

基于改进的粒子群和蚁群算法的高光谱森林聚类研究

李琰¹, 邢艳秋², 王立海^{2*}

(1. 东北林业大学信息与计算机工程学院, 黑龙江哈尔滨 150040; 2. 东北林业大学森林作业环境研究中心, 黑龙江哈尔滨 150040)

摘要 高光谱图像分类可分为监督分类与非监督分类, 聚类分析进行非监督分类是一种现今比较受研究者广泛关注的技术。粒子群算法具有自适应、自组织性、可同时进行局部和全局搜索等特点; 蚁群算法通过智能个体间不断进行信息交流和传递, 具有较强的发现最优解的能力。提出一种基于改进的粒子群和蚁群算法的高光谱图像聚类方法, 设计其模型并将其应用在森林类型分类问题上, 提高分类精度, 减少人工干预。以吉林省汪清林业局为研究区, 通过修改粒子群的惯性系数, 得出最优解集, 然后利用蚁群寻优的过程对阔叶林、针叶林、混交林、水体进行聚类分析, 区分精度达到 85% 证明, 该方法能较好地识别森林类型。

关键词 粒子群优化算法; 蚁群算法; 遥感图像; 高光谱; 聚类; 森林类型

中图分类号 S126 文献标识码 A 文章编号 0517-6611(2014)27-9615-04

The Research of Hyperspectral Image Clustering for Forest Type Based on Ant Colony Algorithm and Improved PSO

LI Yan¹, XING Yan-qiu, WANG Li-hai^{2*} (1. College of Information and Computer Engineering, Northeast Forestry University, Harbin, Heilongjiang 150040; 2. Forest Operation and Environmental Research Center of Northeast Forestry University, Harbin, Heilongjiang 150040)

Abstract Hyperspectral image classification can be divided into supervised classification and unsupervised classification. Unsupervised classification was conducted on cluster analysis, which is a relatively modern technology by researchers attention. PSO algorithm with adaptive, self-organization, can be used for local and global search simultaneously. Ant colony algorithm has a strong ability to find the optimal solution through continuous exchanging the information and transmission between intelligent individual. This paper presents an improved method for hyperspectral image clustering based on improved PSO and ant colony algorithm, designs the model and applies it in the forest type classification problems for improving classification accuracy, reducing manual intervention. Taking Wangqing Forestry Bureau in Jilin Province as the study area, by modifying the inertia coefficient of PSO, the optimal solution set was obtained. Using optimization process of the ant colony, clustering for forest types can distinguish better between broad-leaved forest, coniferous forest, mixed forest and water bodies. The classification accuracy is 85%. The method can identify forest types.

Key words Particle swarm optimization algorithm; Ant colony algorithm; Remote sensing; Hyperspectral image; Clustering; Forest type

森林资源是陆地生态系统的主体, 对改善空气质量、涵养水源、减少风沙危害等起着重要作用, 在林业资源保护和林业资源动态监测中森林类型识别占有重要地位, 但在传统的林业资源调查中, 不但要耗费大量的人力、物力和财力, 而且很难实现大空间尺度的调查。与传统方法相比, 遥感技术有宏观性强、周期短、可重复处理等优点, 被广泛应用到森林分类中^[1]。高光谱遥感数据有着图谱合一和光谱连续的优势, 能够实现地物的准确识别和精细分类^[2]。但高光谱的缺陷是上百个波段构成的海量数据导致处理难度大, 因此学者们针对如何提高森林类型的分类效率和精度, 尝试不同方法。其中群智能算法是新近研究的热点。通常开始分析遥感图像时, 用需要人工交互较少的非监督分类方法来研究数据的本来结构及自然点群的分布情况, 降低处理数据量^[3]。聚类分析是非监督分类的一种主要方法。仲青青利用粒子群算法进行多光谱聚类研究, 只考虑了对参数选取的改进^[4]。王志辉等应用 EO-1 Hyperion 数据对浙江农林大学植物园内的树种进行森林类型识别, 表明用高光谱遥感数据对树种分类能达到较高的分类精度^[5]。但现有的多数研究是基于国外的遥感卫星, 对于国内的环境与灾害监测预报小卫星关于森林树种识别的研究鲜有报道。同时, 粒子群与蚁群

算法结合鲜有应用于遥感领域。蚁群算法收敛速度快, 比传统进化算法简洁^[6]。为此, 笔者在改进粒子群算法的基础上, 采用两段式将粒子群与蚁群算法融合, 应用到基于高光谱遥感图像的森林类型识别上, 提高分类精度, 减少分类时间, 降低专业理解难度, 减少人工干预, 并用试验进行验证其有效性。

1 数据收集与处理

1.1 研究区域 研究区为吉林省延边朝鲜族自治州东北部汪清县东南部汪清林业局经营区, 地处 129°56' ~ 131°04'E、43°05' ~ 43°40'N。总面积 30.4 万 hm², 其中林地面积为 18.7 万 hm², 为较典型的天然次生林。森林类型以针阔混交林为主, 带状分布。针叶树主要有红松 (*Pinus koraiensis*)、云杉 (*Picea*)、臭松 (*Symplocarpus Salisb*)、落叶松 (*Larix kaempferi*)、阔叶树多为水曲柳 (*Fraxinus maudschurica*)、胡桃楸 (*Juglans maudshurica*)、蒙古栎 (*Quercus mongolica*)、椴树 (*Tilia*)、色木 (*Acermono*)、榆树 (*Ulmus pumila*)、白桦 (*Betula platyphylla*)、杨树 (*Populus*) 和枫桦 (*Betula davurica*) 等^[7]。

1.2 高光谱数据 环境与灾害监测预报小卫星 A、B 星 (HJ1A/B 星) 上搭载 CCD 相机和世界上第 1 个基于空间调制干涉成像原理设计的民用超光谱成像仪 (Hyperspectral Imaging Radiometer, HSI)。HJ-1A 图像幅宽大于 50 km, 地面像元分辨率为 100 m。HJ-1A 共有 115 个波段, 分为蓝色波段 (1~6)、青色波段 (7~19)、绿色波段 (20~40)、黄色波段 (41~49)、橙色波段 (50~58)、红色波段 (59~88)、近红外波段 (89~115)。光谱范围为 0.45~0.9 μm, 平均光谱分辨率为

基金项目 国家自然科学基金面上项目 (41171274); 国家林业“948”项目 (2014-4-78)。

作者简介 李琰 (1980 -), 女, 黑龙江哈尔滨人, 讲师, 在读博士, 从事森林工程方面的研究。* 通讯作者, 教授, 博士, 博士生导师, 从事森林工程方面的研究。

收稿日期 2014-08-05

4.32 nm,且具备 $\pm 30^\circ$ 侧视能力及星上定标功能。重访周期为96 h,可实现对研究区快速重复观测。此次处理的高光谱数据的获取时间为2010年6月9日,光谱景序列号321381,各波段增益0.7^[8]。

1.3 野外调查数据 该研究在2006~2013年进行了4次野外地面调查,按照森林类型、坡度等情况随机布设79个水平投影面积为500 m²的圆形样地,对样地内立木进行每木测量,测量参数包括胸径、树高、树种和郁闭度。野外调查样地的类型分为阔叶林、混交林和针叶林3个类型组^[9]。野外调查的阔叶林、针叶林、混交林的样地数分别为53、11和15个。再将野外数据调查结果与吉林省规划设计院提供的汪清林业局的森林二类调查资料成果相结合,选出验证样本数各为20个。

1.4 数据处理 HIS高光谱数据为2级产品,它的1~20以及113、114和115波段是坏数据,所以实际参与分析的波段为92个。其中21~88波段为可见光部分,90~112波段为近红外部分^[10]。数据首先经过对数残差大气纠正,消除大气和光照等因素对地物反射的影响。然后用地面控制点进行几何精校正。

2 研究方法

2.1 改进的粒子群算法 基于文献[11]描述的基本粒子群算法,研究发现粒子群算法应用于高维复杂问题优化时,容易遇到早熟收敛的难题,该点可能是局部极小点,或局部极小点邻域的一个点。粒子群优化算法早期收敛速度快,但到寻优的后期,算法缺乏有效的机制,使收敛陷入局部极小。

为改善粒子群算法的收敛性能,对公式(1)中惯性权重 ω 进行调整,即:

$$V_{ij}(t+1) = \omega V_{ij}(t) + c_1 r_1 [P_{ij} - X_{ij}(t)] + c_2 r_2 [P_{ij} - X_{ij}(t)] \quad (1)$$

$$X_{ij}(t+1) = X_{ij}(t) + V_{ij}(t+1) \quad (2)$$

式中, $0 < \omega < 1$,基本粒子群算法是一种惯性权重 $\omega = 1$ 的特殊情况。 ω 使得粒子保持运动惯性,拥有扩展空间的能力。 ω 较大时,粒子群具有较好的全局收敛能力, ω 较小时具有较强的局部搜索能力。单纯调整 ω ,当惯性权重递减时,粒子群前期具有较好的搜索能力,但局部搜索能力较弱。惯性权重递增时,后期局部搜索能力较弱。该研究将采用增加加速系数 Φ ,与惯性权重结合,灵活地调整使 ω 先增后减,让粒子群算法即拥有前期收敛较快的优点,又克服后期跳出局部搜索僵局的缺点。令 $\Phi_1 = c_1 r_1$, $\Phi_2 = c_2 r_2$,惯性权重满足 $\Phi_1 + \Phi_2 < 2(\omega + 1)$ 时粒子群优化算法收敛^[12]。 Φ 可随机取值,参考文献[13],提出加速系数:

$$\Phi_1 = (\omega + 1) \times \text{rand1} \quad (3)$$

$$\Phi_2 = (\omega + 1) \times (1 - \text{rand1}) \times \text{rand2} \quad (4)$$

令迭代次数与最大迭代次数之比为 k ,据经验设定:

$$\omega(i) = \begin{cases} k + 0.4 \\ -k + 1.4 \end{cases} \quad (5)$$

式中, i 为迭代次数。该方法理论上能使粒子群算法在前期可以较快收敛,后期局部搜索能力也较好。

2.2 基于改进的粒子群算法与蚁群融合的高光谱聚类方法 蚁群算法解决离散问题较好,而粒子群优化算法解决连续问题较好,该研究提出把这两种算法结合分段使用,先利用粒子群算法进行粗聚类,再将遥感数据离散化,选已初步聚类的较优解集进行精聚类,降低高维数据处理时间,提高分类精度。该方法主要思想:①利用粒子群算法有效的全局搜索特性,进行粗聚类,得到聚类原型,可以对精度不做要求,只要得到聚类原型处于最优解的附近邻域即可;②在①中得到的聚类原型的基础上,利用蚁群算法精确求解得到最优聚类。

2.2.1 粒子个体编码结构。设像元集为 $X = \{X_i, i = 1, 2, 3, \dots, N\}$, X_i 表示某一像元的总特征,为 D 维向量。聚类就是要找到一个划分 $C = \{C_1, C_2, C_3, \dots, C_K\}$,使总的类内离散度和达到最小。

$$J_c = \sum_{j=1}^k \sum_{i \in C_j} d(X_i, C_j) \quad (6)$$

式中, C_j 为第 j 个聚类的中心; $d(X_i, C_j)$ 为像元到对应聚类中心的距离。 J_c 为各类样品到对应聚类中心距离的总和。设粒子群的每个粒子为一个可行解,其解为聚类中心的集合。

每个粒子的结构设计如图1所示。

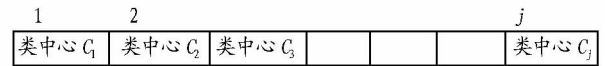


图1 粒子编码表示

每个粒子的位置由 k 个聚类中心组成。当聚类中心确定,聚类的划分可由最邻近法决定。即对被分类像元 X_i, C_j 满足公式(7),则 X_i 属于类 j 。

$$d(X_i, C_j) = \text{mind}(X_i, C_j) \quad (7)$$

适应度值为 J_c 的倒数。

2.2.2 聚类实现。具体步骤如下。

(1)初始化粒子群,给定聚类数目 $centernum$,粒子数量 $N = 20$,将要分类像元随机指派为某一类,作为最初的聚类划分。计算各类的聚类中心,将其作为粒子的位置编码。计算适应度 $1/J_c$ 。粒子的初始速度设为0。反复 M 次。

(2)根据初始粒子群得到粒子的个体最优位置 P_{id} 和全局最优位置 P_{gd} 。

(3)根据公式(1)、(2)更新所有粒子的位置和速度。 ω 按公式(5)取值。

(4)对每个待分像元,根据粒子的聚类中心编码,按公式(7)的最邻近法确定聚类划分。

(5)对每个粒子,按相应的聚类划分计算新的聚类中心,更新粒子的适应度值。

(6)对每个粒子,比较它的适应度值和它经历过的最好位置 P_{id} 和 P_{gd} 的适应度值,如果好于现有的 P_{id} 和 P_{gd} ,则更新 P_{id} 和 P_{gd} 。

(7)不断迭代,直到满足终止条件,如设置的最大迭代次数或解符合某种阈值时,退出。否则继续重复步骤(3)~(6)。

(8)依据上步得出的聚类划分模型,初始化蚂蚁,即从图像上的几个粗分后的类别中各自任意选取一个像素作为该

类别的遗传基因。初始聚类中心^[14]。

蚁群的个体结构编码如图 2 所示。

1	2	k+1				
波段信息	波段信息	类别	接受阈值	信息交换标签	接受程度	年龄

图 2 蚁群编码表示

遥感图像上的每一个像素是一只蚂蚁,每一只蚂蚁都是一个多维向量。波段信息包含 k 个波段,类别标签用 L 表示,初始时候,蚂蚁并未归于任何一类,所以 L 为 0。接受阈值为 T 可动态变化,范围为 $0 \sim 1$,可用相似性函数表示:

$$T = [\text{aver}(i) + \max(i)] / 2 \quad (8)$$

信息交换标签用 M 表示,表示该蚂蚁 i 与遇到的其他蚂蚁交换信息是否成功,成功一次 M 增加 1。蚂蚁幼虫时还没遇到其他蚂蚁,此时 $M = 0$,之后 M 不断增加, M_i 表示了第 i 只蚂蚁所在巢的大小。 M_i^+ 表示蚂蚁 i 在巢中被其他蚂蚁接受的程度。两只蚂蚁彼此接受时 M_i^+ 变大。年龄 A 表示蚂蚁被接受的次数。

每只蚂蚁与随机的任意其他蚂蚁相遇,所有蚂蚁均完成一次搜索后,根据类别标签 L 完成聚类中心更新。此时,没有巢的蚂蚁被分到和它相似的蚁穴中。

(9) 如果没有达到预想设定的最大迭代次数,转向步骤(8)循环执行;否则,输出最优聚类解集。

3 结果与分析

该研究选择环境与灾害监测预报小卫星 A 星的 HIS 高光谱影像,随机将研究区样地指派为针叶林、阔叶林、混交林、非林地、水体等 5 类,并按公式(6)、(7)计算个体的适应度。利用优化算法迭代 20 次,其迭代次数与适应度收敛情况如图 3 所示。

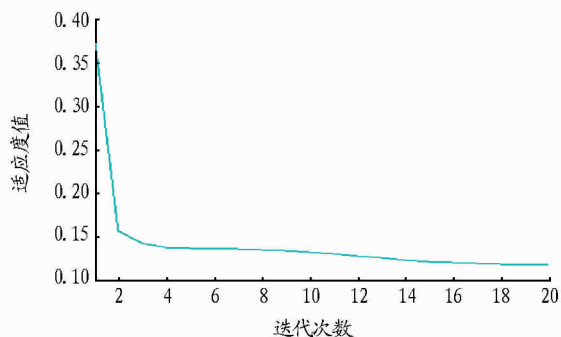


图 3 迭代 20 次与适应度关系

对研究区进行分类,选取迭代第 5 次与第 20 次的不同分类结果展示,如图 4、5 所示。

改进的粒子群与蚁群结合的优化算法中,取粒子个数为 20,当算法第 5 次迭代时适应度值为 0.143 8,第 20 次迭代,适应度值为 0.130 7。选取样地点并辅以试验区小斑数据,计算分类精度。第 5 次迭代的计算结果的总体精度为 72%, $Kappa$ 系数为 0.65,精度如表 1 所示。经过 20 次迭代,并结合森林类型的光谱数据,最终聚类总体精度为 85%, $Kappa$ 系数为 0.812 5,精度如表 2 所示。

从表 1 可以看出,100 个验证样本中有 72 个样本落在正

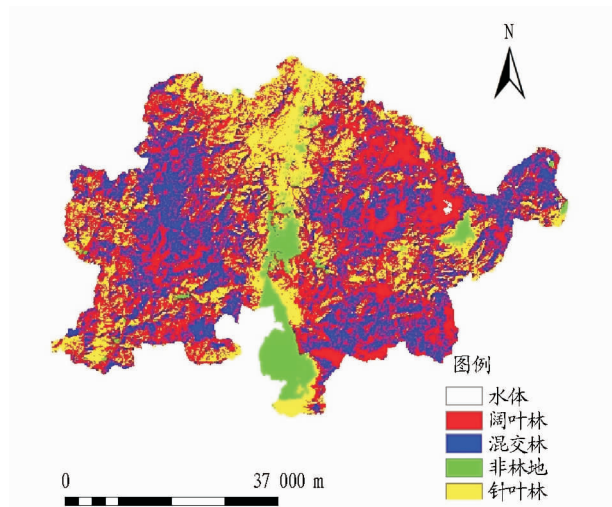


图 4 第 5 次迭代聚类结果

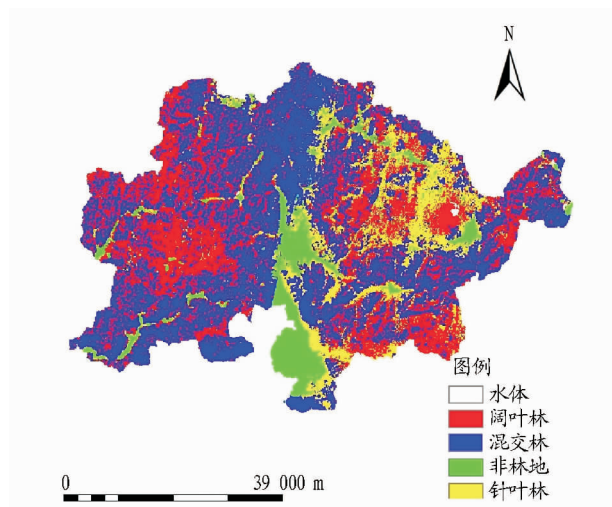


图 5 第 20 次迭代聚类结果

确分类区内,总体分类精度为 72%。其中,混交林的样本中,有 13 个样本落在混交林分类区,3 个落在针叶林分类区,4 个落在阔叶林分类区,分类精度为 65%;针叶林精度达 75%;阔叶林 11 个验证样本分类正确,4 个落在混交林分类

表 1 第 2 次迭代分类精度

类别	混交林	针叶林	阔叶林	水体	非林地	制图精度
	个	个	个	个	个	
混交林	13	3	4	1	2	65
针叶林	3	15	3	0	1	75
阔叶林	4	2	11	1	2	55
水体	0	0	0	18	0	90
非林地	0	0	2	0	15	75
总数	20	20	20	20	20	

注:总体精度 = 72%, $Kappa$ 系数 = 0.650 0。

表2 第20次迭代分类精度

类别	混交林 针叶林 阔叶林 水体 非林地					制图精度
	个	个	个	个	个	%
混交林	15	1	3	1	2	75
针叶林	2	18	0	0	1	90
阔叶林	3	0	17	0	1	85
水体	0	0	0	19	0	95
非林地	0	1	0	0	16	80
总数	20	20	20	20	20	

注:总体精度=85%;Kappa系数=0.8125。

区,3个落在针叶林分类区,分类精度为55%;水体与非林地分类精度分别为90%和75%。

从表2可以看出,经过20次迭代,100个样本中85个样本被正确聚类,总体分类精度为85%。其中,混交林验证样本中,15个落在混交林分类区,2个样本落在针叶林分类区,3个被错分成阔叶林,分类精度为75%;针叶林的20个样本中,有18个样本落在针叶林分类区,1个落在混交林分类区,1个落在非林地分类区,分类精度为90%;阔叶林的20个样本中,有17个样本落在阔叶林分类区,3个落在混交林分类区,精度85%;水体的20个验证样本中,有19个分类正确,1个落在混交林分类区,精度为95%;非林地16个落在非林地分类区,2个落在混交林分类区,1个落在针叶林分类区,1个落在阔叶林分类区,分类精度为80%。

混交林、针叶林、阔叶林、水体、非林地的制图精度在迭代20次后比迭代初级阶段第5次时分别有所提高。

研究发现,水体的反射率值与森林区别较大,混淆程度较低,容易区分,精度相对很高。野外调研多数选在9、10月份,针叶林与阔叶林及混交林差异较大,较容易区分。阔叶林与混交林灰度值相差不大,它们各自均有一些样本混淆,影响了精度。随着迭代次数的增加,适应度收敛,越来越接近最优聚类结果。HIS高光谱数据的质量也影响了精度的提升。但改进的粒子群与蚁群结合的算法进行聚类为森林类型识别提供了一种新思路。

4 结论

该研究在对粒子群引入加速系数、改进惯性系数的基础

上,采用蚁群生化特性,分多阶段将粒子群与蚁群智能算法结合,并设计出了适合于高光谱图像聚类的蚁群和粒子群编码规则及个体结构,将其应用到林业遥感的聚类方法中,将高光谱影像的聚类过程映射为群智能算法的寻优过程,利用不同地物的光谱特征,从HJ/1A遥感影像的115个波段中挑选出类内距离较小的集合,验证样本验证结果表明,其能较好地区分阔叶林、针叶林和混交林等样地类型。该优化算法拥有自学习性、鲁棒性强等特点,可为遥感应用问题提出新的解决思想。

参考文献

- [1] TONG Q X, ZHANG B, ZHENG L F. Hyperspectral remote sensing[M]. Beijing: Higher Education Press, 2006.
- [2] FAZAKAS Z, NILSSON M, OLSSON H. Regional forest biomass and wood volume estimation using satellite data and ancillary data[J]. Agricultural and Forest Meteorology, 1999, 98/99: 417-425.
- [3] 舒清志, 唐守正. 国际森林资源监测的现状与发展趋势[J]. 世界林业研究, 2005, 18(3): 33-37.
- [4] 仲青青. 基于粒子群优化的遥感图像聚类研究[D]. 杭州: 浙江大学, 2011: 4-8.
- [5] 王志辉, 丁丽霞. 基于叶片高光谱特性分析的树种识别[J]. 光谱学与光谱分析, 2010, 30(7): 1826-1830.
- [6] TIMOTHY M K, TALAAT M R. Structural controls on Neoproterozoic mineralization in the South Eastern Desert, Egypt: An integrated field, Landsat-TM, and SIR-C/X SAR approach[J]. Journal of African Earth Sciences, 2002, 35: 107-112.
- [7] 李俊明. 基于森林类型光谱特征的最佳波段选择研究[J]. 森林工程, 2013, 29(4): 4-5.
- [8] 韩瑞梅. 环境星HSI影像处理关键技术的研究山[D]. 长沙: 中南大学, 2010.
- [9] 李俊明, 邢艳秋, 杨超, 等. 基于环境与灾害监测预报小卫星的树种识别[J]. 东北林业大学学报, 2013(11): 42-43.
- [10] FEMANDES C, RAMOS V, ROSA A C. Self-regulated artificial ant colonies on digital image habitats[J]. International Journal of Lateral Computing, 2005, 2(1): 1-8.
- [11] 王东. 基于粒子群智能的遥感找矿方法研究[D]. 长沙: 中南大学, 2008: 24-28.
- [12] LOAN CHRISTIAN T I. The particle swarm optimization algorithm: Convergence analysis and parameter selection[J]. Information Processing Letters, 2003, 85: 317-325.
- [13] 周丽娟. 改进粒子群算法和蚁群算法[J]. 长春工业大学学报: 自然科学版, 2009, 30(3): 342-345.
- [14] 周爽. 蚁群算法在高光谱图像降维和分类中的[D]. 哈尔滨: 哈尔滨工业大学, 2010: 76-82.

(上接第9512页)

- [17] GASSAN J. A study of fibre and interface parameters affecting the fatigue behaviour of natural fibre composites[J]. Composites Part A: Applied Science and Manufacturing, 2002, 33(3): 369-374.
- [18] 周兴平, 解孝林, LI R K Y. PP/PMMA接枝剑麻纤维复合材料的结构和性能[J]. 材料研究学报, 2002, 16(5): 551-556.
- [19] 周兴平, 解孝林, LI R K Y. PP/PMMA接枝剑麻纤维复合材料-(II) SF表面处理对PP/SF复合材料结构和性能的影响[J]. 高分子材料科学与工程, 2004, 20(4): 138-145.
- [20] 王璐琳, 何莉萍, 田永, 等. 表面处理对剑麻纤维增强聚丙烯复合材料力学性能的影响[J]. 机械工程材料, 2008, 32(5): 58-61.
- [21] 汤芬, 甘厚磊, 邹汉涛, 等. 碱处理对剑麻连续长纤维增强聚丙烯复合材料力学性能的影响[J]. 武汉纺织大学学报, 2011, 24(3): 11-15.
- [22] 席建岭, 吴宏武. 剑麻纤维的改性方法对其形态结构和力学性能的影响[J]. 塑料, 2010, 39(1): 42-57.
- [23] 王娟, 李敏, 顾轶卓, 等. 剑麻纤维表面特性及其浸润行为[J]. 复合材料学报, 2012, 29(4): 69-74.

- [24] 李津, 王春红, 贺文婷, 等. 洋麻纤维的表面改性及其在聚丙烯基复合材料中的应用[J]. 工程塑料应用, 2014, 42(2): 6-10.
- [25] GASSAN J, BLEZDKI A K. The influence of fibre-surface treatment on the mechanical properties of jute-polypropylene composites[J]. Composites Part A: Applied Science and Manufacturing, 1997, 28(12): 1001-1005.
- [26] DOAN T T L, GAO S L, MADER E. Jute/polypropylene composites I: Effect of matrix modification[J]. Composites Science and Technology, 2006, 66(7/8): 952-963.
- [27] 管映亭, 董政娥, 温桂清, 等. 苘麻切皮纤维研究[J]. 纺织学报, 2003, 24(6): 572-573.
- [28] JOHN M J, ANANDJIWALA R D. Recent Developments in Chemical Modification and Characterization of Natural Fiber-Reinforced Composites[J]. Polymer Composites, 2008, 29: 187-207.