

基于雷达反射率图像特征的冰雹暴雨识别*

路志英^{1)†} 刘海¹⁾ 贾惠珍²⁾ 尹静¹⁾

1) (天津大学电气与自动化工程学院, 天津市过程检测与控制重点实验室, 天津 300072)

2) (天津市气象局, 天津 300074)

(2014年1月17日收到; 2014年5月16日收到修改稿)

为了实现冰雹暴雨天气的识别与分类, 提出了一种基于雷达反射率图像特征的自动识别方法. 对雷达回波反射率图像中冰雹回波区域和暴雨回波区域的图像特征进行提取, 通过分析冰雹暴雨间单一特征的差异性和不同特征之间的分类互补性, 确定了识别冰雹暴雨的有效图像特征 (包括强度特征和纹理特征). 将提取出的样本有效特征与探空数据 ($0\text{ }^{\circ}\text{C}$ 和 $-20\text{ }^{\circ}\text{C}$ 温度层高度) 结合, 利用粗糙集理论进行数据挖掘, 进而建立了冰雹暴雨天气的客观识别模型. 通过对 362 个测试样本的测试与统计, 冰雹击中率达到 93.29%, 暴雨的击中率达到 89.27%, 并且两者均具有较低的误警率. 实验结果与传统 PUP 系统比较, 表明利用雷达反射率图像特征实现对冰雹暴雨天气的识别与分类具有较好的效果.

关键词: 特征提取, 粗糙集理论, 数据挖掘

PACS: 92.60.Nv, 92.60.Qx, 92.60.Wc

DOI: 10.7498/aps.63.189201

1 引言

冰雹和暴雨是由强对流天气系统引起的两种重要气象灾害, 准确地识别和预报冰雹暴雨天气是气象研究人员的重点研究目标之一. 针对强对流天气雷达回波的识别技术, 国外从二十世纪六七十年代就开始了大量研究, 并提出了一些识别方法: Rinehart 和 Garvey^[1,2] 应用模式识别中相关分析的方法对整幅回波图像做处理, 并介绍了使用单一常规气象雷达的反射率因子数据获取风暴的三维运动信息. 1982年, Austin 和 Bellon^[3] 提出了三维矩心识别方法, 即把每一个风暴单体看作是一个具有三维连续结构的整体, 通过计算出的相关物理特征量进行识别. Rosenfeld^[4] 以及 Michael 和 Gerry^[5] 对这一方法进行改进和应用, 通过对风暴形状的几何推理来处理其分裂和合并, 但当出现多个风暴相距较近、风暴移动不规则以及风暴出现了合并和分裂时, 误差仍然较大. 1998年, Johnson 和

Pamela^[6] 进一步提出利用 7 个不同的反射率因子识别阈值来替代此前的惟一阈值, 同时采取特征核抽取技术, 该方法在美国实地测试过程中取得了较好的结果.

近十几年来, 随着天气雷达技术的发展, 国内以雷达资料为基础的对流天气临近预报技术的研究进展很快^[7-9]. 胡文东等^[10] 利用宁夏中北部 51 个强对流降水过程反射率图像资料, 分析了不同地区强降水过程在反射率图像上所表现出的基本纹理特征差异. 张家国等^[11] 在对近几年来武汉的暴雨天气过程进行多普勒雷达图像分析, 总结了 β 中尺度暴雨回波系统的形态、强度、结构等基本特征, 以及强降水与诸多因素的关系. 刘黎平等^[12] 提出以面积较大的回波块为研究对象, 通过计算等高面中云体的中心、面积、强度等特征实现风暴识别的二维矩心识别法. 王瑾^[13] 通过使用雷达网的多雷达拼图资料, 分析雷达回波的统计特征, 建立了冰雹等强对流天气的有效识别方法.

新一代多普勒天气雷达作为识别强对流天气

* 天津市自然科学基金 (批准号: 14JCYBJC21800) 和气象关键技术集成与应用项目 (批准号: CMAGJ2013M02) 资助的课题.

† 通讯作者. E-mail: luzy@tju.edu.cn

的重要平台,其反射率因子图和径向速度图能够直接展现对流云团的结构、形态以及对流场的分布,从而促进了人们对冰雹、暴雨等强对流天气的特点及其演变规律的认识^[14].上述国内外的研究表明:利用雷达回波的图像数据,发掘冰雹暴雨天气回波的特征,可以实现对冰雹暴雨等强对流天气的识别.

然而利用国外识别方法对中国地区的强对流天气进行识别,其结果出现较大偏差;国内的研究则大多是针对强对流天气的个例进行分析.基于上述分析,本文以雷达回波反射率产品作为信息源,利用反射率因子产品生成雷达反射率图像,并以雷达回波反射率图为基础,首先在图像上确定强回波天气区域,提取出特定区域的图像统计特征;然后将提取出来的特征进行基于粗糙集理论的数据挖掘,获得区分强天气过程冰雹和暴雨的判别规则;最后利用这些规则对所关注的天气过程进行识别.结果表明,本文的方法可以有效地识别强对流天气中冰雹与暴雨过程,并且实现冰雹和暴雨两者的分类.

2 雷达回波反射率图

本文中所使用的雷达为CINRAD-SA多普勒天气雷达,其基本产品包括:反射率因子、平均径向速度和速度谱宽^[15,16],其中,反射率因子的大小反映了气象目标内部降水粒子的尺度和密度分布,用来表示气象目标的强度,产品上数据的单位用“dBZ”表示^[17,18].一般情况下,反射率强度越高,出现强对流天气的可能性越大^[19].为了便于观察,通常将反射率范围量化为15个等级.反射率图像通过15种不同的颜色来反映反射率的强度差异,每一种颜色对应一个反射率强度范围,用黑色表示无效回波,反射率强度与颜色对应关系如图1所示.

雷达反射率图像来源于基本产品中的反射率因子,是将基数据转化为一个512×512的8位位图.反射率图中以雷达所处位置为原点,坐标系中每一坐标点上的颜色取值用于反映相应空间位置

的雷达回波反射率强度,如图2所示.

3 数据资料

本文的数据来源于天津市气象台提供的雷达基数据以及天气状况分型表.从连续的长时间序列资料中,选取2004年6月21日至2010年8月18日期间所有数据信息完整的冰雹暴雨过程的雷达基数据进行分析和处理,其中包括85个冰雹过程(冰雹降落直径大于5 mm)和81个暴雨过程(降水累积量大于20 mm·h⁻¹)共1206个样本.每个过程中的数据均为从冰雹暴雨降落开始时刻到降落结束时刻的全部样本,代表了一个连续的降水演化过程.依据天气状况分型表的记录,不同年份的冰雹过程数和暴雨过程数如表1所示.

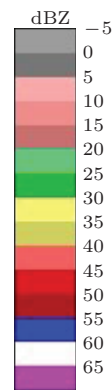


图1 (网刊彩色) 雷达反射率图基本色与反射率强度的对应关系

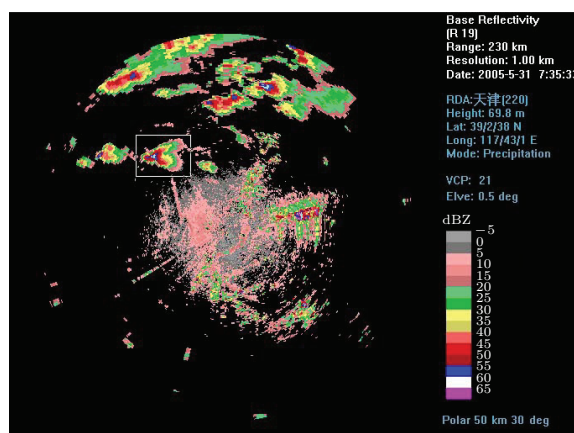


图2 (网刊彩色) 反射率图像

表1 各年份冰雹过程数与暴雨过程数

年份	2004	2005	2006	2007	2008	2009	2010	总数
日期	6.21—9.15	4.19—8.16	4.17—8.25	5.15—8.25	6.14—7.17	4.13—9.26	5.3—8.18	
冰雹过程数/次	13	28	16	9	10	5	4	85
暴雨过程数/次	8	12	17	13	9	12	10	81

4 基于反射率图的回波强度特征提取

反射率图像上每一点的颜色都与一个固定的反射率强度范围相对应, 即不同的颜色本质上代表不同的回波反射率强度. 因此, 可以通过统计反射率图像上不同颜色的分布情况来反映回波强度特征. 本文引入强度直方图来统计所关注回波区域的不同强度所占比例, 进而获取各回波区域的强度特征. 借助图像的强度直方图实现回波强度特征提取的方法简单有效, 并且对图像本身的尺寸和方向的依赖性较小, 具有较强的鲁棒性.

4.1 强度特征

4.1.1 强度直方图

强度直方图是对图像内容的一种直观的描述方式, 它描述了雷达反射率图中不同强度在图像中所占的比例. 描述如下:

设一幅图像包含 M 个像素, 反射率图像的强度空间被量化为 N 种不同的强度, 不同强度出现的概率为

$$H = (h_1/M, h_2/M, \dots, h_N/M), \quad (1)$$

其中, h_i 为第 i 种强度在整幅图像中具有像素数, $\sum_{i=1}^N h_i = M$.

不难看出, 强度直方图表现的是反射率图像的一种全局特性, 对图像的旋转以及平移具有很好的不变性, 故对于每幅反射率图像仅具有惟一的强度直方图.

4.1.2 强度特征定义

把直方图的包络线看作一条曲线, 则可以将其表示成 1-D 函数 $f(r)$, 自变量 $r \in [1, L]$, L 为回波强度的量化级数, 故 $(r, f(r))$ 表示曲线上的任意点^[20]. 使用一阶统计测度如均值、方差等作为不同类别间的特征差异, 具体表示如下.

1) 均值: 是回波强度量化后的量化级别的平均值, 表征了回波反射率的强度大小,

$$Mean = \sum_{r=1}^L r \cdot f(r). \quad (2)$$

2) 方差: 是强度对比度的度量, 表达了曲线相对于均值的分布情况, 描述了直方图的相对平滑程

度, 进一步描述了回波反射率强度的分散程度,

$$Variance = \sum_{r=1}^L (r - m)^2 f(r). \quad (3)$$

3) 偏度: 是用于度量分布的倾斜程度及偏向的指标, 直观看来就是密度函数曲线在其中一个方向上的尾部有拉长趋势的程度. 计算公式为

$$Skewness = \frac{1}{(Variance)^{3/2}} \sum_{r=1}^L (r - m)^3 f(r), \quad (4)$$

正态分布偏度为 0, 左右两侧是对称的. 当 $Skewness < 0$, 称分布具有负偏离, 直观表现为分布在左方向的尾部比在右方向的尾部有拉长的趋势; 反之 $Skewness > 0$, 称分布为正偏离, 表现为分布在右方向的尾部比在左方向的尾部有拉长的趋势, 如图 3 所示.

4) 峰度: 又称峰态系数, 是一个反映随机变量分布形状的量, 它度量了分布尾部的长度. 计算公式为

$$Kurtosis = \frac{1}{(Variance)^2} \sum_{r=1}^L (r - m)^4 f(r), \quad (5)$$

正态分布峰度为 3. $Kurtosis < 3$ 称分布具有不足的峰度, 表示其尾部比正态分布尾部短; $Kurtosis > 3$ 称分布具有过度的峰度, 表示其尾部比正态分布尾部长, 如图 4 所示.

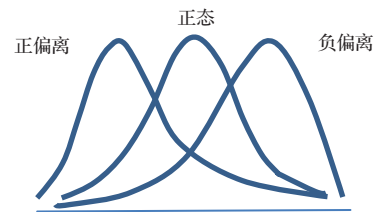


图 3 偏度示意图

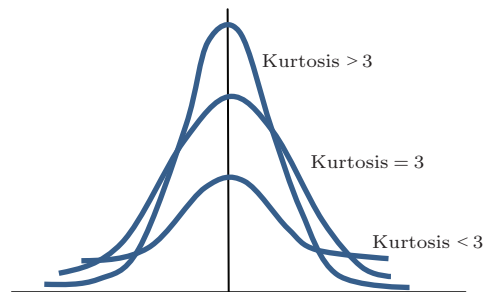


图 4 峰度示意图

4.2 回波强度特征的气象含义

雷达发射的电磁波束在大气中传播, 遇到云和降水粒子时被散射, 其中一部分向后的散射波返回

雷达方向, 被雷达天线接收. 粒子散射电磁波的能力和粒子的大小、形状以及其电磁特性有关.

一般情况下, 雷达反射率因子的大小与散射粒子的强度成正比. 对于冰雹和暴雨, 它们的粒子为冰球和水球, 由于水球的尺寸超过一定限度就会自行破裂, 故冰球的粒子尺寸一般要大于水球. 因此在强对流天气中, 冰雹单体在大部分情况下反射率强度要高于暴雨单体, 这就说明了在强度特征中冰雹的均值要高于暴雨. 其次, 冰雹的层次结构比暴雨要明显 (参见图 5), 即单体核反射率很高并且向外逐渐降低. 所以冰雹反射率强度在不同强度等级分布较分散, 而暴雨相对集中, 因此在强度特征中冰雹的方差要高于暴雨. 对于偏度和峰度, 依据定义并结合反射图像分析可知, 冰雹单体通常具有较高的值, 因此在强度特征中, 偏度和峰度是对均值和方差的重要补充.

4.3 冰雹暴雨回波的强度特征提取

本文将索引值 1—15 对应反射率强度 -5—65 dBZ, 用索引值 0 表示无效区域的强度 (黑色). 直方图中, 统计不同强度的索引值的出现概率, 并利用得到的直方图计算均值、方差等回波的强度特征.

根据气象学理论及预报经验, 雷雨大风、冰雹、暴雨等天气现象发生在存在“单体”的对流云系中. 单体反射率强度具有层次结构, 反射率大于等于 40 dBZ 的连通点的集合可以包含绝大部分单体的主体部分. 经大量测试与实验, 对强单体的主体部分进行特征提取具有较好的效果. 由于 40 dBZ 对应的索引值为 10, 故强度直方图的横坐标取值为 10—15 即可.

本文选取 2005 年 5 月 31 日海淀地区的降雹过程中某一时刻反射率图像为例, 此次降雹尺寸为 10 mm, 降雹区域的冰雹回波如图 5(a) 所示, 图 5(b) 为提取出的冰雹回波的主体. 图 5(c) 为 2008 年 7 月 1 日天津汉沽地区暴雨过程中某一时刻的反射率图, 此次累积降雨量达到 31.3 mm, 图 5(d) 为提取出的暴雨回波主体.

分别统计图 5(b) 和 (d) 所示回波主体内各强度对应像素点的出现概率, 进而得到两者的强度直方图, 如图 6 所示. 从直方图中可以看出: 冰雹回波总体反射率强度较高, 在高于 50 dBZ (索引值为 12) 的强度范围内, 仍占有一定的比重; 暴雨回波的

反射率强度等级大部分集中在 10—12 级之间, 对应反射率强度为 40—50 dBZ.

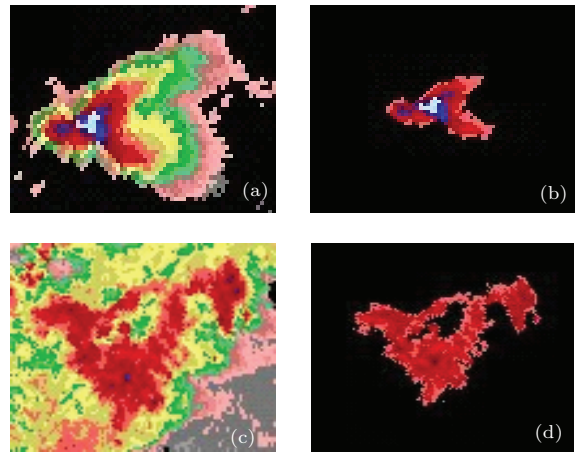


图 5 回波反射率 (a) 冰雹回波; (b) 冰雹回波主体; (c) 暴雨回波; (d) 暴雨回波主体

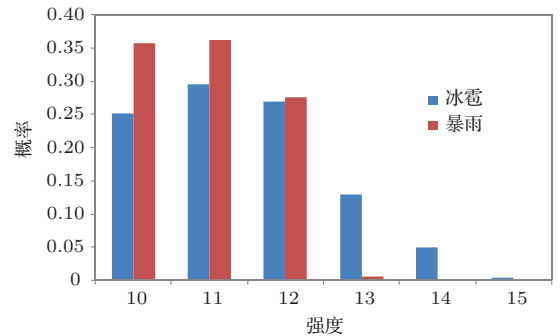


图 6 (网刊彩色) 强度直方图

根据两个强度直方图分别计算两组回波强度特征, 包括均值、方差、偏度和峰度, 两者对比如表 2 所示.

表 2 冰雹暴雨强度特征对比表

样本	天气类型	均值	方差	偏度	峰度
图 5	冰雹	11.4424	1.3546	0.481583	2.51802
图 6	暴雨	10.9283	0.647272	0.185759	1.71357

从两组数据对比可以看到: 冰雹与暴雨之间的强度特征存在明显差异, 对于均值、方差、偏度、峰度四个强度特征, 冰雹单体在数值上一般大于暴雨单体, 故借助强度特征可以实现较好的分类效果.

5 基于反射率图的回波纹理特征提取

纹理是对图像各像素灰度的空间分布的一种描述, 属于低级别特征. 纹理特征是图像处理和

模式识别的主要特征之一, 它很好地兼顾了宏观和细部结构两个方面. 常用的纹理特征描述方法包括统计分析方法、结构分析方法、频谱分析方法等, 本文采用的是统计分析方法中空间灰度共生矩阵法^[21,22].

5.1 纹理特征

5.1.1 空间灰度共生矩阵

空间灰度共生矩阵法是建立在估计图像的二阶组合条件概率密度函数基础上的, 是人们公认的一种重要的纹理分析方法. 灰度共生矩阵表示了灰度的空间依赖性, 描述了在一种纹理模式下的像素灰度的空间关系.

空间灰度共生矩阵生成方法: 首先, 将一幅图像进行灰度化, 并且将每个像元上的灰度量化为 N 级, 即灰度级划分数目为 N , 则共生矩阵就是 $N \times N$ 矩阵. 然后定义一个方向和一个固定步长, 那么该物体共生矩阵 $\mathbf{P}(N \times N)$ 的第 (i, j) 元素值等于灰度级 i 和 j 在物体内沿该方向相距该指定步长的两个像素同时出现的次数再除以 M , 其中 M 是对矩阵 \mathbf{P} 有贡献的像素对的总数.

例如, 图 7 为一个 4×4 , 灰度层为 4 的图像. 由于灰度层为 4, 故其灰度共生矩阵为 4×4 的矩阵. 取步长 $d = 1$, 方向为 0° , 即统计图像中同一行前后两个相邻像素点的灰度状态出现次数. 若采用对称共生矩阵, 则可以计算出图 7 中 $(0, 0)$ 共出现 4 次, 同理, $(1, 1)$ 出现 4 次; $(2, 2)$ 出现 6 次; $(0, 1)$ 、 $(1, 0)$ 和 $(3, 3)$ 各出现 2 次; $(0, 2)$ 、 $(2, 0)$ 、 $(2, 3)$ 和 $(3, 2)$ 各出现 1 次. 因此, 可以得到其空间灰度共生矩阵

$$\mathbf{P} = \begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix},$$

归一化后

$$\mathbf{P}' = \begin{bmatrix} 1/6 & 1/12 & 1/24 & 0 \\ 1/12 & 1/6 & 0 & 0 \\ 1/24 & 0 & 1/4 & 1/24 \\ 0 & 0 & 1/24 & 1/12 \end{bmatrix}.$$

5.1.2 纹理特征定义

根据灰度共生矩阵, 可以定义出大量的纹理特征, Haralick 等用灰度共生矩阵提取了 14 种特征,

分别为角二阶矩、熵、对比度、相关性、均匀性、逆差矩、最大概率、纹理方差、共生和均值、共生和方差、共生和熵、共生差均值、共生差方差、共生差熵. 由于灰度共生矩阵计算量很大, 并且考虑特征对于反射率图的分类效果, 本文选择以下 6 种纹理特征用于实现对强对流天气的识别.

1) 角二阶矩

$$ASM = \sum_i \sum_j \mathbf{P}(i, j)^2. \quad (6)$$

角二阶矩是灰度共生矩阵各元素的平方和, 又称能量, 它是图像纹理灰度变化均匀性的度量, 反映了图像灰度分布均匀程度和纹理粗细程度. 从图像整体来看, 纹理越粗, ASM 值相应越大, 反之, ASM 值较小.

2) 对比度

$$CON = \sum_i \sum_j (i - j)^2 \mathbf{P}(i, j). \quad (7)$$

对比度是主对角线附近的惯性矩, 它度量矩阵的值如何分布以及影像中局部变化的多少, 反映了图像的清晰度和纹理的沟纹深浅. 对于粗纹理, $\mathbf{P}(i, j)$ 的大数值集中于主对角线附近, 此时 $(i - j)$ 较小, 相应的 CON 值也较小; 对于细纹理则有较大的值.

3) 相关性

$$COR = \sum_i \sum_j (i - \mu_x)(j - \mu_y) \times \mathbf{P}(i, j) / \sigma_x \sigma_y, \quad (8)$$

$$\mu_x = \sum_i i \sum_j \mathbf{P}(i, j),$$

$$\mu_y = \sum_j j \sum_i \mathbf{P}(i, j), \quad (9)$$

$$\sigma_x = \sum_i (i - \mu_x)^2 \sum_j \mathbf{P}(i, j),$$

$$\sigma_y = \sum_j (j - \mu_y)^2 \sum_i \mathbf{P}(i, j). \quad (10)$$

相关性用来度量空间灰度共生矩阵的元素在行或列方向上的相似程度, 因此, 相关值的大小反映图像中的局部灰度相关性.

4) 熵

$$ENT = - \sum_i \sum_j \mathbf{P}(i, j) \log \mathbf{P}(i, j). \quad (11)$$

熵用来度量图像纹理的随机性, 反映了图像中纹理的非均匀程度或复杂程度. 当灰度共生矩阵中

所有值均相等时,它取得最大值;相反,如果灰度共生矩阵中的值非常不均匀时,其值比较小。

5) 逆差距

$$IDM = \sum_i \sum_j P(i, j) / [1 + (i - j)^2]. \quad (12)$$

逆差距反映图像纹理的同质性,度量图像纹理局部变化的多少,其值大则说明图像纹理的不同区域间缺少变化,局部非常均匀。

6) 灰度平均值

$$GM = \sum_i \sum_j i \cdot P(i, j). \quad (13)$$

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

图7 图像灰度分布

5.2 纹理特征的气象含义

从雷达的反射率强度来看,冰雹单体和暴雨单体具有不同的梯度特性。冰雹单体的过渡带的平均宽度较暴雨单体要窄,即由高反射率到低反射率的空间距离短。因此,冰雹单体具有明显的层次结构,一般会呈现出一种细纹理。暴雨单体的区域较大,并且过渡带比较宽,在雷达反射率图像中,往往呈现出大片相同强度的区域,强度集中在较小的区间内。因此,对比冰雹单体,暴雨单体一般会呈现出一种粗纹理。

从纹理特征的定义来看,纹理越粗,角二阶矩越大,而对比度越小。由于冰雹单体层次结构明显,图像整体更加杂乱无章,随机性较强,故冰雹具有较大的熵值。暴雨单体图像局部纹理变化较小,故逆差距的值较大。对于平均值,与图像强度有一定的关系,而相关值的大小反映图像中的局部灰度相关性,两者均具有一定的分类特性。

5.3 冰雹暴雨回波的纹理特征提取

本文中雷达反射率图中共包含15个强度(参见图1),黑色代表无效区域,因此反射率强度可量化为16级,即0级表示黑色无效区域,1—15级分

别表示-5—65 dBZ强度。提取回波纹理特征的步骤如下:

步骤1 在雷达图像中提取关注区域轮廓,并把轮廓内的每个像素点进行反射率强度量化;

步骤2 对上述关注回波区域生成空间灰度共生矩阵,其步长 $d = 1$,方向分别为 0° (水平)、 45° (主对角线)、 90° (垂直)、 135° (次对角线),如图8所示;

步骤3 将步骤2中生成4个空间灰度共生矩阵分别进行归一化处理,并利用归一化的共生矩阵计算特征值,每个特征值取四个方向的平均值。

选取2.4中的冰雹和暴雨样本,如图9(a), (b)所示,分别计算两者的纹理特征,如表3所示。

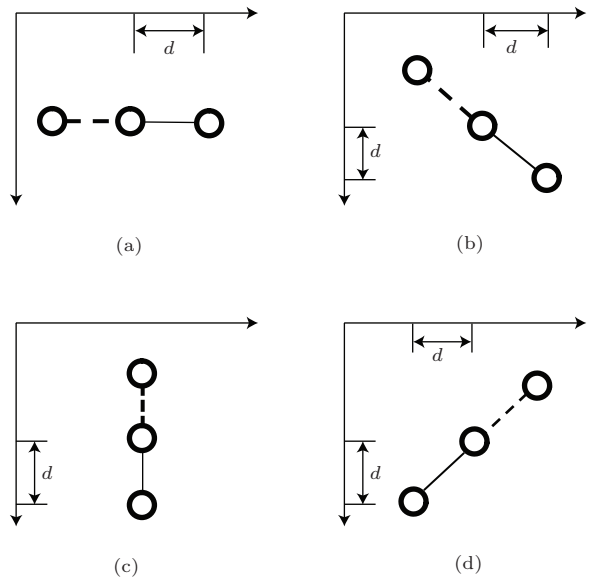


图8 生成空间灰度共生矩阵的方向 (a) 0° 方向; (b) 45° 方向; (c) 90° 方向; (d) 135° 方向

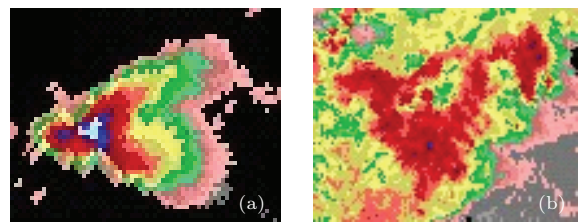


图9 冰雹与暴雨反射率图 (a) 冰雹; (b) 暴雨

在强对流天气的类型识别中,冰雹和暴雨的纹理特征有很好的区分性,可以实现很好的分类效果。通过表3可以看出:对于角二阶矩和逆差距两个特征,冰雹的特征值要小于暴雨;对于对比度、相关性、熵和平均值四个特征,冰雹的特征值要大于暴雨。

表3 纹理特征对比

时间	类型	角二阶矩	对比度	相关性	熵	逆差距	平均值
图9(a)	冰雹	0.081494	1.18366	0.035623	2.98062	0.703396	3.56927
图9(b)	暴雨	0.086044	0.608682	0.023509	2.65397	0.767139	2.09813

本文提取了85个冰雹过程的627个样本以及81个暴雨过程中的579个样本的纹理特征. 为了更好地观察每个纹理特征的分类效果, 本文从样本点的六个纹理特征中任意选取四个并分成两组, 以二维图像的方式呈现出来, 如图10和图11所示.

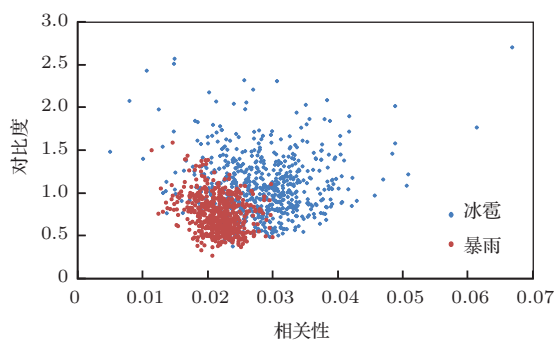


图10 (网刊彩色) 分类效果图(相关性-对比度)

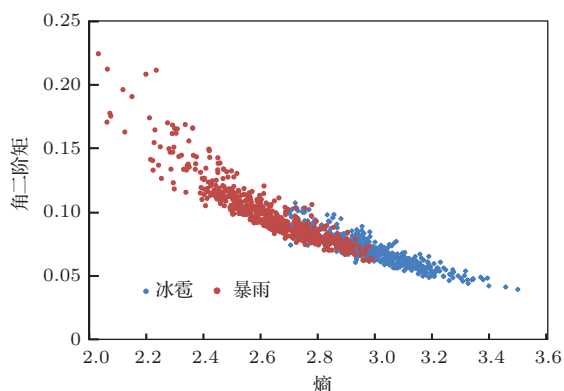


图11 (网刊彩色) 分类效果图(熵-角二阶矩)

将图10和图11中二维平面上的点分别投射到横轴与纵轴, 则可以得到对于单一特征样本点的分类特性. 由图像可以看出, 对于每一个单独的特征, 不同类型样本点之间都会存在一定差异, 具有一定

的分类性, 但也都有较大的重叠, 而从一维特征提升到二维, 可以大大降低样本间的重叠, 增强分类效果. 若把相关性、对比度和熵这三维特征映射到三维图像中, 则如图12所示, 冰雹和暴雨的分类效果进一步增强. 因此可以说明提取的纹理特征是有效的, 并且不同特征间具有一定的互补性, 随着维数继续增加, 分类效果会逐渐增强.

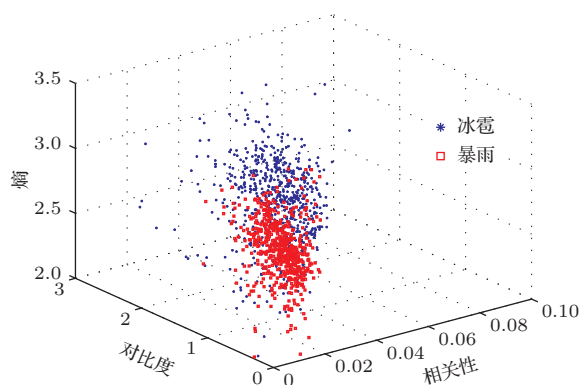


图12 (网刊彩色) 三维特征分类效果图

6 特征数据库建立

在冰雹和暴雨天气的识别过程中, 除强度特征和纹理特征之外, 单体的回波顶高度以及 0°C 和 -20°C 温度层高度数据在区分冰雹和暴雨两类天气过程中也起到了重要作用, 这三个特征与不同天气的关系如表4所示. 三个特征对于冰雹和暴雨天气差异明显, 故本文将回波顶高度、 0°C 温度层高度以及 -20°C 温度层高度三个特征加入特征数据库, 进而建立更加全面的冰雹暴雨天气的客观识别模型.

表4 冰雹暴雨天气中强回波与 0°C 和 -20°C 高度之间的关系

特征	天气	冰雹	暴雨
0°C 高度		3—4.5 km	3.5—5.2 km
-20°C 高度		5.8—7.6 km	6.8—8.6 km
回波顶高度		-20°C 以上	大部分在 -20°C 以下, 回波顶高于 -20°C 层时, 0°C 层基本在4.5 km以上, 融化效应明显

本文从天津市气象台的历史资料中选取 85 个冰雹过程和 81 个暴雨过程共 1206 个样本数据. 将 1206 个样本数据分别提取 13 个特征, 包括 4 个强度特征(均值、方差、偏度、峰度)、6 个纹理特征(角二阶矩、对比度、相关性、熵、逆差距和平均值)、单体的回波顶高度、0 °C 和 -20 °C 温度层高度. 将每个样本数据的 13 个特征及其天气类型存入到 MySQL 数据库中, 形成训练样本特征数据库, 部分数据如表 5 所示.

7 基于粗糙集的客观模型建立

以训练样本特征数据库为基础, 利用数据挖掘技术, 挖掘数据库中隐含的和有意义的知识, 形成知识库, 进而建立客观模型 [23]. 数据挖掘常用技术包括: 模糊方法、粗糙集理论、人工神经网络、遗传算法、云理论等. 本文采用粗糙集理论进行数据挖掘.

粗糙集理论最早是由波兰科学家 Pawlak 于 1982 年提出的 [24,25]. 该理论用于处理含糊性和不确定性的数学问题, 具有很多独特的优越性 [26]. 目前, 粗糙集理论被广泛应用于机器学习、过程控制、模式识别、专家系统、冲突分析等领域.

粗糙集作为一种数学工具, 能从数学上严格地处理数据分类问题, 尤其适用于数据具有噪声、不完全或不精确的非线性问题. 此外, 粗糙集理论除了需要待处理问题的数据集之外, 不需要为其提供其他任何先验信息. 本文要实现基于雷达反射率图的冰雹和暴雨之间的分类问题, 本身是一种不完整信息问题, 并且雷达数据具有噪声, 因此采用粗糙集理论实现客观模型的建立具有很大的优势.

7.1 粗糙集基本概念

1) 信息系统: 信息系统是一个四元组 $IS = (U, A, V, f)$, 其中, U 是对象的有界非空集合, 也称为域; A 为属性的有界非空集合; V 是属性的值域集合; f 是映射函数, 即 $f : U \times A \rightarrow V$.

2) 等价关系: 信息系统 $IS = (U, A, V, f)$, 任意属性集 $B \subseteq A$, 关于 B 的等价关系为

$$IND(B) = \{(x, y) \in U \times U | f_b(x) = f_b(y), for b \in B\}, \quad (14)$$

可见, 若 $(x, y) \in IND(B)$, 则对象 x, y 在属性集 B 上是等价的或不可区分的.

3) 上近似和下近似: 设 $X \subseteq U, B \subseteq A$, 则 X 对 B 的下近似 $B_-(X)$ 可定义为 X 所包含的关于 B 的所有等价类的并集, 即

$$B_-(X) = \cup\{[x]_B \subseteq X\}$$

或 $B_-(X) = \{x \in U | [x]_B \subseteq X\}, \quad (15)$

X 对 B 的上近似 $B^-(X)$ 可定义为与 X 交集非空的关于 B 的所有等价类的并集, 即

$$B^-(X) = \cup\{[x]_B \cap X \neq \emptyset\}$$

或 $B^-(X) = \{x \in U | [x]_B \cap X \neq \emptyset\}. \quad (16)$

4) 边界区域和粗糙集

边界区域: 设 $X \subseteq U, B \subseteq A$, 对象集 X 关于属性集 B 的边界区域定义为

$$BN_B(x) = B^-(X) - B_-(X). \quad (17)$$

粗糙集: 若 $BN_B(x) \neq \emptyset$, 则称 $BN_B(x)$ 是对象集 X 关于属性集 B 的粗糙集. 粗糙集基本概念如图 13 所示.

表 5 特征数据库中部分数据

强度特征				纹理特征							0 °C 高度	-20 °C 高度	回波顶高度	天气类型
均值	方差	偏度	峰度	角二阶矩	对比度	相关性	熵	逆差距	灰度平均值					
12.02	1.913	0.010	2.071	0.064	1.574	0.0403	3.117	0.647 7	3.8079	3.75	6.75	11.56	冰雹	
11.21	1.203	0.649	2.649	0.068	1.057	0.0273	3.059	0.7234	2.7162	3.98	6.86	9.71	冰雹	
11.607	1.402	0.119	1.933	0.054	1.197	0.0322	3.174	0.6753	3.0062	2.96	5.82	5.97	冰雹	
11.189	1.101	0.583	2.486	0.062	0.989	0.0259	3.023	0.7088	2.5072	3.69	7.06	10.81	冰雹	
...	
10.869	0.579	0.239	1.804	0.089	0.741	0.0225	2.653	0.7518	2.2389	4.30	7.64	6.53	暴雨	
10.631	0.279	-0.07	1.979	0.136	0.876	0.0148	2.348	0.7074	2.1647	4.99	8.69	9.15	暴雨	
10.942	0.532	0.089	1.881	0.121	0.652	0.0233	2.468	0.7662	2.4601	4.86	7.89	6.57	暴雨	

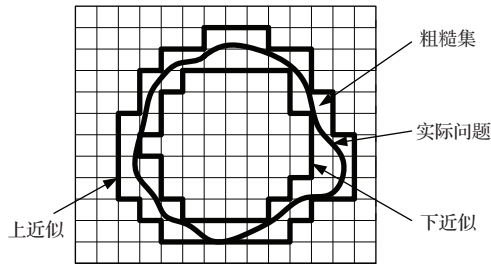


图 13 粗糙集概念示意图

7.2 客观模型的建立

由于使用粗糙集理论进行数据挖掘,故首先要对数据进行离散化.离散化就是对连续属性的取值范围或取值区间进行划分,将一个大区间分为若干个小区间.本文采用的等频离散化方法是一种简单的无监督和单变量离散化算法.它是将每个特征属性值根据设定的频数 k ,将测试样本分为 k 个子区间,每个子区间包含相同的样本数.

本文选取 85 个冰雹过程和 81 个暴雨过程共 1206 个样本数据,随机选取其中的 70% 共 844 个样

本作为学习样本,362 个作为测试样本,每个属性值的等频离散频数为 6.

为了处理智能数据,粗糙集将具有条件属性和决策属性(二者的集合为属性集)的信息系统描述为决策表.针对决策表,需要进行决策表约简,化简决策表中的条件属性和属性值,使决策表在保持决策能力的同时,具有较少的条件属性和属性值[27].本文采用 Johnson 算法进行决策表的约简,并生成规则库,其中共 182 条规则,部分规则如表 6 所示.

表 6 中,每一行代表一项规则的基本信息,第一列为规则条件,第二列和第三列为学习样本中满足该条件的冰雹样本个数和暴雨样本个数,最后一列表示满足该条件的样本数占学习样本总数的比例.以表中最后一项规则为例,对规则表的解释如下:当 Variance (方差)在 0.5118 和 0.6217 之间, Kurtosis (峰度)在 2.1993 和 2.3479 之间时,有 2 个样本为冰雹,16 个样本为暴雨,满足该条件的样本数共为 18 个,占总样本数的 2.13%.

表 6 规则数据库中部分规则

条 件	不同类型的支持样本数		样本覆盖度/%
	冰雹	暴雨	
Variance $\in [1.35423, +\infty)$	134	6	16.59
0 °C 高度 $\in (3.70, 3.97)$ & ENT $\in [3.037, +\infty)$	39	0	4.62
Kurtosis $\in [2.1993, 2.3479)$ & GM $\in (-\infty, 2.0361)$	0	24	2.84
...
Variance $\in [0.5118, 0.6217)$ & Kurtosis $\in [2.1993, 2.3479)$	2	16	2.13

7.3 决策方法

针对 7.2 节中得到的规则,采用标准投票的方法对实况天气进行决策,具体决策方法如下:

- 1) 定义冰雹和暴雨两个变量 hailstone 和 rainstorm,并且均初始化为 0;
- 2) 对待识别的天气样本计算各项特征数据;
- 3) 将 2) 中特征数据依次与客观模型中的分类规则进行匹配,匹配过程中,如果符合某条规则,则将 RHS Support 中的表示冰雹与暴雨的数目分别加到变量 rainstorm 和 hailstone 上,否则直接跳过本规则并对下一条规则进行匹配,直到全部规则匹

配结束;

- 4) 比较两个变量 hailstone 和 rainstorm 的大小,若 hailstone > rainstorm,则识别结果为冰雹;若 rainstorm > hailstone,则识别结果为暴雨;若没有规则可以匹配,则归为正常天气.

8 识别结果与分析

为了将结果进行统计分析,气象学上通常引入命中率(percent of doom, POD)、误警率(false alarm rate, FAR)和临界成功指数(critical success index, CSI)来进行评价.以冰雹天气为例,对命中率、误警率、临界成功指数定义如下.

$$POD = \frac{\text{正确识别的冰雹样本数}}{\text{冰雹样本总数}}, \tag{18}$$

$$\text{FAR} = \frac{\text{错误识别的冰雹样本数}}{\text{正确识别的冰雹样本数} + \text{错误识别的冰雹样本数}}, \quad (19)$$

$$\text{CSI} = \frac{\text{正确识别的冰雹样本数}}{\text{冰雹样本总数} + \text{错误识别的冰雹样本数}}, \quad (20)$$

根据命中率 POD, 可计算漏报率 $= 1 - \text{POD}$.

本文针对 362 个测试样本进行了测试, 可以得到以下结果: 179 个冰雹样本中, 能够正确识别出 167 个, $\text{POD} = 93.29\%$, $\text{FAR} = 10.69\%$, $\text{CSI} = 83.92\%$; 183 个暴雨样本中, 能够正确识别出 163 个, $\text{POD} = 89.07\%$, $\text{FAR} = 6.86\%$, $\text{CSI} = 83.58\%$. 综合两类天气预报情况, 准确识别出的样本共 323 个, $\text{POD} = 91.16\%$, $\text{FAR} = 8.84\%$, $\text{CSI} = 83.76\%$, 如表 7 所示.

表 7 识别结果

	POD/%	FAR/%	CSI/%
冰雹	93.29	10.69	83.92
暴雨	89.07	6.86	83.58
综合	91.16	8.84	83.76

利用传统的 PUP 系统对本文的测试样本进行识别, 统计结果为: $\text{POD} = 91.89\%$, $\text{FAR} = 85.77\%$, $\text{CSI} = 14.05\%$. 将传统 PUP 系统的统计结果与表 7 中的冰雹统计结果进行比较, 不难发现: 在击中率上传统 PUP 系统方法略高于本文方法, 但其误警率远高于本文方法, 因此从综合成功指数角度分析, 本文方法优势明显, 总体效果优于传统 PUP 系统方法.

9 结 论

本文从天津市气象台的 2004—2010 年的历史资料中, 选取 85 个冰雹过程和 81 个暴雨过程共 1206 个样本 (测试样本与训练样本没有交叉) 的雷达基数据进行分析和处理, 可以得到以下结论.

1) 将雷达基数据中的反射率强度转化为反射率图像, 并且利用图像特征实现对强对流天气中冰雹和暴雨的识别具有一定的有效性.

2) 基于强度直方图的强度特征和基于空间灰度共生矩阵的纹理特征对于冰雹和暴雨天气的分类识别具有较好的效果. 就单个特征而言, 对冰雹和暴雨两类回波均具有较好的区分能力, 但两类之间会存在一定的重叠. 图 10—12 所示结果表明, 不

同特征之间具有一定的互补性, 故随着特征维数的增加, 冰雹和暴雨之间的分类效果会逐渐增强.

3) 除了强度特征和纹理特征, 本文还提取了单体回波顶高度, 得到了回波顶高、 0°C 与 -20°C 温度层高度三个特征与冰雹和暴雨天气的关系, 体现了它们在冰雹暴雨的识别与分类过程中的重要作用.

4) 利用粗糙集理论, 对提取出的 13 个特征进行数据挖掘, 建立了基于图像特征的冰雹暴雨识别的客观模型. 通过标准投票决策方法, 在对强对流天气中冰雹和暴雨的识别过程中取得了良好的效果 (如表 7 中, 冰雹 $\text{POD} = 93.29\%$, $\text{FAR} = 10.69\%$, $\text{CSI} = 83.92\%$; 暴雨 $\text{POD} = 89.07\%$, $\text{FAR} = 6.86\%$, $\text{CSI} = 83.58\%$).

本文方法主要是利用雷达回波反射率因子产品, 而借助雷达的径向速度产品可以获取垂直速度、中层径向辐合等信息, 将其与本文方法有机结合可以完善冰雹暴雨天气的客观识别模型, 提高识别模型的准确性, 这也是本文后续工作的重点.

参考文献

- [1] Rinehart R E 1981 *Atmos. Technol.* **13** 119
- [2] Rinehart R E, Garvey E T 1978 *Nature* **273** 287
- [3] Austin G L, Bellon 1982 *Nowcasting* (London: Academic Press) pp177–190
- [4] Rosenfeld D 1987 *J. Atmos. Ocean. Technol.* **4** 422
- [5] Michael D, Gerry W 1993 *J. Atmos. Ocean. Technol.* **10** 785
- [6] Johnson J T, Pamela L 1998 *Wea. Forecasting* **13** 263
- [7] Zhao Y L, Huang S X, Du H D, Zhong J Q 2011 *Acta Phys. Sin.* **60** 079202 (in Chinese) [赵延来, 黄思训, 杜华栋, 仲跻芹 2011 物理学报 **60** 079202]
- [8] Wang P, Pan Y 2013 *Acta Phys. Sin.* **62** 069202 (in Chinese) [王萍, 潘跃 2013 物理学报 **62** 069202]
- [9] Huang S X, Zhao X F, Sheng Z 2009 *Chin. Phys. B* **18** 5084
- [10] Hu W D, Li Y C, Zheng G F, Ding J J, Yang Y L, Sun Y C 2007 *J. Desert Res.* **27** 331 (in Chinese) [胡文东, 李艳春, 郑广芬, 丁建军, 杨有林, 孙银川 2007 中国沙漠 **27** 331]
- [11] Zhang J G, Wang Y, Zhou J L, Yue Y 2008 *Torrential Rain and Disasters* **27** 326 (in Chinese) [张家国, 王珏, 周金莲, 岳阳 2008 暴雨灾害 **27** 326]

- [12] Liu L P, Xu B X, Wang Z J 1991 *Plateau Meteorol.* **10** 317 (in Chinese) [刘黎平, 徐宝祥, 王致君 1991 高原气象 **10** 317]
- [13] Wang J 2008 *Ph. D. Dissertation* (Beijing: Graduate University of Chinese Academy of Sciences) (in Chinese) [王瑾 2008 博士学位论文 (北京: 中国科学院研究生院)]
- [14] Wang P, Niu Z Y 2014 *Acta Phys. Sin.* **63** 019201 (in Chinese) [王萍, 牛智勇 2014 物理学报 **63** 019201]
- [15] Yu X D 2011 *Adv. Meteorol. Sci. Technol.* **1** 30 (in Chinese) [俞小鼎 2011 气象科技进展 **1** 30]
- [16] Yu X D 2006 *Doppler Radar Principle and Business Applications* (1st Ed.) (Beijing: China Meteorological Press) p3 (in Chinese) [俞小鼎 2006 多普勒天气雷达原理与业务应用第一版 (北京: 气象出版社) 第3页]
- [17] Liao Y F, Pan Z X, Guo Q 2006 *Sci. Meteorol. Sin.* **26** 564 (in Chinese) [廖玉芳, 潘志祥, 郭庆 2006 气象科学 **26** 564]
- [18] Chen Q P, Chen Q C, Liu J X, Li B L 2011 *Meteorol. Sci. Technol.* **39** 545 (in Chinese) [陈秋萍, 陈齐川, 刘锦绣, 李白良 2011 气象科技 **39** 545]
- [19] Xie X N, Liu X D 2010 *Chin. Phys. B* **19** 109201
- [20] Zhang Y J 2004 *J. Image Graph.* **9** 513 (in Chinese) [章毓晋 2004 中国图象图形学报 **9** 513]
- [21] Yue P, Liu X Y, Guo L C, Lü C H 2005 *Arid Meteorol.* **23** 50 (in Chinese) [岳平, 刘晓云, 郭良才, 吕长虹 2005 干旱气象 **23** 50]
- [22] Tian Y Q, Guo P, Lu H Q 2004 *Comput. Sci.* **31** 162 (in Chinese) [田艳琴, 郭平, 卢汉清 2004 计算机科学 **31** 162]
- [23] Wang G H, Jiang P 2004 *J. Tongji Univ.* **32** 246 (in Chinese) [王光宏, 蒋平 2004 同济大学学报 **32** 246]
- [24] Pawlak Z 2002 *Inform. Sci.* **147** 1
- [25] Pawlak Z 1998 *Cybernetics and Systems* **29** 661
- [26] Han Z X, Zhang Q, Wen F S 1999 *Control Theory and Applications* **16** 153 (in Chinese) [韩祯祥, 张琦, 文福拴 1999 控制理论与应用 **16** 153]
- [27] Wang W S 2011 *Artificial Intelligence* (1st Ed.) (Beijing: Posts and Telecom Press) p139 (in Chinese) [王万森 2011 人工智能 (第一版) (北京: 人民邮电出版社) 第139页]

Recognition of hail and rainstorm based on the radar reflectivity image features*

Lu Zhi-Ying^{1)†} Liu Hai¹⁾ Jia Hui-Zhen²⁾ Yin Jing¹⁾

1) (*Department of Electrical Engineering and Automation, Tianjin University, Tianjin Key Laboratory of Process Measurement and Control, Tianjin 300072, China*)

2) (*Tianjin Meteorological Bureau, Tianjin 300074, China*)

(Received 17 January 2014; revised manuscript received 16 May 2014)

Abstract

Based on radar reflectivity image features, an automatic recognition method is proposed to identify the hail and rainstorm. We extract the image features of hail echo areas and rainstorm echo areas from radar reflectivity images. By analyzing both the differences in single feature between hail and rainstorm and the classified complementarity among different features, we determine the effective image features, including the intensity and texture features, to identify the hail and rainstorm. The hail and rainstorm objective recognition model can be established through the data mining of the extracted sample features and sounding data by using the Rough Set Theory. Through the test and identification of the 362 test samples, the hit ratio of hail reaches 93.29% and the hit ratio of rainstorm reaches 89.07%. The false alarm ratios of hail and rainstorms can be also at a low level. Compared with those from the PUP system, the experimental results from the present system show that it has a good effect to identify and classify hail and rainstorm by using the radar reflectivity image features.

Keywords: features extraction, Rough Set Theory, data mining

PACS: 92.60.Nv, 92.60.Qx, 92.60.Wc

DOI: [10.7498/aps.63.189201](https://doi.org/10.7498/aps.63.189201)

* Project supported by the Natural Science Foundation of Tianjin, China (Grant No. 14JCYBJC21800) and the Meteorological Key Technology Integration and Application Item, China (Grant No. CMAGJ2013M02).

† Corresponding author. E-mail: luzy@tju.edu.cn