

文章编号:2096 - 5389(2021)06 - 0099 - 03

基于决策树模型的贵州降雹识别研究

彭宇翔^{1,2}, 文继芬^{1,2}, 李 皓^{1,2}, 刘 涛^{1,2}, 唐辟如^{1,2}, 郭 茜³

(1. 贵州省人工影响天气办公室, 贵州 贵阳 550081;

2. 贵州省气象灾害防御技术中心, 贵州 贵阳 550081; 3. 贵州省气象信息中心, 贵州 贵阳 550002)

摘 要:以 FY-2G 卫星反演产品为输入参数建立决策树模型, 对 2020 年贵州降雹进行识别研究。收集了 2020 年贵州 68 个降雹点数据和 68 个未降雹点数据, 从中随机选取 58 组降雹点和 58 组未降雹点数据用于建立决策树模型, 剩余 10 组降雹点和 10 组未降雹点数据用于检验所建立模型的识别效果。结果表明, 所建模型降雹识别准确率为 80%, 其中对 10 个降雹点识别准确率为 70%, 对 10 个未降雹点识别准确率为 90%。

关键词:决策树; 冰雹; 识别; 检验

中图分类号:P426.64 **文献标识码:**B

Hail Recognition in Guizhou Based on Decision Tree Model

PENG Yuxiang^{1,2}, WEN Jifen^{1,2}, LI Hao^{1,2}, LIU Tao^{1,2}, TANG Piru^{1,2}, GUO Xi³

(1. Guizhou Weather Modification Office, Guiyang 550081, China;

2. Guizhou Meteorological Disaster Prevention Technology Center, Guiyang 550081, China;

3. Guizhou Meteorological Information Center, Guiyang 550002, China)

Abstract: In this paper, based on the FY-2G satellite inversion products as input parameters, a decision tree model is established to identify hail in Guizhou in 2020. Data of 68 hail spots and 68 non-hail spots in Guizhou in 2020 were collected. The data of 58 groups of hail spots and 58 groups of non-hail spots were selected at random to establish the decision tree model, and the remaining 10 groups of hail spots and 10 groups of non-hail spots were used to suggest the recognition effect of the established model. The results show that the recognition accuracy of the model is 80%, among which the recognition accuracy of 10 hail spots is 70%, and the recognition accuracy of 10 non-hail spots is 90%.

Key words: decision tree; hail; identification; test

0 引 言

FY-2G 卫星观测资料是目前我国人工影响天气业务主要使用的卫星资料, 中国气象局人工影响天气中心基于 FY-2G 卫星提供的云顶高度、云顶温度、过冷层厚度、光学厚度、有效粒子半径、液水

路径、黑体亮温等 7 项反演产品在我国人工影响天气监测预警业务中发挥了重要作用。

近年来, 很多学者利用卫星红外云图、可见光、中波红外等资料针对对流云的识别开展了研究^[1-5], 倪煜淮等^[3]利用 FY-2D 逐小时亮温资料对 2019 年一次冰雹天气过程进行了分析研究; 针对

收稿日期: 2021-06-09

第一作者简介: 彭宇翔(1990—), 男, 工程师, 主要从事人工影响天气业务与冰雹识别研究工作, E-mail: 1070792379@qq.com。

通讯作者简介: 文继芬(1963—), 女, 正研, 主要从事人工影响天气科研业务工作, E-mail: 55579059@qq.com。

资助项目: 贵州省科技计划项目(黔科合基础-ZK[2021]一般 217); 基于风云卫星观测资料的冰雹天气识别研究; 贵州省气象局科研业务项目(黔气科登[2020]07-13 号); 基于 Logistic 回归模型的贵州降雹识别指标研究; 中国气象局人工影响天气中心业务项目: FY3/4 卫星云特性产品在西南防雹增雨中的应用示范(一期)。

卫星反演产品也有相关研究,安晓存等^[5]分析了云顶亮温、云顶亮温梯度特征与冰雹的关系;孙玉稳等^[7]分析了 1987 年 5 次降雹个例的云系的云顶温度与地面雹雨分布特征;刘小艳等^[8]利用 CPAS 系统统计分析了冰雹个例中 FY-2G 卫星反演产品的特征参数及其时间变化。但是,现阶段还没有学者利用 FY-2G 卫星反演产品进行降雹识别研究。FY-2G 卫星提供的反演产品虽已用于贵州人工影响天气监测预警业务,但针对冰雹云的监测还主要是依靠地面雷达,卫星反演产品主要是对全省云系发展进行较大范围的宏观监测,且主要是依靠个人经验进行主观定性判断,针对反演产品对降雹是否具有实质性的指示作用还没有进行过深入研究。因此,本研究主要基于 FY-2G 卫星反演产品,建立决策树降雹识别模型,验证卫星反演产品对冰雹的识别作业。

本文以 FY-2G 卫星的反演产品为输入参数建立决策树模型,利用检验集数据检验模型识别效果,检验卫星反演产品对降雹识别研究的有效性。

1 数据与方法

1.1 FY-2G 卫星数据

FY-2G 是风云二号(03 批)卫星中的第二颗卫星,于 2014 年 12 月 31 日成功发射,自 2015 年 7 月 1 日开始定位于 105°E 赤道上空,并提供观测服务,是目前我国人工影响天气业务主要使用的卫星资料,中国气象局人工影响天气中心基于 FY-2G 卫星提供的云顶高度、云顶温度、过冷层厚度、光学厚度、有效粒子半径、液水路径、黑体亮温等 7 项反演产品在我国人工影响天气监测预警业务中发挥了重要作用。

本文收集了 2020 年 3—5 月 11 个冰雹日 136 组 FY-2G 卫星反演产品数据,其中包括了 68 个降雹点数据和 68 个未降雹点数据,每个降雹点选取降雹时段之前或者之后 15 min 内的反演产品作为该时段的反演数据,选取该 11 个冰雹日中未降雹点的反演数据作为对比。将该数据集分为模型训练集和模型检验集。从中随机选取 116 组数据作为模型训练集用于训练模型(其中包括 58 组降雹点和 58 组未降雹点数据),剩余 20 组数据作为模型检验集(其中包括 10 组降雹点和 10 组未降雹点数据)。

1.2 决策树模型

决策树是一个预测模型,他代表的是对象属性与对象值之间的一种映射关系。树中每个节点表示某个对象,每个分叉路径代表某个可能的属性

值,每个叶结点对应从根节点到该叶节点所经历的路径所表示的对象的值。决策树仅有单一输出,若欲有复数输出,可以建立独立的决策树以处理不同输出。数据挖掘中决策树是一种经常要用到的技术,可以用于分析数据,同样也可以用来作预测。

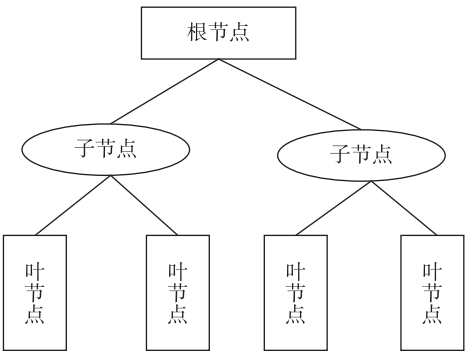


图 1 决策树模型示意图
Fig. 1 Schematic diagram of decision tree model

从数据产生决策树的机器学习技术叫做决策树学习。每个决策树都表述了一种树型结构,由它的分支来对该类型的对象依靠属性进行分类。每个决策树可以依靠对源数据库的分割进行数据测试。决策树如果依靠数学的计算方法可以取得更加理想的效果。

2 结果与讨论

2.1 决策树模型建立

在 2020 年 3—5 月 11 个冰雹日 136 组 FY-2G 卫星反演产品数据中随机选取 116 组数据作为模型训练集用于训练模型(其中包括 58 组降雹点和 58 组未降雹点数据),利用 MATLAB 软件完成决策树模型建立。

2.2 模型识别效果检验

2020 年 3—5 月 11 个冰雹日 136 组 FY-2G 卫星反演产品数据中建模随机选取 116 组数据后剩余的 20 组数据作为模型检验集(其中包括 10 组降雹点和 10 组未降雹点数据)对模型识别效果进行检验,结果显示:所建决策树模型对训练集中 10 个降雹点准确识别了 7 个,3 个降雹点被识别为未降雹点;对 10 个未降雹点准确识别了 9 个,1 个未降雹点被识别为了降雹点。

可见,所建决策树模型对降雹与未降雹点总的识别准确率为 80%,其中对 10 个未降雹点识别准确率为 90%,对 10 个降雹点识别准确率为 70%。这表明该模型对所建立数据集中的未降雹点识别

准确率高于降雹点的识别准确率;将降雹点识别为未降雹会导致实际业务中的漏报,而将未降雹点识别为降雹点会导致空报,因此,所建立的模型空报率为 10%,漏报率为 30%,空报率低于漏报率。

3 结论

本文以 FY-2G 卫星的反演产品为输入参数建立决策树模型,对 2020 年贵州冰雹云进行识别研究,所使用的 7 项反演产品包括:云顶高度、云顶温度、过冷层厚度、光学厚度、有效粒子半径、液水路径、黑体亮温。将收集的 2020 年 3—5 月 11 个冰雹日 136 组 FY-2G 卫星反演产品数据,其中包括了 68 个降雹点数据和 68 个未降雹点数据分为模型训练集和模型检验集。从中随机选取 116 组数据作为模型训练集用于训练模型,完成决策树模型建立,利用剩余 20 组数据作为模型检验集,验证模型识别效果。结果表明,所建模型冰雹识别准确率为 80%,其中对 10 个降雹点识别准确率为 70%,对 10 个未降雹点识别准确率为 90%,这表明该模型对所建立数据集中的未降雹点识别准确率高于降雹点的识别准确率。将降雹点识别为未降雹会导致实际业务中的漏报,而将未降雹点识别为降雹点会导

致空报,所建立的模型空报率(10%)低于漏报率(30%),这表明该模型如果投入人工影响天气冰雹识别业务,则该模型的空报率较低,但存在一定程度的漏报。

参考文献

- [1] 刘健,李云. 风云二号静止气象卫星的云相态识别算法[J]. 红外与毫米波学报,2011,30(4):322-327.
- [2] 白洁,王洪庆,陶祖钰. GMS 卫星红外云图强对流云团的识别与追踪[J]. 热带气象学报,1997(2):63-72.
- [3] 倪煜淮,何宏让,陈涛. 基于雷达和卫星资料对一次冰雹天气过程的中尺度特征分析[J]. 海洋技术学报,2019,38(3):59-63.
- [4] 尹跃,李万彪,姚展予,等. 利用 FY-2C 资料对西北太平洋海域云分类的研究[J]. 北京大学学报(自然科学版),2009,45(2):257-263.
- [5] 安晓存,董晶,景学义. 利用静止卫星云图特征对黑龙江省冰雹天气进行估计[J]. 黑龙江气象,2004(4):5-6.
- [6] 张杰,李文莉,康凤琴,等. 一次冰雹云演变过程的卫星遥感监测与分析[J]. 高原气象,2004,23(6):758-763.
- [7] 孙玉稳,孙霞,韩洋,等. 雹云顶温度分布与地面降雹、雨的相关性观测研究[J]. 科学技术与工程,2014,14(8):120-125.
- [8] 刘小艳,刘国强,王兴菊,等. 基于卫星云参数监测产品的贵州冰雹云指标分析[J]. 中低纬山地气象,2020,44(1):10-14.