食品、农产品检测中的数据处理和分析方法

赵杰文 林 颢 编著

科学出版社

食品、农产品检测中的数据处理和分析方法

赵杰文　林　颢　编著

科学出版社

北京

内 容 简 介

本书是一部系统介绍和归纳食品、农产品无损检测数据处理和分析的科研论著，基于对所获数据进行处理和分析，从而建立有效的识别模型. 各章节介绍了各种数据处理和分析方法的基本原理，并通过大量的实例阐述这些方法如何应用于食品、农产品的品质检测和分析. 本书精华部分大多来自作者在食品、农产品检测方面多年研究成果的积累，并结合国内外食品、农产品无损检测中数据处理和分析的最新方法，为相关领域科研人员接触数据处理和分析方法的最新动态提供帮助，具有鲜明的特征和实用性.

本书可供从事食品工程、食品分析、农业工程方面的教学及科研工作者参考.

图书在版编目(CIP)数据

食品、农产品检测中的数据处理和分析方法/赵杰文，林颢编著. —北京：科学出版社，2012

ISBN 978-7-03-034274-4

Ⅰ. ①食… Ⅱ. ①赵… ②林… Ⅲ. ①食品检验-数据处理-分析方法 ②农产品-检验-数据处理-分析方法 Ⅳ. ①TS207.3 ②S37

中国版本图书馆 CIP 数据核字(2012) 第 091910 号

责任编辑：伍宏发 曾佳佳 黄 海 / 责任校对：张凤琴

责任印制：赵德静 / 封面设计：许 瑞

科学出版社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2012 年 5 月第 一 版 开本：B5(720 × 1000)

2012 年 5 月第一次印刷 印张：15 1/4

字数：300 000

定价：48.00 元

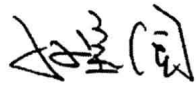
(如有印装质量问题，我社负责调换)

序

无损检测是食品、农产品检测的前沿方向和热门课题, 涉及多个交叉学科. 伴随着各种新的检测技术和方法的不断涌现与应用, 处理和分析大量的、物理意义不同的数据信息是获得精确检测结果和建立高精度识别模型不可或缺的环节. 许多从事食品、农产品品质无损检测技术研究的学者将计算机科学和应用数学的最新研究成果消化、吸收并应用到各自的研究课题, 在这方面取得了一系列可喜的研究成果, 甚至食品、农产品无损检测的数据处理和分析本身也呈现出成为新的研究方向的趋势, 但到目前为止, 国内外还没有相关的论著对其进行系统介绍和归纳.

该书作者赵杰文教授从事食品、农产品无损检测技术研究 20 余年, 受国内 20 余所高校邀请做过专题学术报告, 研究成果曾获 2008 年度国家技术发明二等奖, 在实验数据处理分析及建模方面也积累了大量的心得. 赵杰文教授有一个优秀的学术团队, 成员年富力强, 均有博士学位和出国深造经历, 了解国内外的最新研究进展, 具备渊博的学识. 以上这些为该书的成功编著提供了强有力的保障.

该书围绕如何对所获取的数据进行处理和分析, 从而建立有效的识别模型展开讨论, 分为数据前处理、变量筛选、特征提取、定性识别模型和定量分析模型等几部分, 是作者在食品、农产品检测方面多年研究成果的积累, 同时也反映了计算机科学和应用数学的最新研究成果在食品、农产品检测方面的应用动态. 该书是国内外第一本系统介绍和归纳关于食品、农产品无损检测数据处理和分析的科研论著. 该书的出版将对食品、农产品品质检测技术的发展起到促进作用, 同时也将为广大从事该领域研究的科学工作者、技术人员和研究生们提供一本内容全面、反映数据处理和分析方法最新动态的技术参考书.



中国工程院院士

2011 年 12 月

前　言

食品、农产品质量及安全检测技术近年来发展得很快, 其中无损检测技术异军突起. 该技术在检测外观形状、表面色泽、风味口感、物理缺陷及病虫害造成的伤害等方面已表现出它的优势, 研究成果涉及方方面面.

食品、农产品无损检测技术是通过给予检测对象一定形式的输入能量, 利用对象本身的光、声、电和力学等特性得到相应的响应信号, 对响应信号中包含的大量反映被测对象品质特性的信息进行分析判断的技术. 由于检测在非破坏条件下进行, 且具有速度快、操作方便和能实现在线检测的优点, 无损检测技术已成为食品、农产品检测的前沿和热门研究领域.

处理和分析大量的、物理意义不同的数据信息是获得精确检测结果和建立高精度识别模型不可或缺的环节, 所以应用数学的最新成果往往很快就在研究过程中得到应用; 对响应信号 (包括检测数据) 的处理、分析、建模等本身也自然成为研究的热点, 成了食品、农产品无损检测技术研究领域的一个重要分支. 但是, 到目前为止还没有一本相关的论著对其进行系统的介绍和归纳.

作者于 20 世纪 80 年代末开始开展食品、农产品无损检测技术的研究, 属于我国从事该领域研究的先行者之一, 并且 20 余年来主持了我国现代农业技术领域中有关无损检测技术的第一个 “863” 专项, 以及 5 项有关食品、农产品无损检测技术的国家自然科学基金项目, 其成果荣获 2008 年度国家技术发明二等奖. 作者还在实验数据处理、分析及建模方面, 发表了多篇相关论文, 并被 SCI 收录, 展示了国内外应用数学最新成果在我国食品、农产品无损检测技术研究领域的应用水平. 20 余年来的探索和实践, 加上对大量文献的阅读领会, 作者产生了编著一本系统介绍食品、农产品无损检测中的数据处理和分析方法专著的想法, 在逐步积累资料的同时, 书的框架和体系也逐渐清晰起来.

本书由绪论、数据前处理、变量筛选、特征提取、定性识别、定量分析、常用软件简介七章组成, 涵盖了检测数据 (信号) 的采集、处理、建模的整个数学过程, 书中介绍的应用实例大多来自作者与其团队长期的研究成果. 值得指出的是, 应用数学的最新成果能在食品、农产品品质检测技术的数据分析、处理中得到很好的应用, 在其他相关行业同样也可以得到应用. 故本书既为从事食品、农产品检测的教学、科研工作者和广大研究生们提供了一本具有鲜明特征和实用性的参考书, 同时也可为从事其他相关行业的科研、生产、管理人员提供借鉴.

食品、农产品无损检测技术涉及计算机工程、信息工程、机械工程、农业工程、

食品工程以及数学、力学等多个学科领域. 囿于理论基础和实践经验, 作者在编著这本书的过程中, 感受到了较大压力, 特别是明显感到数学功底不够扎实, 故常有“心有余而力不足” 之感, 疏漏和错误在所难免, 衷心希望同行和读者不吝指正.

编著者

2011 年 12 月

目　　录

第一章　绪　　论

第一节　食品、农产品品质无损检测技术及其特点

食品、农产品的品质检测是指运用数学、物理、化学、生物等学科的基本理论及各种科学技术，对检测对象包括生产原料、辅助材料、半成品、成品、副产品等的状态和主要成分含量及微生物状况进行分析检测. 在对食品、农产品品质进行检测时，因检测目的不同，且检测对象的性质和状态差异较大，所选择的检测方法也各不相同.

无损检测又称为非破坏检测，是近年发展起来的一种新技术，是指在不破坏样品的情况下对其进行品质评价的方法. 食品、农产品无损检测技术涉及光学、力学、电学、磁学等学科，范围广泛，其基础更是涉及材料科学、计算机技术、生物技术、信息技术等诸多领域，其中以光学检测发展最快.

近年来，随着科学技术的发展，无损检测技术也得到长足发展，目前已呈现出两个重要的发展趋势. 其中一个重要发展趋势即无损检测技术与传感器技术、纳米技术及计算机技术结合得越来越紧密，使无损检测分析仪器不但具有越来越强大的“智能”，而且正沿着落地式—台式—移动式—便携式—手持式—芯片实验室的方向发展，越来越小型化、微型化、智能化；检测分析仪器和专用计算机的界限在今后也将变得越来越模糊，许多检测分析仪器实际上是具有某种检测分析功能的计算机. 目前，食品、农产品检测系统多数仍处于智能化的低级阶段，系统只能把计算机技术与传统的食品、农产品检测分析结合起来，仅能适应被测参数的变化、自动补偿、自动选择量程等. 因此，将无损检测技术与传感器技术、纳米技术及计算机技术等多学科交叉结合，开发出能识别与解释各种光学、力学谱图的食品、农产品“智能化”检测系统，成为当前食品、农产品检测智能化研究的热点.

无损检测技术发展的另一个重要趋势，是数据处理与分析方法在检测中显示出越来越重要的作用. 科学技术的发展对现代食品、农产品无损检测技术提出了更高的要求，人们不仅要求及时、精确、可靠地获得有关待测样品品质的数据，而且要求全面快速地分析. 无损检测技术不仅要解决有关测量数据的获取问题，更需要解决从大量数据中提取有用信息的问题并建立相应的模型. 尤其是近年来在食品、农产品无损检测技术中所采用仪器的精密度越来越高，所获取的数据量也越来越大. 在一些实验中，一个样品测试一次即可获取几千甚至数十万个数据，对数据信号的

前处理、数据的精简、数据变量的筛选、特征信息的提取以及识别模型的建立成为无损检测技术研究的热门课题, 模糊数学、统计学、信号处理、化学计量学及模式识别方法等数学方法也越来越多地应用于无损检测数据处理中.

第二节　食品、农产品品质无损检测中的数据处理与分析

随着仪器精密度不断提高, 食品、农产品无损检测仪器所获得的数据量不断增大, 如何运用恰当的数据处理方法, 在庞大的数据量中挖掘出能准确描述检测对象的有用信息, 建立鲁棒性强的数学模型, 已成为无损检测研究的热门课题. 在数据的处理与分析中, 数据的前处理、变量筛选、特征提取、定性识别模型和定量分析模型的建立是其重要组成部分.

一、数据前处理

在食品、农产品品质无损检测中, 检测器所获取的数据信号除含样品待测成分信息外, 还包括各种仪器噪声, 如高频随机噪声、基线漂移、杂散信号、样品背景等. 因此, 在数据分析前, 首先应针对特定的信号测量和样品体系进行合理的处理, 减弱甚至消除各种非目标因素对检测信号信息的影响, 为建立稳定、可靠的数学模型奠定基础. 常用的数据前处理方法有数据标准化处理 (均值中心化、归一化、标准正态变量变换等)、高频噪声滤除 (卷积平滑、傅里叶变换、小波变换等)、信号的微分求导和基线校正等. 数据标准化处理是将原始数据矩阵中各元素减去该列元素均值后, 再除以所在列元素的方差. 其特点是数据矩阵的一列元素权重相同, 均值都为 0, 方差和标准化都为 1. 卷积平滑法是基于最小二乘法原理, 保留分析信号中的有用信息, 消除随机噪声, 但是过度的平滑将会造成检测信号中部分有用信息的丢失. 基线校正主要是扣除仪器背景或漂移对信号的影响, 可以采取偏置扣减、微分求导处理和基线倾斜等方法. 采用微分求导可以较好地净化谱图信息, 在降低噪声的同时也可放大检测信号, 但需注意的是, 微分求导窗口数据点的大小对结果有一定影响, 因此在微分求导的时候需对窗口大小做出合理的选择.

二、变量筛选

变量筛选 (也称为变量选择或特征选择) 是指从原始变量中挑选出一些有代表性的特征变量, 代替原始变量进行数据分析和处理. 在食品、农产品无损检测实验中, 检测仪器每次可获取大量的数据, 但对应着不同原始变量的数据对待测样品品质信息的贡献率不尽相同, 有些变量反映的信息量较为丰富, 有些变量反映的信息量较少, 甚至与待测样品基本无关. 如果将检测器所获取的数据都用于建模, 则建模计算过程将极为烦琐, 计算量很大, 建立的模型也较为复杂, 鲁棒性差. 研究表

明, 通过特定的变量筛选方法对自变量进行优选, 不仅可简化模型, 更重要的是可剔除不相关或非线性变量, 得到预测能力强、稳健性好的训练模型. 常用的变量筛选方法有穷尽搜索法、区间 (如前向区间、后向区间、联合区间) 筛选法和以某种算法 (如模拟退火、遗传算法、无变量信息消除) 为指导思想的随机性质搜索方法. 穷尽搜索法从原始数据集合中将所有可能组合都搜索一遍, 这种方法一定能得到一个最优子集, 但这一般很少用, 因为由它带来的计算量可能让人难以承受. 区间筛选法即将原始数据的变量分为若干个区间, 对每个区间或某几个区间的变量建立相应的模型, 选取最优区间. 相比穷尽搜索, 区间筛选大大减少了工作量, 但同样会存在所选取区间的变量间对待测样品信息贡献率不同的问题. 以模拟退火算法为代表的随机性质搜索方法可选择与检测对象品质相关的信息, 但该方法在搜索特征变量时搜索的范围太广, 带有一定的盲目性, 易陷入局部最优. 因此, 将区间筛选法和随机搜索方法结合, 可减少搜索的盲目性, 提高变量筛选效率.

三、特征提取

特征提取是指通过映射 (或变换) 的方法对原始数据进行重组, 以期用较少的特征值描述原始数据中所包含的大部分信息. 由于对所处理的信息了解不深刻, 且还有许多因素之间的关系及相关程度亦不能肯定, 人们往往先根据化学 (或物理) 的选择标准, 尽可能地把一切相关、又容易获取的特征变量都提取出来, 然后借助于数学方法, 筛选出对模型的建立起较大作用的特征变量. 在实际计算中, 一些不相关的特征变量会降低模型的鲁棒性, 因此研究人员总是力图抛弃那些对建模作用不大的特征变量, 在保证模型精度的前提下, 使特征变量数减到最少. 常用的特征提取方法有直接从原始数据中提取一些特征参数 (如均值、标准差、最大值、极差等) 法, 从原始数据中提取的特征参数比较直观、简便, 但所反映的信息较为粗糙; 还有以某种算法 (如主成分分析法、独立分量分析法等) 为依据, 从原始数据中提取一些特征变量. 主成分分析是把多个指标化为几个综合指标的一种统计方法, 它沿着协方差最大方向由多维数据空间向低维数据空间投影, 各主成分向量之间相互正交. 通过选择合理的主成分既可以达到降维的目的, 又不会过多地丢失原始数据信息, 同时可以减少原始数据中的冗余信息. 主成分分析可保证分解出的分量互相正交; 独立分量分析利用信号的高阶统计量, 要求分解出的各分量尽可能独立, 在信号的特征提取中则表现出更大的优势.

四、定性识别

定性识别 (也称为模式识别) 是指对表征事物或现象的各种形式 (如数值、文字和逻辑关系等) 的信息进行处理和分析, 以便对事物或现象进行描述、辨认、分类和解释的过程. 按照识别时是否需要具有识别样本的先验知识, 模式识别方法可

分为 “监督学习分类” 和 “非监督学习分类”. 在模式识别的特征空间里, 如果所分类的情况是已知的, 在此基础上, 可以选择一个合适的距离尺度, 以得到有关这些类的分布形状以及典型模式的信息, 这种方法称为 “监督学习分类”. 常见的模式识别方法多为 “监督学习分类” 方法, 如线性判别分析、人工神经网络、偏最小二乘分类、支持向量机等. 在模式识别中, 也会遇到不能事先获取任何关于样本的先验知识 (很多时候需要在无监督情形下将很多东西分类) 的情况. 因此, 分类系统必须先通过一种有效的方法去发现样本的内在相似性, 然后指导同类检测对象的分类, 这种方法称为 “非监督学习分类”. 最常见的非监督学习分类法为聚类分析法, 包括 “树聚类” 和 “K 均值聚类” 等. 此外, 按照使用的分类函数, 模式识别方法可分为线性和非线性判别分析方法. 线性判别分析方法有欧氏距离、马氏距离、费歇尔投影法和 K 最近邻法等; 非线性判别分析方法有人工神经网络、支持向量机和支持向量数据描述等.

五、 定量分析

定量分析是指分析一个被研究对象所包含成分的数量关系或所具备性质间的数量关系; 也可以对几个对象的某些性质、特征、相互关系从数量上进行分析比较, 研究的结果也用 “数量” 加以描述. 定量分析是依据统计数据, 建立数学模型, 并用数学模型计算出分析对象的各项指标及其数值的一种方法. 相比定性识别而言, 定量分析更加科学, 需运用到更多的数学计算; 定性识别虽然较为粗糙, 但在数据资料不够充分或分析者数学基础较为薄弱时比较适用. 这两种分析方法对数学知识的要求虽然有高有低, 但并不能就此把定性识别与定量分析划分开来. 事实上, 现代定性识别方法同样要采用数学工具进行计算, 而定量分析则必须建立在定性预测基础上, 二者相辅相成, 定性是定量的依据, 定量是定性的具体化, 二者结合起来灵活运用才能取得最佳效果. 线性回归法 (包括一元线性回归和多元线性回归) 是最早采用的定量分析法. 由于食品、农产品无损检测所获取的数据量比较大, 主成分回归及在主成分分析基础上进行的偏最小二乘分析法已越来越多地运用在食品、农产品品质无损检测中; 随着非线性方法研究的不断拓展, 一些传统的定性识别方法经过改造, 也逐步应用到定量分析中, 如人工神经网络回归和支持向量回归法.

第三节　数据处理和分析在食品、农产品无损检测中的应用趋势

一、 多学科知识交叉

在食品、农产品无损检测中, 数据处理和分析已凸显出越来越重要的作用, 其

内涵和外延也不断扩展, 各门学科中与数据处理相关的最新研究成果经消化、吸收、更新, 不断地被应用于食品、农产品无损检测的数据处理和分析中. 在数学领域, 应用数学学科的最新研究成果 (如小波分析、随机过程相关科研成果) 正越来越快地被应用于数据信号的处理和随机信号的分析中; 在通信领域, 卡尔曼滤波、最小均方自适应滤波、高阶谱分析等方面的最新研究成果也在最短时间内被应用于数据信号的平滑、滤波去噪等信号前处理和分析中; 在化学计量学领域, 遗传算法和模拟退火算法等方面的最新研究成果也不断被应用于数据矩阵特征提取和变量筛选中; 在模式识别领域, 各种模式识别方法的最新成果也被应用于数据信号的定性识别中, 各种软件也越来越多地被用到食品、农产品无损检测的数据处理和分析中去. 各个学科间的相互交叉是现代科学技术发展的重要趋势, 各个学科之间的界限越来越模糊, 在食品、农产品无损检测数据处理和分析中表现得尤为明显. 计算机、通信、化学计量学等领域的数据处理和分析都是以数学为基础, 有一定的交叉但又明显带有各自学科的特色, 这些学科的知识又可以相互交叉应用于无损检测中的数据处理与分析, 如一批实验样品的检测数据, 其数据前处理部分采用的是通信领域的信号处理, 特征变量提取与筛选则应用到化学计量学的知识, 而数据的识别则需采用模式识别方法. 随着科技的发展, 各个学科之间的交叉将会越来越明显, 且各个学科中的最新研究成果也将会越来越快地被应用到食品、农产品无损检测数据处理与分析中.

二、计算机和数据处理软件作用凸显

在食品、农产品无损检测中, 实验所获取的数据量越来越大, 一批实验结果获取到几十或几百万个数据量的现象十分常见, 且随着仪器检测精度、速度进一步提高, 所获取的数据量将会更庞大. 检测仪器获取的大量 (甚至可谓海量) 数据必将伴随着极为烦琐的数据处理过程, 纯粹依赖人工不可能完成如此庞大的数据计算量, 借助计算机来处理数据已经成为必然趋势, 且由于处理数据量过大, 一些简易的计算软件也难以处理如此庞大的数据. 因此, 现代数据处理和分析对计算机提出了更高的要求, 即选择快速、简便、适宜的软件处理数据以缩短计算过程和时间, 并保证计算结果的准确性. 目前, 可用于数据处理和分析的通用软件有 Excel、SPSS(statistical product and service solutions)、SAS(statistical analysis system)、MATLAB(matrix laboratory)、VB(visual basic)、VC++(microsoft visual C++) 等, 还有一些专用型的实验数据优化、图形可视化和数值分析软件, 如 Design-expert、Origin、Probit 等. 其中, 最常用的是 Excel、SPSS、SAS 和 MATLAB. Excel 以表格的方式输入与管理数据, 能进行简单的数据处理和分析, 如数据排序、回归分析等, 操作比较简单、易学; SPSS 是 IBM 公司开发的最早采用图形菜单驱动界面的统计软件, 采用类似 Excel 表格的方式输入与管理数据, 数据接口较为通用, 能方便地从其他数据

库中读入数据, 其统计过程包括常用的、较为成熟的统计过程, 可以满足非统计专业人士的工作需要; SAS 是由美国北卡罗来纳州州立大学开发的统计软件, 功能及操作方法与 SPSS 基本一致; MATLAB 是矩阵实验室 (matrix laboratory) 的简称, 是美国 MathWorks 公司出品的商业数学软件, 用于算法开发、数据可视化、数据分析以及数值计算的高级技术计算语言和交互式环境. MATLAB 可以进行矩阵运算、绘制函数和数据、实现算法、创建用户界面、连接其他编程语言的程序. 几乎所有食品、农产品无损检测中数据处理与分析, 都可以在 MATLAB 中实现. 随着科学技术的发展, 各种软件版本也不断更新, 目前 SPSS 和 SAS 软件已更新了十多个版本, 而 MATLAB 软件更新了二十多个版本. 在食品、农产品数据处理量不断增大且精度要求越来越高的趋势下, 各种软件将不断优化更新, 而另外一些专用型软件也将不断推出. 此外, 对计算机硬件也将提出更高的要求, 或许不远的将来, 处理食品、农产品无损检测数据的专用型计算机也会诞生.

主要参考文献

陈斌, 黄星奕. 2004. 食品与农产品无损检测新技术. 北京: 化学工业出版社.

林颢. 2010. 基于敲击振动、机器视觉和近红外光谱的禽蛋品质无损检测研究. 镇江: 江苏大学食品与生物工程学院.

许禄, 邵学广. 2004. 化学计量学方法. 北京: 科学出版社.

应义斌, 韩东海. 2005. 农产品无损检测技术. 北京: 化学工业出版社.

赵杰文, 孙永海. 2008. 现代食品检测技术. 2 版. 北京: 中国轻工业出版社.

Berrueta L A, Alonosa-Salces R M, Heberger K. 2007. Supervised pattern recognition in food analysis. Journal Chromatography A, 1158: 196–214.

Foody G M, Mathur A, Sanchez-Hernandez C, et al. 2006. Training set size requirements for the classification of a specific class. Remote Sensing Environment, 104: 1–14.

Hermmateenejad B, Miri B, Akhond M, et al. 2002. QSAR study of the calcium channel antagonist activity of some recently synthesized dihydropyridine derivatives. An application of genetic algorithm for variable selection in MLR and PLS methods. Chemometrics and Intelligent Laboratory Systems, 64: 91–99.

Zou X B, Zhao J W, Huang X Y, et al. 2007. Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of 'Fuji' apple based on different PLS models. Chemometrics and Intelligent Laboratory Systems, 87: 43–51.

Zou X B, Zhao J W, Malcolm, et al. 2010. Variables selection methods in near-infrared spectroscopy. Analytica Chimica Acta, 667: 14–32.

第二章　数据前处理

在食品、农产品的品质检测时, 获取的信息中除含有待测样品的原始信息外, 还包含各种外在的干扰信息, 这些噪声信息会导致测得的数值和真实值之间存在一定差异. 为尽可能消除误差, 应保持实验时的环境因素尽量一致. 除此之外, 必须运用各种数据处理方法来减弱甚至除去各种干扰因素的影响, 为下一步的数据处理奠定基础. 常用的数据处理方法有数据归一化处理、数据平滑和去噪处理等, 下面将对这些方法进行逐一介绍.

第一节　标准化处理

一、均值中心化

均值中心化 (mean centering, MC) 处理的思想是将每个传感器所获取数据矩阵中各项平均能量去除掉, 对于第 j 个样本的数据值 $GP(j,i)(i=1,2,\cdots,n)$, 中心化处理公式如下:

$$GP'(j,i)=GP(j,i)-\frac{1}{n}\sum_{k=1}^{n}GP(j,k),\quad i=1,2,\cdots,n \tag{2-1}$$

其中, n 为数据的变量总数. 中心化处理将每个数据矩阵减去平均值, 可以简化后续数据处理部分的计算并使之稳定.

二、极小/极大归一化

归一化处理 (normalization) 是为了使所有数据都处于一个相同的范围内, 使变量和平均值的分布更加均衡. 最常用的极值归一化 (Min/Max) 是指把数据矩阵的每一行数据与该行最小值的差都除以极差 (即最大值与最小值的差), 得到的新数据范围为 0~1. 在食品、农产品检测中, 传感器的响应信号值之间数据的差异性可能会很大, 有时会相差几个数量级. 因此, 在提取传感器响应信号特征值的基础上, 还得将其归一化, 即使传感器响应特征值处于 [0, 1] 或 [−1, 1], 几种常见的归一化方法见表 2-1. 表中 $P_{ij}(i=1,2,\cdots,8;j=1,2,3,4)$ 为已提取的响应信号特征值, y_{ij} 为归一化后的响应信号特征值, 把 P_{ij} 看做一个矢量, $P_{kij}(k=1,2,\cdots,n;i=1,2,\cdots,8;j=1,2,3,4)$ 就是矢量 P_{ij} 的第 k 个值, $P_{ij}^{\max}$, $P_{ij}^{\min}$, $\overline{P}_{ij}$ 和 σ_{ij} 为该特征值多次测量后统计的最大值、最小值、平均值和方差.

一般归一化的公式计算简单, 可以让所有特征值都映射到 [0, 1] 区间上, 适合于各种情况. 随着检测的进行, 样本特征的平均值和方差都会改变, 对所筛选的特