doi:10. 3969/j. issn. 1671-3168. 2023. 05. 033

基于改进目标进化算法的城市园林 GIS 数据均衡化技术研究

朱海苍,康 骏,杨志伟
(中煤科工重庆设计研究院(集团)有限公司,重庆 渝中 400016)

摘要:传统城市园林 GIS 数据均衡化处理技术忽略了对调整参数的全面寻优,导致部分数据仍存在不均衡问题。为此,提出基于改进目标进化算法的城市园林 GIS 数据均衡化技术。针对城市园林 GIS 数据,结合 SMOTE 算法与遗传算法,形成一种新的改进目标进化算法——GSA 算法,对 GIS 数据完成二次采样,通过鸡群算法求解目标函数最优解,以此得到多个最优调整参数,使图像灰度分布更加均匀,实现城市园林 GIS 数据均衡化处理。实验结果表明,与传统技术相比,改进目标进化算法的 GIS 数据均衡化效果更好,证明该技术具有理想应用性能。

关键词:改进目标进化算法;城市园林;GIS 数据;均衡化技术;GSA 算法;鸡群算法

中图分类号:S731. 2;P208;S711 **文献标识码**:A **文章编号**:1671-3168(2023)05-0201-06

引文格式:朱海苍,康骏,杨志伟. 基于改进目标进化算法的城市园林 GIS 数据均衡化技术研究[J]. 林业调查规划,2023,48(5):201-206. doi:10. 3969/j. issn. 1671-3168. 2023. 05. 033

ZHU Haicang, KANG Jun, YANG Zhiwei. Data Equalization Technology of Urban Landscape GIS Based on Improved Objective Evolutionary Algorithm[J]. Forest Inventory and Planning, 2023, 48(5): 201-206. doi: 10. 3969/j. issn. 1671-3168. 2023. 05. 033

Data Equalization Technology of Urban Landscape GIS Based on Improved Objective Evolutionary Algorithm

ZHU Haicang, KANG Jun, YANG Zhiwei
(CCTEG Chongqing Engineering (GROUP) Co., Ltd., Chongqing 400016, China)

Abstract: The traditional urban landscape GIS data equalization processing technology ignores the comprehensive optimization of the adjustment parameters, which leads to the imbalance of some data. Therefore, a data equalization technology of urban landscape GIS based on improved objective evolutionary algorithm was proposed. A new objective evolutionary algorithm——GSA algorithm was formed by combining SMOTE algorithm and genetic algorithm for urban landscape GIS data. The secondary sampling on GIS data was completed, and the optimal solution of the objective function was solved through chicken swarm algorithm to get more optimal adjustment of parameters, making the image gray scale distribution more uniform and realizing the city garden equalization processing of GIS data. The experimental results showed that compared with the traditional technique, the GIS data equalization effect of the proposed technique was better, which proved that the proposed technique had ideal application performance.

Key words: improved objective evolutionary algorithm; urban landscape; GIS data; equalization technology; GSA algorithm; chicken swarm algorithm

收稿日期:2022-04-18.

第一作者:朱海苍(1982-),男,四川大竹人,硕士,高级工程师. 从事建筑技术和信息技术管理工作.

责任作者:杨志伟(1993-),男,甘肃人,硕士,助理工程师. 从事城市规划及建筑设计研究.

随着社会经济的发展,城市化进程不断加快,城市人口不断增多,因此导致了一系列城市污染问题,如雾霾、热岛效应、水污染等。为缓解城市污染问题,城市园林建设必不可少,不仅如此,园林建设可美化城市环境,为人们提供休闲娱乐场所。为更好地建设城市园林,需要通过 GIS 实时获取城市园林 GIS 数据[1]。GIS 数据,即空间地理数据。通过城市园林 GIS 数据可明确园林建设范围、园林植物种类、园林时空变化规律等。然而,GIS 采集到的数据受各种因素的影响,经常导致数据不均衡,直接影响到数据分析的准确性和效率性[2]。

为优化 GIS 数据的均衡化处理效果,相关领域学者已取得了一些较好的研究成果。曹鹏,李博,栗伟等人提出一种基于 K 近邻和粒子群优化的不均衡数据处理算法,该方法通过优化算法来求解以不均衡数据分类评价准则作为目标的函数,从而得到最优的重采样倍率,以此实现数据均衡化。赵伟则针对遥感图像 GIS 数据,利用一种新型的启发式算法对图像进行均衡化增强处理,使处理后的图像灰度分布更加均匀,提高了图像的对比度和图像细节程度。余发军,刘义才利用改进的粒子群算法设计了一类新颖的灰度图像增强方法,实现了在直方图均衡化的基础上进一步改进图像质量,平衡图像灰度系数。

虽然目前该领域的研究成果有很多,但这些研究均只针对一种类型的 GIS 数据进行研究,具有一定的局限性。本文在前人研究的基础上,利用两种改进目标进化算法,分别对 GIS 数据和 GIS 图像进行均衡化处理,以提高后期 GIS 数据利用率,提高 GIS 数据分析质量,为城市园林建设提供可靠依据。

1 城市园林 GIS 数据重采样

在城市园林 GIS 数据类型后期分析过程中,需保证各类样本数量之间保持在同一水平上才能使分析结果满足要求。然而,实际情况下城市园林 GIS 数据样本类型很难保持均衡。数据均衡化是指增加少数类样本或者减少多数类样本,使不同类型数据样本数量大致相同[3]。由上述描述可知,GIS 数据均衡化问题即解决数据样本的不平衡问题,而样本的不平衡问题的最有效解决方式就是重采样,采用的方法包括决策树、支持向量机(Support Vector Machine,SVM)、k 近邻(k-Nearest Neighbor,k-NN)等,但这些方法在不断产生新的 GIS 样本数据过程中并不能改变原有样本分布的外围轮廓特征,这就意味着对分类问题中分类边界的影响比较小[4]。基于上述传统方法存在的问题,本研究将 SMOTE 算法与遗传算法结合在一起,形成一种新的改进目标进化算法,对城市园林 GIS 数据类型的数据进行均衡化处理。

1.1 SMOTE 算法

SMOTE 算法,英文全称 Synthetic Minority Oversampling Technique,中文为基于人工合成样本的过采样算法[5]。该算法具体过程如下:

步骤 1:假设有一个少数类 GIS 数据样本 Smin,通过 K-最近邻算法在数据集中搜索其到全体样本集中最近的另外 k 个样本。K-最近邻算法基本流程如下:

step1:输入 GIS 数据样本集;

step2:对 GIS 数据样本进行去噪、清洗、降维等预处理;

step3:从 GIS 数据样本集中随机选取一个测试样本点;

step4:计算测试样本点到 GIS 数据样本集中其他每个样本点的距离。计算公式有 3 种:

1)欧式距离:

$$D(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{1}$$

2)曼哈顿距离:

$$D(x,y)=\sum_{i=1}^{n}|x_i-y_i| \tag{2}$$

3)闵可夫斯基距离:

$$D(x,y)=\sqrt[p]{\sum_{i=1}^{n}|x_i-y_i|^p} \tag{3}$$

式中:x 和 y 为两个 n 维向量。

step5:根据计算出来的距离,按照从小到大的顺序进行排序;

step6:选择出距离测试样本点最近的 k 个点,即得到 k 个样本;

step7:根据少数服从多数的原则,完成样本分类。

步骤 2:根据样本分类结果,确定 GIS 数据样本集中各类别数据样本的不平衡比例。

步骤 3:根据不平衡比例设置一个少类样本的采样倍率。

步骤 4:从少数样本集中循环选择样本,记为 x_i。

步骤 5:对 x_i 的 k 个最近邻中的每一个样本按照下述公式构建新的样本。

$$x_{new} = x_i + \text{rand}(0,1) \cdot (x_{old} - x_i) \quad (4)$$

式中:rand(0,1)为(0,1)的随机常数。

步骤 6:将新样本与原有少数样本进行混合,生成新的样本,使样本数量与 GIS 数据集中其它样本数量相同[6]。

1.2 遗传算法

遗传算法[8-9]是在达尔文进化论的基础上衍生出来的一种算法,优者胜利、劣者淘汰,最终留下来的都是优秀的基因,因此,遗传算法也可以说是目标进化算法[7]。该进化算法基本流程如图 1 所示。

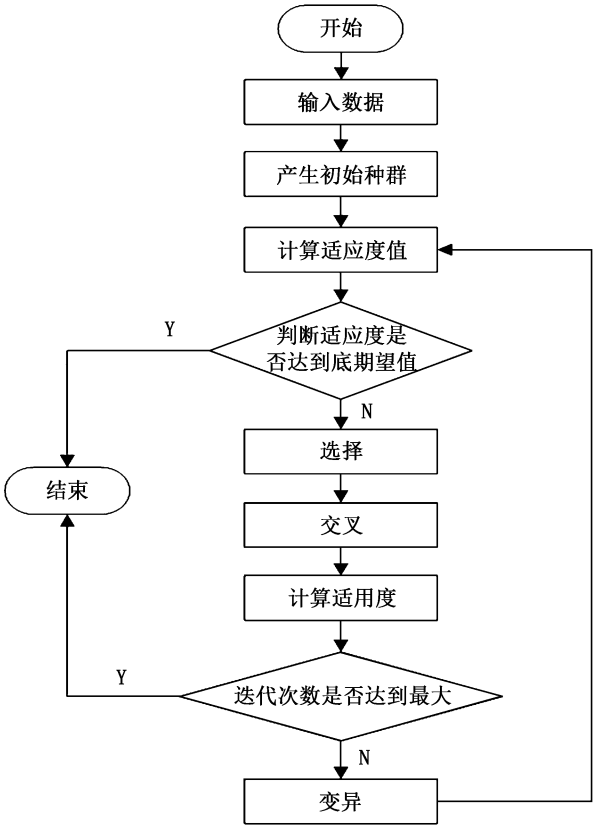

图 1　遗传算法基本流程

Fig. 1　Basic flow of genetic algorithm

1.3 基于 GSA 算法的 GIS 图像重采样

基于上述研究的两种单一算法,在本章节将两种算法集合在一起,构成改进的目标进化算法——GSA 算法。该算法对不平衡 GIS 数据样本进行均衡化的具体过程如下:

步骤 1:输入原始 GIS 数据样本,并设置合成新样本的数量以及循环变量。

步骤 2:对 GIS 数据样本进行预处理。

步骤 3:判断循环变量是否等于合成新样本的数量? 若等于,则结束操作,否则,则进行如下操作:

Step1:进行遗传算法中的选择操作,选中待重构的少数类样本;

Step2:从待重构的少数类样本邻域中随机选择一个近邻样本;

Step3:判断近邻样本与待重构的少数类样本是否属于同一类别? 若属于同一类别,则利用交叉算子合成新样本,否则利用变异算子合成新样本。

步骤 4:保存合成的新样本,并标记为少数类。

步骤 5:重复执行上述步骤,直至循环变量等于合成新样本的数量,算法结束,得到若干个新样本。

步骤 6:将合成的新样本添加到原始 GIS 数据样本中,得到均衡化后的 GIS 数据集。

2 基于鸡群算法的城市园林 GIS 数据均衡化的实现

在采集过程中,GIS 图像受到光线、自然天气以及采集设备本身的影响,图像像素分布并不均衡,严重影响到后续图像识别的准确性和效率,因此需要对 GIS 图像进行均衡化处理,使增强后的图像灰度分布更加均匀[10]。在此采用鸡群算法进行城市园林 GIS 图像均衡化处理,其基本原理框架如图 2 所示。

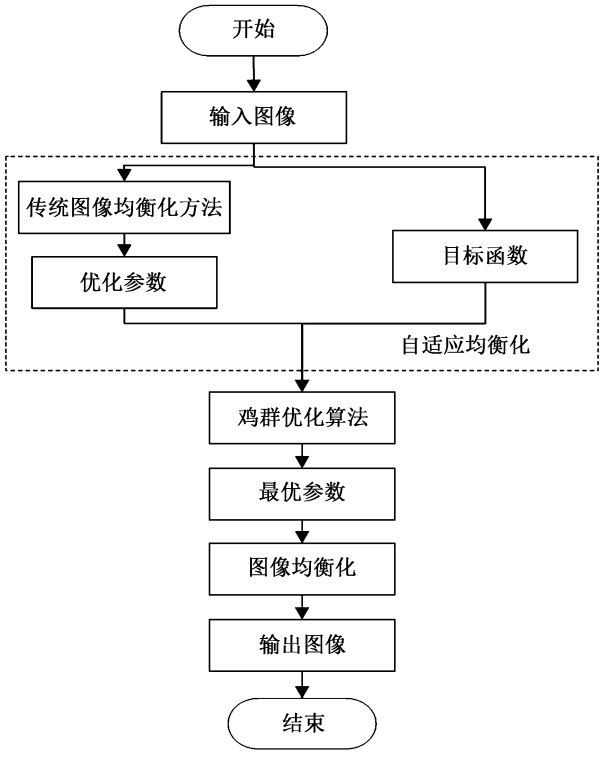

图 2　基于鸡群算法的城市园林 GIS 图像均衡化处理

Fig. 2　Image equalization of urban landscape GIS based on chicken swarm algorithm

2.1 GIS 图像均衡化的目标函数

GIS 图像均衡化的目标函数是一个以图像灰度信息熵为目标的方程,公式表达式为:

$$S = -\sum_{i=0}^{l-1} p_e(i) lb[p_e(i)] \tag{5}$$

式中:$p_e(i)$为像素值为 i 的概率;lb 为图像的灰度级数目,取值为 256[11]。

2.2 鸡群算法求解目标函数

鸡群算法是改进目标进化算法中的一种。利用该算法对上述公式进行求解,具体过程如下:

步骤 1:输入原始 GIS 图像。

步骤 2:颜色空间转换,即 RGB 空间与 YUV 空间之间的转换[12]。转换公式为:

$$\begin{cases} Y = 0.299R + 0.587G + 0.114B \\ U = 0.147R + 0.289G + 0.436B \\ V = 0.615R + 0.515G + 0.100B \end{cases} \tag{6}$$

$$\begin{cases} R = Y + 1.14V \\ G = Y - 0.39U - 0.58V \\ B = Y + 2.03U \end{cases} \tag{7}$$

步骤 3:利用不同尺度的高斯函数滤波器对 GIS 图像进行滤波,去除噪声,公式表达式为:

$$R(x,y) = \sum_{i=1}^{N} w_i \{\log I(x,y) - \log[f_i(x,y) \cdot I(x,y)]\} \tag{8}$$

式中:$f_i(x,y)$为不同尺度的高斯函数滤波器;N 为高斯核的个数;w_i 为对应的权重;$R(x,y)$为滤波后的图像[13-14]。

步骤 4:对滤波后的 GIS 图像进行 Tetrolet 变换[15],得到高频分量系数。

步骤 5:参数设置,包括鸡群的规模、维度[16]。

步骤 6:计算初始化鸡群的位置,计算公式为:

$$x_{ij} = \text{rand}(p_j - L_j) + L_j \tag{9}$$

式中:p_j、L_j 分别为个体所处位置在第 j 维空间取值的最大值和最小值;x_{ij} 为第 i 个个体在第 j 维空间的取值;rand 为 0 到 1 之间的随机数[17]。

步骤 7:基于上述步骤,对高频分量系数进行非线性处理,公式表达式为:

$$\widetilde{Q}_k^l(i) = \begin{cases} 0, Q_k^l(i) < T_k^l \\ f[\overline{Q}_k^l(i)] \geqslant T_k^l \end{cases} \tag{10}$$

式中:$\widetilde{Q}_k^l(i)$ 为非线性处理后的高频子带系数;T_k^l 为第 l 个尺度上第 k 分量中选取的阈值[18];l 为尺度;k 为高频子带索引。

步骤 8:利用 Tetrolet 进行重构[19]。

步骤 9:Gamma 校正,校正公式为:

$$R'(x,y) = [R(x,y)]^{\frac{1}{r}} \tag{11}$$

式中:$R'(x,y)$为校正后的图像;r 为校正参数,大于 0。

步骤 10:计算种群位置的适应度值[20]。

步骤 11:基于上一步骤结果划分种群中个体种类。

步骤 12:位置更新[21],更新公式如(9)、(12)和(13)。

$$x_{ij}(t+1) = x_{ij}(t) \cdot [1 + \text{Randn}(0,d)] \tag{12}$$

式中:Randn(0,e)是均值为 0、标准差为 d 的随机数。

步骤 13:个体适应值更新。

步骤 14:是否满足子群的种类更新代数?满足,则更新子群个体成员之间的关系;否则,则回到步骤 10。

步骤 15:是否满足最大迭代次数?若不满足,则继续迭代。若满足,则输出最优参数,包括非线性调整强度系数 k、非线性调整系数 a 和非线性调整偏移系数 b,并通过下述公式进行非线性调整[22]。

$$f(x) = k\text{sgn}(x) \frac{1}{1 + e^{-ax}} + b \tag{13}$$

式中:x 为高频子带系数;sgn(x)为系数的符号。

步骤 16:图像重构。

步骤 17:Gamma 校正[23],得到均衡化的 GIS 图像。

3 仿真实验分析

为验证所研究的基于改进目标进化算法的城市园林 GIS 数据均衡化技术的有效性,在 MATLAB 2017 b 平台上进行仿真测试。测试包括 GIS 数据测试和 GIS 图像测试两部分。

3.1 实验准备

3.1.1 实验数据

实验所需要的 GIS 数据如表 1 所示。

表 1 GIS 实验数据

Tab. 1 Experimental data of GIS

GIS 数据样本	样本数量	特征数目	少数类	多数类
样本 1	1665	3	69	656
样本 2	5469	9	354	6565
样本 3	4365	7	284	5422
样本 4	6146	8	401	5781
样本 5	4652	5	350	3548
样本 6	534	3	52	554

3.1.2 参数设置

GSA 算法参数设置见表 2。

表 2　GSA 算法参数设置

Tab. 2　GSA algorithm parameters

名称	参数	名称	参数
种群大小	8.00	最大终止代数	30.00
交叉概率	0.85	终止遗传代数	100.00
变异概率	0.10	近邻样本数量	7.00

3.1.3 实验图像

试验图像如图 3 所示。

图 3　实验 GIS 图像

Fig. 3　Experimental GIS images

3.2 不同技术的应用性能测试

鸡群算法参数设置见表 3。

表 3　鸡群算法参数设置

Tab. 3　Parameters of chicken swarm algorithm

名称	参数	名称	参数
鸡群大小	10.00	阈值系数	0.50
公鸡的比例	0.15	子群更新代数	5.00
母鸡的比例	0.70	最大进化代数	50.00
小鸡的母鸡比例	0.50	高斯函数滤波器尺度	250/80/20
维数	4.00		

GIS 图像均衡化评价指标包括熵(E)、标准差(Std)、均亮度(ML)3 个。计算公式为:

1)信息熵(E)

$$E = -\sum_{m=0}^{L-1} T(m)\log_2 T(m) \tag{14}$$

2)标准差(Std)

$$\mathrm{Std} = \sqrt{\frac{\sum_{l=1}^{L}\sum_{i=1}^{M}\sum_{j=1}^{N}[f^l(i,j) - \bar{f}^l]}{\mathrm{LMN}}} \tag{15}$$

3)均亮度(ML)

$$\mathrm{ML} = \frac{\sum_{l=1}^{L}\sum_{i=1}^{M}\sum_{j=1}^{N} f^l(i,j)}{\mathrm{LMN}} \tag{16}$$

式中:$T(m)$是指灰度级 m 在整幅图像灰度值中出现的概率;M、N 为影像的行列像素数;L 表示灰度级范围;f^l 为(i,j)位置处的灰度值;$\bar{f}^l$ 为灰度平均值。

信息熵越大,标准差越小,均亮度越大,像素分布越均匀,均衡化效果越好。

相同仿真条件下,利用引言中提到的赵伟法,王旖旎法对 GIS 图像进行均衡化处理,统计信息熵(E)、标准差(Std)、均亮度(ML)3 个指标,不同技术下各项指标统计结果如表 4 所示。

表 4　不同技术下信息熵、标准差、均亮度统计

Tab. 4　Statistical results of information entropy, standard deviation and average brightness

技术	信息熵(E)	标准差(Std)	均亮度(ML)
改进目标进化算法	4.658	36.31	154.290
K 近邻和粒子群优化算法	3.859	55.15	130.460
改进量子进化算法	3.968	51.95	149.251

从表 4 可以看出,采用改进目标进化算法,信息熵更大、标准差更小、均亮度更大,由此说明基于改进目标进化算法的城市园林 GIS 图像均衡化在技术上是有效的。

4 结　论

为更好地掌握城市园林时空变化情况,GIS 技术被广泛应用,因此产生了很多 GIS 数据。为此,本研究进行基于改进目标进化算法的城市园林 GIS 数据均衡化技术研究。该研究通过仿真测试,证明了本方法无论在 GIS 数据还是 GIS 图像处理上均得到良好的均衡效果,且与其他方法相比表现更为优秀,证明了改进目标进化算法的有效性。

参考文献：

[1] 李发明,李娜,王婷婷,等. 基于 BD-GIS 耦合模型的北京奥林匹克森林公园景观优化策略研究[J]. 中国园林,2020,36(1):65-69.

[2] 刘慧民,宫思羽,邹铁安,等. 基于 GIS 技术规划城市节水型绿地系统的研究[J]. 农业工程学报,2019,35(6):287-295.

[3] 钟昌康. 基于 K 近邻和粒子群优化的特征选择算法[J]. 现代计算机,2020(9):21-24.

[4] 余发军,刘义才. 基于改进量子进化算法的稀疏特征提取方法[J]. 北京理工大学学报,2020,40(5):56-62.

[5] 李国和,张腾,吴卫江,等. 面向机器学习的训练数据集均衡化方法[J]. 计算机工程与设计,2019,40(3):812-818.

[6] 张宗堂,王森,孙世林. 一种针对不平衡数据分类的集成学习算法[J]. 山东大学学报（工学版),2019,49(4):32-37.

[7] 向伟,王新维. 基于多类邻域三支决策模型的不平衡数据分类[J]. 计算机科学,2020,47(5):111-117.

[8] 苏俊宁,叶东毅. 基于样本密度峰值的不平衡数据欠抽样方法[J]. 计算机应用,2020,40(1):83-89.

[9] 张忠林,曹婷婷. 基于重采样与特征选择的不均衡数据分类算法[J]. 小型微型计算机系统,2020,41(6):209-215.

[10] 徐玲玲,迟冬祥. 面向不平衡数据集的机器学习分类策略[J]. 计算机工程与应用,2020,56(24):18-33.

[11] 赵春丽,董静薇,徐博,等. 融合直方图均衡化与同态滤波的雾天图像增强算法研究[J]. 哈尔滨理工大学学报,2019,24(6):97-101.

[12] 赵爱玲,张鹏程,杨一鸣,等. 改进的直方图均衡化与双边滤波的 X 射线图像对比度增强算法[J]. 中北大学学报(自然科学版),2020,41(6):90-96.

[13] 林昌,陈武,周海峰. 基于全局亮度自适应均衡化的海上图像带色彩恢复的多尺度 Retinex 算法[J]. 科学技术与工程,2020,20(33):209-214.

[14] 王小春,葛雨凡,杨晨箫,等. 基于多尺度几何分解的光照不均图像增强方法[J]. 数学的实践与认识,2019,49(21):151-158.

[15] LV M. Agricultural climate change and multilingual GIS database translation system based on embedded database and artificial intelligence[J]. Arabian Journal of Geosciences,2021,14(11):1-20.

[16] 车森,杨辉,葛磊,等. 一种面向在线地图叠加的 GIS 数据几何校正方法[J]. 测绘科学技术学报,2020,37(4):95-98,107.

[17] 魏芬娟,陈崇成,叶晓燕,等. 森林资源调查与督查的移动 GIS 研发及应用[J]. 计算机工程与应用,2020,56(6):237-244.

[18] 董秀军,许强,佘金星,等. 九寨沟核心景区多源遥感数据地质灾害解译初探[J]. 武汉大学学报(信息科学版),2020,45(3):119-128.

[19] JAKOVLJEVIC G,GOVEDARICA M,ALVAREZ-TABOADA F. Waterbody mapping: a comparison of remotely sensed and GIS open data sources[J]. International Journal of Remote Sensing,2019,40(7-8):2936-2964.

[20] ASHTEKAR A S,MOHAMMED-ASLAM M A,MOOSVI A R. Utility of normalized difference water index and GIS for mapping surface water dynamics in sub-upper krishna basin[J]. Journal of the Indian Society of Remote Sensing,2019,47(8):1431-1442.

[21] 夏小科,贾庆仁,杨泉,等. 一种面向三维 WebGIS 的空间数据加载优化方法[J]. 武汉大学学报(信息科学版),2020,45(12):172-179.

[22] MOHAMED S A. Application of satellite image processing and GIS-Spatial modeling for mapping urban areas prone to flash floods in Qena governorate,Egypt[J]. Journal of African Earth Sciences,2019,158:1-103507.

[23] 史北祥,杨俊宴. 基于 GIS 平台的大尺度空间形态分析方法——以特大城市中心区高度、密度和强度为例[J]. 国际城市规划,2019,34(2):111-117.

责任编辑：许易琦

(上接第 130 页)

参考文献：

[1] 克德尔汗,吴金莲. 新疆夏尔希里自然保护区生态质量评价研究[J]. 新疆环境保护,2008,30(1):25-28.

[2] 赵金发,张宓,郑进烜,等. 云南省自然保护地建设发展研究[J]. 林业调查规划,2020,45(3):69-75.

[3] 国家林业局. 自然保护区自然生态质量评价技术规程:LY/T 1813-2009[S]. 北京:中国标准出版社,2009.

[4] 云南省林业调查规划院生态分院. 永善县五莲峰市级自然保护区总体规划(2021—2030 年)[Z]. 2021.

[5] 国家环境保护局,国家技术监督局. 自然保护区类型与级别划分原则:GB/T 14529—93[S]. 北京:中国标准出版社,1994.

责任编辑：杨竣焓