

计算机视觉视域中水稻叶片叶绿素含量的数学建模

陈诚, 廖桂平*, 史晓慧, 李锦卫

(湖南农业大学 农业信息研究所, 湖南 长沙 410128)

摘要: 利用计算机视觉技术获取水稻叶片的颜色指标 DGCI、Hv、I2、I3、(2G-R-B)/L*和 Hv*Diff, 结合 BP 网络、多元回归模型和遗传算法, 建立叶绿素相对含量(SPAD 值)的预测模型, 对叶片 SPAD 值进行数学模拟, 模型的数学表达式为 $Y = \text{purelin}[W_5 * \text{tansig}(W_4 * X, B_4), B_5]$ 。利用所建立的模型对叶绿素相对含量(SPAD 值)进行预测, 结果与大田实测数据相对误差率仅为 3.355 7%。

关键词: 计算机视觉; 叶绿素相对含量; BP 神经网络; 多元回归; 遗传算法; 聚类分析; 水稻

中图分类号: S511; TP183 文献标志码: A 文章编号: 1007-1032(2011)05-474-05

The mathematical modeling of chlorophyll content in rice in view of computer vision

CHEN Cheng, LIAO Gui-ping*, SHI Xiao-hui, LI Jin-wei

(Agricultural Information Institute, Hunan Agricultural University, Changsha 410128, China)

Abstract: DGCI, Hv, I2, I3, (2G-R-B)/L* and Hv*Diff, which are the color index of leaves, were acquired by using computer vision technology. Combining those indexes with BP network, multiple regression models and genetic algorithms, the predictive model for chlorophyll relative content was established. The value of SPAD for leaves is simulated by mathematics, and the mathematical expressions of model is $Y = \text{purelin}[W_5 * \text{tansig}(W_4 * X, B_4), B_5]$. Using the established model to predict the value of SPAD, relative error rate is 3.355 7 % between the results and measured data in field.

Key words: computer vision; chlorophyll relative content; BP neural network; multiple regressions; genetic algorithm; cluster analysis; rice

叶绿素含量是作物生长各阶段的光合作用能力、营养胁迫和生长发育的良好指示剂。作物叶片叶绿素含量的高低, 既表明作物的生长状况, 又是作物与外界发生物质能量交换的重要条件; 因此, 实时监测作物叶片中叶绿素含量的变化, 成为实现作物长势监测和产量估算的重要手段^[1-2]。鉴于目前实时测定叶绿素含量仍存在器材昂贵、过程繁琐、不利于推广实施等问题, 笔者设想通过结合遗传算法的全局寻优特性、BP神经网络的非线性映射、自适应能力和学习能力以及多元线性回归在处理线性关系的优点, 建立叶绿素相对含量(SPAD值)

与叶色的数学模型, 测算叶绿素含量, 实现作物叶片叶绿素含量的自动诊断, 以实时、无损地获取作物生长信息。

1 材料及方法

1.1 水稻大田试验

大田试验于2009年在湖南省浏阳市永安镇进行。供试水稻汕优63, 由国家杂交水稻工程研究中心提供。以不同施N量构建不同群体冠层。采用顺序区组设计, 3次重复, 小区面积39 m²。设计6个施氮

收稿日期: 2011-02-23

基金项目: 国家自然科学基金项目(31071328); 湖南省研究生科技创新基金项目(CX2010B280)

作者简介: 陈诚(1986—), 男, 安徽蚌埠人, 硕士研究生, 主要从事智能信息处理研究, ccasxh@126.com; *通信作者, lgpxf@yahoo.com.cn

处理 :N0 ,全生育期不施氮肥 ;N1 ,施氮60 kg/hm² ; N2 , 120 kg/hm² ; N3 , 180 kg/hm² ; N4 , 240 kg/hm² ; N5 , 300 kg/hm²。全生育期按高产栽培技术进行田间管理 , 全程进行病虫害防治。移栽后从分蘖中期至齐穗期 , 每 7 d 对水稻叶片随机取样 , 用手持式叶绿素仪 (SPAD—502) 测定叶绿素相对含量。

1.2 叶片颜色指标的获取

移栽后从分蘖中期至齐穗期 , 每 7 d 对水稻叶片随机取样 , 移至光箱内拍照 , 利用 DIPS 1.0 (湖南农业大学农业信息研究所基于 Matlab R2006 开发) 进行图像处理 , 提取水稻冠层的各项颜色指标 , 拟

选取颜色指标为 :DGCI、Hv、I2、I3、(2G-R-B)/L*、Hv*Diff。在正式选取这 6 个指标为预测指标之前 , 需要对这 6 个指标进行两两间的相关性分析 , 以确定指标间是否存在显著相关性。在 SPSS18.0 平台上做 6 个指标间的相关性分析 , 发现 DGCI、Hv、I2、I3、(2G-R-B)/L* 和 Hv*Diff 之间不存在相关性 ; 因此 , 在室内颜色特征分析和田间颜色特征分析均可采用这 6 个颜色特征值作为预测依据。选取叶绿素相对含量 (SPAD 值) 为输出集。实验的训练样本容量为 84 (表 1) , 测试样本容量为 42 (表 2)。

表 1 训练数据

Table 1 Training data

编号	DGCI	Hv	I2	I3	(2G-R-B)/L*	Hv*Diff	大田 SPAD 值
1	0.449 16	82.079 70	0.136 36	0.148 49	0.007 14	36.157 59	38.00
2	0.456 95	82.383 39	0.129 46	0.160 56	0.006 94	36.934 36	40.20
3	0.448 40	81.842 90	0.129 68	0.136 74	0.006 99	30.817 84	34.00
4	0.432 25	80.266 68	0.139 45	0.137 32	0.007 05	33.436 27	36.10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
81	0.515 07	86.758 16	0.100 75	0.144 00	0.006 41	39.760 01	37.75
82	0.492 02	83.926 62	0.111 05	0.125 11	0.006 33	39.027 32	40.25
83	0.496 32	84.746 05	0.106 89	0.144 61	0.006 41	39.905 63	39.75
84	0.500 71	86.657 26	0.107 33	0.165 10	0.007 07	38.424 72	42.65

表 2 测试数据

Table 2 Test data

编号	DGCI	Hv	I2	I3	(2G-R-B)/L*	Hv*Diff	大田 SPAD 值
1	0.437 86	81.410 06	0.138 56	0.144 72	0.007 30	34.732 64	38.70
2	0.437 98	79.896 40	0.136 20	0.134 03	0.006 94	32.168 10	36.85
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
41	0.490 67	83.874 46	0.111 27	0.125 26	0.006 35	30.061 28	38.85
42	0.497 09	84.927 93	0.106 25	0.124 63	0.006 43	29.912 39	39.90

2 数学建模

将大田所测 SPAD 值和叶片颜色指标分成 S1 和 S2 两个部分 , S1 包含自变量 (叶片颜色指标) 与因变量 (SPAD) 呈线性关系的数据 , S2 包含自变量与因变量呈非线性关系的数据。以 S1 为数据集建立多元线性回归模型 , 以 S2 为数据集建立 BP 神经网络模型。再以 BP 神经网络为纽带 , 将建立的多元线性回归模型和 BP 神经网络模型融合在一起 , 利用遗传算法的全局寻优性对建立的杂合 BP 神经网络模型进行优化。

2.1 多元线性回归建模

以 S1 为处理数据 , 建立多元线性回归模型 ,

模型的数学表达式为 : $Y_1 = W_1 * X_1 + C$, 其中 W_1 为权值矩阵 , C 为常数项 , X_1 由因变量及与之呈线性关系的自变量组成。

2.2 BP 神经网络建模

以 S2 为处理数据 , 建立 BP 神经网络模型 , 网络输入层的节点数为自变量数 , 输出层的节点数为 1 (因变量数) , 隐含层的节点数根据训练样本的大小而定。隐含层的激活函数使用可导函数 tansig , 输出层使用线性函数 purelin。为保证网络的稳定性和精度 , 采用 0、3 为最终的学习速率。模型的数学表达式为 : $Y_2 = \text{purelin}[W_3 * \text{tansig}(W_2 * X_2, B_1), B_2]$, 其中 W_2 、 W_3 为权值矩阵 , B_1 、 B_2 为阈值矩阵 , X_2 由因变量及与之呈非线性关系的自变量组成^[3]。

2.3 杂合 BP 神经网络建模

设 BP 神经网络的输出值为 Y_1 ，多元线性回归模型的输出值为 Y_2 。以 Y_1 和 Y_2 为自变量，建立一个含有 2 个输入节点、3 个隐层节点、1 个输出节点的 BP 神经网络。隐含层的激活函数使用可导函数 $tansig$ ，输出层使用线性函数 $purelin$ ，采用 0、3 为最终的学习速率。杂合 BP 神经网络模型的数学表达式为： $Y=purelin[W_5* tansig(W_4*X, B_4), B_5]$ ，其中 W_4 、 W_5 为权值矩阵， B_4 、 B_5 为阈值矩阵， $X=\{Y_1, Y_2\}$ 。

2.4 杂合 BP 神经网络的遗传算法优化

将 BP 神经网络的所有初始权值归结为 1 个种群^[4]，设置遗传算法的遗传代数为 1 000，种群规模 100，适应度函数 $val=1/SE(SE$ 为 BP 网络输出值与真实值的误差平方和)。最终得到基于遗传算法优化的杂合 BP 神经网络模型。

2.5 试验数据的清理

在建模过程中，为加快网络的训练速度和模型精度，引入聚类分析^[5]的方法，对神经网络的输入和输出数据进行预处理，以达到剔除奇异数据的目的。

在 Matlab 软件平台上，通过系统聚类分析发现，在测试样本中，有 4 个叶节点所包含的个体属

于奇异数据，即第 20、21、23 和 28 个叶节点，输出这 4 个叶节点，发现它们包含的个体为第 20、21、23、24、62、63、65 和 66 个样本个体，应予以剔除，所以最终采用的训练样本数为 76 个。

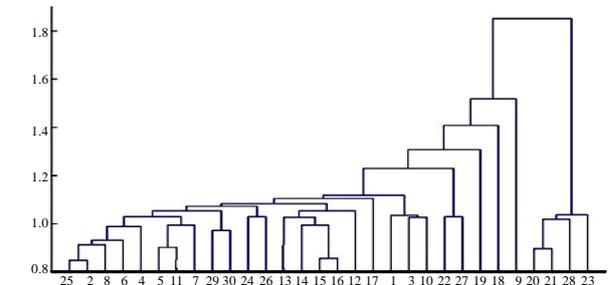


图 1 部分训练样本聚类结果
Fig. 1 Clustering results of training samples

2.6 数据集的相关性分析

在 SPSS 18.0 平台上分别作 SPAD 与 DGCI、SPAD 与 H_v 、SPAD 与 I_2 、SPAD 与 I_3 、SPAD 与 $(2G-R-B)/L^*$ 、SPAD 与 H_v*Diff 的相关性分析，以确定模型中的线性模型部分和非线性模型部分的变量构成，结果见图 2。作为自变量的 6 个颜色指标中， I_3 、 H_v*Diff 与 SPAD 显著相关，呈线性关系；DGCI、 H_v 、 I_2 、 $(2G-R-B)/L^*$ 则与 SPAD 呈非线性关系。

	VAR7	VAR1
VAR7 Pearson 相关性	1	.109
显著性(双侧)		.350
N	76	76
VAR1 Pearson 相关性	.109	1
显著性(双侧)	.350	
N	76	76
a		
	VAR7	VAR1
VAR7 Pearson 相关性	1	.080
显著性(双侧)		.492
N	76	76
VAR2 Pearson 相关性	.080	1
显著性(双侧)	.492	
N	76	76
b		
	VAR7	VAR1
VAR7 Pearson 相关性	1	-.068
显著性(双侧)		.560
N	76	76
VAR3 Pearson 相关性	-.068	1
显著性(双侧)	.560	
N	76	76
c		
	VAR7	VAR1
VAR7 Pearson 相关性	1	-.219
显著性(双侧)		.057
N	76	76
VAR5 Pearson 相关性	-.219	1
显著性(双侧)	.057	
N	76	76
d		
	VAR7	VAR1
VAR7 Pearson 相关性	1	.385**
显著性(双侧)		.001
N	76	76
VAR4 Pearson 相关性	.385**	1
显著性(双侧)	.001	
N	76	76
e		
	VAR7	VAR1
VAR7 Pearson 相关性	1	.347**
显著性(双侧)		.002
N	76	76
VAR6 Pearson 相关性	.347**	1
显著性(双侧)	.002	
N	76	76
f		

**在.01 水平(双侧)上显著相关。

**在.01 水平(双侧)上显著相关。

a SPAD 与 DGCI 的相关性；b SPAD 与 H_v 的相关性；c SPAD 与 I_2 的相关性；d SPAD 与 $(2G-R-B)/L^*$ 的相关性；e SPAD 与 I_3 的相关性；f SPAD 与 H_v*Diff 的相关性。

图 2 数据集的相关性分析结果

Fig. 2 The chart results of correlation analysis

3 建模仿真

以自变量 I3、Hv*Diff 和因变量 SPAD 值建立多元线性回归模型；分别以自变量 DGCI、Hv、I2、(2G-R-B)/L*和因变量 SPAD 为输入、输出，建立 BP 神经网络，因为训练数据的样本容量为 76 个，本着隐层节点数尽量接近样本容量的原则，选取隐层节点数为 75^[6]个。最后，用所建立的 2 个模型替

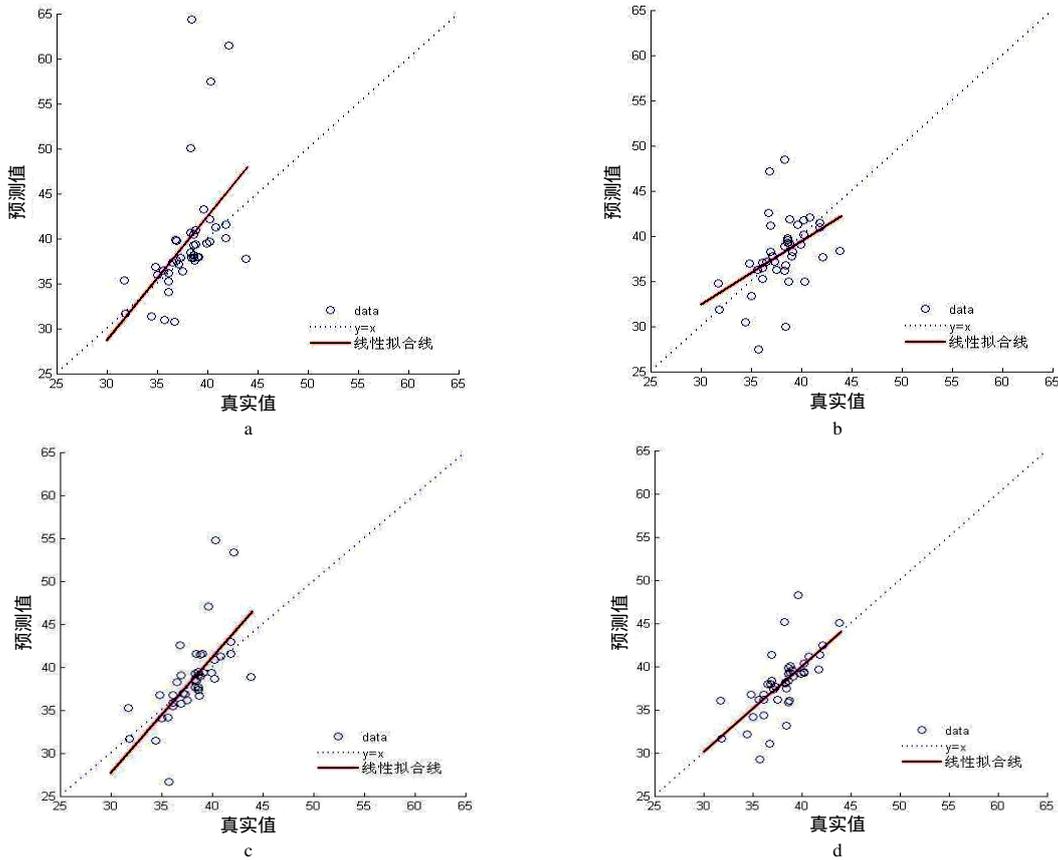
代最终模型中的相应部分。

设置网络的训练目标为总均方差(SSE)小于 0.01。以 MATLAB 7.0 作为系统仿真平台，进行建模仿真。统计各个模型的预测值与真实值之间的相对误差率(表 3)，分别绘制 4 种方法的输出值与真实值的拟合图(图 3)。

表 3 预测值与真实值的相对误差率

Table 3 Relative errors between predicted and actual values

数据集	相对误差率/%			
	BP 网络	杂合 BP 网络	GA-BP 网络	GA-杂合 BP 网络
训练数据	4.873 4	4.065 2	2.932 8	2.605 3
测试数据	7.943 5	6.872 0	4.832 2	3.355 7



a BP 神经网络；b 杂合 BP 神经网络；c 遗传算法优化的 BP 神经网络；d 遗传算法优化的杂合 BP 神经网络。

图 3 模型拟合结果

Fig.3 The fitting results

结合表 3 和图 4 可以看出，笔者所建立的预测模型，在选取 6 个颜色指标 DGCI、Hv、I2、I3、(2G-R-B)/L*、Hv*Diff 为影响因子的情况下，都能够较好地模拟水稻冠层的叶绿素相对含量，遗传算

法优化后的杂合 BP 神经网络算法的拟合线更接近于 Y=X，分布更加均匀、集中，拟合效果要明显优于其他 3 种模型^[7]。处理线性和非线性杂合数据时，模型的精度和泛化能力较以往的 BP 神经网络和遗

传算法优化的 BP 神经网络都有了一定的提高。

利用 Wilcoxon 符号秩检验^[8]分析新算法与传统算法之间的显著性差异,以确定方法间的优劣性,结果(表 4)表明,基于遗传算法优化的杂合 BP 网络模型明显优于其他 3 种方法,它们的优劣顺序为 GA-杂合 BP 网络、GA-BP 网络、杂合 BP 网络、BP 网络。

表 4 检验结果
Table 4 The results of test

	BP 网络	杂合 BP 网络	GA-BP 网络	GA-杂合 BP
BP 网络		-	-	-
杂合 BP 网络	+		-	-
GA-BP 网络	+	+		-
GA-杂合 BP	+	+	+	

参考文献:

- [1] 武仙山, 昌小平, 景蕊莲, 等. 小麦旗叶绿素含量及荧光动力学参数与产量的灰色关联度分析[J]. 作物学报, 2010, 36(2): 217-227.
- [2] LI Jin-wei, LIAO Gui-ping, OU Zhong-bin, et al. Rapeseed seeds classification by machine vision[C]// Proceedings of Workshop on Intelligent Information Technology Application. Zhangjiajie :Control Conference, 2007: 222-226.
- [3] 陈博, 欧阳竹. 基于 BP 神经网络的冬小麦耗水预测[J]. 农业工程学报, 2010, 26(4): 81-86.
- [4] 王金武, 葛宜元, 王金峰. 基于遗传算法的水稻整株秸秆还田埋草弯刀的设计与试验[J]. 农业工程学报, 2010, 26(1): 166-169.
- [5] 张建庭, 吕进, 刘辉军, 等. 基于聚类分析的局部建模方法在茶叶近红外光谱分析中的应用[J]. 理化检验——化学分册, 2010, 46(2): 125-129.
- [6] LIN Y C, ZHANG Jun, ZHONG Jue. Application of neural networks to predict the elevated temperature flow behavior of a low alloy steel[J]. Computational Materials Science, 2008, 43(4): 752-758.
- [7] SUN Hong-wei, FANG Jian-cheng, LI Jian-li. Temperature errors modeling for micro inertial measurement unit using multiple regression method[C]// Proceedings of the International Symposium on Intelligent Information Systems and Applications. Qingdao: Academy Publisher, 2009: 411-415.
- [8] Ruvie Lou Maria C Martinez, Joshua D Naranjo. A pretest for choosing between logrank and wilcoxon tests in the two-sample problem[J]. International Journal of Statistics, 2010, 68(2): 111-125.

责任编辑: 罗慧敏