

文章编号: 1007-4619(2006)02-0184-07

多光谱卫星云图的高维特征聚类与降水天气判别

洪 梅¹, 张 韧^{1, 2}, 孙照渤²

(1. 解放军理工大学 气象学院海洋气象系, 江苏 南京 211101; 2. 南京信息工程大学 大气科学博士后流动站, 江苏 南京 210044)

摘 要: 基于静止气象卫星(GMS-5)多光谱云图的天气采样数据, 分别对各样本数据在红外、水汽及可见光通道的灰度、梯度和纹理高维特征空间的投影点进行聚类分析, 以确定诸天气样本在特征空间中的类属区域, 进而用其对云图进行天气区的判别分类。针对传统聚类方法存在的缺点, 本文采用了模糊 C 均值聚类(FCM)、遗传算法(GA)和模糊减法聚类(FSC)相互交叉、优势互补的思想, 既克服了 GA/FCM 算法局部/全局寻优的不足, 又可客观确定出聚类中心数目。对高维特征空间中的重叠和交叉部分的样本点类属, 通过计算其与空间中各聚类中心点的欧氏距离来予以甄别, 最后得到高维特征空间中各天气的类属域, 实况云图中诸像素点通过计算和判断其灰度梯度特征量在高维空间中的投影点落区位置, 即可确定其天气类属, 进而实现对天气区的自动分类。试验结果表明, 该方法具有良好的分类效果, 判别结果与天气实况基本一致。

关键词: 卫星云图; 天气判别; 遗传算法; 模糊 C 均值聚类; 减法聚类

中图分类号: P407/TP751.1 **文献标识码:** A

A High-dimension Feature Spaces Clustering and Corresponding Weather Classification for Multi-spectral Satellite Images

HONG Mei¹, ZHANG Ren^{1, 2}, SUN Zhao-bo²

(1. Institute of Meteorology, PLA University of Science and Technology, Jiangsu, Nanjing 211101, China;

2. Nanjing University of Information Engineering, Jiangsu, Nanjing 210044, China)

Abstract: According to weather sampling data from static GMS-5 images, the projections of IR₁, IR₂, VSWV in high-dimension feature spaces such as gray degree, grade degree and veins can be clustered. In this way, we can get to know the subject area of each weather sample in the feature spaces, so that we can get the weather classification of each nephogram. In view of the disadvantages of the conventional clustering algorithm, we develop an idea to combine FCM, GA with FSC mutually. In this way, we can not only overcome the local/the global optimum of GA/FCM algorithm, but also confirm the number of clustering centers objectively. Especially to estimate the classifications of the overlapped samples in high-dimensional feature spaces, we can calculate the distance between these samples and the clustering centers to determine their classifications. The type of the pixels in original cloud images can be found out which group in high-dimensional feature spaces the pixels belong to. So that we can make sure its weather area to accomplish the automatic classifications of the weather area. A lot of experimental test to our method has shown good classification effect and the estimated outcome basically conforms to the weather fact.

Key words: satellite image; weather classification; genetic algorithm; FCM; subtractive algorithm

收稿日期: 2004-08-09; 修订日期: 2004-11-01

基金项目: 卫星课题(参作气字 2002 第 35 号)及国家自然科学基金(40375019)资助。

作者简介: 洪梅(1982-), 女, 解放军理工大学气象学专业本科毕业, 现为该校气象学专业硕士研究生, 主要研究方向为大气与海洋

卫星遥感应用, 已发表论文 3 篇。E-mail: flowerainhm@126.com.

1 引 言

静止气象卫星可对多种尺度的天气系统进行有效的全天候同步监测,已成为大气信息获取和气象要素反演的重要手段。基于静止卫星云图的降水判别,国外已发展出一些有效的方法途径^[1,2],如云指数方法、云生命史方法、双光谱方法等。云指数方法是一项较为简单、但便捷可行的降水估计方法。其基本思想是:一个区域的降水总量为覆盖该区域的云体面积和云类的函数。该方法存在的主要问题,一是认为云类相同则降水指数相同;二是估计对流性降水时,不能区分出主要降水中心;三是没有考虑降水的其他条件。云生命史方法的基本思想是在可见光和红外云图上,愈亮和面积愈大的地区,表示云层厚且上升运动强、大气中水汽丰富,产生的降水也越大,因而可根据云的亮度和面积判断和估计降水。其缺点是仍然没有考虑相同云类在不同发展阶段时降雨强度的变化。双光谱方法^[3]综合应用了红外和可见光双光谱卫星云图信息,可更合理地确定降水强度,减少了虚警率。但该方法主要用于区分雨区与非雨区,且难应用于无雷达探测资料的地区,其应用性受到很大限制。

国内的相关研究工作包括陈渭民、郁凡基于云顶温度和反照率等,通过建立云类降水回归方程,用云图估计梅雨锋降水^[4]和用双光谱云图阈值和单位特征空间对降水进行分类^[5];王立志等^[6]在静止卫星云图云分类基础上,根据一维云模式得到的对流云顶核心温度与降水间的关系,对层云和对流云进行了降水估计;李伟钢等^[7]采用小波分析对卫星图像进行预处理,结合气象雷达资料,用神经网络进行了降水量估算试验;师春香等人的图像分割与人工神经网络相结合的降水分类研究等^[8]。

由于静止气象卫星分辨率相对较低,且只有红外、水汽和可见光等有限的光谱通道,可资利用的信息较少,特别是缺少微波通道的信息,这就大大限制了静止卫星云图在天气、特别是降水分析中的应用。但静止气象卫星具有时间分辨率高、应用广泛和信息存储传输方便等优点,因此积极开展静止卫星云图中降水等天气的特征识别和分类研究,不仅有学术探索意义,更有实际应用价值。本文拟基于云图天气样本,通过构造高维特征信息空间和采用模糊 C 均值聚类、遗传算法与模糊减法聚类相结合的方法,对多光谱卫星云图进行天气类别的分析判别研究。

2 样本特征空间

与云分类(主要表现为形态学特征判别)不同,降水等天气是大气中多种热力和动力因子共同作用的产物。在云图上则表现为多种严格的条件约束,常规的低维(<3)特征空间聚类分析,揭示的多是该类天气的云图类属必要条件,而非充分必要条件。因此,要从云图上分割和判别天气,需要引入更多的云图有效信息,构造更高阶的灰度、梯度和纹理特征空间,采用基于地面观测资料和同时次卫星云图,并迭加地理定位显示的多光谱云图天气采样系统,对 1998—1999 年夏季(6—8 月)雾、小雨、中雨、大雨和雷暴等天气共采集约 3000 个 4×4 像素的 GMS-5 红外 1、红外 2、水汽和可见光云图样本(各天气类样本数不等)。

相关研究和试验分析表明,在可用的云图信息中,IR1、IR2、VS、WV 云图样本的灰度均值和梯度值以及 IR1—IR2 灰度差值共计 9 个特征量能够较好地描述和刻画不同天气特征。考虑到计算量和运算速度等情况,选取以下特征量分别构成白天和夜间的特征聚类空间:

白天:IR1 灰度、IR1 梯度、VS 灰度、VS 梯度、WV 灰度和 IR1—IR2 灰度差值等 6 个特征量构成六维特征空间。

夜间:由于缺少可见光信息,故选择 IR1 灰度、IR1 梯度、WV 灰度和 IR1—IR2 灰度差值等 4 个特征量构成四维特征空间。

3 研究技术路径

模糊 C 均值聚类(Fuzzy C Means FCM)是一种局部搜索和目标逼近能力强、应用广泛的聚类算法。由于各类天气之间并没有严格的界限,加上观测资料本身和采样过程中的一些误差,因此 FCM 算法及其隶属度概念,可合理地用于刻画和描述上述模糊性特征。但 FCM 存在两个固有缺点:迭代收敛过程中的误差局部极小和聚类结果对初值的敏感性。为此,人们提出割平面、分支定界、神经网络等数学规划解决方案,但效果均不理想。遗传算法(Genetic Algorithm, GA)是一种基于生物自然选择与自然遗传机理的全局优化算法,具有全局搜索优势,但其局部优化能力有限。因此,FCM 与 GA 算法存在很好的优势互补性。但是,无论 FCM 算法还是 GA 算

法,都需要预先人为地给出聚类数目,这既缺乏充分的客观性,也不利于云图天气的自动分类实现。为此,我们可以借助于模糊减法聚类(Fuzzy Subtractive Clustering, FSC)。FSC是一种快速的单次(One-pass)聚类算法。该方法将每个数据点均作为可能的聚类中心,根据该数据点周围的样本数据点密度来估算该点作为聚类中心的可能性。被选为聚类中心的数据点周围具有最高的样本数据分布密度,同时该点附近的数据点将被排除作为聚类中心的可能;第一个聚类中心选出后,从剩余的可能作为聚类中心的数据点中,继续采用类似方法选择下一个聚类中心。该过程一直持续到所有剩余数据点作为聚类中心的可能性低于某一阈值时为止。减法聚类的上述思想使其适宜于进行实时云图的云类数的客观估算。但由于减法聚类只是单次判别算法,因此用它来搜索和逼近各聚类中心的最佳位置的效果并不好。

综合 FCM, GA 和 FSC 方法各自的优势和缺点,我们本着优势互补、取长补短的思想,提出了下述技术途径和解决方案:首先采用 FSC 方法来客观估算和确定出各天气样本在特征空间中的聚类数;之后利用 GA 全局搜索优势进行聚类分析,确定出具有较好全局结构特性的聚类结果;随后再利用 FCM 局部寻优特性对 GA 聚类结果进行调整优化(GA 搜索出的聚类中心替代常规 FCM 算法中随机选取的初始聚类中心);最后综合确定出各类天气的聚类中心和相应的特征空间类属域。模糊 C 均值聚类、遗传算法和模糊减法聚类均是成熟的计算方法,其模型结构和算法原理可参考相关的文献专著,不再赘述。

4 特征聚类分析

由于大气运动的复杂性、观测资料的误差及采样过程中不可避免的人为因素,各类天气样本在特征空间中的投影点表现得较为散乱,若未经处理即映射于同一空间中,则特征投影点交叉重叠,变得杂乱无章,根本无法进行聚类分析。

为此,我们基于第 3 节所述的研究技术路径,首先进行各类天气样本的独立聚类分析,在此基础上将各类天气独立的聚类结果进行集成并作必要的分割处理,最后得到高维特征空间中各类天气的隶属区域(各类天气可以有多个独立的特征隶属区域),并将其作为实况云图中天气分类的判别依据。

第一阶段。先对各类天气样本在高维特征空间

中的投影点集单独进行聚类,以找出其空间结构的分布特征,确定出各类天气在特征空间中的聚类中心和隶属范围,并通过设定适当的临界隶属度,将远离聚类中心的疏散点作为噪声去除。具体实施步骤为:

(1)首先对特征空间中的样本数据进行模糊减法聚类,客观地确定出聚类中心数目。经减法聚类确定的各类天气的聚类中心为:雾和小雨各 3 类;中雨、大雨和雷暴各 2 类。上述多聚类中心现象表明,不同天气的产生具有多种物理机制和发生途径,云图上相应表现为多特征量组合的多个聚类中心族。

(2)基于减法聚类确定的聚类数,引入遗传算法分别对各天气样本进行聚类搜索,确定出对应各聚类中心的隶属区域范围,确定适当的临界隶属度值,去除边远零散点。遗传算法的聚类结果从全局意义上较好地找出和确定了各天气样本的聚类中心和隶属区域范围,但聚类的局部效果仍有改进与完善的余地和必要。

(3)采用模糊 C 均值聚类方法对 GA 的聚类结果进行调整,即用 GA 确定的聚类中心替代常规 FCM 算法中随机选取的初始聚类中心,作进一步的聚类优化。

经上述步骤处理后,可得到 5 类天气分别对应的 5 个高维特征空间及空间中表现各类天气属性的若干聚类中心和隶属域。显然,这些独立的特征空间和类属区域尚不能用于实况云图的天气划分,因此必须将它们放入同一特征空间中进行综合集成,并作必要的界定和分割处理。

第二阶段。将上述 5 类天气对应的 5 个独立特征空间以及空间中各类天气的聚类中心和隶属域合并于同一特征空间,旨在最终界定和分割出各类天气在同一高维特征空间中各自的类属区域范围。为便于直观显示,高维特征空间合并后的各类天气样本的聚类中心和隶属域,分别被分解为若干三维图像形式表示。分析结果表明,同一高维特征空间中,5 类天气样本的聚类中心位置彼此相互分离,各类天气的隶属域基本处于相对独立的位置(图 1)。这表明所选的云图特征量能够较好甄别和区分 5 类天气,所建立的特征空间聚类结果具有较好的代表性和可分性。由于各类天气之间不同程度地存在一些样本重叠(图 1),对这些分界交叉域重叠点的类型归属,我们计算和比较了该点与邻近聚类中心的欧式距离长短,并以最短距离的聚类中心作为该点的天气类属。处理后各类天气的聚类区完全独立(图 2)。

邻界重叠点经甄别修订后的高维特征空间中,

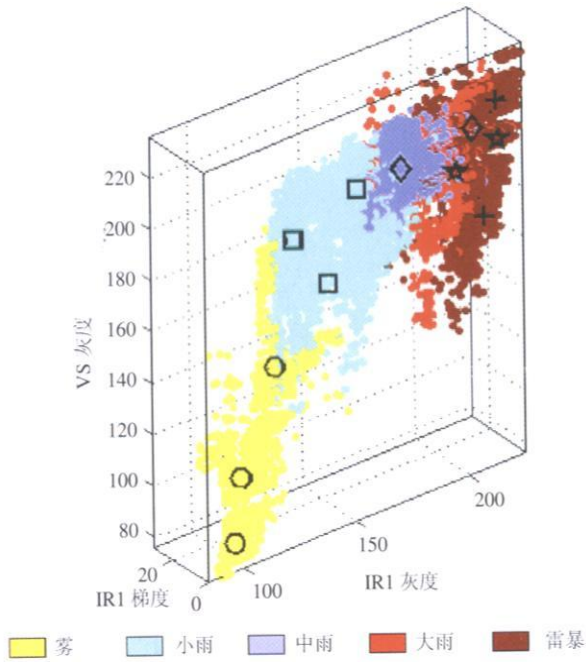


图1 集成的IR1灰度-IR1梯度-VS灰度天气聚类图
 Fig.1 Composite weather clustering pattern with IR1-gray,IR1-grads,VS-gray

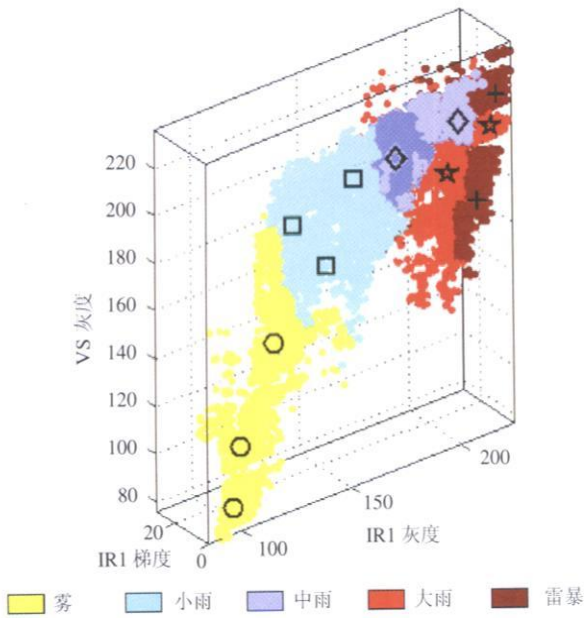


图2 修订后的IR1灰度-IR1梯度-VS灰度天气聚类图
 Fig.2 Revised pattern of Fig.1

其他特征量分解结构中的各类天气的聚类区,同样可以完全独立。由于夜间可见光云图信息不可用,但 IR1灰度、IR1梯度、WV灰度和 IR1-IR2灰度差值等 4 个特征量构成的四维特征空间,对各类天气仍表现出较好的可分性,因此仍然可以应用于夜间天气的判别分析。

经以上两个阶段处理后所得的高维特征空间中,5类天气合成的聚类中心和隶属区域即可作为卫星云图天气划分的判据。实际应用时,通过计算实时云图中逐像素点的红外、水汽和可见光通道的灰度、梯度值,并判断其特征值集映射点在特征空间中落区位置的属性,即可判定该像素点的天气类性,进而实现基于卫星云图的天气区域判别划分。

5 天气分类试验

为检测建立的高维特征空间聚类模型的天气分类效果,选择两次较为典型的天气个例进行判别分析。图 3和图 5分别是 1999-07-08 14: 32(白天,个例 A)和 1999-07-11 20: 32(傍晚,个例 B)等角投影的 GMS-5红外 1云图 (R2、VS、WV云图略),为验证模型分类效果,云图中还叠加了实况天气的观测值(注意:所叠加的实况天气为 14时和 20时正点,观测值与云图存在约半小时的时差)。

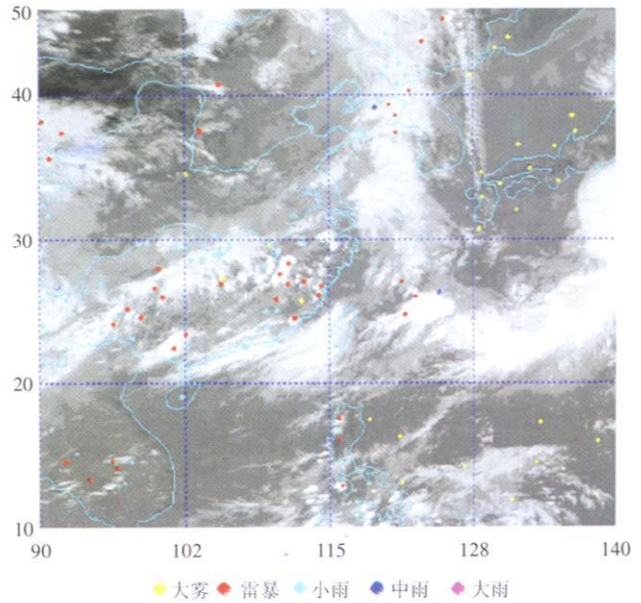


图 3 个例 A 的红外 1云图及对应的天气实况
 Fig 3 Infrared satellite image of sample A and corresponding actual weather

分别计算四通道云图中逐像素点的 6 个特征值 IR1,梯度 IR1,VS,梯度 VS,WV和 IR1-IR2(个例 A)和三通道云图中逐像素点的 4 个特征值 IR1,梯度 IR1,WV,IR1-IR2(个例 B)。将上述特征值构成的高维点集分别投影到前面建立的特征空间天气分类模型中,通过判断其落区位置进行天气类型甄别。个例 A、B 的模型分类判别结果分别如图 4和图 6。

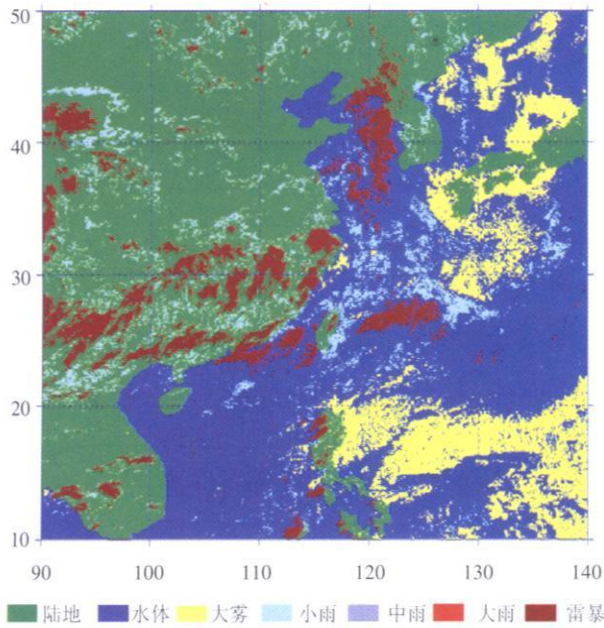


图4 个例A云图的天气判别分类结果
Fig.4 The weather classification of sample A

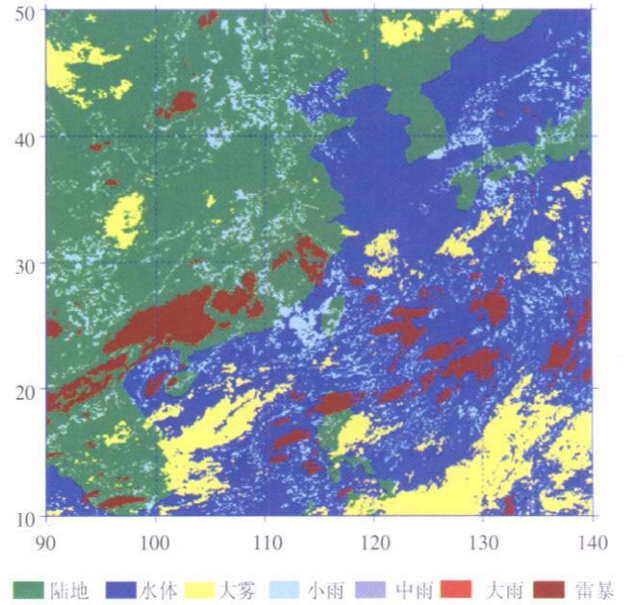


图6 个例B云图的天气判别分类结果
Fig.6 The weather classification of sample B

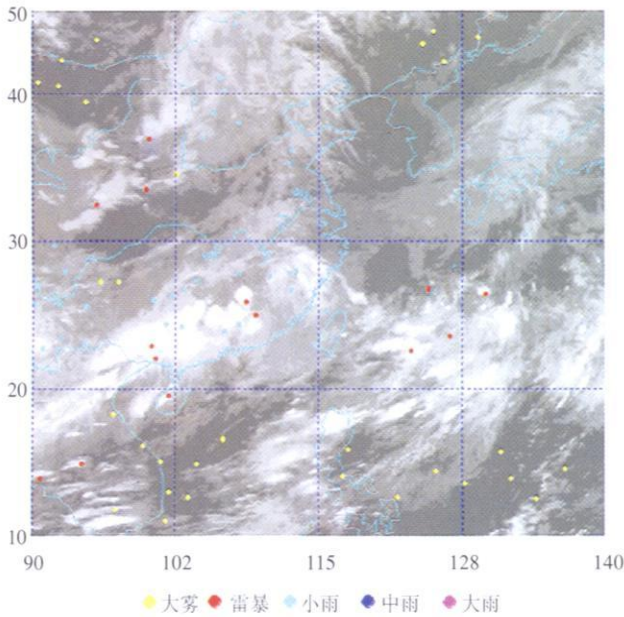


图5 个例B的红外1云图及对应的天气实况
Fig.5 Infrared satellite image of sample B and corresponding actual weather

实况天气与分类模型判别输出的对比结果表明, 个例 A, B 的模型分类结果, 基本上能够正确描述和定位出主要的天气区域, 模型划分结果符合天气实际。

此外, 为定量检测高维聚类模型的天气分类效果, 我们又选择了 1999 年夏季 (6—8 月) 多个云图个例及其相应的若干站点天气实况进行了判别和对照比较, 得到的统计分析结果如表 1 和表 2。

表 1 天气分类判别结果统计 (白天)

Table 1 The statistical analysis of weather classification (in day)

	大雾	雷暴	小雨	中雨	大雨
	共 281 个 站点	共 67 个 站点	共 249 个 站点	共 9 个 站点	共 4 个 站点
大雾	268	0	9	0	0
雷暴	0	66	0	0	0
小雨	13	0	231	4	0
中雨	0	1	5	5	0
大雨	0	0	0	0	2
分类准确率 /%	95.37	98.51	92.77	55.56	50

白天的分析结果表明, 对大雾, 雷暴, 小雨等发生几率较大的天气现象, 模型分类准确率较高 (如大雾、雷暴和小雨的分类准确率均达到 90% 以上)。而对中雨、大雨这些发生几率较小的天气现象, 模型的分类准确率较低。其原因与此类天气发生较少, 相应的云图天气样本较少, 以致在模型聚类空间中未能形成充分合理划分该类天气的聚类域有关。但就总体分类效果而言, 这种聚类方法的天气分类结果仍是切实可行的。

夜间的天气分类准确率普遍低于白天, 大雨、中

表 2 天气分类判别结果统计(夜间)

Table 2 The statistical analysis of weather classification (in night)

	大雾	雷暴	小雨	中雨	大雨
	共计 218 站点	共计 54 站点	共计 125 站点	共计 4 站点	共计 3 站点
大雾	199	0	16	0	0
雷暴	0	40	0	0	0
小雨	23	3	105	1	0
中雨	0	0	0	2	1
大雨	0	0	0	0	1
分类准确率 /%	91.28	74.07	84	50	33.33

雨判别率依然较差。但夜间大雾、雷暴、小雨等天气的分类准确率相对较高(均在 70% 以上,大雾和小雨达到 80% 以上)。缺少了可见光云图这个重要的信息源是导致夜间天气分类效果低于白天分类效果的重要原因。尽管如此,夜间的大雾、雷暴、小雨等天气的分类结果仍有较好的实用意义。

上述判别分析中存在的不足和误差原因主要表现在以下方面:

(1)表 1和表 2中仅分析比较了模型天气分类与站点实况天气之间的对比(即实况天气的命中率),由于云图范围内的大部分区域(尤其是海洋地区)缺乏测站或部分时效报文缺测,相当部分的分类结果无法与实况天气进行对比,即缺乏天气判别的虚警率分析。

(2)由于试验样本的实况天气均为正点报文,而云图均为准半小时接收,因此两者存在约半小时时差,对于夏季一些发展或移动较为迅速的对流云团天气,其判别结果的对照检测可能会有一些位相偏差。

(3)雾、降水和雷暴等天气是大气中动力、热力过程共同作用的产物,具有复杂的物理机理和成因。本文仅从云图形态学的角度进行了特征提取(而静止气象卫星有限的光谱通道信息,尤其是缺少微波通道,使得这种特征提取尚不充分)和分类识别,缺少必要的内在机理支持,因此改进和完善天气分类效果,必须进一步扩展和挖掘有效信息,并在模型中注入云物理机理的考虑,这也是我们下一步努力的方向。

6 结 论

基于 GMS-5 静止气象卫星多光谱云图天气采样数据库,构造了红外、水汽和可见光等多通道云图信息映射的高维特征空间,通过各类天气对应的云图样本数据在上述灰度、梯度和纹理等高维特征空间投影点的聚类分析,确定各天气样本在特征空间中的类属区域,进而用其对云图进行天气区的判别分类。针对传统聚类方法存在的缺点,本文采用了模糊 C 均值聚类、遗传算法和模糊减法聚类相互交叉、优势互补的思想和途径;对合成后高维特征空间中重叠和交叉部分的样本点属性,通过计算其与空间中各聚类中心点的欧氏距离来予以确认,最后得到高维特征空间中各类天气的隶属域判据。实际应用时,通过计算和判断多光谱云图中逐像素点的灰度梯度特征值映射点在特征聚类空间中的落区位置类属,即可确定其天气类性,进而实现对天气区(小雨、中雨、大雨、雷暴、大雾等五类)的自动分类。实况云图的试验结果表明,该方法能够较为客观准确地定位和划分云图中主要的天气区域,判别结果基本符合天气实际。

本方法与引言中所述的目前较为通用的静止气象卫星云图天气判别方法(如云指数方法、云生命史方法、双光谱方法和神经网络方法)相比较,其优点主要表现在以下方面:(1)可以对样本库中的采样误差进行有效修正和对混淆的天气样本进行分离,以保证典型天气的正确识别;(2)可以实时补充和增加天气样本,升级和更新天气分类模型;(3)可以对大雾、雷暴、大雨、中雨、小雨等多种天气现象进行同步分类判别,尤以这方面的优势表现得更为突出。

此外,本文的技术途径还可方便地移植应用于 FY-2 静止卫星及极轨卫星等多光谱云图资料的天气分类模型的建立,具有较好的普适性和实用性。

参 考 文 献 (References)

- [1] Koffler R, Decotius A G, Rao P K. A Procedure for Estimating Cloud Amount and Height from Satellite Infrared Radiation Data [J]. Mon Wea Rev 1973, 101: 240-243
- [2] Desbois M, Seze G, Szejwach G. Automatic Classification of Clouds on METEOSAT Imagery Application to High-Level Clouds [J]. J Appl Meteor 1982, 21: 401-402
- [3] Welch R M, Navar M S, Sengupta S K. The Effect of Resolution Upon Texture-Based Cloud Field Classification [J]. J Geophys Res, 1989, 94: 14767-14781

- [4] Chen W M, Yu F. A Primary Discussion of Estimating the Meiyu Front Precipitation Using the GMS Satellite Data [J]. Journal of Nanjing Institute of Meteorology, 1994, 17 (1): 79—85 [陈渭民, 郁凡. 用 GMS 数字化卫星资料估计梅雨锋降水的初步探讨 [J]. 南京气象学院学报, 1994, 17 (1): 79—85]
- [5] Yu F. The Characteristic Extraction from GMS Multi-Spectrum Satellite Images and Its Application in Mid-Scale Numerical Forecast Model [D]. Dr Degree Dissertation of Nanjing University, 1998 [郁凡. 多光谱 GMS 卫星图像气象特征量提取及在中尺度数值预报模式中的应用 [D]. 南京大学博士学位论文, 1998]
- [6] Wang L Z, Li J, Zhou F X. The Auto Classification of 4-Channels Satellite Images and Its Application in Precipitation Estimation [J]. J. Atmo. Sci Sinica, 1998, 22 (3): 371—378 [王立志, 李俊, 周凤仙. GMS-5 四通道云图的自动分类及其在定量降水估算中的应用 [J]. 大气科学, 1998, 22 (3): 371—378]
- [7] Li W G. A Research of Estimating Precipitation on Satellite Images by Using Wavelet and ANN Technology [J]. Journal of Nanjing Institute of Meteorology, 2000, 23 (2): 277—282 [李伟钢. 气象卫星云图的多分辨小波分解及神经网络降水估计研究 [J]. 南京气象学院学报, 2000, 23 (2): 277—282]
- [8] Shi C X, Wu R Z, Xiang X K. The Cloud Auto Separation Experiment on Satellite Images by Using Multi-Threshold and ANN Method [J]. Acta Applied Meteorology Sinica, 2001, 12 (1): 70—78 [师春香, 吴蓉璋, 项续康. 多阈值和神经网络卫星云图云系自动分割试验 [J]. 应用气象学报, 2001, 12 (1): 70—78]