

基于 Spark 的支持向量机在小麦病害 图像识别中的应用

林中琦¹,牟少敏^{1*},时爱菊²,孙肖肖¹,李磊¹
(1. 山东农业大学 信息科学与工程学院,山东 泰安 271018;
2. 山东农业大学 化学与材料科学学院,山东 泰安 271018)

摘要: 为了提高小麦病害图像分类的效率,提出了一种基于 Spark 的并行式支持向量机算法。首先对小麦病害图像进行滤波去噪、灰度压缩等处理,利用灰度共生矩阵、不变矩矩阵等从颜色、纹理和形状 3 个方面提取 49 个特征向量;然后通过数据集的切分和并行框架的支持,将大数据并行处理技术 Spark 与支持向量机结合,运用 Scala 语言实现了串行支持向量机算法的并行化,并将其应用于小麦病害图像识别。针对小麦锈病和白粉病的图像分类测试结果表明,当测试图像分别是 2 600、3 900、5 120 张时,该算法对锈病的分类精度依次是 76.03%、81.18%、77.82%,对白粉病的分类精度依次是 83.27%、85.91%、83.14%,比串行支持向量机分类精度有所提升。分类时间依次是 13 928.0、18 506.1、24 897.2 ms,明显低于串行支持向量机的分类时间。改进的算法实现了小麦病害分类精度的小幅度提升,明显提高了处理速度,具有较快的学习收敛速率。
关键词: 小麦病害; 图像分类; Spark; 支持向量机; 大数据; 并行计算; 图像特征提取
中图分类号: S435.121;TP391.41 **文献标志码:** A **文章编号:** 1004-3268(2017)07-0148-06

Classification of Wheat Disease Images Using Parallelized Support Vector Machine Based on Spark

LIN Zhongqi¹, MU Shaomin^{1*}, SHI Aiju², SUN Xiaoxiao¹, LI Lei¹
(1. College of Information Science and Engineering, Shandong Agricultural University, Taian 271018, China;
2. College of Chemistry and Material Science, Shandong Agricultural University, Taian 271018, China)

Abstract: In order to improve the efficiency of image classification for wheat diseases, a parallelized support vector machine algorithm based on Spark was proposed. First of all, the wheat disease images were denoised by filtering and compressed at gray-scale. Gray level co-occurrence matrix and invariant matrix and others were used to extract 49 feature vectors from color, texture and shape. Secondly, we combined the Spark with support vector machine through the support of segmentation of data sets and parallel framework. Finally, Scala language was used to realize the parallel processing of single support vector machine, and it was applied in the recognition of wheat disease images. The experimental results on the image classification of wheat diseases showed that the classification accuracies of wheat leaf rust were 76.03%, 81.18%, 77.82%, and the classification accuracies of powdery mildew were 83.27%, 85.91%, 83.14%, when the numbers of test images were 2 600, 3 900 and 5 120, respectively. The classification accuracy had been improved compared with the single support vector machine. The classification times were 13 928.0 ms, 18 506.1 ms, 24 897.2 ms respectively, which were obviously lower than that of the single support vector machine. The improved algorithm could make the classification accuracy of wheat diseases get a small increasing while the processing speed get a obvious ascension. The

收稿日期:2016-12-22
基金项目:山东省自然科学基金项目(ZR2012FM024); 2013 年山东省农业重大应用技术创新课题
作者简介:林中琦(1992-),男,山东泰安人,在读硕士研究生,研究方向:机器学习、图像处理、大数据。
E-mail:18764881210@163.com
* 通讯作者:牟少敏(1964-),男,山东泰安人,教授,博士,主要从事机器学习、数字图像处理、计算机视觉和大数据处理技术研究。E-mail:msm@sdau.edu.cn

proposed algorithm owns a faster convergence rate.

Key words: wheat diseases; image classification; Spark; support vector machine; big data; parallel computing; image characteristics extraction

目前,小麦病害识别基本上是凭借植保专家和务农人员的专业知识和工作经验进行判断和归类,费时费力,效率较低,严重影响小麦病害防治工作的精准性和时效性。近年来,随着数字图像处理技术的快速发展,通过计算机进行作物病害识别体现出前所未有的高效性。众多经典的模式识别算法,如支持向量机、神经网络等,为农业病害防治奠定了坚实的理论基础。备受关注的大数据并行处理技术,更是大大加快了计算机运行速度。因此,将三者结合,实现小麦病害种类快速而精准的识别分类,能够带来巨大的农业经济效益和广阔的开发应用前景。

支持向量机(SVM)是由 Vapnik^[1]提出的一种机器学习分类算法,因其严密的数学推理和较好的实践结果而备受关注,被广泛应用到农作物病害识别中。田有文等^[2]将 SVM 应用到葡萄病害的图像识别中。宋凯等^[3]实现了基于 SVM 的玉米叶片的病害识别。张建华等^[4]改进了最优二叉树 SVM,并将其应用于蜜柚叶面的病害识别。任东等^[5]利用线性组合核函数的 SVM 处理黄瓜病害图像,取得了较好的试验结果。Camargo 等^[6]以 SVM 为理论基础,以植物病害的 52 个特征参数为试验材料,得出其最优分辨率。

SVM 适用于小样本,在处理大规模数据时,算法的时间、空间复杂度急剧增加,传统串行 SVM 无法较好地处理大规模数据集。针对 SVM 处理海量数据开销大、速度慢的缺点,研究人员寻求解决方法,目前主要有 2 种途径:原始算法改进和并行计算^[7-10]。Woodsend 等^[11]将信息传递接口(MPI)与开放式信息传递接口(OpenMP)混合,实现了多核 SVM;Graf 等^[12]提出基于级联算法的层叠 SVM;Chang 等^[13]利用置信区间牛顿法并行化 Linear SVM,加快了运算速度;张奕武^[14]实现了基于 Hadoop 的 SVM,但依赖于 MapReduce^[15],速度提升不明显;唐振坤^[16]利用 Spark 实现 SVM 等算法,较大幅度降低了算法时间复杂度。随着近年来信息规模的增长,大数据处理技术快速兴起,在 Hadoop 基础上发展起来的内存式并行计算框架 Spark 能够较好地应对大规模迭代运算,Zaharia 等^[17]通过数学推理和试验证明对其做了详细解释。目前,尚未见基于 Spark 的并行式 SVM 应用于大规模小麦病害图像分类识别的研究。因此,本研究将 Spark 与 SVM 结合,以实现并行式 SVM,并用于小麦病害图像的

识别,在保证不降低分类精度的前提下,提高小麦病害图像的识别效率,对于小麦病害的自动快速诊断具有一定借鉴价值,同时对促进机器学习、图像处理、大数据等技术在农业工程中的融合应用具有重要实际意义。

1 基于 Spark 的支持向量机

将大数据并行处理技术 Spark 与 SVM 结合,实现串行算法的并行化,在不降低分类精度的前提下,可以明显提高算法的分类效率。

1.1 简介

SVM 适用于小样本,处理大规模数据时,传统串行 SVM 处理速度会降低。针对此问题,本研究将并行框架 Spark 与 SVM 结合,实现了基于 Spark 的 SVM。

算法过程分为 Map、Combine 和 Reduce 3 个阶段,采用分而治之的思想:算法利用 Spark 框架,将数据集分割为若干数据子块,并分配给各线程 Executor,进行局部性 SVM 的并行训练,再将各子分类器整合。通过数据子块训练出的子分类器具有局部性,因此要对整合后的全局分类器进行再训练。由于 Executor 的并行训练已使分类器较快地收敛,全局分类器只需在 Reduce 阶段微调即可,节省运算时间。

该算法除了提高运算速度外,还能提升分类精度,这是由于采用多道设计,每个线程能精细利用分割后的数据集子块,进行局部分类子模型的训练,规避了传统串行 SVM 无法较好利用样本局部信息的缺陷^[18-19],一定程度上提升了分类器的性能。

1.2 算法设计与实现

首先是数据集切分。将数据样本上传到分布式文件存储系统(HDFS),根据特征维度和指定分块数,通过 Spark 所提供的 partition 类中的分割规则,将数据集切分转换成弹性分布式数据集(RDD),并分布到各 Executor 上。

切分块数可根据集群节点数和机器性能等因素自行决定,由于本研究每个 Worker 设置 2 个 Executor,集群共有 14 台机器,所以分块数为 28 块。之后,每个子块调用 randomSplit() 将数据集按照 7:3 的比例随机划分成训练样本和测试样本。

1.2.1 Map 阶段 每个 Executor 根据目标函数建立分类器,调用 train() 方法,通过数据子块对分类器进行迭代训练。为适应大规模训练样本,本研究

采用随机梯度下降算法 (SGD) 进行参数调优,每次迭代只有部分样本参与计算,内存开支小,耗时低。

本研究创建 setter 对象,实现对 train() 的重写:设置迭代次数 numIterations 为 300,迭代步长 step-Size 为 2,每次迭代样本参与比例 miniBatchFraction 为 2,为防止过拟合,引入岭回归 L2 regularization 作为修正函数,正则因子 regRaram 为 0.1。

1.2.2 Combine 阶段 Combine 是 Map 和 Reduce 之间的过渡阶段,通过 Combiner 对象,将所有数据子块以及 Map 阶段训练出的局部分类器合并,交予 Reduce 阶段。

Combine 阶段不是简单的线性拼接,它通过实例化 shuffledRDD 类的对象调用 repartition(),实现原数据与子分类器洗牌后的拼接。

1.2.3 Reduce 阶段 接收 Combine 阶段返回的结果,调用 predictPoint() 对合并后的分类器测试。计算测试样本每条记录的预测值,与原始数据对比,获得误分率,并通过接收器操作特性曲线 (ROC) 打分。若分值小于阈值,分类器不合格,继续优化,分值大于阈值,分类器达到标准,测试阶段结束,打印全局最优分类器对测试集每条记录的预测值和隶属程度、分类精度、分类时间、ROC 系数等。基于 Spark 的 SVM 数据流如图 1 所示。

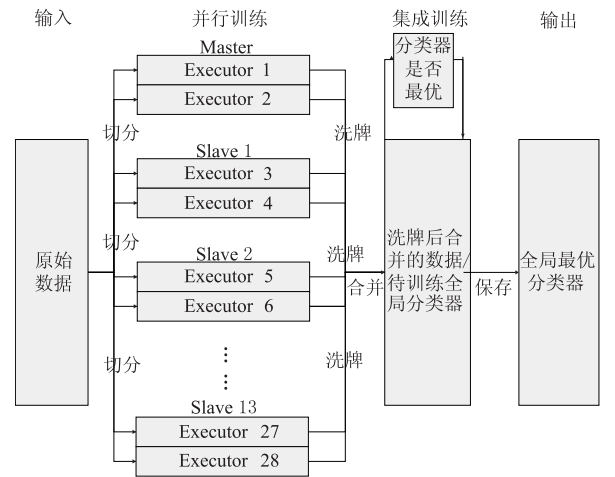


图 1 基于 Spark 的 SVM 数据流

2 小麦病害图像的识别

2.1 试验环境

试验基于 14 台计算机所构成的 Spark 完全分布式环境,每台机器硬盘容量为 1.5 TB,可用内存 3.8 GB,双核处理器,处理器型号为 Intel(R) Pentium(R) CPU G645@ 2.90 GHz 2.90 GHz。其中,1 台机器作为 Master,其余 13 台机器作为 Slave,运行模式为 Standalone。每台机器软件信息如表 1 所示。

表 1 Spark 集群机器软件信息

软件名称	软件版本
Linux operating system	Ubuntu 14.0
Hadoop	Hadoop 1.20.2
Spark	Spark v1.2.2
Jdk	jdk - 8u60 - linux - i586
Scala	Scala 2.9.3
Intellij IDEA	idealC - 15.0.3

2.2 小麦病害图像的采集和预处理

为验证该算法在处理速度和分类精度上的优势,从山东农业大学和山东省泰安市周边地区的济麦 20 号、烟农 19 号小麦生产基地以及试验田,人工对小麦发病叶片进行高质量图像采集,采集工具为数码单反相机 Canon EOS 760D。

由于受到光照不均、设备抖动等因素的影响,在采集到的小麦病害彩色图像中,病斑的边缘、颜色发生弥散、模糊、反光等现象,图像降质会影响特征提取,进而降低后期试验的准确性和说服力。为去除噪声,强化病斑对比度,本研究采用同态滤波对图像进行增强处理。同态滤波将图像的照度模型由乘积形式变为加和形式,通过照度范围的压缩和对比度的增强来提升图像质量。同态滤波的具体步骤如图 2 所示。

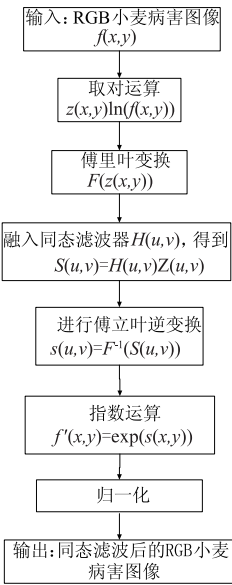


图 2 同态滤波流程

本研究利用 Matlab 7.0 对小麦病害图像进行同态滤波处理,效果如图 3 所示。图 3(a)是一幅曝光度过高的小麦白粉病叶部原始图像,由于采光较强,图像整体发白模糊,此外,叶片自身弯曲度使叶部存在阴影部分,影响对病斑的颜色、形状等图像特征的提取。图 3(b)是同态滤波去除噪声后的图像,同态滤波可以降低入射光照的低频分量,增加反射光照的高频分量,压缩图像照度值域的同时,又提高相邻

区域像素的对比度,消除因光照过强引发的图像模糊现象,并对阴影区域进行有效地增强。

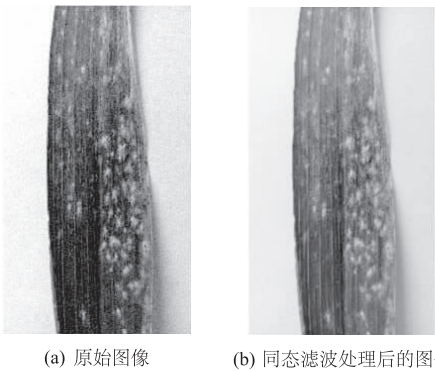


图 3 小麦白粉病叶部图像同态滤波处理效果

2.3 小麦病害图像的特征提取

由于众多病害机制的差异,不同病害下的小麦叶片呈现出不同外表,合理的图像特征提取能最大化突出每种叶部病害外观的特点,有利于计算机对小麦病害的类别进行精准识别。

本研究以预处理后的 5 120 张小麦病害图像为

试验样本,图像样本如表 2 所示。图像大小为 5 312×2 988,格式为 JPEG,24 位图,分为锈病和白粉病 2 类病害。

表 2 小麦叶部病斑样本实例

项目	小麦锈病	小麦白粉病
病症外观		

获得小麦叶片的病变区域后,从颜色、纹理和形状 3 个方面提取 49 个特征向量,其中颜色特征 6 个,纹理特征 32 个,形状特征 11 个。具体的小麦病害图像特征提取参数如表 3 所示。

表 3 小麦锈病和白粉病图像特征提取参数

特征类别	参数	数量/个	总数/个
颜色特征	HSV 颜色空间一阶矩灰度值(H, S, V)	3	6
	HSV 颜色空间二阶矩灰度值(H', S', V')	3	
纹理特征	RGB、HSV 颜色空间 $\theta=0^\circ$ 灰度共生矩阵中能量 E 、熵 H 、惯性矩 I 、相关性 C	8	32
	RGB、HSV 颜色空间 $\theta=45^\circ$ 灰度共生矩阵中能量 E 、熵 H 、惯性矩 I 、相关性 C	8	
	RGB、HSV 颜色空间 $\theta=90^\circ$ 灰度共生矩阵中能量 E 、熵 H 、惯性矩 I 、相关性 C	8	
	RGB、HSV 颜色空间 $\theta=135^\circ$ 灰度共生矩阵中能量 E 、熵 H 、惯性矩 I 、相关性 C	8	
	Hu 不变矩阵; $m(1), m(2), m(3), m(4), m(5), m(6), m(7)$	7	
形状特征	面积(S)	1	11
	周长(L)	1	
	圆度(C)	1	
	复杂度(E)	1	

2.3.1 颜色特征提取 以 Matlab 7.0 为提取平台,先通过遍历算法读取指定文件夹下所有 JPEG 格式的小麦病害图像,获得纹理区域后,将各颜色分量转化为灰度值,获得灰度图像,再将 RGB 空间转换成 HSV 空间^[20],得到一、二阶矩上 H 、 S 、 V 3 个分量的值。

2.3.2 纹理特征提取 采用 RGB、HSV 颜色空间。为减少计算量并提高处理精度,对原始图像进行灰度等级压缩,将 Gray 量化成 16 级。再计算 4 个灰度共生矩阵,取距离为 1,角度分别是 0° 、 45° 、 90° 、 135° ,再对归一化的共生矩阵求解 4 个方向上的能量 E 、熵 H 、惯性矩 I 和相关性 C 。

能量 E 用来衡量图像灰度分布的均匀程度,计算公式为:

$$E = \sum_{i=0}^{P-1} \sum_{j=0}^{P-1} p^2(i, j)$$

式中, $p(i, j)$ 表示在大小为 $P \times P$ 的灰度共生矩

阵中,处于坐标 (i, j) 位置上的元素值,下同。

熵 H 用来衡量病害图像纹理的非均匀分布的程度,其计算公式为:

$$H = \sum_{i=0}^{P-1} \sum_{j=0}^{P-1} p(i, j) \log_2 p(i, j)。$$

惯性矩 I 用来衡量截面抵抗弯曲的能力,计算公式为:

$$I = \sum_{i=0}^{P-1} \sum_{j=0}^{P-1} (i - j)^2 p(i, j)。$$

相关性 C 用来衡量灰度共生矩阵的所有元素在行列方向分布的相似度,其计算公式为:

$$C = \frac{\sum_{i=0}^{P-1} \sum_{j=0}^{P-1} ij p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

式中, μ_x 和 μ_y 表示 $p_x(i)$ 和 $p_y(j)$ 的均值, σ_x 和 σ_y 表示 $p_x(i)$ 和 $p_y(j)$ 的标准差。

2.3.3 形状特征提取 先用 Canny 边缘检测

法^[21-22]提取图像边缘,保留边缘灰度图像,再依次计算面积 S 、周长 L 、圆度 C 、复杂度 E 和 Hu 不变矩阵。面积 S 是边缘封闭区域内像素的总和,计算公式为:

$$S = \sum_{x=0}^N \sum_{y=0}^N f(x,y)$$

式中, $f(x,y)$ 是二值图像函数,可求得图像矩阵中位于坐标 (x,y) 位置上像素点的面积。

周长 L 是保留下来的边缘的长度,计算公式为:

$$L = \sum_{i=1}^N \Delta l_i$$

式中, Δl_i 表示病斑图形的微元长度。

圆度 C 用来衡量病斑边缘的拟圆程度,计算公式为:

$$C = \frac{4\pi S}{L^2}$$

式中, S 表示病斑面积, L 表示病斑周长,下同。

复杂度 E 用来衡量小麦病变叶面区域的离散程度,计算公式为:

$$E = \frac{L^2}{S}。$$

用 Otsu^[23-24] 阈值法为每一幅图像选定阈值,并用该阈值对图像进行二值化处理。为减少计算过程中的精度丢失,将像素矩阵中各元素的数据类型转换成双精度,之后依次计算灰度图像的零阶几何矩阵,图像的二阶、三阶几何矩阵,图像的二阶、三阶中心矩阵,图像的归一化中心矩阵,最后将各阶中心矩阵组合获得 Hu 的 7 个不变矩阵。求得 Hu 不变矩阵算法^[25] 流程如图 4 所示。

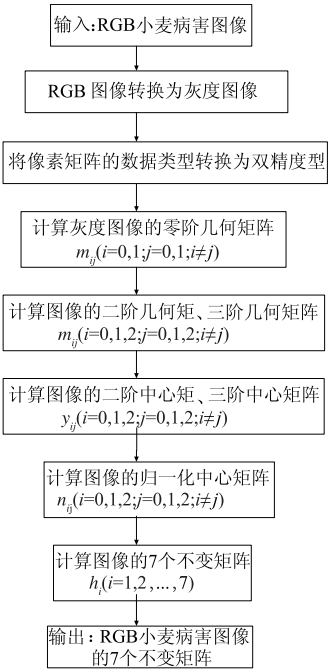


图 4 Hu 不变矩阵算法流程

2.4 并行 SVM 与串行 SVM 图像识别对比试验

将基于 Spark 的并行 SVM 与串行 SVM 作对比试验,两者的优化目标均为:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \zeta_i$$

$$\text{subject to } y_i [(wx_i) + b] \geq 1 - \zeta_i (i = 1, 2, \cdots, l) \\ \zeta_i \geq 0$$

式中,惩罚参数 C 设为 10,松弛变量 ζ 取 0.001。

核函数均采用目前广泛应用的径向基核函数 (RBF)^[26-28],它可将一组无法直接进行线性分割的数据集映射到高维空间中,较好地解决种类和属性值之间非线性的情况。径向基核函数为:

$$\text{radial basis function: } K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

$$\text{subject to } \gamma > 0$$

式中,参数 γ 设为试验样本特征向量总数的倒数,即 $\gamma = 1/49 \approx 0.02$,核函数的宽度 q 取 3。

试验共采用 5 120 张小麦病害图像,根据随机抽出样本数量的不同分为 4 组,分别命名为 Sample_1、Sample_2、Sample_3 和 Sample_4,将每组样本的 70% 作为训练集,30% 作为测试集,数据样本的具体情况如表 4 所示。

表 4 试验样本分组

样本名称	数据样本数量/条	训练样本数量/条	测试样本数量/条
Sample_1	1 300	910	390
Sample_2	2 600	1 820	780
Sample_3	3 900	2 730	1 170
Sample_4	5 120	3 584	1 536

本研究将改进的并行式 SVM 与串行 SVM 在相同条件下进行试验,通过处理样本数量依次递增的 4 组小麦病害图像数据集,得到两者的分类精度和分类时间,分析样本规模大小对 2 种算法的影响。通过表 5 可知,在处理 4 组不同规模大小的数据样本时,基于 Spark 的并行式 SVM 比串行 SVM 的分类精度有所提升,原因是并行式支持向量机采用多道设计和数据集分割,每一线程能够精细利用分割后的部分数据集,规避了传统串行 SVM 无法较好利用样本局部信息的缺点^[18-19],比起串行 SVM 直接从全部样本中训练分类器,学习效果更好。

表 5 基于 2 种算法的试验分类精度 %

试验样本	并行式 SVM		串行 SVM	
	锈病	白粉病	锈病	白粉病
Sample_1	91.20	93.82	89.76	91.64
Sample_2	76.03	83.27	74.94	82.63
Sample_3	81.18	85.91	80.06	84.09
Sample_4	77.82	83.14	76.01	82.58

通过表 6 可知,当试验数据集规模较小时,两者

处理速度持平,甚至串行 SVM 的效率略高于并行式 SVM。这是由于样本过少,Spark 集群在数据集切分和任务调度上投入过多时间,并行效果不明显,不如串行 SVM 直接训练分类器效率高。当试验样本数量上涨到一定程度,串行 SVM 所用时间明显增加,而并行式 SVM 所用时间平稳增加,可见当测试样本规模较大时,Spark 集群的并行处理效果显著,从整体上节省了算法运行时间。

表 6 基于 2 种算法的试验分类时间 ms		
试验样本	并行式 SVM	串行 SVM
Sample_1	11 014.5	10 456.7
Sample_2	13 928.0	14 845.1
Sample_3	18 506.1	20 510.1
Sample_4	24 897.2	38 298.1

3 结论与讨论

复杂多变的致病机制决定了小麦叶部病害外观的多样性,对大量原始病发图像的高清采集,中期图像的去噪处理,以及后期从颜色、纹理、形状 3 个方面提取 49 个图像特征,是本研究基于小麦病征进行病害种类划分的依据。

SVM 是基于大量迭代的分类算法,适用于小样本环境,在处理大规模数据集时,时间复杂度较高。本研究提出了基于 Spark 的 SVM 的设计思想,并在 Spark 平台上实现算法,可将大规模数据集切分,并应用到分类器的并行训练上。通过对大量小麦病害图像的分类试验得出结论:与传统串行 SVM 相比,基于 Spark 的并行式 SVM 的分类精度有所提升,分类时间明显缩减,体现出大数据技术 Spark 并行处理的优势。

参考文献:

[1] Vapink V. The nature of statistical learning theory [J]. New York:Springer-Verlag,1995,8(6):988-999.

[2] 田有文,牛妍. 支持向量机在黄瓜病害识别中的应用研究[J]. 农机化研究,2009(3):36-39.

[3] 宋凯,孙晓艳,纪建伟. 基于支持向量机的玉米叶部病害识别[J]. 农业工程学报,2007,23(1):155-157.

[4] 张建华,孔繁涛,李哲敏,等. 基于最优二叉树支持向量机的蜜柚叶部病害识别[J]. 农业工程学报,2014,30(19):222-231.

[5] 任东,于海业,王纪华. 基于线性组合核函数支持向量机的病害图像识别研究[J]. 农机化研究,2007(9):41-43.

[6] Camargo A,Smith J S. Image pattern classification for the identification of disease causing agents in plants [J]. Computers and Electronics in Agriculture,2009,66(2):121-125.

[7] Liu H Y,Yuan Q Q,Wang B B. Survey of parallel algorithms for data mining [J]. Electronic Science and Technology,2006,19(1):69-72,77.

[8] 刘泽桑,潘志松. 基于 Spark 的并行 SVM 算法研究

[J]. 计算机学报,2016,43(5):238-242.

[9] 张巍,张功萱,王永利,等. 基于 CUDA 的 SVM 算法并行化研究[J]. 计算机科学,2013,40(4):69-72.

[10] 郭欣欣. 基于分布式计算的 SVM 算法优化[D]. 西安:西安电子科技大学,2014.

[11] Woodsend K,Gondzio J. Hybrid MPI/OpenMP parallel linear support vector machine training [J]. Journal of Machine Learning Research,2009,10(12):1937-1953.

[12] Graf H P,Cosatto E,Bottou L,et al. Parallel support vector machines:The cascade SVM[C]//Advances in neural information processing systems(NIPS). Cambridge,MA:MIT Press,2004:521-528.

[13] Chang C C,Lin C J. LIBSVM:A library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology,2011,2(3):75-102.

[14] 张奕武. 基于 Hadoop 分布式平台的 SVM 算法优化及应用[D]. 广州:中山大学,2012.

[15] Dean J,Ghemawat S. MapReduce:Simplified data processing on large cluster [J]. Communications of the ACM,2008,51(1):107-113.

[16] 唐振坤. 基于 Spark 的机器学习平台设计与实现[D]. 厦门:厦门大学,2014.

[17] Zaharia M,Chowdhury M,Franklin M J,et al. Spark: Cluster computing with working sets [C]//Proceedings of the 2nd USENIX conference on hot topics in cloud computing. Berkeley:USENIX Association,2010:10.

[18] Cheng H,Tan P N,Jin R. Efficient algorithm for localized support vector machine [J]. IEEE Transactions on Knowledge & Data Engineering,2010,22(4):537-549.

[19] Segata N,Blanzieri E. Fast and scalable local kernel machines [J]. Journal of Machine Learning Research,2010,11(6):1883-1926.

[20] 林庆,徐柱,王士同,等. HSV 自适应混合高斯模型的运动目标检测 [J]. 计算机科学,2010,37(10):254-256.

[21] 张帆,彭中伟,蒙水金. 基于自适应阈值的改进 Canny 边缘检测方法 [J]. 计算机应用,2012,32(8):2296-2298.

[22] 薛丽霞,李涛,王佐成. 一种自适应的 Canny 边缘检测算法[J]. 计算机应用研究,2010,27(9):3588-3590.

[23] 胡敏,李梅,王荣贵. 改进的 Otsu 算法在图像分割中的应用 [J]. 电子测量与仪器学报,2010,24(5):443-449.

[24] 范九伦,赵凤. 灰度图像的二维 Otsu 曲线阈值分割法 [J]. 电子学报,2007,35(4):751-755.

[25] 张伟,何金国. Hu 不变矩的构造与推广 [J]. 计算机应用,2010,30(9):2449-2452.

[26] Chapelle O,Vapnik V. Choosing multiple parameters for support vector machines [J]. Machine Learning,2002,46:131-159.

[27] Osuna E E,Freund R,Girosi F. Support vector machines:Training and applications [D]. Cambridge:Massachusetts Institute of Technology,1998.

[28] Dong C X,Yang S Q,Rao X,et al. An algorithm of estimating the generalization performance of RBF-SVM [C]//Proceedings of the fifth international conference on computational intelligence and multimedia applications. USA:IEEE,2003:61-66.