

doi:10.3969/j.issn.2095-6002.2017.03.001

文章编号:2095-6002(2017)03-0001-15

引用格式:陈峰,李鹤东,王亚棋,等.化学计量学方法在食品分析中的应用[J].食品科学技术学报,2017,35(3):1-15.



CHEN Feng, LI Hedong, WANG Yaqi, et al. Applications of chemometrics in food analysis[J]. Journal of Food Science and Technology, 2017,35(3):1-15.

化学计量学方法在食品分析中的应用

陈峰¹, 李鹤东², 王亚棋¹, 付海燕^{1,2,*}, 郑福平^{1,3}

(1. 克莱姆森大学 食品营养和包装科学系, 美国 克莱姆森 29634; 2. 中南民族大学 药学院, 湖北 武汉 430074; 3. 北京工商大学 北京食品营养与人类健康高精尖中心, 北京 100048)

摘要: 化学计量学是由数学、统计学、化学及计算机科学交叉形成的具有独特魅力的重要新兴学科。作为其核心理论体系的模式识别和多维校正方法,在海量数据信息挖掘与处理、分析信号分辨与解析中,体现了突出的优势,并能攻克传统分析方法难以处理的复杂难题,已广泛应用于食品分析领域。系统介绍了化学计量学核心理论方法及其在食品分析中的应用及研究进展,详细分析了模式识别和多维校正方法的基本原理及优缺点,指出了目前化学计量学方法在食品分析领域研究中需要解决的问题并对其未来发展方向进行了展望。

关键词: 化学计量学; 食品分析; 模式识别; 多维校正

中图分类号: TS201.2

文献标志码: A

食品分析的核心任务是通过反映食品特征品质(如食品营养活性成分、外源的农药兽药残留、非法添加的违禁物质、内源的霉菌毒素等指标,以及食品分级分类、掺伪假冒等)的量测数据性质和信息进行甄别和定性定量分析,进而为评判食品质量好坏和食用安全性提供依据^[1]。食品种类繁多、成分复杂,各种食品安全与质量问题屡有发生,因此,发展快速、高效、简便的食品分析方法至关重要。目前,一些现代分析技术,如色谱-质谱联用技术^[2]、近红外光谱技术^[3]、中红外光谱技术^[4]、原子光谱技术^[5]、电子鼻技术^[6]、电子舌技术^[7]等,均已成功应用在现代食品分析领域。但归根结底,无论需要处理的食品问题怎么变化,各种现代分析仪器或检测技术获得的是食品中具有不同内涵性质的各种类型的信息数据,而化学计量学针对数据信息挖掘与处理,特别是对复杂体系隐含信息的提取,能提供强大的手段,因此,化学计量学必然会在解决食品分析

问题中发挥重要的作用。

化学计量学是20世纪70~80年代兴起的一门交叉学科。它基于数学、统计学、化学及计算机科学的方法和原理,设计优化实验,挖掘与处理、分辨与解析实验量测数据的信息,从而获取复杂分析体系中隐含的有用信息^[8]。化学计量学包括实验设计、信号解析、模式识别、多元分辨和多维校正等理论体系,拥有广阔的应用领域,其中,模式识别和多维校正理论及其应用是化学计量学中最为重要并且最具发展潜力的领域之一。

本文介绍了化学计量学处理的分析对象即数据结构类型,重点阐述用于处理相应数据结构类型的化学计量学中经典的模式识别和多维校正方法的基本原理,评述各种方法间的适用性和优缺点,详细介绍了化学计量学方法在食品分析中的应用,简要分析了化学计量学方法在食品分析领域中还需进一步解决的问题并对其未来的发展进行展望。

收稿日期: 2017-02-02

基金项目: 国家自然科学基金资助项目(21576297);国家重点研发计划项目(2016YFD0400500)。

作者简介: 陈峰,男,教授,博士生导师,主要从事食品风味营养与质量安全控制方面的研究;

*付海燕,女,副教授,主要从事化学计量学及其在复杂体系中创新性应用方面的研究,通信作者。

1 化学计量学方法处理的不同数据结构类型

化学计量学这一学科的创始人之一,美国化学家 Kowalski 教授,基于张量代数理论,对实验过程中所获得的各种不同量测数据进行过简单的分类。随着日益复杂的数据产生,其张量数据结构分类也在不断地得到总结和归纳^[9-11]。图1形象地描述了各种仪器和检测技术所获取的原始张量数据结构类型,并区分了数据的“阶”和“维”的概念,以及模式识别与多维校正所适应于处理的相应张量数据类型。这里针对于单个样品数据,而针对于具有相同结构类型的多个样品在一个维度方向延展组合的数据集为“维”(见图1),故“维”比“阶”高1个维度,而提取不同结构类型数据集中隐含信息的方法需要依靠不同的化学模式识别及多维校正方法。当“零阶仪器”(如单波长光度计、离子选择性电极等)量测一个实验样本只产生一个响应值时,该数据类型为标量,即零阶张量;多个实验样本下该结构类型数据组合的样本数据集为矢量,相对应处理该数据集的化学计量学方法为零阶校正方法,如标准曲线

法;类似地,“一阶仪器”(如紫外-可见分光光度计、荧光分光光度计、红外或近红外光谱仪、拉曼光谱仪以及核磁共振等各类光谱仪、单通道 HPLC、电化学扫描仪或传感器阵列等)量测一个实验样本获得的数据类型是矢量,即一阶张量,多个实验样本下该结构类型数据组合的样本数据集为矩阵,相对应处理该数据集的化学计量学方法为一阶校正方法或模式识别方法;同样地,“二阶仪器”(如各种色谱 HPLC/GC、液相色谱毛细管电泳分离技术(HPLC/CE)与二极管阵列检测器联用仪(HPLC/DAD)、激发-发射荧光仪(EX-EM)等)量测一个实验样本获得的数据类型是矩阵,即二阶张量,多个实验样本下该结构类型数据组合的样本数据集为三维数阵,相对应处理该数据集的化学计量学方法为二阶校正方法。近年新发展起来的“三阶仪器”(如全二维气相色谱或全二维液相色谱与二极管阵列或质谱联用仪(GC-GC/LC-LC-DAD/MS)等)的使用,使量测一个实验样本获得的数据类型变成一个三维响应数阵,即三阶张量,而多个实验样本下该结构类型数据堆叠组合的样本数据集为四维数阵,相应的处理该数据集的化学计量学方法为三阶校正方法。三阶校正甚至更高阶校正方法的理论体系是化学计量学中的难点和盲点,目前仍处于起步阶段。

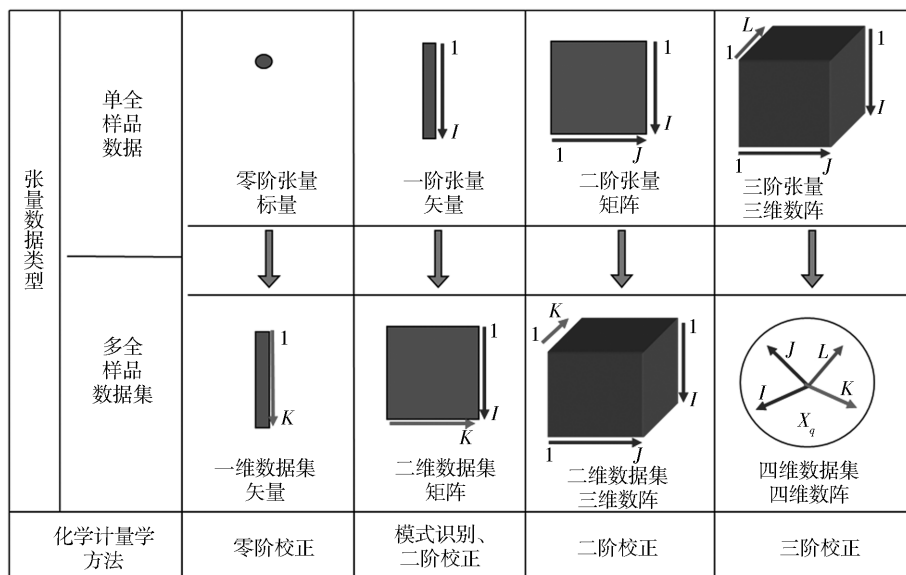


图1 化学计量学方法处理的分析数据结构类型示意

Fig.1 Schematic diagram of data structure processed by chemometric method

1.1 模式识别

模式识别是利用统计学、信号处理、数学算法等工具,根据量测对象的特征性质进行数据挖掘和信

息提取,进而对所研究的物质或现象进行聚类 and 分类判断的方法^[12]。模式识别方法适用于样品数据集为矩阵数据的处理。根据方法基本原理,可分为

多类聚类方法、判别分析方法和一类分类的类模型方法。多类聚类方法是不同类别样品根据不同的计算依据,如样本在空间分布的距离进行相似类聚、相异分离;判别分析方法是根据训练集样品进行学习,将两类或多类样品按照训练的目标函数进行多类样品识别和分类,但当未知的样品属于训练集之外的类别时,判别分析方法将导致错误的判别结果;一类分类的类模型方法则只需训练集中仅包含目标类样品,通过建立目标类的类模型,达到从非目标类样品中的目标类样品识别。图2给出了判别分析方法对两类或多类和类模型对一类分类示意图。图2将化学计量学方法中常用的模式识别方法基本原理和优缺点进行了归纳,为在食品分析应用中选择相应的模式识别方法分析和解决问题提供了理论支持和依据。

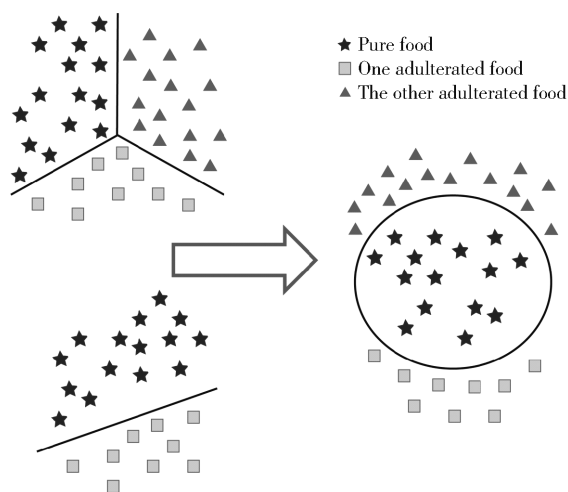


图2 判别分析方法对两类或多类和类模型对一类分类示意

Fig.2 Schematic diagram of two or multi-class by discrimination analysis method and one-class classification by class-modeling techniques

1.1.1 多类聚类和判别分析方法

1) 主成分分析

主成分分析(principle component analysis, PCA),是一种基于将原始变量转换为互不相关的新特征变量之间的线性组合,并使其包含最大样本方差,用较少的指标和变量特征反映和描述样本的固有变异性及原始变量的特征,从而达到简化和降低特征空间维数的建模方法^[13]。PCA方法虽然可通过较少的几个显著主成分,最大化地保留和表征原始变量特征,但因其降维后的新组合信息量描述的是每个样品的矢量空间分布,因而该算法在降维这

一过程中,有时会丢失有助于判别各样品间差异信息的变量特征,导致类内和类间的样品分类出现混乱。

2) K最近邻分类算法

K最近邻(K-nearest neighbor, KNN)分类算法是一种原理最为简单的聚类分类技术^[14]。该算法原理为将样本定义在距离特征空间里,如果一个样本的最接近的K个“邻居”(即相邻样本)中的绝大多数均属于同一个类别,则可决策该样本也属于这个类别,且具有该类别中其他样本的特性。由于KNN方法对于样品所属类别的判定不是通过判别类域的方法进行决策,而是通过依靠该样品周围少量的相邻样本,因而不需要估计参数,不需要训练,适合对交叉重叠、稀有事件进行分类。该方法简单、易于理解和实现,但该算法在分类时的主要缺点是当每类样本数量不平衡时,如一个类的样本数很多,而其他类样本数很少,则可能导致每当输入一个新样本时,该样本的K个邻居样本中,数量多的类的样本将会占多数;另一个缺点是在大训练集中寻找最近邻点时,需要对每一个待分类的样本计算其到全体已知样本的距离,才能求得它的K个最近邻点。此过程所需时间可能很长,容易造成“维数灾”,导致分类结果精度低。

3) 偏最小二乘判别分析

偏最小二乘判别分析(partial least squares discriminant analysis, PLSDA)方法是经典的多类判别分析方法,该方法基于同时分解预测量测矩阵 X 和类别矩阵 Y 提取因子,并将因子按它们间相关性从大到小进行排列,再通过交互验证确定隐变量个数,参与回归建模。该方法将样品集编码为类别矩阵 $Y_{m \times n}$ (类别矩阵中的所有第 i 类样品的第 i 个元素编码均为1,其余元素编码均为0),然后基于类别矩阵的每一列关联预测矩阵 $X_{m \times v}$,进行回归建模,求解回归矩阵 $R_{v \times n}$,见式(1)。

$$Y_{m \times n} = X_{m \times v} R_{v \times n} \quad (1)$$

由求得的回归矩阵 $R_{v \times n}$ 与所需预测的未知样品的量测矩阵再进行回归关联,即可计算得到未知样品的类属矩阵 Y_p ,见式(2)。

$$Y_p = X_p R_{p \times k} \quad (2)$$

最后,再根据已知类别矩阵的编码位置,解码类属矩阵中最大元素所在的位置,若最大元素出现在第 i 个位置,则可将该样品甄别为第 i 类^[15]。

该方法在多类分类判别分析中性能稳健,可以

减少变量间多重共线性产生的影响,但是有时会由于样品的组成和变量性质间缺乏相关关系,使模型受到无用信息变量的干扰,从而导致判别分析精度降低甚至错误。

4) 移动窗口偏最小二乘判别分析法

移动窗口偏最小二乘方法(moving window partial least-squares discriminant analysis, MWPLSDA)是在偏最小二乘判别方法基础上引入变量挑选的策略。该方法的原理是基于虚拟设置一个尺寸大小为包含变量数为 w 的窗口在量测矩阵变量区间连续移动(即移动的窗口内所包含的建模变量数 w 为从窗口起点 j 到终点 $(j+w-1)$ 个变量),基于移动窗口中的变量构建一系列的PLSDA子模型,并计算在不同隐变量数时,这一系列窗口中分类子模型的残差平方和,最后将不同隐变量PLSDA子模型的残差平方和对窗口作图,即可依据残差平方和最小的原则挑选出少而精的量测矩阵变量区间,重新组合构建新的PLSDA模型^[16]。该方法能最有效提取有利于分类的变量特征差异信息,从而消除一些非线性物理扰动或非成分相关等各种因素干扰,既能减小模型复杂度,又能提高模型精度,还能极大提高运算建模速度。但是,该方法不能完全避免和克服变量区间中存在变量信息重叠的问题,从而有时可能对PLSDA模型分类精度改善不明显。

5) 人工神经网络

人工神经网络(artificial neural network, ANN)是通过模拟人脑结构,以对大脑的生理研究成果为基础,并基于数理方法进行简化抽象和模拟的一种智能仿生的自适应运算模型和信息处理系统^[17-20]。该算法作为分类聚类运算时可分为有监督学习和无监督学习两种方式。其中,有监督学习的原理是将训练样本输入到网络,并将网络输出和期望的目标值相比较,获得误差值,据此控制连接权重值连接强度的调整,经过多次训练中不断朝误差减小的方向收敛,自适应计算出一个确定的权重值,进而获得满意精度的网络输出值;无监督学习的原理则不给输入网络训练样本,直接根据已建立的聚类进行自组织的权重值调整,学习规律的变化服从连接权值的演变方程。该算法不依赖于精确的数学模型,能拟合强的非线性关系,具有并行性、容错性、非线性、自组织学习和自适应处理能力等特点,但该算法常易陷入过拟合和局部最优,在某种程度上限制了它的实际使用。

6) 支持向量机

支持向量机(support vector machine, SVM)是Vapnik等^[21]于1995年首先提出的基于统计学理论的一种新的非常有潜力和广阔应用前景的分类模式识别方法^[22-24]。该方法的本质是解决一个二次规划问题,并最终可转化为凸优化问题,且在运算过程中基于统计学习理论的VC理论和采用结构风险最小化原理,能保证算法收敛的全局最优性,从而使得该方法建模时基本不受到训练样本数的影响,即使在其训练样本建模数量较少时,该算法依然表现出良好的泛化能力。但是,该方法通常都需要计算和存储Hessian矩阵,当数据量较大时,训练速度极大地受到训练集规模的影响,求解过程所需计算资源很大,有时会出现计算阻滞,这样使得SVM建模求解时更适合于处理样本数较小的实验体系。此外,该方法性能直接受到模型中所选择的核函数的影响,而对于选择最佳的核函数构建针对特定问题的SVM模型仍是一个难题。因此,作为一种新兴的技术,SVM目前在实际应用中也有一定的局限性。

1.1.2 一类分类的类模型方法

模式识别方法中的多类聚类和判别分析方法通常是对两类或几类、多类样品进行学习和预测。在实际分析中,这些方法难以对层出不穷的掺杂、假冒样品进行完备的采样和学习,因此对两类、多类样品的分类远远不能满足无目标检测的要求。特别是由于非法添加物数不胜数、层出不穷,以及受原产地域保护产品的干扰影响,也不可能对所有其他产地与产品完全采样,很难建立全面的常规检测体系和制度。这种食品安全检验中的缺陷在国内外引发了多次严重的食品安全危机,引起公众对农产品和食品安全性的普遍担忧,因此,发展“无目标”检测方法,是帮助解决此类问题的一个有效补充手段。化学计量学中的一类分类(one-class classification)的类模型方法(class-modeling techniques)^[25-29]对于该类问题的解决提供了强有力的分析手段。该类方法仅对我们感兴趣的一类样品(如地理标志产品、纯正食品)的代表性样品进行充分采样、检测和表征,对测量信号建立类模型,通过对未知样品测量信号的分析,进一步发现假冒、掺杂引起的(很可能是比较细微的、仅凭肉眼难以分辨的)信号差异,最终目的是尽可能全面地检出已知或未知的各种掺杂。传统的一类分类的类模型方法主要有以下几种:1)基于软独立模型类簇法(soft independent modelling

of class analogy, SIMCA)^[30]及其变种;2)基于多元正态分布和 Hotelling 的 TZ 统计量的非均匀分布类(unequal dispersed classes, UNEQ)模型^[31];3)基于各种势函数和概率密度估计的类模型方法^[32];4)基于人工神经网络(artificial neural network, ANN)的类模型方法^[33]。其中,传统的一类分类的类模型方法中最为经典的 SIMCA 算法由于其简便性、有效的数据降维和对误差分布的较少假定,获得了广泛的应用,已被收入大多数化学计量学教科书中。但是, SIMCA 在实际应用中仍然遇到一些困难,主要在于:1)显著主成分数对其性能影响很大,太少的主成分数不能充分描述化学成分的变化,太多的主成分数则往往导致模型的过拟合;2)SIMCA 决策区间的估计也存在一定困难,主要原因是残差的相关性和由此引起的自由度估算不准确;3)SIMCA 的隐变量侧重于解释类内样品的最大方差,并未考虑同类样品的投影应尽可能互相接近而有效避免与类外样品相重叠,造成了隐变量降维与分类性能优化的脱节,以及 SIMCA 模型性能的退化。

在传统一类分类的类模型基础上,近年来提出了一种基于支持向量机(SVMs)的类模型方法,即支持向量数据描述算法(SVDD)^[34]。SVDD 在模型参数规范化的同时,使类内样品的主体到类中心的距离最小化。由于 SVDD 立足于已获得广泛认可的 SVMs,又有比较清晰的几何意义,受到较多关注。但是,SVDD 也存在一些问题^[35]:1)SVDD 中的参数,如规范化参数、高斯核函数的宽度参数在很大程度上依靠经验选取,尚没有客观的估计方法;2)SVDD 容易发生拟合。此外,还提出了以经典的 PLS 回归为基础的一类偏最小二乘(one-class partial least square, OCPLS)方法^[36],OCPLS 隐变量是类内样品向类中心的投影,并考虑了训练集样品特征数据阵的方差解释和类内样品投影的紧密性,并对未知样品在给定显著性水平下的决策区间进行一类分类,该方法目前发展了 MATLAB 的 OCPLS 工具包^[37]。但是,该方法在隐变量的构造和回归计算中,若存在奇异样品,可能会导致模型的严重偏差。

1.2 多维校正

多维校正方法是化学计量学最为核心理论部分,它立足于创新性地构建定量分析的基础理论,发展针对不同结构类型量测数据的相应校正方法,通过构建的多维校正模型从不同结构类型数据的海量信息中最大限度地提取隐含的有用信息进行定量表

征,从而解决各种复杂的实际分析体系中的问题和难题^[38]。将化学计量学方法中常用的多维校正方法基本原理和优缺点进行归纳,为在食品分析应用中选择相应的多维校正方法分析和解决问题提供理论支持和依据。

1.2.1 一阶校正

一阶校正方法用于处理和解析样本集为矩阵结构类型的量测数据,是目前化学计量学理论中最为成熟且应用最广泛的校正方法^[38]。一阶校正方法的代表性算法主要有 P-矩阵法、K-矩阵法、主成分回归(PCR)和偏最小二乘回归(PLS)等。其中 P-矩阵法和 K-矩阵法在求解过程中涉及到矩阵求逆,这就要保证量测矩阵与样品性质响应矩阵均为满秩矩阵这一前提,而实际量测数据变量数通常都会多于样品中的组分数,这就存在矩阵亏秩而导致不确定性的解析结果。针对矩阵亏秩,PCR 是基于量测矩阵的奇异值分解(SVD),然后采用显著的奇异值重构量测矩阵,从而能获得较为稳健的解析结果。PLS 是在 PCR 基础上进一步发展,该方法基于量测矩阵和响应矩阵同时进行 SVD,这种求解方式能克服一些非线性扰动因素影响,从而使得模型精度更高和稳健性更好。但是,一阶校正因其在对矩阵分解时,存在矩阵双线性分解旋转的不确定性,其缺点是如果校正集样品中未对预测集样品中的其他干扰进行校正建模,则会出现依干扰程度的不同产生不同程度的预测偏差和错误。因此,为保证一阶校正给出有物理意义的可靠解,必须要满足校正集样品包含预测集样品的所有响应组分和性质。

1.2.2 二阶校正

二阶校正方法用于处理样本集为三维数阵的结构类型的数据,是目前化学计量学理论中发展潜力巨大的校正方法。二阶校正方法针对一阶校正所出现的不足,利用基于特殊性结构的三维数阵分解具有唯一性的优点,从而可以得到具有明确物理意义的唯一解,即可从未知和未校正的背景中提取目标分析物的信号。因此,二阶校正方法比一阶校正方法具有一个明显的优势,无需要求校正集样品包含预测集样品中的各种未知干扰,即著名的数学分离“二阶优势”^[38]。依据三维数阵分解的不同原理,可将二阶校正方法分为主要的三类:第一类是基于特征值分析的非迭代类算法,典型的代表性算法有秩消失因子分析(rank annihilation factor analysis, RAFA)算法^[39]、广义秩消失因子分析(GRAM)算

法^[40]和直接三线性分解算法(direct trilinear decomposition, DTLD)^[41]。这类算法的优点是对组分估计不敏感,可以实现未知干扰下目标分析物的定量分析,但缺点是因该类算法对不严格服从三线性成分模型或包含较大噪声的三维数阵进行三线性分解时方差较大,故对信噪比低的分析体系建模解析时,可能因复数解而出现严重偏差。第二类是基于展开形式迭代的最小二乘(ALS)拟合类算法,平行因子分析(parallel factor analysis, PARAFAC)是该类算法中的经典代表^[42]。基于严格意义上的最小二乘原理的PARAFAC优点是对噪声影响不敏感,能对三线性成分模型进行最优拟合,模型偏差较小。缺点是该算法对初始值和组分敏感,易陷入局部最优,且通过展开形式计算,收敛速度较慢。第三类是基于切片形式的矩阵双广义逆运算方式类算法,其经典代表有交替三线性分解(alternating trilinear decomposition, ATLD)算法^[43],该方法基于对切片矩阵对角元素和广义逆求解结合的方式,充分结合了PARAFAC和DTLD的优点,兼具对组分不敏感和收敛速度快的双重优点。在该算法基础上,研究者针对不同三维数阵数据特点,发展了一系列性能优良的基于切片形式的ATLD变种算法^[44],例如:自加权交替三线性分解(self-weighted alternating trilinear decomposition, SWATLD)算法^[45]、交替惩罚三线性分解(alternating penalty trilinear decomposition, APTLD)算法^[46]等。

1.2.3 三阶校正及高阶校正

三阶校正及高阶校正用于处理样本集为四维数阵甚至更高维数阵的结构类型的数据,是目前化学计量学多维校正领域中的前沿难点理论。三阶校正方法是通过引入额外一维的维度优势,克服二阶校正方法在解决复杂体系中存在高共线或严重基体效应时会出现较大偏差甚至错误的问题^[47]。目前,已报道的方法有基于数阵展开形式的四维平行因子分析(four-way parallel factor analysis, Four-way PARAFAC)法^[48]、基于数阵切片形式的交替惩罚四线性分解(alternating penalty quadrilinear decomposition, APQLD)法^[49]、归一化自加权交替四线性分解(regularized self-weighted alternating quadrilinear decomposition, RSWAQLD)法^[50]及基于数阵不完全展开和切片形式相结合的交替加权残差约束四线性分解(alternating weighted residual constraint quadrilinear decomposition, WRCQLD)算法^[51]等。三阶校正

方法在充分发挥“二阶优势”的基础上,还具有二阶校正不能比拟的维度优势。

由于能直接产生四阶及更高阶张量数据的仪器比较少见,而通过加入时间或pH值等维度构建更高维数据的实验工作量很大,所以关于更高阶校正方法的研究相对较少。现有的高阶校正大多拓展自迭代类的三阶校正方法,如交替五线性分解(alternating quinquelinear decomposition, AQQLD)^[52]、约束交替多线性分解(constrained alternating multilinear decomposition, CAMLD)^[53]、展开的偏最小二乘结合残差多线性(unfolded partial least-squares with residual multilinearization, U-PLS/RML)^[54]等。更高维校正方法与三维校正方法一样具有“二阶优势”,同时也具有一些额外的其他优势——“高阶优势”。由于目前研究有限,高维校正方法的更多潜能还有待在理论和实际分析应用中进一步探索和发现。

2 化学计量学方法在食品分析中的应用

2.1 多类聚类 and 判别分析方法在食品分类和判别分析中应用

多类聚类 and 判别分析方法在解决食品品质分级分类和判别分析问题中具有很高的应用价值,如用于食品品质精细分级和不同品牌分类、原产地溯源识别研究等;不同等级或劣质产品通常可根据光谱、色谱、传感等现代分析仪器获得的样品指纹图谱或指标成分参数,再利用多类聚类 and 判别分析方法进行聚类分析或特征提取,构建模型来识别出不同样品的属性类别。Fu等^[55]利用FT-NIR结合偏最小二乘判别分析(PLSDA)实现了不同货架期及霉变葵花籽、大豆的判别分析,此外,还利用FT-NIR结合群策策略的PLSDA方法实现了25种不同原产地的龙井的快速判别^[56];Liu等^[57]利用“Turn-off”荧光传感结合PLSDA实现了29种不同物种、不同产区 and 不同等级绿茶的高灵敏识别。Wu等^[58]利用HPLC结合主成分分析(PCA)和逐步线性判别分析(S-LDA)实现不同种类茶叶的判别分析;Lu等^[59]利用UPLC-MS和FIMS指纹图谱结合PLSDA用于判别不同产地的中国枸杞;Bajoub等^[60]利用HPLC-ESI-IT MS结合线性判别分析(LDA)实现不同产地橄榄油的判别分析;研究者们利用FT-NIR、E-nose、UV-VIS结合多类聚类 and 判别分析方法实现了不同产地松仁、藏红花,不同种类苹果,不同品

牌的烈酒和醋的分类和判别^[61-64]；Xiao 等^[65]利用 SBSE-TD-GC-MS 和 E-nose 结合主成分分析及聚类分析 (CA) 实现不同产地黄酒的判别分析；Yu 等^[66]基于 PCA 和判别分析 (DA) 分析 GC/MS、E-nose 数据实现不同年限黄酒的判别；Chen 课题组利用 GC-MS 结合主成分分析对不同制备方法下柚皮精油感官评价数据降维,分析原始精油组以及经过光照作用、氧气作用和热处理等精油组间风味特

征的差异,并通过聚类分析对各组精油的香气值进一步分析,将各组精油中挥发性风味成分分为 5 类,阐释了不同精油组成分变化与风味变化的关系^[67]；该课题组利用相同的化学计量学方法还成功研究了不同处理方式的柚子汁风味成分变化与风味关系^[68],不同提取方式的柚子皮的香气物质变化特性^[69]。表 1 列举了模式识别中多类聚类和判别分析方法在食品分类和判别分析中的一些应用。

表 1 多类聚类和判别分析方法在食品分类和判别分析中的应用

食品类别	检测技术	化学计量学方法	参考文献
不同货架期及霉变葵花籽、大豆	FT-NIR	PLSDA	[55]
不同原产地的龙井	FT-NIR	基于 ES 的 LCNC	[56]
不同物种、产地及等级的绿茶	“Turn-off”荧光	PLSDA	[57]
不同种类的茶	HPLC	PCA、S-LDA	[58]
不同产地的中国枸杞	UPLC-MS、FIMS	PLS-DA	[59]
不同产地的橄榄油	HPLC-ESI-IT MS	PCA、LDA	[60]
不同产地的松仁	FT-NIR	DPLS	[61]
不同种类的苹果	FT-NIR	PCA、FDCM	[62]
不同品牌的烈酒、醋	E-nose	KFDA	[63]
不同产地的藏红花	UV-VIS	PCA、LDA	[64]
不同产地的黄酒	SBSE-TD-GC-MS	PCA、CA	[65]
不同年限的黄酒	GC/MS、E-nose	PCA、DA	[66]
不同制备方法的柚皮精油风味成分	GC-MS	PCA、CA	[67]
不同处理方式的柚子汁风味成分	GC-MS	PCA、CA	[68]
不同提取方式的柚子皮的香气物质	GC-MS	PCA、CA	[69]

2.2 一类分类的类模型方法在食品掺假无目标分析中应用

食品非法添加、掺杂与假冒的检测是世界性难题。“有目标检测”方法,只有在预先获知或怀疑某成分可能存在时,才能展开相关的“被动”检测。针对“有目标”检测的不足,发展“无目标”检测方法是一种解决食品掺假问题的有效途径。Xu 等利用光谱技术结合 OCPLS 的算法对花生油^[70]、芝麻油^[71]、糯米粉^[72]、全脂奶粉^[32]、蜂胶^[73]、西湖藕粉^[74]、酸奶^[75]、板蓝根^[76]、葛根粉^[77]和道地香菇^[78]等中的各种非法添加的掺假物进行无目标检测,取得了较满意的结果。Zhang 等^[79]应用一类支持向量机对芝麻油进行无目标检测。一类分类的类模型方法在食品掺假无目标分析应用中的例子^[80],见表 2。

2.3 一阶校正方法在食品定量分析中的应用

一阶校正方法是多维校正理论体系中最成熟的校正方法。一阶校正可以同时分析多个待分析物,并且不要求分析检测技术对待分析物的响应信号具有完全的选择性。一阶校正方法构建的数学模型比较简单,数学原理容易理解并且计算方法容易实现,在食品定量分析中得到广泛的应用。表 3 列举一阶校正方法在食品中定量分析的一些应用实例。Fan 等^[81]利用“Turn-off”荧光传感结合基于改进的支持向量机,实现在绿茶和污水等复杂基质中多种农药残留的同时定量表征和分析；Yang 等^[82]利用傅里叶变换近红外光谱 (FT-NIR) 结合偏最小二乘回归 (PLSR) 测定中草药中农残的含量；Hou 等^[83]利用表面增强拉曼光谱 (SERS) 结合 PLSR 定

表 2 一类分类的类模型方法在食品掺假无目标分析中的应用

Tab.2 Summary of studies on untargeted detection of illegal adulterations in foods by class modeling techniques				
纯净食物	掺假物质	检测技术	化学计量学方法	参考文献
花生油	其他食用油	MIR 光谱技术	PLSCM	[70]
芝麻油	廉价油	FTIR 光谱技术	SIMCA、PLSCM	[71]
糯米粉	小麦粉、滑石粉	NIR 光谱技术	SIMCA、OCPLS	[72]
全脂奶粉	工业明胶、蛋白粉	NIR 光谱技术	OCPLS	[36]
蜂胶	杨树橡胶	NIR 光谱技术	OCPLS	[73]
西湖藕粉	廉价淀粉	NIR 光谱技术	SIMCA、PLSCM	[74]
酸奶	工业明胶、蛋白粉	NIR 光谱技术	OCPLS	[75]
板蓝根	苹果皮	NIR 光谱技术	SNV-OCPLS SNV-PLSDA	[76]
葛根粉	增白剂、滑石粉	NIR 光谱技术	PLSCM	[77]
道地香菇	非道地香菇	NIR 光谱技术	PLSDA OCPLS-DA	[78]
芝麻油	混合食用油	GC-MS	OC-SVM	[79]
木薯粉	马来酸	NIR 光谱技术	OCPLS	[80]

量分析单组分和混合的食品防腐剂含量;有文献报道,利用 NIR 高光谱成像结合包括 PCR 和 PLSR 两种多元回归方法对加拿大小麦中的蛋白质含量和硬度值进行建模分析,并比较两种一阶校正方法对其预测性能的差异^[84],以及鱼油微囊补充剂中蛋白质含量^[85];Qin 等^[86]利用太赫兹光谱结合 PLSR 对抗生素盐酸四环素粉末和溶液进行无损预测分析;许多研究者们用高光谱成像(HSI)成功实现了在食品

中的各种营养成分、色素成分等的高灵敏的定量分析^[87-90];Xiao 课题组利用 PLSR 很好地表征了甜橙精油、蔓越莓、榴莲酒中香气物质与感官评价得分间的相关关系^[91-93]。值得指出的是,目前一阶校正正在食品定量分析应用中,共性的特点是需要有足够代表性的样品校正集来应付未知待测样本所有的预期变化。

表 3 一阶校正方法在食品定量分析中的应用

Tab.3 Summary of studies employing first-order calibration for food quantitative analysis				
被分析物	检测技术	化学计量学方法	参考文献	
绿茶中农药残留量	荧光光谱	PSO-OWLS-SVM	[81]	
中草药中农药残留量	FT-NIR 光谱	PLSR	[82]	
食品中农药防腐剂量	SERS 光谱	PLSR	[83]	
加拿大小麦中蛋白质含量、硬度值	NIR 高光谱成像	PLSR、PCR	[84]	
鱼油微囊补充剂中蛋白质含量	ATR-FTIR 光谱	PLSR	[85]	
粉末和溶液形式的盐酸四环素	THz 光谱	PLSR	[86]	
鱼肌肉组织的化学性质	NIR 和 HSI 光谱	PLSR	[87]	
鸡肉中的羟脯氨酸含量	HSI 光谱	PLSR	[88]	
红肉中总色素量	HSI 光谱	SPA-PLSR	[89]	
草鱼片中硫代巴比土酸值	HSI 光谱	PLSR	[90]	
甜橙精油中主要香气物质含量	感官评审得分	PLSR	[91]	
蔓越莓中主要香气物质含量	感官评审得分	PLSR	[92]	
榴莲酒中主要香气物质含量	感官评审得分	PLSR	[93]	

2.4 二阶及高阶校正在食品复杂基质未知干扰共存下对目标分析物的应用

二阶校正具有“数学分离”代替或简化“物理和化学分离”的“二阶优势”,可以实现食品安全检测所要求的现场、快速和实时分析,因此具有广阔的应用前景。近年来,具有“二阶优势”的化学计量学多维校正方法已被广泛应用于食品复杂基质中目标分析物的定量研究,表 4 列举了一些二阶及高阶校正方法在各种食品复杂基质中多组分被分析物的同时定量分析的研究工作。例如:El-Sheikh 等^[94]采用紫外光谱(UV)结合 pH 值变化产生三维数据,基于平行因子分子(PARAFAC)和基于残差双线性的双线性最小二乘(bilinear least squares/residual bilinearization, BLIS/RBL)方法,实现了在未知干扰共存下 4 种软饮料中人工合成色素柠檬黄、诱惑红、日落黄、胭脂红和亮蓝的直接同时定量分析,并与色谱分析的结果相比较,结果表明,分析所得的定量结果准确和可靠。Alarcón 等^[95]采用激发-发射矩阵(excitation-emission matrix, EEM)荧光结合基于残差双线性的展开偏最小二乘方法(unfolded partial least squares/residual bilinearization, U-PLS/RBL)和 PARAFAC 方法实现了在未知干扰共存情况下食用橄榄油和葵花籽油中 7 种多环芳烃的定量分析,获得了与液相色谱-荧光检测一致的定量结果,这种数学分离的方法具有简单、快速和绿色的优势,与色谱分析相比,极大地缩短了分析时间,减少了有机溶剂的消耗。Boeris 等^[96]利用 HPLC-DAD 测定了果汁、水果和蔬菜等复杂食品基质中 5 种农药的残留量。在简单的等度洗脱条件下,复杂食品样品的色谱洗脱曲线中出现了色谱峰共流出、洗脱时间漂移、色谱峰形变化和未知背景干扰等。即便如此,利用

基于二阶校正的多元曲线分辨交替最小二乘(MCR-ALS)无需将食品中各种内源干扰与目标分析物 5 种农残色谱峰进行峰的精细优化和分离,依然得到了准确的定量结果。Lozano 等^[97]采用超高速高效液相色谱与激发-发射矩阵(EEM)荧光光谱检测器联用(UPLC-EEM)技术对多个样本测量并构建四维数据阵,结合 PARAFAC、基于残差三线性的展开偏最小二乘(unfolded partial least squares/residual trilinearization, U-PLS/RTL)和基于残差三线性的多维偏最小二乘(multi-way partial least squares/residual trilinearization, N-PLS/RTL)方法对数据进行解析,无需样本物理和化学繁琐的分离预处理步骤,即使在有荧光响应的其他干扰物共存的情况下,成功实现了橄榄油中叶绿素 a 和 b 以及脱镁叶绿素 a 和 b 的直接和快速同时定量分析。Rubio 等^[98]基于标准加入法引入实际样品提取物不同稀释倍数作为第四维度,构建与三维荧光相结合的四维数据阵,采用 four-way PARAFAC 进行数据分析,在分析物光谱互相重叠且与基质背景光谱也重叠的情况下,实现了卷心菜中氨基甲酸酯类农药的定量分析,同时还识别出了不同稀释倍数下卷心菜基质的响应和影响情况。Yu 等^[99]在简单的等度洗脱条件下,采用 HPLC-DAD 结合三线性分解方法测定了蜂蜜中的 12 种喹诺酮类物质的含量。尽管各被分析物喹诺酮类物质间出现了严重的峰共流出问题,同时蜂蜜基质中也存在各种未知干扰物,借助二阶校正方法的“二阶优势”依然实现了对蜂蜜中 12 种喹诺酮类物质的直接同时定量分析。Zhang 等^[100]采用 HPLC-DAD 结合二阶校正方法的策略,极大地简化了样本的预处理过程,实现了 5 种蜂蜜中 9 种多酚类物质的同时定量分析。Wang 等^[101]在不经提取,直接稀

表 4 二阶及高阶校正方法在食品复杂基质中多组分被分析物的同时定量分析
Tab. 4 Summary of studies on second-order or higher-order calibration for quantitative analysis of analytes in foods

被分析物	检测技术	化学计量学方法	参考文献
饮料中人工合成色素	UV 光谱	PARAFAC 和 BLIS/RBL	[94]
橄榄油和葵花籽油中 7 种多环芳烃	EEM 荧光	U-PLS/RBL 和 PARAFAC	[95]
果汁、水果和蔬菜中 5 种农药残留	HPLC-DAD	MCR-ALS	[96]
橄榄油中叶绿素 a 和 b 以及脱镁叶绿素 a 和 b	UPLC-EEM	PARAFAC、U-PLS/RTL 和 N-PLS/RTL	[97]
卷心菜中氨基甲酸酯类农药	EEM 荧光	four-way PARAFAC	[98]
蜂蜜中 12 种喹诺酮类物质	HPLC-DAD	ATLD	[99]
蜂蜜中 9 种多酚类物质	HPLC-DAD	ATLD	[100]
植物油中 10 种抗氧化剂	HPLC-DAD	APTLD	[101]

释植物油样本后采用 HPLC-DAD 方法,在 6 min 内使 10 种抗氧化剂全部流出,然后结合交替惩罚三线分解 (APTLD) 方法对获得的 HPLC-DAD 数据进行解析,在有重叠峰、未知背景干扰和基线漂移的情况下,快速实现了 5 种植物油样本中 10 种抗氧化剂的同时数学分辨和定量表征。上述二阶及高阶校正方法在各种复杂食品分析问题中的研究表明,与传统的化学分析方法相比,基于二阶及高阶校正的化学计量学方法应用于复杂食品体系分析时有诸多优势,必将在食品化学、食品分析等学科中获得越来越多的实际应用。

3 评论和展望

在食品分析研究的过程中,化学计量学方法在数据解析和分析中发挥了重要的作用,为分析复杂的食品体系提供了一个有别于传统食品研究的新思路。特别是基于二阶校正及更高阶校正的化学计量学方法,能够利用数学分离代替或部分代替物理及化学分离,进而直接进行食品中多个组分的同时分辨和直接定量表征和分析,为满足食品实际生产过程和质量安全监控等要求提供了一条简单、快速、有效的绿色节能新途径。然而,目前采用化学计量学进行食品分析的研究和应用还有待进一步扩大深度和空间。因此,新的检测技术和新的化学计量学分析方法对丰富食品分析的内容,显得极其重要。新的检测技术例如高光谱成像技术、纳米材料传感新技术,为食品分析注入了新的活力,并在食品领域有着极大的潜在应用前景。除此之外,进一步扩大各种分析仪器的“数据融合技术”,并根据分析样品不同维度的物理和化学信息,反映出相比于单独仪器技术更全面的样品信息,对于深度阐释和揭示食品的整体理化特性和作用机制等具有重要意义。化学计量学将一些优化策略如变量筛选、稳健分析计算等引入到现有的经典算法中,提出一些新的具有普适性应用模式的化学计量学方法,可以改善现有方法的一些缺点,简化分析模型、缩短分析时间、扩大模型的适用性、提高模型的准确度和精度,也将进一步推动化学计量学在食品分析中各种实际应用的发展。

4 结 论

化学计量学是分析化学的前沿研究方向之一,

并被成功引入食品分析领域,为解决食品分析中的诸多问题和难题提供了强有力的数据解析工具和分析方法。目前,化学计量学在食品分析中的应用日趋广泛,但由于食品种类繁多和成分复杂,仍然有待做更多和更深入的研究工作,以期在实际应用中更好地完成食品分析的任务和目标。鉴于化学计量学方法的各种优势,研究者们应充分借助化学计量学的不同方法,扩大其在食品分析中的应用范围,开拓新的应用领域,促进食品分析方法的发展和实际应用研究;同时,研究者们也应发展更多适应于处理各种现代新仪器分析方法和新检测技术所产生的、海量的、不同结构类型数据的化学计量学新理论,使其在食品分析领域中具有更为广阔的应用前景。

参考文献:

- [1] DZANTIEV B B, BYZOVA N A, URUSOV A E, et al. Immunochromatographic methods in food analysis [J]. Trends in Analytical Chemistry, 2014, 55(55): 81-93.
- [2] GARCÍA-REYES J F, MOLINA-DÍAZ A, FERNANDEZ-ALBA A R. Identification of pesticide transformation products in food by liquid chromatography/time-of-flight mass spectrometry via “fragmentation-degradation” relationships [J]. Analytical Chemistry, 2015, 79(1): 307-321.
- [3] POREP J U, KAMMERER D R, CARLE R. On-line application of near infrared (NIR) spectroscopy in food production [J]. Trends in Food Science & Technology, 2015, 46(2): 211-230.
- [4] KAROUI R, DOWNEY G, BLECKER C. Mid-infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships-a review [J]. Chemical Reviews, 2010, 110(10): 6144-6168.
- [5] BUTLER O T, CLOUGH R, COOK J M, et al. Current trends: a perspective from 30 years of atomic spectrometry updates [J]. Journal of Analytical Atomic Spectrometry, 2016, 31: 32-34.
- [6] LOUTFI A, CORADESCHI S, MANI G K. Electronic noses for food quality: a review [J]. Journal of Food Engineering, 2015, 144: 103-111.
- [7] KANG B S, LEE J E, PARK H J. Electronic tongue-based discrimination of Korean rice wines (makgeolli) including prediction of sensory evaluation and instrumental measurements [J]. Food Chemistry, 2014, 151: 317-323.
- [8] LAVINE B, WORKMAN J. Chemometrics [J]. Analytical Chemistry, 2010, 82(12): 4699-4711.

- [9] SANCHEZ E, KOWALSKI B R. Tensorial calibration: I. first-order calibration [J]. *Journal of Chemometrics*, 1988, 2(4): 247–263.
- [10] WU H L, NIE J F, YU Y J, et al. Multi-way chemometric methodologies and applications: a central summary of our research work [J]. *Analytica Chimica Acta*, 2009, 650(1): 131–142.
- [11] OLIVIERI A C, ESCANDAR G M, GOICOECHEA H C, et al. Fundamentals and analytical applications of multi-way calibration [M]. Amsterdam: Elsevier, 2015: 7–8.
- [12] BRERETON R G. Multivariate pattern recognition in chemometrics [J]. *Chemometrics & Intelligent Laboratory Systems*, 2015, 149: 90–96.
- [13] WOLD S, ESSENSEN K, GELADI P. Principal component analysis [J]. *Chemometrics and Intelligent Laboratory Systems*, 1987, 2(1–3): 37–52.
- [14] LUO W, HUAN S, FU H, et al. Preliminary study on the application of near infrared spectroscopy and pattern recognition methods to classify different types of apple samples [J]. *Food Chemistry*, 2011, 128(2): 555–561.
- [15] FU H Y, HUAN S Y, XU L, et al. Moving window partial least-squares discriminant analysis for identification of different kinds of bezoar samples by near infrared spectroscopy and comparison of different pattern recognition methods [J]. *Journal of Near Infrared Spectroscopy*, 2007, 15(2): 291–297.
- [16] FU H Y, HUAN S Y, XU L, et al. Construction of an efficacious model for nondestructive identification of traditional Chinese medicines Liuweidihuang pills from different manufacturers using near infrared spectroscopy and moving window partial least-squares discriminant analysis [J]. *Analytical Sciences*, 2009, 25(9): 1143–1148.
- [17] BOGER Z, GUTERMAN H. Knowledge extraction from artificial neural network models [J]. *Journal of Renewable and Sustainable Energy*, 2015, 4(5): 3030–3035.
- [18] DEBSKA B, GUZOWSKA-ŚWIDER B. Application of artificial neural network in food classification [J]. *Analytica Chimica Acta*, 2011, 705(1/2): 283–291.
- [19] CUBEDDU A, RAUH C, DELGADO A. Hybrid artificial neural network for prediction and control of process variables in food extrusion [J]. *Innovative Food Science & Emerging Technologies*, 2013, 21: 142–150.
- [20] KHAJEH M, BARKHORDAR A. Modelling of solid-phase tea waste extraction for the removal of manganese from food samples by using artificial neural network approach [J]. *Food Chemistry*, 2013, 141(2): 712–717.
- [21] VAPNIK V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995: 267–290.
- [22] BARBOSA R M, NELSON D R. The use of support vector machine to analyze food security in a region of Brazil [J]. *Applied Artificial Intelligence*, 2016, 30(4): 318–330.
- [23] BAO Y, LIU F, KONG W, et al. Measurement of soluble solid contents and pH of white vinegars using VIS/NIR spectroscopy and least squares support vector machine [J]. *Food and Bioprocess Technology*, 2014, 7(1): 54–61.
- [24] SOHRABI M R, DARABI G. The application of continuous wavelet transform and least squares support vector machine for the simultaneous quantitative spectrophotometric determination of myricetin, kaempferol and quercetin as flavonoids in pharmaceutical plants [J]. *Spectrochimica Acta Part A: Molecular & Biomolecular Spectroscopy*, 2016, 152: 443–52.
- [25] BRERETON R G. One-class classifiers [J]. *Journal of Chemometrics*, 2011, 25(5): 225–246.
- [26] FORINA M, OLIVERI P, LANTERI S, et al. Class-modeling techniques, classic and new, for old and new problems [J]. *Chemometrics & Intelligent Laboratory Systems*, 2008, 93(2): 132–148.
- [27] FORINA M, OLIVERI P, JÄGER H, et al. Class modeling techniques in the control of the geographical origin of wines [J]. *Chemometrics & Intelligent Laboratory Systems*, 2009, 99(2): 127–137.
- [28] OLIVERI P, EGIDIO V D, WOODCOCK T, et al. Application of class-modelling techniques to near infrared data for food authentication purposes [J]. *Food Chemistry*, 2011, 125(4): 1450–1456.
- [29] MARINI F, MAGRÌ A L, BUCCI R, et al. Class-modeling techniques in the authentication of Italian oils from Sicily with a protected denomination of origin (PDO) [J]. *Chemometrics & Intelligent Laboratory Systems*, 2006, 80(1): 140–149.
- [30] WOLD S, SJÖSTRÖM M. Comments on a recent evaluation of the SIMCA method [J]. *Journal of Chemometrics*, 1987, 1(4): 243–245.
- [31] FORINA M, LANTERI S, SARABIA L. Distance and class space in the UNEQ class-modeling technique [J]. *Journal of Chemometrics*, 1995, 9(2): 69–89.
- [32] FORINA M, ARMANINO C, LEARDI R, et al. A class-modelling technique based on potential functions [J]. *Journal of Chemometrics*, 1991(5): 435–453.
- [33] POLONOVSKI M. Multilayer feed-forward artificial neural networks for class modeling [J]. *Chemometrics & Intelligent Laboratory Systems*, 2007, 88(1): 118–124.

- [34] FORINA M, OLIVERI P, CASALE M, et al. Multivariate range modeling, a new technique for multivariate class modeling: the uncertainty of the estimates of sensitivity and specificity[J]. *Analytica Chimica Acta*, 2008, 622(1):85–93.
- [35] SRIVASTAVA C. Support vector data description[J]. *Machine Learning*, 2004, 54(1):45–66.
- [36] XU L, YAN S M, CAI C B, et al. One-class partial least squares (OCPLS) classifier[J]. *Chemometrics & Intelligent Laboratory Systems*, 2013, 126(11):1–5.
- [37] XU L, GOODARZI M, SHI W, et al. A MATLAB toolbox for class modeling using one-class partial least squares (OCPLS) classifiers[J]. *Chemometrics & Intelligent Laboratory Systems*, 2014, 139:58–63.
- [38] ARANCIBIA J A, DAMIANI P C, ESCANDAR G M, et al. A review on second and third-order multivariate calibration applied to chromatographic data[J]. *Journal of Chromatography B*, 2012, 910(23):22–30.
- [39] LORBER A. Features of quantifying chemical composition from two-dimensional data array by the rank annihilation factor analysis method[J]. *Analytical Chemistry*, 1985, 57(12):2395–2397.
- [40] SANCHEZ E, KOWALSKI B R. Generalized rank annihilation factor analysis[J]. *Analytical Chemistry*, 1985, 58(2):496–499.
- [41] LI S, GEMPERLINE P J. Eliminating complex eigenvectors and eigenvalues in multiway analyses using the direct trilinear decomposition method[J]. *Journal of Chemometrics*, 1993, 7(2):77–88.
- [42] FABER N M, BRO R, HOPKE P K. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review[J]. *Chemometrics & Intelligent Laboratory Systems*, 2003, 65(1):119–137.
- [43] WU H L, SHIBUKAWA M, OGUMA K. An alternating trilinear decomposition algorithm with application to calibration of HPLC–DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons[J]. *Journal of Chemometrics*, 1998, 12(1):1–26.
- [44] ZHANG S, WU H, YU R. A study on the differential strategy of some iterative trilinear decomposition algorithms: PARAFAC–ALS, ATLD, SWATLD, and APTLD[J]. *Journal of Chemometrics*, 2015, 29(3):179–192.
- [45] CHEN Z P, WU H L, YU R Q. On the self-weighted alternating trilinear decomposition algorithm-the property of being insensitive to excess factors used in calculation[J]. *Journal of Chemometrics*, 2001, 15(5):439–453.
- [46] XIA A L, WU H L, FANG D M, et al. Alternating order calibration with application to interference-free analysis of excitation-emission matrix fluorescence data[J]. *Journal of Chemometrics*, 2005, 19(2):65–76.
- [47] SMILDE A K, TAULER R, HENSHAW J M, et al. Multi-component determination of chlorinated hydrocarbons using a reaction-based chemical sensor. 3. medium-rank second-order calibration with restricted Tucker models[J]. *Analytical Chemistry*, 1994, 66(20):3345–3351.
- [48] RUBIO L, SARABIA L A, ORTIZ M C. Standard addition method based on four-way PARAFAC decomposition to solve the matrix interferences in the determination of carbamate pesticides in lettuce using excitation-emission fluorescence data[J]. *Talanta*, 2015, 138:86–99.
- [49] XIA A L, WU H L, LI S F, et al. Alternating penalty quadrilinear decomposition algorithm for an analysis of four-way data arrays[J]. *Journal of Chemometrics*, 2007, 21(3/4):133–144.
- [50] KANG C, WU H L, YU Y J, et al. An alternative quadrilinear decomposition algorithm for four-way calibration with application to analysis of four-way fluorescence excitation-emission-pH data array[J]. *Analytica Chimica Acta*, 2013, 758(1):45–57.
- [51] FU H Y, WU H L, YU Y J, et al. A new third-order calibration method with application for analysis of four-way data arrays[J]. *Journal of Chemometrics*, 2011, 25(8):408–429.
- [52] QING X D, WU H L, ZHANG X H, et al. A novel fourth-order calibration method based on alternating quinquelinear decomposition algorithm for processing high performance liquid chromatography-diode array detection-kinetic-pH data of naptalam hydrolysis[J]. *Analytica Chimica Acta*, 2015, 861:12–24.
- [53] CHAO K, WU H L, SONG J J, et al. A flexible trilinear decomposition algorithm for three-way calibration based on the trilinear component model and a theoretical extension of the algorithm to the multilinear component model[J]. *Analytica Chimica Acta*, 2015, 878:63–77.
- [54] MAGGIO R M, PEÑA A M D L, OLIVIERI A C. Unfolded partial least-squares with residual quadrilinearization: a new multivariate algorithm for processing five-way data achieving the second-order advantage. Application to fourth-order excitation-emission-kinetic-pH fluorescence analytical data[J]. *Chemometrics & Intelligent Laboratory Systems*, 2011, 109(2):178–185.
- [55] FU H Y, JIANG D, ZHOU R, et al. Predicting mildew contamination and shelf-life of sunflower seeds and soybeans by Fourier transform near-infrared spectroscopy and chemometric data analysis[J/OL]. (2016–11–

- 18)[2017-02-21]. <http://link.springer.com/article/10.1007/s12161-016-0726-9>.
- [56] FU H Y, YIN Q B, XU L, et al. Challenges of large-class-number classification (LCNC): a novel ensemble strategy (ES) and its application to discriminating the geographical origins of 25 green teas[J]. *Chemometrics & Intelligent Laboratory Systems*, 2016, 157:43-49.
- [57] LIU L, FAN Y, FU H Y, et al. "Turn-off" fluorescent sensor for highly sensitive and specific simultaneous recognition of 29 famous green teas based on quantum dots combined with chemometrics[J/OL]. (2017-02-08)[2017-02-11]. <http://www.sciencedirect.com/science/article/pii/S000326701730123X>.
- [58] WU Q J, DONG Q H, SUN W J, et al. Discrimination of Chinese teas with different fermentation degrees by stepwise linear discriminant analysis (S-LDA) of the chemical compounds [J]. *Journal of Agricultural & Food Chemistry*, 2014, 62(38):9336-9344.
- [59] LU W, JIANG Q, SHI H, et al. Partial least-squares-discriminant analysis differentiating Chinese wolfberries by UPLC-MS and flow injection mass spectrometric (FIMS) fingerprints [J]. *Journal of Agricultural & Food Chemistry*, 2014, 62(37):9073-9080.
- [60] BAJOUB A, AJAL E A, FERNÁNDEZ-GUTIÉRREZ A, et al. Evaluating the potential of phenolic profiles as discriminant features among extra virgin olive oils from Moroccan controlled designations of origin[J]. *Food Research International*, 2016, 84:41-51.
- [61] LOEWE V, NAVARRO-CERRILLO R M, GARCÍA-OLMO J, et al. Discriminant analysis of Mediterranean pine nuts (*Pinus pinea* L.) from Chilean plantations by near infrared spectroscopy (NIRS)[J]. *Food Control*, 2017, 73: 634-643.
- [62] WU X, WU B, SUN J, et al. Classification of apple varieties using near infrared reflectance spectroscopy and fuzzy discriminant C-means clustering model [J/OL]. (2016-01-12)[2017-02-21]. <http://onlinelibrary.wiley.com/doi/10.1111/jfpe.12355/pdf>.
- [63] YIN Y, HAO Y, BAI Y, et al. A Gaussian-based kernel Fisher discriminant analysis for electronic nose data and applications in spirit and vinegar classification[J]. *Journal of Food Measurement & Characterization*, 2017, 11(1): 24-32.
- [64] D'ARCHIVIO AA, MAGGI M A. Geographical identification of saffron (*Crocus sativus* L.) by linear discriminant analysis applied to the UV-visible spectra of aqueous extracts[J]. *Food Chemistry*, 2017, 219:408-413.
- [65] XIAO Z, DAI X, ZHU J, et al. Classification of Chinese rice wine according to geographic origin and wine age based on chemometric methods and SBSE-TD-GC-MS analysis of volatile compounds[J]. *Food Science & Technology Research*, 2015, 21(3):371-380.
- [66] YU H, DAI X, YAO G, et al. Application of gas chromatography-based electronic nose for classification of Chinese rice wine by wine age [J]. *Food Analytical Methods*, 2014(7):1489-1497.
- [67] SUN H, NI H, YANG Y, et al. Investigation of sunlight-induced deterioration of aroma of pummelo (*Citrus maxima*) essential oil [J]. *Journal of Agricultural and Food Chemistry*, 2014, 62(49): 11818-11830.
- [68] NI H, HONG P, JI H F, et al. Comparative analyses of aromas of fresh, naringinase-treated and resin-absorbed juices of pummelo by GC-MS and sensory evaluation[J]. *Flavour and Fragrance Journal*, 2015, 30(3): 245-253.
- [69] SUN H, NI H, YANG Y, et al. Sensory evaluation and gas chromatography-mass spectrometry (GC-MS) analysis of the volatile extracts of pummelo (*Citrus maxima*) peel [J]. *Flavour and Fragrance Journal*, 2014, 29(5): 305-312.
- [70] XU L, CAI C, DENG D. Multivariate quality control solved by one-class partial least squares regression: identification of adulterated peanut oils by mid-infrared spectroscopy[J]. *Journal of Chemometrics*, 2011, 25(10):568-574.
- [71] DENG D H, XU L, YE Z H, et al. FTIR spectroscopy and chemometric class modeling techniques for authentication of Chinese sesame oil[J]. *Journal of the American Oil Chemists' Society*, 2012, 89(6):1003-1009.
- [72] XU L, YAN S M, CAI C B, et al. Untargeted detection of illegal adulterations in Chinese glutinous rice flour (GRF) by NIR spectroscopy and chemometrics: specificity of detection improved by reducing unnecessary variations[J]. *Food Analytical Methods*, 2013(6):1568-1575.
- [73] XU L, YAN S M, CAI C B, et al. Untargeted detection and quantitative analysis of poplar balata (PB) in Chinese propolis by FT-NIR spectroscopy and chemometrics [J]. *Food Chemistry*, 2013, 141(4):4132-4137.
- [74] XU L, SHI P T, YE Z H, et al. Rapid analysis of adulterations in Chinese lotus root powder (LRP) by near-infrared (NIR) spectroscopy coupled with chemometric class modeling techniques [J]. *Food Chemistry*, 2013, 141(3): 2434-2439.
- [75] XU L, YAN S M, CAI C B, et al. The feasibility of using near-infrared spectroscopy and chemometrics for untargeted detection of protein adulteration in yogurt;

- removing unwanted variations in pure yogurt [J/OL]. (2013-06-01) [2018-02-11]. <https://www.hindawi.com/journals/jamc/2013/201873/abs/>.
- [76] XU L, FU X S, FU H Y, et al. Rapid detection of exogenous adulterants and species discrimination for a Chinese functional tea (banlangen) by Fourier-transform near-infrared (FT-NIR) spectroscopy and chemometrics[J]. Journal of Food Quality, 2015, 38(6):450-457.
- [77] XU L, SHI W, CAI C B, et al. Rapid and nondestructive detection of multiple adulterants in kudzu starch by near infrared (NIR) spectroscopy and chemometrics[J]. LWT-Food Science and Technology, 2015, 61(2):590-595.
- [78] XU L, FU H Y, YANG T M, et al. Enhanced specificity for detection of frauds by fusion of multi-class and one-class partial least squares discriminant analysis; geographical origins of Chinese shiitake mushroom[J]. Food Analytical Methods, 2016, 9(2):451-458.
- [79] ZHANG L, LI P, NA W, et al. Multivariate adulteration detection for sesame oil[J]. Chemometrics & Intelligent Laboratory Systems, 2017, 161:147-150.
- [80] FU H Y, LI H D, XU L, et al. Detection of unexpected frauds: screening and quantification of maleic acid (MA) in cassava starch (CS) by Fourier transform near-infrared (FT-NIR) spectroscopy[J]. Food Chemistry, 2017, 227:322-328.
- [81] FAN Y, LIU L, SUN D, et al. "Turn-off" fluorescent data array sensor based on double quantum dots coupled with chemometrics for highly sensitive and selective detection of multicomponent pesticides[J]. Analytica Chimica Acta, 2016, 916:84-91.
- [82] YANG T M, ZHOU R, JIANG DU, et al. Rapid detection of pesticide residues in Chinese herbal medicines by Fourier transform infrared spectroscopy coupled with partial least squares regression[J/OL]. (2016-11-14) [2017-02-11]. <https://www.hindawi.com/journals/jspec/2016/9492030/abs/>.
- [83] HOU M, HUANG Y, MA L, et al. Quantitative analysis of single and mix food antiseptics basing on SERS spectra with PLSR method[J]. Nanoscale Research Letters, 2016, 11(1):1-8.
- [84] MAHESH S, JAYAS D S, PALIWAL J, et al. Comparison of partial least squares regression (PLSR) and principal components regression (PCR) methods for protein and hardness predictions using the near-infrared (NIR) hyperspectral images of bulk samples of Canadian wheat[J]. Food and Bioprocess Technology, 2015, 8(1):31-40.
- [85] VONGSVIVUT J, HERAUD P, ZHANG W, et al. Rapid determination of protein contents in microencapsulated fish oil supplements by ATR-FTIR spectroscopy and partial least square regression (PLSR) analysis[J]. Food and Bioprocess Technology, 2014, 7(1):265-277.
- [86] QIN J, XIE L, YING Y. Determination of tetracycline hydrochloride by terahertz spectroscopy with PLSR model[J]. Food Chemistry, 2015, 170:415-422.
- [87] CHENG J H, SUN D W. Partial least squares regression (PLSR) applied to NIR and HSI spectral data modeling to predict chemical properties of fish muscle[J/OL]. (2016-07-12) [2017-02-11]. <http://link.springer.com/article/10.1007/s12393-016-9147-1>.
- [88] XIONG Z, SUN D W, XIE A, et al. Potential of hyperspectral imaging for rapid prediction of hydroxyproline content in chicken meat[J]. Food Chemistry, 2015, 175:417-422.
- [89] XIONG Z, SUN D W, XIE A, et al. Quantitative determination of total pigments in red meats using hyperspectral imaging and multivariate analysis[J]. Food Chemistry, 2015, 178:339-345.
- [90] CHENG J H, SUN D W, PU H B, et al. Suitability of hyperspectral imaging for rapid evaluation of thiobarbituric acid (TBA) value in grass carp (*Ctenopharyngodonidella*) fillet[J]. Food Chemistry, 2015, 171(4):258-265.
- [91] XIAO Z, MA S, NIU Y, et al. Characterization of odour-active compounds of sweet orange essential oils of different regions by gas chromatography-mass spectrometry, gas chromatography-olfactometry and their correlation with sensory attributes[J]. Flavour and Fragrance Journal, 2016, 31(1):41-50.
- [92] ZHU J C, CHEN F, WANG L Y, et al. Characterization of the key aroma volatile compounds in cranberry (*Vaccinium macrocarpon* Ait.) using gas chromatography-olfactometry (GC-O) and odor activity value (OAV)[J]. Journal of Agricultural and Food Chemistry, 2016, 64(24):4990-4999.
- [93] ZHU J C, CHEN F, WANG L Y, et al. Comparison of aroma-active compounds and sensory characteristics of durian (*Duriozibethinus* L.) wines using strains of *Saccharomyces cerevisiae* with odor activity values and partial least-squares regression[J]. Journal of Agricultural and Food Chemistry, 2015, 63(7):1939-1947.
- [94] EL-SHEIKH A H, AL-DEGS Y S. Spectrophotometric determination of food dyes in soft drinks by second order multivariate calibration of the absorbance spectra-pH data

- matrices[J]. *Dyes & Pigments*, 2013, 97(2):330–339.
- [95] ALARCÓN F, BÁEZ M E, BRAVO M, et al. Feasibility of the determination of polycyclic aromatic hydrocarbons in edible oils via unfolded partial least-squares/residual bilinearization and parallel factor analysis of fluorescence excitation emission matrices [J]. *Talanta*, 2013, 103(21):361–370.
- [96] BOERIS V, ARANCIBIA J A, OLIVIERI A C. Determination of five pesticides in juice, fruit and vegetable samples by means of liquid chromatography combined with multivariate curve resolution [J]. *Analytica Chimica Acta*, 2014, 814(9):23–30.
- [97] LOZANO V A, PEÑA A M D L, DURÁN-MERÁS I, et al. Four-way multivariate calibration using ultra-fast high-performance liquid chromatography with fluorescence excitation-emission detection: application to the direct analysis of chlorophylls a, and b, and pheophytins a, and b, in olive oils[J]. *Chemometrics & Intelligent Laboratory Systems*, 2013, 125(1):121–131.
- [98] RUBIO L, SARABIA L A, ORTIZ M C. Standard addition method based on four-way PARAFAC decomposition to solve the matrix interferences in the determination of carbamate pesticides in lettuce using excitation-emission fluorescence data[J]. *Talanta*, 2015, 138:86–99.
- [99] YU Y J, WU H L, SHAO S Z, et al. Using second-order calibration method based on trilinear decomposition algorithms coupled with high performance liquid chromatography with diode array detector for determination of quinolones in honey samples [J]. *Talanta*, 2011, 85(3):1549–1559.
- [100] ZHANG X H, WU H L, WANG J Y, et al. Fast HPLC–DAD quantification of nine polyphenols in honey by using second-order calibration method based on trilinear decomposition algorithm [J]. *Food Chemistry*, 2013, 138(1):62–69.
- [101] WANG J Y, WU H L, CHEN Y, et al. Fast analysis of synthetic antioxidants in edible vegetable oil using trilinear component modeling of liquid chromatography–diode array detection data[J]. *Journal of Chromatography A*, 2012, 1264:63–71.

Applications of Chemometrics in Food Analysis

CHEN Feng¹, LI Hedong², WANG Yaqi¹, FU Haiyan^{1,2,*}, ZHENG Fuping^{1,3}

(1. *Department of Food, Nutrition and Packaging Sciences, Clemson University, Clemson, SC 29634, USA;*

2. *School of Pharmaceutical Sciences, South-Central University for Nationalities, Wuhan 430074, China;*

3. *Beijing Advanced Innovation Center for Food Nutrition and Human Health, Beijing Technology and Business University, Beijing 100048, China)*

Abstract: Chemometrics is a novel and unique interdisciplinary technique that integrates mathematics, statistics, chemistry and computer science. Pattern recognition and multivariate calibration method, as the core aspects of the technique, have exhibited overwhelming advantages in massive data mining and processing, as well as resolution and analysis of analytic signals. In addition, it is able to solve complex problems which are difficult to be solved by conventional analytic methods, resulting in the method to be widely used in many research fields, including food analysis. This paper reviews the fundamental principles of chemometrics in regards of its advantages and disadvantages, applications, and recent progresses. Furthermore, the review has pointed out problems which need to be solved in current researches, and highlighted the prospective aspects of the technique.

Keywords: chemometrics; food analysis; pattern recognition; multivariate calibration

(责任编辑:叶红波)