态分解 (ensemble empirical mode decomposition, EEMD)对振动信号数据提取 IMF,然后利用 MTS 进行特征选择.

Lu 等<sup>[52]</sup>针对非线性、非平稳的信号,首先利用小波包变换 (wavelet packet transforms, WPT) 和自回归 (autoregression, AR) 模型构建特征矩阵,然后利用奇异值分解提取特征向量,最后利用 MTS 进行特征向量选择和诊断. Zhao 等<sup>[53]</sup>提出利用拉普拉斯特征映射 (Laplacian eigenmaps, LE) 对高维非线性数据进行特征提取,然后利用 MTS 进行特征选择. Ohkubo 等<sup>[54]</sup>提出利用稀疏主成分分析 (sparse principal component analysis, SPCA) 对高维数据进行特征提取,然后利用 MTS 进行特征选择.

通过以上梳理,对 MTS 的高维数据环境、相应的特征提取方法以及代表性文献进行归纳,具体见表 3.

表3 高维数据环境、特征提取方法及代表性文献

数据环境	特征提取方法	代表性文献
高维数据	稀疏主成分分析	[54]
高维非线性数据	拉普拉斯特征映射	[53]
时间序列数据	多重分形去趋势分析	[46]
	多重分形谱和广义分形维数	[47]
振动信号数据	经验模态分解和奇异值分解	[48]
	经验模态分解	[50]
	集成经验模态分解	[51]
非线性、非平稳的信号	小波包变换和自回归模型	[52]

2.7 应用研究进展

MTS 作为一种产生于质量工程学领域的模式识别方法,既可以通过马氏距离构建测量尺度以达到诊断产品质量的目的,也可以基于田口正交试验设计原理进行特征选择,使产品质量特性波动小,稳健性好,从而达到提高产品质量的目的. 因此,MTS 被广泛用于产品质量诊断和提高产品质量,如 Riho 等<sup>[55]</sup>利用 MTS 对晶片质量进行诊断; Hsiao 等<sup>[56]</sup>利用 MTS 对萨克斯风音色进行诊断; Das 等<sup>[57]</sup>利用 MTS 筛选热轧钢的化学成分,以提高热轧钢的铸造质量; Shakya 等<sup>[58]</sup>采用 MTS 对机器轴承进行故障诊断.

除了以上传统应用领域,MTS 还被拓展到其他应用领域,如 Nagao 等<sup>[59]</sup>利用 MTS 进行人脸识别; Cudney 等<sup>[60]</sup>利用 MTS 对汽车驾驶满意度进行评价; Mahalakshmi 等<sup>[61]</sup>利用 MTS 制定虾类养殖业标准; Lee 等<sup>[24]</sup>利用 MTS 预测台湾电子行业的金融风险; Su 等<sup>[62-63]</sup>利用 MTS 进行褥疮和阻塞性睡眠窒息症诊断; Ketkar 等<sup>[64]</sup>利用 MTS 对 MBA 入学成绩进行综合评价; Hwang 等<sup>[13]</sup>利用 MTS 构建了一种多目标优化算法; 常志朋等<sup>[9,12,14,65-68]</sup>利用 MTS 计算属性权重、模糊测度以及处理区间属性信息等.

通过梳理,对 MTS 的应用领域、对应的 MTS 功能以及代表性文献进行总结,具体见表 4.

表4 MTS 应用领域、对应功能及代表性文献

应用领域	对应的 MTS 功能	代表性文献
产品质量诊断	异常值识别、特征选择	[55-56]
提高产品质量	特征选择	[57]
机械故障诊断	异常值识别、特征选择	[58]
人脸识别	异常值识别、特征选择	[59]
满意度评价	异常值识别、特征选择	[60]
标准制定	特征选择	[61]
风险识别	异常值识别、特征选择	[24]
医疗诊断	异常值识别、特征选择	[62-63]
综合评价	异常值识别、特征选择	[64]
优化算法	异常值识别	[13]
多属性决策	特征选择	[9,12,14,65-68]

3 结论与展望

MTS 作为一种面向不平衡数据的模式识别方法,具有独特的异常值识别和特征选择原理. 该方法不仅可以应用在产品质量诊断、医疗诊断、风险识别、多属性决策等领域,未来还可以在恐怖分子识别、多维贫困识别、社会风险监测等领域有着广泛的应用. 本文对 MTS 的基本原理进行了介绍,并对 MTS 自 2000 年正式提出以来的研究进展进行了梳理. 从梳理的结果看,大致得出以下结论:

1) MTS 已经由二分类判别拓展到多分类判别<sup>[10,69]</sup>,由监督学习拓展到非监督学习<sup>[18]</sup>,基本形成了相对完整的理论框架,见图 7.

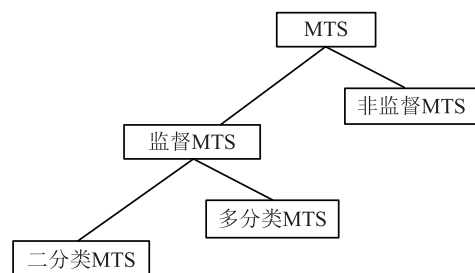


图7 MTS 理论框架

2) 从研究内容看,最近几年更偏向于应用研究(见图8),而理论研究的重点主要集中在马氏距离、信

噪比、马氏空间构建、特征选择方法、阈值确定方法、数据环境拓展等领域,见图9。

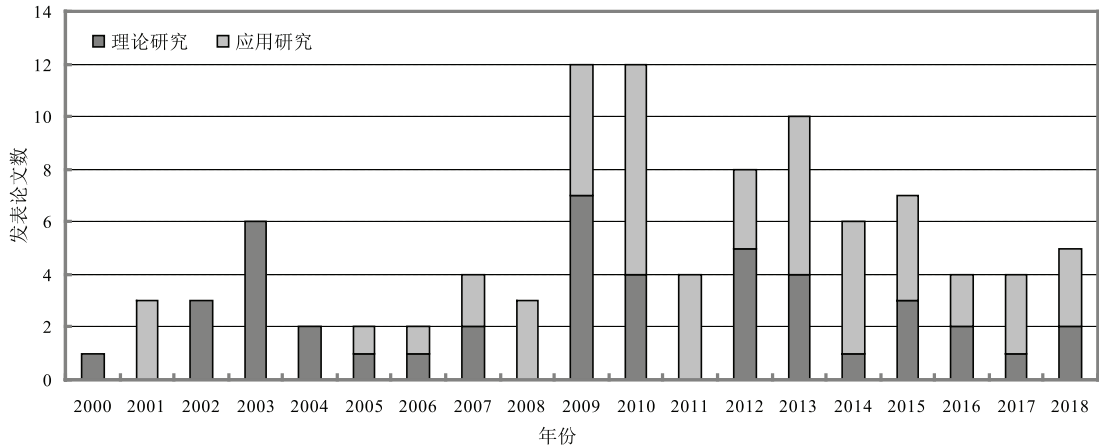


图8 MTS理论研究和应用研究文献分布情况

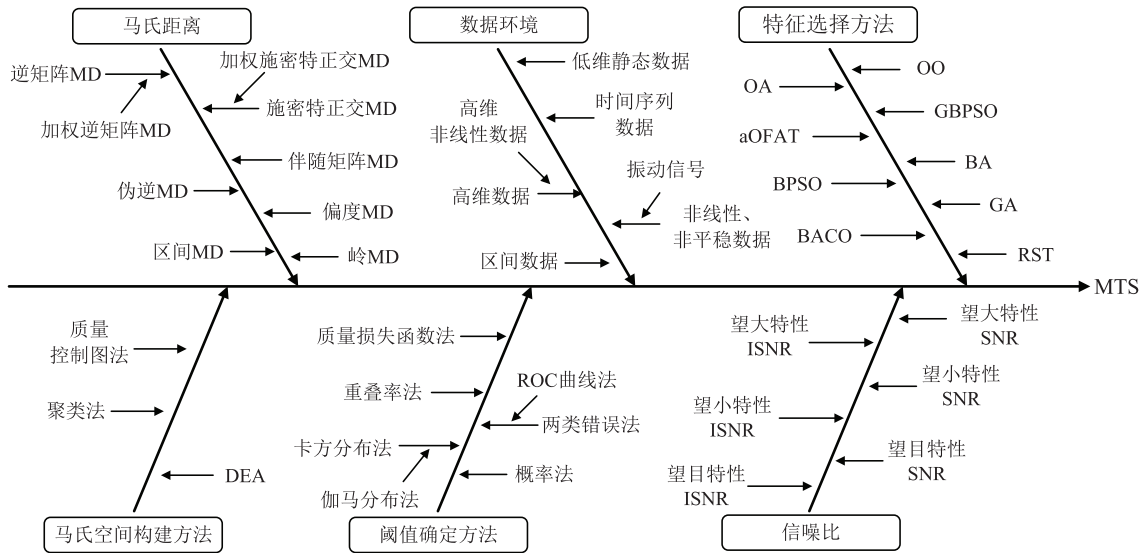


图9 MTS理论研究进展框架

经过近20年的发展,MTS虽然在理论研究和应用研究都取得了一定进展,但同时也存在一些不足,如采用智能搜索算法替代正交表进行特征选择,虽然提高了MTS的性能,但很大程度上增加了MTS计算的复杂性,有悖于MTS提出的初衷:原理简单、易于操作。因此,如何在提高MTS性能与降低计算复杂性之间作出平衡,是未来应该重点关注的。从现有的研究进展看,MTS的异常值识别主要还是依靠阈值,这种识别方式对一些线性不可分数据识别效果较差,因此如何通过构建判别函数来识别异常值,是一个值得深入研究的问题。

随着大数据时代的到来,以及深度学习技术的快速发展,MTS还有很大的研究空间,未来应重点关注以下两个问题:一是MTS在多源异构数据环境中的应用问题。目前,MTS在动态、高维数据环境中的应用研究已经取得了一定进展,但是随着“互联网+”

时代的到来,数据来源不但多样,而且存在很大的异构性。如何将异构数据同构化,使传统的MTS基础理论框架适用于多源异构数据环境,将是一个很大的挑战。二是MTS与深度学习技术的交叉融合问题。近些年,深度学习领域涌现出很多前沿的新方法、新理论,如果将其与MTS深度融合,则将为MTS的理论研究提供更广阔的空间。

参考文献(References)

[1] Taguchi G, Jugulum R. New trends in multivariate diagnosis[J]. The Indian Journal of Statistics, 2000, 62B(2): 233-248.

[2] 郑称德, 韩之俊. MTS原理及其设计模型[J]. 管理工程学报, 2000, 14(3): 43-47. (Zheng C D, Han Z J. Theory and design model of Mahalanobis-Taguchi system[J]. Journal of Industrial Engineering and Engineering Management, 2000, 14(3): 43-47.)

- [3] 李昭阳, 韩之俊. 一种新的判别预测方法——马田系统(MTS)[J]. 管理工程学报, 2000, 14(2): 54-55.  
(Li Z Y, Han Z J. A new discriminant and prediction method—Mahalanobis-Taguchi system[J]. Journal of Industrial Engineering and Engineering Management, 2000, 14(2): 54-55.)
- [4] Taguchi G, Chowdhury S. The Mahalanobis-Taguchi system[M]. New York: McGraw-Hill, 2001.
- [5] Taguchi G, Jugulum R. The Mahalanobis-Taguchi strategy: A pattern technology system[M]. New York: John Wiley & Sons, 2002.
- [6] Jeff Wu, Hamada Michael. Experiments planning, analysis, and parameter design optimization[M]. Beijing: China Statistics Press, 2003: 518-520.
- [7] 韩之俊, 章渭基. 质量工程学——线内, 线外质量管理[M]. 北京: 科学出版社, 1991: 113-120.  
(Han Z J, Zhang W J. Quality engineering—Outside and inside quality management[M]. Beijing: Science Press, 1991: 113-120.)
- [8] Mahalanobis P C. On the generalized distance in statistics[C]. Proceedings of the National Institute of Sciences of India. New Delhi, 1936, 2: 49-55.
- [9] 常志朋, 程龙生. 基于马田系统和模糊积分的多属性决策方法[J]. 管理工程学报, 2015, 29(3): 107-115.  
(Chang Z P, Cheng L S. Multi-attribute decision making method based on Mahalanobis-Taguchi system and fuzzy integral[J]. Journal of Industrial Engineering and Engineering Management, 2015, 29(3): 107-115.)
- [10] Su C T, Hsiao Y H. Multiclass MTS for simultaneous feature selection and classification[J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(2): 192-205.
- [11] 陶建波, 程龙生, 王会灵, 等. 基于岭估计和 AMOGA 的马田系统分类方法[J]. 系统工程, 2017, 35(4): 137-142.  
(Tao J B, Cheng L S, Wang H L, et al. Classification approach using MTS based on ridge estimation and AMOGA[J]. Systems Engineering, 2017, 35(4): 137-142.)
- [12] 常志朋, 程龙生. 基于马田系统与 TOPSIS 的区间数多属性决策方法[J]. 系统工程理论与实践, 2014, 34(1): 168-175.  
(Chang Z P, Cheng L S. Multiple attribute decision making method with intervals based on Mahalanobis-Taguchi system and TOPSIS method[J]. Systems Engineering—Theory & Practice, 2014, 34(1): 168-175.)
- [13] Hwang I J, Park G J. A multi-objective optimization using distribution characteristics of reference data for reverse engineering[J]. International Journal for Numerical Methods in Engineering, 2011, 85: 1323-1340.
- [14] 常志朋, 程龙生, 崔立志. 基于马田系统的区间 Choquet 模糊积分多属性决策方法[J]. 控制与决策, 2016, 31(1): 180-186.  
(Chang Z P, Cheng L S, Cui L Z. Interval Choquet fuzzy integral multi-attribute decision making method based on Mahalanobis-Taguchi system[J]. Control and Decision, 2016, 31(1): 180-186.)
- [15] Yang T, Cheng Y T. The use of Mahalanobis-Taguchi system to improve flip-chip bumping height inspection efficiency[J]. Microelectronics Reliability, 2010, 50(3): 407-414.
- [16] Wang N, Saygin C, Sun S D. Impact of Mahalanobis space construction on effectiveness of Mahalanobis-Taguchi system[J]. International Journal of Industrial and Systems Engineering, 2013, 13(2): 233-249.
- [17] Soylemezoglu A, Jagannathan S, Saygin C. Mahalanobis-Taguchi system as a prognostics tool for rolling element bearing failures[J]. Journal of Manufacturing Science and Engineering, 2010, 132(5): 1-12.
- [18] Das P, Datta S. Developing an unsupervised classification algorithm for characterization of steel properties[J]. International Journal of Mechatronics and Manufacturing Systems, 2012, 29(4): 368-383.
- [19] Liparas D, Laskaris N, Feldt R. Applying the Mahalanobis-Taguchi strategy for software defect diagnosis[J]. Automated Software Engineering, 2012, 19(2): 141-165.
- [20] Liparas D, Laskaris N, Angelis L. Incorporating resting state dynamics in the analysis of encephalographic responses by means of the Mahalanobis-Taguchi strategy[J]. Expert Systems with Applications, 2013, 40(7): 2621-2630.
- [21] 叶芳羽, 单汨源, 韩之俊, 等. 基于马田系统与数据包络分析的工业运行质量评价研究[J]. 管理学报, 2018, 15(5): 767-773.  
(Ye F Y, Shan M Y, Han Z J, et al. The industry operation quality evaluation research based on MTS and data envelopment analysis[J]. Chinese Journal of Management, 2018, 15(5): 767-773.)
- [22] Woodall W H, Koudelik R, Tsui K L, et al. A review and analysis of the Mahalanobis-Taguchi system[J]. Technometrics, 2003, 45(1): 1-15.
- [23] Das P, Datta S, Bhattacharyay B K. Classifying tensile strength of HSLA steel: An investigation through neural networks using Mahalanobis distance[J]. International Journal of Mechatronics and Manufacturing Systems, 2010, 4(6): 600-614.
- [24] Lee Y C, Teng H L. Predicting the financial crisis by Mahalanobis-Taguchi system—Examples of Taiwan's electronic sector[J]. Expert Systems with Applications, 2009, 36(4): 7469-7478.
- [25] Huang C L, Chen Y H, Wan T L J. The Mahalanobis-Taguchi system—Adaptive resonance theory neural network algorithm for dynamic product designs[J]. Journal of Information and Optimization Sciences, 2012, 33(6): 623-635.

- [26] Huang C L, Hsu T S, Liu C M. The Mahalanobis-Taguchi system—Neural network algorithm for data-mining in dynamic environments[J]. *Expert Systems with Applications*, 2009, 36(3): 5475-5480.
- [27] Mahmoud E B. Modified Mahalanobis-Taguchi system for imbalance data classification[J]. *Computational Intelligence and Neuroscience*, 2017, 2017: 1-15.
- [28] Su C T, Hsiao Y H. An evaluation of the robustness of MTS for imbalanced data[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2007, 19(10): 1321-1332.
- [29] 牛俊磊, 程龙生. 一种基于改进马田系统的不平衡数据分类方法[J]. *管理工程学报*, 2012, 26(2): 85-93.  
(Niu J L, Cheng L S. Development of a methodology for imbalanced data classification using improved Mahalanobis-Taguchi system[J]. *Journal of Industrial Engineering and Engineering Management*, 2012, 26(2): 85-93.)
- [30] Huang J C Y. Reducing solder paste inspection in surface-mount assembly through Mahalanobis-Taguchi analysis[J]. *IEEE Transactions on Electronics Packaging Manufacturing*, 2010, 33(4): 265-274.
- [31] Kumar S, Chow T W S, Pecht M. Approach to fault identification for electronic products using Mahalanobis distance[J]. *IEEE Transactions on Instrumentation and Measurement*, 2010, 59(8): 2055-2064.
- [32] Johnson R A, Wichern D W. Applied multivariate statistical analysis[M]. New Jersey: Prentice Hall, 1992: 210-258.
- [33] Das P, Datta S. A statistical concept in determination of threshold value for future diagnosis in MTS: An alternative to Taguchi's loss function approach[J]. *International Journal for Quality research*, 2010, 4(2): 95-103.
- [34] Rai B K, Chinnam R B, Singh N. Prediction of drill-bit breakage from degradation signals using Mahalanobis-Taguchi system analysis[J]. *International Journal of Industrial and Systems Engineering*, 2008, 3(2): 134-148.
- [35] Abraham B, Asokan Mulayath V. Discussion—A review and analysis of the Mahalanobis-Taguchi system[J]. *Technometrics*, 2003, 45(1): 22-25.
- [36] Jagulum R, Taguchi G, Taguchi S, et al. Discussion—A review and analysis of the Mahalanobis-Taguchi system[J]. *Technometrics*, 2003, 45(1): 16-21.
- [37] Foster C R, Jugulum R, Frey D D. Evaluating an adaptive one-factor-at-a-time search procedure within the Mahalanobis-Taguchi system[J]. *International Journal of Industrial and Systems Engineering*, 2009, 4(6): 600-614.
- [38] Ramlie F, Jamaludin K, Dolah R, et al. Optimal feature selection of Taguchi character recognition in the Mahalanobis-Taguchi system using bees algorithm[J]. *Global Journal of Pure and Applied Mathematics*, 2016, 12(3): 2651-2671.
- [39] Pal A, Maiti J. Development of a hybrid methodology for dimensionality reduction in Mahalanobis-Taguchi system using Mahalanobis distance and binary particle swarm optimization[J]. *Expert Systems with Applications*, 2010, 37(2): 1286-1293.
- [40] Edgar R, Luis A, Georgina S. Binary ant colony optimization applied to variable screening in the Mahalanobis-Taguchi system[J]. *Expert Systems with Applications*, 2013, 40(2): 634-637.
- [41] Edgar R. Mahalanobis-Taguchi system applied to variable selection in automotive pedals components using Gompertz binary particle swarm optimization[J]. *Expert Systems with Applications*, 2013, 40(7): 2361-2365.
- [42] Yadira R C, Cecilia M G, Edgar R F. Optimal variable screening in automobile motor-head machining process using metaheuristic approaches in the Mahalanobis-Taguchi system[J]. *The International Journal of Advanced Manufacturing Technology*, 2018, 95: 3589-3597.
- [43] 牛俊磊, 程龙生. 基于全方位优化算法的改进马田系统分类方法[J]. *系统工程理论与实践*, 2012, 32(6): 1324-1336.  
(Niu J L, Cheng L S. Classification using improved Mahalanobis-Taguchi system based on omni-optimizer[J]. *Systems Engineering—Theory & Practice*, 2012, 32(6): 1324-1336.)
- [44] Iquebal A S, Pal A, Ceglarek D, et al. Enhancement of Mahalanobis-Taguchi system via rough sets based feature selection[J]. *Expert Systems with Applications*, 2014, 41(17): 8003-8015.
- [45] Liparas D, Pantraki E. An evolutionary improvement of the Mahalanobis-Taguchi strategy and its application to intrusion detection[C]. *International Conference on Advanced Information Systems Engineering*. Greece: Thessaloniki, 2014: 16-28.
- [46] Lv Y Q, Gao J M, Gao Z Y, et al. Multifractal information fusion based condition diagnosis for process complex systems[J]. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, 2012, 227(3): 178-184.
- [47] Hu J Q, Zhang L B, Liang W. Dynamic degradation observer for bearing fault by MTS-SOM system[J]. *Mechanical Systems and Signal Processing*, 2013, 36(2): 385-400.
- [48] Wang Z P, Liu C. Fault diagnosis and health assessment for bearings using the Mahalanobis-Taguchi system based on EMD-SVD[J]. *Transactions of the Institute of Measurement and Control*, 2013, 35(6): 798-807.
- [49] 陈俊洵, 程龙生, 余慧, 等. 基于EMD-SVD与马田系统的复杂系统健康状态评估[J]. *系统工程与电子技术*, 2017, 39(7): 1542-1548.  
(Chen J X, Cheng L S, Yu H, et al. Health status assessment for complex systems based on EMD-SVD and Mahalanobis-Taguchi system[J]. *Systems Engineering*

- and Electronics, 2017, 39(7): 1542-1548.)
- [50] 陈俊洵, 程龙生, 胡绍林, 等. 基于EMD的改进马田系统的滚动轴承故障诊断[J]. 振动与冲击, 2017, 36(5): 151-156.  
(Chen J X, Cheng L S, Hu S L, et al. Fault diagnosis of rolling bearings using modified Mahalanobis-Taguchi system based on EMD[J]. Journal of Vibration and Shock, 2017, 36(5): 151-156.)
- [51] Chen J X, Cheng L S. Rolling bearing fault diagnosis and health assessment using EEMD and the adjustment Mahalanobis-Taguchi system[J]. International Journal of Systems Science, 2018, 49(1): 147-159.
- [52] Lu C, Cheng Y, Wang Z, et al. An approach for tool health assessment using the Mahalanobis-Taguchi system based on WPT-AR[J]. Journal of Vibroengineering, 2014, 16(3): 1424-1433.
- [53] Zhao S, Huang Y X, Wang H R. A modified Mahalanobis-Taguchi system analysis for monitoring of ball screw health assessment[C]. IEEE International Conference on Prognostics and Health Management. Ottawa: IEEE, 2016: 1-7.
- [54] Ohkubo M, Nagata Y. Anomaly detection in high-dimensional data with the Mahalanobis-Taguchi system[J]. Total Quality Management & Business Excellence, 2018, 29(10): 1213-1227.
- [55] Riho T, Suzuki A, Oro J, et al. The yield enhancement methodology for invisible defects using the MTS+ method[J]. IEEE Transactions on Semiconductor Manufacturing, 2005, 18(4): 561-568.
- [56] Hsiao Y H, Su C T. Multiclass MTS for saxophone timbre quality inspection using waveform-shape-based features[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2009, 39(3): 690-704.
- [57] Das P, Datta S. Exploring the effects of chemical composition in hot rolled steel product using Mahalanobis distance scale under Mahalanobis-Taguchi system[J]. Computational Materials Science, 2007, 38(4): 671-677.
- [58] Shakya P, Kulkarni M S, Darpe A K. Bearing diagnosis based on Mahalanobis-Taguchi-Gram-Schmidt method[J]. Journal of Sound and Vibration, 2015, 337: 342-362.
- [59] Nagao M, Yamamoto M, Suzuki K, et al. A face identification system based on the Mahalanobis-Taguchi system[J]. International Transactions in Operational Research, 2001, 8(1): 31-45.
- [60] Cudney E A, Paryani K, Ragsdell K M. Applying the Mahalanobis-Taguchi system to vehicle handling[J]. Concurrent Engineering, 2006, 14(4): 343-354.
- [61] Mahalakshmi P, Ganesan K. Mahalanobis-Taguchi system based criteria selection for shrimp aquaculture development[J]. Computers and Electronics in Agriculture, 2009, 65(2): 192-197.
- [62] Su C T, Wang P C, Chen Y C, et al. Data mining techniques for assisting the diagnosis of pressure ulcer development in surgical patients[J]. Journal of Medical Systems, 2012, 36(4): 2387-2399.
- [63] Su C T, Chen K H, Chen L F, et al. Prediagnosis of obstructive sleep apnea via multiclass MTS[J]. Computational and Mathematical Methods in Medicine, 2012, 2012: 1-8.
- [64] Ketkar M, Vaidya O S. Evaluating and ranking candidates for MBA program: Mahalanobis-Taguchi system approach[J]. Procedia Economics and Finance, 2014, 11: 654-664.
- [65] 常志朋, 程龙生. 基于马田系统和 $\phi_s$ 转换的模糊积分多属性决策方法[J]. 系统工程与电子技术, 2013, 35(8): 1702-1710.  
(Chang Z P, Cheng L S. Fuzzy integral multi-attribute decision making method based on Mahalanobis-Taguchi system and  $\phi_s$  transformation[J]. Systems Engineering and Electronics, 2013, 35(8): 1702-1710.)
- [66] 常志朋, 程龙生. 基于施密特正交马田系统和 $\phi_s$ 转换的灰模糊积分关联度决策模型[J]. 控制与决策, 2014, 29(7): 1257-1261.  
(Chang Z P, Cheng L S. Grey fuzzy integral correlation degree decision model based on Mahalanobis-Taguchi Gram-Schmidt and  $\phi_s$  transformation[J]. Control and Decision, 2014, 29(7): 1257-1261.)
- [67] 常志朋, 程龙生. 灰模糊积分关联度决策模型[J]. 中国管理科学, 2015, 23(11): 105-111.  
(Chang Z P, Cheng L S. Grey fuzzy integral correlation degree decision model[J]. Chinese Journal of Management Science, 2015, 23(11): 105-111.)
- [68] 常志朋, 程龙生, 刘家树. 基于马田系统的2可加Choquet积分多属性决策方法[J]. 管理工程学报, 2016, 30(1): 133-139.  
(Chang Z P, Cheng L S, Liu J S. Multi-attribute decision making method based on Mahalanobis-Taguchi system and 2-additive Choquet integral[J]. Journal of Industrial Engineering and Engineering Management, 2016, 30(1): 133-139.)
- [69] 郑称德. MTS多类判别研究[J]. 管理工程学报, 2003, 17(4): 106-109.  
(Zheng C D. The study of Mahalanobis-Taguchi system for multi-class recognition[J]. Journal of Industrial Engineering and Engineering Management, 2003, 17(4): 106-109.)

#### 作者简介

常志朋(1978—), 男, 教授, 博士, 从事模式识别、多属性决策、公共管理等研究, E-mail: changzhp@126.com.

(责任编辑: 李君玲)