

基于介电特性及 ANN 的油桃糖度无损检测方法

商亮¹, 谷静思², 郭文川^{1*}

(1. 西北农林科技大学机械与电子工程学院, 杨凌 712100; 2. 西北农林科技大学食品科学与工程学院, 杨凌 712100)

摘要: 为了探索利用果品的介电特性无损预测内部品质的可能性, 该文采用矢量网络分析仪测量了 10 d 贮藏期间, 300 个 99-1 油桃在 20~4 500 MHz 频率下的相对介电常数和介电损耗因子, 以糖度作为内部品质指标, 基于 x - y 共生距离的样本划分法确定了含 243 个样本的校正集和 57 个样本的预测集; 建立了预测油桃糖度的偏最小二乘、支持向量机及极限学习机模型, 并综合比较了采用全频谱以及利用无信息变量消除法和连续投影算法分别提取的特征变量作为各模型输入变量时, 对各模型拟合效果的影响。结果表明: 连续投影算法结合极限学习机预测效果最好(预测相关系数为 0.887, 预测均方根误差为 0.782); 与全频谱和无信息变量消除法相比, 连续投影算法在简化模型及提高模型稳定性方面性能良好。该研究结果表明, 基于油桃介电特性无损检测糖度是可行的, 可为应用介电特性无损检测果品的内部品质指标提供了一种新方法。

关键词: 无损检测, 介电特性, 模型, 油桃, 可溶性固形物含量, 人工神经网络, 支持向量机, 极限学习机

doi: 10.3969/j.issn.1002-6819.2013.17.033

中图分类号: S183

文献标志码: A

文章编号: 1002-6819(2013)-17-0257-08

商亮, 谷静思, 郭文川. 基于介电特性及 ANN 的油桃糖度无损检测方法[J]. 农业工程学报, 2013, 29(17): 257-264.

Shang Liang, Gu Jingsi, Guo Wenchuan. Non-destructively detecting sugar content of nectarines based on dielectric properties and ANN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(17): 257-264. (in Chinese with English abstract)

0 引言

糖度是衡量水果品质的主要指标。糖度的主要成分是可溶性固形物含量(soluble solids content, SSC), 因此常用可溶性固形物含量评价糖度。通常用阿贝折射仪、折光仪或数字式折射仪测量果品果汁的可溶性固形物含量。这种方法虽然准确, 但因破坏样品, 只适合于抽样检验, 无法用于在线检测每个样品的糖度。因此, 寻找无损检测果品糖度的方法一直是国内外工作者的研究内容。

介电特性是指物质在外加电场中表现出的电学性质, 是电介质的固有性质^[1-2], 是物质内部组织结构和组成成分的宏观表现。果品是一种电介质, 在基于介电特性检测其内部品质方面已经取得了一定的研究进展^[3-5]。刘亚平等研究了葡萄冷藏期间

电学参数与质地的变化规律, 并发现在 0.1 kHz 下电参数与其内部品质具有较强相关性^[6]; Soltani 等通过香蕉的介电参数评估成熟度, 发现在 100 kHz 下成熟度与介电参数有很好的相关性(决定系数为 0.94)^[7]; Nelson 等对西瓜^[8]、蜜瓜^[9]的内部品质与介电参数的关系进行了研究, 并发现密瓜果肉和果汁的 ϵ'/SSC 与 ϵ''/SSC (ϵ' 和 ϵ'' 分别表示相对介电常数和介电损耗因子) 具有很好的线性关系, 但在西瓜中并没有发现该规律; 郭文川等在对桃^[10]和苹果^[11]的介电参数与可溶性固形物含量关系的研究中也没有发现此规律。因此, 寻找表达介电特性与内部品质关系的模型就成为果品品质无损检测研究领域的难题。

随着人工神经网络技术(artificial neural network, ANN)的发展, 误差反向传播网络、径向基网络、支持向量机网络、极限学习机等机器学习模型以其学习能力强, 预测精度高, 建模效果稳定等优点已被广泛应用于近红外光谱数据的分析中^[12-14]。为了探索机器学习模型在基于果品的介电频谱无损检测果品糖度方面的可行性, 本文以 99-1 油桃为对象, 以糖度为检测指标, 根据 20~4 500 MHz 间油桃的介电频谱, 应用化学计量学和人工神经网络技术, 用全频谱、经无信息变量消除和连续投影算法分别提取的特征变量, 分别建立预测油桃糖度的偏最小二乘回

收稿日期: 2013-03-30 修订日期: 2013-08-04

基金项目: 国家自然科学基金项目(31171720); 中央高校基本科研业务费专项资金项目(ZD2012017)

作者简介: 商亮(1989-), 男, 辽宁铁岭人, 主要从事于农产品品质无损检测技术研究。杨凌 西北农林科技大学机械与电子工程学院, 712100。Email: yuliang0225@sina.com

*通信作者: 郭文川(1969-), 女, 陕西临潼人, 博士, 教授, 中国农业工程学会会员(E041200348S), 主要从事农产品和食品品质无损检测技术研究。杨凌 西北农林科技大学机械与电子工程学院, 712100。

Email: guowenchuan69@126.com

归、支持向量机及极限学习机模型,通过比较模型的复杂性和精度,确定基于介电频谱无损检测油桃糖度的最佳方法。

1 材料与方法

1.1 试验材料

本研究所用的 99-1 油桃于 2012 年 6 月 17 日采摘于陕西杨凌某农户的果园中。采后当天将每 20~25 个样品装于聚乙烯塑料袋中,密封保存于 2℃ 的冷藏室中。每隔 1 天取样 1 次,每次随机选取 45~60 个无缺陷样品。共取样 6 次,300 个样品。

1.2 试验仪器及处理软件

E5071C 型网络分析仪、85070E 末端开路同轴探头和 85070 软件(Agilent Technology, 马来西亚槟城); PR101 α 型数字折射仪(Atago Co. Ltd., 日本东京); 家用压蒜器; 数据处理软件 Matlab 7.14.0.739 (R2012a, Math Works, 美国马萨诸塞州); 数据处理软件 Unscrambler v9.8 (CAMO, 挪威奥斯陆)。

1.3 测试步骤

试验前将油桃从冷藏室中取出,用自来水清洗干净、擦干后置于室温(24 \pm 2)℃下 5 h 左右,以使样品温度回升至室温。预热 E8071C 网络分析仪 1 h 后,分别用开路、短路及 50 Ω 负载对其校准;启动 85070 软件,并设定测量频率范围为 20~4 500 MHz,在对数坐标下等间距选取 101 个频率。然后用开路(即空气)、短路子及 25℃去离子水校准 85070E 探头。测量 25℃去离子水的介电参数,并与参考数据比较以判断是否需要重新校准。详细的校准过程可参考文献^[15-16]。

将一个完整无损的油桃样品横放在小型支架上,提升支架使油桃表皮与垂直向下的探头紧密接触,测量该接触处无损油桃在 20~4 500 MHz 间的介电特性,随后相继间隔约 90°,选取另外 3 点测量。4 个点处介电参数的平均值作为该油桃样品的测定结果。随后用家用削皮器削去测量点处的果皮,取出适量果肉,用家用压蒜器压汁,然后用 PR101 α 数字折射仪测量果汁的可溶性固形物含量,每点重复 3 次,4 点 12 次重复的平均值作为样品的糖度。

本研究以 ϵ' 的 101 个值作为样本的前 101 个变量, ϵ'' 的 101 个值为样本的后 101 个变量,每个样本的总变量数为 202 个。

1.4 数据分析及处理方法

1.4.1 样本划分方法

本研究采用的样本划分方法为基于 x - y 共生距离的样本划分方法(sample set partitioning based on joint x - y distances, SPXY)。SPXY 法是一种比较

有效的样本划分方法,已经广泛运用于近红外光谱的样本划分中^[17]。它是经典的 Kennard-Stone 算法为基础,综合考虑了 x 变量和 y 变量的欧氏距离,进而完成样本集的划分工作^[18]。其最大优势在于能够有效的覆盖多维空间,划分获得的校正集具有较强的代表性,可大大提高模型的预测效果。

1.4.2 介电谱预处理方法

1) 连续投影算法(successive projection algorithm, SPA)在 SPA 中,首先选择一个初始频率,每次迭代后增加一个频率,直到频率数达到设定数后停止运算^[19]。通过多次迭代,选择出冗余信息最少的变量组,进而解决信息重叠,共线性等问题。经该算法处理后的数据去除了大量的重复信息,选择的信息更具有代表性^[20-21]。

2) 无信息变量消除法(uninformative variables elimination, UVE) UVE 是基于偏最小二乘回归系数选择变量,最终达到降维的目的^[22]。在筛选变量时, UVE 综合考虑了噪声和目标浓度信息,因此比较直观实用。但经该算法筛选的变量维数依然很大,因此其效果受到了一定限制^[23]。

1.4.3 建模方法

1) 偏最小二乘(partial least squares, PLS) PLS 是一种经典的线性拟合方法,它通过最小化偏差平方和进而实现对曲线的拟合。

2) 支持向量回归(support vector regression, SVR) SVR 是基于支持向量机网络理论发展起来的一种非线性拟合方法,有较强的预测和泛化能力^[24]。支持向量机网络是基于风险最小化思想的机器学习算法,通过建立超平面实现正例与反例之间的隔离最大化,从而达到分类的目的^[25]。

3) 极限学习机(extreme learning machine, ELM) ELM 是一种针对单隐含层前馈网络的新算法,其克服了传统学习方法存在的训练速度慢、容易陷入局部极小及对学习率敏感等缺点,因此具有学习速度快、泛化性能好等特点^[26]。ELM 网络的输入层与隐层间的连接权值和神经元间的阈值都是随机确定的,因此只需设置各层神经元的个数便可以经校正集训练后获得最优解,因此在许多领域得到了广泛应用^[27-28]。

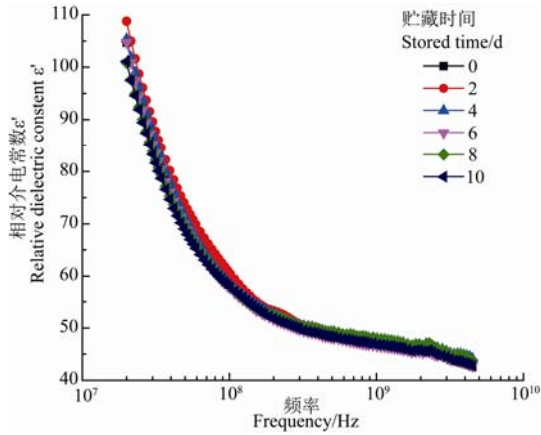
1.5 模型的评判

以校正集样品的校正相关系数(R_c)和校正均方根误差(root mean square error of calibration set, $RMSEC$)以及预测集样品的预测相关系数(R_p)和预测均方根误差(root mean square error of predication set, $RMSEP$)作为评价模型性能的指标。所建模型的 R_c 和 R_p 越接近 1, $RMSEC$ 和 $RMSEP$ 越小,则模型精度越高。

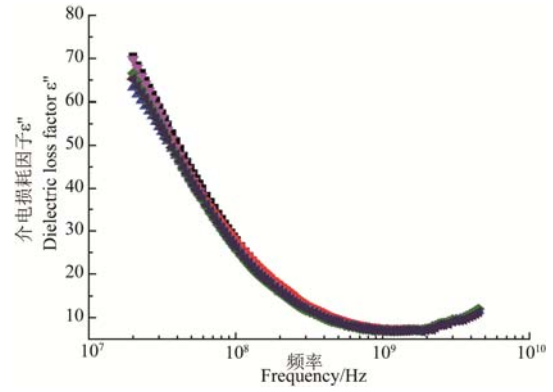
2 结果与分析

2.1 测量频率对贮藏期间油桃介电参数影响的方差分析

10 d 的贮藏期间油桃的 ϵ' 和 ϵ'' 随频率变化的关系如图 1 所示。



a. 频率对 ϵ' 的影响
a. Influence of frequency on ϵ'



b. 频率对 ϵ'' 的影响
b. Influence of frequency on ϵ''

图 1 频率对采后不同贮藏时间油桃的 ϵ' 和 ϵ'' 的影响
Fig.1 Influence of frequency on ϵ' and ϵ'' of postharvest nectarines at different storage days

图 1 表明, ϵ' 随频率的增大而单调减小; 而 ϵ'' 先减小后增大。贮藏时间和频率对介电参数影响的方差分析见表 1。

表 1 贮藏时间和频率对油桃介电参数影响的显著性分析

Table 1 Significance analysis on dielectric parameters influenced by storage and frequency

方差来源 Source of variation		平方和 Sum of square	自由度 Degree of freedom	均方 Mean square	F 值 F value	P 值 P value	显著性 Significance
相对介电常数 ϵ' Relative dielectric constant ϵ'	频率 Frequency	130444.15	100	1304.44	1691.86	<0.01	**
	贮藏时间 Stored time	339.16	5	67.83	87.98	<0.01	**
	误差 Error	385.51	500	0.77			
	合计 Total	131168.82	605				
介电损耗因子 ϵ'' Dielectric loss factor ϵ''	频率 Frequency	185977.01	100	1859.77	3680.89	<0.01	**
	贮藏时间 Stored time	170.19	5	34.04	67.37	<0.01	**
	误差 Error	252.63	500	0.51			
	合计 Total	186399.83	605				

注: **: $P < 0.01$, 表示极显著。

Note: ** means the item is significant at $P < 0.01$.

表 1 说明, 在 0.01 的显著水平上频率及贮藏时间均对介电参数有显著性的影响。因此, 以不同频率和不同贮藏时间下油桃的介电参数作为后续油桃糖度模型的输入数据是可行的。

2.2 校正集与预测集的划分

机器学习模型可以从复杂的原始数据中提取相关信息进而建立准确可靠的数学模型。如果可以在样本集中选取具有代表性的样本作为校正集, 则会使模型的预测效果及稳定性大幅提升。为此, 本研究选取了目前划分效果较好的 SPXY 法划分样本, 校正集与预测集的样本比例为 4:1, 最终划分结果见表 2。

由表 2 可以看出, 不管是校正集还是预测集, 贮藏对样本 SSC 平均值的影响规律不明显, 但并不能说明个体的介电参数和 SSC 之间没有关系。

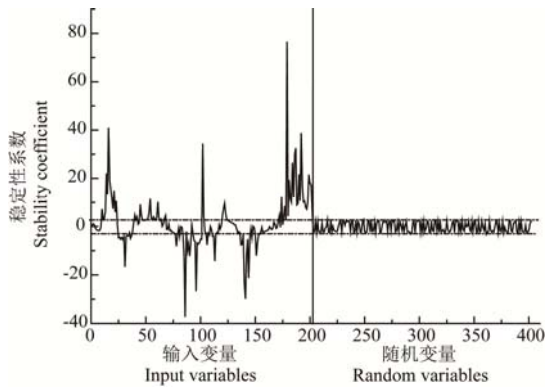
2.3 应用无信息变量消除法筛选变量

将 UVE 中随机产生的变量数设置为输入变量的维数 202(101 个频率点下的 ϵ' 和 ϵ'')。本研究选取主成分数为 15, UVE 选择结果见图 2。

竖线左侧为 202 个输入变量的稳定性曲线, 右侧为 202 个随机变量的稳定性曲线。图中 2 条阈值线之间的变量认为是无信息变量, 予以剔除。最后确定了 108 个变量作为输入变量, 其中有 55 个 ϵ' 和 53 个 ϵ'' 。选择结果如图 3 所示。

表 2 校正集和预测集样本 SSC 的统计结果
Table 2 Statistic results of SSC in calibration and predication sets

贮藏时间 Stored time/d	校正集 Calibration set					预测集 Predication set				
	样本数 Sample number	平均值 Mean/%	最大值 Max/%	最小值 Min/%	标准偏差 Standard deviation /%	样本数 Sample number	平均值 Mean/%	最大值 Max/%	最小值 Min/%	标准偏差 Standard deviation /%
0	41	11.7	16.2	6.7	2.0	10	12.5	14.9	9.0	1.7
2	40	11.4	15.2	8.3	1.5	10	11.6	14.4	10.1	1.2
4	40	10.8	16.0	7.1	2.1	9	11.4	15.1	9.1	1.6
6	46	11.9	15.9	8.4	1.8	10	12.2	15.2	9.9	1.6
8	36	11.6	16.5	7.4	2.1	9	11.9	13.3	9.8	1.3
10	40	11.0	14.7	7.4	1.8	9	11.4	13.4	9.9	1.1
合计 Total	243	11.4	16.5	6.7	1.9	57	11.8	15.2	9.0	1.5



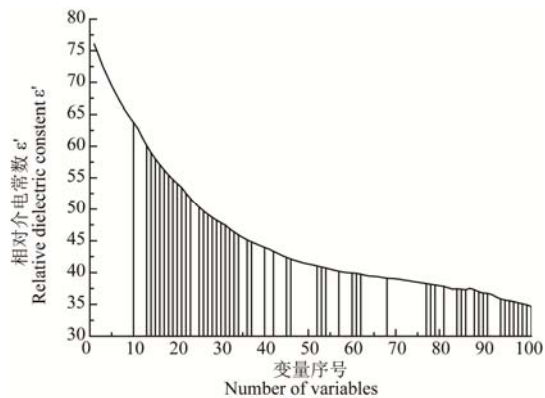
注：虚线为变量选择的阈值。根据随机变量的稳定性系数，取值为 4。
Note: Dotted lines mean the threshold values of selected variable. According to the stability factor, the threshold was 4.

图 2 主成分数为 15 时无信息变量消除法的稳定性分布曲线
Fig.2 Stability distribution of uninformative variables elimination algorithm considering with 15 main components

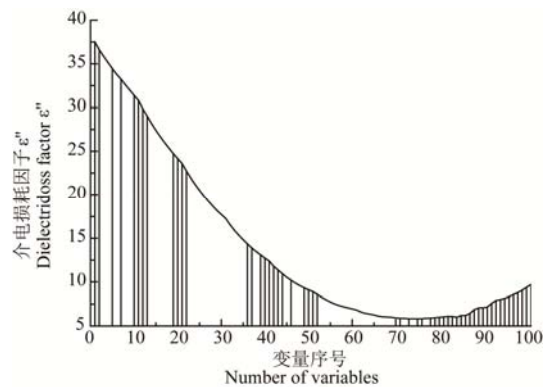
2.4 应用连续投影算法选取特征变量

设定 SPA 的特征变量数为 3~25。RMSEC 值随特征变量数的增加而不断减小，以其不再显著减小时的变量数作为最佳特征变量数。RMSEC 随 SPA 选取的特征变量数的变化曲线如图 4 所示。

由于变量数大于 14 的 RMSEC 减小不显著，而变量数过多会增加模型的复杂性，综合考虑，确定最佳特征变量数为 14。所选择的 14 个特征变量及各特征变量所处的频率见表 3。



a. 选择的 55 个 ϵ'
a. Selected 55 variables of ϵ'



b. 选择的 55 个 ϵ''
b. Selected 53 variables of ϵ''

图 3 UVE 法选择的 55 个 ϵ' 和 53 个 ϵ'' 变量
Fig.3 Selected 55 variables of ϵ' and 53 variables of ϵ'' by UVE

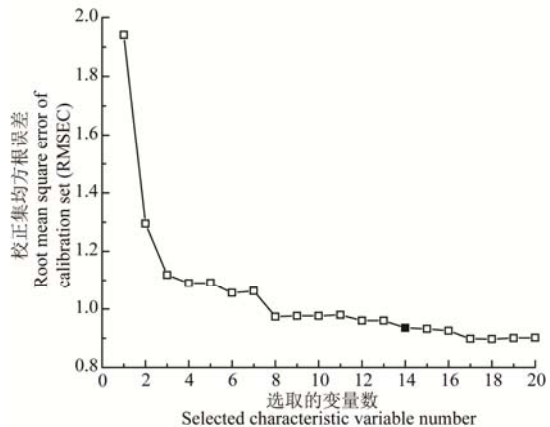


图 4 RMSEC 随 SPA 选取的特征变量数的变化曲线
Fig.4 Changed RMSEC with selected characteristic variables by SPA

2.5 模型训练参数的选择

本文采用交叉验证的方法确定偏最小二乘的因子数。选取逼近速度快，效率高的径向基核函数用于建立支持向量机模型。根据留一交互原则选取惩罚因子 (c) 和松弛变量 (g)。将 ELM 网络的激励函数设为 sigmoidal 函数。根据多次重复建模确定 ELM 网络的隐层节点数。在基于全频谱 (full frequency range, FF) 和经 UVE 和 SPA 处理下用

于建模的 PLS 的因子数、SVR 的参数及 ELM 网络各层神经元数设定见表 4。

表 3 应用连续投影算法选择的 14 个特征变量及其所在的频率
Table 3 Selected 14 characteristic variables by successive projection algorithm (SPA) and frequencies where the variables located

变量号 Variable number	变量 Variable	变量所在 频率/ Frequency where variable located /MHz	变量号 Variable number	变量 Variable	变量所在 频率/ Frequency where variable located /MHz
1	ϵ'	123.5	8	ϵ''	886.3
2	ϵ'	494.1	9	ϵ''	1077.4
3	ϵ'	3935.4	10	ϵ''	1632.8
4	ϵ'	4123.6	11	ϵ''	2289.7
5	ϵ''	76.1	12	ϵ''	2618.2
6	ϵ''	224.7	13	ϵ''	2806.3
7	ϵ''	589.8	14	ϵ''	4500.0

表 4 偏最小二乘、支持向量回归和极限学习机建模参数的设定
Table 4 Modeling parameters of partial least squares(PLS), support vector regression (SVR) and extreme learning machine (ELM)

变量处理方法 Pretreatment method on variables	偏最小二乘 PLS		支持向量回归 SVR		极限学习机 ELM	
	因子数 Factor Number	松弛变 量 c	惩罚因 子 g	输入 节点数 Input nodes	隐层 节点数 Hidden layer nodes	输出 节点数 Output nodes
全频率 FF	6	1.000	0.063	202	30	1
无信息变量 消除法 UVE	5	2.000	0.063	108	20	1
连续投影算 法 SPA	3	5.657	0.250	14	10	1

2.6 建模效果的比较分析

在基于 SPXY 法划分的校正集样品的全频谱以及经 UVE 或 SPA 算法提取的特征变量，应用 PLS、SVR 及 ELM 对 99-1 油桃的糖度进行建模，并用预测集的样品对模型进行验证，其建模效果如表 5 所示。

表 5 偏最小二乘、支持向量回归及极限学习机建模效果的比较
Table 5 Comparison of modeling results using partial least squares(PLS), support vector regression (SVR) and extreme learning machine (ELM)

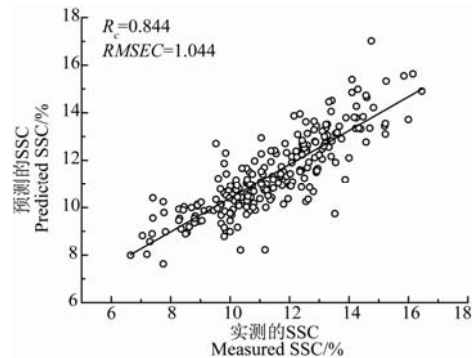
变量处理 方法 Pretreatment method on variables	建模方法 Modeling method	校正 相关 系数 R_c	校正集 均方根 误差 $RMSEC$	预测 相关 系数 R_p	预测集 均方根 误差 $RMSEP$
全频率 FF	PLS	0.864	0.753	0.853	0.824
	SVR	0.896	0.870	0.827	1.144
	ELM	0.795	1.181	0.856	0.857
无信息变量 消除法 UVE	PLS	0.872	0.959	0.854	0.818
	SVR	0.896	0.869	0.841	1.212
	ELM	0.846	1.038	0.879	0.794
连续投影算 法 SPA	PLS	0.856	1.012	0.803	0.922
	SVR	0.887	0.901	0.867	1.274
	ELM	0.844	1.044	0.887	0.782

由于 ELM 的参数是随机选取的，因此结果有

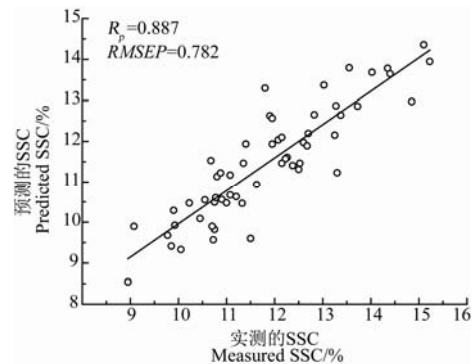
一定的波动。本研究采取重复建模 50 次，选取一个最好的模型作为最终结果。

由表 5 可以看出，在全频谱下用 SVR 建模和经 UVE 处理用 SVR 建模具有最大的 R_c ，均为 0.896，而 UVE 处理用 SVR 建模的 R_p 较好 (0.841)。

全频谱虽完好地保留了数据的原始信息，但依然存在噪声干扰、数据重叠等问题。由于采用所有变量 (202 个) 建模，因此所建模型十分复杂。UVE 虽然消除了大量的无信息变量，所建模型的 R_c 和 R_p 略优于全频谱，但所用变量依然很大 (108 个)，因此，模型同样比较复杂，难于应用于工业化处理中。同全频谱和 SUV 处理相比，SPA 处理能较好地表达原始频谱的 202 个变量中筛选出 14 个特征变量，所建模型比较简单，在工业化应用中有较好的前景。在经 SPA 处理，分别采用 PLS、SVR 和 ELM 所建的 3 种模型中，采用 ELM 建模具有最大的 R_p (0.887) 和最小 $RMSEP$ (0.782)，说明 SPA-ELM 模型具有最好的预测性能。而用 SPA-UVR 所建模型虽具有最大的 R_c (0.887) 和最小的 $RMSEC$ (0.901)，但是其 $RMSEP$ 最大 (1.274)。采用 PLS 所建模型的 R_p 最小，说明 SPA-PLS 所建模型预测性能最差。SPA-ELM 模型的校正及预测结果如图 5 所示。



a. 校正集
a. Calibration set



b. 预测集
b. Prediction set

图 5 校正集和预测集中 99-1 油桃 SSC 的实测值与采用 SPA-ELM 预测值的比较

Fig.5 Measured SSC vs. predicted SSC for 99-1 nectarines in calibration set and prediction set by SPA-ELM

除了本文中的 UVE 和 SPA 变量处理方法外,笔者也曾尝试应用主成分分析法对数据进行处理,但其处理效果一般。这表明主成分分析在介电频谱数据的降维过程中可能遗失了样本的某些重要信息,导致模型效果变差。而 UVE 及 SPA 均能在较好地保持原始频谱数据的基础上达到降维的目的,尤其是 SPA 大大减小了输入变量数,简化了模型,降低了运算量。

3 结 论

1) 在 20~4 500 MHz 间,随着测试信号频率的增大,油桃的相对介电常数单调递减,而介电损耗因子先减小后增大;贮藏时间和频率均对油桃的介电参数有显著的影响。

2) 在无信息变量消除法和连续投影算法处理下,分别从全频谱的 202 个变量中提取出了表征全频谱信息的 108 个和 14 个变量。同无信息变量消除法相比,连续投影算法更适用于对介电频谱数据进行预处理。其处理效果比较稳定,在保证模型预测精度的情况下极大地降低了输入变量的维数,有效地简化了模型。

3) 连续投影算法对于提高支持向量回归和极限学习机的建模效果比较明显。其中连续投影算法结合极限学习机所建模型具有最大的预测相关系数(0.887)和最小的预测均方根误差(0.782)。

[参 考 文 献]

- [1] 郭文川. 果蔬介电特性研究综述[J]. 农业工程学报, 2007, 23(5): 284—289.
Guo Wenchuan. Review of dielectric properties of fruits and vegetables [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2007, 23(5): 284—289. (in Chinese with English abstract)
- [2] 郭文川, 朱新华. 国外农产品及食品介电特性测量技术及应用[J]. 农业工程学报, 2009, 25(2): 308—312.
Guo Wenchuan, Zhu Xinhua. Foreign dielectric property measurement techniques and their applications in agricultural products and food materials[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2009, 25(2): 308—312. (in Chinese with English abstract)
- [3] Skierucha W, Wilczek A, Szyplowska A. Dielectric spectroscopy in agrophysics[J]. International Agrophysics, 2012, 26(2): 187—197.
- [4] Nelson S O, Trabelsi S. Factors influencing the dielectric properties of agricultural and food products[J]. Journal of Microwave Power and Electromagnetic Energy, 2012, 46(2): 93—107.
- [5] 王瑞庆, 周永洪, 张继澍. 温度对火柿电学特性的影响[J]. 农业工程学报, 2012, 28(4): 269—274.
Wang Ruiqing, Zhou Yonghong, Zhang Jishu. Effects of temperature on electrical parameters for 'Huoshi' persimmon fruit[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(4): 269—274. (in Chinese with English abstract)
- [6] 刘亚平, 刘兴华, 李红波. 葡萄冷藏中电学参数与质地特性变化规律[J]. 农业工程学报, 2011, 27(10): 343—348.
Liu Yaping, Liu Xinghua, Li Hongbo. Law of dielectric parameters and texture properties for grape during cold storage[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2011, 27(10): 343—348. (in Chinese with English abstract)
- [7] Soltani M, Alimardani R, Omid M. Evaluating banana ripening status from measuring dielectric properties[J]. Journal of Food Engineering, 2011, 105(4): 625—631.
- [8] Nelson S O, Guo W, Trabelsi S, et al. Sensing quality of watermelons through dielectric permittivity[J]. Measurement Science and Technology, 2007, 18(7): 1887—1892.
- [9] Nelson S O, Trabelsi S, Kays S J. Dielectric spectroscopy of honeydew melons from 10 MHz to 1.8 GHz for quality sensing[J]. Transactions of the ASABE, 2006, 49(6): 1977—1981.
- [10] 郭文川, 朱新华, 郭康权, 等. 桃的电特性及新鲜度识别[J]. 农业机械学报, 2007, 38(1): 112—115.
Guo Wenchuan, Zhu Xinhua, Guo Kangquan, et al. Electrical properties of peaches and its application in sensing freshness[J]. Transactions of the Chinese Society for Agricultural Machinery, 2007, 38(1): 112—115. (in Chinese with English abstract)
- [11] Guo Wenchuan, Zhu Xinhua, Yue Rong, et al. Dielectric properties of fuji apples from 10 to 4500 MHz during storage[J]. Journal of Food Processing and Preservation, 2011, 35(6): 884—890.
- [12] Chia K, Abdul Rahim H, Abdul Rahim R. Neural network and principal component regression in non-destructive soluble solids content assessment: a comparison[J]. Journal of Zhejiang University: Science B, 2012, 13(2): 145—151.
- [13] 赵龙莲, 李军会, 张文娟, 等. 基于 SVR 的傅里叶变换型近红外光谱仪间数学模型传递的研究[J]. 光谱学与光谱分析, 2008, 28(10): 2299—2303.
Zhao Longlian, Li Junhui, Zhang Wenjuan, et al. Calibration transfer between two FTNIR spectrophotometers using SVR[J]. Spectroscopy and

- Spectral Analysis, 2008, 28(10): 2299 — 2303. (in Chinese with English abstract)
- [14] 邓万宇, 郑庆华, 陈琳, 等. 神经网络极速学习方法研究[J]. 计算机学报, 2010, 33(2): 279—287.
Deng Wanyu, Zheng Qinghua, Chen Lin, et al. Research on extreme learning of neural networks[J]. Chinese Journal of Computers, 2010, 33(2): 279 — 287. (in Chinese with English abstract)
- [15] Zhu Xinhua, Guo Wenchuan, Wu Xiaoling, et al. Dielectric properties of chestnut flour relevant to drying with radio—frequency and microwave energy[J]. Journal of Food Engineering, 2012, 113(1): 143—150.
- [16] Zhu Xinhua, Guo Wenchuan, Wu Xiaoling. Frequency- and temperature—dependent dielectric properties of fruit juices associated with pasteurization by dielectric heating[J]. Journal of Food Engineering, 2012, 109(2): 258—266.
- [17] Zhan Xiaori, Zhu Xiangrong, Shi Xinyuan, et al. Determination of hesperidin in tangerine leaf by near— infrared spectroscopy with SPXY algorithm for sample subset partitioning and monte carlo cross validation[J]. Spectroscopy and Spectral Analysis, 2009, 29(4): 964—968.
- [18] Galvão R K H, Araujo M C U, José G E, et al. A method for calibration and validation subset partitioning[J]. Talanta, 2005, 67(4): 736—740.
- [19] Zhang J, Rivard B, Rogge D M. The successive projection algorithm (SPA), an algorithm with a spatial constraint for the automatic search of endmembers in hyperspectral data[J]. Sensors, 2008, 8(2): 1321—1342.
- [20] 洪涯, 洪添胜, 代芬, 等. 连续投影算法在砂糖橘总酸无损检测中的应用[J]. 农业工程学报, 2010, 26(14): 380—384.
Hong Ya, Hong Tiansheng, Dai Fen, et al. Successive projections algorithm for variable selection in nondestructive measurement of citrus total acidity [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(14): 380—384. (in Chinese with English abstract)
- [21] Ye Shengfeng, Wang Dong, Min Shungeng. Successive projections algorithm combined with uninformative variable elimination for spectral variable selection[J]. Chemometrics and Intelligent Laboratory Systems, 2008, 91(2): 194—199.
- [22] Centner V, Massart D L, de Noord O E, et al. Elimination of uninformative variables for multivariate calibration[J]. Analytical Chemistry, 1996, 68(21): 3851—3858.
- [23] 吴迪, 吴洪喜, 蔡景波, 等. 基于无信息变量消除法和连续投影算法的可见-近红外光谱技术白虾种分类方法研究[J]. 红外与毫米波学报, 2009, 28(6): 423—427.
Wu Di, Wu Hongxi, Cai Jingbo, et al. Classifying the species of exopalaemon by using visible and near infrared spectra with uninformative variable elimination and successive projections algorithm[J]. Journal of Infrared and Millimeter Waves, 2009, 28(6): 423 — 427. (in Chinese with English abstract)
- [24] 王晓丹, 王积勤. 支持向量机训练和实现算法综述[J]. 计算机工程与应用, 2004, 40(13): 75—78, 175.
Wang Xiaodan, Wang Jiqin. A survey on support vector machines training and testing algorithms[J]. Computer Engineering and Applications, 2004, 40(13): 75—78, 175. (in Chinese with English abstract)
- [25] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121—167.
- [26] Huang Guangbin, Zhang Hongming, Ding Xiaojian, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE Transactions on Systems Man and Cybernetics Part B: Cybernetics, 2012, 42(2): 513—529.
- [27] Lan Yuan, Hu Zongjiang, Soh Y C, et al. An extreme learning machine approach for speaker recognition[J]. Neural Computing & Applications, 2013, 22(3/4): 417—425.
- [28] Yang Juncheng, Xie Shanjuan, Yoon S, et al. Fingerprint matching based on extreme learning machine[J]. Neural Computing & Applications, 2013, 22(3/4): 435—445.

Non-destructively detecting sugar content of nectarines based on dielectric properties and ANN

Shang Liang¹, Gu Jingsi², Guo Wenchuan^{1*}

(1. College of Mechanical and Electronic Engineering, Northwest A & F University, Yangling 712100, China; 2. College of Food Science and Engineering, Northwest A & F University, Yangling 712100, China)

Abstract: Sugar content is the main attribute of a fruit's internal qualities. It is usually determined by soluble solids content (SSC). Since the traditional method used in detecting SSC with an Abbe-refractor is destructive, it is not suitable for on-line detection. To find a method to measure SSC nondestructively and quickly, the dielectric properties (relative dielectric constant and dielectric loss factor) of intact postharvest 99-1 nectarines at 1 day intervals during 10 days' storage were measured with a vector network analyzer (E5071C) and an open-ended coaxial-line probe (85070E) at 101 discrete frequencies over the frequency range of 20~4500 MHz. The sugar content of the fruit juice of each sample was measured with a digital refractometer. Altogether, 300 nectarines were used in the study. One-hundred and one relative dielectric constant values and 101 dielectric loss factor values at 101 discrete frequencies for each sample were used as variables to build models. A significance analysis was done to investigate whether frequencies and storage time had a significant influence on the values of permittivities. Sample set partitioning based on joint x-y distances (SPXY) was used to subset partitioning. Uninformative variables elimination (UVE) and successive projection algorithm (SPA) were applied to extract the characteristic variables from the original dielectric spectra of dielectric constant and dielectric loss factor. The modeling methods, such as partial least squares (PLS) and artificial neural network technology, such as support vector regression (SVR) and extreme learning machine (ELM) were applied to establish models for predicting SSC from permittivities. The experimental results showed that as the frequency increased, the relative dielectric constant of nectarines decreased, but the dielectric loss factor changed from decreasing to increasing. The analysis of variance indicated that storage time and frequency had a significant influence on dielectric properties. Based on SPXY, 243 samples were partitioned to a calibration set and 57 samples to a prediction set. One-hundred and eight variables were selected from 202 variables with UVE, and 14 characteristic variables were extracted with SPA. SPA was more effective than UVE in selecting useful information from the whole spectra of dielectric constant and dielectric loss factor and in simplifying a model for predicting SSC. Contrasted with PLS, SVR and ELM had better performance in predicting SSC under the SPA pretreatment method. SPA-ELM gave the highest correlation coefficient of predication set (0.887) and the lowest root mean square error of predication set (0.782). The study indicates that dielectric spectra combined SPA and artificial neural network technology could be applied in determining the sugar content of nectarines. It offers useful knowledge for developing nondestructive sensors for fruits' sugar content based on the frequency spectra of permittivities.

Key words: nondestructive examination, dielectric properties, models, nectarine, soluble solids content, artificial neural network, support vector regression, extreme learning machine

(责任编辑: 刘丽英)