

苹果品种及损伤苹果的 FT-NIR 鉴别研究

李光辉¹, 任亚梅^{1*}, 任小林², 赵玉¹, 李帅¹, 苏晋文¹, 刘朵¹

(1.西北农林科技大学食品科学与工程学院, 陕西 杨凌 712100; 2.西北农林科技大学园艺学院, 陕西 杨凌 712100)

摘要: 用傅里叶近红外光谱技术(FT-NIR)对不同品种的苹果以及损伤嘎啦和完好嘎啦进行快速、无损检测, 比较不同判别方法对所建立的区分苹果品种及苹果损伤模型的影响。结果表明: 损伤嘎啦和完好嘎啦的近红外图谱经小波分析预处理后, 用 12000~4000 cm^{-1} 波数范围的前 5 个主成分分别结合多层感知神经网络、径向基神经网络、Fisher 判别 3 种方法所建立的判别模型对未知样本的正确判别率分别为 97.8%、87.2% 和 84.8%, 基于权重法用多元线性回归(MLR)所选择的特征波长所建立的 Fisher 判别模型对未知样本的正确判别率为 89.1%; 用偏最小二乘判别(PLS-DA)所建立的判别模型对未知样本的正确判别率为 100%, 由于 PLS-DA 模型对训练集和验证集的正确判别率均为 100%, 因此 PLS-DA 模型优于其他模型。不同品种苹果的光谱经平滑预处理后, 用全波数范围 12000~4000 cm^{-1} 的前 6 个主成分所建立的判别模型优于经验波数范围 8000~4500 cm^{-1} 所建立的判别模型, 其较优模型对建模集和验证集的正确判别率分别为 90.9% 和 92.1%。近红外光谱技术结合化学计量学可以快速、无损鉴别苹果是否有损伤以及不同品种的苹果。

关键词: 苹果; 近红外技术; 神经网络; 偏最小二乘判别; Fisher 判别

Discrimination and Identification of Bruised Apples and Apple Varieties by FT-NIR

LI Guang-hui¹, REN Ya-mei^{1*}, REN Xiao-lin², ZHAO Yu¹, LI Shuai¹, SU Jin-wen¹, LIU Duo¹

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

2. College of Horticulture, Northwest A & F University, Yangling 712100, China)

Abstract: Near-infrared (NIR) spectroscopy was applied to rapidly and non-destructively distinguish between bruised and intact apples and identify different apple varieties. Besides, the effect of different discrimination methods on the distinguishing and identification models obtained was investigated. The results indicated that discrimination models were developed based on the first five principal components in the range from 12000 cm^{-1} to 4000 cm^{-1} from the NIR spectra of bruised and intact apples subjected to wavelet pretreatment using three different discrimination methods, multilayer perceptron (MLP) neural network, radial basis function (RBF) neural network and Fisher line discriminant analysis (Fisher-DA) with discrimination accuracy rates of 97.8%, 87.2% and 84.8%, respectively for unknown samples. The Fisher linear discriminant analysis model established based on multiple linear regression and the loading weights showed a discrimination accuracy rate of 89.1% compared with 100% for the model established using partial least squares discriminant analysis (PLS-DA). The discrimination accuracy rates of the PLS-DA model for the training and validation sets were both 100%, and therefore the PLS-DA model was superior to others. A better discrimination model was established based on the first six principal components of the NIR spectra of different apple varieties subjected to smoothing pretreatment in the full wavelength range of 12000—4000 cm^{-1} than in the empirical range of 8000—4500 cm^{-1} with discrimination accuracy rates of 90.9% and 92.1% for the predication and validation sets, respectively. In conclusion, the combination of NIR and chemometrics can provide a rapid and non-destructive approach to discriminate whether apples are bruised and identify different apple varieties.

Key words: apple; NIR spectroscopy; artificial neural network; PLS-DA; Fisher-DA

中图分类号: S123; TS255.3

文献标识码: A

文章编号: 1002-6630(2012)16-0251-06

我国是苹果生产大国, 栽培苹果品种繁多; 不同的苹果品种, 其口感、品质和大小等差异较大, 随着

苹果采后商品化处理和加工技术的发展, 苹果分级将越来越重要, 其中重要方面为不同品种苹果鉴别和损伤苹

收稿日期: 2011-07-03

基金项目: 国家苹果产业技术体系专项(NYCYTX-08-05-02)

作者简介: 李光辉(1985—), 男, 硕士研究生, 研究方向为果蔬贮藏、加工和无损检测。E-mail: hnlgh1228@163.com

* 通信作者: 任亚梅(1970—), 女, 副教授, 博士, 研究方向为果蔬深加工和贮藏。E-mail: yameiren@yahoo.com

果识别。因此,研究一种简单、方便、快速识别苹果品种以及损伤苹果的方法对实际生产将具有重要的意义。

目前,近红外光谱无损检测技术在苹果方面的定量分析有糖度^[1-2]、病害果^[3-4]和酸度^[5]等,定性分析有苹果品种^[6]鉴别和损伤苹果^[7-9]识别等。但是,近红外技术对苹果品种和损伤苹果的鉴别研究仍不够成熟,表现为建模方法单一,未有区别苹果的品种和损伤苹果的建模方法比较。本研究用傅里叶近红外技术(fourier transform near infrared spectroscopy, FT-NIR)无损检测不同品种的苹果以及损伤嘎啦和完好嘎啦,比较不同判别分析方法对所建立的区分不同品种苹果模型以及苹果是否有损伤模型的影响,为近红外光谱技术无损鉴别损伤苹果以及苹果品种提供理论基础。

1 材料与方法

1.1 材料

供试材料‘嘎啦’、‘红星’和‘乔纳金’于2010年9月7日采自陕西省永寿县。选取成熟度一致、无病虫害、无机械损伤和个体大小在100~350g之间的果实为试验材料,其中嘎啦230个,红星和乔纳金各120个。

1.2 苹果损伤方法

取嘎啦110个,逐个将苹果拿在手中,果实赤道垂直手心,手心向下,使苹果自然的从0.90m高度落到水泥地面,果实赤道处触地,当果实落地反弹时将果实接住,避免二次损伤。

1.3 光谱采集

果实损伤的第2天将完好嘎啦、红星和乔纳金以及损伤的嘎啦苹果从冷库($0^{\circ}\text{C} \pm 1^{\circ}\text{C}$, RH 85%~90%)中取出,放置24h,使苹果整体温度与室温($25^{\circ}\text{C} \pm 1^{\circ}\text{C}$)一致;在每个苹果上编上号码并在果实赤道处标记等距离的3个光谱采集点(损伤嘎啦的光谱采集点其中一个为损伤部位)。然后,用MPA型傅里叶变换近红外光谱仪(德国Bruker光学仪器公司)采集完好和损伤苹果的近红外光谱;光谱采集时的仪器参数为:固定光纤探头,分辨率为 8cm^{-1} ,扫描次数为64,波长范围 $12000\sim 4000\text{cm}^{-1}$;然后平均每个果实的3个点的近红外图谱,得到每个样本的代表光谱。

2 结果与分析

2.1 损伤嘎啦的鉴别

2.1.1 损伤嘎啦和完好嘎啦的近红外图谱分析

由图1可知,完好嘎啦和损伤嘎啦的近红外光谱曲线形状基本相似,曲线在5300、5708、7186、7501、8752、10495 cm^{-1} 等波数处出现峰值,完好嘎啦和损伤嘎啦的光谱交叉重合,因此很难直接从光谱曲线中区分嘎啦苹果是否有损伤。

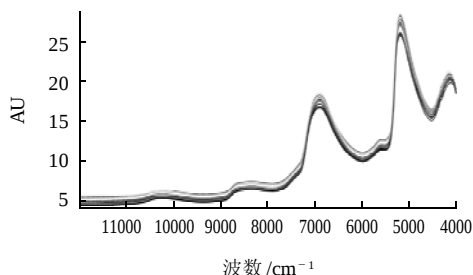


图1 损伤嘎啦和完好嘎啦的近红外光谱图
Fig.1 NIR spectra of non-bruised and bruised Gala apples

2.1.2 基于主成分分析的损伤嘎啦的鉴别

主成分分析的目的是把多个指标化为少数几个综合指标;用主成分分析处理近红外图谱,解决了图谱重叠而无法分析的问题。用OPUS 5.5软件将230个嘎啦苹果的图谱在 $12000\sim 4000\text{cm}^{-1}$ 范围的吸光度导出,放入Excel中形成 230×2074 数据矩阵,然后用Matlab 2010对数据矩阵进行小波分析(db5小波、5尺度)预处理,最后用SPSS 19.0对预处理后的数据矩阵进行主成分分析。由主成分分析结果可知,主成分1(X_1)、主成分2(X_2)、主成分3(X_3)、主成分4(X_4)、主成分5(X_5)的累计贡献率分别为77.612%、97.815%、99.472%、99.793%、99.882%,前5个主成分的累计贡献率已达99.882%,因此前5个主成分可以代表小波分析预处理后光谱的主要信息,使数据矩阵从 230×2074 减少到 230×5 。

2.1.2.1 多层感知器(multilayer perceptron, MLP)神经网络

(1) 模型参数的确定

神经网络模型由3部分组成即输入层、隐含层和输出层。隐含层的激活函数有S型(Sigmoid)函数和双曲正切函数,输出层的激活函数有Softmax函数、恒等函数、S型函数和双曲正切函数;在建模过程中需要确定神经网络模型的隐含层激活函数、隐含层数和输出层的激活函数。本研究建立3层MLP神经网络模型,因此隐含层为1层,由于提取的前5个主成分可以代表光谱的主要信息,故神经网络的输入层为5,隐含层单位数由软件计算确定,输出层节点数为2,即“1”代表正常嘎啦苹果,“2”代表损伤嘎啦苹果。经SPSS 19.0计算,得出的最佳模型参数为:隐含层激活函数为双曲正切函数,隐含层单位数为2和输出层的激活函数为Softmax函数。

(2) MLP神经网络模型对训练集和检验集的判别

据(1)确定的模型参数,建立MLP神经网络模型,训练集(184个)和检验集(46个)判别结果见表1。可见所建立的MLP神经网络模型对训练集和检验集的正确判别率分别为95.1%和97.8%,说明主成分分析结合MLP神经网络对损伤嘎啦苹果的鉴别可行。

表1 MLP神经网络模型对训练集和检验集的判别结果

Table 1 Discrimination results obtained for the training and validation sets using the MLP neural network model

组别	类型	各类别 样本总数	预测到各类别总数		正确判 别率/%
			正常嘎啦	损伤嘎啦	
训练集	正常嘎啦	91	89	2	97.8
	损伤嘎啦	93	7	86	92.5
	合计	184	96	88	95.1
检验集	正常嘎啦	29	28	1	96.6
	损伤嘎啦	17	0	17	100
	合计	46	28	18	97.8

2.1.2.2 径向基函数(radical basis function, RBF)神经网络

(1) 模型参数的确定

本研究建立3层RBF神经网络模型,以提取的前5个主成分为神经网络的输入层,隐含层为1层,由软件计算确定隐含层单位数,输出层节点数为2(即“1”代表正常嘎啦,“2”代表损伤嘎啦)。经SPSS 19.0计算,得出的最佳模型参数为:隐含层激活函数为Softmax函数,隐含层单位数为10及输出层的激活函数为恒等函数。

(2) RBF神经网络模型对训练集与检验集的判别

据(1)确定的模型参数,建立RBF神经网络模型,训练集(191个)和检验集(39个)的判别结果见表2。可见所建立的RBF神经网络模型对训练集和检验集的正确判别率分别为91.6%和87.2%,低于MLP神经网络,尤其是模型对验证集中损伤嘎啦的判别率仅为80.0%,判别率较低,因此,主成分分析结合RBF神经网络对损伤嘎啦的识别有待进一步研究。

表2 RBF神经网络模型对训练集和检验集的判别结果

Table 2 Discrimination results obtained for the training and validation sets using the RBF neural network model

组别	类型	各类别样 本总数	预测到各类别总数		正确判 别率/%
			正常嘎啦	损伤嘎啦	
训练集	正常嘎啦	101	92	9	91.1
	损伤嘎啦	90	7	83	92.2
	合计	191	99	92	91.6
检验集	正常嘎啦	19	18	1	94.7
	损伤嘎啦	20	4	16	80.0
	合计	39	22	17	87.2

2.1.2.3 Fisher判别

用SPSS 19.0建立Fisher判别模型^[10-11],以提取的前5个主成分为自变量,对于训练集的184个样本,“1”和“2”(“1”代表正常嘎啦,“2”代表损伤嘎啦)为因变量。通过软件计算,得到Fisher判别函数为:正常嘎啦 $Y_1 = -0.043X_1 + 0.083X_2 - 0.0040X_3 + 0.074X_4 - 0.088X_5 - 1.656$;损伤嘎啦 $Y_2 = 0.047X_1 -$

$0.091X_2 + 0.044X_3 - 0.081X_4 + 0.096X_5 - 1.839$ 。所得判别函数对训练集(184个)和检验集(46个)的判别结果见表3。

表3 Fisher判别模型对训练集和验证集的判别结果

Table 3 Discrimination results obtained for the training and validation sets using the Fisher linear discriminant analysis model established based on principal component analysis

组别	类型	各类别样 本总数	预测到各类别总数		正确判 别率/%
			正常嘎啦	损伤嘎啦	
训练集	正常嘎啦	96	93	3	96.9
	损伤嘎啦	88	10	78	88.6
	合计	184	103	81	92.9
检验集	正常嘎啦	24	23	1	95.8
	损伤嘎啦	22	6	16	72.7
	合计	46	29	17	84.8

由表3可见,所建立的Fisher判别模型对训练集和检验集的正确判别率分别为92.9%和84.8%,低于MLP神经网络;原因是模型对建模集和验证集中损伤嘎啦的判别率较低,分别为88.6%和72.7%,因此,主成分分析结合Fisher判别对损伤嘎啦苹果的鉴别有待于进一步的研究。

2.1.3 基于指纹图谱的损伤苹果的鉴别

嘎啦苹果的红外图谱,经小波分析预处理后,随机将光谱分为建模集(184个)和验证集(46个);将全波数范围 $12000 \sim 4000\text{cm}^{-1}$ 分为3个区间即波数范围I: $12000 \sim 7000\text{cm}^{-1}$ 、波数范围II: $7000 \sim 5208\text{cm}^{-1}$ 和波数范围III: $5208 \sim 4000\text{cm}^{-1}$ ^[12]。各波数范围的权重分析和数学模型的建立均在SPSS 19.0完成。

对于波数范围I: $12000 \sim 7000\text{cm}^{-1}$,用权重法选取的波长为 10120cm^{-1} ,以此波长为自变量,“1”和“2”(“1”代表正常嘎啦,“2”代表损伤嘎啦)为因变量进行MLR回归分析,其模型的相关系数 R 为0.714。表明:所建立的区分损伤苹果的模型效果较好,可用于区分苹果是否损伤。

对于波数范围II: $7000 \sim 5208\text{cm}^{-1}$,用权重法选取的2个波长分别为 6479.15cm^{-1} 和 6425.586cm^{-1} ,以这2个波长为自变量,“1”和“2”(“1”代表正常嘎啦,“2”代表损伤嘎啦)为因变量进行MLR回归分析,其模型的相关系数 R 为0.743,相关性较好,可用于区分苹果是否有损伤。

对于波数范围III: $5208 \sim 4000\text{cm}^{-1}$,用权重法选取4个波长分别为 4231.887 、 4242.015 、 4223.46cm^{-1} 和 4258.18cm^{-1} ,以这4个波长为自变量,“1”和“2”(“1”代表正常嘎啦,“2”代表损伤嘎啦)为因变量进行回归分析,其模型的相关系数 R 为0.809,相关性较好,可用于区分苹果是否损伤。

将上述选取的7个波长作为自变量,“1”和“2”

(“1”代表正常嘎啦,“2”代表损伤嘎啦)为因变量,用逐步回归法剔除无用波长,选出信息较丰富的波长。经软件多次计算,选出有用的波长为3个,分别为10120、6479.15 cm^{-1} 和4231.887 cm^{-1} ,此时MLR模型的相关系数 R 为0.870。以这3个波长为自变量,“1”和“2”(“1”代表正常嘎啦,“2”代表损伤嘎啦)为因变量,在SPSS 19.0中进行Fisher判别,所得判别函数为:正常嘎啦 $Y_3 = 50.457X_1 + 337.746X_2 - 26.955X_3 - 201.858$;损伤嘎啦 $Y_4 = 174.032X_1 + 322.056X_2 - 42.573X_3 - 220.838$ 。

判别函数对训练集和验证集的判别结果见表4。可见用权重法基于MLR所选择的波长建立的Fisher判别模型对训练集和检验集的正确判别率分别为92.4%和89.1%,低于MLP神经网络,其主要原因是所建立的模型对损伤嘎啦的正确判别率较低,因此,基于权重法用MLR所选择的特征波长建立的Fisher判别模型对损伤嘎啦的识别有待于进一步的研究。

表4 Fisher判别对训练集和验证集的判别结果

Table 4 Discrimination results obtained for the training and validation sets using the Fisher linear discriminant analysis model established based on fingerprinting analysis

组别	类型	各类别 样本总数	预测到各类别总数		正确判 别率/%
			正常嘎啦	损伤嘎啦	
训练集	正常嘎啦	96	93	3	96.9
	损伤嘎啦	88	11	77	87.5
	合计	184	104	30	92.4
检验集	正常嘎啦	24	23	1	95.8
	损伤嘎啦	22	4	18	81.8
	合计	46	27	19	89.1

2.1.4 苹果有无损伤的PLS判别

PLS-DA是一种基于PLS回归的一种判别方法,需先将光谱数据与分类变量进行线性回归,然后判别,其基本思想是首先建立训练集样本的分类变量,然后用PLS法建立分类变量与光谱数据间的回归模型,最后根据训练集与分类变量的PLS-DA模型,计算验证集的分类变量 Y_p ,根据 Y_p 和偏差判定样本属于哪一类^[13]。

用OPUS 5.5建立PLS-DA模型,其中分类变量为“1”和“2”(“1”代表正常嘎啦,“2”代表损伤嘎啦),通过软件自身优化,在原始光谱、波数范围为11995.4~7498.1 cm^{-1} 和6101.9~5446.2 cm^{-1} 、主成分为5时所建立的PLS-DA判别模型效果较好,其预测时的相关系数 R^2 和RMSEE分别为91.76%和0.153,验证时的 R^2 和EMSECV分别为85.85%和0.18,此时模型对训练集和验证集的正确判别率均为100%。PLS-DA模型对验证集的判别结果见表5。

表5 PLS-DA模型对验证集的判别结果

Table 5 Discrimination results obtained for the validation set using the PLS-DA model

序号	真实值	预测值	偏差	序号	真实值	预测值	偏差
1	1	0.095	0.044	23	1	0.878	0.122
2	1	1.23	-0.23	24	1	1.251	-0.251
3	1	1.23	-0.23	25	2	1.568	0.432
4	1	0.951	0.049	26	2	2.130	0.130
5	1	1.052	-0.052	27	2	2.264	-0.264
6	1	0.929	0.071	28	2	1.826	0.174
7	1	0.744	0.256	29	2	1.602	0.398
8	1	0.818	0.182	30	2	1.956	0.044
9	1	0.985	0.015	31	2	2.030	-0.030
10	1	0.928	0.072	32	2	1.958	0.042
11	1	1.114	-0.114	33	2	2.090	-0.090
12	1	1.072	-0.072	34	2	2.011	-0.011
13	1	1.319	-0.319	35	2	1.789	0.211
14	1	1.067	-0.067	36	2	2.271	-0.271
15	1	1.119	-0.119	37	2	1.785	0.215
16	1	1.009	-0.009	38	2	2.167	-0.167
17	1	1.131	-0.131	39	2	2.150	-0.150
18	1	1.245	-0.245	40	2	1.804	0.196
19	1	1.06	-0.0597	41	2	1.944	0.056
20	1	1.118	-0.118	42	2	2.108	-0.108
21	1	0.954	0.046	43	2	1.645	0.355
22	1	1.099	-0.099	44	2	1.859	0.141

2.2 苹果品种的识别

2.2.1 不同品种苹果的 NIR 光谱分析

在采集样本的光谱时,由于样本的均匀程度、仪器以及环境等因素的影响,图谱易出现噪音和基线漂移等现象,因此,在建模分析时,就需要对苹果的近红外光谱进行预处理。本研究选取平滑(9点)对光谱进行预处理(图2),平滑能较好的将图谱的噪音滤除,消除了基线的偏移,减少了外界因素对光谱的影响。

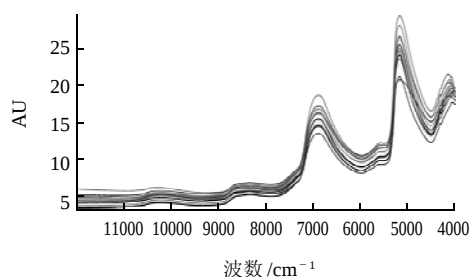


图2 经平滑预处理后的3个品种苹果的近红外图谱

Fig.2 Smoothed NIR spectra of non-bruised and bruised Gala apples

由图2可知,3个品种苹果的近红外图谱经平滑9点预处理后在11076.20、8313.76、7000、6888.60 cm^{-1} 和51620 cm^{-1} 等波峰或波谷处有明显的差异,这些差异为鉴定苹果不同的品种提供了理论基础。

2.2.2 主成分分析

首先,通过 OPUS 5.5 将完好嘎啦、红星、乔纳金各 120 个近红外光谱进行平滑(9 点)预处理;然后,将平滑预处理后的吸光度导出,并保存到 Excel 中;最后,分别选取全波数范围 12000~4000 cm^{-1} 和经验波数范围 8000~4500 cm^{-1} 的吸光度在 SPSS 19.0 软件中进行主成分分析。

表 6 可知,全波数范围 12000~4000 cm^{-1} 的前 6 个主成分累计贡献率已达 99.91%,并且前 6 个主成分的特征根都大于 1 和经验波数范围 8000~4500 cm^{-1} 的前 4 个主成分累计贡献率已达 99.89%,并且前 4 个主成分的特征根都大于 1,因此,全波数范围 12000~4000 cm^{-1} 的主要信息可以用这前 6 个主成分表示,经验波数范围 8000~4500 cm^{-1} 的主要信息可以前 4 个主成分表示。

表 6 不同波数范围的主成分累积贡献率

Table 6 Discriminant analysis results of testing samples predicted by PLS-DA model

类型	波数范围/ cm^{-1}	主成分					
		PC ₁	PC ₂	PC ₃	PC ₄	PC ₅	PC ₆
累计贡献率/%	12000~4000	62.03	96.15	99.06	99.75	99.85	99.91
	8000~4500	90.14	97.45	99.59	99.89	—	—
贡献率/%	12000~4000	62.03	34.12	2.91	0.70	0.10	0.06
	8000~4500	90.14	7.31	2.14	0.30	—	—

2.2.2.1 用 12000~4000 cm^{-1} 波数范围内提取的前 6 个主成分建立多层感知器(MLP)神经网络模型

(1) 模型参数的确定

本研究建立 3 层 MLP 神经网络模型即输入层-隐含层-输出层,由于提取的前 6 个主成分可以代表光谱的主要信息,故 MLP 神经网络的输入层为 6,隐含层为 1 层,输出层节点数为 3(即苹果不同品种的名义变量,“1”代表嘎啦,“2”代表红星,“3”乔纳金),隐含层单位数、隐含层和输出层函数由软件确定。经 SPSS19.0 软件多次计算,得出的最佳模型为隐含层激活函数为双曲正切函数,隐含层单位数为 4 和输出层的激活函数为 Softmax 函数。

(2) MLP 神经网络模型的训练结果与检验结果

根据(1)的参数,建立 MLP 神经网络模型,训练集(284 个)和检验集(76 个)的判别结果如表 7。可见模型对训练集样本的总的正确识别率为 90.8%,其中第 1 类(嘎啦)样本的正确识别率最高,达到 96.6%;第 3 类(乔纳金)的正确识别率良好,为 90.9%。验证集样本总的识别率为 92.1%,其中第 1 类(嘎啦)样本的正确识别率最高,为 93.9%。因此,利用全波段光谱的主成分结合 MLP 神经网络可以区分不同品种的苹果。

表 7 12000~4000 cm^{-1} 波数范围内 MLP 神经网络模型对 360 个样品的识别结果Table 7 Recognition results obtained for 360 samples using the MLP neural network model based on the first six principal components in the range of 12000 — 4000 cm^{-1}

类型	原始类别	预测到各类别总数			正确率/%
		1	2	3	
训练集	1	84	2	1	96.6
	2	4	84	10	85.7
	3	0	9	90	90.9
检验集	1	31	1	1	93.9
	2	0	20	2	90.9
	3	0	2	19	90.5

2.2.2.2 用 8000~4500 cm^{-1} 波数范围内提取的前 4 个主成分建立多层感知器(MLP)神经网络模型

(1) 模型参数的确定

本研究建立 3 层 MLP 神经网络模型即输入层-隐含层-输出层,因此隐含层为 1 层;由于提取的前 4 个主成分可以代表光谱的主要信息,故 MLP 神经网络的输入层为 4,输出层节点数为 3(即苹果不同品种的名义变量,“1”代表嘎啦,“2”代表红星,“3”乔纳金),隐含层单位数和隐含层、输出层函数由软件确定。经 SPSS19.0 软件多次计算,得出的最佳模型为隐含层激活函数为双曲正切函数,隐含层单位数为 3 和输出层的激活函数为 Softmax 函数。

(2) 模型对训练集和检验集的判别结果

根据(1)的参数,建立 MLP 神经网络模型,训练集(284 个)、检验集(76 个)判别结果见表 8。可见训练集样本中第 1 类(红星)样本的正确识别率最高,为 82.4%;第 2 类(红星)正确率识别率最差,为 76.5%,不符合检测的基本要求。训练集和验证集样本的总的正确识别率分别为 79.2% 和 75%,因此,利用 8000~4500 cm^{-1} 波数范围的主成分结合 MLP 神经网络对苹果品种的识别有待于进一步的研究。

表 8 8000~4500 cm^{-1} 波数范围内 MLP 神经网络模型对 360 个样品的识别结果Table 8 Recognition results obtained for 360 samples using the MLP neural network model based on the first four principal components in the range of 8000 — 4500 cm^{-1}

类型	原始类别	预测到各类别总数			正确率/%
		1	2	3	
训练集	1	75	5	11	82.4
	2	14	78	10	76.5
	3	11	8	72	79.1
检验集	1	20	2	7	69.0
	2	2	15	1	83.3
	3	4	3	22	75.9

3 结 论

3.1 完好嘎啦和损伤嘎啦的近红外图谱预处理后,用

全波段光谱的主成分分别结合 MLP 神经网络、RBF 神经网络、Fisher 判别, 以及偏最小二乘判别等建模方法, 建立了区分嘎啦苹果有无损伤的 5 个判别模型, 结果为: 近红外图谱经小波分析预处理后, 用 $12000\sim 4000\text{cm}^{-1}$ 波数范围的前 5 个主成分分别结合 MLP 神经网络、RBF 神经网络、Fisher 判别 3 种方法所建立的判别模型对未知样本的正确判别率分别为 97.8%、87.2% 和 84.8%, 基于权重法用 MLR 所选择的特征波长所建立的 Fisher 判别模型对未知样本的正确判别率为 89.1%; 用 PLS-DA 判别法所建立的判别模型对未知样本的正确判别率为 100%, 由于 PLS-DA 模型对训练集和验证集的正确判别率都为 100%, 因此 PLS-DA 模型优于其他模型。表明近红外技术结合化学计量学可以快速、无损鉴别嘎啦苹果是否有损伤。

3.2 用 $8000\sim 4500\text{cm}^{-1}$ 和 $12000\sim 4000\text{cm}^{-1}$ 两个波数范围的近红外光谱提取的主成分分别结合 MLP 神经网络建立了区分苹果品种的 2 个判别分析模型。结果表明: 用全波数范围 $12000\sim 4000\text{cm}^{-1}$ 的主成分所建立的模型优于经验波数范围 $8000\sim 4500\text{cm}^{-1}$ 建立的判别模型, 其对建模集和验证集的正确判别率分别为 90.9% 和 92.1%, 说明: 全波段的前 6 个主成分才能完整代表这 3 个苹果品种的特征, 用全波段的主成分结合 MLP 神经网络可以区分不同品种的苹果。

参考文献:

- [1] 周丽萍, 胡耀华, 陈达, 等. 苹果可溶性固形物含量的检测方法 - 基于可见光近红外光谱技术[J]. 农机化研究, 2009(4): 104-107.
- [2] 王加华, 韩东海. 便携式近红外水果糖度分析模型简化研究[J]. 食品安全质量检测技术, 2009, 1(1): 32-38.
- [3] 韩东海, 刘新鑫, 赵丽丽, 等. 苹果水心病的光学无损检测[J]. 农业机械学报, 2004, 35(5): 143-146.
- [4] 韩东海, 刘新鑫, 鲁超, 等. 苹果内部褐变的光学无损检测研究[J]. 农业机械学报, 2006, 37(6): 86-88.
- [5] 应义斌, 刘燕德, 傅霞萍. 苹果有效酸度的近红外漫反射无损检测[J]. 农业机械学报, 2004, 35(6): 124-126.
- [6] 何勇, 李晓丽. 用近红外光谱鉴别杨梅品种的研究[J]. 红外与毫米波学报, 2006, 25(3): 192-194.
- [7] XING Juan, BRAVO C, MOSHOU D, et al. Bruise detection on "Golden Delicious" apples by VIS/NIR spectroscopy[J]. Computers and Electronics in Agriculture, 2006, 52(1/2): 11-20.
- [8] XING Juan, SAEYS W, De BAERDEMAEKER J. Combination of chemometric tools and image processing for bruise detection on apples[J]. Computers and Electronics in Agriculture, 2007, 56(1): 1-13.
- [9] 韩东海, 刘新鑫, 赵丽丽, 等. 受损苹果颜色和组织的近红外光谱特性[J]. 农业机械学报, 2003, 34(6): 112-115.
- [10] 李亮. 近红外光谱技术在原料奶掺假快速检测及新鲜度检测中的应用研究[D]. 杨凌: 西北农林科技大学, 2010.
- [11] 唐启义, 冯明光. DPS 数据处理系统 - 实验设计、统计分析及模型的优化[M]. 北京: 科学出版社, 2006.
- [12] 李桂峰, 赵国建, 刘兴华, 等. 苹果质地品质近红外无损检测和指纹分析[J]. 农业工程学报, 2008, 24(6): 189-173.
- [13] 李剑锋. 基于近红外光谱漫反射技术检测鸡蛋新鲜度的研究[D]. 武汉: 华中农业大学, 2009.