

doi:10.6041/j.issn.1000-1298.2013.02.035

基于 BPSO 的棉花异性纤维目标特征快速选择方法*

王金星 李恒斌 王蕊 刘双喜 曹维时 闫银发

(山东农业大学机械与电子工程学院, 泰安 271018)

摘要: 针对现有棉花异性纤维目标特征选择方法迭代次数多、速度慢等问题, 提出了一种基于改进粒子群优化算法的棉花异性纤维目标特征快速选择方法。使用离散型粒子群优化算法作为特征选择算法, 利用支持向量机算法作为分类器对最优特征集进行验证。实验结果表明, 在分类准确率与蚁群算法相当的情况下, 能减少 26% 的运行时间。

关键词: 棉花 异性纤维 支持向量机 离散型粒子群优化算法 特征选择

中图分类号: TP491.4 **文献标识码:** A **文章编号:** 1000-1298(2013)02-0188-04

A Fast Feature Selection for Cotton Foreign Fiber Objects Based on BPSO

Wang Jinxing Li Hengbin Wang Rui Liu Shuangxi Cao Weishi Yan Yinfa

(Mechanical and Electronic Engineering College, Shandong Agricultural University, Taian 271018, China)

Abstract: A fast feature selection for cotton foreign fiber objects based on binary particle swarm optimization was presented, for the current feature selection of cotton foreign fiber having more iteration times and slow speed. Binary particle swarm optimization (BPSO) was used to select feature in the method, and the support vector machine algorithm was used to verify the optimal feature set. Experimental results showed that the running time could reduce by 26%, when the classification accuracy was almost with other algorithms.

Key words: Cotton Foreign fiber SVM Binary particle swarm optimization Feature selection

引言

棉花异性纤维是指在棉花采摘、摊晒、收购、储存、运输、加工过程中混入棉花中的对棉花及其制品质量有严重影响的非棉纤维和色纤维, 如塑料布、布条、麻绳、羽毛和丙纶丝等, 俗称“三丝”^[1]。异性纤维在棉花中的含量虽然很少, 但一旦混入棉花中, 会导致断纱、瑕疵, 使棉纺织能力下降, 严重的还会影响织物的外观^[2]。目前国内大部分企业普遍采用人工挑拣来剔除棉花异性纤维, 成本高且效率低^[3]。为了解决棉花异性纤维的困扰, 国内外在异性纤维检测方面进行了多年研究, 并已经取得了很多成果, 采用的主要技术方法之一是利用机器视觉技术^[4]。目标分类是机器视觉技术的关键^[5], 而特征选择可以选出最有利于分类的特征, 加快算法的

运行速度、消除冗余、提高分类准确率^[6]。

特征选择是指从一组经过预处理的特征数据中挑选出最有效的特征子集, 从而降低特征空间维数、并且可以简化分类器设计、提高分类速度^[5]。粒子群优化算法 (Particle swarm optimization, PSO) 是群体智能算法的一种, 由 Kennedy 和 Eberhart 于 1995 年提出^[7~8]。粒子群通过个体间的协作来寻找最优解, 其概念简单、收敛速度快、容易实现, 目前已经广泛应用于特征选择和模式分类等领域^[9], 但不能用来解决连续空间的优化组合问题。现应用于棉花异性纤维特征选择的算法有遗传算法^[10]和蚁群算法 (IACA)^[11]等, 但普遍存在参数多、搜索速度慢的问题。

因此, 本文提出一种基于粒子群优化算法的棉花异性纤维目标特征快速选择方法。

收稿日期: 2012-02-08 修回日期: 2012-03-13

* 国家自然科学基金资助项目 (30971693) 和新世纪优秀人才计划资助项目 (NCET-09-0731)

作者简介: 王金星, 教授, 主要从事数字图像处理研究, E-mail: jinxiangw@163.com

1 棉花异性纤维目标特征

1.1 图像分割

样本图像来源于基于自动视觉检测(AVI)的棉花异性纤维检测系统,从实验平台实时采集的图像中选出含有异性纤维的4000像素×128像素真彩色图像(图1)。

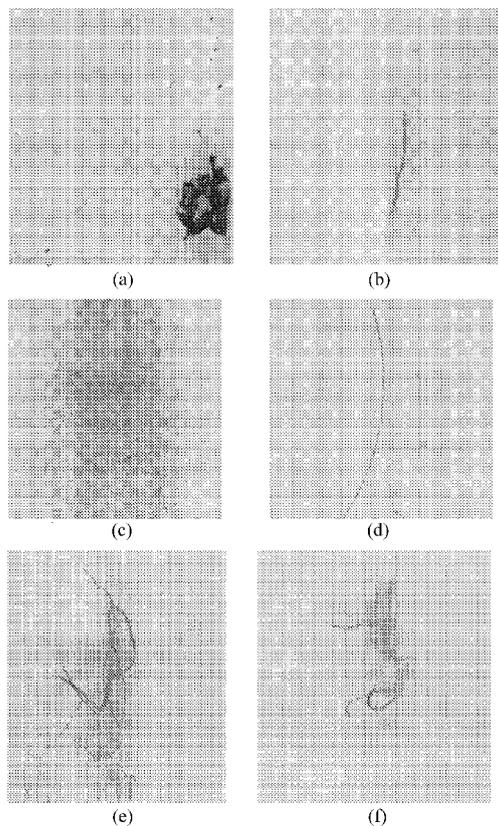


图1 原始棉花异性纤维图像

Fig. 1 Original image of cotton foreign fiber

(a) 塑料布 (b) 布条 (c) 麻绳
(d) 头发 (e) 丙纶丝 (f) 羽毛

对样本图像进行噪声分析,去除噪声;通过对灰度直方图分析,确定棉花背景以及异性纤维目标的灰度范围,应用阈值对样本图像进行分割处理,对分割出的小目标进行分类,分为塑料布、布条、麻绳、头发、丙纶丝和羽毛6类,每类选择100个小目标,对小目标进行特征提取,得到600个75维样本组成的特征数据集。

1.2 数据描述

描述目标最常用的特征是颜色特征、形状特征和纹理特征,但由于棉花异性纤维的种类很多、颜色与纹理各不相同,单独使用某一种特征几乎无法对棉花异性纤维进行准确分类,为了提高目标分类的准确率,需组合使用颜色特征、形状特征和纹理特征^[10]。

棉花异性纤维采用75维的特征向量进行表示,

特征向量包括75个特征,其中包括24个颜色特征、43个纹理特征、8个形状特征^[11]。

2 基于改进BPSO的特征选择算法

2.1 算法基本原理与改进

2.1.1 离散型PSO算法

PSO和遗传算法等其他群算法相似,基于群体的迭代搜索,但是PSO没有交叉、变异算子,它通过每个个体之间的协作来搜索最优解。每个粒子的位置和运动速度是由当前粒子的最优解和粒子群体的最优解决定的^[13]。粒子在解空间追随最优的粒子进行搜索,因此能够提高搜索效率。PSO不能用来解决连续空间的优化组合问题,因此Kennedy等提出了用于解决离散问题的离散型PSO算法(Binary particle swarm optimization, BPSO)^[16]。BPSO算法具有收敛性差和易陷入局部极值点等缺点。本文针对BPSO的缺点并结合棉花异性纤维的特征作了如下改进:采用动态参数来进行问题求解,引入惯性权重系数和收缩因子用以较好地控制粒子的飞行速度,能有效平衡算法全局最优解和局部最优解的寻找,从而解决算法收敛性差和易陷入局部极值点等缺点。

2.1.2 改进BPSO算法

每一个粒子表示了 N 维解空间中的一个点,它的下一个运动位置由粒子当前的位置和速度所决定。第 i 个粒子在第 k 次迭代下的位置可以表示为 $X_i(k) = [x_{i1}(1) \ x_{i2}(2) \ \dots \ x_{id}(k)]^T$,速度可以表示为 $V_i(k) = [v_{i1}(1) \ v_{i2}(2) \ \dots \ v_{id}(k)]^T$ 。

设 $p_i(k) = [p_{i1}(1) \ p_{i2}(2) \ \dots \ p_{id}(k)]^T$ 表示 X_i 的当前最优解, $g_i(k) = [g_{i1}(1) \ g_{i2}(2) \ \dots \ g_{id}(k)]^T$ 表示全局最优解,则第 $k+1$ 次迭代中,所有粒子的位置和速度更新公式为

$$v_{id}(k+1) = v_{id}(k) + c_1 r_1(k) (p_{id}(k) - x_{id}(k)) + c_2 r_2(k) (g_{id}(k) - x_{id}(k)) \quad (1)$$

$$x_{id}(k+1) = x_{id}(k) + \beta v_{id}(k+1) \quad (2)$$

式中 c_1, c_2 ——学习因子,取值在 $[0, 2]$ 之间

r_1, r_2 ——随机数,取值在 $[0, 1]$ 之间

β ——约束因子

在BPSO中每个位置分量 x_{id} 取0或1,故 v_{id} 不再表示位置变化的大小,它反映的是 v_{id} 为 x_{id} 取1的概率,通过转换函数限制概率值在 $[0, 1]$ 之间,通常都采用Sigmoid函数,即

$$\text{sig}(v_{id}(k)) = \frac{1}{1 + \exp(-v_{id}(k))}$$

故BPSO的粒子位置的更新公式为

$$x_{id}(k+1) = \begin{cases} 1 & (r_{id} \leq \text{sig}(v_{id}(k+1))) \\ 0 & (\text{其他}) \end{cases} \quad (3)$$

式中 r_{id} ——随机数

引入惯性权重系数 w 和收缩因子 l , 将速度公式更新为

$$v_{id}(k+1) = l[wv_{id}(k) + c_1 r_1(k)(p_{id}(k) - x_{id}(k)) + c_2 r_2(k)(g_{id}(k) - x_{id}(k))] \quad (4)$$

($w \in [w_{\max}, w_{\min}]$)

其中 $l = \frac{2}{|2 - m - \sqrt{m^2 - 4m}|}$

($m = r_1 c_1 + r_2 c_2$ 且 $m > 4$)

通过调节惯性权重系数 w 能够控制粒子的速度, 从而影响算法的收敛速度和收敛性。通常 w 越大, 粒子的飞行速度就越高, 这样就更容易搜索全局最优解, 但是容易忽略局部最优解; w 越小, 粒子的速度也就越低, 这样就更容易搜索局部最优解, 但是容易忽略全局最优解。因此选择最优值时应该兼顾全局最优解和局部最优解的搜索, 这样可以快速寻找到最优解。

2.2 适应度函数设计

适应度函数模型为

$$f = \left(\frac{1}{N} \sum_{k=1}^N a_k^2 \right)^{-\frac{1}{2}} \quad (5)$$

式中 a_k ——编号为 k 的粒子

f ——粒子 a_k 生成解的适应度

N ——选择的特征个数

适应度函数是判断粒子优劣性能评定标准, 适应度越大, 表明选择的特征子集表现越好, 即利用较少的特征获得较高的分类正确率。

2.3 支持向量机分类识别

支持向量机 (Support vector machine, SVM) 处理高维数据时的具有灵活性和计算有效性, 被广泛应用于数据挖掘的分类问题中^[17]。设计采用高斯径向基函数作为 SVM 的核函数, 核函数为

$$K(x, x_i) = \exp \left(- \frac{\|x - x_i\|^2}{\sigma^2} \right) \quad (6)$$

所得支持向量机分类函数为

$$f(x) = \text{sgn} \left[\sum y_i \alpha_i \exp \left(- \frac{\|x - x_i\|^2}{\sigma^2} \right) - b \right] \quad (7)$$

用 BPSO 算法提取棉花异性纤维的最优特征子集作为 SVM 的输入向量, 对数据集进行分类学习, 应用 SVM 分类器来验证特征选择的优劣。

2.4 算法流程

算法循环次数达到最大迭代次数时本算法终止, 输出最优特征集。算法流程如图 2 所示。

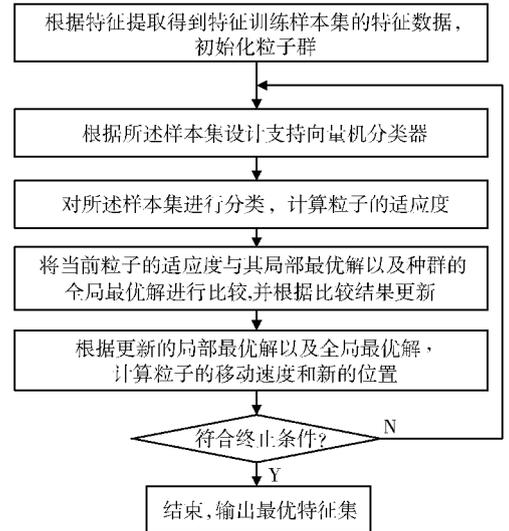


图 2 基于 BPSO 的特征选择算法流程图

Fig. 2 Flow chart of feature selection based on BPSO

3 实验结果与讨论

算法利用 Java 编程实现。原始数据集被随机等分成 10 份, 保持每份中的每种特征的数量和异性纤维的类型数量相等。每次取其中 1 份作为测试集, 另 9 份合并为训练集。参数设置为: $c_1 = c_2 = 2$; $w \in [0.9, 0.7]$, 在程序中设定从 0.9 逐步减小到 0.7; 粒子数 $P_N = 20$; 最大迭代次数 $N_m = 1000$ 。

实验对 BPSO 算法和蚁群算法^[11] (IACA) 在棉花异性纤维特征选择方面的性能进行比较。两算法实验所用的图像源自同一平台, 但本实验应用的系统有所改进。对 BPSO 算法和 IACA 算法各运行 10 次, 然后求平均, 结果如表 1 所示。

表 1 原特征集与选择后特征集分类准确率比较

Tab. 1 Comparison of classification performance of original feature sets and selected feature sets %

种类	原始特征	IACA 特征选择后	BPSO 特征选择后
塑料布	92	99	99
布条	82	94	95
麻绳	78	88	89
头发	82	89	89
丙纶丝	94	100	100
羽毛	74	85	87
总计	83.67	92.50	93.17

原始特征数为 75, 分类准确率 83.67%; IACA 算法选择后特征数为 22, 平均分类准确率为 92.50%; BPSO 算法选择后特征数为 20, 平均分类准确率为 93.17%。与原始特征相比, BPSO 算法选择后的特征能够提高目标分类准确率, 与 IACA 算法相比, BPSO 算法选择后的特征不会降低目标特征分类准确率。

表2为IACA算法和BPSO算法搜索效率比较,BPSO迭代次数比IACA减少31%,运行时间缩短26%,BPSO特征选择算法能够有效提高算法的搜索效率。

表2 IACA算法和BPSO算法搜索效率比较

Tab.2 Comparison of search efficiency of IACA and BPSO

特征选择方法	迭代次数/次	运行时间/s
IACA算法	1 472	5 152
BPSO算法	1 015	3 812

为了考察基于改进粒子群优化算法的棉花异性纤维目标特征快速选择方法对于提高特征选择速度的有效性,对经过2种算法选择出的特征集分类性

能和特征选择效率进行了比较。实验结果表明,在不影响目标分类精度的前提下,BPSO算法能够有效提高特征选择效率。

4 结束语

提出了一种基于改进粒子群优化算法的棉花异性纤维目标特征快速选择方法。该方法能在一定程度上解决离散型粒子群优化算法收敛性差和易陷入局部极值点的缺点,提高特征选择速度,兼顾全局最优解和局部最优解的搜索,在不损失目标分类精度的前提下,使特征选择的时间大大减少。与蚁群算法搜索效率的比较中,BPSO算法和IACA算法的运行时间各为3 812 s和5 152 s,BPSO算法的运行时间缩短了26%。

参 考 文 献

- 冯万众. 棉花异性纤维的定义和评定方法探讨[J]. 中国纤检,2000(1):156~160.
- 陈兆顺. 谈棉花中异性纤维对纺织企业的影响[J]. 中国纤检,1996(10):2 203~2 207.
- 朱仲堂. 解决棉花中异性纤维的建议[J]. 中国纤检,2007(6):42~43.
- 郑文秀,王金星,刘双喜,等. 棉花异性纤维的动态识别技术研究[J]. 仪器仪表学报,2009,30(10):2 203~2 207. Zheng Wenxiu, Wang Jinxing, Liu Shuangxi, et al. Research on dynamic identification technology of cotton foreign fibers[J]. Chinese Journal of Scientific Instrument, 2009,30(10):2 203~2 207. (in Chinese)
- 边肇祺,张学工. 模式识别[M]. 北京:清华大学出版社,2000.
- 黄炜,黄志华. 一种基于遗传算法和SVM的特征选择[J]. 计算机技术与发展,2010,20(6):21~24.
- Kennedy J, Eberhart R. Particle swarm optimization[C]//Proc. of IEEE Int. Conf. on Neural Networks, Perth, Australia. Piscataway, NJ: IEEE Press, 1995: 1 942~1 948.
- Eberhart R, Kennedy J. A new optimizer using particle swarm theory[C]//Proc. of the Sixth International Symposium on Micro Machina and Human Science, Nagoya, Japan. Piscataway, NJ: IEEE Press, 1995: 39~43.
- John G H, Kohavi R, Pflieger K. Irrelevant features and the subset selection problem[C]//Proceedings of the Eleventh International Conference on Machine Learning, New Brunswick, NJ, USA: Morgan Kaufmann, 1994: 121~129.
- 杨文柱,李道亮,魏新华,等. 基于改进遗传算法的棉花异性纤维目标特征选择[J]. 农业机械学报,2010,41(4):173~178. Yang Wenzhu, Li Daoliang, Wei Xinhua, et al. Feature selection for cotton foreign fiber objects based on improved genetic algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2010,41(4):173~178. (in Chinese)
- 赵学华,李道亮,杨文柱,等. 基于改进蚁群算法的棉花异性纤维目标特征选择方法[J]. 农业机械学报,2011,42(4):168~173. Zhao Xuehua, Li Daoliang, Yang Wenzhu, et al. Feature selection for cotton foreign fiber objects based on improved ant colony algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011,42(4):168~173. (in Chinese)
- Uncu Özge, Türksen I B. A novel feature selection approach: combining feature wrappers and filters[J]. Information Sciences, 2007, 177(2): 449~466.
- 杨维,李岐强. 粒子群优化算法综述[J]. 中国工程科学, 2004, 6(5): 87~92.
- 杨文柱,李道亮,魏新华,等. 基于自动视觉检测的棉花异性纤维分类系统[J]. 农业机械学报,2009,40(12):177~181. Yang Wenzhu, Li Daoliang, Wei Xinhua, et al. AVI system for classification of foreign fibers in cotton[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(12): 177~181. (in Chinese)
- 刘峤,秦志光,陈伟,等. 基于零范数特征选择的支持向量机模型[J]. 自动化学报,2011,37(2):252~256. Liu Qiao, Qin Zhiguang, Chen Wei, et al. Zero-norm penalized feature selection support vector machine[J]. Acta Automatica Sinica, 2011, 37(2):252~256. (in Chinese)
- Kennedy J, Eberhart R. A discrete binary version of the particle swarm algorithm[C]//IEEE International Conference on Systems, Man and Cybernetics, 1997: 4 104~4 108.
- Vapnik V N. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995: 1~15.
- Ji Ronghua, Li Daoliang. Classification and identification of foreign fibers in cotton on the basis of a support vector machine[J]. Mathematical and Computer Modelling, 2010,51(11~12): 1 433~1 437.