

编者按:营养、健康和美味的食品一直是人类永恒的追求。优质美味的食品,可使人们获得更多的愉悦感和满足感,享受生活带来的美好感受。人类对食物的感知,包括视觉、味觉、嗅觉、触觉、听觉 5 个方面。食品感知科学与技术是一门系统研究人类感官与食物相互作用形式与规律的交叉学科,涉及物理学、生理学、心理学及大脑神经科学等。其从科学的角度帮助解决人们对食物感官的更高层次需求。本刊开设“食品感知技术”特约专栏,连续刊载有关文章。

(本刊编辑部。本栏目得到嘉兴未来食品研究院的支持。客座编辑:江南大学食品学院钟芳教授。)

数据挖掘与建模技术在食品嗅觉和味觉感知与情绪认知中的应用

王 蓓¹, 王 颖¹, 王亚东¹, 刘 帅^{2*}, 江 滔^{3*}

¹北京工商大学食品与健康学院 北京 100048

²北京工商大学数学与统计学院 北京 100048

³勃艮第大学 里昂神经科学研究中心 法国布龙 69500

摘要 随着食品嗅觉和味觉感官及情绪认知相关的科学技术水平的不断提升,越来越多的仪器分析方法和实验设备用于上述领域的研究。检测方法的多样化、全面化以及检测精度的提升,也伴随着风味感知相关数据规模的扩增。如何从食品风味仪器分析结果以及消费者的情绪认知行为相关研究中收集的大量数据中获得关键信息,并建立数据间的关联,越来越被研究人员所关注。食品领域的数据挖掘与建模技术是利用食品生产及流通过程中获得的大量数据,实时、准确地监控食品产业链的各个环节中的物理化学变化,并预测这些变化可能对消费者感官特征及情绪认知产生的影响。目前在食品嗅觉和味觉感知领域,数据挖掘与建模技术可为食品科研人员和消费者提供前所未有的洞察力和分析能力。本文在针对经典的机器学习方法中的有监督和无监督的数据挖掘与建模方法以及深度学习方法的基础上,对食品感官属性研究以及情绪认知方面的最新应用进展进行分析,并展望数据挖掘与建模技术在食品嗅觉和味觉感知领域的应用前景,助力食品行业的科技进步和产业升级。

关键词 数据挖掘与建模;机器学习;风味感知;感官属性;情绪认知

文章编号 1009-7848(2024)07-0001-13 **DOI:** 10.16429/j.1009-7848.2024.07.001

如何生产出消费者愿意接受或者追求的,具有良好感官特性的产品,可以给消费者带来正面情感体验,是所有食品生产、销售的终极目标之一。食品嗅觉与味觉的感官体验对消费者的偏好性和选择性及整个消费过程,起着至关重要的作用^[1]。当前对大部分食品嗅觉和味觉特性的认知,是基于个人对食品引起的感官反应的直接测量,取决于评价者个人区分、感受及表达感官属性的

能力^[2]。随着人们对感官属性认知需求的增加,各种感官评价新技术和新方法不断涌现。随着感官评价方法的创新,越来越多的测量食品嗅、味觉等相关属性的仪器和设备被开发出来。食品的嗅、味觉感知是一种复杂的体验,受食物属性(包括气味、滋味、质构等)、个人因素(如基因、生理、情绪状态)及饮食环境乃至文化背景的影响^[3-4]。为了指导食品生产的发展及创新,研究人员需要综合多种数据来源,了解消费者的感官需求,例如:通过仪器测量得到的食品理化性质的数据,或通过问卷调查获得消费者的感官评价数据以及模拟人体感觉的数据等^[5]。通过不同来源多模态数据的整合,可以获得更加全面、准确的风味感知信息,为

收稿日期:2024-07-22

基金项目:国家自然科学基金项目(32072345)

第一作者:王蓓,女,博士,教授

通信作者:刘帅 E-mail: ls@btbu.edu.cn

江滔 E-mail: tao.jiang@univ-lyon1.fr

常用的方法。机器学习是一门涉及统计学、凸分析、概率论和近似理论的交叉学科^[7]。机器学习可以在给定的数据集上拟合数学或统计函数，并应用于预测化合物的功能特性^[8]，它使机器能够自动从数据和过去的经验中学习，以识别模式并在最少的人为干预下作出预测。通常来说，按照任务类型不同，可以将机器学习模型分为有监督和无监督两类。

机器学习技术是实现人工智能(Artificial intelligent, AI)的重要手段之一。为了制造更高性能的人工智能,科学家们建立了更加复杂的机器学习方法,由此产生了深度学习。深度学习是机器学习的前沿领域,它将机器学习推到一个新的高度。目前的深度学习基本由多层次深度神经网络来实现。接下来,按照经典机器学习方法与深度学习 2 个方面对常用的数据挖掘与建模技术进行介绍,如表 1 所示。

表 1 常用的数据挖掘与建模技术在食品中的应用

类型	方法	原理	适用任务	局限性	具体应用	参考文献
无监督	主成分分析(Principal component analysis, PCA)	正交变换,方差最大化旋转,降维	降维、特征提取、相关性、数据分布、描述型分析	提取的主成分难以解释其含义,所选择的变量会影响主成分的构成	采用主成分分析对咖啡的挥发性成分进行特征提取,确定影响感官特征的主要因素	[9]~[11]
	层次聚类(Hierarchical clustering, HC)	根据产品轮廓的相似度,以不同的阈值划分所属聚类,形成聚类路径图	聚类、探索性分析	不能给出明确的分类信息	根据啤酒的口感维度,采用层次聚类对啤酒分级	[12]
	k 均值聚类(k-means)	迭代聚类中心,判别样本所属类别	聚类、探索性分析	需要人工指定聚类的个数,聚类的选取直接影响聚类效果	研究味觉感知时,采用 k 均值聚类对受试者按照口味偏好划分亚群,研究不同群体的感官特征	[13],[14]
有监督	自编码器 (Auto encoder, AE)	神经网络,特征转换和压缩	降维、特征提取、数据压缩	方法较为复杂,使用难度高	用自编码器将面部表情图片转化为特征向量	[15]
	K 最近邻法(K-nearest neighbor, KNN)	计算样品间距离,根据最近几个样品判定其类别	分类	需要人工确定近邻范围,近邻范围的选取直接影响分类效果	对电子舌采集的数据进行 K 最近邻法分类,识别物质类别	[9]
	线性回归 (Linear regression, LR)	线性方程描述变量相关性	预测、相关性分析	只能拟合线性关系,不适用于非线性复杂数据	建立活性炭和异味去除率的线性回归方程,预测水体异味去除率	[16],[17]
	逻辑回归 (Logistic regression)	二分类问题的线性方程	预测、相关性分析	只能拟合线性关系,不适用于非线性复杂数据	使用逻辑回归模型对牛奶进行分类	[13]
	支持向量机 (Support vector machine, SVM)	映射核函数到超空间,求解超平面决策边界	分类	计算复杂度较高,对模型参数较为敏感	用自编码器将面部表情图片转化为特征向量,再输入支持向量机模型进行感官分类	[13],[15]

(续表 1)

类型	方法	原理	适用任务	局限性	具体应用	参考文献
	偏最小二乘法 (Partial least squares, PLS)	多元线性回归分析+典型相关分析+主成分分析	预测、相关性分析	只能拟合线性关系, 不适用于非线性复杂数据, 对本依赖度高	使用偏最小二乘法对质谱特征进行分类, 并对液相色谱和气相色谱数据进行回归	[10], [11], [18], [19]
	判别分析 (Discriminant analysis, DA)	建立判别函数, 判定样品属于哪个类别	预测和分类	不适用非线性数据, 对异常值和噪声敏感	根据咖啡成分预测感官评分, 主成分分析用于特征诊断, 将相关成分输入偏最小二乘回归进行建模, 预测感官评分	[9], [11], [20]
	随机森林 (Random forest, RF)	特征抽样, 大量决策树进行组合决策	预测、分类、特征选择	黑箱难以解释, 不适用小样本	在基于液相色谱-质谱联用仪的数据分析中, 运用偏最小二乘回归和随机森林方法提炼出重要特征, 确定了与柑橘提取物老化相关的新型化合物	[13]
	梯度提升决策树 (Gradient boosting decision tree, GBDT)	迭代的决策树, 将基分类器层层叠加, 每一层对前一层分错的样本给予更高的权重	预测、分类、特征选择	不能并行计算, 耗时长	测量啤酒的化学成分, 使用梯度提升决策树根据化学成分预测消费者评分	[13]
深度学习	人工神经网络 (Artificial neural network, ANN)	全连接, 1个输入层, 若干隐含层, 1个输出层	预测、分类	不适用小样本	使用人工神经网络对牛奶进行分类	[13]
	卷积神经网络 (Convolutional neural networks, CNN)	图像卷积, 强化局部特征, 输入后面的网络	预测、分类	不适用小样本, 依赖模型结构设计	使用卷积神经网络将激光速度、激光能量通量和面团水分作为输入, 预测该激光会对面团造成什么样的褐变, 并输出预测的褐变图像	[15], [21]
	循环神经网络 (Recurrent neural networks, RNN)	在神经网络中引入循环连接, 在序列的演进方向进行递归	预测、分类	不适用于长序列数据	通过循环神经网络来识别潜在的鲜味肽, 预测鲜味肽的存在	[22]
	长短期记忆结构 (Long short time memory, LSTM)	循环神经网络的改进, 增加长期记忆细胞	预测、分类	不适用小样本	品尝饮料, 用长短期记忆结构从面部表情视频中识别喜好度	[15]

1.1 经典方法

通常来说,按照需要解决问题的类型不同,可以将机器学习模型分为有监督和无监督两类。无监督方法,即没有指定的输出参数(指定的目标),只是为了了解事物本身,寻找出数据包含的事物的结构及关系进行的统计学分析方法。无监督方法更多是用于探索事物的特性,如描述性分析。有监督方法,即有明确的输出参数、指标和类别,参与预测过程本身的统计学分析方法,大多用来预测或分类。

1.1.1 无监督方法 无监督方法针对的数据包含若干自变量 X ,而没有因变量 Y 。因为没有有一个明确的分类标签 Y ,所以无监督方法属于没有指定要建立类别的方法,也无法量化模型效果。通常来说,无监督方法分为降维和聚类两大类。降维是指缩减自变量 X 的数量,以达到简化数据或提取特征的目标,主要包括主成分、奇异值分解等。聚类则是根据所测变量的数据计算出的样品的空间距离或者相似度,将样品聚合成若干组,在没有预先指定样品类别的前提下,将相似样品归为同组,主要包括主成分分析、 k 均值聚类、层次聚类等。

主成分分析是一种经典的无监督方法,适用于无明确因变量的数据,其任务目标是对数据进行重新组合维度,以使得大部分信息投射在靠前的几个维度上(即常说的降维)。在不明显损失数据信息的前提下,用少量线性独立的主成分来代替原始的自变量 X ,由此实现降维和提取特征的目的,并且能够满足某些模型的独立性假设。层次聚类和 k 均值聚类,其任务目标是筛选出可以将样品进行归类的因素。在由多个自变量 X 组成的数据表上,层次上升聚类从每个样品被视作单个类别开始,对每个样品计算出最接近的另一个样品,两两聚类,层层上升,以聚成1个总类,由研究人员确定在哪个层次的聚类已充分包含了数据所带来的信息,继而判断出将产品进行区分的因素。而 k 均值聚类法则是由研究人员根据已掌握的对待分析产品的认知,在分析前就确定了最终所有产品的分类个数,以分析达到所指定的分类个数的因素,所有测量的变量及它们在算法中的权重都可能影响样本相似性的估算。聚类分析可以根据测定得到的多元数据,将样品划分为若干群体,

能够用来细分群体或识别异常样品。在食品风味相关研究中,主成分分析和 k 均值聚类这两种方法是最常用的无监督方法,常用于提取关键风味特征,寻找某一感官特征对应的关键风味化合物等。

1.1.2 有监督方法 有监督方法则针对有明确标签的数据,即数据中除了有对自变量 X 的测量,还必须包含对因变量 Y 的测量,以作为分类或预测的标签数据或目标变量。其任务目标是根据所测的 X 和 Y 的数据建立数学模型,找到一种最准确、最有效的规则,以便在只有 X 数据的情况下就可以预测 Y 的状态。若 Y 为连续的数值型数据,则模型一般为预测性的回归模型;若 Y 为分类型数据,则模型一般为判定类别的分类模型。有监督方法包括线性回归及其衍生(岭回归等)、决策树及其衍生(随机森林、梯度提升决策树等)、支持向量机、 K 最邻近法等。例如常用的偏最小二乘回归就是一种特殊的多元线性回归,它改进了线性回归的求解过程,能够在自变量 X 存在高度相关性的条件下正常建模,因此很适合用于食品风味组分与感官特征这类存在高度内部相关性的数据分析。

1.2 深度学习

除了上述提及的可以对基于所获得数据利用统计描述及概率分布进行学习、训练以对事物作出描述、预测及决策的经典的机器学习方法以外,近年来,基于对大脑神经结构及运作过程的认知,神经网络如何介入生物体的学习,以及记忆过程指导生物体的行为认知的机器学习也衍生出一个重要的分支——深度学习。深度学习是在机器学习技术的基础上,结合大数据技术建立起的一种新的学习理念。由于任务目标和数据蕴含的信息高度复杂化,模型需要有足够多的数据量及变量个数,才能够实现准确及完整的逻辑推理及判断,因此深度学习模型从体量和复杂程度上都迅速上升,训练机器进行建模也就需要有足够多的数据量。随着可积累的数据越来越多,计算机的运算能力的重大提升,深度学习在食品风味研究中的应用也越来越多。

人工神经网络是一种模仿脑神经结构设计的深度学习算法。输入数据会在若干个运算节点中

模拟神经信号的传递方式,每个运算节点都会根据“经验”对输入数据进行不同的加工,再传输到下一个节点,最终产生一个更符合现实场景的输出。20世纪末,在人工神经网络的基础上,出现了卷积神经网络的研究,并由此深度学习方法得到迅速发展^[23]。卷积神经网络最大的特点是在网络结构中加入了卷积层。卷积是一种特殊的积分方式,在求和时赋予每个分割区间1个权重,利用数据挖掘与建模技术进行数据学习或训练的过程就是不断调整权重的过程。图像的卷积就相当于多张图像叠加在一起,每张叠加的图像先按照某个权重进行变换。卷积神经网络的卷积结构能够非常有效地提取图像特征,因而非常适用于图像处理与识别任务。当给定一个由二维数值矩阵表达的图像时,卷积神经网络能够有效地识别图像中有什么,进而对图像进行分类。在食品科学中也经常会应用到图像数据并建立模型,例如对食品图片的分类^[21]、人面部表情分析^[15]等,以此判断消费者基于视觉对食品的选择性以及消费者对食品风味的享乐程度。

2 数据挖掘与建模技术在食品嗅觉和味觉研究中的应用

随着食品检测技术的不断发展,从食品中获得越来越多的数据,而数据处理以及关键信息的获取逐渐成为难题,利用数据挖掘与建模技术可以很好地解决这个问题。数据挖掘与建模技术可以在已知食品的气味、滋味所对应数据的基础上建立模型,从而对食品的感官特性进行预测和分析,如图2所示。近年来,数据挖掘与建模技术在食品风味感知研究中的应用主要分为3种:1)在食品嗅觉感官属性研究中的相关应用研究;2)在食品味觉感官属性研究中的相关应用研究;3)在情绪认知方面的应用研究。

2.1 数据挖掘与建模技术在嗅觉方面的应用

食品的嗅觉感知是食品主观感知中不可或缺的部分,它们影响对食品的整体品质评估,关系消费者的接受度和满意度。气味分子在食品中用于改善嗅觉,以刺激消费者的食欲^[24]。食品风味感知可被视为复杂生化系统的一种突现特性^[25],很难解释和预测,而利用数据挖掘与建模技术,研究人

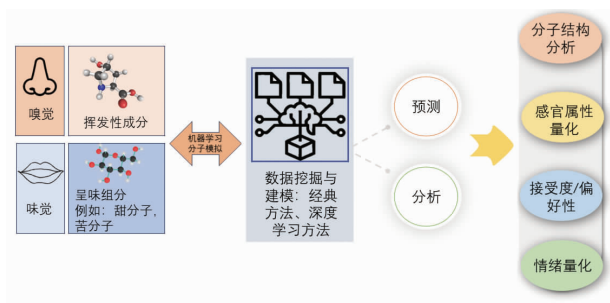


图2 数据挖掘和建模技术在食品嗅觉和味觉研究中的应用

Fig.2 Application of data mining and modeling techniques in food olfactory and gustatory research

员已经找到可以解释或预测一些感官属性的算法。目前研究主要集中在通过已知气味分子结构特征来预测嗅觉感知,以及利用仪器分析检测样品的定性、定量结果来预测嗅觉感知两个方面。

2.1.1 基于气味分子结构特征预测嗅觉感知 物质的分子结构决定了其与嗅觉受体结合的特征,从而导致具有不同的气味特点。气味分子并不与某一个嗅觉受体特异性结合。相反,一个气味分子可以具有几个不同的活性基团,因此可以结合几个不同的嗅觉受体,其亲和力取决于它们的物理化学性质。与气味分子结合后,特定组合的嗅觉感觉细胞向大脑传递信息,形成了人体感知到的气味特征^[26]。分子模拟是嗅觉和味觉研究中常用的方法,主要包括分子动力学和分子对接^[27],其中分子动力学是一个复杂的生物系统的计算模拟,描述了分子运动、分子间相互作用和原子水平的动力学^[28]。分子对接是一种基于锁与钥匙理论的技术,通过计算风味分子与受体之间的相互作用,预测它们可能的结合模式^[29]。目前,分子模拟的应用主要集中于研究受体和小分子之间的相互作用以预测食品感官特征,改善食品感官特性的方面,而对利用分子模拟技术建立的模型主要是基于有监督的经典建模方法和深度学习的建模方法进行分析。

在气味方面,已有学者基于不同算法开发出将食品的风味分子结构与预测食品感官特征结合起来的模型。例如,Chacko等^[30]将196个二维RDKit的分子描述符作为模型的输入值,利用多种数据挖掘与建模算法(如随机森林模型、梯度提升树和支持向量机)对机器进行训练,开发了用于

预测气味特征的机器学习模型。另外,葡萄酒的感官分析也可以从近几十年分子模拟取得的强劲发展中受益,能够从仪器分析得到的挥发性化合物的有效信息中,使用数据挖掘与建模技术来产生能够预测葡萄酒感官特征的模型^[18]。Cardoso等^[18]利用分子模拟技术设计了一个包含葡萄酒的风味分子结构的定量结构-气味关系(Quantitative structure-odour relationship, QSOR)模型来预测葡萄酒的感官分析。模型将一组“描述因子”(如与分子化学结构相关的已知变量)与特定气味感官属性相互关联,从而把风味分子结构与预测食品中的某种特征感官属性相联系^[31]。由于不同分子之间的分子结构差异微妙而复杂,分子特征描述符众多,相互作用复杂^[21],因此寻找气味分子的化学结构和物理化学性质与嗅觉感知预测之间的关系仍是一项十分具有挑战性的任务。随着深度学习的发展,卷积神经网络或深度神经网络等数据驱动方法逐渐被用于定量结构-气味关系的预测。Wang等^[32]提出基于侧抑制原理的特征金字塔动态卷积网络,以特征金字塔网络为主干网络提取气味分子结构特征,能够很好地处理多尺度变化,它们模仿动物嗅觉的侧抑制机制,在动态卷积中加入基于侧抑制原理的注意力图,提高了嗅觉感知预测的准确性。此外,他们还提出在深度神经网络中加入“注意力分解”机制,通过基于注意力机制的加权深度特征组合,获得分子描述特征,这将有助于食品行业的产品设计和质量评估。

此外,近年来食品气味研究相关领域越来越关注在分子特征信息与其对应特定感官属性间建立相关联模型的可能性,从而实现在分子水平有针对性的改变,达到改变与创新产品感官特征的目的^[33-34]。例如,从源头开始进行分子设计案例,为人工香料的设计提供了一种重现度高的方法,这类方法由于算法相对复杂,因此大多会采用深度学习的方法,通过建立神经网络模型来预测具有所需气味特性的分子,与人类利用专业知识设计的分子相比,该算法具有更高的效率和精确度,也具有较好的应用前景^[35]。因此,利用分子生成技术设计人工香料分子在香料研发领域具有广阔前景^[36]。

2.1.2 基于仪器定性、定量结果预测气味特征 食品中风味化学组分的种类和含量会对食品的感官

属性产生较大影响。利用数据挖掘与建模技术,分析食品嗅觉化合物定性、定量的结果来预测主观感知,也越来越被人们所关注。目前,该技术主要集中在预测食品的味觉感官特征以及对食品味觉品质进行分类两个方面。

首先利用气-质谱联用方法、气相色谱-离子迁移谱联用方法等挥发性风味组分分析仪器,检测样品气味组分的定性、定量结果,结合感官定量描述结果,基于不同种类的统计学方法建立模型,可用于预测某种食品的感官特征。目前这类研究大部分采用经典统计学中的有监督方法。Leygeber等^[37]应用气-质谱联用方法对番茄汤的挥发性成分进行定性、定量分析,在此基础上,利用随机森林对气-质谱联用方法的数据进行回归预测,进一步采用偏最小二乘判别分析对番茄汤的气味感官属性建立预测模型,估计质谱特征对模拟感官反应和区分产品成分的重要性。这种方法可将仪器分析检测样品的定性、定量结果与有监督的建模方法关联起来,以选择与感官属性和组合属性相关的特征,从而获得预测风味的新见解。随着深度学习建模技术的发展,人工神经网络模型也逐渐在预测食品气味特征中有所应用。例如,Fuentes等^[38]利用气相色谱和化学计量分析确定的黑皮诺葡萄酒香气谱,与来自黑皮诺葡萄园的天气和水管管理信息作为输入数据,使用高精度的人工神经网络模型来预测葡萄酒的香气特征。

其次,对于同一类型的食品,其风味化学组分种类和含量的区别导致其质量品质差别较大,因此食品挥发性风味组分中的关键特征值,如关键气味成分、浓度、阈值等特征数据的深度挖掘,可以用于食品品质分类、分级的依据。例如:基于食品中气相色谱-质谱分析中鉴定得到风味物质数据,建立判别模型,可以为同一香气类型食品的快速鉴别,提供新的策略。Mildner等^[39]采用主成分分析和线性判别分析两种经典统计学方法,对不同植物油的热重分析和气-质谱联用分离和融合的数据进行化学计量学分析,发现将样品的热重分析数据与挥发性风味组分数据相结合后,可以有效提高主成分分析和线性判别分析模型的识别能力。将该方法应用于大豆油掺假橄榄油的检测中,可将橄榄油样品与掺假量在25%~75%之间的掺

假油样品准确区分。基于热重-气相色谱-质谱联用技术结合的多模态数据挖掘与建模技术,为植物油的鉴定提供了新的思路。Zhou等^[40]选择大米中的独特香气化合物,成功地将数据挖掘与建模的簇类独立软模式法(Soft independent modeling of class analogies, SIMCA)建模技术应用于大米起源识别,对于了解中国不同地区水稻风味特征差异,研究具有地理保护特性的产品感官品质特征具有重要的参考价值。Zhu等^[41]将利用静态顶空-气相色谱-离子迁移谱法(SHS-GC-IMS)分析长相思葡萄酒得到的香气化合物数据建立数据集,以非针对性的方式,对数据集进行计算机建模,通过预测模型将香气化学与葡萄酒感官质量分级联系起来,比较了6种机器学习模型(包括支持向量机、极端梯度增强和人工神经网络等)对葡萄酒等级的预测能力,其中人工神经网络的预测准确率为95.4%,表现最为理想。

需要注意的是由于感官评价者的疲劳性和主观性会影响感官评价的科学性和客观性。鉴于市场需求和智能化发展,近年来电子鼻技术也越来越受到关注。电子鼻技术采集到的数据是复杂气味的混合数据,并且它的信号通常由大量的原始数据组成,因此,为了提高测量结果的可解释性,降低数据中的噪声,需要通过降低数据维数并保留有用信息,最常用的数据简化算法是主成分分析,主要功能是用于样品的识别和监测^[42]。此外,为了更好地挖掘电子鼻传感器采集到的复杂数据,很多深度学习技术,如人工神经网络也得到广泛应用。此外,如何将电子鼻数据与感官特征数据进行更好地关联,用于感官属性的预测和感官品质的判别分析,也是未来智能感官数据挖掘的重要方向。

2.2 数据挖掘与建模技术在味觉方面的应用

2.2.1 基于滋味分子的结构特征预测味觉特征

味觉是人类进化出来的一种复杂的感觉,主要是为了对食物衍生的呈味物质作出反应,不同的滋味分子与味觉受体特异性结合,会引起味觉细胞膜上不同味觉受体的响应,导致形成不同的味觉^[43]。目前,味觉分子模拟主要集中在甜味和苦味这两种滋味的研究上,而分子模拟技术被广泛用于研究受体和小分子之间的相互作用。

随着分子模拟技术的发展,越来越庞大的味觉分子数据库被不断挖掘,目前已建立多个基于甜味化合物和苦味化合物分子结构特征的数据库。2019年,Zheng等^[44]整理了530种甜味剂和850种非甜味剂组成甜味初级数据库,并从文献中收集352种具有相对甜度的甜味剂,组成次级甜味数据库。针对这些试验数据库,他们采用偏最小二乘判别分析、k近邻、多元线性回归和支持向量机4种机器学习方法和分子指纹,分别通过共识策略推导出甜味剂及其相对甜度预测的分类和回归模型,并以上述技术为基础,开发了e-Sweet平台,用于自动预测甜味剂及其相对甜度水平,为设计新的甜味剂提供了有价值的信息。Goel等^[45]利用Material Studio v6.0平台中的遗传函数逼近和人工神经网络结合,混合二维和三维分子描述符,在包含455种甜味剂的数据集上建立了回归模型,用于甜味剂及其相对甜度预测的回归。Tuwani等^[46]利用来自BitterDB14和SuperSweet23两个数据库的苦味和甜味化合物,开发了基于分子结构特征的随机森林模型,并提出实现随机森林和决策树融合的机器学习模型,实现了对苦味分子和甜味分子的预测和分类。综上所述,数据挖掘与建模技术在利用食品滋味分子结构对食品甜味及苦味感官进行预测、分类及回归方面已有广泛应用,该技术在未来新型甜味剂的开发以及甜味剂复配技术等方面也有广阔的应用前景。

2.2.2 基于仪器定性、定量结果预测滋味特征

食品中的滋味化学组分的种类和含量会对食品的感官属性产生较大影响,利用数据挖掘与建模技术分析食品呈味化合物定性、定量的结果,预测主观感知,越来越为人们所关注。

首先,滋味组分的数据发掘和建模方法可以利用液质等可用于滋味组分研究的仪器提供的样品化合物定性、定量信息,然后将其与感官评价小组数据联系起来,从而获得预测风味的新见解。Leygeber等^[37]利用液相色谱-质谱对27种不同加工方法制备的番茄汤进行非挥发组分的化学分析,利用这些数据建立了26种不同滋味感官属性的偏最小二乘回归预测模型,并通过这些模型发现番茄汤的滋味和代谢物之间的关联,达到对不同加工条件下番茄汤的滋味预测的目的。

其次, 滋味化学组分种类和含量的区别也会导致食品品质的差别, 因此, 对食品滋味中的关键特征组分检测与分析可以作为食品品质分类、分级的依据。利用数据挖掘与建模技术得到的模型, 可以为不能通过人体主观鉴别区分的食品提供快速鉴别的新策略。例如, Agorastos 等^[12]对 24 份啤酒样品进行了感官评价和主成分分析, 将啤酒带来的口腔感觉分为收敛感、刺激感和干涩感 3 个维度, 并选取具有统计学意义 ($P < 0.05$) 的主成分分析结果进行聚类分析, 将啤酒样品按 3 个不同口感维度进行分类。

最后, 同气味分析一样, 以电子舌为代表的智能传感仪器在食品滋味分析中的应用也越来越多。与电子鼻相比, 电子舌与食品中各种呈味属性的相关程度更好^[42]。大部分电子舌信号采用多频大幅度脉冲信号伏安法测量, 因此基于电子舌传感器系统采集到的数据信息, 结合各种机器学习技术, 可以更好的解决样品滋味的预测和分类问题。2010 年, Oliveri 等^[47]利用 SciFinder Scholar 数据库进行了一项研究, 证明大多数电子舌传感器的信号处理工作在特征提取阶段大多使用主成分分析, 分类阶段使用人工神经网络, 回归阶段采用偏最小二乘回归。电子舌分析在食品分析中表现出较高的客观性和较少的时间消耗, 并且对测试样品中多组分的定量是非特异性的, 这使得它相对灵活, 可以使用模式识别或多元数据分析等算法进行数据处理。为了能在食品样品上复制预期结果, 智能感知算法的性能还有待提高。

2.3 消费者对食品情绪认知数据挖掘与建模技术的应用

近年来, 越来越多的研究者关注消费者的整体体验, 研究食物和消费者之间的关系, 以及理解情绪如何驱动人类行为, 从而捕捉消费者在享乐喜好相关的反应^[48]。消费者对食物选择可以受到许多内在和外在因素的影响, 这些因素可以相互作用, 形成一系列潜在的复杂变量。数据挖掘与建模技术则可以很好地分析这些复杂变量, 实现消费者对食品选择的预测, 帮助消费者找到理想的食品。

大部分的情绪测量可以采用自我陈述性问卷、生理心理参数测量(如皮肤电导和心率、面部

表情、眼动仪等方法)^[49](图 3), 然而通过这些方法收集到的数据都具有数据冗余较大且不易获得关键信息的特点, 而数据挖掘与建模技术则可以通过收集和分析大量关于食品与情绪关联的数据, 来揭示这种关联的模式和规律。

采用主观变量进行情绪测量时, 一般会用到情绪量表。情绪量表可以评估消费者在看到或摸到食物、闻到食物气味、品尝到食物味道时的有意识的情绪感受, 是研究消费者在食品风味感知中常用的方法^[50]。通过数据挖掘与建模, 可以建立情绪模型来了解情绪在食品选择中的作用, 描述情绪与食品之间的联系。例如, Elali 等^[49]采用包含 25 个词汇的喜好度量表对 7 种价格的葡萄酒诱发的情绪进行评估, 利用对应分析直观地展示了每个消费者对所有样品的情感词汇的使用情况, 并计算每个术语被参与者使用的次数与可能使用该术语的总次数的关系, 从而进一步建立感官人员情绪与葡萄酒感官之间的联系。

采用“客观”变量进行情绪测量时, 可以采用多源变量分析方法。例如面部肌肉运动测量, 面部表情分析以及记录与自主神经活动有关的生理参数, 如心跳、呼吸几率、皮肤电活动等, 在获得数据后, 通过数据挖掘与建模技术进行分析, 将食品与情绪进行关联, 并揭示这种关联的模式和规律。这些数据可能包括人们对不同食物的喜好, 摄入食物后的情绪变化, 以及食物成分与情绪状态之间的相关性等。例如, Yamamoto 等^[51]使用消费者的面

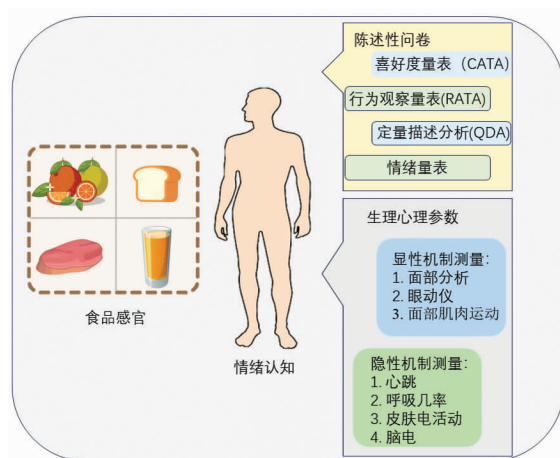


图 3 常见的情绪测量方法

Fig.3 Common measures of emotion

部表情这一“客观”变量用来评估消费者对食品的喜好。他们将每张面部表情图片上传到人工智能应用程序,以获得不同种类情绪评分的结果,并对参与者的享乐情况进行评分,再进一步进行多元线性回归分析,以获得预测感知偏好性的模型,该模型检验结果表明,预测得到的享乐等级与消费者主观评价得到的享乐等级具有良好的相关性和一致性。Dibeklioglu 等^[19]建立了一种从面部表情中自动估计饮料口味喜好度的方法,通过利用面部运动的速度和加速度等表情,动态揭示味觉诱发反应的隐藏模式。该方法将面部表情图片标记、划分为4个不同的区域,即:眉毛/前额、眼睛、脸颊和嘴巴,使用长短时记忆结构对每帧区域面部动态进行深度学习,并将其编码为每个区域的 Fisher 向量来描述视频。然后用线性支持向量机分类器对区域 Fisher 向量进行连接和分类,区分对样品的3种喜好度(不喜欢、中性、喜欢),准确率达到70.37%。

食品情绪感知通常涉及不同类型的多个数据来源,上述多模态数据的存在进一步增加了风味分析的复杂性。人体主观参数测量虽是最快捷的方法,但它受太多不可控因素的影响,造成结果的误差往往很大。如何绕过人体的主观测量,利用客观的理化测量来预测人体的主观反应,一直是产品生产者最终的目标。目前通过数据挖掘与建模技术,可以初步构建出与情绪认知相关的预测模型,这些模型可以根据个人的口味偏好、情绪状态以及食品成分等信息,预测某种食物可能带来的情绪效果,是该类研究的一个发展趋势。需要注意的是,食品情绪感知是一个相对主观和复杂的领域,受到多种因素的影响。因此,上述技术在食品情绪感知方面的应用需要更加谨慎,并且最好结合其它研究方法进行综合比较。

3 结论与展望

近年来,数据挖掘与建模技术在食品风味研究中的应用越来越广泛,为食品科研人员提供了一种新的视角和方法,以深入理解和控制食品的感官特性。一方面,研究人员已成功地利用经典数据挖掘与建模技术,建立感官属性和风味化合物之间的关系。具体而言,回归模型有助于根据系数

来识别关键因素/风味化合物,在分析数据上应用分类方法可以预测食品的感官,对食品进行鉴别分析,评估消费者对食品的接受性和偏好性。最终可以通过应用适当的数据挖掘与建模技术来实现预测食品风味。另一方面,由于未来工作的复杂性,将关注点集中在更先进的深度学习建模技术上。例如利用分子模拟和机器学习技术,将多种来源的数据(如风味分子的分子结构、感官属性、仪器量化数据等)用于风味感知研究中的建模。作为一种计算机模拟方法,上述模型有助于将大量化合物按照其所需的风味特性进行优先排序,从而显著减少用于详细感官分析的候选化学物质的数量,数据建模的可行性和效率得到广泛认可。

虽然数据挖掘与建模技术在食品气味研究中具有诸多优势,但是在实际应用过程中仍面临一些挑战。例如,数据的质量和数量对模型的性能具有重要影响,需要保证数据的准确性和完整性,同时,更新不及时、数据不可访问和代码不公开的问题仍然存在。因此,在发表研究过程中,公开所有数据和代码的呼声也越来越高。未来,随着技术的不断发展和数据的不断积累,数据挖掘与建模技术在食品风味感知领域的应用将更加广泛和深入,也会进一步推动食品产业走向另一个数智化新纪元。

参 考 文 献

- [1] VINCEKOVIĆ M, VISKIĆ M, JURIĆ S, et al. Innovative technologies for encapsulation of mediterranean plants extracts[J]. Trends in Food Science & Technology, 2017, 69: 1-12.
- [2] SHI L F, ZHOU J Y, GUO J Y, et al. Starch inclusion complex for the encapsulation and controlled release of bioactive guest compounds[J]. Carbohydrate polymers, 2021, 274: 118596.
- [3] GHANDEHARI-ALAVIJEH S, CAN KARACA A, AKBARI-ALAVIJEH S, et al. Application of encapsulated flavors in food products: Opportunities and challenges [J]. Food Chemistry, 2024, 436: 137743.
- [4] PREMJIIT Y, PANDHI S, KUMAR A, et al. Current trends in flavor encapsulation: A comprehensive review of emerging encapsulation techniques, flavour

- release, and mathematical modelling[J]. *Food Research International*, 2022, 151: 110879.
- [5] CUI Z Y, WU B, BLANK I, et al. TastePeptides-EEG: An ensemble model for umami taste evaluation based on electroencephalogram and machine learning[J]. *Journal of Agricultural and Food Chemistry*, 2023, 71(36): 13430-13439.
- [6] KOU X R, SHI P Q, GAO C K, et al. Data-driven elucidation of flavor chemistry[J]. *Journal of Agricultural and Food Chemistry*, 2023, 71(18): 6789-6802.
- [7] ACHARY P G R, TOROPOVA A P, TOROPOV A A. Combinations of graph invariants and attributes of simplified molecular input-line entry system (smiles) to build up models for sweetness[J]. *Food Research International*, 2019, 122: 40-46.
- [8] HAN M Y, LIU S, ZHANG D C, et al. Addicted-chem: A data-driven integrated platform for new psychoactive substance identification[J]. *Molecules*, 2022, 27(12): 3931.
- [9] LEON-MEDINA J X, CARDENAS-FLECHAS L J, TIBADUIZA D A. A data-driven methodology for the classification of different liquids in artificial taste recognition applications with a pulse voltammetric electronic tongue[J]. *International Journal of Distributed Sensor Networks*, 2019, 15(10): 812340928.
- [10] PÉREZ -JIMÉNEZ M, SHERMAN E, POZO -BAYÓN M A, et al. Application of untargeted volatile profiling and data driven approaches in wine flavoromics research[J]. *Food Research International*, 2021, 145: 110392.
- [11] BRESSANELLO D, MARENGO A, CORDERO C, et al. Chromatographic fingerprinting strategy to delineate chemical patterns correlated to coffee odor and taste attributes[J]. *Journal of Agricultural and Food Chemistry*, 2021, 69(15): 4550-4560.
- [12] AGORASTOS G, KLOSSE B, HOEKSTRA A, et al. Instrumental classification of beer based on mouthfeel[J]. *International Journal of Gastronomy and Food Science*, 2023, 32: 100697.
- [13] ACHARYA S, DANDIGUNTA B, SAGAR H, et al. Analyzing milk foam using machine learning for diverse applications[J]. *Food Analytical Methods*, 2022, 15(12): 3365-3378.
- [14] GERVIS J E, CHUI K K H, MA J T, et al. Data-driven clustering approach to derive taste perception profiles from sweet, salt, sour, bitter, and umami perception scores: An illustration among older adults with metabolic syndrome[J]. *The Journal of nutrition*, 2021, 151(9): 2843-2851.
- [15] DIBEKLIOGLU H, GEVERS T. Automatic estimation of taste liking through facial expression dynamics[J]. *IEEE Transactions on Affective Computing*, 2020, 11(1): 151-163.
- [16] BERTONE E, CHANG C, THIEL P, et al. Analysis and modelling of powdered activated carbon dosing for taste and odour removal[J]. *Water Research*, 2018, 139(2018): 321-328.
- [17] FARRUGGIA M C. Data-driven analyses in obesity and chemosensory neuroimaging [D]. New Haven: Yale University, 2022.
- [18] CARDOSO SCHWINDT V, COLETTI M M, DÍAZ M F, et al. Could qsr modelling and machine learning techniques be useful to predict wine aroma? [J]. *Food and Bioprocess Technology*, 2023, 16(1): 24-42.
- [19] MOSER B, JANDRIC Z, TROYER C, et al. Evaluation of spectral handheld devices for freshness assessment of carp and trout fillets in relation to standard methods including non-targeted metabolomics[J]. *Food Control*, 2023, 152: 109835.
- [20] MAKRIDIS G, HEYRMAN E, KOTIOS D, et al. Evaluating machine learning techniques to define the factors related to boar taint[J]. *Livestock Science*, 2022, 264: 105045.
- [21] KHOSLA M, RATAN MURTY N A, KANWISHER N. A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition[J]. *Current Biology*, 2022, 32(19): 4159-4171.
- [22] QI L L, DU J L, SUN Y, et al. Umami-mrnn: deep learning-based prediction of umami peptide using rnn and mlp[J]. *Food Chemistry*, 2023, 405: 134935.
- [23] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time-series [J]. *Handbook of Brain Theory and Neural Networks*, 1995, 3361(10): 1995.
- [24] EELAGER M P, MASTI S P, CHOUGALE R B, et al. Evaluation of mechanical, antimicrobial, and antioxidant properties of vanillic acid induced chitosan/poly (vinyl alcohol) active films to prolong the

- shelf life of green chilli[J]. *International Journal of Biological Macromolecules*, 2023, 232: 123499.
- [25] COHEN S M, EISENBRAND G, FUKUSHIMA S, et al. FEMA GRAS assessment of natural flavor complexes: Citrus-derived flavoring ingredients[J]. *Food and Chemical Toxicology*, 2019, 124(2019): 192–218.
- [26] KELLER A, GERKIN R C, GUAN Y, et al. Predicting human olfactory perception from chemical features of odor molecules[J]. *Science*, 2017, 355(6327): 820–826.
- [27] VIDAL-LIMON A, AGUILAR-TOALAÁ J E, LICEAGA A M. Integration of molecular docking analysis and molecular dynamics simulations for studying food proteins and bioactive peptides[J]. *Journal of Agricultural and Food Chemistry*, 2022, 70(4): 934–943.
- [28] ACEVEDO W, GONZÁLEZ-NILO F, AGOSIN E. Docking and molecular dynamics of steviol glycoside-human bitter receptor interactions[J]. *Journal of Agricultural and Food Chemistry*, 2016, 64(40): 7585–7596.
- [29] YU Y D, XU S Q, HE R, et al. Application of molecular simulation methods in food science: Status and prospects[J]. *Journal of Agricultural and Food Chemistry*, 2023, 71(6): 2684–2703.
- [30] CHACKO R, JAIN D, PATWARDHAN M, et al. Data based predictive models for odor perception[J]. *Scientific Reports*, 2020, 10(1): 17136.
- [31] ENCISO M, MEFTAH N, WALKER M L, et al. Biopsy: An open-source platform for qsar/qspr analysis[J]. *PLoS One*, 2016, 11(11): e0166298.
- [32] WANG Y, ZHAO Q L, MA M Y, et al. Olfactory perception prediction model inspired by olfactory lateral inhibition and deep feature combination[J]. *Applied Intelligence*, 2023, 53(16): 19672–19684.
- [33] CAPOZZI F, BORDONI A. Foodomics: A new comprehensive approach to food and nutrition[J]. *Genes & Nutrition*, 2013, 8(1): 1–4.
- [34] PAVAGADHI S, SWARUP S. Metabolomics for evaluating flavor-associated metabolites in plant-based products[J]. *Metabolites*, 2020, 10(5): 197.
- [35] BILODEAU C, JIN W G, JAAKKOLA T, et al. Generative models for molecular discovery: Recent advances and challenges[J]. *WIREs Computational Molecular Science*, 2022, 12(5): e1608.
- [36] WALTERS W P, BARZILAY R. Applications of deep learning in molecule generation and molecular property prediction[J]. *Accounts of Chemical Research*, 2021, 54(2): 263–270.
- [37] LEYGEBER S, GROSSMANN J L, DIEZ-SIMON C, et al. Flavor profiling using comprehensive mass spectrometry analysis of metabolites in tomato soups[J]. *Metabolites*, 2022, 12(12): 1194.
- [38] FUENTES S, TONGSON E, TORRICO D D, et al. Modeling pinot noir aroma profiles based on weather and water management information using machine learning algorithms: A vertical vintage analysis using artificial intelligence[J]. *Foods*, 2019, 9(1): 33.
- [39] MILDNER-SZKUDLARZ S, JELEŃ H H. The potential of different techniques for volatile compounds analysis coupled with pca for the detection of the adulteration of olive oil with hazelnut oil[J]. *Food Chemistry*, 2008, 110(3): 751–761.
- [40] ZHOU Y M, CHEN X G, ZHU S Y, et al. Understanding the flavor signature of the rice grown in different regions of China via metabolite profiling[J]. *Journal of the Science of Food and Agriculture*, 2022, 102(7): 3010–3020.
- [41] ZHU W Y, BENKWITZ F, KILMARTIN P A. Volatile-based prediction of sauvignon blanc quality gradings with static headspace-gas chromatography-ion mobility spectrometry (shs-gc-ims) and interpretable machine learning techniques[J]. *Journal of Agricultural and Food Chemistry*, 2021, 69(10): 3255–3265.
- [42] LU L, HU Z Q, HU X Q, et al. Electronic tongue and electronic nose for food quality and safety[J]. *Food Research International*, 2022, 162: 112214.
- [43] ISHIMARU Y, INADA H, KUBOTA M, et al. Transient receptor potential family members pkd113 and pkd211 form a candidate sour taste receptor[J]. *Proceedings of the National Academy of Sciences - PNAS*, 2006, 103(33): 12569–12574.
- [44] ZHENG S Q, CHANG W P, XU W X, et al. E-sweet: A machine-learning based platform for the prediction of sweetener and its relative sweetness[J]. *Frontiers in Chemistry*, 2019, 7: 35.
- [45] GOEL A, GAJULA K, GUPTA R, et al. In-silico prediction of sweetness using structure-activity relationship models[J]. *Food Chemistry*, 2018, 253: 127–131.

- [46] TUWANI R, WADHWA S, BAGLER G. Bitter-sweet: Building machine learning models for predicting the bitter and sweet taste of small molecules [J]. *Scientific Reports*, 2019, 9(1): 7155.
- [47] OLIVERI P, CHIARA CASOLINO M, FORINA M. Chapter 2 – chemometric brains for artificial tongues [M]//TAYLOR S L. *Advances in Food and Nutrition Research*. New York: Academic Press, 2010: 57–117.
- [48] YQ LOW J, JANIN N, TRAILL R M, et al. The who, what, where, when, why and how of measuring emotional response to food. A systematic review [J]. *Food Quality and Preference*, 2022, 100: 104607.
- [49] ELALI I, M BAILARA K, SANDERS V, et al. Is wine an emotional object? Measurements of the subjective and automatic components of emotions in a wine-tasting situation[J]. *OENO One*, 2023, 57(4): 7634.
- [50] XU J X, GUO X Y, LIU M Y, et al. Self-construal priming modulates sonic seasoning[J]. *Frontiers in Psychology*, 2023, 14: 1041202.
- [51] YAMAMOTO T, MIZUTA H, UEJI K, et al. Analysis of facial expressions in response to basic taste stimuli using artificial intelligence to predict perceived hedonic ratings[J]. *PLoS One*, 2021, 16(5): e0250928.

Application of Data Mining and Modeling Techniques in Food Olfactory and Taste Perception and Emotional Cognition

Wang Bei¹, Wang Ying¹, Wang Yadong¹, Liu Shuai^{2*}, Jiang Tao^{3*}

¹*School of Food and Health, Beijing Technology and Business University, Beijing 100048*

²*School of Mathematics and Statistics, Beijing Technology and Business University, Beijing 100048*

³*Centre for Neuroscience Research, University of Burgundy, Lyon, Bron 69500, France*

Abstract With the continuous improvement of science and technology related to food olfactory and taste senses and emotional cognition, more and more instrumental analysis methods and experimental equipment are used in the research of the above fields. The diversification and comprehensiveness of detection methods and the improvement of detection accuracy are accompanied by the expansion of data scale related to flavor perception. How to obtain key information from the analysis results of food flavor instruments and the large amount of data collected in the research of consumers' emotional cognition and behavior, and establish the correlation between the data, has been paid more and more attention by researchers. Data mining and modeling technology in the field of food is to use a large amount of data obtained in the process of food production and circulation to monitor the physical and chemical changes in each link of the food industry chain in real time and accurately, and to predict the impact of these changes on consumers' sensory characteristics and emotional cognition. In the field of food smell and taste perception, data mining and modeling techniques can provide unprecedented insight and analytical power to food researchers and consumers. Based on the supervised and unsupervised data mining and modeling methods and deep learning methods of classical machine learning methods, this paper analyzes the latest application progress in the research of food sensory attributes and emotional cognition, and looks forward to the application prospect of data mining and modeling technology in the field of food olfactory and taste perception. Help the scientific and technological progress and industrial upgrading of the food industry.

Keywords data mining and modeling; machine learning; flavor perception; sensory attributes; emotional cognition