

文章编号: 1007-4619 (2000) 01-0066-05

基于 GIS 的贝叶斯统计推理技术在印度 野牛生境概率评价中的应用

张洪亮¹, 李芝喜¹, 王人潮², 张军³, 孟鸣³

(1. 西南林学院 资源学院, 云南 昆明 650224; 2. 浙江大学 环境与资源学院, 浙江 杭州 310029;

3. 云南省地理研究所, 云南 昆明 650223)

摘要: 目前, GIS 技术已被广泛应用在野生动物生境研究中。但是, 作为空间数据分析和处理工具, GIS 缺乏进行启发式推理的能力。因此, 与擅长于此的贝叶斯统计推理技术相结合则是解决这一问题的重要途径。以西双版纳纳板河流域生物圈保护区为试验区, 综合应用 GIS 技术和多元统计技术建立印度野牛生境的两个逻辑斯蒂多元回归模型: 趋势表面模型和环境模型, 第一个模型的自变量是位置坐标, 第二个模型的自变量是一组环境因子, 然后应用贝叶斯统计合并这两个模型产生贝叶斯综合模型。结果表明, 贝叶斯综合模型优于环境模型, 可应用于野生动物生境概率评价。

关键词: 地理信息系统; 贝叶斯统计推理; 生境; 西双版纳

中图分类号: TP79/Q958 **文献标识码:** A

1 引言

生境(habitat)一词首先由美国的 Grinnel 于 1917 年提出, 指的是生物的居住场所, 即生物个体、种群或群落能在其中完成生命过程的空间。自然界的生物都有它特定的生活环境, 都有各自要求的适宜的环境条件。传统生态学和动物地理学的研究手段常为如何利用空间属性和时间坐标来收集信息的问题所困扰, 而且在收集具有时空属性特征的数据以及在分析、解释和表达结果的手段上都存在很大的局限性。

随着地理信息系统的迅速发展, 为野生动物生境的研究提供了极大的潜力。国内外许多研究事例都表明, GIS 技术的引入使野生动物生境的研究水平有了较大的提高^[1,2]。但是, 作为空间数据分析和处理工具, GIS 缺乏进行启发式推理的能力。为了使野生动物生境的研究上升到一个更高的水平, 本文在野生动物生境的 GIS 模型中引入贝叶斯统计推理成功地对印度野牛的生境概率进行了评价。

2 研究地区概况

西双版纳位于云南省南部边缘, 与缅甸、老挝接壤。因为其独特的地理位置、地貌形态和气候条件, 使之成为地球上热带林最北端的分布区, 同时也是中国生物多样性荟萃地。试验区为纳板河自然保护区, 它位于西双版纳州景洪市与勐海县接壤处, 地理坐标为 100°32'—100°44' E, 22°04'—22°16' N, 土地总面积为 261 km², 纳板河自北向南穿过保护区中部, 流入澜沧江。

保护区内动植物资源十分丰富, 具有西双版纳所分布的各种植被类型, 即: 热带雨林、热带季雨林、亚热带常绿阔叶林、落叶阔叶林、暖性针叶林等。

3 研究方法

本研究的技术路线是在印度野牛活动位置信息和与印度野牛生境相关因子数据度的支持下, 综合应用 GIS 技术和统计技术建立印度野牛生境的两个逻辑斯蒂多元回归模型, 趋势表面模型和环境模

收稿日期: 1998-12-28; 修订日期: 1999-04-19

基金项目: 云南省科委资助项目(项目编号: 98C013Q)。

作者简介: 张洪亮(1969—), 男, 博士。主要从事遥感与 GIS 应用领域研究, 发表论文 7 篇。

型。以第一个模型产生的概率作为贝叶斯统计推理的先验概率,用环境模型的信息去修正这些概率产生贝叶斯综合模型,进行其生境概率评价。

3.1 贝叶斯统计推理技术

贝叶斯定理

某一对象被观测以前,我们仅仅知道这一对象的先验概率。观测之后,不管怎样,我们应该能利用此观测值和条件概率密度函数去改进我们关于对象种类从属关系的知识。对象被观测之后,这一对象属于种类 i 的所谓后验概率由贝叶斯定理给定如下,即

$$P(C_i | x) = \frac{P(x | C_i)P(C_i)}{\sum_{i=1}^n P(x | C_i)P(C_i)} \quad (i = 1, 2, \dots, n)$$

贝叶斯定理允许我们合并种类从属关系先验概率,用条件概率密度函数和观测值计算被观测对象属于某一类别的概率。在给定这些信息的条件下,我们可有选择地为每一对象赋予它最有可能的种类。

贝叶斯估计

贝叶斯估计方法把未知参数作为随机变量。而且,事前可以提供未知参数的某些附加信息。贝叶斯估计假设在抽样试验以前,未知参数假定有一个先验概率密度函数。在训练集被观测之后,使用贝叶斯定理利用抽样值更新先验概率密度函数。结果产生未知参数值的后验概率密度函数。有关贝叶斯估计的步骤如下,首先,我们为未知参数假定一个先验概率密度函数;第2,我们靠观测训练集从总体中获取样本值;第3,利用抽样值,我们使用贝叶斯定理更新先验概率密度函数成后验概率密度函数;最后,形成样本和未知参数的联合密度,并对后者作积分运算以保留概率密度函数的理想估计值。

3.2 印度野牛生境的贝叶斯综合模型建立

工作环境

该项研究的硬件环境以计算机 SGI 工作站局域网为核心,包括 5 台 INDY 工作站及微机终端若干,Contex FSS 5200 扫描仪,HP Designjet 750C 绘图仪等。地理信息系统软件选用目前世界上最为流行的 ARC/INFO Version 7.0.3 和 ARCVIEW Version 3.0。

数据库的建立^[3]

印度野牛生境数据库包括印度野牛活动位置信息(225 个活动点和 321 个非活动点)数据库;与地形

有关的生境因子(海拔、坡度、坡向)数据库;与生物物理相关的生境因子(硝塘、森林植被类型)数据库及与人类活动相关的生境因子(居民点)数据库 4 个组成部分。

建立图形数据库时,使用 Contex FSS 5200 扫描仪进行图形输入。为了保证图形质量,在纳板河保护区选择了 6 个控制点,使控制点在地形图(1:50,000)上的残余误差小于 0.2mm。在保证图形精度的前提下,进行坐标转换、图形编辑和拓扑关系的建立,最后输入属性数据。

属性数据库,除了基于 Coverage 的数据类型,自动产生诸如周长、面积、长度之类的空间图形信息之外,还需要增加用户自定义的属性数据。以便识别地表特征并建立图形与属性之间的对应关系。

先验概率的准备——趋势表面模型(TSM)的建立

印度野牛生境适宜性分析的趋势表面模型是一个利用栅格单元的 x, y 坐标作为解释变量的 n 阶多项式的一个多元逻辑斯蒂回归模型。我们在 ARC/INFO 系统的 grid 模块下用 trend 命令生成。Trend 表面插值器利用多元回归使一个最小的正方形表面逼近输入点,以找出最佳适配方程来生成整个表面。

随着多项式阶数增加,表面的处理会更复杂,阶数越高并不一定能提高表面生成的精确性。最好的阶数选择取决于数据。列出的 RMS 错误越小则插值生成的表面更接近于输入点。所以最终采用二阶多项式生成了印度野牛生境的趋势表面模型。

修正信息源的准备——环境模型(EM)的建立

环境模型的建立使用 Logistic 多元回归模型。Logistic 回归模型是一个二分因变量对一组包括数字和类别自变量的回归,回归的结果预测因变量的概率。这种方法已经被成功地应用在野生动物生境研究中^[4-6]。在本研究中,印度野牛活动位置数据(活动点或非活动点)是因变量。自变量包括一组与印度野牛生境相关的环境变量。

印度野牛活动位置点的森林植被类型的确定是通过点对多边形叠加操作产生。居民点和硝塘对印度野牛生境的影响是通过印度野牛活动位置点到居民点和硝塘的欧几里得距离确定。在 ARC/INFO 的 grid 模块下用 Eudistance 命令实现。

Logistic 回归分析以前,首先对每个自变量进行了显著性检验。结果表明,森林植被名词性变量和硝塘、居民点两个数字性变量为显著自变量¹⁾。这

1)张洪亮,李芝喜,王人潮等. 印度野牛生境模型研究,待发表.

些自变量被用来建立印度野牛生境的环境模型。逻辑斯蒂回归要求 n 个名词变量转换成 $n - 1$ 个二分亚变量和一个额外的参考类^[7]。此种情况下,5类森林植被的“其它类”被当作参考类,其余4种森林植被类型被转换成4个二分亚变量(某一森林植被类型的存在为1,不存在为0)。在546个样点中,随机选取2/3的样点数据(364)用来建立模型,而其余1/3的样点数据(182)则被用来进行后面的模型验证。结果得到印度野牛生境的环境模型如下:

$$Y = -1.825 - 0.001X_1 + 0.001X_2 + 1.170X_3 + 2.521X_4 + 3.011X_5 + 0.710X_6 \quad (1)$$

其中: X_1 为至硝塘的距离; X_2 为至居民点的距离; X_3 为热带季节雨林、季雨林的存在; X_4 为亚热带季风常绿阔叶林的存在; X_5 为竹林混交林、次生植被的存在; X_6 为亚热带山地苔藓常绿阔叶林的存在。

$$P = \frac{e^y}{e^y + 1} \quad (2)$$

其中, P 为印度野牛在某一位置的生境适宜性概率, Y 为逻辑斯蒂方程指数,也是自变量的加权线性组合。输出概率的范围为0—1。

贝叶斯综合模型(BIM)

利用贝叶斯统计推理技术合并环境模型和趋势

表面模型产生贝叶斯综合模型。此技术在遥感领域广泛使用,例如,地形信息为已知植被类型的像元提供先前概率,使用光谱信息修正这些概率,导致植被分类精度的提高^[8]。对印度野牛生境适宜性分析来说,趋势表面模型产生先验概率,环境模型作为其它信息源去修正这些概率。此步骤的数学公式如下^[9]:

$$P_{BIM} = \frac{1}{1 + e^{\{\log[(1 - P_{EM})/P_{EM}] - \log[P_{TSM}/(1 - P_{TSM})]\}} \quad (3)$$

P_{BIM} 为贝叶斯综合模型生成的预估概率, P_{EM} 和 P_{TSM} 分别是由环境模型和趋势表面模型产生的概率预估值。

4 BIM 和 EM 的比较

在 ARC/INFO 的 grid 模块下,根据所有 GIS 层的最小多边形确定栅格大小为 $50m \times 50m$ (这样以便在最大水平上保留空间信息),然后应用方程(1),(2),(3)对有关自变量的 GIS 层进行运算,得到了整个研究区由贝叶斯综合模型产生的印度野牛生境概率图(封三图版 I 图 1)和由环境模型产生的印度野牛生境概率图(封三图版 I 图 2)。

贝叶斯综合模型执行统计(图3(a))和(3(b))

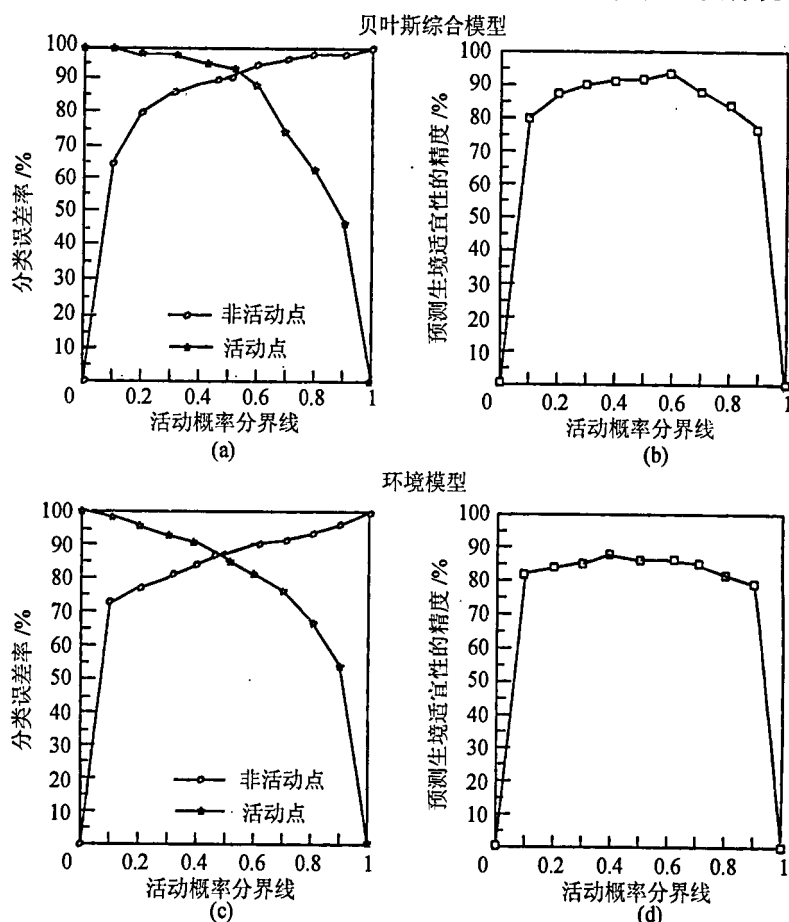


图 3 贝叶斯综合模型,环境模型,分类误差率和预测生境适宜性的精度

Fig. 3 Classification error rates and predictive accuracy of habitat suitabilities for bayesian integrated model and environmental model

表明概率 0.6 是最适分界线,环境模型执行统计(图 3(c)和 3(d))表明概率 0.4 是最适分界线,以此分界线分别得到由 BIM 和 EM 产生的印度野牛生境适宜性评价图(封三图版 I 图 4,5)。此时,我们利用其余 1/3 样点数据(75 个活动点和 107 个非活动点)来计算两个模型的分精度,即将栅格化的印度野牛生境适宜性评价图与栅格化的 1/3 样点数据图进行叠加,分别统计得到在印度野牛生境适宜区和非适宜区中活动点和非活动点的数量。结果得到两个模型的分精度(表 1)。从表 1 可看出贝叶斯综合模型对印度野牛生境适宜性的分精度比单独使用环境模型提高了 5.5%。同时,比较封三图版 I 图 1 与图 2、封三图版 I 图 4 与图 5,可看出贝叶斯综合模型得到微小的生境适宜区。贝叶斯综合模型的状态揭示了分量模型的影响,在高概率区域具有清晰的平滑度,在低概率过渡区域仍然是呈锯齿状的,这主要是由于环境模型的控制。

表 1 贝叶斯综合模型(BIM)与环境模型(EM)分类精度比较

Table 1 A comparison of the classification accuracy of BIM and EM

实际类别	预测类别				精度/%	
	活动点		非活动点			
	BIM	EM	BIM	EM	BIM	EM
活动点(75)	66	68	9	7	88.0	90.7
非活动点(107)	5	17	102	90	95.3	84.1
总验证样点(182)					92.3	86.8

5 讨论

有效的辅助决策是 GIS 发展的重要方向。本文在 GIS 的基础上应用贝叶斯统计推理技术建立了印度野牛生境的贝叶斯综合模型,并与环境模型相比,结果是令人满意的,为野生动物生境的定量研究提供了一种新的研究方法。同时,在 GIS 基础上建立野生动物生境评价的贝叶斯概率算法专家系统有待

于进一步研究和探讨。

致谢:参加该项目的还有西双版纳纳板河自然保护区曹光宏,杨正兴,田维祥等;云南省环境科学研究所夏峰,曾广权提供了有关资料;此外,云南省地理研究所“3S”中心在设备上给予了大力支持,在此一并致谢。

参考文献 (References)

- [1] Li Zhixi, Li Hongga, Lu Feng. Evaluation of Asian elephant's habitat. *Remote Sensing of Environment* [J]. 1996, 11(2): 108—115. (in Chinese) [李芝喜,李红脊,陆峰. 亚洲象生境评价[J]. 环境遥感,1996,11(2):108—115.]
- [2] Pereira, J. M. C., R. M. Itami. GIS-based habitat modeling using logistic multiple regression: A study of the Mt. Graham red squirrel [J]. *PE&RS*, 1991, 57(11): 1475—1486.
- [3] Zhang Hongliang, Zhang Jun, Meng Ming. Naban River sustainable development information system in Xishuangbanna [J]. *Yunnan Geographic Environment Research*, 1998, 10(2): 80—85. [张洪亮,张军,孟鸣. 西双版纳自然保护区纳板河可持续发展信息系统. 云南地理环境研究,1998,10(2):80—05.]
- [4] Ling Bian, Eric West. GIS modeling of ELK calving habitat in a prairie environment with statistics [J]. *PE&RS*, 1997, 63(2): 161—167.
- [5] Johnston. K. M., Using statistics regression analysis to build three prototype GIS wildlife models. proceedings [A]. GIS/LIS'92: annual conference and exposition [C]. San Jose, California, 1992, 374—386.
- [6] Mladenoff. D. J., T. A. Sickley, R. G. Haight, A. P. Wydeven. A regional landscape analysis and prediction of favorable gray wolf habitat in the northern Great lakes region [J]. *Conservation Biology*, 1995, 9(2): 279—294.
- [7] Demaris, A. Logit modeling: Practical application, sage publications [M], Newbury Park, California, 1992.
- [8] Strahler, A. H. The use of prior probabilities in maximum likelihood classification of remotely sensed data [J]. *Remote Sensing of Environment*, 1980, 10: 135—163.
- [9] Maynard, P. F. The logit classifier: A general maximum likelihood discriminant for remote sensing applications [D]. Unpublished M. A. Thesis, Department of Geography, university of Californis, Santa Barbara, California, 1981.

Application of Bayesian Statistics Inference Techniques Based on GIS to the Evaluation of Habitat Probabilities of *Bos Gaurus Readei*

ZHANG Hong-liang¹, LI Zhi-xi¹, WANG Ren-chao², ZHANG Jun³, MENG Ming³

(1. Southwest Forestry College, Kunming 650224, China; 2. Zhejiang University, Hangzhou 310029, China;

3. Yunnan Geography Institute, Kunming 650223, China)

Abstract: At present, GIS has been widely applied to the study of wildlife habitat. However, GIS, which is a tool of spatial data analysis and processing, lacks of the capacity of heuristic reasoning. Therefore, it is an important way to solve this problem by the integration of Bayesian statistics inference with GIS. In this article, the Naban river nature reserve of Xishuangbanna was taken as an experimental area, GIS and multivariate statistical techniques were applied to the development of two logistic multiple regression models for *Bos gaurus readei* habitat: trend surface model and environmental model. Independent variables were locational coordinates in the first model, and a set of environmental factors in the second model. Bayesian statistics were then used to integrate the two models into a Bayesian integrated model. The results show that the Bayesian integrated model is superior to the environmental model and can be applied to wildlife habitat evaluation.

Key words: GIS; bayesian statistics inference; habitat; Xishuangbanna

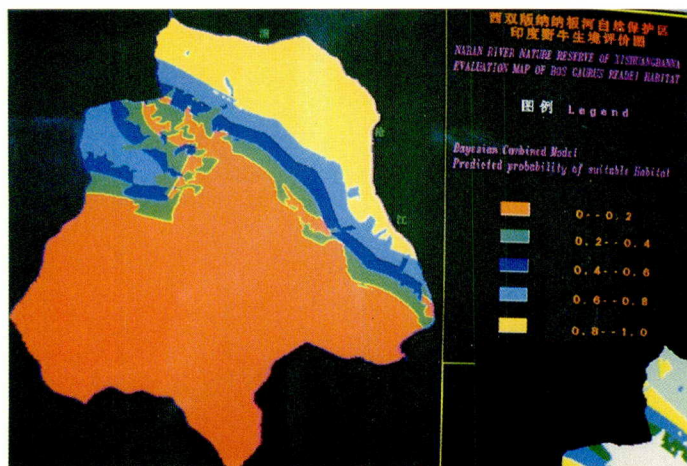


图 1 贝叶斯综合模型产生的印度野牛生境概率图

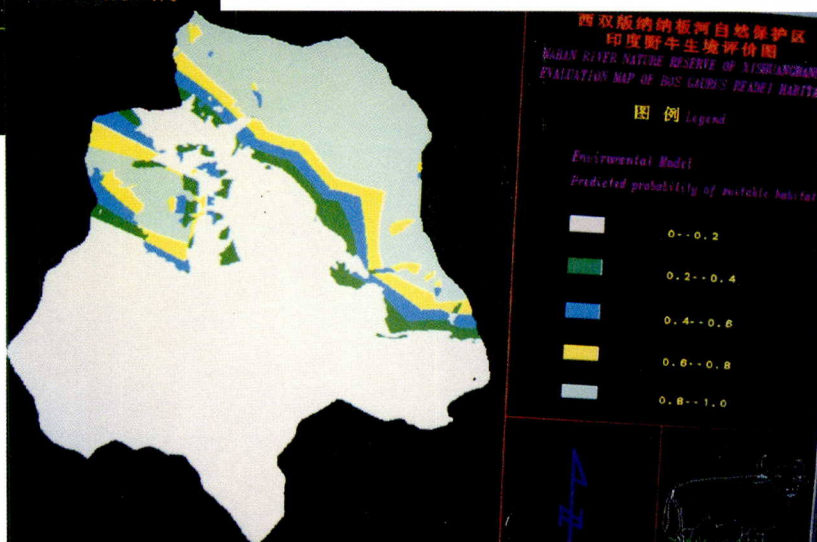


图 2 环境模型产生的印度野牛生境概率图



图 4 BIM 产生的印度野牛生境适宜性评价图



图 5 EM 产生的印度野牛生境适宜性评价图