

一种基于离散属性 BNT 的双偏振气象 雷达降水粒子分类方法

李 海¹, 尚金雷¹, 孙婷逸¹, 冯 青¹, 庄子波²

(1. 中国民航大学天津市智能信号与图像处理重点实验室, 天津 300300; 2. 中国民航大学飞行技术学院, 天津 300300)

摘 要: 针对传统降水粒子分类算法存在的过度依赖专家经验和模型预设误差问题, 本文提出了一种基于离散属性贝叶斯网络 (Bayesian NeTwork, BNT) 的双偏振气象雷达降水粒子分类 (Hydrometeor Classification, HC) 方法. 首先对双偏振气象雷达获取的偏振参量取值进行离散化处理生成离散化标准, 并根据离散化标准制作训练数据集; 然后使用训练数据集对贝叶斯网络进行结构学习学得贝叶斯网络结构, 以及参数学习学得与贝叶斯网络结构匹配的条件概率表; 最后加入附加信息计算出每种降水粒子类先验概率, 与贝叶斯网络结构和条件概率表共同组成贝叶斯网络分类器. 训练好的贝叶斯网络分类器根据最大后验概率准则完成对测试数据的降水粒子分类, 与模糊逻辑算法对比评价结果. 实验证明: 该方法能有效区分不同的降水粒子得到准确的降水粒子分类结果.

关键词: 双偏振雷达; 降水粒子分类; 贝叶斯网络; 贝叶斯分类; 不确定知识; 离散化; BNT

中图分类号: TN959.4 **文献标识码:** A **文章编号:** 0372-2112 (2021)03-0619-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20191217

A BNT Hydrometeor Classification Algorithm for Dual-Polarization Radar

LI Hai¹, SHANG Jin-lei¹, SUN Ting-yi¹, FENG Qing¹, ZHUANG Zi-bo²

(1. Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China, Tianjin 300300, China;

2. Flight Technology College, Civil Aviation University of China, Tianjin 300300, China)

Abstract: The over-reliance on expert experience and model preset errors in traditional precipitation particle classification algorithms are discussed. This paper proposes a dual-polarization hydrometeor classification (HC) method based on discrete attribute Bayesian NeTwork (BNT). Firstly, the value of polarization parameters obtained by the dual-polarization meteorological radar is discretized to generate a discretization standard, and the training data set is made according to the discretization standard. Then the training data set is used to learn the structure of the Bayesian network and the conditional probability table matching the structure of the Bayesian network. At last, additional information is added to calculate the prior probability of each precipitation particle class, and the Bayesian network classifier is composed of Bayesian network structure and conditional probability table. The trained Bayesian network classifier classifies the precipitation particles according to the maximum posterior probability criterion and compares the evaluation results with the fuzzy logic algorithm. Experiments show that this method can effectively distinguish different precipitation particles.

Key words: dual polarization radar; hydrometeor classification; Bayesian networks; Bayesian classification; uncertain knowledge; discretization; BNT (Bayesian NeTwork)

1 引言

双偏振气象雷达作为一种新型天气信号回波探测

工具^[1], 能交替发射和接收水平与垂直偏振方向的信号, 通过对回波数据进行处理, 可以提取出有关降水粒子的大小、形状、空间取向及相态等信息, 与传统气象雷

达相比,在定量估测降水、粒子相态识别、预测灾害性天气等方面具有明显优势^[2],在天气预报、防洪减灾、航空危险预警、航路规划以及人工影响天气等方面有重要应用价值^[3].

对双偏振气象雷达降水粒子分类的研究始于1996年,根据不同类型的降水粒子的偏振参量具有不同取值范围确定相应的隶属度函数(Membership Function, MF),首次将模糊逻辑算法(Fuzzy Logic Algorithms, FLA)应用到降水粒子分类中来.借助于模糊逻辑对不确定性问题的准确描述,成功实现了降水粒子的分类^[4].在此之后对模糊逻辑算法的研究大部分专注于对隶属度函数(Membership Function, MF)的改进^[5]、对特定降水粒子的识别^[6]以及提高模糊逻辑算法的鲁棒性^[7].2018年中国科学院大气物理研究所肖辉^[8]等利用模糊逻辑算法对2015年8月7日北京一次较大范围的降雹进行水凝物粒子分类识别结果与真实情况一致.模糊逻辑算法成为主要的双偏振气象雷达降水粒子分类算法,但其也存在缺陷:过度依靠专家经验,模糊逻辑算法中每一种降水粒子对应每一个雷达偏振参量都需要定义一个梯形、Beta、高斯或半高斯隶属度函数,而每个隶属度函数的参数及相关权重的确定都需要依赖专家经验.另外隶属度函数本身形状是预定的,因此可能会对分类结果带来误差.不仅如此,传统的贝叶斯分类方法也存在概率密度函数模型预设误差问题^[9].

本文提出了一种基于离散属性贝叶斯网络的双偏振气象雷达降水粒子分类算法.首先对雷达偏振参量的取值进行离散化处理生成离散化标准,然后通过构建一种离散属性贝叶斯网络(BNT)分类器实现降水粒子的分类.其中构建BNT分类器包括结构学习、参数学习和计算类先验概率三部分:根据离散化标准生成训练数据集,使用训练数据集对贝叶斯网络进行结构学习学得贝叶斯网络结构;确定结构后继续进行参数学习求出网络结构中每个节点的条件概率表;使用包含融化层信息、异常值信息以及场景影响信息作为附加信息确定贝叶斯网络类先验概率,完成BNT分类器的构建.根据贝叶斯网络结构、条件概率表以及类先验概率计算不同降水粒子的后验概率,依据最大后验概率准则实现降水粒子分类.本文方法使用条件概率表存储概率信息,无需进行概率模型或隶属度函数模型假设就可以表示概率空间中任意分布的降水粒子.

2 数据预处理:雷达偏振参量取值的离散化处理

选择雷达偏振参量包括雷达反射率因子(Z_H)、差分反射率因子(Z_{DR})、互相关系数(ρ_{HV})以及差分相移率(K_{DP})作为降水粒子分类的属性依据.构建离散属性

贝叶斯网络分类器前需要进行雷达偏振参量取值离散化处理,雷达偏振参量的离散化处理即找到适宜的断点集来表示原本连续取值的雷达偏振参量,采取基于粗糙集信息熵的离散化算法^[10]对雷达偏振参量取值进行离散化处理,该算法由模糊逻辑分类算法提供的降水粒子样本集合进行雷达偏振参量取值离散化.共分9类降水粒子(地杂波也认为是一种降水粒子类别)如下表所示:

表1 降水粒子输出类别标签数据集

类标签(C)	降水类型	英文名称	简记
$C=1$	地杂波	Ground Clutter	GC
$C=2$	冰晶	Ice Crystals	IC
$C=3$	干雪	Dry Snow	DS
$C=4$	湿雪	Wet Snow	WS
$C=5$	雨	Rain	RA
$C=6$	暴雨	Heavy Rain	HR
$C=7$	大雨滴	Big Drop Rain	BD
$C=8$	霰	Graupel	GR
$C=9$	冰雹	Hail	HA

基于粗糙集信息熵的离散化算法使用决策表描述由雷达偏振参量数据与对应降水粒子类标签组成的粗糙集,通过计算最小信息熵求得雷达偏振参量离散化结果.决策表示意图如表2所示.

表2 决策表示意图

论域	条件属性				决策属性
	Z_H	Z_{DR}	ρ_{HV}	K_{DP}	
U					C
u_1	$u_{1,1}$	$u_{1,2}$	$u_{1,3}$	$u_{1,4}$	$C_1 (C_1 \in \{1, \dots, 9\})$
u_2	$u_{2,1}$	$u_{2,2}$	$u_{2,3}$	$u_{2,4}$	$C_2 (C_2 \in \{1, \dots, 9\})$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
u_N	$u_{N,1}$	$u_{N,2}$	$u_{N,3}$	$u_{N,4}$	$C_N (C_N \in \{1, \dots, 9\})$

其中 N 为论域 $U = \{u_1, u_2, \dots, u_N\}$ 中的样本个数,表2中样本的决策属性为 $j(j \in \{1, 2, \dots, 9\})$ 的样本个数记为 $G(j)$,则该决策表的信息熵如式(1).

$$H(U) = - \sum_{j=1}^9 q \log_2 q, \quad \text{其中 } q = \frac{G(j)}{N} \quad (1)$$

下面以条件属性 Z_H 为例介绍离散化算法:首先对 U 中的样本按 Z_H 取值大小进行排序和去重处理,生成集合 U' ,初始化候选断点集合 D_{hx} 为 U' 中相邻元素的中点,初始化断点集合 $D = \emptyset$,设 U 中条件属性 Z_H 的值为 $R(R \in [l_z, r_z])$,候选断点集合 D_{hx} 中的某一候选断点 $d_{hx}(d_{hx} \in D_{hx})$ 将 R 分为 $[l_z, d_{hx}) \cup [d_{hx}, r_z]$,论域 U 中样本条件属性 Z_H 取值属于 $[l_z, d_{hx})$ 的样本集记为 U_l ,论域 U 中样本条件属性 Z_H 取值属于 $[d_{hx}, r_z]$ 的样

本集记为 U_r , 用 $G(U_l)$ 和 $G(U_r)$ 分别表示集合 U_l 和 U_r 的样本个数, 则加入候选断点 d_{hx} 后 U 的信息熵计算方式如式(2).

$$H^d(U) = \frac{G(U_l)}{N}H(U_l) + \frac{G(U_r)}{N}H(U_r) \quad (2)$$

设 U 中样本的条件属性 Z_H 小于 d_{hx} 且决策属性为 $j_l(j_l \in \{1, 2, \dots, 9\})$ 的样本个数记为 $G(j_l)$, 条件属性 Z_H 大于等于 d_{hx} 且决策属性为 $j_r(j_r \in \{1, 2, \dots, 9\})$ 的样本个数记为 $G(j_r)$. 那么式(2)中 $H(U_l)$ 和 $H(U_r)$ 计算方法如下式(3).

$$H(U_l) = - \sum_{j_l=1}^9 q_l \log_2 q_l, \quad \text{其中 } q_l = \frac{G(j_l)}{G(U_l)}$$

$$H(U_r) = - \sum_{j_r=1}^9 q_r \log_2 q_r, \quad \text{其中 } q_r = \frac{G(j_r)}{G(U_r)} \quad (3)$$

根据式(2)计算集合 D_{hx} 中所有候选断点针对 U 的信息熵, 选取具有最小信息熵的候选断点 d_{hx}^{min} 作为结果断点加入到集合 D 中, 并从 D_{hx} 中去掉 d_{hx}^{min} . 完成了一个断点的选取工作.

迭代计算更新后 D_{hx} 中候选断点针对论域 U 的信息熵, 选出具有最小信息熵的候选断点加入到集合 $D = \{d_1, d_2, \dots, d_a\}$ 中. 当迭代第 a 次时, D 将 ZY 划分为 $[l_z, d_1] \cup [d_1, d_2] \cup \dots \cup [d_a, r_z]$ 记为 U_1, U_2, \dots, U_{a+1} , 根据式(2), 则这时候候选断点 d_{hx}^a 针对 U 的信息熵计算变为下式:

$$H^d(U) = H^d(U_1) + H^d(U_2) + \dots + H^d(U_{a+1}) \quad (4)$$

每次迭代后验证决策表相容性作为迭代是否终止的条件. 如果决策表相容则结束, 如果决策表不相容, 则继续迭代直到整个决策表相容为止. 最终输出集合 D 称为离散化标准, 作为后续训练数据集和测试数据集的离散化依据.

3 基于离散属性贝叶斯网络降水粒子分类算法

基于 BNT 的双偏振气象雷达降水粒子分类算法首先根据离散化标准对训练数据集进行离散化; 其次使用按离散化标准处理后的训练数据集进行贝叶斯网络学习学得贝叶斯网络结构和条件概率表, 引入附加信息确定不同降水粒子的类先验概率, 与贝叶斯网络结构和条件概率表共同构成贝叶斯网络分类器. 最后依据最大后验概率准则完成降水粒子分类. 下面进行详细阐述:

3.1 贝叶斯网络学习

贝叶斯网络学习^[11]包括结构学习学得有向无环图(Directed Acyclic Graph, DAG)和参数学习学得条件概率表(Conditional Probability Table, CPT).

BNT 结构学习按照最小欧式距离准则, 将雷达采集的实测数据按照离散化标准进行离散化制作训练数

据集. 贝叶斯网络分类器可以理解为朴素贝叶斯分类器(Naive Bayes Classifier, NBC)的网络依赖扩展, 允许属性节点(即雷达偏振参量属性节点)之间形成任意有向图. 朴素贝叶斯分类器认为每个属性节点仅以类节点作为父节点, 不同属性节点之间相互独立.

结构学习即在朴素贝叶斯结构基础上继续确定各属性节点之间的依赖关系, 基本思想是利用信息论的方法定量的分析偏振参量属性节点间的依赖关系求得表述偏振参量属性节点关系的网络结构. 首先计算两属性节点间的互信息, 选取大于门限值 $e(e = 0.25)$ 的节点对之间建立无向边. 然后对已建立的无向边定向, 定向原则为由与类节点 C 互信息值大的属性节点指向与类节点 C 互信息值小的属性节点. 如果求得的贝叶斯网络中存在成环的部分, 则去掉环中互信息最小的边, 最终得到一个有向无环图, 完成贝叶斯网络结构学习. 学得贝叶斯网络结构如图 1 所示.

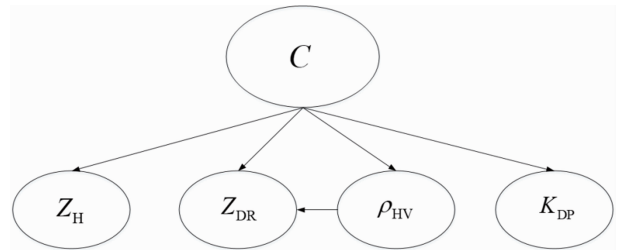


图1 贝叶斯网络结构

BNT 参数学习的目的是学得贝叶斯网络结构中属性节点的条件概率表. 通过概率统计的方式确定每个属性节点取值情况的概率.

3.2 引入附加信息确定类先验概率

在没有任何附加信息的情况下, 各降水粒子类先验概率相同. 传统降水粒子分类算法提出引入融化层信息、温度信息以及异常值信息中一种或几种作为附加信息^[12]进行分类优化. BNT 方法将附加信息加入类先验概率中, 利用类先验概率可以为 0 的特点, 当某降水粒子类先验概率为 0 时, 认为分类器的输出结果不包括该类降水粒子. 由此达到针对每一个雷达回波单元构建不同分类器的目的, 更好地完成降水粒子分类. 确定 BNT 类先验概率的具体描述如下:

3.2.1 融化层信息

如图 2^[13]根据雷达主瓣波束宽度和雷达仰角信息将雷达扫描区域按径向距离分为 5 个融化区域, 不同融化区域内允许存在的降水粒子如图所示. 图中不同融化区域内标出的降水粒子类型为在该区域可能存在该种降水粒子, 否则认为不可能存在该种降水粒子, 若某雷达融化区域内不允许某种类型降水粒子存在, 将该降水粒子类先验概率置为 0, 否则为 1.

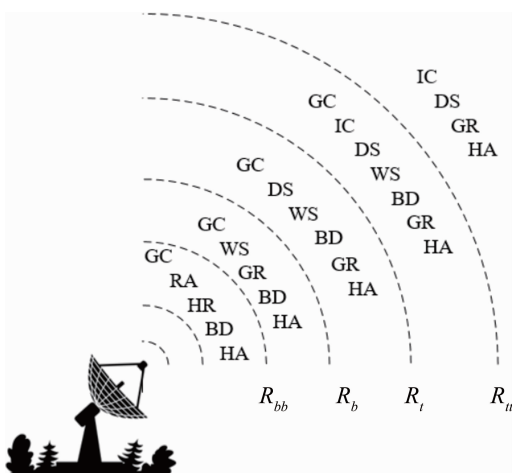


图2 不同融化区域内允许存在降水粒子信息

3.2.2 异常值信息

异常值信息结合了降水粒子类型的特征,根据不同降水粒子偏振参量取值范围不同确定类先验概率.比如暴雨(HR)的反射率 Z_H 不可能小于 30dBZ,因此当 $Z_H > 30\text{dBZ}$ 时认为暴雨的类先验概率为 0,更多异常值信息参见表 3.

表 3 属性特定范围内禁止存在粒子类别

极化参量	取值范围	禁止存在类别
Z_H	$> 40\text{dBZ}$	CR
Z_H	$> 50\text{dBZ}$	RA
Z_H	$< 30\text{dBZ}$	HR
Z_H	$< 40\text{dBZ}$	HA
Z_H	$< 20\text{dB}$	WS
Z_H	< 10 或 $> 60\text{dBZ}$	GR
Z_{DR}	$< 0\text{dB}$	WS

3.3 BNT 分类器降水粒子分类原理

设贝叶斯网络分类器输入为 $[Z_H, Z_{DR}, \rho_{HV}, K_{DP}] = [x_1, x_2, x_3, x_4]$, 根据最大后验概率准则, BNT 实现降水粒子分类问题可以描述为式(5)所示的最大后验概率问题:

$$\arg \max_j P(c_j | x_1, x_2, x_3, x_4), \quad j \in \{1, 2, \dots, 9\} \quad (5)$$

其中 j 表示类别标签, c_j 即第 j 类降水粒子. 根据贝叶斯式可将式(5)转化为:

$$\begin{aligned} & \arg \max_j P(c_j | x_1, x_2, x_3, x_4) \\ &= \arg \max_j \frac{P(c_j, x_1, x_2, x_3, x_4)}{P(x_1, x_2, x_3, x_4)} \\ &= \arg \max_j \frac{P(c_j) \prod_{i=1}^4 P(x_i | \pi_i, c_j)}{P(x_1, x_2, x_3, x_4)}, \quad j \in \{1, 2, \dots, 9\} \end{aligned} \quad (6)$$

其中 x_i 表示第 $i (i \in \{1, 2, 3, 4\})$ 号雷达偏振参量节点, π_i 表示节点 x_i 的父节点配置^[14] (即存在哪些父

节点)根据贝叶斯网络结构确定, $P(x_1, x_2, x_3, x_4)$ 为常数, 式(5)所示后验概率转化为先验概率和似然函数的乘积, 如式(7)所示, 最终求得的 j 对应的降水粒子类型即为降水粒子分类结果.

$$\arg \max_j \left\{ P(c_j) \prod_{i=1}^4 P(x_i | \pi_i, c_j) \right\}, \quad j \in \{1, 2, \dots, 9\} \quad (7)$$

综上所述, 基于离散属性 BNT 的降水粒子分类算法流程图如图 3 示.

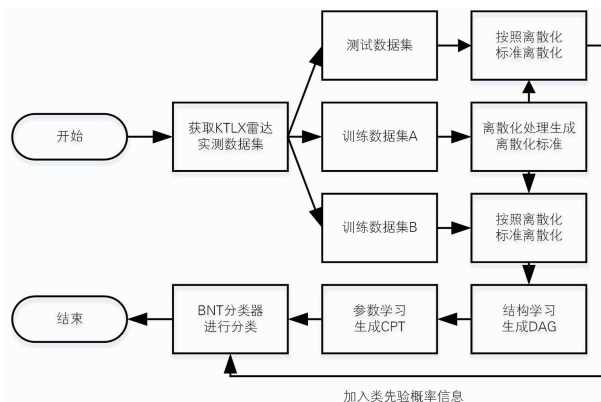


图3 基于离散属性BNT的降水粒子分类算法流程图

4 实验验证

4.1 实测数据验证

数据采集自美国国家海洋和大气管理局 (National Oceanic and Atmospheric Administration, NOAA) 官方网站, 选取 WSR-88D 雷达网中坐落于俄克拉荷马市 (OKLAHOMA) 的一部双偏振气象雷达 KTLX. 采集了 KTLX (0.5°仰角) 在 2018 年 6 月 7 日 15:14 的一次雷达回波数据, 进行降水粒子分类, 然后给出了 BNT 算法和 NOAA 的分类结果 (作为模糊逻辑算法的分类结果), 如图 4 所示. 不同类别降水粒子数量统计如表 4. 引入基于灰度共生矩阵的纹理特征统计量^[15] 来定量评估分类结果, 如表 5 所示.

通过图 4(a) 可以看出 BNT 分类算法可以很好完成降水粒子分类. 比较图 4 发现 BNT 算法图像相邻分类单元连续性强, 物理聚类特征表现更明显, 这在表 5 特征统计量中得到验证, 本文方法分类结果各个方向的同质性均高于模糊逻辑分类结果. 通过表 4 发现, 在中型降雨天气下, BNT 分类结果中降水粒子总量比 NOAA 模糊逻辑算法多, 说明对于一些未知的输入数据, 在模糊逻辑算法不能给出分类结果的情况下, BNT 算法给出了降水粒子分类结果. 这主要得益于雷达获得的实测数据 BNT 算法首先根据离散化标准对其进行离散化处理, 使得 BNT 算法对未知数据的处理能力更强, 证明 BNT 算法具有良好的泛化性.

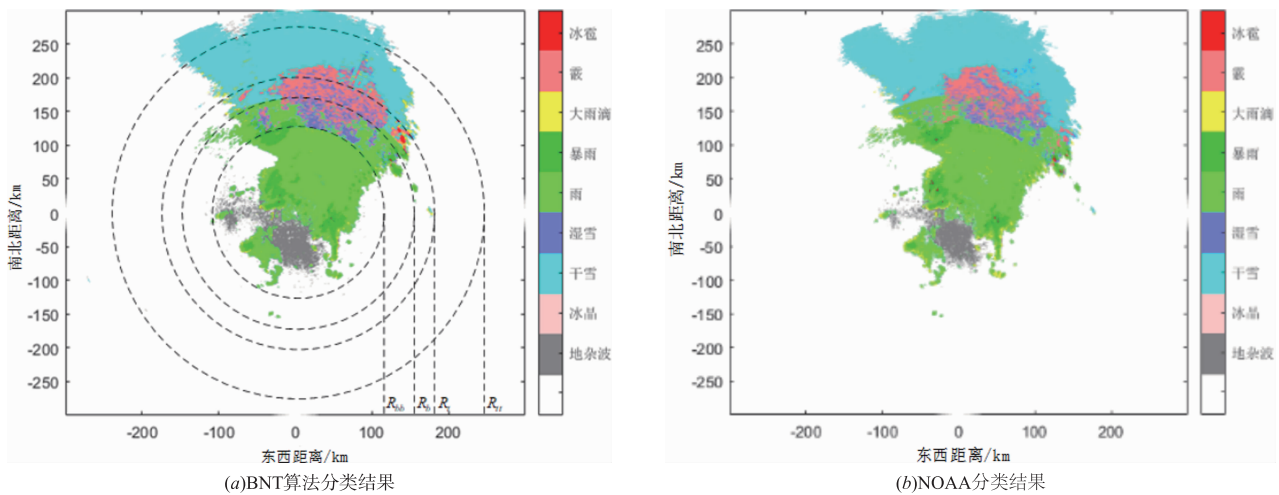


图4 分类结果图(2018/06/07 15:14)

表 4 各类粒子数量及占比(2018/06/07 15:14)

类别	BNT 算法		NOAA	
	数据个数	占比(%)	数据个数	占比(%)
地杂波	32661	18.63	27281	16.57
冰晶	891	0.51	452	0.27
干雪	22185	12.66	24470	14.86
湿雪	4905	2.8	5430	3.30
雨	81520	46.51	80637	48.96
暴雨	14744	8.41	10777	6.54
大雨滴	7760	4.43	6631	4.03
霰	10118	5.77	8326	5.06
冰雹	488	0.28	682	0.41
合计	175272	100	164686	100

表 5 灰度共生矩阵统计量特征(2018/06/07 15:14)

方向	BNT 算法		NOAA	
	熵	同质性	熵	同质性
0°	1.7942	0.9526	1.8557	0.9505
45°	2.0640	0.8858	2.1303	0.8793
90°	2.0540	0.8879	2.1185	0.8819
135°	2.0627	0.8856	2.1278	0.8795

5 结束语

本文提出了一种基于离散属性贝叶斯网络的双偏振气象雷达降水粒子分类算法,通过多个实验证明,使用条件概率表描述雷达偏振参量概率分布,避免了传统方法中概率模型或隶属度函数模型预设带来的误差,同时离散化过程使分类器具有较好的泛化性。

参考文献

[1] 楼小凤,师宇,李集明.云降水和人工影响天气催化数值模式的发展及应用[J].气象科技进展,2016,6(3):75-82.

[2] 张磊.C波段双线偏振雷达数据处理及其在粒子分类和降水估测中的应用[D].南京:南京信息工程大学,2017.

[3] 杨通晓,袁招洪.多波段双偏振天气雷达识别降水类型的模拟研究[J].高原气象,2017,36(1):241-255.

Yuan Zhaohong. Simulation research on hydrometeor classification by multi-wavelength dual linear polarization Doppler radar[J]. Plateau Meteorology, 2017, 36(1):241-255. (in Chinese)

[4] 曹俊武,刘黎平,葛润生.模糊逻辑法在双线偏振雷达识别降水粒子相态中的研究[J].大气科学,2005(05):159-168.

[5] Hassan, et al. A new fuzzy logic hydrometeor classification scheme applied to the french X-, C-, and S-band polarimetric radars[J]. Journal of Applied Meteorology & Climatology, 2013, 52(10):2328-2344.

[6] Roberto, et al. Multi-sensor analysis of convective activity in central Italy during the, HyMeX SOP 1. 1 [J]. Atmospheric Measurement Techniques, 2016, 8(9):9241-9287.

[7] Bechini, et al. A semisupervised robust hydrometeor classification method for dual-polarization radar applications[J]. Journal of Atmospheric and Oceanic Technology, 2015, 32(1):22-47.

[8] 冯亮,肖辉,孙跃.X波段双偏振雷达水凝物粒子相态识别应用研究[J].气候与环境研究,2018,23(03):366-386.

A study on hydrometeor classification and application based on X-band dual-polarization radar measurements [J]. Climatic and Environmental Research, 2018, 23(3):366-

386. (in Chinese)

- [9] Marzano, et al. Supervised classification and estimation of hydrometeors from C-band dual-polarized radars: a bayesian approach [J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(1) : 85 – 98.
- [10] 谢宏,程浩忠,牛东晓. 基于信息熵的粗糙集连续属性离散化算法[J]. 计算机学报, 2005(9) : 1570 – 1574.
- [11] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016. 156 – 157.
- [12] Brandes, et al. Freezing-level estimation with polarimetric radar. Journal of Applied Meteorology [J], 2004, 43(11) : 1541 – 1553.
- [13] Park, et al. The hydrometeor classification algorithm for the polarimetric WSR-88D; description and application to an MCS [J]. Weather and Forecasting, 2009, 24(3) : 730 – 748.
- [14] 张剑飞. 贝叶斯网络学习方法和算法研究[D]. 吉林: 东北师范大学, 2005.
- [15] 葛静祥. 图像纹理特征提取及分类算法研究[D]. 天津: 天津大学, 2010.

作者简介



李海 (通信作者) 男, 1976 年出生, 天津人, 教授, 博士, 主要研究方向为机载气象雷达信号处理、分布式目标检测与参数估计、自适应信号处理、阵列信号处理、动目标检测与参数估计.

E-mail: haili@cauc.edu.cn



尚金雷 男, 1993 年生, 山东威海人, 硕士研究生, 主要研究方向为机器学习、机载气象雷达信号处理.

E-mail: 18660039930@163.com



孙婷逸 女, 1995 年出生, 河北张家口人, 主要研究方向为机载气象雷达信号处理.

E-mail: 812901035@qq.com



冯青 女, 1980 年生, 陕西蒲城人, 讲师, 硕士, 主要研究方向为气象雷达信号处理、阵列信号处理.

E-mail: fengqing_tj@163.com



庄子波 男, 1980 年生, 山东潍坊人, 副教授, 硕士, 主要研究方向为航空气象.

E-mail: zbzhuang@cauc.edu.cn