

人工智能在冰雹识别及临近预报中的初步应用^{*}

张文海¹ 李磊²ZHANG Wenhai¹ LI Lei²

1. 深圳市强风暴科学研究院, 深圳, 518057

2. 深圳市气象局, 深圳市国家气候观象台, 深圳, 518040

1. *Shenzhen Academy of Severe Storms Science, Shenzhen 518057, China*2. *Meteorological Bureau of Shenzhen Municipality/Shenzhen National Climate Observatory, Shenzhen 518040, China*

2018-04-10 收稿, 2018-07-17 改回.

张文海, 李磊. 2019. 人工智能在冰雹识别及临近预报中的初步应用. 气象学报, 77(2):282-291

Zhang Wenhai, Li Lei. 2019. A preliminary application of artificial intelligence on the detection and nowcasting of hail weather. *Acta Meteorologica Sinica*, 77(2):282-291

Abstract Based on three-dimensional mosaic reflectivity data from ten S-band Doppler radars in Guangdong Province, an artificial intelligence (AI) algorithm for automatic hail detection and nowcasting is developed using the machine learning technology. The training dataset used to develop the algorithm includes vertical and horizontal slice data of the mosaic radar reflectivity, in which the slice data of hail clouds are taken as positive samples, and other data are used as negative samples. The Bayes classifier method is used during the training process of machine learning to establish the capability of AI on recognizing the characteristics of the hail cloud. The data during the period of 2008–2013 and 2015–2016 are taken as training sets, while the observed data during the 12 hail weather processes in 2014 are used to validate the capability of AI. The result of comparative validation is encouraging, and the AI method is 9% higher than the traditional Conceptual Model method on identifying the hit rate. The current study preliminarily shows the strong capability of AI on identifying the nonlinear strong weather processes.

Key words Hail detection, Nowcasting, Artificial intelligence, Machine learning, Bayes classifier

摘 要 基于广东 10 部 S 波段多普勒天气雷达的三维拼图资料, 利用机器学习技术开发了一种冰雹识别和临近预报的人工智能算法。算法设计时以雷达回波反射率的垂直和水平扫描数据为基础训练集, 将冰雹云的雷达反射率扫描数据作为正样本, 将其他雷达反射率扫描数据作为负样本, 通过贝叶斯分类法对正、负样本数据集进行机器学习, 训练人工智能识别冰雹云内在规律的能力。训练时以广东省 2008—2013 和 2015—2016 年的数据作为训练集, 使用了 2014 年广东省 12 次冰雹过程的数据做检验。对比检验的结果表明, 人工智能法比传统的概念模型法击中率高 9 个百分点。研究结果表明了人工智能对冰雹这类非线性强天气过程具有较强的识别能力。

关键词 冰雹识别, 临近预报, 人工智能, 机器学习, 贝叶斯分类

中图法分类号 P412.25

1 引 言

冰雹灾害是由强对流天气系统引起的一种剧烈的气象灾害, 它突发性强, 并常伴随着狂风、强降水

等灾害性天气过程, 易给农业、建筑、通讯、电力、交通以及人民生命财产带来较大损失。由于冰雹过程具有突发性、局地性强以及生命期短的特点, 其预报效果长期以来都不理想, 哪怕只是提前量为半小时

^{*} 资助课题: 国家重点研发计划(2016YFC0203600)、国家自然科学基金(41575005)、中国气象局华南区域中心项目(GRMC2015M02)。

作者简介: 张文海, 主要从事短时临近预报研究。E-mail: 17350048@qq.com

通信作者: 李磊, 主要从事城市气象研究。E-mail: chonp@163.com

的临近预报,其进展也十分缓慢,这其中最大的困难就在冰雹回波的识别上。近年来,随着雷达气象学的不断发展以及其他学科先进成果的不断引进,强对流天气临近预报总体上已经取得了长足进步(俞小鼎等,2006,2012;兰红平等,2009; Sánchez, et al,2013;曹春燕等,2015)。在冰雹云识别领域也有相当的进展,其核心技术路线是对冰雹云雷达回波的三维结构特征进行总结,建立冰雹云的概念模型,并将特征用定量指标予以描述(Witt, et al,1998;王瑾等,2010;Skripniková, et al,2014),或者利用双偏振雷达的特殊性能进行判别(Smyth, et al,1999;苏德斌等,2011),这些研究成果已经在很大程度上推进了冰雹云的自动识别和临近预报。然而,由于冰雹云特征指标与强对流云雷达三维观测数据存在非线性关系,概念模型法基于阈值判断是否为冰雹云显得较为“武断”,且相同的错误会反复出现,除非人工优化修订阈值,模型本身没有自适应和自我改进能力,在实际业务应用时仍然会发生较多的空报、漏报现象。因此,在冰雹自动识别和临近预报这一科学领域仍然存在较大的提升空间。

近年来,人工智能(Artificial Intelligence, AI)在图像识别、数据挖掘及医疗领域中的应用越来越广泛,尤其是2016年AlphaGo(Silver, et al,2016)在围棋比赛中战胜职业选手,进一步引发了人们对人工智能的关注,人工智能在天气预报领域中的可用性也开始引起更为广泛的关注。事实上,20世纪80年代,一些早期的人工智能技术已经在天气预报领域得到了初步应用,但限于当时的人工智能水平,应用主要限于不同种类资料的协同综合,或在资料时、空密度不足时利用模糊逻辑进行推断等工作(Moninger, et al,1987;曾晓梅,1999)。随后,基于人工神经网络(张承福,1994)和支持向量机(陈永义等,2004;冯汉中等,2004)的机器学习方法也被用于天气预报,这些研究虽然从新的视角为天气预报提供了思路,但在实际业务工作中的应用却相对有限,尤其是对于冰雹这样高度非线性的突发性强天气过程暂时未能表现出令人信服的预报能力。本研究尝试汲取当前人工智能的研究成果,利用机器学习方法找出珠江三角洲地区冰雹云的内在规律、特征,建立智能化的冰雹云识别预报机器人,以期提高冰雹天气的临近预报、预警水平,从而为提升区域防灾、减灾能力和促进地方经济发展服务。

2 资料和方法

2.1 资料及前处理

本研究使用广东省10部S波段天气雷达(广州、深圳、韶关、珠海、清远、阳江、河源、汕尾、梅州、湛江)0.5—20 km高度20层的反射率因子拼图资料。为下一步计算方便,在雷达资料三维体扫描原始数据基础上,利用最近邻居法和垂直方向线性内插法相结合插值到三维笛卡尔坐标系中(肖艳姣等,2006),形成三维雷达回波网格数据集,该数据集为本研究下一步工作的资料基础。

为了提取冰雹云的特征向量,在三维雷达回波网格数据集基础上做进一步处理,生成几个重要的导出量,分别是雷达反射率因子(Radar Reflectivity Factor, RRF)、组合反射率(Composite Reflectivity, CR)、固定高度回波强度(Constant Altitude Plan Position Indicator, CAPPI)、柱液态水含量(Vertical Integrated Liquid, VIL)和回波顶高(Echo Top, ET)。其中,柱液态水含量定义为单位面积上垂直柱体中的总含水量,按照吴书君等(2006)的方法计算得到;回波顶高定义为高反射率核上空18.3 dBz回波的高度(肖艳姣等,2009)。

除了雷达数据外,还有冰雹观测及预警数据,主要从广东省气象业务网上获取,2008—2016年,广东省气象业务网总共记录有548个冰雹样本;除此之外,本研究根据灾情报告以及移动互联网报告检索等手段筛选出有目击冰雹观测报告9 a中总共216个样本,上述这些样本为本研究提供了重要的基础数据。其中,2008—2013年和2015—2016年的数据在本研究中用于人工智能的机器学习,而2014年的数据用于检验,因为这一年冰雹报告案例总数最多,有利于获得统计上可信的检验结果。

2.2 总体技术路线

本研究将冰雹的自动识别归结为计算机视觉机器学习问题,所需要解决的根本问题就是利用三维雷达数据在指定的区域内找出冰雹将要发生的区域——在水平方向上识别出冰雹可能出现的区域,即菜豆型的雷达回波;在垂直方向识别出冰雹的悬垂结构,综合二者形成可自动识别出冰雹区域的人工智能机器人。研究所采用的技术路线如图1所示,其总体思路如下:(1)将雷达回波资料分为正样本(即样本时次之后2 h以内有冰雹观测报告)与负

样本(即样本时次之后 2 h 以内无冰雹观测报告); (2)确定回波在水平方向上的取样范围,对于正样本为包含冰雹云回波的边长为 50 km 的正方形范围,对于负样本为包含反射率最大值周边的 50 km 正方形范围;(3)将取样范围内不同高度上的回波水平扫描数据输入人工智能分类器,同时告诉分类器每个数据所对应的是正样本还是负样本;(4)将取样范围内不同方位的回波垂直剖面数据输入人工智能分

类器,同时告诉分类器每个数据所对应的是正样本或是负样本;(5)分类器根据不同样本的两类数据(回波水平扫描和垂直扫描)进行机器学习,寻找正样本区别于负样本的内在特征规律,最终形成可以自动识别冰雹云的人工智能机器人。在形成人工智能机器人之后,用 2014 年的数据进行对照检验,确定人工智能方法的可靠性。

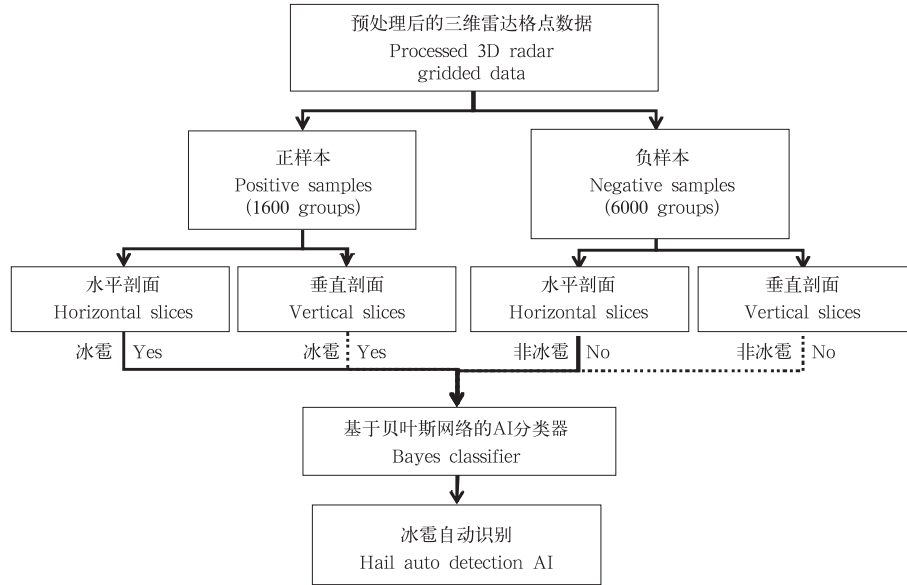


图 1 建立冰雹自动识别人工智能的技术路线

Fig. 1 The method used to establish the hail automatic recognition based on artificial intelligence

2.3 正样本提取方法

对于所有接到冰雹报告的三维雷达回波网格数据集,首先提取 CAPPI-2.5 km 的拼图进行图像识别,将冰雹云中的典型“豆状”回波或者钩状回波(胡家美等,2008)识别出来,取“豆状”回波图像质心周围边长为 50 km 区域内的数据作为正样本取样的水平范围,如图 2 所示。针对所有正样本,按照式(1)计算该样本对应的强对流指数(G),一般认为 $G > 40$ 的格点,是对流旺盛的区域(周德平等,2007),当一张图上存在连续 4 个相邻的 $G > 40$ 的格点时,这组格点群被标记为一个冰雹角点,然后按照图 3 所示的方法,以冰雹角点为轴截取三维雷达回波网格数据集的垂直扫描,每旋转 1° 取一个扫描,每个冰雹角点可获得 360 个垂直扫描,形成一组正样本。作为训练时段的 7 a 中,总共提取了 80 次冰雹过程

的数据,从每次过程的雷达回波网格数据中提取 20 组样本,故训练集共有 1600 组正样本。

$$G(x, y) = \text{Vil}(x, y) + \text{CAPPI}(x, y, 5 \text{ km}) + \text{ET}(x, y) \times 3 \quad (1)$$

2.4 负样本提取方法

如图 1 所示,负样本是小时雨量不低于 20 mm 的强降水样本(但确认未收到冰雹报告),之所以选择 20 mm 以上的强降水是因为这类降水引发的回波足够强,与冰雹云的强回波有一定可比性,但由于没有冰雹云特有的悬垂结构,对于机器学习中总结结构特征有帮助。对于负样本求出区域内 G 的最大值,以其为中心选取周边 50 km 为负样本的水平范围,截取该范围内的雷达回波水平扫描和以 G 最大值为轴的 360 个垂直扫描负样本。共有 6000 组负样本。

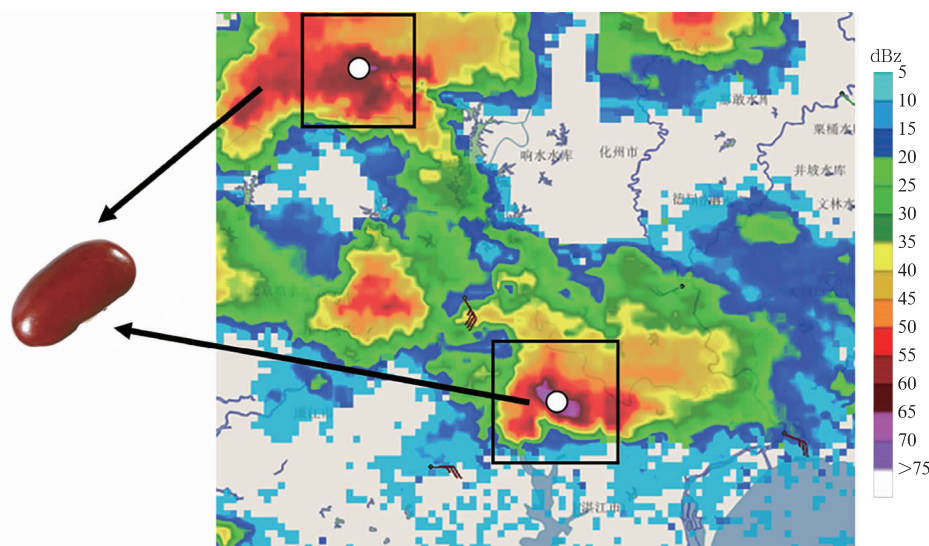


图2 根据 CAPPI-2.5 km 雷达数据确定正样本水平取样范围
Fig.2 An example of how to choose a horizontal positive sample based on the CAPPI-2.5 km radar data

2.5 机器学习过程

机器学习过程也即分类模型训练阶段,其核心任务是利用分类器根据输入数据产生冰雹判别的人工智能,但在这一过程中并不提供任何客观指标给分类器。这一过程恰如教会一个并不知道什么是老虎的人认识老虎:教师给他 10000 张照片,其中有 1000 张是老虎的照片,另外 9000 张是其他随机的动物照片,教师在每张老虎照片上标识出“这是老虎”(正样本),但并不给出老虎的具体身长、体重、斑纹颜色等客观指标,一个人会自动根据所获取的 1000 张老虎照片的视觉信息,总结形成对老虎的视觉判识标准,然后便可以判识任意图片上的动物是不是老虎。对于本阶段的研究工作而言,实际任务是计算冰雹云与非冰雹云在训练样本中的出现频率,以及每个冰雹案例的特征属性划分对出现冰雹和不出现冰雹的条件概率估计,并存储结果。这一阶段因为要处理海量数据,需要在高性能计算机上并行迭代处理。

本研究所使用的分类器是正态贝叶斯分类器(Normal Bayes Classifier, NBC),NBC 方法假设所有的特征向量之间是相互独立的,这个特点对冰雹的特征属性是适合的,比如垂直扫描的形状与水平面上回波强度并没有必然的关联。实践表明,NBC 方法在适合的场景能获得惊人的表现(朱军等, 2015)。具体而言,冰雹单体是由具有一组特征属性

的雷达回波立体网格数据组成,则每个冰雹云单体用一个 n 维特征向量来描述,包括取样范围内的 10 层水平扫描和 360 个垂直扫描($n = 370$),这个 n 维特征向量实际上隐含了很多冰雹云特征信息,如不同高度的回波强度、液态水含量、回波顶高、水平扫描形态、垂直扫描形态等。将输入样本分别记为冰雹类(C_1)和非冰雹类(C_2)(如欲更精细还可以分为小尺寸冰雹、中等尺寸冰雹、大尺寸冰雹和非冰雹,原理相同),分类器可根据输入的样本类型,统计出 n 维特征向量任一分量是冰雹云的先验概率,记为 $P(x_1|C_1), P(x_2|C_1), \dots, P(x_n|C_1)$,总体记为 $P(X|C_1)$,这些先验概率在自动识别未知回波是否为冰雹云时将起到关键作用。由此可见,人工智能经过训练获得的知识并非是像概念模型那样指标阈值固定的分类标准,而是 n 维特征向量中任一分量为冰雹云的先验概率,在判别新输入的某未知三维雷达数据时,人工智能将利用这些先验概率知识去计算它是否是冰雹的后验概率。

上述人工智能算法的实现采用了 OpenCV 的计算机视觉库,它是一个开源的计算机视觉库,为图像处理、模式识别、三维重建、物体跟踪、机器学习提供各种基础算法。本研究基于雷达回波的冰雹识别综合运用了这几方面的知识和算法,包括上文所述的 NBC 即为 OpenCV ML 库中的一种算法。具体处理流程为:(1)将事先准备的雷达数据集转换为

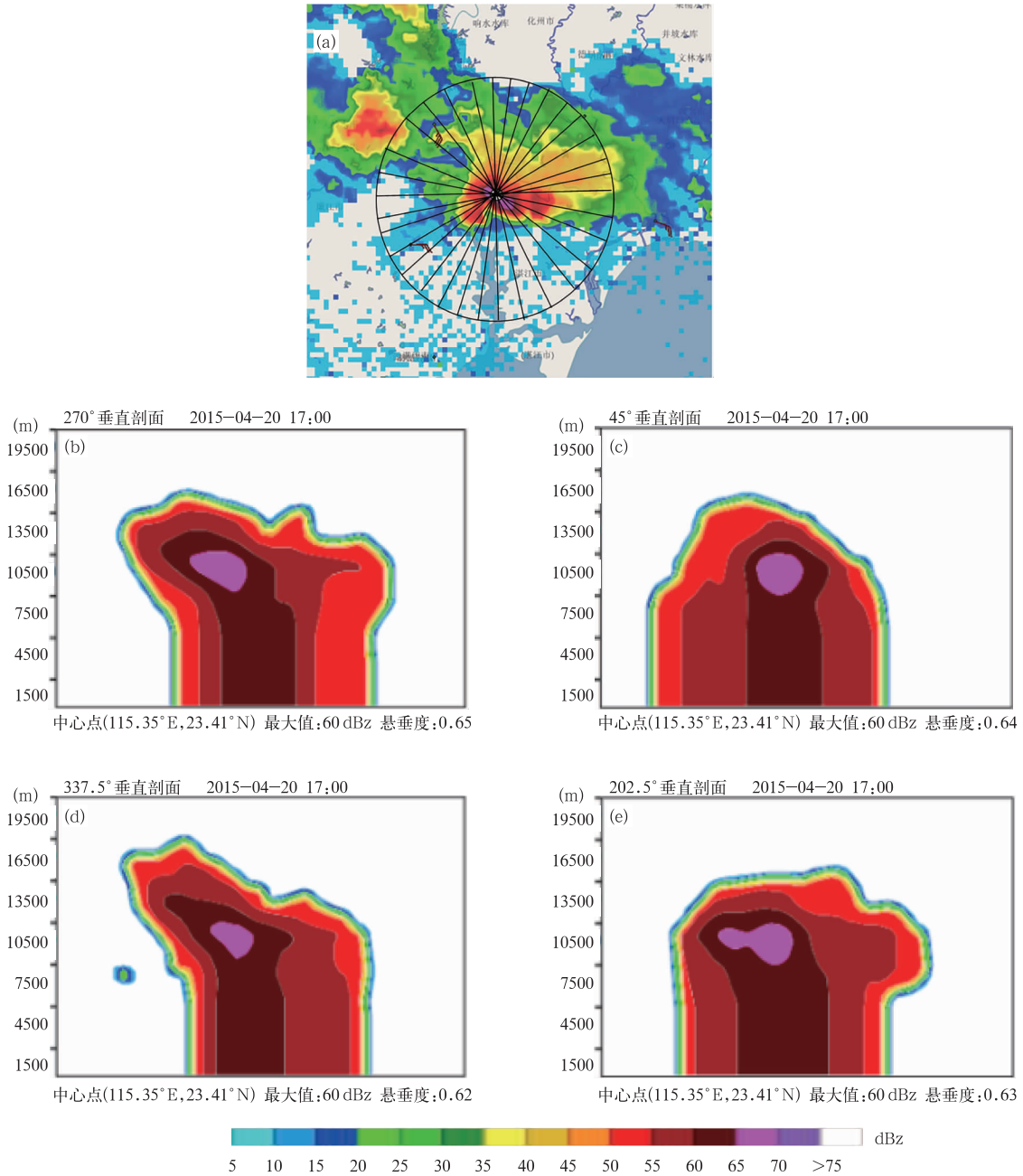


图3 提取正样本垂直剖面

Fig. 3 An example of how to choose a group of vertical positive samples

NBC 训练函数支持的数据格式;(2)调用 train 函数进行训练(训练过程需要很大计算量,持续时间较长);(3)训练结束后,调用 save 函数将训练模型存储为 XML 或者 YAML;(4)模型检验或实际运行阶段,调用 load 函数装入 XML 或者 YAML 格式的训练模型,随后将雷达测试数据集或实时数据集转换为预测函数特定格式,调用 predict 函数进行冰雹识别输出。

2.6 AI 识别流程

经过训练获得 $P(X|C_1)$ 先验概率的分类器,在获得一个新的未知三维回波网格数据时,将采用图 4 所示步骤进行识别。其中,前 3 步与机器学习阶段类似,需要根据 CAPPI-2.5 km 的水平分布寻找强回波区,然后进行水平和垂直扫描,再输入 AI 分类器中进行判断。

分类器会根据输入的 n 维变量(记为 $X = \{x_1,$

$x_2, \dots, x_n\}$), 判断是冰雹和非冰雹的概率, 也即 $P(C_1|X)$ 和 $P(C_2|X)$ 。贝叶斯定理如式(2), 它意味着要计算后验概率 $P(C_1|X)$ 可转化为计算最大化先验概率 $P((x_1, x_2, \dots, x_n) | C_1) \times P(C_1)$, 而先验概率 $P(x_1|C_1), P(x_2|C_1), \dots, P(x_n|C_1)$ 已从训练数据集先期计算得到。

$$P(C_1 | X) = \frac{P((x_1, x_2, \dots, x_n) | C_1) \times P(C_1)}{P(x_1, x_2, \dots, x_n)} \quad (2)$$

不难看出, 这种根据 n 维特征向量计算后验概率的方式, 比单纯依靠概念模型进行武断识别的方法要好得多, 这是因为人工智能分类器会综合考虑 n 维特征向量的先验概率, 并且在计算 $P(C_1|X)$ 时

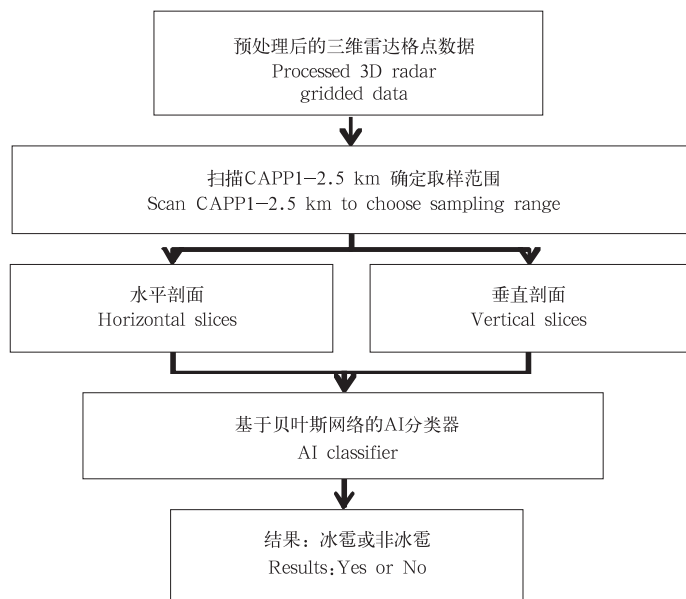


图4 对未知回波数据的分类流程

Fig. 4 The flowchart of classifying unknown echo data

会不断根据优先级较高的特征分量(例如先计算水平扫描是冰雹云的概率), 调整低优先级特征分量是冰雹云的概率, 可以很好地解决模棱两可的冰雹类型。

3 结果验证

3.1 统计对比

在验证基于人工智能的冰雹云自动识别机器人时, 选取了2014年12个冰雹天气案例, 每个案例前后选取了10个时次的雷达回波数据, 共240个正样本(表1), 分别利用机器学习法和概念模型法进行冰雹云识别定位; 同时选取了700个没有冰雹但小时雨量不低于20 mm的其他过程或弱回波过程作为负样本。利用击中率(POD)、空报率(FAR)、临界成功指数(CSI)来评价识别效果。将冰雹识别结果的位置提取出来, 结合冰雹报告的位置、冰雹大小做评分检验。由于冰雹落区是通过媒体获取, 定位描述有一定偏差, 规定如果某位置上预报出现冰雹, 而实际观测或报告也出现冰雹, 位置偏差在10 km

以内, 则判断为该点成功; 如果某位置上预报没有冰雹, 实况出现冰雹, 判断该点漏报; 如果某位置识别出现冰雹, 实况没有出现冰雹, 判断该点空报。击中率(POD)、空报率(FAR)和临界成功指数(CSI)的计算方法如式(3)~(5)

表1 检验所用的2014年广东12个冰雹过程案例

Table 1 The 12 cases used to validate the established method

案例日期及时段	地区
2014年8月1日16—17时	深圳
2014年8月1日16—17时	东莞
2014年3月30日20—21时	深圳
2014年3月30日19—20时	香港
2014年3月30日06—07时	广州
2014年3月30日06—07时	佛山
2014年3月30日11—12时	云浮
2014年7月23日18—19时	韶关
2014年4月1日10—11时	惠州
2014年4月3日00—01时	惠州
2014年8月1日17—18时	惠州
2014年4月2日02—03时	开平

$$POD = \frac{n_{成功}}{n_{成功} + n_{漏报}} \quad (3)$$

$$FAR = \frac{n_{空报}}{n_{成功} + n_{空报}} \quad (4)$$

$$CSI = \frac{n_{成功}}{n_{成功} + n_{漏报} + n_{空报}} \quad (5)$$

检验结果表明:人工智能对于冰雹自动识别的击中率为 86%,而概念模型法击中率仅为 77%,人工智能的击中率较传统概念模型法提高了 9 个百分点;与此同时,人工智能的空报率降低了 9 个百分

点,而临界成功指数提升了 14 个百分点。对比分析表明,人工智能在冰雹自动识别上展示出令人鼓舞的结果。

3.2 个例分析

为直观地说明人工智能与概念模型法的区别,筛选了 2 个案例用 2 种方法进行了回算,并对结果进行对比分析。

(1) 2014 年 8 月 1 日过程

如图 5a 所示,2014 年 8 月 1 日 17—18 时(北京时,下同)的 1 h 内,分散的强对流云团不断在惠州、

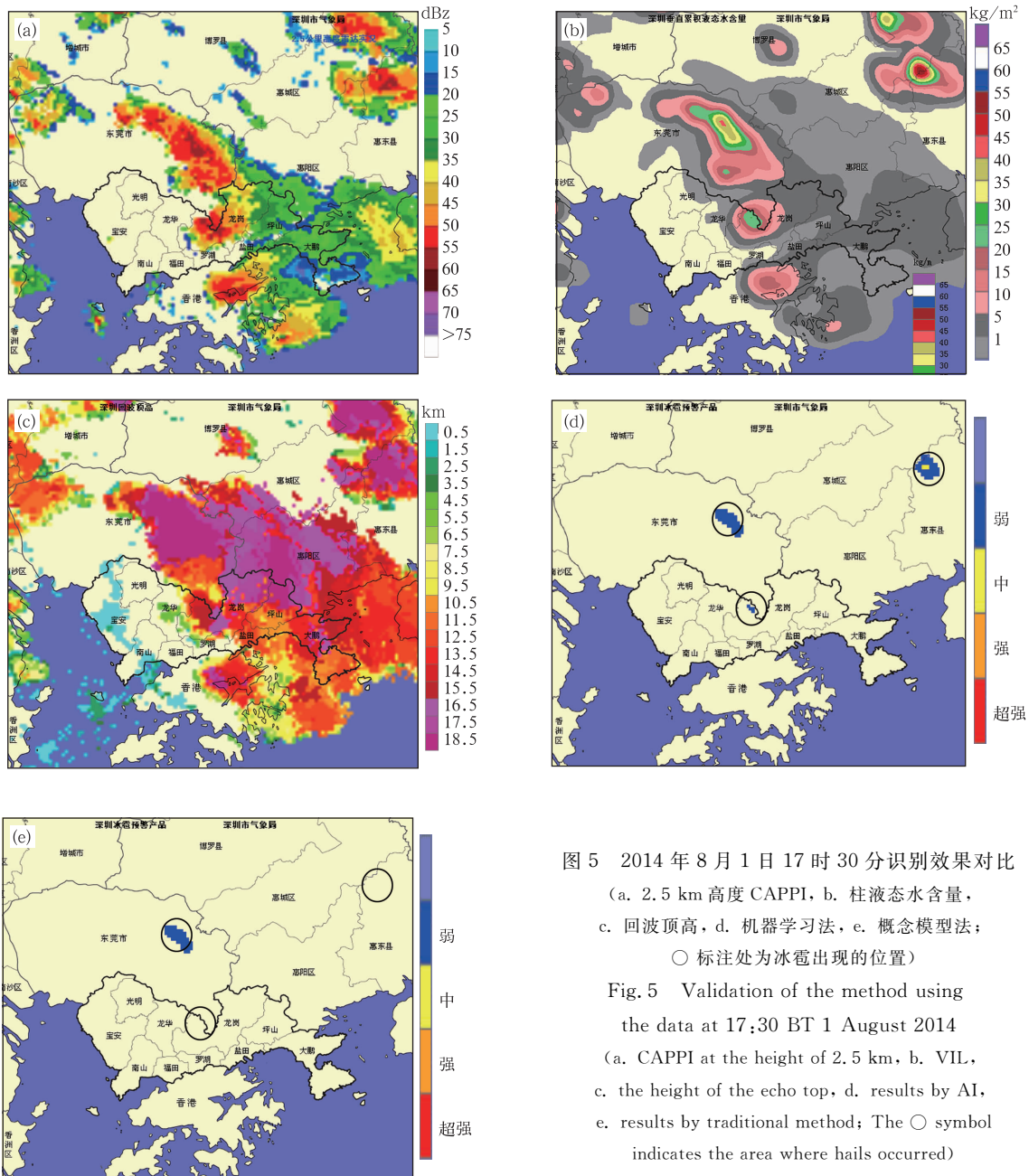


图 5 2014 年 8 月 1 日 17 时 30 分识别效果对比

(a. 2.5 km 高度 CAPPI, b. 柱液态水含量, c. 回波顶高, d. 机器学习法, e. 概念模型法; ○ 标注处为冰雹出现的位置)

Fig. 5 Validation of the method using the data at 17:30 BT 1 August 2014

(a. CAPPI at the height of 2.5 km, b. VIL, c. the height of the echo top, d. results by AI, e. results by traditional method; The ○ symbol indicates the area where hails occurred)

东莞、深圳移动增强,并在惠州惠东县、东莞市、深圳龙岗区记录有降雹发生。选取当日 17 时 30 分的识别结果,人工智能识别的落区与媒体提供的目击报告落区分布接近,没有明显的漏报和空报(图 5d 中蓝色区域);概念模型仅在东莞区域有较大面积冰雹识别(图 5e),在惠州和深圳境内的降雹没有有效识别。根据回波强度、回波顶高、柱液态水含量及回波剖面等数据分析表明,位于惠州和深圳龙岗漏报区域的强回波悬垂特征数据接近但未达到判识阈值,概念模型因而识别失败,从而造成漏报。由此可知,

人工智能可以动态解决概念模型单一阈值设置造成的漏报问题,而无需改动程序。随着案例的逐步累积,人工智能识别的效果还会有进一步的提高,而概念模型法则停在原有水平,并且同样的错误会反复出现。

(2) 2016 年 7 月 30 日过程

如图 6a 所示,2016 年 7 月 30 日 17 时,一股强对流云团影响广州及顺德部分地区,并形成了冰雹过程。深圳和广州雷达完整地记录到了这次冰雹的演变过程。机器学习法对本次强对流过程进行了每

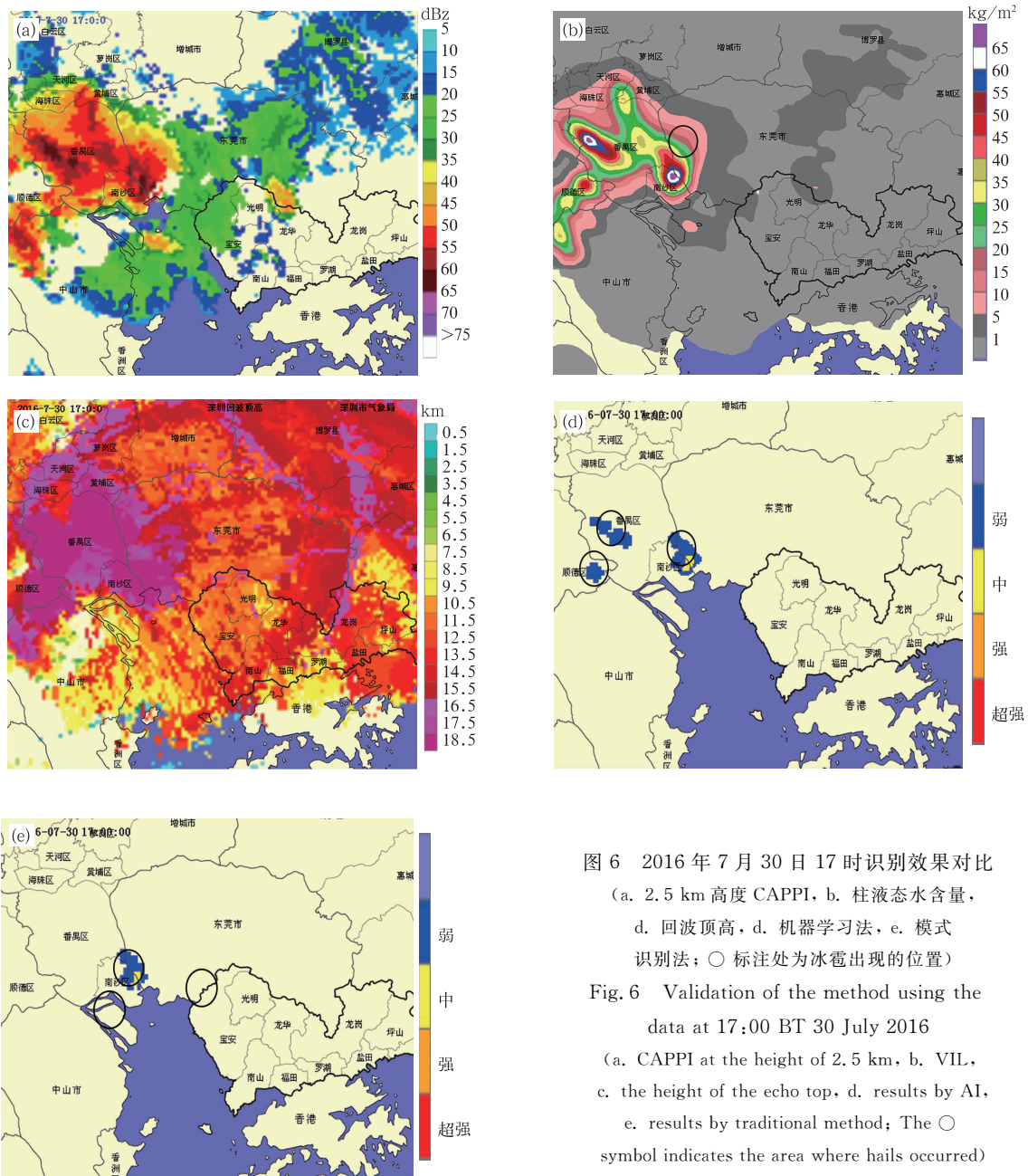


图 6 2016 年 7 月 30 日 17 时识别效果对比

(a. 2.5 km 高度 CAPPI, b. 柱液态水含量, d. 回波顶高, d. 机器学习法, e. 模式识别法; ○ 标注处为冰雹出现的位置)

Fig. 6 Validation of the method using the data at 17:00 BT 30 July 2016

(a. CAPPI at the height of 2.5 km, b. VIL, c. the height of the echo top, d. results by AI, e. results by traditional method; The ○ symbol indicates the area where hails occurred)

隔 6 min 一次的冰雹监测。选取当日 17 时的识别结果,对这次降雹过程的落区、强度等进行对比试验。从过程识别结果和实况的对比可知,人工智能识别的落区有 3 处,即南沙、番禺、顺德,与实况报告十分接近,而概念模型法识别由于阈值设置偏大,导致只识别出南沙 1 处冰雹落区,番禺和顺德存在漏报。

为进一步理解人工智能得到的判别知识相较于概念模型法的区别,对判别的中间环节做了具体分析,发现:在悬垂回波分类方面,人工确认的悬垂结构,人工智能与概念模型均可确认;在菜豆型回波的分类方面,人工智能给出的判别相对于人工判别更为宽松,约 10% 判定结果与人工确认以及概念模型法的判定结果不符。考虑到水平扫描(豆形回波是通过水平扫描识别的)本身是 n 维向量中优先级较高的分量,故这可能是人工智能识别表现优于概念模型的重要原因。近几年的个例分析证明,该方法的运用可以有效避免缺乏经验的预报员犯低级错误,提高整体预报水平。

4 结论与展望

文中介绍了人工智能技术在珠三角地区冰雹自动识别和临近预报中的初步应用,取得了令人鼓舞的效果。基于机器学习的人工智能在自动识别冰雹云方面展现出了较强的能力,在本研究案例中其识别能力的击中率比传统的概念模型法提高了 9 个百分点。研究表明,在大数据的训练过程中,当输入特征向量数据与目标存在非线性关系时,机器学习能非常好地训练出大规模特征和目标结果存在的隐含关系,而强天气过程通常是高度复杂和非线性的。因此,可以根据输入特征向量进行自我适应的人工智能优于需要人为设置固定阈值的概念模型法,展现出了人工智能在天气预报中的良好前景。

尽管如此,在本领域仍然有可以继续深入的地方,尤其是当前三维雷达数据预处理,对冰雹云特征提取至关重要,直接决定训练集和测试集数据的质量,今后的提升工作需要在这方面继续努力,寻找新的雷达导出变量来更加准确地刻画冰雹云的特征。

需要特别指出的是,特征向量的选取是开展人工智能自动识别冰雹云研究的关键,而本研究对于特征向量的选取主要基于已有业务工作的经验,还有进一步优化和提升的空间,具体包括:(1)利用试

验来评估特征向量的选择对识别效果的影响。比如,增加最大反射率因子、 -20°C 等温线对应等高面上反射率因子、冰雹融化层高度等要素作为特征量,重新制作训练集和测试集,建立改进后的识别模型(俞小鼎等,2006,2014)。这些可能的特征向量有的可以直接从三维雷达数据中提取,有的还需通过建立要素与雷达数据的关系间接提取。(2)特征量阈值评估试验,实现回波顶高和对冰雹识别的敏感性分析,选择 30 dBz 作为回波顶阈值,与目前的 18.3 dBz 阈值进行对比,进一步提高模型对冰雹的敏感性。(3)目前数据集的制作方法是多部雷达先拼图后,在统一的三维坐标空间上进行冰雹云识别,优点是识别算法相对简单、速度快,缺点是雷达拼图经过内插后精度降低;今后将尝试应用单部雷达资料制作测试集和训练集,建立基于单部雷达的冰雹云识别模型;在实时运行中,将根据单部雷达确定的降雹区域在统一的坐标空间中进行拼接,形成整个广东省区域内降雹区临近预报。

冰雹个例库的建设是另一项可深入开展的工作,数据集的建设是人工智能开发工作的核心,其工作量占比在一半以上,而冰雹个例库是建立训练集和测试集的重要且唯一依据,因此,建成一个相对完整的冰雹个例库对这项工作非常重要。由于冰雹缺乏客观的观测手段,且最大的冰雹往往不是出现在站点上,需要通过气象局灾情直报系统和互联网上相关报道结合雷达回波确认冰雹个例,今后将充分利用这两种途径逐步完善冰雹个例库,为模型训练奠定更加坚实的基础。

参考文献

- 曹春燕, 陈元昭, 刘东华等. 2015. 光流法及其在临近预报中的应用. 气象学报, 73(3): 471-480. Cao C Y, Chen Y Z, Liu D H, et al. 2015. The optical flow method and its application to now-casting. Acta Meteor Sinica, 73(3): 471-480 (in Chinese)
- 陈永义, 俞小鼎, 高学浩等. 2004. 处理非线性分类和回归问题的一种新方法(I): 支持向量机方法简介. 应用气象学报, 15(3): 345-354. Chen Y Y, Yu X D, Gao X H, et al. 2004. A new method for non-linear classify and non-linear regression I: Introduction to support vector machine. J Appl Meteor Sci, 15(3): 345-354 (in Chinese)
- 冯汉中, 陈永义. 2004. 处理非线性分类和回归问题的一种新方法(II): 支持向量机方法在天气预报中的应用. 应用气象学报, 15(3): 355-365. Feng H Z, Chen Y Y. 2004. A new method for non-linear classify and non-linear regression II: Application

- of support vector machine to weather forecast. *J Appl Meteor Sci*, 15(3): 355-365 (in Chinese)
- 胡家美, 刘峰, 黄奕铭等. 2008. 冰雹和低空风切变天气的多普勒雷达回波特征. *广东气象*, 30(1): 24-28. Hu J M, Liu F, Huang Y M, et al. 2008. Signatures of hail and low level shear from Doppler radar products. *Guangdong Qixiang*, 30(1): 24-28 (in Chinese)
- 兰红平, 孙向明, 梁碧玲等. 2009. 雷暴云团自动识别和边界相关追踪技术研究. *气象*, 35(7): 101-111. Lan H P, Sun X M, Liang B L, et al. 2009. An automatic tracking and recognition algorithm for thunderstorm cloud-cluster (TRACER). *Meteor Mon*, 35(7): 101-111 (in Chinese)
- 苏德斌, 马建立, 张蔷等. 2011. X波段双线偏振雷达冰雹识别初步研究. *气象*, 37(10): 1228-1232. Su D B, Ma J L, Zhang Q, et al. 2011. Preliminary research on method of hail detection with X band dual linear polarization radar. *Meteor Mon*, 37(10): 1228-1232 (in Chinese)
- 王瑾, 李明元, 汪华. 2010. 基于多雷达三维插值格点强冰雹诊断因子的强冰雹识别方法. *高原气象*, 29(6): 1533-1545. Wang J, Li M Y, Wang H. 2010. A new hail detection algorithm of hail diagnostic parameters based on 3D mosaicing data of multi-radar. *Plateau Meteor*, 29(6): 1533-1545 (in Chinese)
- 吴书君, 王凤娇, 刘昭武等. 2006. 两种垂直累积液态含水量产品的应用对比分析. *气象*, 32(10): 34-40. Wu S J, Wang F J, Liu Z W, et al. 2006. Applied comparison of two products of vertically integrated liquid water content. *Meteor Mon*, 32(10): 34-40 (in Chinese)
- 肖艳姣, 刘黎平. 2006. 新一代天气雷达网资料的三维格点化及拼图方法研究. *气象学报*, 64(5): 647-657. Xiao Y J, Liu L P. 2006. Study of methods for interpolating data from weather radar network to 3-D grid and mosaics. *Acta Meteor Sinica*, 64(5): 647-657 (in Chinese)
- 肖艳姣, 马中元, 李中华. 2009. 改进的雷达回波顶高、垂直积分液态含水量及其密度算法. *暴雨灾害*, 28(3): 210-214. Xiao Y J, Ma Z Y, Li Z H. 2009. Advanced algorithms of ET, VIL and VIL density for the CINRAD. *Torr Rain Dis*, 28(3): 210-214 (in Chinese)
- 俞小鼎, 姚秀萍, 熊廷南等. 2006. 多普勒天气雷达原理与业务应用. 北京: 气象出版社, 314pp. Yu X D, Yao X P, Xiong T N, et al. 2006. The Principle and Operational Application of Doppler Weather Radar. Beijing: China Meteorological Press, 314pp (in Chinese)
- 俞小鼎, 周小刚, 王秀明. 2012. 雷暴与强对流临近天气预报技术进展. *气象学报*, 70(3): 311-337. Yu X D, Zhou X G, Wang X M. 2012. The advances in the nowcasting techniques on thunderstorms and severe convection. *Acta Meteor Sinica*, 70(3): 311-337 (in Chinese)
- 俞小鼎. 2014. 关于冰雹的融化层高度. *气象*, 40(6): 649-654. Yu X D. 2014. A note on the melting level of hail. *Meteor Mon*, 40(6): 649-654 (in Chinese)
- 曾晓梅. 1999. 国外人工智能技术在天气预报中的应用综述. *气象科技*, (1): 4-10. Zeng X M. 1999. A review on the application of artificial intelligence in weather forecast abroad. *Meteor Sci Technol*, (1): 4-10 (in Chinese)
- 张承福. 1994. 人工神经网络在天气预报中的应用研究. *气象*, 20(6): 43-47. Zhang C F. 1994. On the application of artificial neural network for weather forecast. *Meteor Mon*, 20(6): 43-47 (in Chinese)
- 周德平, 杨洋, 王吉宏等. 2007. 冰雹云雷达识别方法及防雹作业经验. *气象科技*, 35(2): 258-263. Zhou D P, Yang Y, Wang J H, et al. 2007. Hail cloud identification method by radar echoes and experiences from hail suppression operation in Liaoning province. *Meteor Sci Technol*, 35(2): 258-263 (in Chinese)
- 朱军, 胡文波. 2015. 贝叶斯机器学习前沿进展综述. *计算机研究与发展*, 52(1): 16-26. Zhu J, Hu W B. 2015. Recent advances in Bayesian machine learning. *J Comp Res Dev*, 52(1): 16-26 (in Chinese)
- Moninger W R, Davis J, Dyer R, et al. 1987. Summary of the first conference on the artificial intelligence research in environmental science (AIRIES). *Bull Amer Meteor Soc*, 68: 793-800
- Sánchez J L, López L, García-Ortega E, et al. 2013. Nowcasting of kinetic energy of hail precipitation using radar. *Atmos Res*, 123: 48-60
- Silver D, Huang A, Maddison C J, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587): 484-489
- Skripniková K, Řezáčová D. 2014. Radar-based hail detection. *Atmos Res*, 144: 175-185
- Smyth T J, Blackman T M, Illingworth A J. 1999. Observations of oblate hail using dual polarization radar and implications for hail-detection schemes. *Quart J Roy Meteor Soc*, 125(555): 993-1016
- Witt A, Eilts M D, Stumpf G J, et al. 1998. An enhanced hail detection algorithm for the WSR-88D. *Wea Forecasting*, 13(2): 286-303