

基于 Ant-Miner 的洪灾风险区划模型及应用*

赖成光^{1,2}, 王兆礼³, 陈晓宏^{1,2}, 黄锐贞^{1,2}, 廖威林³, 吴旭树³

(1. 中山大学水资源与环境研究中心, 广东 广州 510275;

2. 华南地区水循环与水安全广东普通高校重点实验室, 广东 广州 510275;

3. 华南理工大学土木与交通学院, 广东 广州 510641)

摘要: 应用蚁群优化算法 (Ant Colony Optimization, ACO) 进行规则挖掘是一个新的研究热点。为解决指标变量与风险级别间非线性关系, 提出一种基于蚁群规则挖掘算法 (Ant-Miner) 的洪灾风险区划模型。在 GIS 技术支持下, 将该模型应用于北江流域洪灾风险区划实例中, 结果表明: ① Ant-Miner 模型可挖掘 15 条适合研究区的洪灾风险分类规则, 这些规则以简单的条件语句形式表现, 便于生成风险区划图; ② Ant-Miner 模型测试精度 (95.1%) 高于相同条件下 BP 神经网络模型的精度 (92.9%), 表明其分类性能更好, 对洪灾风险区划具有更好的适用性; ③ 研究区高风险区主要集中于降雨量较大、地势平缓低洼、人口财产密集的地区, 与历史洪灾风险情况较吻合, 表明所构建的模型科学合理, 可为流域洪灾风险评价提供了新思路。

关键词: 洪灾; 风险区划; 蚁群优化算法; 规则挖掘; 北江流域

中图分类号: TV877 **文献标志码:** A **文章编号:** 0529-6579 (2015) 05-0122-08

Flood Risk Zoning Model Based on Ant-Miner and Its Application

LAI Chengguang^{1,2}, WANG Zhaoli³, CHEN Xiaohong^{1,2}, HUANG Ruizhen^{1,2}, LIAO Weilin³, WU Xushu³

(1. Center for Water Resources and Environment, Sun Yat-sen University, Guangzhou 510275, China;

2. Key Laboratory of Water Cycle and Water Security in Southern China of Guangdong High Education Institute, Sun Yat-sen University, Guangzhou 510275, China;

3. School of Civil and Transportation Engineering, South China University of Technology, Guangzhou 510641, China)

Abstract: Using Ant Colony Optimization (ACO) to mine rules is a research hotspot nowadays. This paper proposed a new zoning model of flood risk based on ant colony rule mining algorithm (Ant-Miner) to solve the non-linear relationship between index and flood risk grade. The model was used in the Beijiang River basin with the support of GIS technique. The assessment results show that ① 15 simple rules expressed in the form of conditional statement were mined by the Ant-Miner model. The rules are appropriate for the study areas and can be easily used for generating a zoning map of flood disaster risk. ② The test accuracy is 95.1% in the Ant-Miner model, 92.9% in BP neural network model, indicating that the discriminative capability and flood risk zoning applicability of the former is stronger than the latter. ③ The high risk areas identified by Ant-Miner are mainly located in the regions with large precipitation, flat and low-lying terrain and dense population and property. These areas match well with the submerged are-

* 收稿日期: 2014-12-17

基金项目: 国家自然科学基金资助项目 (51210013, 51479216, 51209095, 41301627); 国家科技支撑计划资助项目 (2012BAC21B0103); 水利部公益资助项目 (201301002-02, 201301071); 中央高校基本科研业务费专项基金资助项目 (2014ZZ0027)

作者简介: 赖成光 (1986 年生), 男, 研究方向: 灾害风险研究; **通讯作者:** 陈晓宏; E-mail: eescxh@mail.sysu.edu.cn

as of historical flood disasters, indicating that the Ant-Miner model is reasonable and practicable and can provide a new method for flood risk assessment.

Key words: Flood disaster; risk zoning; ant colony optimization; rule mining; the Beijiang River basin

洪水灾害是一种突发性强、发生频率高、危害严重的自然灾害^[1-2]。洪灾风险评价是确定洪灾风险相对大小的重要手段，是一种定性、半量化且考虑自然属性和社会属性的综合性评价。洪灾风险区划是在洪水风险评价基础上把研究区划分为不同风险级别的区域，其目的是为了更准确地掌握洪灾风险的空间分布格局及内在规律。开展洪灾风险区划工作对洪灾风险管理、灾害预警预报、洪水保险评估等工作具有重要指导意义。由于洪灾风险区划受自然和社会诸多因素影响，评价过程涉及多个评价指标，因而一直是国内外灾害学研究的热点和难点之一。构建风险区划模型常用的方法有层次分析法^[3]、模糊综合评价^[4]、集对分析法^[5]等。

随着洪灾风险区划方法多样化发展，一些智能算法也逐渐被应用，如决策树^[6-7]、贝叶斯网络^[8]、随机森林模型^[9]、人工神经网络^[10]等。Colormi 和 Dorigo 等^[11]于 1991 年提出了一种新的模拟蚂蚁群体智能行为的蚁群优化算法 (Ant Colony Optimization, ACO)，该算法具有较强的鲁棒性、自适应性、正反馈和优良的分布式计算机制、易于与其他算法结合的优点。基于蚁群算法的规则挖掘 (Ant-Miner) 最初由 Parpinelli^[12]于 2002 年提出，目的是从训练数据中挖掘分类规则，然后应用分类规则对待测数据进行分类，已被成功应用于蛋白质层次分类^[13]、声频信号分类^[14]、遥感影像分类^[15]、元胞机转移规则发现^[16]等领域。上述研究均表明 Ant-Miner 能有效解决非线性分类问题，且具有较高的精度与效率。

洪灾风险区划需综合考虑研究区指标变量与最终风险级别相关关系，显然，风险区划过程实质上也是一个解决多变量分类问题的过程。各指标变量与风险级别之间的关系往往是复杂的非线性关系，而 Ant-Miner 可以有效地解决此类问题，且能保持较高的精度与效率。因此基于 Ant-Miner 的分类方法理论上可被应用到洪灾风险区划领域然而却鲜见相关报道。鉴于此，本文将以北江流域为例构建洪灾风险区划指标体系，借助 GIS 技术建立基于 Ant-Miner 的洪灾风险区划模型，以探索洪灾风险评价新途径，以为流域防洪减灾和灾害评估工作提供参考依据。

1 基于 Ant-Miner 洪灾风险区划模型

洪灾风险区划指根据研究区洪水危险性特征，参考区域承灾能力及社会经济状况，把研究区划分为不同风险等级的区域，区划过程受自然和社会众多因素影响，而这些因素往往通过风险指标变量体现。构建合适的洪灾风险区划模型可以方便分析风险空间分布特性。模型的核心任务是根据指标变量对风险进行综合分类，即对研究区内的风险水平进行分类并区划出不同级别。

ACO 算法主要通过模拟自然界中蚂蚁集体寻径行为而提出的一种基于群体智能的启发式仿生进化算法^[17]。而基于蚁群算法的规则挖掘 (Ant-Miner) 的目标则是从训练数据中挖掘分类规则，每个蚂蚁的任务是不断寻找分类规则，最终整个蚁群能够挖掘出一个非常满意的规则库^[18]。同理，基于 Ant-Miner 的洪灾风险区划模型的基本思想是利用该算法卓越的分类性能对洪灾风险水平进行分级，主要原理是利用蚁群觅食原理在数据库中搜索最优规则，通过模仿蚂蚁寻找食物的方式来构造洪灾风险级别识别的规则，并利用这些规则对待测数据进行分级。

1.1 Ant-Miner 原理

定义路径为指标节点和风险级别节点的连线，其中每个指标节点最多只出现一次且必须有风险级别节点^[17]。图 1 中给出了两条可能的路径，每条路径对应着一条分类规则，分类规则的挖掘可以看成对路径的搜索。然而搜索的并非最短路径，而是最优路径，此最优路径表示最优的分类规则。可以用路径对应规则的分类能力 (有效性) 和长短 (简洁性) 来衡量路径的优劣。

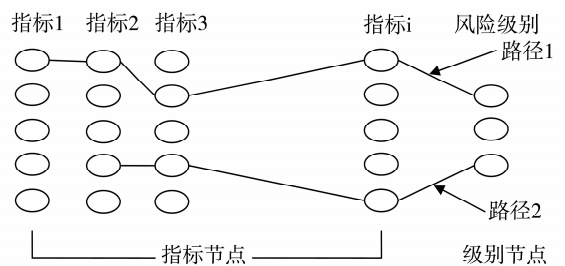


图 1 分类规则对应的路径

Fig. 1 Classification rules corresponding to routes

蚂蚁构造规则的过程体现为构造一条路径, 可分为 3 个阶段^[17]: 首先从一条空路径开始重复选择路径节点增加到路径上 (模仿蚂蚁的爬行过程), 直到得到一条完整路径, 也即一条分类规则; 其次进行规则的剪枝, 以解决分类规则过度拟合问题; 最后更新所有路径上的外激素浓度, 对下一只蚂蚁构造规则施加影响 (模拟蚂蚁间的信息交流)。三个阶段间的具体协作如图 2 所示。

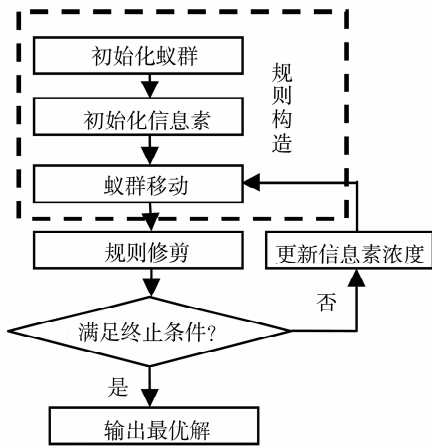


图 2 ACO 规则构建流程图

Fig. 2 Flow chart of ACO rule construction

1) 规则构造

定义蚁群搜索路径为指标节点和等级节点的连线, 即每只蚂蚁走过的一条路径对应于一条分类规则。如图 1 所示, 风险等级可以认为是食物源, 各离散指标值到风险等级之间构成了一段路径, 每条路径对应一条分类规则, 分类规则的挖掘可以当作是对最优路径的搜索。刚开始时, 随机生成一条规则, 规则的形式为: IF < term₁ AND term₂ AND... > THEN < grade >, 其中 term_i 为条件项, 条件组合用 < 指标名称, 操作符号 (即 “> =” 和 “< =” 两种符号), 指标值 > 表示, grade 为满足该规则的洪灾风险等级。需要注意的是指标样本是连续值, 需对数据进行离散化处理^[19]。

规则构造具体包括初始化蚁群、初始化信息素和蚁群移动 3 部分内容:

初始化蚁群: m 只蚂蚁随机分布在第一个评价指标的某节点上。

初始化信息素: 所有路径节点的信息素浓度被初始化为相同的值:

$$\tau_{ij}(0) = \frac{1}{\sum_{i=1}^a b_i} \quad (1)$$

其中, τ_{ij} 为条件项 term _{j} 的信息素浓度, a 为数据库中指标 (不包括等级数) 总数, b_i 为指标 i 所有可能取值的数据, j 为风险等级数。

蚁群移动: 蚂蚁根据以下公式选择下一个节点。

$$P_{ij}(t) = \frac{\tau_{ij}(t) \cdot \eta_{ij}(t)}{\sum_{i=1}^a \sum_{j=1}^{b_i} \tau_{ij}(t) \cdot \eta_{ij}(t)} \quad (2)$$

上式中采用赌轮机制用来选择指标节点^[12], 对每个指标来说, 其所属节点 term _{j} 被选择的概率为 $P_{ij}(t)$ 。其中, $\tau_{ij}(t)$ 为条件项 term _{j} 的信息素浓度, η_{ij} 为条件项 term _{j} 的启发式函数值。

$$\eta_{ij} = \frac{\max(\sum_n \text{freq}T_{ij}^1, \sum_n \text{freq}T_{ij}^2, \dots, \sum_n \text{freq}T_{ij}^k)}{\sum_n T_{ij}} \quad (3)$$

其中, T_{ij} 为满足条件项 term _{j} 的样本数, $\text{freq}T_{ij}^k$ 为 T_{ij} 中风险等级为 k 的频数。在规则挖掘过程中, 每得到一条最终规则后, 要移除符合规则条件部分的记录, 因而, $\max(\sum_n \text{freq}T_{ij}^1, \sum_n \text{freq}T_{ij}^2, \dots, \sum_n \text{freq}T_{ij}^k)$ 和 $\sum_n T_{ij}$ 在得到一条最终规则后其值会发生变化, 需要对其进行动态更新。

2) 规则修剪

根据指标节点的选择标准, 在每个指标中选择一个节点后得到一条最原始的规则, 规则的有效性通过公式 (4) 进行计算。

$$Q = \frac{\text{TruePos}}{\text{TruePos} + \text{FalseNeg}} \times \frac{\text{TrueNeg}}{\text{FalsePos} + \text{TrueNeg}} \quad (4)$$

对于等级 c_i 的分类结果, TruePos 表示实际属于该等级且分类模型认为属于该等级的样本数; TrueNeg 表示实际属于该等级但分类模型认为不属于该等级的样本数; FalsePos 表示实际不属于该等级但分类模型认为属于该等级的样本数; FalseNeg 表示实际不属于该等级且分类模型认为不属于该等级的样本数。

修剪方法是依次移去能使规则有效性得到最大提高的指标节点, 即移除多余的指标节点, 直到任一指标节点的移除都会降低规则的有效性。若达到结束条件 (足够好的规则或最大迭代次数) 则结束, 否则再次返回蚁群移动的步骤。当若干蚂蚁连续搜索到同一路径时, 则认为搜索收敛, 该路径进行规则修剪后成为一条最终规则; 或当迭代次数达到指定的次数时, 在迭代过程搜索到的规则中选择质量最好的规则作为最终规则。由于在迭代过程中, 质量较好的规则, 由于其信息素浓度逐渐增

强，从而能够得以保留，并被视为最终分类规则，其它质量较差的规则被丢弃。

3) 信息素更新

人工蚂蚁在指标节点选择过程中的正反馈机制是通过改变指标节点上的信息素浓度来实现的。当一次迭代中的人工蚂蚁构造的规则经过修剪得到分类规则后，所有路径节点的信息素浓度都将依据这种分类规则的效率进行更新，指标节点的信息素浓度更新公式如下：

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + \Delta\tau_{ij} \quad (5)$$

$$\Delta\tau_{ij} = \sum_{k=1}^m \tau_{ij}^k \quad (6)$$

$$\Delta\tau_{ij}^k(t) = \begin{cases} \frac{Q_k}{1 + Q_k}, & \text{若第 } k \text{ 只人工蚂蚁在该次迭代} \\ & \text{中经过节点 } term_{ij} \\ 0, & \text{否则} \end{cases} \quad (7)$$

其中 ρ 为信息素浓度的衰减系数，通常设 $\rho < 1$ ，以避免路径上信息素的无限累加。 Q_k 为本次迭代过程中第 k 只人工蚂蚁找到的分类规则的质量， $\Delta\tau_{ij}^k(t)$ 为第 k 只人工蚂蚁在本次迭代中留在节点 $term_{ij}$ 上的信息量，所有指标节点的信息素尝试被更新。

1.2 基于 Ant-Miner 的洪灾风险评价流程

本研究提出基于 Ant-Miner 的洪灾风险区划模型，基本思路是从风险的训练数据中挖掘出风险分类规则，并对这些规则进行精度测试，再应用这些规则对研究区进行分类和区划以获取洪灾风险区划图，评价流程如图 3 所示。模型总体上可以划分为训练、测试和分类 3 部分，其中训练过程是规则构建的核心过程。训练过程得到的分类规则用于样本测试检验和待测数据分类。

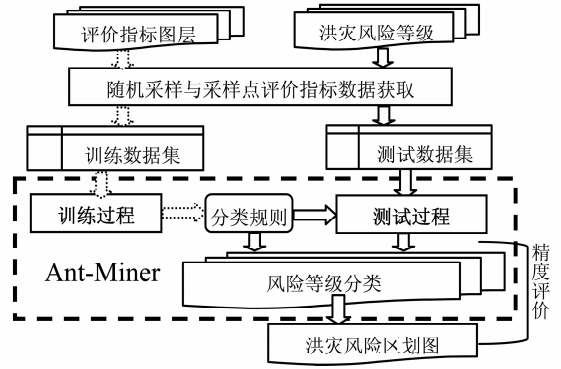


图 3 基于 Ant-Miner 的洪灾风险评价流程图

Fig. 3 Assessment flow chart of flood risk based on Ant-Miner

2 实例分析

2.1 实例流域及数据预处理

北江是珠江第二大河流，流域总面积 46 170 km²，一级支流有翁江、琶江、锦江等。由于暴雨量大，流域坡陡，水系又是叶脉分布，洪水容易集中，洪峰上涨速度快，具有山区洪水特点：峰型尖瘦，涨落较快，洪峰一般呈连续的多峰型。历史上该流域洪水泛滥频繁，损失较严重，对其洪灾风险进行评价具有重要现实意义。

依据灾害系统理论，遵循代表性系统性、客观性与可操作性等原则^[21-23]，选取 10 个评价指标，包括 6 个反映自然属性的指标：最大 3 d 降雨量 (mm)，台风年频次 (年/次)，坡度 (°)，河流缓冲区 (m)，径流深 (mm)，数字高程模型 (DEM) (m)；4 个反映社会属性的指标：距公路距离 (m)，耕地面积百分比 (%)，人口密度 (人/km²)，GDP 密度 (万元/km²)。各指标具体分布如图 4 所示。

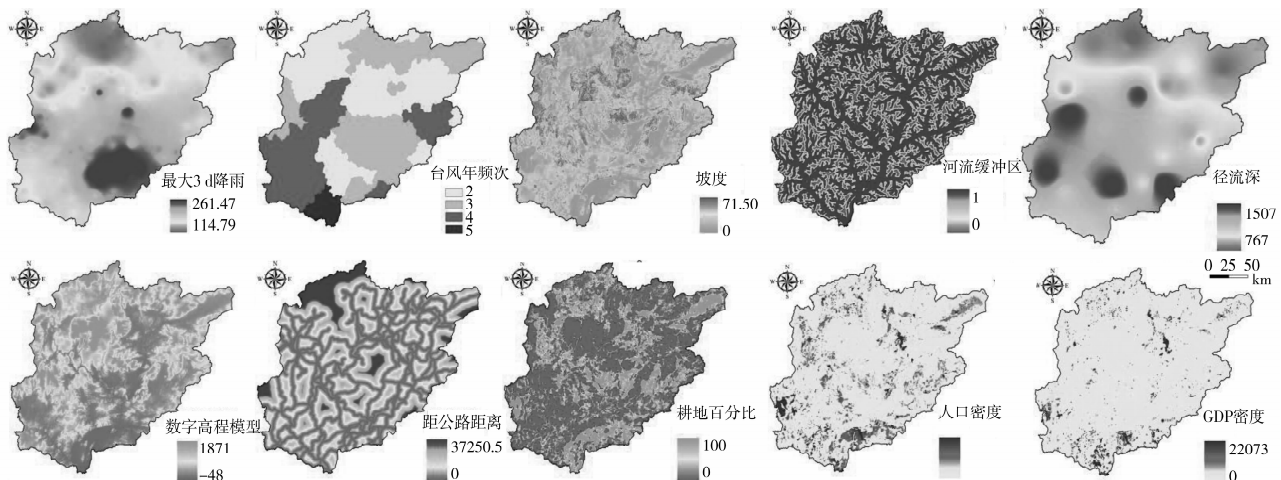


图 4 各指标图层分布特征

Fig. 4 Characteristic distributions of evaluation indexes

数据来源与预处理: 1961-2005 年最大 3 d 降雨量和径流深数据来源于广东省水文局 (<http://www.gdsw.gov.cn/wcm/gdsw/index.html>), 分别利用流域内降雨站和水文站数据, 采用克里金插值法插值获得; 台风数据来源于广东省气象局 (<http://www.grmc.gov.cn/>), 台风年频次是指各县市遭受台风袭击次数的多年平均值, 且以县(市)为单元进行赋值; 数字高程模型 (DEM) 数据来源于美国地质调查局 (<http://data.geocomm.com/dem/>), 用于表征流域地形情况; 坡度和河流数据是运用 GIS 技术通过 DEM 提取, 其中河流缓冲区是根据提取的河流的不同级别乘以 500m 后得到缓冲区, 再利用 Arc GIS 中的 Euclidean Distance 工具进行插值获得; 公路数据 (2005) 来源于广东省公路局 (<http://www.gdhighway.gov.cn/>), 到距公路距离指标的处理类似于河流缓冲区, 利用 Euclidean Distance 工具进行插值可得; 人口和 GDP 密度数据来源于国家基础地理信息共享网站 (<http://www.ngcc.cn/>); 耕地面积百分比来源于中国科学院共享网站 (<http://www.cas.cn/>)。在 ArcGIS9.3 软件 Spatial Analyst 模块支持下, 利用栅格计算器 (Raster Calculator) 的地图代数功能 (Map Algebra) 对上述 10 个评价指标图层进行极差标准化处理, 生成 10 个 grid 栅格数据图层, 空间分辨率为 100 m × 100 m。

2.2 实例应用步骤

为验证 Ant-Miner 可以被应用到洪灾风险区划领域中, 本文结合图 3 通过样本选取、分类规则挖掘 (训练)、精度检验 (测试)、风险等级识别 (分类) 4 个步骤构建基于 Ant-Miner 的洪灾风险区划模型, 详述如下:

1) 样本选取。训练样本选择是蚁群学习的关键步骤, 直接关系到所获规则质量。根据实际调查和现有的评价成果^[10,20-22], 利用 ArcGIS9.3 空间分析模块的 Sample 命令采用分层随机采样的方法来获取训练样本和测试数据集。训练数据集的样本数为 2 000, 测试数据集的样本数为 1 029。基于 WEKA (Waikato Environment for Knowledge Analysis) 平台利用基于熵的离散化方法对样本数据进行离散化处理^[15]。

2) 分类规则挖掘。洪灾风险等级分类规则通过 Ant-Miner 从训练数据中挖掘, Ant-Miner 是根据文献 [15] 提供的代码并作改进后在 VisualBasic6.0 环境中编程实现。在运用 Ant-Miner 挖掘规则时, 离散化后的各指标值作为蚂蚁路径的属性节

点, 洪灾风险等级作为蚂蚁路径的风险级别节点, 每条路径对应一条分类规则。分类规则的挖掘可以当作是蚂蚁对最优路径的搜索。本研究中, 对于 2 000 个训练样本, Ant-Miner 的参数设置如下: 蚁群数量 = 60; 每个规则覆盖的最少样本数 = 5; 最大未覆盖样本数 = 5; 收敛规则数 = 10; 迭代次数 = 1 000 可收敛。在这 5 个参数中, 蚁群数量和每个规则覆盖的最少样本数对分类精度较为敏感。训练结果表明: 模型总体分类精度随着蚁群数量的增加而改善, 但蚁群数量大于 60 以后基本稳定; 分类精度随着每个规则覆盖的最少样本数的增加而降低, 当其等于 5 时达到最佳。

3) 精度检验。应用 Ant-Miner 挖掘的分类规则前, 需利用测试数据集进行精度验证, 精度满足要求才能应用所挖掘的规则。输入 1 029 个测试数据检测所挖掘规则, 并生成混淆矩阵。

4) 风险等级分类。获取挖掘得到的分类规则后, 利用 ArcGIS9.3 栅格计算器的条件语句对风险等级进行分类, 并生成北江流域洪灾风险区划图。

2.3 结果与分析

通过实例应用的 4 个步骤后, 结合北江流域的 10 个指标, 借助 GIS 技术构建的基于 Ant-Miner 洪灾风险区划模型可获得以下主要结果:

1) 分类规则。利用构建的 Ant-Miner 模型对训练数据进行挖掘后共获得 15 条分类规则, 这些分类规则能覆盖所有风险等级。表 1 列出了模型参数及部分分类规则, 其中, 规则的置信度是根据该规则对测试数据的分类精度来确定。从表 1 可以看出, 该模型所挖掘的分类规则较简单, 均以条件语句形式表现, 毋需通过数学公式来表达, 能更方便和准确地描述评价指标间的复杂关系, 显然比数学公式更容易让人理解。与此同时, 根据上述规则可以地利用 ArcGIS9.3 中栅格计算器极其方便地生成风险区划图, 省略了其它智能模型的“待测栅格数据→点数据→判别→还原栅格数据”的转化与判别步骤, 大大节省了计算时间。

2) 测试精度。训练后构建的 Ant-Miner 模型需要满足精度要求才能进行下一步应用。根据生产的混淆矩阵 (表 2) 可知, 模型平均测试精度高达 95.1%, 表明建立的模型精度满足要求 (一般大于 80% 即可), 所挖掘的规则可被应用到全流域进行风险区划。为了突出 Ant-Miner 区划模型与其他模型的判别性能, 选用基于 BP 神经网络的洪灾风险区划模型^[10]进行训练与测试, 测试过程采用与 Ant-Miner 模型相同的训练与测试数据, 最后产生

表 1 Ant-Miner 挖掘的部分规则¹⁾
Table1 Parts of the rules mined by Ant-Miner

序号	IF 〈conditions〉	THEN 〈grade〉	置信度	Ant-Miner 参数
规则 1	a2 > =201 and a2 < =297 and a6 > =1850. 4 and a6 < =2313 and a8 > =1329.4	5	1.0	蚂蚁数：60
规则 2	a2 > =125 and a4 > =32 and a9 > =0.72 and a9 < =0.95	3	0.99	每个规则覆盖的
规则 3	a1 < =195.5 and a4 < =25 and a9 < =0.88 and a10 < =3.45	1	0.97	最少样本数：5
规则 4	a6 > =231.3 and a6 < =3932.1 and a9 > =1.00	4	0.91	最大未覆盖样本数：5
规则 5	a9 < =0.99	2	0.90	最大迭代数：1 000
.....				

1) a1 ~ 10 分别表示最大 3d 降雨量、DEM、GDP 密度、耕地面积百分比、人口密度、距公路距离、径流深、坡度、河流缓冲区、台风年频次 10 个指标的标准化数值；grade 为洪灾风险等级，1 ~ 5 分别表示安全、较安全、中等、较危险、危险。

测试混淆矩阵（表 3）。对比两混淆矩阵可知，BP 神经网络模型的平均测试精度为 92.9%，略低于 Ant-Miner 模型的结果，表明 Ant-Miner 模型的区划性能更好，在洪灾风险区划领域有较好的适用性。

3) 风险空间分布特征。根据挖掘的 15 条分类规则，利用 ArcGIS9.3 栅格计算器输入以条件语句表现的分类规则对风险级别进行分类，并生成北

江流域洪灾风险区划图（图 4）。根据风险区划图分布特征可知，危险区域主要集中于清远市区、四会市、南雄县中部、阳山县中部、怀集县南部、佛冈县、英德市中部等泛洪区；较危险区域主要位于乐昌市、仁化县、广宁县、翁源县、怀集县南部、英德市南部等泛洪区。对比图 4 指标分析可知，危险与较危险区域降雨较为丰富，且多集中于地势平

表 2 Ant-Miner 模型测试精度
Table 2 Test accuracy of Ant-Miner model

等级	1	2	3	4	5	合计	产生精度/%
1	197	7	0	0	0	204	96.6
2	7	186	3	0	0	196	94.9
3	0	0	167	20	0	187	89.3
4	0	0	0	191	7	198	96.5
5	0	0	0	6	238	244	97.5
合计	204	193	170	217	245	1 029	
使用精度/%	96.6	96.4	98.2	88.0	97.1		
总精度 = 95.1%							

表 3 BP 神经网络模型测试精度
Table 3 Test accuracy of BP neural network model

等级	1	2	3	4	5	合计	产生精度/%
1	194	8	2	0	0	204	95.1
2	7	185	4	0	0	196	94.4
3	0	12	157	18	0	187	84.0
4	0	0	3	186	9	198	93.9
5	0	0	1	9	234	244	95.9
合计	201	205	167	213	243	1 029	
使用精度/%	96.5	90.2	94.0	87.3	96.3		
总精度 = 92.9%							

缓的平原或洼地地区。这些地区能较快地汇集雨水并形成径流, 极易形成洪水或内涝; 与此同时, 这些地区人口与财产较为密集且农业较为发达, 因而导致风险较高。为了便于与洪灾实际情况进行比较, 选取北江流域 4 次典型的极端洪水灾害, 即 1915、1949、1982 和 1994 年大洪水, 以洪水淹没范围为基础^[20-22], 并结合灾情记录, 发现北江流域洪灾较重及风险较高的区域与图 5 中危险、较危险区域较为一致。验证结果表明图 5 能较好地反映北江流域存在的洪灾风险实际情况。



图 5 基于 Ant-Miner 模型的北江流域洪灾风险区划图

Fig. 5 Zoning map of flood disaster risk based on Ant-Miner in the Beijiang River basin

3 结论

本文试图将蚁群算法的规则挖掘 (Ant-Miner) 引入到洪灾风险区划中, 并建立基于 Ant-Miner 的洪灾风险区划模型, 在 GIS 技术支持下将此模型应用于北江流域, 结果表明: 1) 基于 Ant-Miner 洪灾风险区划模型在训练过程中共挖掘 15 条适合研究区的洪灾风险分类规则, 这些规则通过简单的条件语句而非复杂的数学公式表现, 极容易与 GIS 技术结合并生成风险区划图; 2) 基于 Ant-Miner 模型的测试精度为 95.1%, 而相同条件下的 BP 神经网络模型精度为 92.9%, 显然前者的分类性能更好, 对洪灾风险区划有更好的适用性; 3) Ant-Miner 识别的高风险区主要集中在降雨量较大、地势平缓低洼、人口和财产密集的地区, 与历史洪灾风险情况较吻合, 表明所构建的模型科学合理, 可

为洪灾风险区划提供新思路。

Ant-Miner 作为一种群智能方法, 在数据处理中具有较大的应用价值, 可为洪灾风险的智能化区划提供一种有效的方法。但毕竟将 Ant-Miner 应用到洪灾风险区划领域才刚刚起步, 在应用时还有一些限制, 如本研究仅选取了 10 个评价指标, 从流域宏观的角度分析洪灾风险分布是合理的, 但因忽略了水库、堤防等水工建筑物的防洪影响, 在微观上风险区划精度不一定能满足要求; 另外对于算法本身, 在数据挖掘过程中所发现的规则是按照顺序排列, 存在计算耗时相对较长等不足。

参考文献:

- [1] HOWARD S W. Flood hazard and management: a UK perspective [J]. *Philosophical Transactions of The Royal Society A*, 2006, 364: 2135 - 2145.
- [2] TAWATCHAI T, MOHAMMED F K. Flood hazard and risk analysis in the southwest region of Bangladesh [J]. *Hydrological Process*, 2005, 19: 2055 - 2069.
- [3] 李林涛, 徐宗学, 庞博, 等. 中国洪灾风险区划研究 [J]. *水利学报*, 2012, 43(1): 22 - 30.
- [4] 王兆礼, 赖成光, 陈晓宏. 基于熵权的洪灾风险空间模糊综合评价模型 [J]. *水力发电学报*, 2012, 31(5): 35 - 40.
- [5] ZOU Q, ZHOU J Z, ZHOU C, et al. Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. *Stoch. Environ Res Risk Assess.* 2013, 27: 525 - 546.
- [6] TINGSANCHALI T, KARIM F. Flood hazard assessment and risk-based zoning of a tropical flood plain: case study of the Yom River, Thailand [J]. *Hydrol Sci J*, 2010, 55(2): 145 - 161.
- [7] KUBALC, HAASE D, MEYER V, et al. Integrated urban flood risk assessment-adapting a multi criteria approach to a city [J]. *Nat Hazards Earth Syst Sci*, 2009, 9: 1881 - 1895.
- [8] LI LF, WANG J F, LEUNG H, et al. Assessment of catastrophic risk using bayesian network constructed from domain knowledge and spatial data [J]. *Risk Analysis*, 2010, 30(7): 1157 - 1175.
- [9] 赖成光, 陈晓宏, 赵仕威, 等. 基于随机森林的洪灾风险评价模型及其应用 [J]. *水利学报*, 2015, 46(1): 58 - 66.
- [10] 赖成光, 王兆礼, 宋海娟. 基于 BP 神经网络的北江流域洪灾风险评价 [J]. *水电能源科学*, 2011, 29(3): 57 - 59 + 161.
- [11] COLORNI A, DORIGO M, MANIEZZO V, et al. Distributed optimization by ant colonies [C] // *Proceedings*

- of European Conference on Artificial Life. Paris, 1991: 134 - 142.
- [12] PAREPINELLI R S, LOPES H S, FREITAS A A. Data mining with an Ant Colony Optimization algorithm [J]. IEEE Transaction in Evolutionary Computation, 2002, 6 (4): 321 - 332.
- [13] COSTA E P, LORENA A C, CARVALHO A, et al. Comparing several approaches for hierarchical classification of proteins with decision trees [C] // 2nd Brazilian Symposium on Bioinformatics (BSB2007). Angra dos Reis, BRAZIL, 2007: 126 - 137.
- [14] OMKAR S N, RAGHAVENDRA K U. Rule extraction for classification of acoustice mission signals using Ant Colony Optimisation [J]. Engineering Applications of Artificial Intelligence, 2008, 21: 1381 - 1388.
- [15] LIU XP, LI X, LIU L, et al. An innovative method to classify remote sensing images using Ant Colony Optimization [J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(22): 4198 - 4208.
- [16] LIU X P, LI X, LIU L, et al. A bottom-up approach to discover transition rules of cellular automata using ant intelligence [J]. International Journal of Geographical Information Science, 2008, 22(11/12): 1247 - 1269.
- [17] 刘小平, 黎夏, 叶嘉安, 等. 利用蚁群智能挖掘地理元胞自动机的转换规则 [J]. 中国科学 D 辑: 地球科学, 2007, 37(6): 824 - 834.
- [18] 戴芹, 刘建波. 基于蚁群规则挖掘算法的多特征遥感数据分类 [J]. 地理研究, 2009, 28(4): 1136 - 1145.
- [19] OTERO F, FREITAS A, JOHNSON C G. cAnt-Miner: An ant colony classification algorithm to cope with continuous attributes [C] // DORIGO M, et al, eds. Springer, Heidelberg: ANTS2008, LNCS: 5217: 48 - 59.
- [20] 宫清华, 黄光庆, 郭敏, 等. 基于 GIS 技术的广东省洪涝灾害风险区划 [J]. 自然灾害学报, 2009, 18(1): 58 - 63.
- [21] 广东省地图集 [M]. 广州: 广东地图出版社, 2003.
- [22] 广东省防汛防旱防风总指挥部. 广东水旱风灾害 [M]. 广州: 暨南大学出版社, 1996.

(上接第 121 页)

- [16] 徐宇丽, 王大伟, 史学伟, 等. 天山冻土产低温脂肪酶菌株的筛选及其多样性分析 [J]. 微生物学报, 2011, 51(2): 233 - 240.
- [17] ZHANG Dianpeng, SPADARO D, GARIBALDI A, et al. Efficacy of the antagonist *Aureobasidium pullulans* PL5 against postharvest pathogens of peach, apple and plum and its modes of action [J]. Biological Control, 2010, 54(3): 172 - 180.
- [18] 高璇璇, 汪建明, 卢星达, 等. 高产普鲁兰多糖的出芽短梗霉的诱变育种 [J]. 中国食品添加剂, 2013(1): 45 - 49.
- [19] KRUPINSKI M, SZEWCZYK R, DŁUGONSKI J. Detoxification and elimination of xenoestrogen nonylphenol by the filamentous fungus *Aspergillus versicolor* [J]. International Biodeterioration & Biodegradation, 2013, 82: 59 - 66.
- [20] 张连庆, 车茜, 李德海, 等. 海洋真菌 *Aspergillus versicolor* ZLQ-43 的次级代谢产物的研究 [J]. 中国海洋药物, 2014, 33(3): 1 - 6.
- [21] MARGESIN R, SCHUNNER F. Low - temperature bioremediation of a waste water contaminated with anionic surfactants and fuel oil [J]. Applied Microbiology And biotechnology, 1998, 49(4): 482 - 486.