

基于流形学习算法的柑橘叶片氮含量光谱估测模型*

岳学军¹ 全东平¹ 洪添胜² 刘永鑫¹ 吴慕春² 段洁利²

(1. 华南农业大学电子工程学院, 广州 510642; 2. 华南农业大学工程学院, 广州 510642)

摘要: 提出了一种基于流形学习算法的柑橘叶片氮含量光谱快速检测方法。分别在萌芽期、稳果期、壮果促梢期和采果期,使用 ASD FieldSpec 3 光谱仪采集了柑橘叶片的反射光谱,并同步采用凯氏定氮法测定叶片的氮含量。首先采用正交试验确定各个生长期小波去噪的最佳参数组合,然后分别采用主成分分析、多维尺度变换、局部线性嵌入、等距映射和拉普拉斯特征映射 5 种流形学习算法对原始光谱和经小波去噪后的光谱数据进行特征提取,将特征数据导入支持向量机回归建立柑橘叶片氮含量预测模型,4 个生长期的最佳验证集模型决定系数依次为 0.901 4、0.934 4、0.895 4 和 0.877 9。试验结果表明,这 5 种流形学习算法都能有效地用于柑橘叶片氮含量预测,为柑橘叶片氮含量快速无损检测、生长态势监测和变量施肥提供了理论依据。

关键词: 柑橘叶片 氮含量 流形学习 光谱

中图分类号: O657.3; S126 **文献标识码:** A **文章编号:** 1000-1298(2015)06-0244-07

Estimation Model of Nitrogen Content for Citrus Leaves by Spectral Technology Based on Manifold Learning Algorithm

Yue Xuejun¹ Quan Dongping¹ Hong Tiansheng² Liu Yongxin¹ Wu Muchun² Duan Jieli²

(1. College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China

2. College of Engineering, South China Agricultural University, Guangzhou 510642, China)

Abstract: Traditional methods of obtaining nitrogen content of citrus leaves are time-consuming, and the process is cumbersome and harmful to citrus leaves, which need proficient experiment techniques and amounts of instruments, equipment and chemical reagents. According to the high dimensionality and redundancy of origin spectral reflectance, a nitrogen content obtaining method of citrus leaves was provided based on manifold learning algorithm which was applied to the high-dimensional spectral vectors for dimension reduction and feature extraction. During four different growth stages, corresponding to germination, stability, bloom and picking stages, spectral reflectance of citrus leaves were measured by the ASD FieldSpec 3 spectrometer, respectively, and at the same time, nitrogen content of citrus leaves was obtained by using Kjeldahl method. For data processing, firstly the parameter combination of wavelet denoising which was used to the high-frequency noise removal was optimized through orthogonal test, and then the principal component analysis (PCA), multidimensional scaling (MDS), locally-linear embedding (LLE), isometric mapping (Isomap) and laplacian eigenmaps (LE) manifold learning algorithms were applied to extract features of original spectrum and denoised spectrum. Finally, the five corresponding support vector regression (SVR) prediction models of nitrogen content for citrus leaves were established based on their features. Experiment results reveal that the five manifold learning algorithms can be effectively used to predict nitrogen content of citrus leaves, which provides theoretical

收稿日期: 2015-01-19 修回日期: 2015-04-24

* 国家自然科学基金资助项目(30871450)、广东省自然科学基金资助项目(S2012010009856)和广州市科技计划资助项目(7414558112697)

作者简介: 岳学军,教授,主要从事农业工程、通信和信息技术应用研究,E-mail: yuexuejun@scau.edu.cn

通讯作者: 洪添胜,教授,主要从事农业工程、机电一体化和信息技术应用研究,E-mail: tshong@scau.edu.cn

basis for obtaining nitrogen content of citrus leaves rapidly and non-destructively, as well as in growth monitoring and variable-rate fertilization.

Key words: Citrus leaves Nitrogen content Manifold learning Spectrum

引言

氮(Nitrogen, N)是果树生长过程中需求量最大的一种必需营养元素,对果树的生理生化、物质代谢、果实产量及品质的形成等有不可替代的作用^[1-2]。我国是柑橘生产大国,但由于管理与技术投入的不足导致其柑橘产业的国际竞争力总体不强^[3],柑橘叶片氮含量的快速无损检测对于柑橘营养元素监测、柑橘产量预测和合理施肥灌溉具有重要意义。

随着光谱技术的迅速发展,国内外研究人员深入研究了柑橘叶片氮含量的光谱检测方法^[4-8]。但通过对国内外相关研究成果的查阅分析,发现研究对象绝大多数处于某单一生长期,并未关注各个重要生长期光谱分析与建模的差异。实际上,果树在生长过程中,营养成分不会维持稳定不变,而是处于不断改变的状态中,因此,基于单一生长期的光谱估测模型对于果树整个生长周期的营养元素含量预测不具有指导性意义^[9]。本文采用流形学习(Manifold learning)算法对萌芽期、稳果期、壮果促梢期和采果期4个柑橘树重要生长期的多维光谱数据进行降维和特征提取,进而获得不同生长期柑橘叶片氮含量较高精度的估测模型,旨在为柑橘营养元素含量的快速无损检测提供参考。

1 材料与方法

1.1 样本管理与采集

本文试验选用的柑橘样本是种植于广东省广州市萝岗区萝岗甜橙种植示范场内的3年生萝岗甜橙树。柑橘园内土壤为有机红壤,所施加的肥料为由尿素 $\text{CO}(\text{NH}_2)_2$ (纯度 46%)、过磷酸钙 $\text{Ca}(\text{H}_2\text{PO}_4)_2 \cdot \text{H}_2\text{O}$ (纯度 12%) 和氯化钾 KCl (纯度 45%) 组成的混合肥。为了使柑橘样本中各营养元素含量具有适当大的范围和合适的含量梯度,使估测模型适用范围更广,氮、磷、钾肥的施用方案采用三因素二水平和三水平复合设计。施氮肥量(换算成纯氮)为 276(N_1)、138(N_2)、69 g/(株·次)(N_3);施磷肥量(换算成纯磷)为 138 g/(株·次)(P_1)和 34.5 g/(株·次)(P_2);施钾肥量(换算成纯钾)为 220.8 g/(株·次)(K_1)和 55.2 g/(株·次)(K_2),共为 12 个处理,增设 1 个不施肥的对照组(CK),总共为 13 个处理,每个处理设置 3 个重复,每个重复的

试验小区均是 5 株树,随机排列,总共有 195 株柑橘树样本。在 2013 年萌芽期(4 月初)、稳果期(6 月初)、壮果促梢期(8 月初)和采果期(12 月初)各施一次氮磷钾混合肥,在每次施肥 30 d 后,按树冠上下层和东西南北方位,采集顶梢往下的第 3、4 片健康鲜叶,每株树共 8 片,其后同步测定叶片反射光谱与氮含量。

1.2 叶片反射光谱测定

光谱测量仪器为美国 ASD 公司 FieldSpec 3 手持便携式光谱分析仪,波段为 350 ~ 2 500 nm,输出间隔为 1 nm。测量前均使用标准白板进行校准,取叶尖、叶中和叶根 3 个部位的反射光谱均值作为该叶片的光谱反射值,将每株树 8 个叶片的光谱反射值均值作为该柑橘样本的光谱描述,故每个生长期均有 195 个样本。在本文中所建立的模型都是将每个生长期所收集的 195 个样本随机选取 80% (156 个样本)的数据作为校正集,用于训练建模,其余 20% (39 个样本)作为验证集,用于评估模型,每个试验方案均取 30 次独立运行结果的均值作为最终结果记录。

1.3 叶片全氮含量测定

本文采用凯氏定氮法对柑橘叶片全氮含量进行测定。在光谱测定后,用脱脂棉将柑橘叶片擦拭干净,放入 105℃ 的干燥箱中杀青 30 min,然后于 80℃ 干燥至质量恒定。使用研钵将干燥样品研磨至粉末状后,称取 0.5 g 干样,在催化剂条件下用浓硫酸进行硝化,冷却后采用丹麦 FOSS 公司生产的 Kjeltec 8400 全自动凯氏定氮仪测定柑橘叶片的全氮含量。

1.4 流形学习算法

原始光谱数据的维度高、冗余性大,并且由于环境或仪器噪声等因素的影响而呈现复杂性^[10],若直接建模分析,不但建模复杂度高,还会降低模型精度。因此,本文采用能够有效使用低维流形结构表示高维空间数据^[11-12]的流形学习算法对多维光谱数据进行降维和特征提取。流形学习算法主要分为线性和非线性两大部分,线性流形学习算法包括主成分分析(Principal component analysis, PCA)^[13]、多维尺度变换(Multidimensional scaling, MDS)^[14]等,非线性流形学习算法包括局部线性嵌入(Locally-linear embedding, LLE)^[15]、等距映射(Isometric mapping, Isomap)^[16]、拉普拉斯特征映射(Laplacian eigenmaps, LE)^[17]等。由于其中部分流形学习算法

对于噪声点过于敏感,故本文采用小波去噪对原始光谱进行预处理。

1.5 小波去噪参数的确定

在小波去噪^[18-19]的去噪参数确定中,选择合适的参数能取得更好的去噪效果^[20],故本文引入正交试验分析,以小波基函数、分解层数和阈值方案为影响因素。通过分析以往的相关研究文献,发现小波基函数 haar、coif2、sym8、db4 较为常用且去噪效果较好,将其作为候选的小波基函数;分解层数一般在大于 7 后,继续增加层数对去噪效果无明显影响^[21],正交试验的因素水平如表 1 所示。

表 1 正交试验因素水平

Tab.1 Factors and levels of orthogonal test

水平	因素		
	小波基函数 A	分解层数 B	阈值方案 C
1	haar	1	rigrsure
2	coif2	3	heursure
3	sym8	5	sqtwolog
4	db4	7	minimaxi

2 结果与分析

2.1 不同生长期柑橘叶片的氮含量分析

在对氮含量数据统计分析时,将 13 个处理中含有同一个氮处理水平的若干个处理合并成一个氮处理(例如 $N_1P_1K_1$ 、 $N_1P_1K_2$ 、 $N_1P_2K_1$ 与 $N_1P_2K_2$ 可合并成 N_1 ,CK 则代表施氮量为 $0\text{ g}/(\text{株}\cdot\text{次})$),则 4 个生长期不同氮处理水平下柑橘样本的叶片氮含量(质量比)统计数据如表 2 所示。从表 2 中分析得出,柑橘叶片在萌芽期不断生长,迅速吸收氮,在稳果期达到最高峰,随着生长周期往后叶片氮含量逐渐下降,叶片氮含量变化趋势从小到大依次为:采果期、萌芽期、壮果促梢期、稳果期。

表 2 不同生长期柑橘样本的氮含量统计数据

Tab.2 Descriptive statistics of nitrogen content

during different growth stages g/kg

处理	萌芽期	稳果期	壮果促梢期	采果期
N_1	28.613 3	30.397 4	29.156 2	26.211 0
N_2	26.322 1	29.464 3	28.564 7	24.987 9
N_3	24.837 2	27.235 1	26.137 1	22.210 8
CK	20.481 9	24.115 6	22.015 6	19.221 7
平均值	25.063 6	27.803 1	26.468 4	23.157 8

2.2 不同施氮水平下柑橘叶片的光谱特征分析

图 1 为在萌芽期内 4 种不同氮处理水平下的柑橘叶片平均反射光谱,在可见光区域,450 nm 和 670 nm 波长附近反射光谱曲线存在 2 处波谷,分析主要是由于柑橘叶片叶绿素分别对蓝光和红光的强

烈吸收;在 550 nm 波长附近反射光谱曲线存在 1 处波峰,分析是由于 550 nm 波长附近是叶绿素对绿光的强烈反射波段所致;在 680 ~ 750 nm 波段反射光谱曲线急剧上升,这是绿色植物普遍存在的红边特征。在近红外区域,750 ~ 1 300 nm 波段附近呈现出强烈的反射状态,这是由于叶肉内的海绵组织结构内有很多大反射表面的空腔,且细胞内的叶绿素呈水溶胶状态,具有强烈的红外反射^[22];在 1 300 nm 后,反射光谱曲线呈现总体缓慢下降的趋势,在 1 450 nm 和 1 900 nm 波长附近存在 2 处波谷,这主要与这 2 个波长是水和二氧化碳的强吸收带有关^[22]。同时从总体上看出,在可见光区域,光谱反射率与柑橘叶片氮含量成反比关系,在叶绿素对绿光的强反射波段附近非常明显;而在近红外区域,光谱反射率与柑橘叶片氮含量则成正比关系,在 750 ~ 1 300 nm 处不同氮含量反射光谱曲线的反射率差异极其明显,这为有效利用光谱技术估测柑橘叶片氮含量提供了依据。

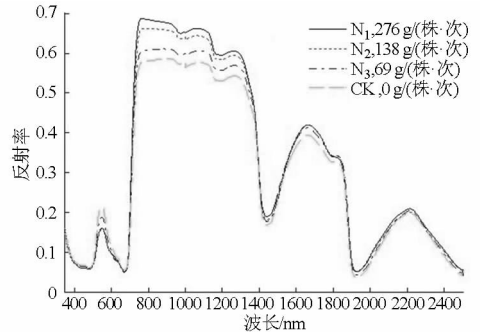


图 1 萌芽期不同氮处理水平的反射光谱

Fig.1 Reflectance spectra with different nitrogen treatments at germination stage

2.3 基于流形学习算法的叶片氮含量建模与分析

2.3.1 萌芽期柑橘叶片氮含量建模分析

在萌芽期内,正交试验采用表 1 的正交试验因素水平表进行小波去噪参数的确定,将经过不同参数组合去噪后的光谱数据直接导入偏最小二乘回归(Partial least square regression, PLSR)中建立回归预测模型,其最佳主成分数选用留一法(Leave-one-out)确定,正交试验的评估指标选为验证集模型决定系数,极差分析结果如表 3 所示。

由表 3 可知,因素 C 对建模性能影响程度最大,因素 A 次之,因素 B 影响程度最小。从表 3 中可看出最佳水平组合是 $A_2B_4C_1$,即小波基函数为 coif2、分解层数是 7、阈值方案是 rigrsure 规则。因此,采用 $A_2B_4C_1$ 水平组合对萌芽期内原始光谱数据进行小波去噪。

以原始光谱数据和经小波去噪后的光谱数据作为输入矢量,分别采用主成分分析(PCA)、多维尺度

表 3 萌芽期小波去噪正交试验的极差分析结果

Tab.3 Range analysis results of wavelet denoising orthogonal test at germination stage

指标	因素		
	A	B	C
K_1	0.761 0	0.765 3	0.787 6
K_2	0.786 0	0.779 6	0.773 6
K_3	0.772 5	0.761 1	0.771 3
K_4	0.766 6	0.780 1	0.753 6
R	0.025 0	0.019 0	0.034 0
优水平	A_2	B_4	C_1

注： K_1 、 K_2 、 K_3 、 K_4 为某一水平的评价指标均值，R 为极差，下同。

变换 (MDS)、局部线性嵌入 (LLE)、等距映射 (Isomap)、拉普拉斯特征映射 (LE) 进行特征提取，针对流形维数人为设定导致建模效果不佳的弊端，采用测地线最小生成树法 (Geodesic minimum spanning tree, GMST)^[23] 确定光谱数据中的本征流形维数，最终将提取的特征数据导入支持向量机回归 (Support vector regression, SVR) 模型中进行建模回归，SVR 的正则项参数 C 和径向基函数 (Radial basis function, RBF) 的核参数 γ 通过格点搜索法 (grid search) 和交叉验证法 (cross validation) 选取最优值，以校正集模型决定系数、校正集均方根误差、验证集模型决定系数和验证集均方根误差作为评估指标，建模结果如表 4 所示。

表 4 萌芽期柑橘叶片氮含量预测模型建模结果

Tab.4 Modeling results of citrus leaves nitrogen content forecasting models at germination stage

光谱形式	回归方法	流形维数	校正集		验证集	
			决定系数	均方根误差	决定系数	均方根误差
原始光谱	PCA-SVR	6	0.895 1	0.560 9	0.842 1	0.668 7
	MDS-SVR	6	0.879 3	0.589 8	0.814 7	0.708 6
	LLE-SVR	6	0.855 3	0.651 0	0.821 4	0.688 3
	Isomap-SVR	6	0.877 0	0.601 3	0.859 6	0.644 2
	LE-SVR	6	0.867 3	0.579 9	0.778 5	0.761 3
去噪光谱	PCA-SVR	5	0.896 9	0.555 5	0.876 7	0.560 9
	MDS-SVR	5	0.898 6	0.550 3	0.860 7	0.625 1
	LLE-SVR	5	0.902 6	0.543 4	0.866 9	0.619 8
	Isomap-SVR	5	0.943 9	0.428 5	0.901 4	0.545 3
	LE-SVR	5	0.885 0	0.561 8	0.872 6	0.573 1

从表 4 中看出，采用经小波去噪后的光谱数据，5 种流形学习算法氮含量预测模型均能有效地对柑橘叶片氮含量进行建模回归预测，其中 Isomap-SVR 的建模效果最好，校正集和验证集的模型决定系数 R^2 分别高达 0.943 9 和 0.901 4，对应的均方根误差为 0.428 5 和 0.545 3，PCA-SVR 模型次之，MDS-SVR 建模效果最差，但其验证集模型决定系

数 R^2 依然达到了 0.860 7，均方根误差为 0.625 1。同时表 4 结果表明采用合适参数的小波去噪对原始光谱进行处理能明显提高氮含量预测模型的建模性能，以 Isomap-SVR 模型为例，校正集和验证集的模型决定系数 R^2 分别比直接用原始光谱建模时提高了 7.63% 和 4.86%，均方根误差分别比直接用原始光谱建模时降低了 28.74% 和 15.35%。采用测地线最小生成树法计算出原始光谱数据和经小波去噪后的光谱数据的本征流形维数不同，前者比后者多了 1 维，分析是由于存在噪声干扰，多出的 1 维对应的是含有噪声的数据部分，影响了建模性能，更加突出了需进行去噪处理的必要性。

2.3.2 稳果期柑橘叶片氮含量建模分析

根据表 5 的正交试验分析结果，选择最佳水平组合 $A_3B_4C_2$ ，即小波基函数为 sym8、分解层数是 7、阈值方案是 heursure 规则，对稳果期内原始光谱数据进行小波去噪。

表 5 稳果期小波去噪正交试验的极差分析结果

Tab.5 Range analysis results of wavelet denoising orthogonal test at stability stage

指标	因素		
	A	B	C
K_1	0.797 0	0.814 3	0.810 2
K_2	0.799 7	0.796 2	0.816 1
K_3	0.828 6	0.804 4	0.812 3
K_4	0.819 1	0.829 4	0.805 7
R	0.031 6	0.033 2	0.010 4
优水平	A_3	B_4	C_2

同样地，以原始光谱数据和经小波去噪后的光谱数据作为输入矢量，分别采用主成分分析 (PCA)、多维尺度变换 (MDS)、局部线性嵌入 (LLE)、等距映射 (Isomap)、拉普拉斯特征映射 (LE) 对其进行特征提取，将特征数据导入支持向量机回归 (Support vector regression, SVR) 建立柑橘叶片氮含量预测模型，建模结果如表 6 所示。

由表 6 可知，与萌芽期不同，在稳果期内，MDS-SVR 模型建立的柑橘叶片氮含量预测模型的建模精度最高，校正集和验证集的模型决定系数 R^2 分别高达 0.967 5 和 0.934 4，对应的均方根误差为 0.321 1 和 0.439 4，Isomap-SVR 模型次之，PCA-SVR 模型的建模效果最差，其校正集和验证集的模型决定系数 R^2 分别为 0.940 4 和 0.892 4，对应的均方根误差为 0.441 5 和 0.537 0。

2.3.3 壮果促梢期柑橘叶片氮含量建模分析

根据表 7 的正交试验分析结果，选择最佳水平组合 $A_3B_3C_1$ ，即小波基函数为 sym8、分解层数是 5、阈值方案是 rigrsure 规则，对壮果促梢期内原始光谱

数据进行小波去噪。

表6 稳果期柑橘叶片氮含量预测模型建模结果

Tab.6 Modeling results of citrus leaves nitrogen content forecasting models at stability stage

光谱形式	回归方法	流形维数	校正集		验证集	
			决定系数	均方根误差	决定系数	均方根误差
			原始光谱	PCA-SVR	7	0.8753
	MDS-SVR	7	0.8685	0.6127	0.8380	0.6734
	LLE-SVR	7	0.8921	0.5665	0.8559	0.6460
	Isomap-SVR	7	0.9011	0.5464	0.8739	0.6027
	LE-SVR	7	0.8925	0.5664	0.8096	0.7285
去噪光谱	PCA-SVR	6	0.9404	0.4415	0.8924	0.5370
	MDS-SVR	6	0.9675	0.3211	0.9344	0.4394
	LLE-SVR	6	0.9479	0.4125	0.9249	0.4673
	Isomap-SVR	6	0.9569	0.3723	0.9283	0.4683
	LE-SVR	6	0.9575	0.3707	0.9102	0.5258

表7 壮果促梢期小波去噪正交试验的极差分析结果

Tab.7 Range analysis results of wavelet denoising orthogonal test at bloom stage

指标	因素		
	A	B	C
K_1	0.7331	0.7308	0.7452
K_2	0.7287	0.7353	0.7327
K_3	0.7457	0.7369	0.7325
K_4	0.7303	0.7348	0.7274
R	0.0170	0.0061	0.0178
优水平	A_3	B_3	C_1

将原始光谱数据和经小波去噪后的光谱数据分别采用上文中的5种流形学习算法进行特征提取,最终将特征数据导入支持向量机回归(Support vector regression,SVR)建立柑橘叶片氮含量预测模型,建模结果如表8所示。

表8 壮果促梢期柑橘叶片氮含量预测模型建模结果

Tab.8 Modeling results of citrus leaves nitrogen content forecasting models at bloom stage

光谱形式	回归方法	流形维数	校正集		验证集	
			决定系数	均方根误差	决定系数	均方根误差
			原始光谱	PCA-SVR	8	0.8781
	MDS-SVR	8	0.8533	0.6493	0.8178	0.7049
	LLE-SVR	8	0.8377	0.6768	0.8034	0.7276
	Isomap-SVR	8	0.8852	0.5629	0.8428	0.6630
	LE-SVR	8	0.8442	0.6610	0.7768	0.7659
去噪光谱	PCA-SVR	7	0.8794	0.5735	0.8446	0.6520
	MDS-SVR	7	0.8844	0.5619	0.8605	0.6357
	LLE-SVR	7	0.9147	0.5390	0.8954	0.5576
	Isomap-SVR	7	0.8968	0.5575	0.8711	0.5843
	LE-SVR	7	0.8849	0.5631	0.8506	0.6500

由表8可看出,在壮果促梢期内,LLE-SVR模型的建模效果最好,校正集和验证集的模型决定系数 R^2 分别为0.9147和0.8954,对应的均方根误差为0.5390和0.5576,Isomap-SVR模型次之,PCA-SVR模型的建模效果最差,其校正集和验证集的模型决定系数 R^2 分别为0.8794和0.8446,对应的均方根误差为0.5735和0.6520。

2.3.4 采果期柑橘叶片氮含量建模分析

根据表9的正交试验分析结果,选择最佳水平组合 $A_4B_4C_4$,即小波基函数为db4、分解层数是7、阈值方案是minimaxi规则,对采果期内原始光谱数据进行小波去噪。

表9 采果期小波去噪正交试验的极差分析结果

Tab.9 Range analysis results of wavelet denoising orthogonal test at picking stage

指标	因素		
	A	B	C
K_1	0.7168	0.7240	0.7282
K_2	0.7272	0.7257	0.7187
K_3	0.7338	0.7277	0.7297
K_4	0.7344	0.7347	0.7355
R	0.0176	0.0107	0.0168
优水平	A_4	B_4	C_4

分别采用上文中的5种流形学习算法对原始光谱数据和经小波去噪后的光谱数据进行特征提取,将特征数据导入支持向量机回归(Support vector regression,SVR)建立柑橘叶片氮含量预测模型,建模结果如表10所示。

表10 采果期柑橘叶片氮含量预测模型建模结果

Tab.10 Modeling results of citrus leaves nitrogen content forecasting models at picking period

光谱形式	回归方法	流形维数	校正集		验证集	
			决定系数	均方根误差	决定系数	均方根误差
			原始光谱	PCA-SVR	9	0.8150
	MDS-SVR	9	0.8285	0.6945	0.7792	0.7586
	LLE-SVR	9	0.8174	0.7046	0.7562	0.7845
	Isomap-SVR	9	0.8412	0.6631	0.8268	0.6846
	LE-SVR	9	0.8527	0.6562	0.7462	0.7978
去噪光谱	PCA-SVR	7	0.8728	0.5779	0.8535	0.6477
	MDS-SVR	7	0.8828	0.5465	0.8416	0.6632
	LLE-SVR	7	0.8817	0.5830	0.8619	0.6340
	Isomap-SVR	7	0.8965	0.5557	0.8779	0.6048
	LE-SVR	7	0.8634	0.6286	0.8349	0.6853

由表10可知,在采果期内,Isomap-SVR模型建立的橘叶片氮含量预测模型的建模精度最高,校正集和验证集的模型决定系数 R^2 分别为0.8965和

0.877 9, 对应的均方根误差为 0.555 7 和 0.604 8, LLE - SVR 模型次之, LE - SVR 模型的建模效果最差, 其校正集和验证集的模型决定系数 R^2 分别为 0.863 4 和 0.834 9, 对应的均方根误差为 0.628 6 和 0.685 3。

2.4 各生长期预测模型比较与分析

由于柑橘叶片营养成分会在生长过程中不断改变, 为了进行全程跟踪式光谱检测与建模分析, 本试验分批次单独对各个生长期进行建模分析, 建立不同生长期柑橘叶片氮含量预测模型。

将每个生长期氮含量最优预测模型参数及建模结果汇总记录, 如表 11 所示, 从中可看出, 每个生长期采用正交试验得到的最佳小波去噪参数组合各不相同, 萌芽期为“小波基函数 *coif2*、分解层数 7、阈值方案 *rigrsure*”, 稳果期为“小波基函数 *sym8*、分解层数 7、阈值方案 *heursure*”, 壮果促梢期为“小波基函数 *sym8*、分解层数 5、阈值方案 *rigrsure*”, 采果期为“小波基函数 *db4*、分解层数 7、阈值方案 *minimaxi*”。造成各个生长期的最佳小波去噪参数组合各不相同的原因是: 光谱数据在采集的过程中, 难免受到仪器随机噪声和环境背景噪声等因素的干扰, 不同批次

光谱数据所包含噪声的统计特性、频率分布和噪声结构等方面都存在一定的差异, 小波去噪参数会随着噪声特性的不同而调整, 因此出现了各个生长期最佳小波去噪参数组合各不相同的情况。

从表 11 中也可看出, 每个生长期的最佳回归预测模型也不尽相同, 萌芽期和采果期的最佳回归预测模型是 *Isomap - SVR*, 稳果期是 *MDS - SVR*, 壮果促梢期是 *LLE - SVR*, 其对应的流形维数也不尽相同。这是因为叶片光谱反射率受叶片内所有生物化学成分的综合影响, 不同生长期叶片内各种化学成分含量及其变化趋势都不一样, 也就是说, 光谱反射率除了受氮含量影响外, 还受叶片内其他化学成分的影响, 因此, 不同批次光谱数据的数据特征、统计特性及噪声分布等方面一定存在不同程度的差异, 所能提取的表征氮含量的特征波段、有效成分、内在特征等也不相同, 采用测地线最小生成树法计算出来的本征维数也不尽相同, 用流形学习算法提取的特征成分也不相同, 进而导致了各个生长期氮含量预测建模结果存在一定差异, 这说明了各个生长期需要分开单独建模的必要性, 此作法有利于提高建模精度。

表 11 不同生长期柑橘叶片氮含量最优预测模型参数及建模结果

Tab. 11 Parameters and results of optimal forecasting models of nitrogen content for citrus leaves during different growth stages

生长期	小波去噪最佳参数组合			最优预测模型		校正集		验证集	
	小波基函数	分解层数	阈值方案	建模模型	流形维数	决定系数	均方根误差	决定系数	均方根误差
萌芽期	<i>coif2</i>	7	<i>rigrsure</i>	<i>Isomap - SVR</i>	5	0.943 9	0.428 5	0.901 4	0.545 3
稳果期	<i>sym8</i>	7	<i>heursure</i>	<i>MDS - SVR</i>	6	0.967 5	0.321 1	0.934 4	0.439 4
壮果促梢期	<i>sym8</i>	5	<i>rigrsure</i>	<i>LLE - SVR</i>	7	0.914 7	0.539 0	0.895 4	0.557 6
采果期	<i>db4</i>	7	<i>minimaxi</i>	<i>Isomap - SVR</i>	7	0.896 5	0.555 7	0.877 9	0.604 8

3 结论

(1) 柑橘叶片在萌芽期内不断生长, 迅速吸收氮, 叶片氮含量不断提高, 在稳果期达到最高峰, 随着生长周期延长叶片氮含量逐渐下降, 叶片氮含量变化趋势从大到小依次为: 采果期、萌芽期、壮果促梢期、稳果期。在可见光区域, 光谱反射率随着柑橘叶片氮含量的提高而不断减小, 而在近红外区域, 光谱反射率随着柑橘叶片氮含量的提高而不断增大。

(2) 对原始光谱数据进行小波去噪处理能提高建模精度, 但不同的小波参数组合对小波去噪的效果有不同的影响, 故本文通过引入正交试验分析确定了萌芽期小波去噪参数应设定为“小波基函数 *coif2*、分解层数 7、阈值方案 *rigrsure*”, 稳果期为“小波基函数 *sym8*、分解层数 7、阈值方案 *heursure*”, 壮果促梢期应为“小波基函数 *sym8*、分解层数 5、阈值

方案 *rigrsure*”, 采果期应为“小波基函数 *db4*、分解层数 7、阈值方案 *minimaxi*”。

(3) 由于原始光谱数据都为高维矢量(本文中的光谱数据为 2 151 维), 不适于直接进行建模分析, 本文引入流形学习算法对原始光谱进行降维和特征提取, 结合支持向量机回归(SVR)进行建模分析, 比较了 *PCA - SVR*、*MDS - SVR*、*LLE - SVR*、*Isomap - SVR* 和 *LE - SVR* 这 5 种柑橘叶片氮含量预测模型的建模结果。试验结果表明, 这 5 种建模方法都能适用于柑橘叶片氮含量的估测, 萌芽期和采果期的最佳回归模型是 *Isomap - SVR*, 稳果期是 *MDS - SVR*, 壮果促梢期是 *LLE - SVR*, 验证集模型决定系数 R^2 分别为 0.901 4、0.877 9、0.934 4 和 0.895 4, 对应的均方根误差为 0.545 3、0.604 8、0.439 4 和 0.557 6。本研究对果园的精准管理具有积极意义。

参 考 文 献

- 1 李文庆,张民,束怀瑞. 氮素在果树上的生理作用[J]. 山东农业大学学报:自然科学版,2002,33(1):96-100.
Li Wenqing, Zhang Min, Shu Huairui. The physiological effects of nitrogen on fruit trees[J]. Journal of Shandong Agricultural University: Natural Science, 2002, 33(1): 96-100. (in Chinese)
- 2 张瑶,郑立华,李民赞,等. 苹果叶片氮素含量快速检测模型[J]. 农业机械学报,2012,43(增刊):300-304.
Zhang Yao, Zheng Lihua, Li Minzan, et al. Rapid determination model of apple leaf nitrogen content[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(Supp.): 300-304. (in Chinese)
- 3 章胜勇,祁春节. 我国柑橘鲜果滞销的原因及对策分析[J]. 华南农业大学学报:社会科学版,2008,7(4):75-83.
Zhang Shengyong, Qi Chunjie. Analysis of the causes and countermeasures against the slow sale of citrus fruit in China[J]. Journal of South China Agricultural University: Social Science Edition, 2008, 7(4): 75-83. (in Chinese)
- 4 Min M, Lee W S. Determination of significant wavelengths and prediction of nitrogen content for citrus[J]. Transactions of the ASAE, 2005, 48(2): 455-461.
- 5 Min M, Lee W S, Burks T F, et al. Design of a hyperspectral nitrogen sensing system for orange leaves[J]. Computers and Electronics in Agriculture, 2008, 63(2): 215-226.
- 6 易时来,邓烈,何绍兰,等. 锦橙叶片氮含量可见近红外光谱模型研究[J]. 果树学报,2010,27(1):13-17.
Yi Shilai, Deng Lie, He Shaolan, et al. Research on nitrogen content of leaf of Jincheng orange cultivar using visible near infrared spectroscopy model[J]. Journal of Fruit Science, 2010, 27(1): 13-17. (in Chinese)
- 7 黄双萍,洪添胜,岳学军,等. 基于高光谱的柑橘叶片氮素含量多元回归分析[J]. 农业工程学报,2013,29(5):132-138.
- 8 李金梦,叶旭君,王巧男,等. 高光谱成像技术的柑橘植株叶片含氮量预测模型[J]. 光谱学与光谱分析,2014,34(1):212-216.
- 9 张瑶,郑立华,李民赞,等. 基于光谱学原理与小波包分解技术预测苹果树叶片氮素含量[J]. 农业工程学报,2013,29(增刊1):101-108.
- 10 黄双萍,洪添胜,岳学军,等. 基于高光谱的柑橘叶片磷含量估算模型实验[J]. 农业机械学报,2013,44(4):202-207,195.
Huang Shuangping, Hong Tiansheng, Yue Xuejun, et al. Hyperspectral estimation model of total phosphorus content for citrus leaves[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(4): 202-207, 195. (in Chinese)
- 11 Hamm J, Lee D. Separating pose and expression in face images: a manifold learning approach[J]. Neural Information Processing—Letters and Reviews, 2007, 11(4): 91-100.
- 12 汪成龙,李小昱,武振中,等. 基于流形学习算法的马铃薯机械损伤机器视觉检测方法[J]. 农业工程学报,2014,30(1):245-252.
- 13 Jolliffe I. Principal component analysis[M]. New York: Springer-Verlag, 1986.
- 14 Borg I, Groenen P J. Modern multi-dimensional scaling: theory and application[M]. 2nd ed. New York: Springer-Verlag, 1997.
- 15 Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.
- 16 Tenenbaum J, Silva V, Langford J. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500):2319-2323.
- 17 Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396.
- 18 马本学,高国刚,王宝,等. 基于双树复小波变换和邻域操作的哈密瓜纹理提取[J]. 农业机械学报,2014,45(12):316-322.
Ma Benxue, Gao Guogang, Wang Bao, et al. Texture extraction of hami melon based on dual-tree complex wavelet transform and neighborhood operation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(12): 316-322. (in Chinese)
- 19 刘伟,王聪华,杨晓波,等. 基于LBV变换与小波变换的OLI图像融合方法[J]. 农业机械学报,2014,45(11):264-271.
Liu Wei, Wang Conghua, Yang Xiaobo, et al. Fusion of OLI image based on LBV transform and wavelet transform[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(11): 264-271. (in Chinese)
- 20 李天华,施国英,魏珉,等. 番茄维生素C含量近红外预测光谱的小波去噪[J]. 农业机械学报,2013,44(增刊1):200-204.
Li Tianhua, Shi Guoying, Wei Min, et al. Wavelet denoising in prediction model of tomato vitamin C content using NIR spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(Supp. 1): 200-204. (in Chinese)
- 21 刘伟,常庆瑞,郭曼,等. 土壤导数光谱小波去噪与有机质吸收特征提取[J]. 光谱学与光谱分析,2011,31(1):100-104.
Liu Wei, Chang Qingrui, Guo Man, et al. Extraction of first derivative spectrum features of soil organic matter via wavelet denoising[J]. Spectroscopy and Spectral Analysis, 2011, 31(1): 100-104. (in Chinese)
- 22 李民赞. 光谱分析技术及其应用[M]. 北京:科学出版社,2006.
- 23 Costa J A, Hero A O. Geodesic entropic graphs for dimension and entropy estimation in manifold learning[J]. IEEE Transactions on Signal Processing, 2004, 52(8): 2210-2221.