

文章编号: 1674-2869(2017)05-0496-07

近红外光谱无损检测技术中数据的分析方法概述

刘军^{1,3}, 吴梦婷^{1,3}, 谭正林², 李威^{1,3}

1. 智能机器人湖北省重点实验室(武汉工程大学), 湖北 武汉 430205;
2. 湖北经济学院烹饪与营养学系, 湖北 武汉 430205;
3. 武汉工程大学计算机科学与工程学院, 湖北 武汉 430205

摘要: 近红外光谱无损检测技术可用于品种鉴别与农产品的定性或者是定量的分析工作. 本文介绍了近红外光谱的基本原理及各类近红外光谱分析方法. 近红外光谱无损检测技术中数据分析方法是通过光谱定量分析找到光谱以及对应浓度的内在关系, 建立相应的数学模型. 这些方法主要有偏最小二乘回归、主成分分析法、BP神经网络算法、支持向量机、K最近邻分类算法和线性判别分析法等. 通过这些分析模型的对比, 研究表明: 支持向量机将是近红外光谱数据分析方法未来一个重要的研究方向.

关键词: 近红外光谱; 无损检测; 数据分析方法

中图分类号: R857.3 文献标识码: A doi: 10.3969/j.issn.1674-2869.2017.05.001

Overview of Data Analysis Methods in Near-Infrared Spectroscopy Nondestructive Testing

LIU Jun^{1,3}, WU Mengting^{1,3}, TAN Zhenglin², LI Wei^{1,3}

1. Hubei Key Laboratory of Intelligent Robot (Wuhan Institute of Technology), Wuhan 430205, China;
2. Department of Cuisine and Nutrition Education, Hubei University of Economics, Wuhan 430205, China;
3. School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

Abstract: Near-infrared spectroscopy nondestructive testing technology can be used for variety identification and the qualitative or quantitative analysis of agricultural products. The basic principle of near-infrared spectroscopy and the methods of near-infrared spectrum analysis were introduced. The data analysis methods in near-infrared nondestructive testing technology aim at finding the relationship between the spectrum and the corresponding concentration through the quantitative analysis of the spectrum, and establishing the corresponding mathematical model, which mainly include partial least squares regression, principal component analysis, back propagation artificial neural network, support vector machine (SVM), K-Nearest neighbor classification algorithm and linear discriminant analysis. The comparison result of these analytical models show that SVM method may be a future research direction in near infrared spectrum data analysis.

Keywords: near-infrared spectroscopy; nondestructive testing; data analysis methods

收稿日期: 2017-05-20

基金项目: 湖北省食品药品监督管理局项目(201610+13); 湖北省智能机器人重点实验室开放基金(HBIR 201608); 武汉工程大学研究生创新基金(CX2016063)

作者简介: 刘军, 博士, 副教授. E-mail: liujun@wit.edu.cn

引文格式: 刘军, 吴梦婷, 谭正林, 等. 近红外光谱无损检测技术中数据的分析方法概述[J]. 武汉工程大学学报, 2017, 39(5): 496-502.

LIU J, WU M T, TAN Z L, et al. Overview of data analysis methods in near-infrared spectroscopy nondestructive testing[J]. Journal of Wuhan Institute of Technology, 2017, 39(5): 496-502.

近红外光谱检测技术的成本低,对于样本无损伤、无污染、方便在线分析^[1].在食品无损检测中得到了广泛的应用.

近红外光谱技术是建立一个稳定可靠的定性或者定量模型来实现数据样本的校准,常用于定性分析^[2].定性分析中常用的方法有主成分分析法(principal component analysis, PCA)、模式识别方法等;定量分析中常用的方法主要是各种多元校正技术,如偏最小二乘回归(partial least squares, PLS)、支持向量机(support vector machine, SVM)以及神经网络算法(back propagation, BP)等.下面将对几种常用算法的原理和应用进行比较分析,以便找到最适合对小样本数据进行分类的算法.

1 近红外光谱技术

红外光谱又可以称为分子振动光谱或者转动光谱.近红外光谱主要是通过投射光谱技术和反射光谱技术获得.近红外光谱主要采集样品的含氢基团的伸缩振动的倍频和合频信息,其中包含了反映物质特性的化学成分、分子结构和状态信息^[3].由于近红外光谱技术在检测样品的过程中,不会对样品造成损伤^[1],因此近红外光谱分析技术可以被用于食品无损检测中.近红外光谱通过合适的化学计量法对于样本进行定性分析,可以将已知的样品划分为子集,用以提高校正模型的预测精度.定量分析是依据实验中得到的数据,通过相应的算法建立数学模型,并且使用模型计算出分析对象各项指标以及其数值的一种方法.与其他的检测方法相比,近红外光谱技术便于实现在线分析和典型的无损分析技术^[3].

2 各类算法原理及分析

2.1 偏最小二乘法

偏最小二乘法(partial least squares, PLS)是一种新的多元校正方法,可以建立多元的回归模型,主要是基于因子分析.偏最小二乘法的原理是首先将样本的光谱矩阵 X 和光谱浓度矩阵 Y 进行分解,同时把包含在样本中的信息引入进来,分解后提取出样本中的 m 个主因子,如公式(1)所示:

$$\begin{cases} X = TP^T + E = \sum_{k=1}^m t_k p_k^T + E \\ Y = UQ^T + F = \sum_{k=1}^m u_k q_k^T + F \end{cases} \quad (1)$$

式(1)中:测量矩阵 X 和 Y 的得分矩阵分别是 T 和 U ; t_k 和 u_k 分别表示光谱矩阵 X 的第 k 个主成

分因子和浓度矩阵的第 k 个主成分因子; p_k 和 q_k 表示主成分因子的载荷, m 为成分因子的个数.

对于 U 和 T 进行线性关联,如公式(2)所示:

$$\begin{cases} U = TB \\ B = (T^T T)^{-1} T^T Y \end{cases} \quad (2)$$

在进行回归预测时,首先求出未知样品光谱矩阵的得分 T_X ,再根据公式(3)得到组分浓度预测值,如公式(3)所示:

$$Y_X = T_X B Q \quad (3)$$

在该算法的整个运算过程中,矩阵的分解和回归分析是同时进行的,计算新的主成分时,首先交换 T 和 U ,使 X 的主成分和测量矩阵 Y 直接关联.

偏最小二乘 PLS 算法分析是通过建立光谱数据和品种分类之间的回归模型来实现的^[4].使用 PLS 分析得到的结果中,不仅可以建立更优化的回归模型,还可以同时进行主成分分析来简化数据结构,观察变量之间的相互关系等研究内容,提供更多的建模信息^[4].

所有的独立变量都可以包含在最终建立的回归模型中,PLS 可以有效解决变量多重性问题.

目前,在光谱分析中多数应用都使用的多元校正方法. Mauer 等^[5]使用近红外光谱和中红外光谱对婴儿奶粉中三聚氰胺的含量进行定量检测,使用偏最小二乘法(PLS)建模可以快速检测出三聚氰胺含量最低值.

2.2 主成分分析

主成分分析是一种线性投影方法,它主要被应用在多元统计分析中,在不损害样本的原始信息的前提下,对高维数据进行降维处理,将其映射到低维空间中. PCA 的基本思路是通过最优化方法简化样本的数据矩阵,降低维数,从原本样本的指标中,得到几个主成分,用来揭示样本中包含的信息.

计算步骤首先是对原始数据矩阵 X 进行拆分,如公式(4)所示:

$$X = TL^T \quad (4)$$

借助投影矩阵 L^T 将 X 投影到多维子空间. T 中的列向量和 L 中的列向量是相互正交的关系,如公式(5)所示:

$$\begin{cases} L_i^T L_j = 0 \\ t_i^T t_j = 0 \end{cases} \quad (i \neq j) \quad (5)$$

重建后的数据变量相互独立,各主成分按照方差顺序进行排列,第一主成分包含了数据方差的绝大部分,排名越靠前的主成分,其所包含的数据信息量越大^[6].经过计算所得到的新的横坐标是样本原来变量的线性组合.第一主成分元素可

以表示为如下形式:

$$\begin{cases} t_{11} = x_{11}l_{11} + x_{12}l_{21} + \dots + x_{1p}l_{p1} \\ t_{21} = x_{21}l_{11} + x_{22}l_{21} + \dots + x_{2p}l_{p1} \\ \dots\dots \\ t_{n1} = x_{n1}l_{11} + x_{n2}l_{21} + \dots + x_{np}l_{p1} \end{cases} \quad (6)$$

主成分分析的对象减少,从而可以有效地降低分析的工作量和误差.分析过程中只使用独立变量,这样可以达到消除噪音的目的.对数据进行降维处理可以帮助信息的提取和聚类分析.

在区分产于绍兴和嘉善的中国米酒的实验中,使用到近红外光谱技术,在区分的过程中,同时使用到主成分分析法和偏最小二乘法建立判别模型,准确率高达100%^[7].

2.3 BP神经网络算法

BP算法又称为误差反向传播算法,是人工神经网络中的一种基于误差反向传播的监督式学习算法^[8].在理论上,BP神经网络算法可以逼近任意函数,基本的结构是由很多非线性变化的单元组成,具有实现输入和输出数据高度非线性映射能力.人工神经网络是对人类大脑结构以及信息处理运作过程进行模拟之后,提出的信息处理系统.因此具有自学习和自适应的能力^[9].

BP神经网络的计算包括正向和反向两个子过程.正向传播过程中,是从输入层逐层转向传出层.如果输出和期望不同的时候,则进行反向传播,把误差沿着正向传播路径的原路径反向传播回去,经过各个神经元的修改,使得误差达到最小.

BP神经网络法的缺点是学习速度慢;在解决全局问题时,很有可能陷入局部最优化,使得训练失败;由于学习能力和泛化能力之间的矛盾,过程中容易出现“过拟合”现象^[25].

Mutlu等^[10]在预测面粉的蛋白质和水分含量等相关参数时,使用到近红外光谱进行检测和人工神经网络进行数据处理,得到的相关系数分别为0.952,0.948,0.933,0.920,0.917和0.832,有效的证明了近红外光谱结合神经网络算法可以很好地解决此类问题.

2.4 支持向量机算法

支持向量机算法的优化原则是采用以结构化风险最小化取代我们常用的经验风险最小化;基本思想是利用核函数代替向高维空间的非线性映射,对于样本,若是非线性可再分,则构建一个或多个甚至无限多个高维的超平面,在经过变换后的高维空间,利用这个超平面来对样本点进行分

类.样本点的与分类边界的距离越大越好,这个距离被称之为间隔(Margin),这些样本点被称之为支持向量.SVM的目标就是要找出使间隔最大的分类超平面^[11].

给定的样本集,步骤一般如下:首先要将样本集分为训练集和测试集,训练集是 $(x_1, y_1)(x_2, y_2) \dots (x_i, y_i)$,其中横坐标 $x_i \in R^n$ 为输入变量的值, $y_i \in R^n$ 为相应的输出值.通过对训练集进行机器学习,寻求最优化的模式 $M(x)$,要求这个模式不仅满足训练集输入和输出的对应关系,而且对于测试集的样本输入变量的值进行预测,同样能得到效果较好的输出.

在支持向量机算法中,使用不同的核函数会产生不同的算法^[25],因为SVM具有优秀的泛化能力,所以是一种适合小样本的学习方法.在SVM进行分类的时候,起到决定性作用的是支持向量.计算的复杂性取决于支持向量的数目,在实验的过程中只需要抓住关键的样本.SVM在运算的过程中很少涉及到概率测度及大数定律等规律,而且支持向量法整体具有较好的“鲁棒”性^[12].

在数据挖掘的实际应用中,一般要解决多类的分类问题^[13].如:将SVM与粗集理论结合,形成一种优势互补的多类问题的组合分类器^[14].

攀玉霞等^[15]在研究猪肉肉糜样本,采用支持向量机回归方法建立相应参数的模型模型.在这个研究中,实验者以鱼糜为研究对象,使用支持向量机(SVM)建立其水分和蛋白质的定量分析模型.

2.5 最小二乘支持向量机算法

最小二乘支持向量机(least square support vector machines, LS-SVM)是支持向量机的一种类型^[16],是在经典SVM的基础上改进得来的.LS-SVM既可以进行线性,也可以进行非线性的多元建模,支持少量的训练样本中高维特征空间的学习,是解决多元建模的一种快速方法^[17].

最小二乘支持向量机具有建模速度快、优化参数少、泛化能力强等优点,因此被广泛应用于近红外光谱技术检测的定量分析中^[17].最小二乘支持向量机同时具有很好的非线性处理能力,可以用于快速建立多元模型^[18].

罗微等^[4]在对白菜种子不同品种的鉴别的实验中,建立最小二乘支持向量机(partial least squares discrimination analysis, PLS-DA)和偏最小二乘判别的模型.

2.6 K最近邻分类算法

在K最近邻分类算法(K-nearest neighbor,

KNN)算法中每个样本都可以用它最接近的 k 个邻居来代表. KNN算法的核心思想是在特征空间中, 如果以一个样本为中心, 它周围最相邻的 k 个样本, 它们大多数属于某一个类别, 则这个样本也属于这个类别, 并且具有这个类别上其他 k 个样本的特性.

如图1所示, 有三类不同样本, 分别以方框、空心圆点和实心圆点表示. 其中以“+”标记出的样本, 需要分析出它所属的类别. 在此图中以该样本作为圆心, 画出两个圆, 如果以小圆为界线, 只有2个实心圆点落在小圆的范围内, 此时按照KNN原理应该将样本划分为实心圆点那一类. 如果以大圆为界线, 共有10个点落在了大圆的范围内, 其中有3个方框类, 有8个实心圆点类, 按照KNN理论, 也应该被划为实心圆点这一类. 这说明 K 值的大小对判别的结果是有一定的影响的. 如果 $K > 1$, 则这 K 个最近邻样本不一定都属于这一类. 如果 $K = 1$, 很自然这一个最近邻样本属于哪类, 那么未知样本就属于此类^[19].

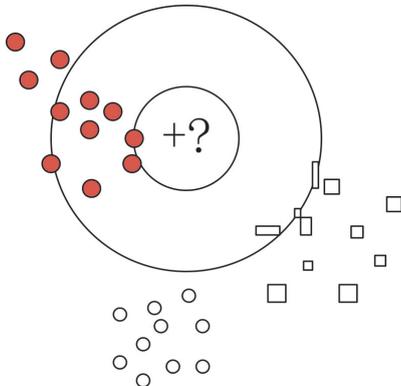


图1 K最近邻算法原理示意图

Fig. 1 Schematic diagram of K-Nearest Neighbor algorithm

KNN分类算法的最大优点是其适合于属性较多或者数据量很大的问题. 在进行数据处理的过程中, 不需要提前设计分类器对训练样本进行分类, 而是可以直接进行分类, 确定每个类的类标识号, 不需要估计参数, 无需训练; 在解决多分类的问题上KNN比SVM表现要好.

KNN在实际的应用中存在下面两个问题: 当对估计参数没有相应的限制时, 训练集的数目又比较大, 这时寻找最近邻所花费的时间较长. 其次, 对于观测集的增长速度有较高要求.

K最近邻分类算法被广泛应用于模式识别、数据挖掘、后验概率统计、相似性分析、计算机视觉和生物信息学等各种人工智能领域^[19]. 倪力军等在鉴别真奶和假奶样品中各成分的实验中, 使用

近红外光谱对所有的样品进行检测, 并使用改进的 K 最近邻结点算法和支持向量机法分别建立判别各类掺假物质的模型^[20].

2.7 线性判别分析法

主要用来判断样品的类别, 即可以对样品中的成分进行定性分析, 线性判别分析法 (linear discriminant analysis, LDA) 在医学诊断、气象学、市场预测、经济学和地质勘探等领域中均已起着重要作用^[20].

LDA算法的基本思想是通过特征向量将已经分组的数据向低维的方向投影, 使得同一个组的数据关系更为紧密, 不同组的之间尽可能的分开. LDA的目标是要根据样本中的 n 个指标变量 $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ 建立一个最优分类判别函数, 判别函数的建立最终是寻找一个矢量 $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_n)^T$, 那建立一个如下的判别模型, 如公式(9)所示:

$$f(\boldsymbol{\omega}, \mathbf{x}) = (\boldsymbol{\omega}^T \cdot \mathbf{x}) + b \quad (8)$$

矢量 $\boldsymbol{\omega}$ 是个常数, 称作阈值权, 相应的决策规则则可表示为, 如公式(10)所示:

$$\begin{cases} \mathbf{x}_{\text{new}} \in \beta_1, f(\boldsymbol{\omega}, \mathbf{x}_{\text{new}}) > 0 \\ \mathbf{x}_{\text{new}} \in \beta_2, f(\boldsymbol{\omega}, \mathbf{x}_{\text{new}}) < 0 \end{cases} \quad (9)$$

$f(\boldsymbol{\omega}, \mathbf{x}) = 0$ 是相应的决策面方程, 如公式(11)所示:

$$\boldsymbol{\omega}^T \mathbf{x} + \boldsymbol{\omega}_0 = 0 \quad (10)$$

$f(\boldsymbol{\omega}, \mathbf{x})$ 也就是该决策面到 d 维空间中任一点 $\boldsymbol{\omega}$ 距离的代数度量, 根据其到该面距离的正负号, 该决策平面将样本分类, 并且确定出相应的类别. $\boldsymbol{\omega}(\boldsymbol{\omega}_0 \neq 0)$ 表示的是决策面在特征空间中的位置, 当其值为0时, 表示这个决策面过原点.

线性判别分析算法是从训练集中训练出判别函数, 当得到新的样品时, 根据相关的判别准则对其与判别函数之间的相似程度进行比较. 由于线性判别分析算法在应用的时候需要样本标签, 属于有监督学习降维.

2.8 连续投影算法

连续投影算法 (successive projections algorithm, SPA) 是在线性空间中, 使共线性达到最小化的一种前向变量选择算法^[21].

设定标集中样品个数为 n , 包含波长数为 m , 组成一个光谱矩阵 $\mathbf{X}_{n \times m}$, SPA从任选一个波长作为起始波长点, 循环进行搜索, 搜索投影中没有被引入波长组合的部分, 每次都最大的投影方向不断地引入, 直到循环 $w (w < n - 1)$ 次以后, 就可以

形成一个波长链,使得此链中的任意相邻两个波长之间线性关系最小^[22].算法如下:

初始化波长:令 $n=1$,任取一个波长 x_j 作为算法循环的起始点,记为 $x_{m(0)}$.把剩下的没有选择到波长链中的所有波长,记为一个集合 $s=\{j, 1 \leq j \leq m, j \notin \{m(0), m(1), \dots, m(n-1)\}\}$

计算 s 中波长与 x_j 的投影向量,如公式(12)所示:

$$P_{x_j} = x_j - (x_j^T - x_{m(n-1)}^T)x_{m(n-1)}(x_{m(n-1)}^T - x_{m(n-1)}^T)^{-1} \quad (12)$$

选择其中最大的投影,并且记录下相应序号作为入选的波长,如公式(13)所示:

$$m(n) = \arg(\max \|P_{x_j}\|, x_j = P_{x_j}, j \in s) \quad (13)$$

当 $n=w$ 时停止循环,否则返回到第二步不断地做投影优选波长.这样总共得到 $w \times m$ 个模型.

连续投影算法的优点是:最大程度的降低了变量之间的共线性,提取最低限度的冗余信息变量组可以减少建立模型所需要变量的个数,提高建模效率和速度^[23].

浙江大学^[21]利用鲜肉的近红外光谱中少量特征波长对其pH值进行预测,可以大幅度降低模型复杂性和计算量,通过连续投影算法(SPA)选择特征波长,并且得到相应的多元线性回归模型.

2.9 算法对比分析

偏最小二乘法是一种新型的多元统计数据分析方法,通常被用于曲线拟合.主成分分析法是一

种降维的统计方法,使用各主成分代替原有的变量进行回归分析.偏最小二乘法在建模的过程中包含了主成分分析的特点,因此在这两种算法中,偏最小二乘法的性能明显优于主成分分析法.吴远远等人在老陈醋可溶性固形物定量分析的实验当中,分别采用PCR和PLS建立SSC的定量分析模型,结果表明,采用5点平滑预处理后,利用PLS建立的老陈醋SSC的定量分析模型最优^[24].

SVM最基本的应用是分类,求解最优分类面,然后用于分类.支持向量机可以较好的解决小样本的分类问题,泛化功能好.张海云等人在对生鲜猪肉水分含量进行无损检测的实验中,对比了PLS和SVM的性能.实验结果分析后得到,采用SVM方法建模要明显好于PLS建立回归模型法^[13].

LDA是一种有效的特征抽取方法,能够使投影后模式样本的类间三部矩阵最大,同时保证类内散布矩阵最小.

BP和SVM是分类领域中的两种重要的方法.神经网络是模拟人脑神经系统的数学模型,具有高度并行性、较强的自学习自适应和联想记忆功能特点^[25].神经网络算法具有强大的非线性处理能力,其中使用最为广泛的就是前馈神经网络^[26].SVM算法在小样本上表现明显优于BP神经网络.KNN分类算法则在多分类问题上表现比SVM要好.算法的优缺点对比,如表1所示.

表1 算法对比

Tab. 1 Algorithm comparison

算法名称 algorithm name	优点 advantage	缺点 disadvantage
偏最小二乘法(PLS)	建立回归模型的同时可以进行主成分分析简化数据.预测性能较好.在近红外光谱检测中最为常用	仅在少数情况下,使用具有优势(例如自变量离差信息利用率较高)
主成分分析(PCA)	有助于信息的提取和用于聚类分析,能够有效地降低误差和消除噪音	当样本中的部分有用变量的相关性很小时,容易遗漏掉这些变量,降低预测模型的可靠性
BP神经网络算法	具有很强非线性映射能力和自学习能力	学习速度慢,可能会陷入局部最优化,容易出现“过拟合”现象
支持向量机算法(SVM)	具有优秀的泛化能力,适合小样本的学习方法.具有较好的“鲁棒性”	对大规模训练样本难以实施,解决多分类问题存在困难
K最近邻分类算法(KNN)	适合于属性较多或者数据量很大的问题.不需要提前设计分类器对训练样本进行分类,在解决多分类问题上表现较好	训练集数目较大,而且对估计参数没有限制时,花费时间较长,对于观测集的增长速度有较高要求
线性判别分析法(LDA)	属于有监督学习降维	不适合对非高斯分布样本进行降维,可能会过度拟合数据
连续投影算法(SPA)	建立模型所需变量个数较少,可以有效将变量之间的共线性降到最低	会损失部分样本精度信息

3 结 语

由于所做实验数据是小样本,因此在选择分类算法时会选择SVM算法.今后的研究方向主要是使用改进的支持向量机算法对近红外光谱无损检测的数据进行分析.与其他分析方法相比,SVM能够很好地克服其他算法中出现的训练时间长,结果不准确等特点.其中LS-SVM可以降低支持向量数目,减少变量存储空间和计算量,可以用来提高实际应用中定量分析的效率.

近年来,支持向量机结合近红外光谱技术,被用于煤种分类、山茶油掺假鉴别、识别树种和无损检测的实验中,均得到了相当不错的结果.

参考文献:

- [1] 黄瑞娟. 红外光谱技术在食品检测中的应用[J]. 现代测量与实验室管理, 2015(1):9-14.
HUANG R J. Application of infrared spectroscopy in food detection [J]. Advanced Measurement and Laboratory Management, 2015(1):9-14.
- [2] MODAREST F, ARAGHINEJAD S. A comparative assessment of support vector machines, probabilistic neural networks, and k -nearest neighbor algorithms for water quality classification [J]. Water Resources Management, 2014, 28(12): 4095-4111.
- [3] 谭正林, 金国梁, 吴梦婷, 等. 奶油中人工色素检测方法概述[J]. 食品安全质量检测学报, 2017, 8(2): 468-474.
TAN Z L, JIN G L, WU M T, et al. Overview on detection methods of artificial pigments in cream [J]. Journal of Food Safety & Quality, 2017, 8 (2) : 468-474.
- [4] 章海亮, 罗微, 杜焱喆. PCA和SPA的近红外光谱识别白菜种子品种研究[J]. 光谱学与光谱分析, 2016, 36(11):3536-3541.
ZHANG H L, LUO W, DU Y Z. Discrimination of varieties of cabbage with near infrared spectra based on PCA and SPA[J]. Spectroscopy and Spectral Analysis, 2016, 36(11): 3536-3541.
- [5] MAUER L J, CHERNYSHOVA A A, HIATT A N, et al. Melamine detection in infant formula powder using near-and mid-infrared spectroscopy [J]. Journal of Agricultural and Food Chemistry. 2009, 57 (10) : 3974-3980.
- [6] 阎吉祥, 王茜蓓, 黄志文, 等. 基于主成分分析和人工神经网络的激光诱导击穿光谱塑料分类识别方法研究[J]. 光谱学与光谱分析, 2012(12):3179-3182.
YAN J X, WANG Q Q, HUANG Z W, et al. Classification of plastics with laser-induced breakdown spectroscopy based on principal component analysis and artificial neural network model [J]. Spectroscopy and Spectral Analysis, 2012(12): 3179-3182.
- [7] 刘洪林. 基于近红外光谱技术(NIRS)对工夫红茶含水率、游离态氨基酸、茶多酚品质成分评价研究[J]. 食品工业科技, 2016, 37(12):67-70.
LIU H L. Research to moisture content, free form amino acids, polyphenols quality ingredients of Kungfu black tea by near infrared spectroscopy [J]. Science and Technology of Food Industry, 2016, 37(12):67-70.
- [8] 徐可欣, 苗静, 曹玉珍, 等. 基于二维相关近红外谱参数化及BP神经网络的掺杂牛奶鉴别[J]. 光谱学与光谱分析, 2013, 33(11):3032-3035.
XU K X, MIAO J, CAO Y Z, et al. Identification of adulterated milk based on two-dimensional correlation near-infrared spectra parameterization and BP neural network [J]. Spectroscopy and Spectral Analysis, 2013, 33(11):3032-3035.
- [9] WANG Y S, LI Q Y. Application study on assistive movement training using BP neural network [C]// Internation Concrence on Computational Science and Engineering. Paris: Atlantis Press, zeger karssen, 2015:274-277.
- [10] MUTLU A C, BOYACI I H, GENIS H E, et al. Prediction of wheat quality parameters using near-infrared spectroscopy and artificial neural networks[J]. European Food Research & Technology, 2011, 233(2): 267-274.
- [11] 白京, 彭彦昆, 王文秀. 基于可见近红外光谱玉米种子活力的无损检测方法[J]. 食品安全质量检测学报, 2016, 7(11):4472-4477.
BAI J, PENG Y K, WANG W X. Discrimination of vitality of maize seeds based on near visible infrared spectroscopy [J]. Journal of Food Safety and Quality, 2016, 7(11):4472-4477.
- [12] 王小燕, 王锡昌, 刘源, 等. 基于SVM算法的近红外光谱技术在鱼糜水分和蛋白质检测中的应用[J]. 光谱学与光谱分析, 2012, 32(9):2418-2421.
WANG X Y, WANG X C, LIU Y, et al. Application of near infrared spectroscopy technique based on support vector machine in the measurement of moisture and protein contents in surimi [J]. Spetroscopy and Spectral Analysis, 2012, 32(9):2418-2421.
- [13] 张海云, 彭彦昆, 王伟, 等. 基于光谱技术和支持向量机的生鲜猪肉水分含量快速无损检测[J]. 光谱学与光谱分析, 2012, 32(10):2794-2798.
ZHANG H Y, PENG Y K, WANG W, et al. Rapid nondestructive detection of water content in fresh pork based on spectroscopy technique combined with

- support vector machine [J]. *Spectroscopy and Spectral Analysis*, 2012, 32(10): 2794–2798.
- [14] 吴习宇, 赵国华, 祝诗平. 近红外光谱分析技术在肉类产品检测中的应用研究进展[J]. *食品工业科技*, 2014, 35(1): 371–374, 380.
WU X Y, ZHAO G H, ZHU S P. Study on the application of near infrared spectroscopy in the meat quality evaluation [J]. *Science and Technology of Food Industry*, 2014, 35(1): 371–374, 380.
- [15] 陈彬, 刘阁, 张贤明. 连续投影算法的润滑油中含水量的近红外光谱分析[J]. *红外与激光工程*, 2013(12): 3168–3174.
CHEN B, LIU G, ZHANG X M. Analysis on near infrared spectroscopy of water content in lubricating oil using successive projections algorithm [J]. *Infrared and Laser Engineering*, 2013(12): 3168–3174.
- [16] ZHANG N, SHETTY D. An effective LS-SVM based approach for surface roughness prediction in machined surfaces [J]. *Neurocomputing*, 2016, 198: 35–39.
- [17] 乔延江, 徐冰, 王星, 等. 基于遗传算法的多目标最小二乘支持向量机在近红外多组分定量分析中的应用[J]. *光谱学与光谱分析*, 2014(3): 638–642.
QIAO Y J, XU B, WANG X, et al. Genetic algorithm based multi-objective least square support vector machine for simultaneous determination of multiple components by near infrared spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2014 (3) : 638–642.
- [18] JI J, ZHANG C S, GUI Y L, et al. New observations on the application of LS-SVM in slope system reliability analysis [J]. *Journal of Computing in Civil Engineering*, 2016(10): 601–602.
- [19] 刘应东, 牛惠民. 基于K-最近邻图的小样本KNN分类算法[J]. *计算机工程*, 2011, 37(9): 198–200.
LIU Y D, NIU H M. KNN classification algorithm based on K-nearest neighbor graph for small sample [J]. *Computer Engineering*, 2011, 37(9): 198–200.
- [20] 倪力军, 钟霖, 张鑫, 等. 近红外光谱结合非线性模式识别方法进行牛奶中掺假物质的判别[J]. *光谱学与光谱分析*, 2014(10): 2673–2678.
NI L J, ZHONG L, ZHANG X, et al. Identification of adulterants in adulterated milks by near infrared spectroscopy combined with non-linear pattern recognition methods [J]. *Spectroscopy and Spectral Analysis*, 2014 (10): 2673–2678.
- [21] 伍学千, 廖宜涛, 樊玉霞, 等. 连续投影算法在猪肉PH值无损检测中的应用[J]. *农业工程学报*, 2010, 26(增刊): 379–383.
WU X Q, LIAO Y T, FAN Y X, et al. Application of successive projections algorithm to nondestructive determination of pork PH value [J]. *Transactions of the Chinese Society of Agricultural Engineering* 2010, 26 (Suppl.): 379–383.
- [22] CHENG J H, SUN D W, PU H, et al. Combining the genetic algorithm and successive projection algorithm for the selection of feature wavelengths to evaluate exudative characteristics in frozen-thawed fish muscle [J]. *Food Chemistry*, 2016(197): 855–863.
- [23] 应义斌, 介邓飞, 谢丽娟, 等. 近红外光谱变量筛选提高西瓜糖度预测模型精度[J]. *农业工程学报*, 2013, 29(12): 264–270.
YING Y B, JIE D F, XIE L J, et al. Improving accuracy of prediction model for soluble solids content of watermelon by variable selection based on near-infrared spectroscopy [J]. *Transaction of the Chinese Society of Agricultural Engineering*, 2013, 29 (12): 264–270.
- [24] 王福杰, 吴远远, 陆辉山, 等. 基于近红外光谱技术的老陈醋可溶性固形物定量分析[J]. *中国酿造*, 2016(8): 69–72.
WANG F J, WU Y Y, LU H S, et al. Quantitative analysis of soluble solids content in mature vinegar based on near infrared spectroscopy technology [J]. *China Brewing*, 2016(8): 69–72.
- [25] 孙剑伟, 王宏涛. 基于BP神经网络和SVM的分类方法研究[J]. *软件*, 2015, 36(11): 96–99.
SUN J W, WANG H T. Research on the classification method based on BP neural network and SVM [J]. *Software*, 2015, 36(11): 96–99.
- [26] 冯先成, 李寒, 周密, 等. 基于前馈神经网络的智慧城市空巢老人识别[J]. *武汉工程大学学报*, 2015, 37(10): 36–39.
FENG X C, LI H, ZHOU M, et al. Recognition of empty-nest elders in intelligent city based on feedforward neural network [J]. *Journal of Wuhan Institute of Technology*, 2015, 37(10): 36–39.