

基于机器学习和可见光光谱的冬小麦叶片氮积累量估算

崔日鲜¹, 刘亚东¹, 付金东^{2*}

1. 青岛农业大学农学与植物保护学院/山东省旱作农业技术重点实验室, 山东 青岛 266109

2. 中国农业科学院作物科学研究所, 北京 100081

摘 要 在拔节期分4次采集了6个施氮水平下的冬小麦冠层图像, 同步进行取样并以凯氏定氮法测定叶片含氮量, 进而计算叶片氮积累量。利用随机森林算法分割冠层图像之后提取冠层覆盖度、可见光波段(R, G和B)三个分量及其衍生的5个色彩指数。以冠层覆盖度外加色彩指数、色彩分量的两种非线性回归, 以及人工神经网络、支持向量回归、随机森林3种机器学习算法建立了冬小麦叶片氮积累量的估算模型。结果表明利用色彩指数的非线性回归模型的估算精度稍低于其他方法, 而随机森林算法的拟合精度最高, 但存在明显的过拟合现象。其他三种方法, 即以冠层覆盖度及色彩分量为输入变量的非线性回归、支持向量回归和人工神经网络方法, 均具有较高的拟合精度和泛化性能。

关键词 冬小麦; 机器学习; 可见光; 色彩指数; 冠层覆盖度; 叶片氮积累量

中图分类号: S126; S512.1 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2016)06-1837-06

引 言

适量增施氮肥能提高作物产量及品质, 而过量施氮会造成产量及品质的下降, 同时由于更多的氮素进入大气、水体加剧气候变暖和水体富营养化。Rockström等^[1]指出全球氮循环通量已经远超其安全边界, 其中很大一部分来源于农业生产消耗。获取作物营养状态是精准农业进行施肥管理的关键步骤。传统的植物组织分析法需要进行破坏性取样和大量的实验室分析。基于光谱学原理的地面遥感技术中, 叶绿素计(SPAD仪)的测量面积较小, 且对过量施肥不敏感, 而GreenSeeker和Yara N-sensor等则对氮饱和状态不敏感, 而且价格较高^[2]。基于作物冠层高光谱反射率的航空、卫星遥感技术则适合于大尺度的作物长势和氮素营养状态监测。

近年来, 数码相机作为可见光光谱地面遥感工具的一种, 因其高效、经济等特点, 被广泛应用于作物生长状况监测研究领域。Lee等^[3]利用逐步回归方法拟合了基于冠层覆盖度和色彩指数的水稻地上部氮积累量估算模型, 表现出较高的估算精度。Li等^[4]和Jia等^[5]分别发现小麦、棉花冠层中提取的冠层覆盖度与从GreenSeeker和Yara N-sensor传感器获取的归一化差异植被指数NDVI显著相关, 与地上部氮素含量呈非线性相关。Tewari等^[6]利用水稻冠层图像提取

的多个色彩指数分别拟合了地上部含氮量的非线性回归方程。张立周等^[7]发现六叶期夏玉米多个冠层图像指数与叶片全氮含量存在极显著相关关系。陈积山等^[8]研究得出苜蓿现蕾期冠层图像的绿光值与叶片SPAD值、植株全氮量呈极显著相关。多数研究利用冠层图像的可见光波段R, G, B分量构建色彩指数, 然后与叶片SPAD值、地上部氮含量、氮积累量等拟合回归方程并从中选择最敏感的色彩指数。

随着计算机的计算能力提升和数据挖掘技术的发展, 机器学习方法在作物农学参数反演研究中的应用逐渐增多^[9-13]。王大成等^[10]和夏天等^[11]利用高光谱数据分别比较了传统回归模型与人工神经网络估算生物量和叶面积指数的效果, 发现人工神经网络较回归分析显著提高了估算精度。梁亮等^[12]和梁栋等^[13]分别利用支持向量回归方法实现了小麦冠层叶绿素含量、叶面积指数的遥感反演。前人大多利用机器学习方法进行作物农学参数的高光谱反演, 而基于作物冠层图像R, G, B分量和色彩指数的农学参数估算方面的研究报道较少。另外, 有关机器学习算法, 如人工神经网络、支持向量回归、随机森林算法等的泛化性能方面的报道也较少。

以冬小麦为研究对象, 利用拔节期叶片氮积累量和由同步采集的冠层图像获取的冠层覆盖度、色彩分量及色彩指数的相关性, 以非线性回归方法、人工神经网络、支持向量回

收稿日期: 2015-04-02, 修订日期: 2015-08-18

基金项目: 农业部(948)项目(2012-Z5)和山东省旱地作物水分高效利用科研创新团队项目资助

作者简介: 崔日鲜, 1963年生, 青岛农业大学农学与植保学院博士 e-mail: chis@qau.edu.cn

* 通讯联系人 e-mail: fujindong@caas.cn

归和随即森林算法分别建立冬小麦拔节期叶片氮积累量的估算模型，并比较估算精度及泛化性能，以期为基于冠层图像分析的冬小麦及其他作物长势监测提供理论依据和方法上的借鉴。

1 实验部分

1.1 田间试验

试验于 2013 年 10 月—2014 年 5 月在青岛农业大学科技示范园试验站进行，试验站位于青岛市胶莱镇。试验地土壤有机质含量 $15.2\text{ g}\cdot\text{kg}^{-1}$ ，碱解氮 $72.67\text{ mg}\cdot\text{kg}^{-1}$ ，有效磷 $26.38\text{ mg}\cdot\text{kg}^{-1}$ ，有效钾 $116.95\text{ mg}\cdot\text{kg}^{-1}$ 。供试品种为济麦 22；试验设 6 个施氮水平，分别为 0，60，120，180，240，300 $\text{kg N}\cdot\text{hm}^{-2}$ 。磷肥、钾肥施用量分别为 $100\text{ kg}\cdot\text{hm}^{-2}\text{ P}_2\text{O}_5$ 和 $120\text{ kg}\cdot\text{hm}^{-2}\text{ K}_2\text{O}$ ，氮、磷、钾肥全部用作基肥施入。小麦行距 18 cm，基本苗 $3.5\times10^6\cdot\text{hm}^{-2}$ 。小区面积为 $80\text{ m}^2(10\text{ m}\times8\text{ m})$ ，随机区组设计，重复 3 次。

1.2 测定项目

分别在 2014 年 3 月 27 日、4 月 6 日、4 月 13 日、4 月 20 日进行四次田间取样。将样本按茎、叶部位分开，并于 $105\text{ }^{\circ}\text{C}$ 下杀青 30 min 后于 $72\text{ }^{\circ}\text{C}$ 下烘干至恒重，称叶片干重后粉碎，用凯氏定氮法测定叶片含氮量，进而计算叶片氮积累量 (leaf nitrogen accumulation, LNA)。

1.3 小麦冠层图像采集与分割

1.3.1 小麦冠层图像采集

田间取样前，用数码相机 (佳能 60D) 拍摄小麦冠层图像，拍摄时间为 11:00—13:00。设定相机 ISO 感光度为 400，采用自动曝光模式，自动白平衡，多点自动对焦。拍摄时将自制的 $72.5\text{ cm}\times45\text{ cm}$ 的方格置于小麦冠层顶部，方格内有 4 行小麦 (见图 1)，相机垂直于地面，距冬小麦冠层顶部约 1 m，取景范围对准方格。用遮阳伞遮住直射太阳光以保证取景框内小麦冠层受光均匀。图像记录像素设置为 S-Raw，约 450 万 (2592×1728)，以 RAW 格式存储，共 72 张。图像每个像素点的亮度值是以 14 位存储的，对它的解码采用 dcraw v9.22 软件^[14]，设定如下选项：白平衡采用拍摄时设置的自动白平衡；去马赛克采用 AHD 算法；RGB 值转换为 sRGB 色彩空间并以 16 位 TIFF 格式保存。

1.3.2 小麦冠层图像分割

图像的分割采用随机森林算法。从图像中复制并拼接出两张 400×400 像素的小麦和土壤图像，读取其红光亮度值 (R, redness intensity)、绿光亮度值 (G, greenness intensity)、蓝光亮度值 (B, blueness intensity) 分量并转换为 CIEL^{*}a^{*}b^{*} 色彩空间的 L^{*}, a^{*}, b^{*} 值，将其作为自变量用 R 软件中的 RandomForest 包^[15] 训练一个含 100 棵分类树的随机森林对图像进行分割，其分割效果见图 1。

1.3.3 冠层图像色彩指数

分割后的图像中提取冠层覆盖度 (CC, Canopy Cover) 及小麦所占像素 R, G, B 分量平均值，然后计算表 1 中所示的各指数。方便起见，将 R, G, B 分量称为色彩分量，由 R, G, B 分量所构造的各个指数称为色彩指数。

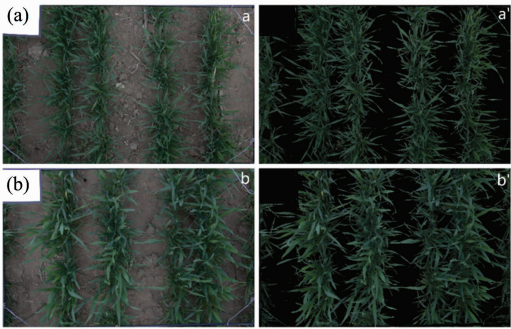


图 1 冬小麦冠层数码图像及分割后的冬小麦图像
(a): 处理 N120 的冠层图像，拍摄时间为 3 月 27 日；(a'): (a) 图像的冬小麦分割图像；(b): 处理 N180 的冠层图像，拍摄时间为 4 月 6 日；(b'): (b) 图像的冬小麦分割图像

Fig. 1 Digital camera images of winter wheat canopy, and segmented winter wheat plant images
(a): Image of treatment N120 taken on March 27; (a'): Segmented image of the image (a); (b): Image of treatment N180 taken on April 6; (b'): Segmented image of the image (b)

表 1 冠层覆盖及色彩指数的计算方法
Table 1 Methods for calculating canopy cover and color indices of winter wheat digital canopy image

色彩指数	计算式	参考文献
CC	$CC = \frac{\text{冬小麦图像像素数}}{\text{冠层图像总像素数}}$	[3]
GMR	$GMR = G - R$	[16]
GTR	$GTR = G / R$	[17]
EXG	$EXG = 2G - R - B$	[18]
<i>g</i>	$g = G / (R + G + B)$	[19]
NDI	$NDI = (R - G) / [R + G + 0.01 \times (R + G + B)]$	[3]

1.4 模型的构建

共获取 72 组数据，其中，剔除 1 组明显离群点数据后，根据叶片氮积累量分层选取 21 组数据用于模型验证，其余 50 组数据用于模型构建。模型的构建利用 R 软件中的 caret 包进行。caret 包提供了一个统一的接口，在模型训练时可以调用对应的函数包，同时具有数据预处理、模型参数优选、变量重要性估计等功能^[20]。为验证模型的泛化性能，训练时分别对模型参数的各个候选值进行 20 次 5 折交叉验证，具体为，首先将数据集等分为 5 个子集，将其中的 4/5 作为训练集，1/5 作为测试集，重复 20 次，记录共 100 次测试的均方根误差数据。最后选择平均均方根误差最小的待选值作为模型参数拟合全部的数据集。利用拟合模型的均方根误差 (RMSE)、相对均方根误差 (RRMSE) 及决定系数 (R^2) 的大小来评估模型拟合程度，模型的交叉验证及外部测试的均方根误差评估模型的泛化性能。

1.4.1 非线性回归

选用高斯-牛顿法^[21] 进行非线性最小二乘回归，该法通过多次迭代来修正模型回归系数使残差平方和达到最小。利用不同的自变量拟合两个系列非线性回归模型，各参数的初始值均由线性变换后最小二乘回归的回归系数计算得到。两

个系列非线性回归模型为：(1)色彩指数回归：所构造的色彩指数包含 2 个或 3 个可见光波段(R, G, B 色彩分量)的信息，将其分别与冠层覆盖度 CC 一起作为 5 个自变量组合；(2)色彩分量回归：将 R, G, B 色彩分量与冠层覆盖度 CC 作为自变量，先用最优子集法进行变量选择以消除自变量间的共线性，选择调整 R^2 最大的组合 CC, G, B 进行非线性最小二乘拟合(NLS)。

1.4.2 机器学习算法

共采用三种不同的机器学习算法，分别为人工神经网络、支持向量回归、随机森林进行拟合。建模时直接用 CC, R, G, B 作为自变量。

人工神经网络通常含输入层、一个或多个隐含层和输出层。通过把输入层的值加权平均传送到隐含层，由输出层输出，然后将误差反向传播以修正权重，如此反复进行直至误差在可接受范围内，从而实现变量的分类或回归。利用 caret 调用 nnet 包^[22]训练单隐含层人工神经网络(ANN)。设定最大迭代次数为 1 000，为避免过拟合采用权重衰减方法，权值衰减常量和隐含层神经元数通过交叉验证确定，候选值分别设为 {0.002, 0.003, 0.004, 0.005, 0.006}, {1, 2, 3, 4, 5}。

支持向量机通过引入核函数将变量映射到高维空间可以实现非线性分类和回归。调用 kernlab 包^[23]训练支持向量回归机(SVR)。训练前对自变量进行标准化预处理，使其均值为 0，方差为 1。核函数选用多项式函数，其中 offset 设为 1。参数 degree、scale 以及惩罚系数 cost 通过交叉验证确定，候选值分别为 {1, 2, 3}, {0.001, 0.01, 0.1}, {0.25, 0.5, 1}。

随机森林算法^[24-25]是 Leo Breiman 和 Adele Cutler 在

2001 年提出的，可用于分类和回归，比它以前所有的方法都精确^[26]。随机森林用于回归时用户自行设定回归树数目 n ，用 bootstrap 重抽样法从训练集中抽取 n 个样本，每个样本训练出一棵回归树，在回归树的每个节点从 m 个(小于自变量个数)随机抽取的自变量中选择分割点对变量空间进行递归二叉分割。训练好的模型用于预测时，每棵回归树单独进行预测，以 n 个预测值的平均值作为最终结果。调用 RandomForest 包^[15]训练随机森林(RF)。回归树数目 n 设为 500。 m 通过交叉验证确定，候选值设为 {1, 2, 3}。

2 结果与讨论

2.1 冬小麦拔节期叶片氮积累量与冠层覆盖度、色彩参数的相关性分析

表 2 为冬小麦拔节期叶片氮积累量与冠层覆盖度等变量的相关分析结果。由表 2 可知：拔节期叶片氮积累量与冠层覆盖度(CC)呈极显著正相关；与可见光波段的红光亮度值(R)、绿光亮度值(G)呈极显著负相关，而与蓝光亮度值(B)呈正相关，但未达到显著水平；叶片氮积累量与冠层图像色彩指数间均呈极显著相关，其中，与差值指数(GMR)、比值指数(GTR)和归一化差值指数(NDI)呈正相关，而与过绿化指数(GTR)和标准化绿光指数(g)呈负相关。相关程度来看，与 CC 的相关性最高，相关系数为 0.889；与 R, G, GTR, EXG 和 NDI 的相关系数也均大于 0.55，达到极显著水平。上述结果表明，由冠层图像分析获取的 CC, R, NDI, EXG 等能较准确地反映冬小麦拔节期叶片氮积累量的变化，即以冠层覆盖度等为自变量的冬小麦拔节期叶片氮积累量估算模型的构建具有一定的可行性。

表 2 叶片氮积累量与冠层覆盖度、色彩分量和色彩指数的相关系数
Table 2 Correlation coefficient between LNA and canopy cover, color components, and color indices

项目	冠层覆盖度 CC	色彩分量			色彩指数				
		R	G	B	GMR	GTR	EXG	g	NDI
LNA	0.889**	-0.627**	-0.451**	0.198	0.385**	0.557**	-0.561**	-0.307**	0.567**

* *: Significant at 0.01 level

2.2 叶片氮积累量估算模型的构建及验证

表 3 为以冠层覆盖度和色彩指数及色彩分量为自变量的非线性回归模型的构建、交叉验证及模型验证结果。由表 3 可知，模型构建和交叉验证的决定系数均大于 0.85，模型验证的决定系数也均大于 0.75，说明基于冠层覆盖度及色彩分量、色彩指数的非线性回归模型能准确地估算拔节期叶片氮积累量。六种非线性回归估算模型中，基于冠层覆盖度及可见光波段 G 和 B 分量的估算模型的估算精度稍高于基于冠层覆盖度及五种色彩指数的估算模型的估算精度。这可能与本研究中色彩指数的计算是由分割后的冠层图像色彩分量有关，色彩指数的构建有消除背景影响及突出特定(目标)信息方面的考虑^[27]，而本研究的图像色彩指数是由分割后的小麦图像可见光波段 R, G, B 分量计算得到，即已经排除了土壤背景的影响。

以冠层覆盖度 CC 和可见光 R, G, B 色彩分量作为自变量，采用三种机器学习算法，即人工神经网络(ANN)、支持向量回归(SVR)和随机森林算法(RF)构建冬小麦拔节期叶片氮积累量的估算模型。通过交叉验证确定的人工神经网络隐含层神经元个数为 2，支持向量回归的 degree、scale 及惩罚系数 cost 的组合为 {3, 0.1, 1}，随机森林的 m 值为 3。表 4 为模型构建、交叉验证及外部测试的结果，为便于比较将估算精度最高的非线性回归(NLS)模型的结果也列于表 4。图 2 为四种估算模型的交叉验证 RMSE 的密度曲线，图 3 为四种模型的预测值与实测值的 1 : 1 关系图。

由表 4 可知，基于随机森林算法的估算模型 RMSE 和 RRMSE 最小， R^2 最大，但其交叉验证和外部测试的 RMSE 和 RRMSE 最大， R^2 最小，表明基于随机森林算法的估算模型的构建存在过拟合现象。四种估算模型交叉验证 RMSE 的

表 3 基于非线性回归的叶片氮积累量估算模型的建立、交叉验证及模型验证结果

Table 3 Calibration, cross-validation and validation results of the nonlinear least squares regression models for estimating winter wheat leaf nitrogen accumulation

回归方程	模型建立			交叉验证		模型验证		
	RMSE	RRMSE/%	R ²	RMSE	R ²	RMSE	RRMSE/%	R ²
$LNA=70.667\times CC^{1.576}\times GMR^{0.658}$	0.801	19.226	0.852	0.830	0.852	0.983	23.775	0.776
$LNA=5.430\times CC^{1.548}\times GTR^{2.230}$	0.713	17.111	0.883	0.728	0.883	0.909	21.976	0.809
$LNA=0.940\times CC^{1.560}\times EXG^{-1.305}$	0.712	17.094	0.883	0.744	0.889	0.888	21.466	0.818
$LNA=0.212\times CC^{1.632}\times g^{-4.655}$	0.780	18.724	0.860	0.813	0.861	0.906	21.897	0.810
$LNA=54.428\times CC^{1.521}\times NDI^{0.868}$	0.709	17.022	0.884	0.726	0.884	0.893	21.585	0.816
$LNA=4.075\times CC^{1.529}\times G^{-2.283}\times B^{1.238}$	0.655	15.719	0.901	0.692	0.901	0.804	19.440	0.851

表 4 基于非线性回归和机器学习算法的叶片氮积累量估算模型的建立、交叉验证及模型验证结果

Table 4 Calibration, cross-validation and validation results of the models for estimating winter wheat leaf nitrogen accumulation

模型	模型建立			交叉验证		模型验证		
	RMSE	RRMSE/%	R ²	RMSE	R ²	RMSE	RRMSE/%	R ²
NLS	0.655	15.719	0.901	0.692	0.901	0.804	19.440	0.851
ANN	0.628	15.063	0.909	0.692	0.897	0.819	19.820	0.845
SVR	0.634	15.212	0.907	0.700	0.897	0.773	18.698	0.862
RF	0.415	9.960	0.960	0.876	0.850	0.956	23.123	0.788

密度曲线(图 2)也反映了这一特征,即随机森林算法交叉验证 RMSE 的分布较其他三种相比明显在右侧。

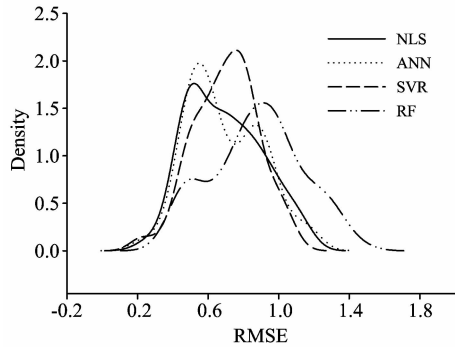


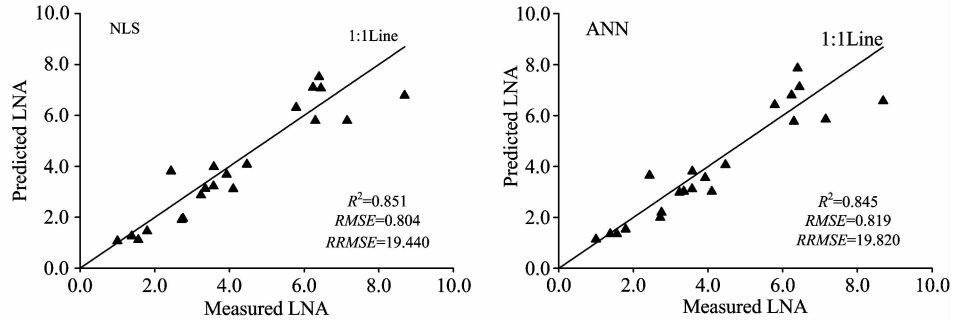
图 2 基于非线性回归和机器学习算法的叶片氮积累量估算模型的交叉验证均方根误差的相对频率

Fig. 2 Relative frequency of root mean square error (RMSE) of cross-validation for the methods of NLS, ANN, SVR, and RF

其他三种估算模型(NLS, ANN 和 SVR)的模型构建及交叉验证方面,基于 ANN 的估算模型建模集的 RMSE 和 RRMSE 最小, R² 最大; NLS 的交叉验证 RMSE 最小, R² 最大; 基于 SVR 估算模型的交叉验证 RMSE, R² 最大,但三种方法间的差异较小。模型验证方面,由表 4 的模型验证结果和 1:1 关系图(图 3)可知,基于 SVR 的模型验证的 RMSE、RRMSE 最小,决定系数最大,但三种方法的差异也较小。模型拟合、交叉验证和模型验证结果表明三种方法(NLS, ANN 和 SVR)均具有较高的估算精度和泛化性能。另外,机器学习算法的部分参数支持自行调试以获得更好的估算效果,本研究中的部分参数采用了软件默认值,因此 ANN 和 SVR 的估算精度还有提升的空间。

3 结 论

(1)冬小麦拔节期叶片氮积累量与由冠层数码图像分析获取的冠层覆盖度及部分色彩分量、色彩指数呈显著相关,其中与冠层覆盖度的相关性最强。



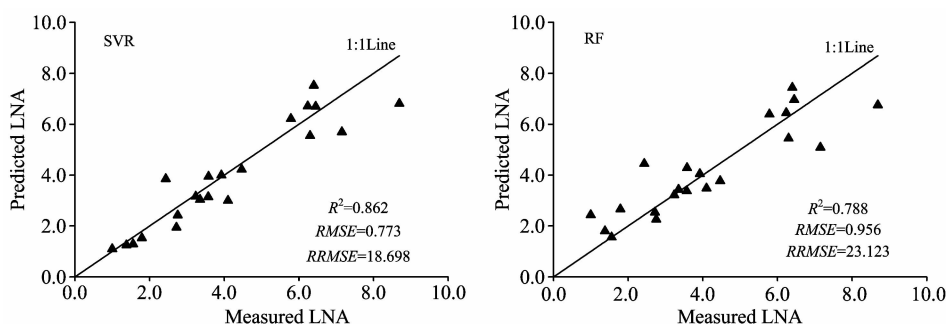


图3 基于非线性回归和机器学习算法的叶片氮积累量估算模型的预测值与实测值之间的关系

Fig. 3 Relationship between predicted leaf nitrogen accumulation based on the methods of NLS with CC, G, and B, ANN, SVR, and RF and measured leaf nitrogen accumulation during elongation stage

(2)基于冠层覆盖度及色彩分量的非线性估算模型较基于冠层覆盖度及色彩指数的非线性估算模型具有更高的估算精度,且两种方法均具有较高的泛化性能。

(3)基于随机森林算法的叶片氮积累量的估算存在明显的过拟合现象;基于NLS, ANN和SVR的叶片氮积累量的估算均具有较高的估算精度和泛化性能,均可用来构建基于冠层覆盖度和色彩分量的拔节期冬小麦叶片氮积累量的估算模型,或反演拔节期冬小麦叶片氮积累量。

线性回归模型中,利用变量选择方法在一定程度上能消

除变量间的多重共线性,从而获得较为稳定的回归方程,但同时也会损失部分信息。这些是在处理高维数据时所采取的折中的办法,而现代数据挖掘的很多算法能应对高维数据的非线性拟合问题,与传统方法相比具有相近或者更高的精度,而且鲁棒性也很强,不必对变量进行组合或剔除。但其缺点是有些机器算法无法生成显性的数学方程或规则,而是将变量之间的关系直接体现在算法当中,因此模型的可解释性较弱^[9]。

References

- [1] Rockstöm J, Steffen W, Noone K, et al. Ecology & Society, 2009, 14(2): 32.
- [2] Muñoz-Huerta R F, Guevara-Gonzalez R G, Contreras-Medina L M, et al. Sensors, 2013, 13(8): 10823.
- [3] Lee K J, Lee B W. European Journal of Agronomy, 2013, 48: 57.
- [4] Li Y, Chen D, Walker C N, et al. Field Crops Research, 2010, 118(3): 221.
- [5] Jia B, He H, Ma F, et al. The Scientific World Journal, 2014. 2014; doi: 10.1155/2014/602647.
- [6] Tewari V K, Arudra A K, Kumar S P, et al. Agricultural Engineering International: CIGR Journal, 2013, 15(2): 78.
- [7] ZHANG Li-zhou, WANG Dian-wu, ZHANG Yu-ming, et al(张立周, 王殿武, 张玉铭, 等). Chinese Journal of Eco-Agriculture(中国生态农业学报), 2010, 18(6): 1340.
- [8] CHEN Ji-shan, ZHU Rui-fen, GAO Chao, et al(陈积山, 朱瑞芬, 高超, 等). Acta Agrestia Sinica(草地学报), 2013, 21(3): 576.
- [9] Breiman L. Statistical Science, 2001, 16(3): 199.
- [10] WANG Da-cheng, WANG Ji-hua, JIN Ning, et al(王大成, 王纪华, 靳宁, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2008, 24(S2): 196.
- [11] XIA Tian, WU Wen-bin, ZHOU Qing-bo, et al(夏天, 吴文斌, 周清波, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2013, 29(3): 139.
- [12] LIANG Liang, YANG Min-hua, ZHANG Lian-peng, et al(梁亮, 杨敏华, 张连蓬, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2012, 28(20): 162, 294.
- [13] LIANG Dong, GUAN Qing-song, HUANG Wen-jiang, et al(梁栋, 管青松, 黄文江, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2013, 29(7): 117.
- [14] LI Feng, YIN Wei-wei(李峰, 印蔚蔚). Computer Engineering(计算机工程), 2011, 37(17): 211.
- [15] Liaw A, Wiener M R. News, 2002, 2: 18.
- [16] WANG Yuan, WANG De-jian, ZHANG Gang, et al(王远, 王德建, 张刚, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2012, 28(17): 131.
- [17] Adamsen F G, Pinter P J, Barnes E M, et al. Crop Science, 1999, 39(3): 719.
- [18] Woebbecke D M, Meyer G E, Von Bargen K, et al. Transactions of the ASAE, 1995, 38(1): 259.
- [19] ZHANG Li-zhou, HOU Xiao-yu, ZHANG Yu-ming, et al(张立周, 侯晓宇, 张玉铭, 等). Chinese Journal of Eco-Agriculture(中国生态农业学报), 2011, 19(5): 1168.
- [20] Kuhn M, Johnson K. Applied Predictive Modeling. New York: Springer, 2013.

- [21] Ritz C, Streibig J C. Nonlinear regression with R. Springer, 2008.
- [22] Venables W N, Ripley B D. Modern Applied Statistics with S. 4th ed. New York: Springer, 2002.
- [23] Karatzoglou A, Smola Am, Hornik K, et al. Journal of Statistical Software, 2004, 11(9): 1.
- [24] James G, Witten D, Hastie T, et al. An Introduction to Statistical Learning. New York: Springer, 2013.
- [25] Hastie T, Tibshirani R, Friedman J, et al. The Elements of Statistical Learning. New York: Springer, 2009.
- [26] WU Xi-zhi(吴喜之). Statistics(统计学). Beijing: Higher Education Press(北京: 高等教育出版社), 2008.
- [27] LIU Ji-ping(刘吉平). Remote Sensing Principle and Remote Sensing Information Analysis(遥感原理及遥感信息分析基础). Wuhan: Wuhan University Press(武汉: 武汉大学出版社), 2012.

Estimation of Winter Wheat Leaf Nitrogen Accumulation using Machine Learning Algorithm and Visible Spectral

CUI Ri-xian¹, LIU Ya-dong¹, FU Jin-dong^{2*}

1. College of Agronomy and Plant Protection, Qingdao Agricultural University, Shandong Provincial Key Laboratory of Dryland Farming Techniques, Qingdao 266109, China
2. Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Abstract In order to study the feasibility of using digital image analysis and machine learning algorithm to estimate leaf nitrogen accumulation (LNA) of winter wheat at canopy level, digital images of winter wheat canopies grown under six levels of nitrogen application rate were taken for four times during the elongation stage. Meanwhile, wheat plants were sampled to measure LNA. The random forest method using CIEL * a * b * components was used to segment wheat plant from soil background and then extract canopy cover, RGB components of sRGB color space and compute five color indices derived from RGB components. Correlation analysis was carried out to identify the relationship between LNA and canopy cover (CC), RGB components, and five color indices. Two kinds of nonlinear least squares regression models (NLS) with different independent variables of color components and color indices, and three machine learning algorithmic of artificial neural network (ANN), support vector regression (SVR), and random forests method (RF) were used to estimate winter wheat leaf nitrogen accumulation. All three machine learning algorithm had four input variables of CC, R, G, and B. The results showed that, CC, R and G component of sRGB color space, and five color indices derived from RGB components showed significant correlations with LNA during the elongation stage. CC revealed the highest correlation with LNA. The lowest accuracy in estimation LNA was achieved by using nonlinear least square model with CC and color indices, and RF had showed the problem of overfitting. The other three methods of LNA with CC and RGB components, ANN, and SVR had showed good performance with higher R^2 (0.851, 0.845, and 0.862) and lower RMSE (19.440, 19.820, and 18.698) for model calibration and validation, revealing good generalization ability.

Keywords Winter wheat; Machine learning algorithm; Visible spectrum; Color indices; Canopy cover; Leaf nitrogen accumulation

(Received Apr. 2, 2015; accepted Aug. 18, 2015)

* Corresponding author