

黄滢,陈燕丽,莫伟华,等. 基于数据关联分析的红树林虫害与气象条件关系研究[J]. 气象研究与应用,2022,43(4):20–25.

Huang Ying, Chen Yanli, Mo Weihua, et al. Data Association analysis to the relationship between mangrove pests and meteorological conditions[J]. Journal of Meteorological Research and Application, 2022, 43(4): 20–25.

基于数据关联分析的红树林虫害与气象条件关系研究

黄滢¹, 陈燕丽^{2*}, 莫伟华², 许文龙¹, 莫建飞²

(1.防城港市气象局, 广西 防城港 538001; 2. 广西壮族自治区气象科学研究所, 南宁 530022)

摘要: 为研究气象条件对红树林病虫害发生的影响, 利用 2016—2020 年广西防城港红树林生态自然保护区内红树林虫害灾情数据和气象数据, 采用数据挖掘技术 Apriori 算法挖掘广州小斑螟、柚木肖弄蝶夜蛾虫害与气象因素之间的关联规则。结果表明, 广州小斑螟虫害发生与降水量、湿度、雨日密切相关, 湿度和雨日的气候异常程度对广州小斑螟虫害级别产生影响; 而柚木肖弄蝶夜蛾的气象影响因子有温度、日照、雨日、降雨量、湿度等多个因子。

关键词: 数据挖掘; Apriori 算法; 广州小斑螟虫害; 柚木肖弄蝶夜蛾虫害; 气象因素

中图分类号: S763 **文献标识码:** A **doi:** 10.19849/j.cnki.CN45-1356/P.2022.4.04

引言

红树林处于海洋与陆地的交错带, 生态系统独特, 病虫害是红树林植物种群变化的重要因素^[1]。天气异常引起的红树林生态演变异常、生态安全事件和生态气象灾害的增加, 近十余年广西红树林几乎每年都不同程度地遭受病虫害危害, 导致红树植物生长受限, 严重情况下出现红树林大面积死亡, 如 2004 年和 2006 年广州小斑螟为主的虫害, 2015 年柚木驼蛾虫害, 造成红树林大面积枯萎死亡, 红树林生境面临着严重的威胁。因此, 研究红树林病虫害监测指标及评估方法, 为红树林病虫害预测预警及科学防治提供科学依据, 对红树林保护工作具有重要意义。

气象因素是影响虫害发生发展的重要因子。虫害的发生和气温的关系研究较多, 研究表明许多害虫会因年平均温度的升高而发生期提前, 世代数也相应增多^[2-5], 一些病害在温度适宜时其发生率最高^[6]。温度对促进或抑制某种病虫害的发生、发展、流行及其危害程度都可能产生显著的影响^[7-10]。一些学者研究发现红树林虫害暴发与降雨偏少、海温变化等

具有相关关系^[11-14]。一些学者研究发现海温、气候异常信号对病虫害的发生流行具有前兆性指示^[15-19]; 叶彩玲等^[20]发现气候变化可能造成病虫害危害的地理范围扩大, 程度加剧。在定量分析方面, 梁燕红等^[21-22]将其应用到广西红树林虫害发生与气候因素的关系挖掘上, 发现广西山口国家红树林生态自然保护区的广州小斑螟虫害级别与气温、降水量密切相关。目前有关气象因子与红树林病虫害的定量分析研究还较少, 虫害和降水量、气温、湿度、日照等气象要素的相关关系还有待于进一步挖掘。

广州小斑螟和柚木肖弄蝶夜蛾是红树植物最普遍、最主要的害虫, 红树林受其袭击, 会导致叶片枯萎, 枝干死亡, 积蓄量大幅减少, 影响下一年繁殖。数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的实际数据中, 提取有用知识的过程, 关联规则是数据挖掘技术中一项重要的数据分析方法, 广泛应用于系统科学、临床医学、应用经济学、作物学、畜牧学等多个领域, 寻找复杂数据之间的关联。防城港面朝东南亚地区, 靠近北部湾, 西南边界与越南接壤, 红树林覆盖面积约有 1500hm², 拥有如北仑河口红树林、东湾红树林、簕山古渔村区域等红树林带, 其

收稿日期: 2022-08-10

基金项目: 广西红树林保护与利用重点实验室开放基金项目 (GKLMC-202004 和 GKLMC-201804)、中国气象局气候变化专项 (CCSF202030)、广西气象科研计划 (桂气科 2019M02)

作者简介: 黄滢 (1981—), 女, 高级工程师, 主要从事气象应用评估工作。E-mail: huangying100@sina.com

* 通讯作者: 陈燕丽 (1982—), 女, 正研级高级工程师, 主要从事生态气象监测评估研究。E-mail: cyl0505@sina.com

中北仑河口属于国家级自然保护区, 现存的红树林面积约 1300hm²。红树林生态环境系统脆弱, 易受到外界因素的影响, 病虫害是近年来影响红树林保护的首要问题, 一般来说, 虫害一旦爆发, 蔓延速度会很快, 影响范围也会比较广, 具有很大的危害。由于红树林的生态较为脆弱, 一旦虫害爆发, 很难大规模的使用化学药剂进行防治。本研究以防城港红树林生态保护区为研究对象, 选取雨量、雨日、湿度、温度、日照等多种气象因子, 运用数据挖掘技术分析影响广州小斑螟和柚木肖弄蝶夜蛾两种主要红树林病虫害的气象因子差异, 结合历年的病虫害情况, 为红树林虫害防治预警提供科学参考依据。

1 资料与方法

1.1 资料与来源

气象数据由防城港市气象局提供, 包括防城港站逐月降雨量、雨日、平均气温、日照时数、相对湿度 5 种气象因子, 计算各气象因子的历史气候值 (1981—2010 年), 同时计算虫害发生年份各气象因子和气候均值的距平或偏差, 用于比较某年气象要素和气候均值的偏离程度。

虫害数据由防城港市林业局提供, 包括 2016—2020 年共 5a 防城港市红树林柚木肖弄蝶夜蛾、广州小斑螟两个主要红树林虫害类型的监测情况、虫害发生面积、成灾面积等灾情数据。

1.2 数据挖掘研究方法

1.2.1 Apriori 计算方法

若两个或多个变量的取值之间存在某种规律性, 就称为关联, 关联规则是寻找在同一个事件中出现的不同项的相关性。Apriori 算法是第一个关联规则挖掘算法, 由 R.Agrawal 和 R.Srikant 于 1994 年提出。关联分析有两个目标: 发现频繁项集和发现关联规则。Apriori 算法的目标是找出支持度大于等于 min-sup 并且置信度大于等于 min-conf 的所有规则, 包含频繁集的产生和规则的产生两个子任务。使用候选项集找出频繁项集的具体过程是: 首先会生成所有单个物品的项集列表; 扫描记录来查看哪些项集满足最小支持度要求, 那些不满足最小支持度的集合会被去掉; 对剩下的集合进行组合以生成包含两个元素的项集; 接下来重新扫描记录, 去掉不满足最小支持度的项集, 重复进行直到所有项集都被去掉。

1.2.2 Apriori 算法数据挖掘

数据之间的联系用关联规则来表示, 表达式为:

$\{X\} \rightarrow \{Y\}$, 把规则前面的项集叫前件, 把规则后面的项集叫后件。

支持度用于表示项集在事务中出现的概率。设置最小支持度, 用于筛选符合需求的项集, 被留下来的项集 (\geq 最小支持度), 称为频繁项集。

置信度用于衡量关联规则的可靠程度, 表示在前件出现的情况下, 后件出现的概率。一般来说, 概率越高, 规则的可靠性越强。设置最小置信度, 用于筛掉一些不符合需求的关联规则。被留下来的关联规则 (\geq 最小置信度), 叫做强关联规则。

提升度用于判断关联规则的促进关系或抑制关系, 评估 X 的出现, 对 Y 出现的影响有多大。提升度的值小于 1, 表示前件对后件是抑制的关系; 提升度大于 1, 表示前件对后件是促进的关系; 提升度的值等于 1 时, 表示前件不影响后件, 两者之间没有关系。

1.3 影响红树林虫害的气象因子选择

根据 2016—2020 年气象数据, 结合防城港林业局提供的历年病虫害数据, 选取气象因子有冬季 (前一年 12 月到次年 2 月)、春季 (3—5 月)、夏季 (6—8 月) 的雨量、雨日、平均气温、日照时数、相对湿度等 15 个气象要素作为信息挖掘数据集。

2 结果与分析

2.1 红树林虫害和气象因子等级划分

为衡量虫害严重程度和气象因素偏离气候值的严重程度, 将虫害面积和气象要素数偏离气候值的程度离散化, 即通过数据映射, 实现了对连续数据分级。原始数据虫害受灾面积按林业局经验划分进行离散化处理划分为 5 个等级 (表 1)。将降水量和降水日数数据和历史气候值相比较, 得出与多年历史气候值对比的降水距平和雨日距平, 再将其离散化, 划分为 8 个等级; 将温度、日照、湿度和历史气候值相比较, 计算其偏差, 然后将其数字离散化, 划分等级 (表 2)。

表 1 虫害情况等级划分

虫害面积/hm ²	虫害面积/hm ²	等级
< 10	< 10.0	等级 1
[10.0, 20.0)	[10.0, 20.0)	等级 2
[20.0, 46.7)	[20.0, 46.7)	等级 3
[46.7, 66.7)	[46.7, 66.7)	等级 4
≥ 66.7	≥ 66.7	等级 5

表 2 和气候值比较降水、雨日、温度、日照、湿度等级划分

降水距平/%	等级	雨日距平/%	等级	温度偏差/℃	等级	日照距平/%	等级	湿度偏差/%	等级
< -50	等级-4	< -50	等级-4	< -0.5	等级-3	< -50	等级-4	< -10	等级-3
[-50 ~ -30)	等级-3	[-50 ~ -30)	等级-3	[-0.5 ~ -0.3)	等级-2	[-50 ~ -30)	等级-3	[-10 ~ -5)	等级-2
[-30 ~ -10)	等级-2	[-30 ~ -10)	等级-2	[-0.3 ~ 0)	等级-1	[-30 ~ -10)	等级-2	[-5 ~ 0)	等级-1
[-10 ~ 0)	等级-1	[-10 ~ 0)	等级-1	[0 ~ 0.3)	等级+1	[-10 ~ 0)	等级-1	[0 ~ 5)	等级+1
[0 ~ 10)	等级+1	[0 ~ 10)	等级+1	[0.3 ~ 0.5)	等级+2	[0 ~ 10)	等级+1	[5 ~ 10)	等级+2
[10 ~ 30)	等级+2	[10 ~ 30)	等级+2	≥ 0.5	等级+3	[10 ~ 30)	等级+2	≥ 10	等级+3
[30 ~ 50)	等级+3	[30 ~ 50)	等级+3			[30 ~ 50)	等级+3		
≥ 50	等级+4	≥ 50	等级+4			≥ 50	等级+4		

预处理后的数据可以反映虫害等级，和各季节
雨量、雨日、温度、湿度偏离历史气候均值的情况

(冬季数据预处理结果如表 3,春季及夏季数据预处
理结果略)。

表 3 冬季数据预处理后结果

年份	虫害等级		冬季				
	小斑螟	夜蛾	雨量 R	雨日 RD	温度 T	日照 S	湿度 H
2016	BM4	YE5	冬季 R+4 Winter R+4	冬季 RD+4 Winter RD+4	冬季 T-3 Winter T-3	冬季 S-1 Winter S-1	冬季 H+1 Winter H+1
2017	BM4	YE1	冬季 R-1 Winter R-1	冬季 RD-2 Winter RD-2	冬季 T+3 Winter T+3	冬季 S+2 Winter S+2	冬季 H+1 Winter H+1
2018	BM1	YE1	冬季 R-2 Winter R-2	冬季 RD+2 Winter RD+2	冬季 T-1 Winter T-1	冬季 S-2 Winter S-2	冬季 H-1 Winter H-1
2019	BM5	YE1	冬季 R+4 Winter R+4	冬季 RD+2 Winter RD+2	冬季 T+3 Winter T+3	冬季 S-3 Winter S-3	冬季 H+2 Winter H+2
2020	BM1	YE1	冬季 R+4 Winter R+4	冬季 RD+1 Winter RD+1	冬季 T+3 Winter T+3	冬季 S+2 Winter S+2	冬季 H+1 Winter H+1

2.2 红树林虫害和气象因子关联分析

设最小支持度为 40%，最小置信度为 60%，根据 Apriori 算法,得到满足强关联规则的挖掘结果如表 4、表 5,表中前件按“季节+气象要素+等级”来显示。Spring、Summer、Autumn、Winter 分别表示春季、夏季、秋季和冬季。气象要素缩写 R、RD、T、S、H 分别代表降水距平、雨日距平、温度偏差、日照距平和湿度偏差。等级用正负数字表示,正值越大,表明超过平均值越多,负值越大,表明低于平均值越多。例如 WinterH+1 表示冬季湿度偏高于气候值等级为 1, SpringRD-2 表示春季雨日偏低低于气候值等级为 2。

对关联规则进行提取及分析,可以发现:

(1)广西防城港红树林的广州小斑螟虫害发生

与降水量、湿度、雨日密切相关,冬春夏秋季的湿度,春季和夏季的降雨量,春季的雨日是对虫害级别最重要的影响因素。如果春秋冬季的湿度偏大、夏季的湿度偏小、春季或夏季的雨量偏少、春季雨日偏少,容易导致小斑螟病虫害的发生。气温、日照和广州小斑螟病虫害发生的关联不大。

(2)湿度和雨量的异常程度对广州小斑螟虫害级别产生影响。虫害级别与春季湿度关联性最强,秋季和冬季湿度是下一个重要的决定因素,当春季或冬季湿度偏大级别为 1 级、秋季湿度偏离程度为 2 时,小斑螟虫害达到严重的 4 级的可能性较大。而当夏季湿度偏小级别为 1 级或 2 级,春季雨日偏少级别为 2 级时,小斑螟虫害有很大的可能为轻度的 R1 级。

表 4 广州小斑螟虫害挖掘规则结果

前件	后件	支持度	置信度	提升度
SpringH+1	BM4	0.4	1.0	2.5
WinterH+1 & SpringR-2	BM4	0.4	1.0	2.5
WinterH+1 & SpringH+1	BM4	0.4	1.0	2.5
SpringR-2 & SpringH+1	BM4	0.4	1.0	2.5
SpringR-2 & AutumnH+2	BM4	0.4	1.0	2.5
AutumnH+2 & SpringH+1	BM4	0.4	1.0	2.5
SpringR-2 & AutumnH+2 & SpringH+1	BM4	0.4	1.0	2.5
WinterH+1 & AutumnH+2 & SpringH+1	BM4	0.4	1.0	2.5
WinterH+1 & SpringR-2 & AutumnH+2	BM4	0.4	1.0	2.5
WinterH+1 & SpringR-2 & SpringH+1	BM4	0.4	1.0	2.5
WinterH+1 & SpringR-2 & AutumnH+2 & SpringH+1	BM4	0.4	1.0	2.5
AutumnH+2	BM4	0.4	0.7	1.7
WinterH+1	BM4	0.4	0.7	1.7
WinterH+1 & AutumnH+2	BM4	0.4	0.7	1.7
SummerH-1	BM1	0.4	0.7	1.7
SpringRD-2	BM1	0.4	0.7	1.7
SpringRD-2 & SummerH-1	BM1	0.4	0.7	1.7
SummerR-2 & SpringRD-2	BM1	0.4	0.7	1.7
SummerR-2 & SummerH-1	BM1	0.4	0.7	1.7
SummerR-2 & SpringRD-2 & SummerH-1	BM1	0.4	0.7	1.7

表 5 柚木肖弄蝶夜蛾挖掘规则结果

前件	后件	支持度	置信度	提升度
SummerH-1	YE1	0.6	1.0	1.3
SpringRD-2	YE1	0.6	1.0	1.3
SummerS-2	YE1	0.6	1.0	1.3
WinterT+3	YE1	0.6	1.0	1.3
SpringRD-2 & SummerH-1	YE1	0.6	1.0	1.3
SummerR-2 & SpringRD-2	YE1	0.6	1.0	1.3
SummerR-2 & SummerH-1	YE1	0.6	1.0	1.3
SummerS-2 & SpringR-2	YE1	0.6	1.0	1.3
SummerR-2 & SpringRD-2 & SummerH-1	YE1	0.6	1.0	1.3

(3)柚木肖弄蝶夜蛾的气象影响因子有春季和夏季的降水、春季的雨日、冬季的温度、夏季的湿度和日照 6 个因子,夏季的湿度偏低(偏离程度 1 级)、春季的雨日偏少(偏离程度 2 级)、夏季日照偏少(偏离程度 2 级),以及冬季的温度偏暖(偏离程度 3 级)易导致柚木肖弄蝶夜蛾虫灾出现,春夏季的雨量雨日湿度偏少的组合因子也会导致柚木肖弄蝶夜蛾虫灾出现。因未满足最小支持度和最小置信度的关联规则,未挖掘到柚木肖弄蝶夜蛾重度虫害的关

联信息。

3 结论与讨论

利用气象要素对比气候均值衡量天气要素异常程度,用关联分析方法挖掘红树林虫害严重程度和气象要素异常程度之间的关联信息,挖掘了广西防城港红树林广州小斑螟、柚木肖弄蝶夜蛾虫害级别与气象因素之间的关联规则。规则表明,广州小斑螟虫害发生与降水量、湿度、雨日密切相关,湿度

和雨日的气候异常程度对广州小斑螟虫害级别产生影响。而柚木肖弄蝶夜蛾的气象影响因子有降雨量、雨日、温度、湿度、日照等多个因子。根据挖掘到关联规则,结合气象预报信息,可提前采取措施防御虫害。

数据关联分析擅长寻找数据之间的关联性,但不能分析数据之间的因果关系,是其方法的短板,需进一步结合害虫的生理生态和生长期,对分析结果进一步进行深入分析。此外,本文收集到的红树林虫害灾情数据时间序列较短,可以作为挖掘数据关联分析使用,可解释虫害的一些相关因素,但如果用于预测虫害灾情的发生还需要积累更长时间序列的虫害灾情数据。

参考文献:

- [1] 胡刚,黎洁,覃盈盈,等.广西北仑河口红树植物种群结构与动态特征[J].生态学报,2018,38(9):3022-3034.
- [2] 丁文广,肖俊豪,汪霞.气候变化对我国森林自然灾害的影响[J].西北林学院学报,2010,25(5):117-120.
- [3] Jonsson A M, Harding S, Barring L, et al. Impact of climate change on the population dynamics of *Ips typographus* in southern Sweden[J].Agricultural and Forest Meteorology, 2007,146(1-2):70-81.
- [4] Massimo Faccoli. Effect of weather on *Ips tyographus* (Coleoptera Curculionidae) phenology, voltinism, and associated spruce mortality in the Southeastern Alps[J]. Environmental Entomology, 2009,38(2):307-316.
- [5] 赵青山.马尾松毛虫发生动态和大发生预测预报的初步研究[J].林业科学,1981(2):123-133.
- [6] 曾柏全,邓子牛,熊兴耀.柑橘产业中病害问题及解决途径[J].安徽农业科学,2008(13):5485-5486.
- [7] 王娟,姬兰柱,Marina Khomutova.黑龙江大兴安岭地区森林害虫发生面积与气象因子的关系[J].生态学杂志,2007(5):673-677.
- [8] 习妍,牛树奎.气候要素对松材线虫病疫情的影响研究[J].林业资源管理,2008(4):70-76.
- [9] 王纯枝,郭安红,王玉玲,等.华北地区杨树烂皮病发生发展气象适宜度预报模型[J].中国农业气象,2011,32(1):139-143.
- [10] 王闫利,王茹琳,姜淦,等.森林害虫发生发展与气象因素关系的研究进展[J].四川林业科技,2012,33(4):20-24.
- [11] 范航清,邱广龙.中国北部湾白骨壤红树林的虫害与研究对策[J].广西植物,2004(6):558-562.
- [12] 刘文爱,范航清.广州小斑螟发生与环境因子的关系[J].生态学报,2011,31(23):335-338.
- [13] 韦江玲,刘文爱,黄琦,等.广西山口红树林保护区近15 a 主要虫害调查[J].福建林业科技,2019,46(4):66-69.
- [14] 范航清,刘文爱,曹庆先.广西红树林害虫生物生态特性与综合防治技术研究[M].北京:科学技术出版社,2012:183-186.
- [15] 赵圣菊.用海温作第一代粘虫发生量和发生期长期预测的初步探讨[J].气象科技,1983(5):46-48.
- [16] 赵圣菊.用海温预测二代粘虫发生区一代成虫迁入期的模式研究[J].气象科学研究院院刊,1987(1):88-96.
- [17] 马树庆,马吉祥,张传亮.用海温和大气环流资料建立粘虫迁入期和发生程度的长期预报模式[J].吉林气象,1993,11(1):32-38.
- [18] 霍治国,叶彩玲,钱拴,等.气候异常与中国小麦白粉病灾害流行关系的研究[J].自然灾害学报,2002(2):85-90.
- [19] 侯婷婷,霍治国,卢志光,等.副热带高压与我国稻飞虱发生关系的研究[J].自然灾害学报,2003,12(2):213-219.
- [20] 叶彩玲,霍治国,丁胜利,等.农作物病虫害气象环境成因研究进展[J].自然灾害学报,2005(1):90-97.
- [21] 梁燕红,梁志清,黄琦,等.C4.5 算法在广州小斑螟发生与气候因素分析中的应用——以广西山口国家红树林生态自然保护区为例[J].中国森林病虫,2019,38(1):34-37.
- [22] 梁燕红.数据挖掘技术在广西红树林病虫害气候因素分析中的应用——以广州小斑螟为例[J].玉林师范学院学报,2017,38(5):131-135.

Data association analysis to the relationship between mangrove pests and meteorological conditions

Huang Ying¹, Chen Yanli^{2*}, Mo Weihua², Xu Wenlong¹, Mo Jianfei²

(1. Fangchenggang Meteorological Bureau, Guangxi Fangchenggang 538001, China;

2. Guangxi Institute of Meteorological Sciences, Nanning 530022, China)

Abstract: Using the mangrove pests data and meteorological data from 2016 to 2020 in Fangchenggang Mangrove Ecological Nature Reserve, the Apriori algorithm was used to analyze the association rule between meteorological factors and the Cantonensis pests and Teak Xiaonong Spodoptera in order to study the effect of meteorological conditions on the occurrence of mangrove pests. The analysis found that the occurrence of the Cantonensis pests was closely related to precipitation, humidity, and rainy days, and the amount of humidity and rainy weather had an impact on the Cantonensis pests level. The meteorological influencing factors of the Teak Xiaonong Spodoptera were influenced by a variety of weather conditions, including temperature, sunshine, rainy days, rainfall, humidity and other factors.

Key words: Data mining; Apriori algorithm; Cantonensis pests; Teak Xiaonong Spodoptera; meteorological factors