#### Vol. 60, No. 2 April 2002

## 用神经网络方法对 NOAA-AVHRR 资料 进行云客观分类

## 师春香 瞿建华

(中国气象局国家卫星气象中心, 北京, 100081)

#### 摘 要

利用 NOAA-A VHRR 5 个通道资料建立了 6 种云类以及陆地和水体的样本数据库,其中包括 8×8 象素样本和单象素样本。A VHRR 的 5 个探测通道都位于大气窗区,吸收物质少,比较透明,可以比较准确地反映探测表面的性质。理论分析和试验结果表明:除了不同性质的云在 5 个通道中有不同的表现外,通道之间的差别也可用于云分类。在理论分析和试验的基础上,对 8×8 象素样本库提取了包括光谱特征、灰度特征、通道差特征、灰度统计量和灰度直方图统计量特征在内的 80 个特征,并利用逐步判别分析方法进行特征筛选,共选出 20 个特征,用神经网络方法对 8 种类型云和地表样本数据库分类,选择网络结构为 20-40-15-4的 BP 网络,利用 3000 多个样本进行神经网络训练,并用其余的 3 万多个独立样本数据进行检验,测试正确率达 79%。类似地,对单象素样本数据,提取了包括光谱特征、灰度特征、通道差特征在内的 20 个特征,用神经网络方法对 8 种类型云和地表分类,选择网络结构为 20-40-15-4的 4 层 B-P 网络,利用 2000 多个样本进行神经网络训练,并用其余的 2 万多个独立样本数据进行检验,测试正确率达 78%。

设计并编写了实际云图客观云分类系统和软件,该系统输入为5个通道的 AVHRR 数据,可自动获取已分类的彩色或灰度图象,其中包括积雨云、浓积云、淡积云、卷云、中云和低云以及陆地和水体及其他类(不属于已定义的8类)共9类。

关键词: AVHRR资料, 云和地表分类, 神经网络。

#### 1 引 言

云对地气系统的辐射收支平衡有重大影响。云与天气气候也存在密切关系,不同的云类反映了不同的大气状况,如对流云的存在预示着可能发生降水或强对流天气,反映了大气的不稳定性;层积云的存在预示着大气中存在着干空气逆温层,这通常是由于天气尺度的环流下沉造成的。因而对云进行正确的分类有助于我们对天气气候变化的理解和正确预报。气象卫星所提供的资料可与中尺度数值模式的分辨率相匹配,在资料同化过程中,云类别、云顶高度和云量及其他云参数资料可用来同化背景场中有关云的信息,改进初始场,也可检验和订正数值模式的预报结果,有助于确定零、云和降水的覆盖范围。

## 2 AVHRR 资料云分类原理

NOAA 极轨气象卫星携带有 5 个探测通道 (CH1, CH2, CH3, CH4, CH5) 的 AVHRR 甚高分辨率扫描辐射仪, 5 个探测通道都位于大气窗区, 吸收物质少, 比较透明, 可以较准确地反映探测表面的性质, 但 也 存在着微弱吸收。CH1 (0. 58 ~ 0. 68  $\mu$ m)位于可见光谱段, 在这一光谱区有臭氧吸收带; CH2 (0. 73~ 1. 1  $\mu$ m)位于近红外光谱段, 包括了 H<sub>2</sub>O 吸收带。可见光云图是太阳辐射经地气系统散射或反射后到达卫星所得到的图像, 该通道的辐射值取决于地表或云的散射或反射系数。由地表和云的典型反照率值可知, 不同的地表(表 1) 和云(表 2) 在可见光波段的辐射是有差别的。另外对

<sup>\*</sup> 初稿时间: 2000 年 5 月 30 日; 修改稿时间: 2000 年 11 月 27 日。 资助课题: 国家重点基础研究项目《我国重大天气灾害形成机理和预测理论研究》(05-01-06)。

干无云的水面,由于分子和烟尘的的散射,使得短波 的后向散射增大,可见光波段(CH1)的反射率是近 红外光谱段(CH2)的两倍,在无云的陆地,CH2和 CH1 相比, 反射率显著增大, 即使下垫面是沙漠, 植 被稀少时也如此;有云时, R<sub>CH2</sub>/R<sub>GH1</sub>≈ 1(0.8~1.1 µm)。CH3 (3.55~3.93 µm)位于红外 波段,该通道的辐射由散射的太阳辐射和地气系统 发射的辐射组成,由于太阳辐射只存在于白天,因而 通道特性在白天和夜间是不同的。这一波长上反射 率强烈地依赖于粒子的大小和相态。CH4(10.3~ 11.3 µm) 和 CH5 (11.3~ 12.5 µm) 均位干红外波 段,辐射来自于地气系统的辐射,与下垫面或云顶 的温度有关。对于无云和不透明的云, 在忽略大气 吸收的情况下, 红外通道的亮温代表了地表和云顶 的温度。但对于半透明的薄云或未充满视场的小 云,卫星的红外亮温比云顶的实际温度高得多。由 于冰和水吸收之间最大的区别在 12 µm 附近, 因而 CH4 和 CH5 的亮温差可以用来区分液滴云和冰晶 云。理论分析和试验结果表明:除了不同性质的云 在5个通道中有不同的表现外,通道之间的差别也 可用于云分类。另外,由于某一种云在一个通道上 的反照率或亮温也随各种状况的不同而有一定的变 化, 而且云类之间本身存在着相互演变的中间状态, 用简单的门阈值方法很难实现对大量云图的正确分 类<sup>[1,2]</sup>。

表 2 云的反照率 表 1 陆表的反照率 地 表 굸 反照率(%) 反照率(%) 薄、碎云: 陆 30 Ci, Čs, Cc 14 35 植被沙地、沙漠 40 27 厚云: 74 冰海 雪: Ac, As, Sc 65 75 新 80

## 3 人工神经网络模型及算法

目前,人工神经网络理论在许多领域得到广泛的应用,人工神经网络模型也已达上百种,本文采用在模式识别领域得到广泛应用的多层前馈网络的B-P 算法(B-P 网络模型)进行云和地表分类试验<sup>[3~5]</sup>。B-P 网络模型是典型的多层网络,它包括输入层、隐层和输出层,层与层之间采用全互连方式,同一层单元之间不存在相互连接。图 1 给出了一个 3 层 B-P 网络结构模型。

网络中神经元的输入是上一层神经元输出的权

重和

$$n_j = \sum_i W_{ij} O_i$$

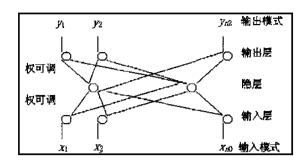


图 1 3层前馈神经网络结构

由网络神经元活化函数(一般为 sigmoid 函数)将权重和转换为神经元输出

$$O_j = 1/\exp(-n_j + \theta_j)$$

其中 () 为常数。

在训练过程中,用下式调整权重

$$\Delta W_{ii}(n+1) = \Re(\delta_i O_i) + \alpha \Delta W_{ii}(n)$$

其中,  $\Delta W_{ji}(n+1)$  为连接神经元 i 和j 的权重被修改后的值,  $\delta_j$  为误差修正项,  $\Pi$  为学习率,  $\alpha$  为动量项。

该网络实现了多层网络学习的设想,当给定网络一个输入模式时,它由输入层单元传到隐层单元, 经隐层单元逐层处理后再送到输出层单元,由输出 层单元处理后产生一个输出模式,这是一个逐层状态更新的过程,称为前向传播,若输出与期望输出模式有误差且误差不满足要求时,误差转入后向传播过程,将误差值沿连接通路逐层传送并修正各层连接权值,这就是误差后向传播的涵义。用给定的一组训练模式,重复对网络训练,也就是重复前向传播和误差后向传播的过程,当各个训练模式都满足要求时,多层前馈人工神经网络学习过程结束,并保留一组连接权系数。已训练好的神经网络模型可以对新的输入模式给出判别结果[6.7]。

## 4 数据库

#### 4.1 样本数据库建立

云和地表样本库的建立是在 Windows 环境下, 利用人机交互式采样软件工具, 由经验丰富的气象 专家根据经验, 目视识别, 获得各类别 NOAA – AVHRR 5 个通道资料的样本数据, 更详细的采样 过程见文献<sup>①</sup>。建立样本库所使用的 AVHRR 云图 的空间分辨率为 0.01°×0.01°, 等经纬度投影, 样本库中所有样本来源于夏季白天的 AVHRR 资料。原始的样本数据为 32×32个象素的方块, 也即样本覆盖面积为 32 km×32 km。考虑到分类结果将应用于中尺度数值预报模式资料同化中, 将 32×32个象素的样本转化为 8×8 象素样本和单象素样本, 每一个 8×8 象素样本对应一个单象素样本, 也就是说单象素样本库也基本涵盖了该类别夏季白天的多数情况。选择了 6 种云以及陆地和水体, 其中两种类型的样本数据量分别为 3 万多个和 2 万多个, 6 种云包括积雨云、浓积云、淡积云、卷云、中云和低云。由于统计识别方法需要大量的样本, 而神经网络分类方法则不同<sup>17</sup>, 因此样本库中的样本数量已能满足神经网络分类的需要。

#### 4.2 样本库合理性检验

样本数据库是由专家通过人机交互方式建立的,虽然已对样本库进行了检查,但还会存在一些人为的错误,因此根据物理意义去掉一些不合理的样本。本文对样本库作了简单的阈值处理,保留了 $CH4亮温(T_4)$ 小于-30  $^{\circ}$   $^{\circ}$ 

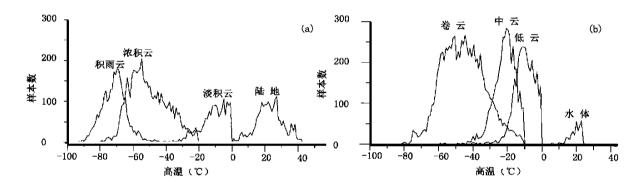


图 2 CH4 最低亮温分布图

(a. 积雨云、浓积云、淡积云和陆地 CH4 最低亮温分布图, b. 卷云、中云、低云和水体 CH4 最低亮温分布)

## 5 特征提取和筛选

#### 5.1 特征提取

在样本库和分类方法相对不变的情况下,特征选取的好坏决定了分类精度。在理解 AVHRR 5 个通道探测地表和云的理论基础和其通道值及其通道值差别所代表的物理意义的基础上,对单象素样本库提取了如下 20 个特征:

#### (1) 光谱特征

5个通道的光谱特征是分类最主要的特征, 包括 CH1 和 CH2 反照率( $R_1$ ,  $R_2$ ), CH3~ CH5 的亮度温度值( $T_3$ ~  $T_5$ )。

#### (2) 灰度特征

灰度图像主要用于人工定性分析卫星云图, 判释云类和地表的许多经验来自于专家定性分析卫星云图的结果。在分类试验中提取了  $CH1 \sim CH5$  的灰度值(  $G_1 \sim G_5$ )。

#### (3) 不同通道间光谱、灰度特征

已有科学家从理论分析、模式计算表明:  $T_3$  –  $T_4$  与  $T_4$  和  $T_3$  –  $T_5$  与  $T_4$  的关系曲线可用来区分液滴云和冰晶云等。在分类试验中提取了如下 10 个通道间光谱和灰度差别特征:  $T_2/T_1$ ,  $T_4$  –  $T_5$ ,  $T_3$  –  $T_4$ ,  $T_3$  –  $T_5$ ,  $(T_3$  –  $T_5)/T_5$ ,  $G_2/G_1$ ,  $G_4$  –  $G_5$ ,  $G_3$  –  $G_4$ ,  $G_3$  –  $G_5$ ,  $(G_3$  –  $G_5$ )/  $G_5$ 。

① 师春香,项续康等. A VHRR 资料云和地表样本库的建立, 921-2 民用遥感器大气遥感文集(第二集), 1996,国家卫星气象中心.

对 8×8 象素样本集提取了如下 80 个特征:

#### (1) 光谱特征

包括 CH1 和 CH2 反照率 ( $R_1$ ,  $R_2$ ) 在 8×8 象 素块中的平均值、最大值和最小值, CH3 ~ CH5 的 亮度温度值 ( $T_3$ ~  $T_5$ ) 在8×8象素块中的最大值、最小值、均值和方差。

#### (2) 灰度特征

提取了 CH1~ CH5 的灰度值( $G_1$ ~  $G_5$ )在 8×8 象素块中的最大值、最小值、均值和方差。

#### (3) 不同通道间光谱、灰度特征

提取了在  $8 \times 8$  象素块中 10 个通道间光谱和灰度差别反映出的特征均值:  $R_2/R_1$ ,  $T_4-T_5$ ,  $T_3-T_4$ ,  $T_3-T_5$ ,  $(T_3-T_5)/T_5$ ,  $G_2/G_1$ ,  $G_4-G_5$ ,  $G_3-G_5$ ,  $(G_3-G_5)/G_5$ 。

### (4) 灰度概率直方图统计量特征

设样本大小  $m \times n$  象素, 灰界  $0 \sim 255$ , N(k) 为灰界等于 k 的象素数, k = 0, 1, 2, ..., 255, 则样本窗口内灰度级直方图为:

$$P_{HS}(k) = N(k)/(m \times n),$$
灰度均值:  $A_{ME} = \sum_{k=0}^{255} k \times P_{HS}(k)$ 
标准差:  $S_{TD} = \left[\sum_{k=0}^{255} k^2 P_{HS}(k) - A_{ME}\right]^{V2}$ 
偏度:  $D_{EF} = (1/S_{TD}^6) \cdot \sum_{k=0}^{255} (k - A_{ME})^3 \cdot P_{HS}(k)$ 
峰度:  $E_{NT} = -\sum_{k=0}^{255} P_{HS}(k) \cdot \lg(P_{HS}(k))$ 
能量:  $E_{NE} = \sum_{k=0}^{255} P_{HS}(k)^2$ 
熵:  $K_{UR} = (1/S_{TD}^8) \cdot \sum_{k=0}^{255} (k - A_{VE})^4 \cdot$ 

每个通道有 6 个灰度直方图统计量, 5 个通道 共 30 个。人工识别云图时除了考虑亮度外, 还考虑 云的空间分布等特性, 这些特征反映了样本的灰度 分布特点。

 $P_{\rm HS}(k)$ 

#### 5.2 特征筛选

为减小神经网络模型的规模, 节约资源, 缩短计算时间, 对 8×8 象素样本库中所提取的 80 个特征进行特征筛选, 选出相对重要的特征, 即用逐步判别

分析方法, 对 3000 多个样本, 从 80 个特征中筛选出如下 20 个特征: CH4 亮温最小值,  $G_4$  –  $G_5$  均值,  $G_2/G_1$  均值,  $(T_3 - T_5)/T_5$  均值,  $G_3$  –  $G_4$  均值, CH1 灰度最大值, CH4 能量, CH1 反照率均值,  $(G_3 - G_5)/G_5$  均值, CH2 灰度最大值, CH1 能量, CH3 熵, CH1 反照率最大值, CH2 反照率最大值, CH4 亮温最大值, CH2 灰度直方图均值, CH4 峰, CH1 熵, CH5 亮温最小值, CH4 灰度最小值。用于网络训练。

对单象素样本库,选取了20个特征,特征数目相对较少,因而没有进行特征筛选。

## 6 分 类

本文采用有监督分类方法, 在样本集中选择了 10%的样本数据对神经网络模型进行训练。网络结构取决于输入特征数, 要分类的类别数, 隐层层数和 各隐层神经元数目的选择。网络输入层神经元个数对应输入特征的数目, 输出层对应要分类的类别数, 为了减小网络结构, 节省资源, 输出的 8 类由 4 个二进制数的组合表示, 见表 3。隐层层数和各隐层神经元数目的选择目前还没有成熟的理论, 在大多数的应用中根据经验和试验确定, 这方面还有待于进一步的研究。

表 3 8 类云和地表对应的理想输出模式

类别	云和地表类别	输出层理想输出值				
	ム和地衣矢別	$y_1$	$y_2$	$y_3$	y 4	
1	积雨云	0	0	0	1	
2	浓积云	0	0	1	0	
3	淡积云	0	0	1	1	
4	卷云	0	1	0	0	
5	中云	0	1	0	1	
6	低 云	0	1	1	0	
7	陆地	0	1	1	1	
8	水体	1	0	0	0	

分类试验中,在输入层和输出层以及训练样本不变的条件下,通过改变隐层数目和隐层神经元数目而改变网络结构,通过试验选择合适的结构。用样本库中随机抽取的 10% 样本训练事先构造好的网络模型,由其余的 90% 的样本进行测试检验。对8×8 象素样本和单象素样本,我们挑选出测试效果最好的网络结构都为: 20-40-15-4,表4为对单象素样本库测试的具体结果。分类总正确率达到77.8%。

表 4	分类试验结果
<del>7</del> 4	分尖压验结果

类别	1	2	3	4	5	6	7	8	9	分对样 本数	各类 总数	正确 率%
1	2207	157	25	335	21	0	0	0	45	2207	2835	77. 0
2	158	1119	41	457	0	0	0	0	158	1119	2091	53.0
3	29	3	684	76	133	223	121	0	0	684	1269	53.0
4	145	214	3	7186	435	0	3	0	194	7186	8374	85. 0
5	38	6	10	450	1822	1	47	0	13	1822	2400	75. 0
6	8	0	554	107	126	1648	413	0	0	1648	2856	57. 0
7	1	0	5	0	2	0	1935	33	6	1935	1982	97. 0
8	0	1	2	0	1	1	28	330	13	330	376	87. 0

注: 行为模式分类结果,列为实际类别

#### 7 应用试验

利用已训练好的云和地表神经网络分类器可对 夏季白天所有 NOAA-AVHRR 图像进行分类, 也可 对 FY-1C 图像进行分类。我们对若干 NOAA-11 和 NOA A-14 的 AVHRR 云图以及 FY-1C 云图做了分 类试验,并请气象专家做了目视检验。试验结果表 明:(1)用8×8象素神经网络分类器对图像分类结 果与用单象素神经网络分类器对图像分类的结果做 对比,后者比前者更好。虽然8×8象素样本比单象 素样本包含了更多的分类信息,可以提取空间分布 特征、纹理特征等,但在实际应用过程中,由于待识 别目标(8×8 象素块)的情况比单象素目标复杂的 多, 使得分类效果较差; (2) 多数薄卷云被识别为低 云。这是因为薄卷云的半透明特性使得卫星探测的 辐射除包含卷云的信息外, 还包含了下垫面部分信 息、因此它的亮温有时接近甚至高于低云亮温、且在 样本集中没有将薄卷云单独设为一类。我们将在方 法改进时加入薄卷云类。目前通过对低云类中反照 率低于 20% 的象素进行修正,海洋上薄卷云分类正 确率有较大提高,但陆地上提高不大:(3)由于样本

集中没有包含多层云系,因而实际云图分类中,多数多层云系被分为未知类别。我们将在方法改进时加入多层云系类: (4) 在水体和陆地交界处,有部分象素被分为未知类别。这是由于该象素为混合象素,象素本身就是由陆地和水体组成的。尽管对云图的实际分类还存在误差,但对水体、陆地、卷云、积雨云和中云等的分类正确率都比较高。应用试验结果表明,用单象素神经网络分类器对云图的分类结果在实际应用中是可行的。图 3a(见封三) 为利用单象素神经网络分类器对 NOAA-11 1992 年 6 月 20 日 05 时(世界时),范围  $30^\circ \sim 35^\circ$  N,  $140^\circ \sim 145^\circ$  E 云图分类结果,图 3b(见附页) 为同时次通道 4 云图(见附页)。

#### 8 小结

利用神经网络技术对 AVHRR 云图做了分类 试验, 结果表明: 基于象素的分类方法在实际应用中是可行的。在下一步的方法改进中, 第 1, 增加更多类别, 比如多层云系、薄卷云等; 第 2, 开发夜间无可见光云图时的分类方法; 第 3, 研究其他季节情况。

#### 参考文献

- 1 Michel Desbois, Genevieve Seze, Gerard Szejwach. Automatic classification of clouds on METEOSAT imagery: Application to high-level clouds. J Appl Meteor, 1992, 21: 401~412
- 2 Toassini C. T. Objective Analysis of Cloud Fields. In: Proceedings of the 2nd Course on Satellite Meteorology of the Mediterraneann 1980. 73~78
- 3 Atkinson PM, Tatnall AR L. Introduction: Neural networks in remote sensing. Int J Remote Sensing, 1997, 18(4): 699~710
- 4 Foody G M, Arora M K. Remote sensing image analysis using a neural network and knowledge-based processing. Int J Remote Sens, 1997, 18 (4): 811~828
- 5 Peak J E, Tag P M. Segmentation of satellite imagery using hierarchical thresholding and neural networks. J Appl Meteor, 1994, 33: 605~616
- 6 杨行峻, 郑君里. 人工神经网络. 北京: 高等教育出版社, 1992. 404pp
- 7 罗四维. 人工神经网络建造. 北京: 中国铁道出版社, 1998. 15~75

# CLOUD CLASSIFICATION FOR NOAA-AVHRR DATA BY USING A NEURAL NETWORK

Shi Chunxiang Qu Jianhua (National Satellite Meteorological Center, China Meteorological Administration, Beijing 100081)

#### Abstract

Sample database of clouds, land and water was built based on NOAA-AVHRR 5- channel data which include more than thirty thousand 8 × 8 pixel samples and more than twenty thousand single pixel samples. The five AVHRR channels are in the atmospheric windows which have a little absorptive mass and are rather transparent. They can show characters of detected surface well. Channel 1 (CH1) is in the visible wave range, CH2 in the near infrared wave range, and CH3, CH4, CH5 are in the infrared wave range. Theoretical analyses and experiments show that not only 5- channel data can be used to distinguish clouds, land and water, but also the difference between channels can do so. For example, CH4- CH5 can be used to distinguish water particle cloud and ice particle cloud because the biggest absorption difference between water particle and ice particle is near 12 \(\mu\)m. On the basis of theoretical analyses and experiments, 80 features were extracted from 5 - channel AVHRR data for 8×8 pixel samples, which involve spectrum feature, gray feature, channel difference feature, the gray statistical feature and the gray histogram statistical feature. 20 features were selected by using the distinguishing analysis step- by- step method. Classification experiment of sample database was done by using neural network method. We designed a neural network model with 20 inputs, 2 hidden layers and 4 outputs (20 -40-15-4). More then three thousand samples selected randomly were used to train the neural network model. The other independent samples were used to test. And we got a testing accuracy of 79%. Similarly, for single pixel samples, 20 features including spectrum feature, gray feature and channel difference feature were extracted from 5- channel AVHRR data. Classification experiment of sample database was made by using neural network method. We designed a neural network model with 20 inputs, 2 hidden layers and 4 outputs (20-40-15-4). More than two thousand samples selected randomly were used to train the neural network model. The other independent samples were used to test. And we got a testing accuracy of 78%. Some different disposals were used to 8 × 8 pixel samples database and single pixel sample database.

We designed and wrote AVHRR automatic cloud classification system and software. The inputs of this system were 5 AVHRR channel data and outputs were classified into gray image and classified color image. Classified image involves cumulonimbus, cumulus congestus, cumulus, cirrus, middle cloud, low cloud and land, water and unknown pixel.

**Key words:** AVHRR data, Classification of cloud and surface, Neural network.

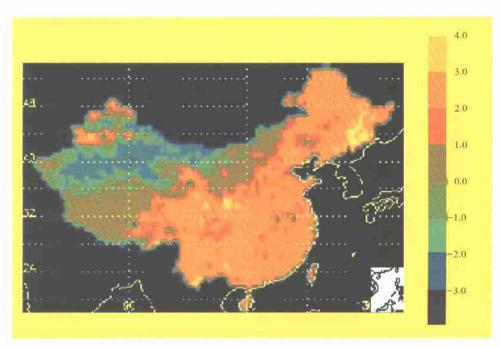


图 la, 归一化植被指数的 EOF1 空间场分布

#### 师春香等:用神经网络方法对NOAA-AVHRR资料进行云客观分类

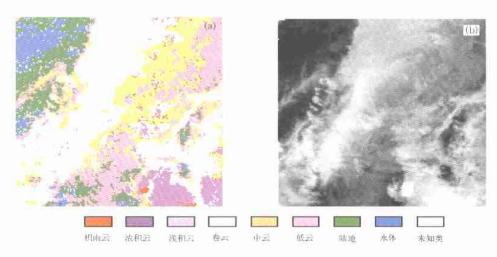


图 3 云和地表分类应用试验结果 (a.NOAA-11 1992 年 6 月 20 日 5 时,(UTC), 范博 30-35\*N, 140-145℃的分类例, b. 同时次通道 4 云图)