

区域洪涝灾害损失评估及预测的 CART 模型研究

——以湖南省为例

吉中会^{1a,1b}, 李 宁^{1a,1b}, 吴吉东^{1a,1b}, 刘雪琴²

(1. 北京师范大学 a. 地表过程与资源生态国家重点实验室;
b. 环境演变与自然灾害教育部重点实验室, 北京 100875; 2. 国家海洋局 国家海洋环境监测中心, 辽宁 大连 116023)

摘要: 选择湖南省 14 个地级市(州)1995—2009 年 4—9 月降水量、降水量的距平百分率、最高水位值以及对应年份的年末耕地面积、农作物受灾面积、受灾人口、人口密度、产值密度 8 类指标作为自变量, 损失率为因变量, 构建指标体系, 利用分类与回归树模型逐级提取出了造成洪涝灾害损失的主要因子, 依次为农作物受灾面积、受灾人口、7 月降水量、6 月降水距平百分率、9 月降水量以及人口密度。同时确定了影响阈值, 当农作物受灾面积不超过 $56.8 \times 10^3 \text{ hm}^2$, 且受灾人口少于 158.2 万人时, 预测损失级别为 1, 属于微灾; 而农作物受灾面积超过 $81.1 \times 10^3 \text{ hm}^2$, 7 月降水量少于 343 mm, 6 月降水距平百分率大于 94.5% 时, 预测损失级别为 5, 属极重灾。CART 模型应用简单、结果客观, 为洪涝灾害损失评估及预测开辟了一条新途径。

关键词: 洪涝灾害; CART 模型; 损失评估; 预测; 湖南省
中图分类号: X4 **文献标志码:** A **文章编号:** 1003-2363(2012)06-0106-05

0 引言

20 世纪 90 年代以来, 全球自然灾害发生频率呈明显上升趋势^[1], 其中洪涝灾害是发生最为频繁的灾种之一, 具有复杂性特征^[2], 近年来的发生频率也越来越高^[3], 损失十分严重。目前, 洪涝灾害损失评估方法较多, 主要集中于致灾因子与损失的关系研究, 国际上通常是建立淹没水深、洪水历时等与损失的函数模型^[4], 基于水文物理过程的动力学模型^[5]等; 国内常用方法则有模糊综合评判法^[6]、神经网络法^[7]、极值模型^[8]等, 而这些方法均具有局限性。传统的多元线性回归由于自变量之间存在共线性, 掩盖了某些因子的作用, 从而影响模型的精度^[9]。

洪涝灾害损失评估一般包括人口、经济、社会及生态等方面, 造成损失的因子众多, 且存在区域差异^[10], 如何从这些复杂的因素中客观地提取出主要因子, 并将其影响程度量化, 是做好洪灾直接和间接损失评估的基础。目前此类研究还较少, 分类与回归树(classification and regression tree, 简称 CART)模型恰好能够解决此类问题, 它通过递归方式将多维自变量进行划分, 提取主要影响变量, 允许混合的标度单位, 同时可以解决回归

方法中假设过于严格的问题^[11], 能更加准确地把握自变量和因变量之间的关系, 并对因变量做出预测, 因而具有良好的应用前景。目前国际上已应用于临床医学诊断^[12]、雪灾风险评估^[13]、生物信息学^[14]等领域; 国内在故障诊断^[15]、矿产分类与预测^[16]等领域也有应用。

本研究以湖南省为例, 尝试将 CART 模型应用于区域洪涝灾害损失评估指标体系, 逐级提取主要影响因子, 并根据分类条件对损失级别做出预测, 结果精度较高。因此, CART 模型可以作为洪涝灾害损失评估及预测的工具, 为洪灾综合风险管理与防范服务。

1 研究区概况

湖南省位于长江中游南部, 东、西、南三面环山, 以山地丘陵为主, 北部为洞庭湖平原, 地势低平。全省属于亚热带季风湿润气候, 河网密集, 5 km 以上河流多达 5 341 条, 河川年径流总量达 $2.136 \times 10^{11} \text{ m}^3$ ^[17], 除少部分属于珠江和赣江水系外, 主要隶属于湘、资、沅、澧四水, 并沿地形呈扇状汇入洞庭湖。特有的自然地理环境是湖南省成为农业大省的重要基础, 同时也是易于发生各类自然灾害的环境条件, 在此环境背景下形成的相对脆弱的人地关系模式, 经常受到山洪灾、渍灾及溃灾等灾害的冲击, 造成严重的损失。洞庭湖区作为全国重要的商品粮基地和工业原料供应地, 人为干扰严重, 洪灾发生频繁, 其他山地丘陵地区如湘西、怀化及娄底等地也经常出现严重洪涝灾情。因此, 选取湖南省作为研究区具有代表性, 其研究结果对其他区域的洪涝灾害研究具有重要的借鉴意义。

收稿日期: 2011-09-18; 修回日期: 2012-10-25
基金项目: 高等学校博士学科点专项科研基金项目(20100003110019); 国家重大科学研究计划项目(2012CB955402); 国际合作项目(2010DFB20880)
作者简介: 吉中会(1984-), 女, 江苏盐城市人, 博士研究生, 主要从事自然灾害的风险管理研究, (E-mail) jizhonghui2006@ yahoo. com. cn。

2 研究方法 with 数据来源

2.1 研究方法

2.1.1 CART 模型概述。CART 模型是由 Breiman 等提出的一种非参数非线性的回归方法^[18], 主要包含两个关键思想: 一是递归划分自变量空间, 另一个是用验证数据对数据进行剪枝。CART 模型由根节点、中间节点和终节点组成, 预测值为终节点中因变量的平均值, 构建步骤为:

将所有样本(X) 作为根节点, 通过分割函数选择分割变量, 并确定分割阈值(S), 即将 $X \leq S$ 或 $X > S$ 作为分割规则。假定样本被分入 j 类的先验概率为 p_j , 其中 $\sum_{j=1}^J p_j = 1, (0 \leq p_j \leq 1, j = 1, 2, \cdots, J)$ 。令 $p_{j/t}$ 为节点 t 中第 j 类样本的比例 $(0 \leq p_{j/t} \leq 1, j = 1, 2, \cdots, J)$, J 为样本类的总个数, 通过搜索特征 $p_{j/t}$, 决定每一个 X 的最佳分枝, 然后在最佳 X 的 t 节点处进行分割。在分类过程中, 属于 j 类的样本被误分到 i 类的成本 $C_{(ij)}$ 称为错判成本^[19], $C_{(ij)} \geq 0$, 当 $i = j$ 时, $C_{ij} = 0 (i = 1, 2, \cdots, J)$ 。

对节点进行分割的目的是不断降低子节点的不纯度, 方法主要有熵值法、Gini 指数法以及误分类法^[20], 本研究选择 Gini 指数法进行划分, 其函数形式为:

$$\psi(t) = \sum_{j=1}^J \sum_{i=1}^{j-1} p_{(j/t)} p_{(i/t)} [C_{(ji)} + C_{(ij)}] \quad (1)$$

式中: $\psi(t)$ 为从根节点到子节点的不纯度总和^[21]。划分直到没有满足分割条件或是节点样本数量很少(通常为 5 个) 为止。样本划分结束之后, 可能会出现对样本数据过度分割的情况, 因此, 需要进行剪枝, 具体步骤如下。

假设树 T 是由 T_i 个终节点组成, 那么定义树 T 的 Gini 不纯度 $D(T)$ 为:

$$D(T) = \sum_{t \in T_i} d(s, t) \quad (2)$$

式中: s 为节点 t 的分割阈值。虽然通过选择最佳分支 X 和最佳分支阈值 S 构建了一个不纯度最小的树, 但不一定是最优的树。

根据树的误分类比例将树的拟合优度作为分类工具, 当误分类成本不均匀时, 可以定义期望误分类成本 $R(t)$ 为:

$$R(t) = \sum_{j=1}^J \sum_{i=1, i \neq j}^J C_{(ij)} Q_{(ij)} P_{(j)} \quad (3)$$

式中: $Q_{(ij)}$ 为 j 类被误分为 i 类的比例。

CART 模型中错判成本是通过交叉验证法来求算的, 首先将所有样本分成 K 个大小相同的子样本, 每次从 $K - 1$ 个子样本的不同组中产生 CART 树, 剩余样本用于检验, 这样的树就可以产生 K 次, 所得的 K 个误分类比例平均后得到交叉验证误分类比例 $Q_{(ij)}^{CV}$, 然后通过公式 (3) 求得总交叉验证误分成本 $R^{CV}(T)$, 由于 $R^{CV}(T)$ 是交叉验证的子样本随机产生的, 因此在修剪树时, 考虑其分布的标准差(SE), 当最小误分类成本在 1 个标准差之内时停止修剪, 此时获得最优树^[19]。

2.1.2 指标体系的构建。从灾害系统论角度考虑了致灾因子、孕灾环境、承灾体社会脆弱性以及灾情 4 个方面, 构建了包括月降水量、降水量的距平百分率、最高水位、耕地面积、洪涝灾害导致的农作物受灾面积和受灾人口、人口密度、产值密度 8 类指标为自变量、损失率为因变量的指标体系(图 1)。湖南省洪涝灾害主要发生在 4—9 月的汛期, 因此降水量、降水量的距平百分率及最高水位指标选取 4—9 月观测值, 其他指标为年值。CART 模型中所用的数据由湖南省 14 个地级市(州) 1995—2009 年 15 年的 23 个自变量($x_1 \sim x_{23}$) 与因变量(y) 共 210 个样本组成。



说明: 产值密度 = 人均 GDP/人口密度; 损失率 = 洪涝灾害直接经济损失/GDP

图 1 洪涝灾害损失评估指标体系
Fig. 1 Indicator system of flood loss evaluation

洪涝灾害损失级别划分方式较多, 各区域差异较大, 因而无统一标准。从湖南省 1995—2009 年 210 个样本的损失率分布曲线(图 2) 看出, 损失率(L) > 0.30 的样本数逐渐逼近于 0, 说明发生极重灾的频次较低, 无需再进行细分, 因此可以将 $L > 0.30$ 作为等级划分的上限, 为 5 级; L 值主要集中在 $0 \sim 0.02$ 之间, 说明属于轻灾等级的样本也是集中分布的, 可以将其单独作为一级, 为 1 级; 其余在 $0.02 \sim 0.3$ 之间再进行细分, 最终将 L 划分为 5 个级别, 分别为 $0 \leq L \leq 0.02$ (134), $0.02 < L \leq 0.10$ (50), $0.10 < L \leq 0.20$ (12), $0.20 < L \leq 0.30$ (7), $L > 0.30$ (7), 括号内数值代表对应范围的样本数量。

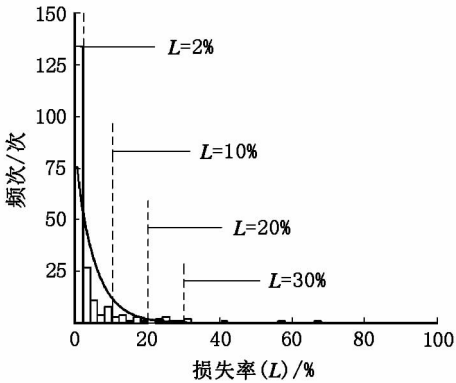


图 2 洪涝灾害损失率直方图
Fig. 2 Histogram of flood loss rate

2.2 数据来源

各月降水数据来源于中国气象数据共享网;各地市(州)水文站最高水位数据、因洪涝灾害造成的农作物受灾面积、受灾人口及直接经济损失数据来源于湖南省防汛抗旱指挥部的《湖南省防汛工作总结》(1995—2009);耕地面积、人口密度以及国民生产总值数据来源于《湖南省统计年鉴》(1996—2010)。

3 结果分析

3.1 洪涝灾害损失评估

为了便于模型构建和保证模型精度,将所有指标作

均值化处理后,进行 CART 建模,模型的综合解释率为 81.33%,所得分类条件的阈值还原为原始量纲,得到的分类与回归树如图 3 所示。从图中可以看出,CART 深度为 5,共 15 个节点,其中终节点 8 个(虚线框),分别为节点 4,6,7,10,11,12,13,14,各节点框中包含分类后样本因变量的平均值、标准方差、样本数量(N)及所占比重(%),终节点中的平均值为该类洪涝灾害损失的预测值,实线框为中间节点(节点 0 为根节点),表示样本需要做进一步的划分。各节点之间的变量为模型所提取出的主要影响因子,相应的分割值(\leq 或 $>$)为这些影响因子的分类阈值,两者综合,即为分类条件。

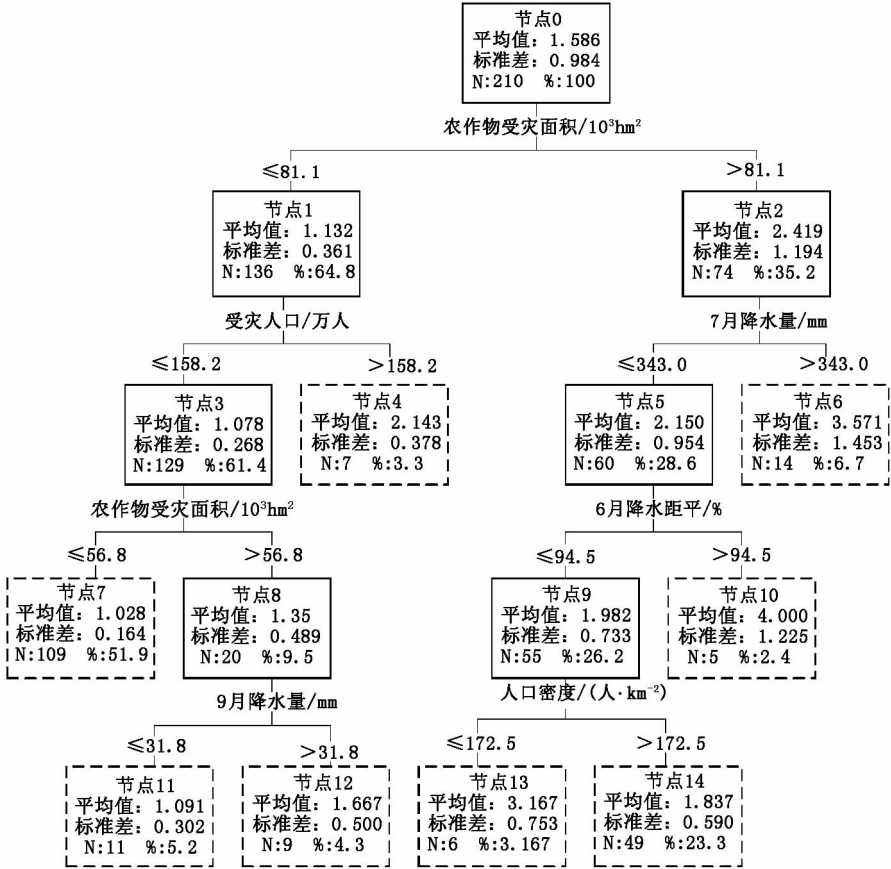


图 3 湖南省洪涝灾害损失评估及预测的 CART 模型

Fig.3 CART model of flood loss evaluation and prediction in Hunan Province

农作物受灾面积因子在模型中两次出现,说明农作物受灾面积是湖南省洪灾损失评估的最主要因子。节点 0 以农作物受灾面积为分类条件,其中不超过 $81.1 \times 10^3 \text{ hm}^2$ 的样本 (64.8%) 被分到节点 1 中,剩余样本 (35.2%) 被分到节点 2 中。同时,节点 3 也是以农作物受灾面积为分类条件的,不超过 $56.8 \times 10^3 \text{ hm}^2$ 的样本 (51.9%) 被分到终节点 7 中,剩余样本 (9.5%) 被分到节点 8 中。农作物受灾面积能够较好地代表农业损失情况,不仅体现了湖南农业大省的地位,同时农业作为农村地区的主要承灾体,其损失在洪涝灾害总损失中所占的比重很高。图 4 为湖南省 1995—2009 年洪涝灾害导致的农业经济损失占总的直接经济损失的比重。可

以看出,比值在 30% ~60% 之间波动,且呈现出一定的上升趋势。因此,在平时的防灾减灾工作中应当重点做好农业的防涝排涝工程,合理规划农业发展布局,在低洼易涝地区选择适宜的作物类型,最大限度地降低洪涝灾害损失。

节点 1 以受灾人口数量为分类条件,大于 158.2 万人的样本 (3.3%) 被分到终节点 4 中,其余样本 (61.4%) 被分到节点 3 中。说明受灾人口也是洪灾损失评估的重要因子,有研究提出地震灾害的经济损失可以直接利用受灾人口进行快速评估^[22],本研究在此也证实了受灾人口是湖南省洪涝灾害损失评估的主要指标。人员伤亡后,不仅需要增加救助投入,还会影响灾后恢

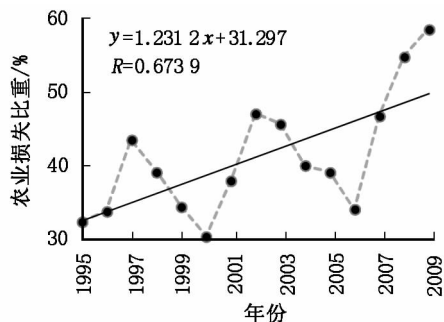


图 4 湖南省农业损失占洪涝灾害总经济损失比重
Fig.4 Percentage of agriculture loss in Hunan Province

复进度,甚至造成更大的间接损失。

节点 2 以 7 月份降水量为分类条件,超过 343 mm 的样本(6.7%)被分配到终节点 6 中,其余样本被分入节点 5 中。说明湖南省 7 月份的降水总量对洪灾的损失有重要影响,降水总值超过一定阈值(>343 mm)就可能产生相对较大的损失。一般情况下,7 月是暴雨频发的月份,而暴雨是洪涝灾害的主要诱发因子,因此,在洪灾损失评估和风险管理中,7 月降水量也是不可忽略的因子。

节点 5 和节点 8 分别以 6 月降水距平和 9 月降水量为分类指标,6 月和 9 月作为长江中下游江南地区的夏季初和秋季初,说明在季节交替时期降水的变化也是洪水导致损失的重要因素。例如,终节点 10 在其他条件相同的情况下(农作物受灾面积 > $81.1 \times 10^3 \text{ hm}^2$, 7 月份降水量 $\leq 343 \text{ mm}$),6 月份降水距平大于 94.5%,洪灾损失达到了最严重的等级,同时也可以看出降水条件和损失之间并非简单的线性正负向关系。同理,终节点 11 和 12 以 9 月降水量为分类条件,当降水量超过一定阈值(31.8 mm)时,预测的损失等级相对较高。

节点 9 以人口密度为分类指标,当区域人口密度不超过 172.5 人/km^2 时,损失相对较大,说明湖南省人口稀疏的地区,洪灾损失的可能更大。湖南省的山地丘陵地区,例如湘、资、沅、澧四水中上游地区,以农村人口为主,经济水平相对落后,脆弱性较高,容易造成严重的洪灾损失。

通过以上对 CART 模型的分析发现,影响湖南省洪涝灾害损失的主要因子依次为农作物受灾面积、受灾人口、7 月降水量、6 月降水距平百分率、9 月降水量以及人口密度。因此,在进行湖南洪涝灾害损失评估时可以主要考虑这些因子。其中农作物受灾面积因子在模型中出现 2 次,充分说明这项指标是洪灾损失评估时要首要考虑的因素。降水指标在模型中有 3 个,它们是形成洪涝的必要条件,是洪涝灾害的主要致灾因子,尤其是夏季 7 月,天气状况极不稳定,发生强降雨的机率很高,应当做好降水的监测和及时预报。作为社会脆弱性和抗灾救灾能力指标的人口密度在一定程度上也影响洪灾损失评估的结果,人口密度较小的区域,社会脆弱性高,

抗灾自救能力较差,因而造成的洪灾损失较大。

3.2 区域洪涝灾害的损失预测

CART 模型逐级提取了洪灾损失的主要影响因子,并确定了影响阈值,基于此可以对分类后样本的损失级别进行预测。表 1 为分类后的预测值及样本分布,纵列总计表示原始观测样本损失级别的分布;横行表示终节点对应损失级别的样本数量,例如终节点 7 和 11,预测级别为 1 的样本有 116 个,级别 2 有 4 个,级别 3,4,5 的样本数为 0,综合预测值为 1.06,损失程度为 1 级,属于微灾,依次可以类推其他终节点中的样本分布。每个终节点代表一个损失级别,预测损失级别 1~5 的样本数量分别为 120,65,6,14,5,对应原始观测样本数量依次为 134,50,12,7,7,通过 10 倍的交叉验证,其标准误差仅为 0.056,表明分类及预测的精度较高。

| 表 1 分类后的预测值及样本分布 | | | | | | | | | |
|--|-------|-----------|----|---|---|---|-----|------|------|
| Tab.1 Distributions and predicted values of samples after classification | | | | | | | | | |
| 终节点 | 评估值 | 分类等级的样本分布 | | | | | 总计 | 预测级别 | 严重程度 |
| | | 1 | 2 | 3 | 4 | 5 | | | |
| 7/11 | 1.060 | 116 | 4 | 0 | 0 | 0 | 120 | 1 | 微灾 |
| 4/12/14 | 1.882 | 16 | 43 | 6 | 0 | 0 | 65 | 2 | 轻灾 |
| 13 | 3.167 | 0 | 1 | 3 | 2 | 0 | 6 | 3 | 中灾 |
| 6 | 3.571 | 2 | 1 | 3 | 3 | 5 | 14 | 4 | 重灾 |
| 10 | 4.000 | 0 | 1 | 0 | 2 | 2 | 5 | 5 | 极重灾 |

根据图 4 和表 1 可以得出,当湖南省农业受灾面积不超过 $56.8 \times 10^3 \text{ hm}^2$ 、受灾人口不超过 158.2 万人时,预测的洪灾损失为微灾级别;而当农作物受灾面积超过 $81.1 \times 10^3 \text{ hm}^2$ 、7 月降水量少于 343 mm、6 月降水距平大于 94% 时,损失为极重灾级别。其他分类条件所预测的灾情等级介于这两者之间。因此,可以将上述条件作为湖南省快速预测洪涝灾害损失级别的依据。

4 结论与讨论

采用 CART 模型提取了湖南省洪涝灾害损失的主要影响因子,依次为农作物受灾面积、受灾人口、7 月降水量、6 月降水距平百分率、9 月降水量以及人口密度。根据这些因子满足的条件(表 2),可以快速地对区域洪涝灾害损失严重程度做出判断。

目前,已有的洪涝灾害损失评估与预测方法主要是对整体的损失状况做出评判,还没有将主要影响因子逐级提取,及确定各因子影响阈值的方法,部分影响因子(如降水条件)与洪灾损失的关系十分复杂,很难用线性的方式来表达,由此可见,CART 模型为洪涝灾害损失评估及预测提供了一个新方法。

今后研究中,还需在两方面进一步深入。首先,CART 模型提取出了 6 个主要影响因子,并不说明其他因子对洪涝灾害损失没有影响,只是影响程度相对较

表 2 湖南省洪涝灾害损失评估及预测标准

Tab.2 Criteria of flood loss estimation and prediction in Hunan Province

| 损失 级别 | 严重 程度 | 判断 条件 | 损失划分标准 | | | | | | |
|----------|----------|----------|---|-------------|---------------|---|------------------|---------------|--------------------------------|
| | | | 农作物受灾面积 /10 ³ hm ² | 受灾人口 /万人 | 7 月降水量 /mm | 农作物受灾面 积/10 ³ hm ² | 6 月降水距 平百分率/% | 9 月降水 量/mm | 人口密度/ (人·hm ⁻²) |
| 1 | 微灾 | 1 | | ≤158.2 | | ≤56.8 | | | |
| | | 2 | ≤81.1 | ≤158.2 | | >56.8 | | ≤31.8 | |
| | | 3 | ≤81.1 | ≤158.2 | | >56.8 | | >31.8 | |
| 2 | 轻灾 | 4 | >81.1 | | ≤343.0 | | ≤94.5 | | >172.5 |
| | | 5 | ≤81.1 | >158.2 | | | | | |
| 3 | 中灾 | 6 | >81.1 | | ≤343.0 | | ≤94.5 | | ≤172.5 |
| 4 | 重灾 | 7 | >81.1 | | >343.0 | | | | |
| 5 | 极重灾 | 8 | >81.1 | | ≤343.0 | | >94.5 | | |

小,已知这 6 个因子时就可以进行损失评估。从众因子中提取出的最重要的农作物受灾面积、受灾人口及 7 月降水量这 3 个因子也是有关部门进行洪灾损失评估时考虑的主要方面,通过本研究也证实了其科学性所在。但是这里 CART 模型所提取的主要因子仅限于湖南省,其他区域对此模型的应用,应结合当地实际情况选择合适的影响因子,不同区域的主导因子亦会有所不同,进而分析结论可能会出现较大差异。今后应当加强对不同区域的研究,并在此基础上,归纳出更为普遍的规律。其次,通过研究可知,CART 模型更适用于非线性的数据结构,洪涝灾害损失影响因子之间的线性关系较弱,因而比较适用,这也是该模型在洪灾损失评估中的优势,但是模型的构建需要大量的样本,否则结果的精度就会受到影响,而灾害数据的获取比较困难,尤其是长时间序列的数据,如果能够解决模型对大样本数据的需求,就可以在灾害损失评估领域中作进一步推广。

参考文献:

[1] 史培军,李宁,叶谦,等. 全球环境变化与综合灾害风险防范研究[J]. 地理科学进展,2009,24(4):428-435.

[2] 魏一鸣,金菊良,周成虎,等. 洪水灾害评估体系研究[J]. 灾害学,1997,12(3):1-5.

[3] 曹诗图,黄昌富. 98 洪灾的反思与对策[J]. 地域研究与开发,1999,18(1):87-89.

[4] Vrisou N,Kok M. Standard Method for Predicting Damage and Casualties as a Result of Floods[M]. Delft, Netherlands: Ministry of Transport, Public Works and Water Management,2001:22-41.

[5] Dutta D,Herath S,Musiak K. A Mathematical Model for Flood Loss Estimation[J]. Journal of Hydrology, 2003, 277(1-2): 24-49.

[6] 吴红华,李正农. 灾害损失评估的区间数模糊综合评判方法[J]. 自然灾害学报,2006,15(6):149-153.

[7] 刘小生,余豪峰. 基于 GIS 和 BP 神经网络的洪灾损失评估模型的研究[J]. 工程勘察,2009(4):72-74.

[8] 刘家福,吴锦,蒋卫国,等. 基于泊松-对数正态复合极值模型的洪水灾害损失分析[J]. 自然灾害学报,2010,

19(6):61-66.

[9] 刘洋,吕一河,郑海峰,等. 用回归树模型分析陕北黄土丘陵沟壑区气候因子对 NDVI 变异的影响[J]. 应用生态学报,2010,21(5):1153-1158.

[10] 商彦蕊. 自然灾害综合研究的新进展——脆弱性研究[J]. 地域研究与开发,2000,19(2):73-77.

[11] 莫春梅,倪宗瓚,高凤琼. 回归树的建模与应用[J]. 中华预防医学杂志,2006,36(5):346-347.

[12] Rovlias A, Kotsou S. Classification and Regression Tree for Prediction of Outcome after Severe Head Injury Using Simple Clinical and Laboratory Variables[J]. Journal of Neurotrauma,2004, 21(7):886-894.

[13] Tachiiri K,Shinoda M,Klinkenberg B,et al. Assessing Mongolian Snow Disaster Risk Using Livestock and Satellite Data[J]. Journal of Arid Environments,2008,72(12):2251-2263.

[14] Che D S,Liu Q, Rasheed K,et al. Decision Tree and Ensemble Learning Algorithms with Their Applications in Bioinformatics[J]. Advances in Experimental Medicine and Biology,2011,696:191-199.

[15] 刘华,王建国. 分类回归树在故障诊断中的应用研究[J]. 设备管理与维修,2005(12):12-14.

[16] 张松林. CART-分类与回归树方法介绍[J]. 火山地质与矿产,1997,18(1):67-75.

[17] 湖南省统计局. 湖南统计年鉴 2009[Z]. 北京:中国统计出版社,2010.

[18] Breiman L,Friedman J,Olshen R,et al. Classification and Regression Trees[M]. Belmont,CA: Wadsworth, 1984:1-357.

[19] David F,Mortgage D. Classification Trees Analysis[J]. Journal of Real Estate Finance and Economics,2005,30(4):369-396.

[20] 什穆列维奇. 基因组信号处理[M]. 北京:科学出版社,2010:139-144.

[21] 温晓霓,蔡汝俊. 分类与回归树及其应用研究[J]. 理论新探,2007(23):14-16.

[22] 周光全,毛燕,施伟华. 云南地区地震受灾人口与经济

Empirical Study on Land Intensive Use and Influencing Factors of Resource-based Cities

Wang Zhongya¹, Chen Weidong², Zhang Guangping³

(1. *Institute of Industrial Economics, Henan Academy of Social Sciences, Zhengzhou*

450002, China; 2. Department of Management and Economics, Tianjin University, Tianjin

300072, China; 3. Bureau of Land and Resources of Suiyang District, Shangqiu City, Shangqiu 476100, China)

Abstract: Through the construction of the intensive use evaluation indicator system for the resource-based cities, the land intensive use levels of 34 classical resource-based cities in China are calculated based on entropy weight method and an empirical study is carried out on the influencing factors of land intensive use. There is significant difference among the different types of resource-based cities. Land intensive use is positive correlation with the economic development level, the city scale, and the industry structure. Finally, the suggestion is put forward that the resource-based cities should take the countermeasure to realize the radical transformation of land utilization way as soon as possible.

Key words: resource-based cities; land intensive use; entropy weight method; influencing factors

(上接第 110 页)

Flood Loss Assessment and Prediction Model Based on CART in Hunan Province

Ji Zhonghui^{1a,1b}, Li Ning^{1a,1b}, Wu Jidong^{1a,1b}, Liu Xueqin²

(1a. State Key Laboratory of Earth Surface Processes

and Resources Ecology; 1b. Key Laboratory of Environmental Change and

Natural Disaster, Ministry of Education of China, Beijing Normal University, Beijing 100875, China;

2. *National Marine Environmental Monitoring Center, State Oceanic Administration, Dalian 116023, China*)

Abstract: Classification and Regression Tree (CART) is a nonparametric and non-linear model that simple and flexible. Taking Hunan Province as the study area, several indicators consisting of precipitation, precipitation anomaly percentage, highest water level, agriculture acreage, crop loss area, flood-hit population, population density, and GDP density were chosen as independent variables, the economic loss rate as the dependent variable correspondingly. Based on the flood loss estimated indicator system, the CART model gradually extracted six major impact factors, including crop loss area, flood-hit population, precipitation of July, precipitation anomaly percentage of June, precipitation of September and population density. If crop loss area less than $56.8 \times 10^3 \text{ hm}^2$ and flood-hit population less than 158.2×10^4 , it can be predicted that the loss is least serious, if crop loss area more than $81.1 \times 10^3 \text{ hm}^2$, precipitation of July less than 343 mm, and precipitation anomaly percentage of June more than 94.5%, the loss can be the most serious. Therefore, CART model is easy for application and the result is objective. It is a new way for flood loss evaluation and prediction.

Key words: flood; CART model; loss evaluation; prediction; Hunan Province