

陈西亮, 张佳华, 艾天成. 基于支持向量机法提取江汉平原三湖农场棉蚜危害程度的空间分布[J]. 江苏农业科学, 2016, 44(9): 157–162.
doi:10.15889/j.issn.1002-1302.2016.09.045

基于支持向量机法提取江汉平原三湖农场棉蚜危害程度的空间分布

陈西亮^{1,2}, 张佳华^{2,1}, 艾天成³

(1. 长江大学地球科学学院, 湖北武汉 430100; 2. 中国科学院遥感与数字地球研究所, 北京 100094; 3. 长江大学农学院, 湖北荆州 434025)

摘要:棉蚜是棉花的主要害虫之一, 目前主要通过田间调查的方式进行测报。将 262 个调查点的棉蚜危害数据平均分成测试集和验证集, 用 3 种方法进行支持向量机分类器参数优化, 结果表明粒子群算法效果最好, 精度达到了 82.035 9%, 最优参数 $c=5.435\ 4$, $g=15.023\ 3$ 。用该参数构建分类模型来识别样点数据空间插值后区域内的蚜虫危害, 最后结合 TM 数据提取的棉田分布得出了江汉平原三湖农场棉蚜危害程度的空间分布, 取得了很好的效果, 对以后的作物虫害研究工作具有很好的指导意义。

关键词:棉蚜; 支持向量机; 粒子群算法; 江汉平原; 空间分布

中图分类号: S127; S435.622⁺.1 **文献标志码:** A **文章编号:** 1002-1302(2016)09-0157-06

棉花是人们重要的生活资料。目前, 棉花不仅仅是我国主要的纺织原料, 也是医学、化学以及国防工业的重要原料。因此, 棉花生产和粮食生产一样在国民经济中占有重要地位, 然而我国棉区辽阔, 自然条件千差万别, 耕作制度复杂, 引发的棉虫种类也越来越多。棉蚜作为最主要的害虫之一, 不仅造成了棉花减产, 同时也严重危害棉花的品质, 从而造成经济上的巨大亏损。棉蚜大数量高密度的繁殖现状除了与自身性质有关外, 还受温度、湿度、土壤养分等多种环境因素的影响。

收稿日期: 2015-09-21

基金项目: 国家自然科学基金(编号: 31571565)。

作者简介: 陈西亮(1991—), 男, 湖北十堰人, 硕士研究生, 从事农业遥感研究。E-mail: 1172113468@qq.com。

通信作者: 张佳华, 研究员, 博士生导师, 主要从事生态环境与全球环境变化遥感方面的研究。E-mail: jhzhang@ceode.ac.cn。

毒性基因 *v1*、*v3*、*v6*、*v9*、*vA*、*Yr3b+4b* 和 *vrSu* 出现频率均在 75% 以上, 与之对应的抗性基因 *Yr1*、*Yr3*、*Yr6*、*Yr9*、*YrA*、*Yr3b+4b* 和 *YrSu* 在贵州不能单独使用作为抗条锈种目标; 当前 CYR32、CYR33 和贵 22-14 在贵州省出现频率依次居前 3 位, 抗这些小种(类型)的小麦品种, 是贵州育种部门的主要育种方向。

为真实全面反映一个地区小麦条锈菌的群体组成, 条锈病标样的采集数量及地点, 是重要的因素。本试验在贵州主要小麦种植区毕节市和黔西南州采集的标样较多, 而在贵州的中部采集的标样较少, 可能会对监测结果有一定影响, 今后会在全境尽可能多地采集条锈病标样, 以期达到更加全面反映贵州条锈菌群体结构。

参考文献:

[1] 刘太国, 王保通, 贾秋珍, 等. 2010—2011 年度我国小麦条锈菌生理专化研究[J]. 麦类作物学报, 2012, 32(3): 574–578.

因此, 为了有效控制棉蚜的危害, 及时、准确、大面积监测和预测未知区域棉蚜的发生情况是有效展开防治工作的前提, 同时对避免棉蚜危害造成棉花减产有重要的作用。传统的棉蚜危害监测预报方法采用田间定点调查或随机调查的方式, 借助放大镜、显微镜等工具或直接肉眼判别棉蚜并统计数量, 这种方法虽然直观、简便, 但需要投入大量的人力和物力、效率低下, 并且调查点有限, 不能反映整个空间的分布。遥感技术和地面数据的结合正好能弥补这一缺陷, 能有效地实现大面积作物病虫害的监测。在国外, Muhammad 等利用卫星影像分析了小麦条锈病的空间影像特征, 将其从正常生长的区域分离出来, 实现了小麦病害的识别^[1]。Mirik 等对俄国麦蚜胁迫下的冬小麦冠层光谱反射特征进行了分析, 提出了蚜虫的准确估测还需依靠相应的光谱指数^[2]。我国学者卢小燕在棉花蚜虫危害主要生育期测试不同危害程度棉叶的光谱, 经过分析指出 434~727 nm 可作为棉叶蚜虫的敏感波段,

[2] 李振岐. 关于小麦品种抗条锈性丧失问题的初步探讨[J]. 西北农业学报, 1980(3): 83–91.

[3] 李振岐. 我国小麦品种抗条锈性丧失原因及其控制策略[J]. 大自然探索, 1998, 17(66): 21–24.

[4] 贾秋珍, 金社林, 曹世勤, 等. 小麦条锈菌生理小种条中 32 号及水源 14 致病类型在甘肃的流行与发展趋势[J]. 植物保护学报, 2007, 34(3): 264–267.

[5] Wan A M, Chen X M, He Z H. Wheat stripe rust in China[J]. Australian Journal of Agricultural Research, 2007, 58(6): 605–619.

[6] 韩德俊, 王琪琳, 张立, 等. “西北—华北—长江中下游”条锈病流行区系当前小麦品种(系)抗条锈病性评价[J]. 中国农业科学, 2010, 43(14): 2889–2896.

[7] 李振岐, 曾士迈. 中国小麦锈病[M]. 北京: 中国农业出版社, 2002: 292–294, 370–373.

[8] 左希, 蒋选利, 李星星, 等. 2009 年贵州小麦条锈菌生理小种的鉴定[J]. 贵州农业科学, 2011, 39(3): 91–93.

从国内外已有的研究情况来看,应用遥感手段研究作物病虫害主要有水稻二化螟、三化螟、稻飞虱、小麦蚜虫、东亚飞蝗以及地下害虫等。害虫侵害作物后,植被的生理或生化组分发生变化,直接表现出来的现象就是光谱发生变化,从而可以直接观测作物的光谱变化来分析病虫害情况。通过棉蚜危害及其环境因子的间接反演病虫害的研究并不多见,因此利用遥感数据和地面数据结合起来间接反演棉蚜危害的方法具有很大的研究潜力。本研究利用地面测量数据和遥感数据结合起来实现对湖北省三湖农场 2004 年棉蚜危害程度空间分布的反演,从而为更大地区棉蚜灾害空间反演提供了可靠的理论依据。

1.1 研究区概况

棉花产业化示范样板建设单位,是国家确定的长江流域优质专用棉生产基地。该地区属北亚热带季风湿润气候区,具有四季分明、热量丰富、光照适宜、雨水充沛、雨热同季、无霜期长等特点。全年日照时数 1 827 ~ 1 897 h、平均气温 16 ~ 16.4 °C、无霜期 246 ~ 262 d、平均降水量 900 ~ 1 100 mm,得天独厚的自然环境和气候为棉花生长提供了充足条件。

三湖农场地区棉花一般在4月中下旬育苗,5月出苗移栽,6月中下旬至7月初现蕾开花,7月末至8月初裂铃吐絮,9月裂铃吐絮收获,10月末至11月拔秆^[5]。根据棉花的生长过程、棉蚜发生时期的不同可将棉蚜分为苗蚜和伏蚜。苗蚜发生在出苗到现蕾以前,个体大,深绿色,适宜偏低温度,气温超过27℃时繁殖受到抑制,虫口迅速下降;伏蚜主要发生在7月中下旬到8月份,伏蚜即夏型蚜,黄绿色,体型小,适宜偏高的温度,在17~28℃下大量繁殖,当平均气温高于30℃时,虫口才迅速减退。棉蚜1年发生10~30代,具有繁殖速度快、适应性强、种群数量大、群聚性等特性。在适合的温度条件下,经过4.5 d就可以发育为成虫,成虫进行孤雌生殖,且繁殖量大,在适合的温度下1 d就可以产10多头,在植物旺盛生长季节、温度15~30℃条件下最适合蚜虫生长发育。

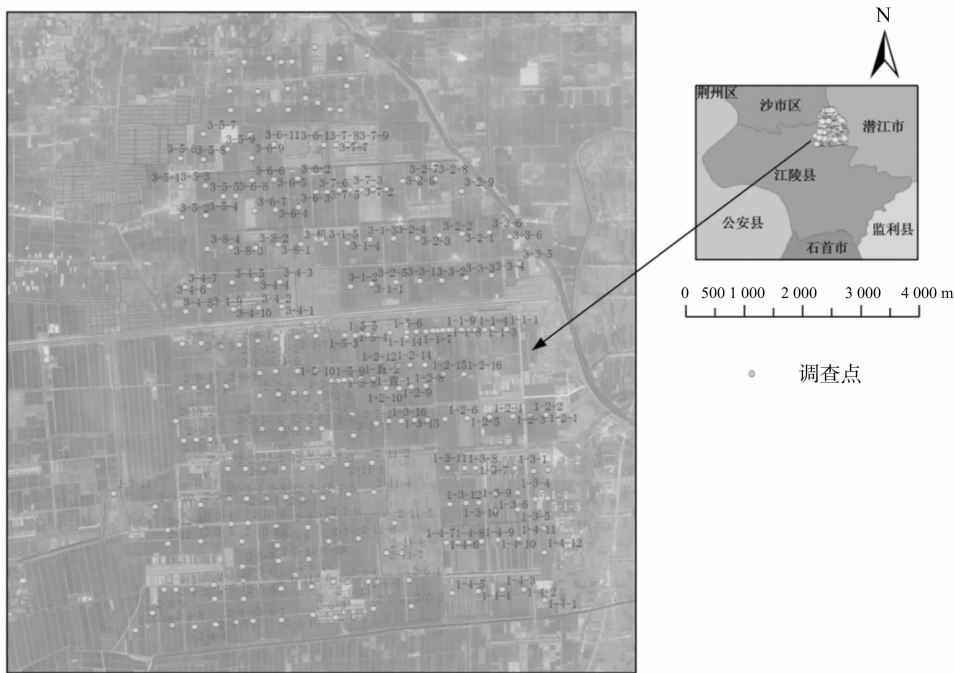


图1 调查点位置

本研究所使用的数据包括地面实测数据和遥感数据。地面数据收集是指各个地面样点一系列数据的收集,地面实测数据是2004年长江大学农学院的艾天成教授对三湖农场3 804 hm²土地的262个土样进行的养分普查数据,包括经纬度坐标、速效氮、速效磷、速效钾和有机质的含量,2004年8月进行的虫害调查数据。样点的NDVI数据和温度数据是长江大学农学院熊勤学教授用相应时相的MOD13Q1植被指数产品和MOD11标准地表温度产品提取的,另外还结合MODIS数据用SEBAL(the surface energy balance algorithm for land)模

型计算出研究区域地面蒸散,并提取样点的蒸散量,蒸散值的大小可以反映出棉田的相对湿度。还用了 TM 数据提取三湖地区棉花种植区域,由于缺乏 2004 年 8 月份有效的 TM 数据,并且 TM 数据仅仅用来提取棉花的种植范围,可以认为一个地区相邻的年份耕作制度不会有很大的变动,最终选用了 2003 年 9 月 2 日的一景覆盖三湖农场的 TM 数据。

1.4.1 调查点数据的空间插值 将调查点数据导入到 ArcGIS 中加上 UTM 投影用克里金插值法插值结果见图 2 至图 9。

从图 2 可以看出 2004 年 8 月份三湖区棉田的温度情况:

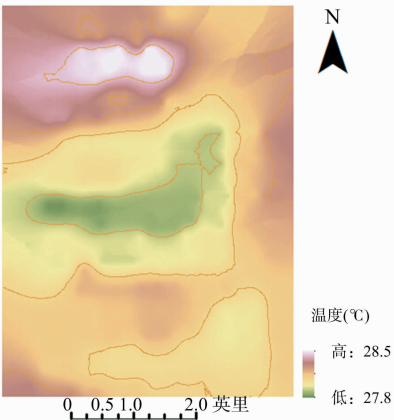


图2 温度分布

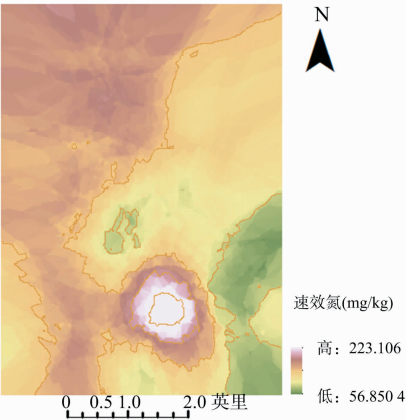


图3 速效氮分布

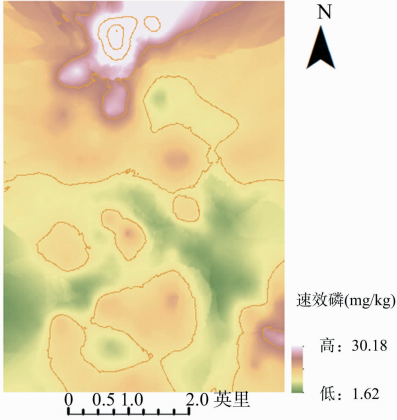


图4 速效磷分布

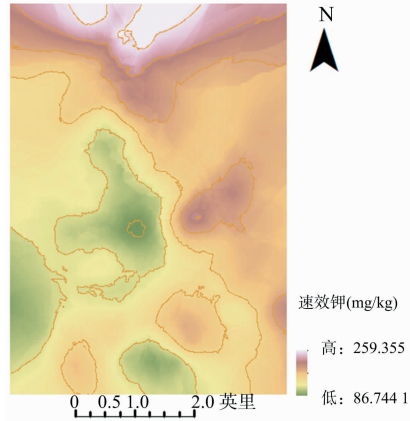


图5 速效钾分布

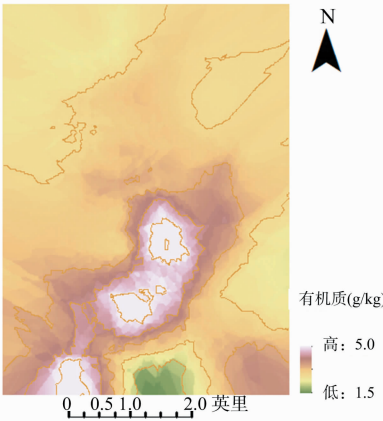


图6 有机质分布

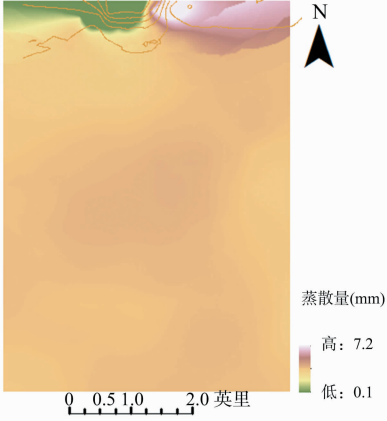


图7 蒸散量分布

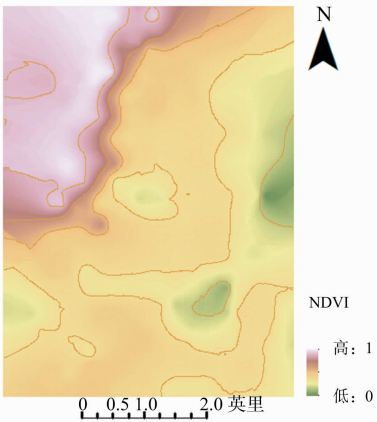


图8 NDVI分布

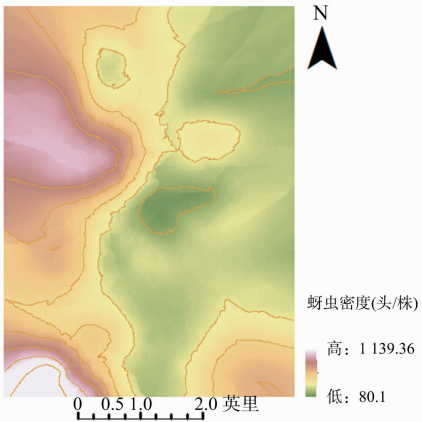


图9 蚜虫密度分布

整个三湖区的温度都稳定在 28.5 ℃ 左右,最低温度在 27.8 ℃ 左右,整体温度差异不大,这可能是因为研究区内水塘多,水体蒸发顺势带走部分热量使温度较周边低,或是研究区内土壤含水量高,水的热容量和热导率较土高,每升高 1 ℃ 需要吸收的热量多,因此不宜升温。从图 5 中能够清楚地发现调查区速效钾的分布状况:在调查区的西部、中部和南部土壤速效钾的含量都不高,中部有些地区甚至只有 50 mg/kg 左右,但在调查区最北部、东北部至东部都有明显的差异,含量较高。从图 7 可以看出调查区蒸散量的空间特征:图中一片深黄色平展开来,反映出了该区域水汽蒸发较均

匀,蒸散量是运用 SEBAL 模型计算出来的,它与土地净辐射通量、土壤热通量、辐射条件等都有密切的关系。NDVI 是应用最为广泛的用于指示植被生长状态和植被覆盖度的因子,从图 8 可以看出调查区植被生长的优劣,大部分地区植被覆盖度不高,而在调查区的西北部少部分地区作物覆盖度极好。图 9 反映了调查区的棉蚜危害程度空间分布特征:可以看出在调查区的中部、东部、南部大部分地区的棉蚜危害程度较轻,在调查区的东北部及西南部地区危害程度比较重,蚜虫密度可高达 1 000 头/株。

1.4.2 数据归一化处理 and 重采样 数据归一化处理是数据

挖掘的一项基础工作,不同评价指标往往具有不同的量纲和量纲单位,这样的情况会影响到数据分析的结果^[6]。为了消除指标间的量纲影响,需要进行数据归一化处理,以解决数据间的可比性。原始数据经过数据标准化处理后,各指标处于同一数量级,适合进行综合对比评价。由插值数据可以看出各种数据的分布范围极不均匀,并且单位也不一样。速效氮和速效钾的值比较大,分布范围也大,其他数据分布范围较小,因此要进行归一化处理。本研究用 min-max 标准化,使结果值映射到 0 和 1 之间。转换函数如下:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \circ$$

式中: x_{\max} 为样本数据的最大值, x_{\min} 为样本数据的最小值, x 为原始数据的值, x^* 为归一化后的值。本研究的调查点数据直接导入到 Matlab 中编程归一化,插值数据导入到 ENVI 中用波段运算代入上式归一化,为了保证插值数据和 TM 数据匹配,将插值数据的分辨率重采样到 30 m,投影也保持一致,最后将数据处理成 Matlab 可以识别的格式,并将样点数据平均分成 2 组,一组作为测试集,另一组作为验证集,代入进行参数优化。

1.4.3 棉蚜危害程度分级 参考前人的研究^[7],根据蚜虫密度将棉蚜危害程度分为 4 类:把蚜虫密度 < 200 头/株的点认定为危害程度轻,200 ~ < 500 头/株的点认定为危害程度中,500 ~ 800 头/株的点认定为危害程度较重,> 800 头/株的点认定为危害程度重。

2 研究方法

本研究将数据预处理后借助于 Matlab 的支持向量机加强版工具箱,用调查点数据进行支持向量机的参数最优化,然后用最优参数建立支持向量机模型,将插值的 7 个影响因子数据代入进行支持向量机分成轻、中、较重、严重 4 类。同时提取 TM 数据的 NDVI,经分析 NDVI 值大于 0.35 的像元为棉田,因此以 NDVI > 0.35 为阈值建立掩膜,将分类结果乘以掩膜数据就得到了三湖农场棉蚜危害程度的空间分布,掩膜外为建筑物和水体。

2.1 支持向量机分类法

支持向量机是由 Vapnik 首先提出的,可用于模式分类和非线性回归。支持向量机的主要思想是建立一个分类超平面作为决策曲面,使得正例和反例之间的隔离边缘被最大化。支持向量机的理论基础是统计学习理论,更准确地说,支持向量机是结构风险最小化的近似实现。

2.1.1 线性可分情形 SVM 算法是从线性可分情况下的最优分类面提出的。所谓最优分类面就是不仅能将 2 类样本点无错误地分开,而且要使分类间隔最大^[7]。

$$f(x) = w^T x + b. \quad (1)$$

式(1)为最优超平面,且最大几何间隔为 $2/\|w\|$,其中 w 是超平面的向量范数。标准的支持向量机学习目标是在特征空间中最大化类别之间的集合间隔,这就等价于最小化如下目标函数:

$$\min_w \left\{ \frac{1}{2} \|w\|^2 \right\}. \quad (2)$$

约束条件为: $y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, n$ 。

如果允许训练误差的存在,式(2)可以改写为:

$$\min_{w, \xi} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \right\} \quad (3)$$

约束条件为: $y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i > 0$ 。

式中: ξ_i 代表训练模式 x_i 的松弛变量, C 表示对误差的惩罚参数。然后定义 Lagrange 函数,转化成凸二次规划的对偶问题求出使上式成立的最优 w 和 b 的值,代入决策函数进行分类。

2.1.2 非线性可分情形 在实际应用中,通常数据在输入空间并不是线性可分的,然而如果原始数据通过非线性映射 $\phi(x)$ 可以被映射到高维特征空间,从而可以在新空间中定义 1 个超平面^[8-9]。对于支持向量机的对偶和原始表示,决策函数为测试样本和训练样本之间的内积组合形式。可以表示成:

$$K(x, x_i) = [\phi(x_i)^T \phi(x)]. \quad (4)$$

式(4)只包含待分类样本与训练样本中支持向量的内积运算,计算的复杂度并没有增加,可见通过核函数映射是解决支持向量机线性不可分问题的一种很好的方案。常用的核函数有:线性核函数、多项式核函数、径向基核函数和两层感知器核函数,还可以自定义核函数。选用不同的核函数可能会导致分类或回归的效果不一样,本研究选用径向基核函数,表达式如下:

$$K(x, x_i) = \exp[-\|x - x_i\|/(2\sigma^2)]. \quad (5)$$

由以上讨论可知要使用 Matlab 的支持向量机工具箱完成棉蚜危害识别,就要确定训练样本的惩罚参数 c 和核函数参数 g ,不同的 c 和 g 分类精度是不同的。为了找到适合本研究的最佳支持向量机模型参数,在后面的研究中使用了网格搜索法、遗传算法和粒子群算法进行参数优化。

2.2 支持向量机分类模型参数寻优

2.1.1 网格搜索法参数寻优 交叉验证是用来验证分类器性能的一种统计方法,基本思想是把原始数据进行分组,一部分作为训练集样本,另一部分作为验证集。用训练集对分类器进行训练得到分类模型,再用模型来识别验证集,将得到的分类精度作为分类器的性能指标^[10-12]。通常都是用 K-fold CV 将原始数据分成 K 组,将每个子集分别做 1 次验证,其余 $K-1$ 组数据作为训练集,这样就得到 K 个分类模型,用这 K 个模型最终验证集分类准确率的平均数作为分类器的性能指标, K 一般取大于 2 的值。网格搜索法的基本原理是让 c 和 g 在一定范围内划分网格并遍历网格内所有点进行取值,对于一组取定的 c 和 g 值,利用 K-fold CV 方法在此参数下验证分类的准确率,最终得到的分类精度最高的那组 c 和 g 作为分类的最优参数^[13-14]。本研究中 c 和 g 的取值范围为 $[2^{-8}, 2^8]$, K 取默认值 5, c 和 g 的步距设为 0.5,利用 Matlab 支持向量机工具箱的 SVMcgForClass 函数进行网格搜索法参数寻优。

2.1.2 粒子群算法参数寻优 粒子群算法(PSO)由 Kennedy 和 Eberhart 在 1995 年提出,该算法模拟鸟群飞行觅食行为,鸟类捕食时每只鸟找到食物最简单有效的方法是追随当前距离食物最近的鸟周围的区域。PSO 是从这种生物种群行为特征中得到启发并用于求解最优化问题的,算法中每个粒子都代表问题的 1 个潜在解,每个粒子对应着 1 个有适应度函数

决定的适应度值。粒子的速度决定了粒子移动的方向和距离,速度随自身及其他粒子的移动经验进行动态调整,从而实现可解空间中的寻优^[15-16]。本研究中的适应度值就是交叉验证分类的准确率。

假设在一个 D 维的搜索空间中,由 n 个粒子组成的种群 $X = (X_1, X_2, \dots, X_n)$, 其中第 i 个粒子表示为一个 D 维的向量 $X_i = (X_{i1}, X_{i2}, \dots, X_{in})^T$, 代表第 i 个粒子在 D 维搜索空间中的位置。根据目标函数即可计算出每个粒子位置 x_i 对应的适应度值。第 i 个粒子的速度为 $v_i = (v_{i1}, v_{i2}, \dots, v_{in})^T$, 其个体极值为 $P_i = (P_{i1}, P_{i2}, \dots, P_{in})^T$, 种群的全局极值为 $P_g = (P_{g1}, P_{g2}, \dots, P_{gn})^T$ 。在每一次迭代过程中,粒子通过个体极值和全局极值更新自身的速度和位置,更新公式为:

$$v_{in+1} = wv_{in} + c_1 r_1 (P_{in} - X_{in}) + c_2 r_2 (P_{gn} - X_{in}); \quad (6)$$

$$X_{in+1} = X_{in} + v_{in+1} \circ \quad (7)$$

式中: w 为惯性因子,调节对解空间的搜索范围。 r_1 和 r_2 是 2 个随机数,取值范围是 $(0, 1)$, c_1 和 c_2 是学习因子,经验取值 $c_1 = c_2 = 2$, 调节学习最大步长。

粒子群算法优化初始值进化代数设置为 100, 种群数量为 20, c 和 g 的范围为 $[0, 100]$, 粒子和速度初始化对初始粒子位置和粒子速度赋予随机值。初始化参数设置好后代入 `psoSVMcgForClass` 函数进行支持向量机分类器参数优化。

2.1.3 遗传算法参数寻优 遗传算法(GA)是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型,是一种模拟自然进化过程搜索最优解的方法,是由美国 Michigan 大学的 Holland 教授提出的。遗传算法模拟了自然选择和遗传中发生的复制、交叉和变异等现象,从任一初始群体出发,通过随机选择、交叉和变异操作,产生一群更适应环境的个体,使群体进化到搜索空间中越来越好的区域,这样一代一代地不断繁衍进化,最后收敛到一群最适应环境的个体,求得问题的最优解^[17-19]。

遗传算法是从代表问题可能潜在的解集的一个种群开始的,而一个种群则由经过基因编码的一定数目的个体组成,因此,第一步需要实现从表现型到基因型的映射,即编码工作。初代种群产生后,按照适者生存和优胜劣汰的原理,逐代演化产生出越来越好的近似解。在每一代,根据问题域中适应度的大小选择个体,并借助自然遗传学的遗传算子进行组合交叉和变异,产生出代表新的解集的种群。这个过程将导致种群像自然进化一样,子代种群比父代种群更加适应环境,末代种群中最优个体经过解码可作为问题近似的最优解。

遗传算法有 3 个基本操作:选择、交叉和变异。选择的目的是为了从当前群体中选出优良的个体,使它们有机会作为父代繁衍子孙。根据各个个体的适应度值,按照一定规则从上一代群体中选出一些优良的个体遗传到下一代,选择的依据是适应性强的个体为下一代贡献 1 个或多个后代的概率大。通过交叉操作可以得到新一代个体,新个体组合了父代的个体特性。将群体中各个个体随机搭配成对,对每一个个体,以交叉概率交换它们之间的部分染色体。对种群中每一个个体,以变异概率改变某一个或多个基因座上的基因值为其他的等位基因,同生物界中一样,变异发生的概率很低,变异为新个体的产生提供了机会。遗传算法参数寻优的初始值设置和粒子群算法一致,代入 `gaSVMcgForClass` 中进行参数

寻优。

3 结果与分析

3.1 参数寻优结果

图 10 是以 $\log_2 c$ 、 $\log_2 g$ 为 x 轴和 y 轴,以交叉验证分类精度为 z 轴的网格参数优化结果的 3D 图,精度值越大表示相应的参数 c 和 g 值越好。当验证分类精度为 80.814% 时,得到最优参数 $c = 16$, $g = 11.3137$ 。图 11 和图 12 都是以进化代数为横轴,适应度为纵轴的参数寻优适应度曲线。粒子群算法在进化代数为 18 左右时最佳适应度达到稳定,达到 82.0359%, 最优参数 $c = 4.0203$, $g = 20.6524$ 。遗传算法进化代数在 20 左右时最佳适应度达到稳定,最佳适应度最大值为 81.4371%, 最优参数 $c = 5.4354$, $g = 15.0233$ 。经比较可知粒子群算法参数寻优的精度最高,可将该组 c 和 g 值用于支持向量机分类器的构造。

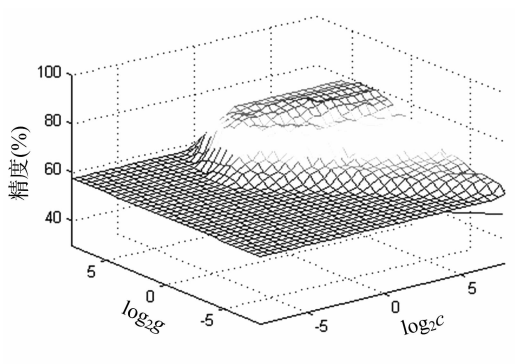


图10 网格搜索法参数寻优结果

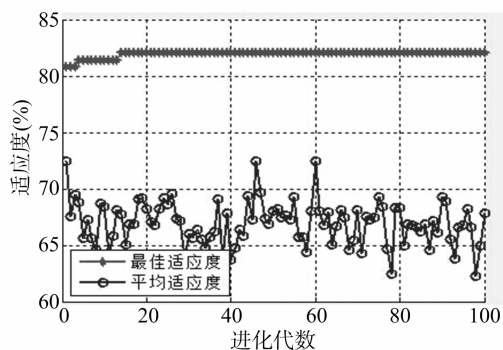


图11 粒子群算法参数寻优适应度曲线

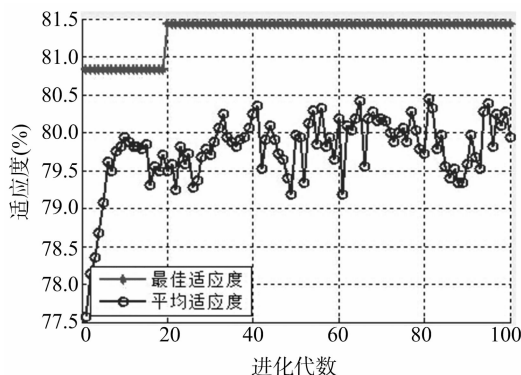


图12 遗传算法参数寻优适应度曲线

3.2 TM 数据提取棉花种植区结果

图 13 是用 TM 数据 NDVI 值大于 0.35 为掩膜提取的三湖农场棉田的种植分布。TM 数据 2,4,3 波段组合形成的假彩色图,农田清晰可见。三湖农场是长江流域优质专用棉生产基地,在 8 月末 9 月初的时候可以认为农田里种的作物全是棉花。图 13 中的黑色背景是用掩膜剔除的建筑用地和水体。

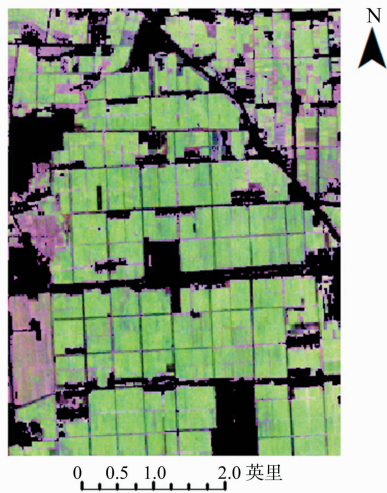


图 13 三湖农场棉田的分布

3.3 支持向量机法提取三湖农场蚜虫危害程度分布的结果

以地面调查的 262 个样点为训练样本,选取径向基核函数,利用粒子群参数寻优得到的优化值 $c = 4.020\ 3$ 、 $g = 20.652\ 4$ 进行支持向量机分类器的构造,得到分类模型。以空间插值并且归一化后的温度、速效氮、速效磷、速效钾、有机质、蒸散量和 NDVI 共 7 个因子数据形成待分类样本,用构造好的分类模型识别待分类样本得到棉蚜危害的分布,再用“3.2”节中建立的掩膜数据乘以分类数据就剔除了非农田的影响(图 14)。

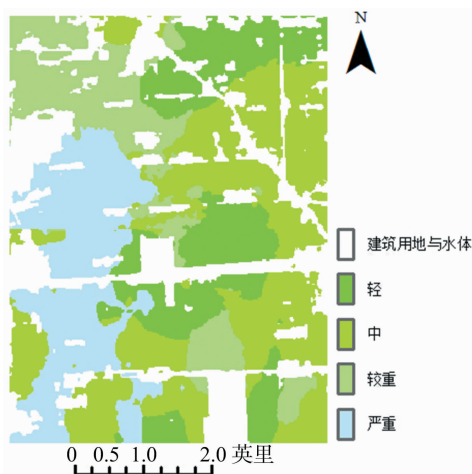


图 14 三湖农场棉蚜危害程度空间分布

由图 14 可以看出三湖农场中部和东部的棉花长势较好,受虫害较轻,受虫害严重的主要集中在西部和西南部;棉蚜危害程度的空间分布整体趋势与直接用蚜虫密度调查点数据空间插值得到的情况一致。

由以上分析可知,本研究所用的方法能很好地反演出三

湖农场棉花受蚜虫危害程度的空间分布,由于获取的样本点有限,导致样点数据的空间插值范围也有限,最终只能局限在三湖农场地区进行虫害情况反演。在以后的研究中可以将该方法推广到更大的区域和其他种类的作物,另外该方法只能利用温度、速效氮、速效磷、速效钾、有机质的含量、蒸散量和 NDVI 的综合效应来识别虫害,并不能明显地反映出每一种因子对棉虫的影响。如果能结合逐步回归分析的方法,找出主要的影响因子以及每个因子的权重,就能利用更少的地面数据进行大面积的虫害反演,进一步提高效率。

参考文献:

- [1] Muhammed H H, Larsolle A. Feature vector based analysis of hyper-spectral crop reflectance data for discrimination and quantification of fungal disease severity in wheat [J]. Biosystems Engineering, 2003, 86(2): 125 - 134.
- [2] Mirik M, Michels Jr G J, et al. Reflectance characteristics of Russian wheat aphid (Hemiptera: Aphididae) stress and abundance in winter wheat [J]. Computers and Electronics in Agriculture, 2007, 57(2): 123 - 134.
- [3] 卢小燕. 棉花蚜虫单叶高光谱特征识别研究 [J]. 新疆农垦科技, 2010, 6(1): 32 - 35.
- [4] 郭永旺, 金晓华, 杨建国, 等. 麦蚜灾害遥感监测技术应用研究 [J]. 植保技术与推广, 2001, 21(3): 3 - 5.
- [5] 苏荣瑞, 熊勤学, 耿一风, 等. 利用多时相 HJ - CCD 影像监测江汉平原南部地区棉花和中稻种植面积 [J]. 长江流域资源与环境, 2013, 22(11): 1441 - 1448.
- [6] 王新志, 陈伟, 祝明坤. 样本数据归一化方式对 GPS 高程转换的影响 [J]. 测绘科学, 2013, 38(6): 162 - 165.
- [7] 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述 [J]. 电子科技大学学报, 2011, 40(1): 2 - 10.
- [8] 顾亚祥, 丁世飞. 支持向量机研究进展 [J]. 计算机科学, 2011, 38(2): 14 - 17.
- [9] 张策, 臧淑英, 金竺, 等. 基于支持向量机的扎龙湿地遥感分类研究 [J]. 湿地科学, 2011, 9(3): 263 - 269.
- [10] 邓蕊, 马永军, 刘尧猛. 基于改进交叉验证算法的支持向量机多类识别 [J]. 天津科技大学学报, 2007, 22(2): 58 - 61.
- [11] 韩萌, 丁剑. 基于交叉验证的 BP 算法的改进与实现 [J]. 计算机工程与设计, 2008, 29(14): 3738 - 3739.
- [12] 胡局新, 张功杰. 基于 K 折交叉验证的选择性集成分类算法 [J]. 科技通报, 2013, 29(12): 115 - 117.
- [13] 王健峰, 张磊, 陈国兴, 等. 基于改进的网格搜索法的 SVM 参数优化 [J]. 应用科技, 2012, 39(3): 28 - 31.
- [14] 王鹏, 朱小燕. 基于 RBF 核的 SVM 的模型选择及其应用 [J]. 计算机工程与应用, 2003, 39(24): 72 - 73.
- [15] 张建科, 刘三阳, 张晓清. 改进的粒子群算法 [J]. 计算机工程与设计, 2007, 28(17): 4215 - 4216.
- [16] 张丹, 韩胜菊, 李建, 等. 基于改进粒子群算法的 BP 算法的研究 [J]. 计算机仿真, 2011, 28(2): 147 - 150.
- [17] 李良敏, 温广瑞, 王生昌. 基于遗传算法的回归型支持向量机参数选择法 [J]. 计算机工程与应用, 2008, 44(7): 23 - 26.
- [18] 王克奇, 杨少春, 戴天虹, 等. 采用遗传算法优化最小二乘支持向量机参数的方法 [J]. 计算机应用与软件, 2009, 26(7): 109 - 111.
- [19] 万源, 童恒庆, 朱映映. 基于遗传算法的多核支持向量机的参数优化 [J]. 武汉大学学报: 理学版, 2012, 58(3): 255 - 259.