

# 多阈值和神经网络卫星云图云系自动分割试验\*

师春香 吴蓉璋 项续康

(国家卫星气象中心, 北京 100081)

## 提 要

卫星云图自动分割是实现卫星云图云系自动识别的基础. 选用 1992~1994 年和 1997~1998 年夏季有典型天气系统的 177 幅 GMS 红外云图建立了云系模型库, 云系分类样本 3079 个, 包含 16 类云系, 云系分割样本 2764 个. 利用云系分割样本集进行神经网络试验, 训练集为从 32 幅云图中抽取的 484 个样本, 测试集为从 145 幅云图中抽取的 2280 个样本, 神经网络模型训练正确率达到 98.8%, 测试正确率为 86.4%. 用 1997 年 7 月 18~21 日和 1998 年 6 月 15~17 日的两组卫星云图做自动分割应用试验, 结果经专家判识, 正确率达到 90% 以上. 本文的工作表明: 用多阈值和人工神经网络相结合方法对卫星云图进行云分割在实际应用中是可行的.

卫星云图自动分割系统的输入是 GMS 红外云图, 输出是分割出的每一个云区, 同时还包括云区的边界链码、起始点、周长、面积, 并保留了原始图像数据. 在下一步的云系识别过程中, 可以在此基础上进行云系分类识别试验.

关键词: 卫星云图 自动分割 人工神经网络 多阈值

## 引 言

卫星云图包含丰富的云信息, 从中可以形象直观地看到云系的形状和演变情况, 但是准确识别出与锋面、气旋、切变线、低压槽和强对流天气等对应的云系需要积累大量的经验. 在利用卫星资料进行降水估计和暴雨短时预报系统中, 也要求输入云图判释后的信息. 为使云图判识客观化、自动化, 扩展卫星云图在天气预报、环境监测中的应用, 进行云图定量自动判识研究是必要的. 云图自动判识包括云图分割和识别两部分. 近年来, 在卫星云图的自动分割方面已有若干研究工作, 王耀生等<sup>[1]</sup>使用数学形态学方法识别卫星云图上的台风云系. Michael Brewer 针对中尺度对流云系, 在像素分类基础上, 提出两种卫星云图自动分割方法<sup>[2]</sup>. Peak 与 Tag 提出多阈值启发式和人工神经网络相结合的方法, 将云图分割成与天气系统相联系的云区<sup>[3,4]</sup>.

卫星云图自动分割的难点在于, 一幅云图上常常有多种不同的云系存在, 而且云系的灰度随云系的种类、地域和时间而变, 同时云系的形状很不规则, 不能简单地用阈值或形状将不同云系分割开, 一般的图像处理方法也面临困难. 本文试验用多阈值和神经网络方法进行卫星云图云系自动分割, 研究发生在中国及附近海域上与夏季大、中尺度天气系统

\* 本研究得到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》(G1998040907) 资助.

1999-09-23 收到, 2000-03-22 收到修改稿.

有关的云系.

## 1 多阈值和人工神经网络卫星云图分割方法

卫星云图上云系的特点可以用图 1 来说明, 图 1 是 1998 年 6 月 20 日 23: 33UTC, GMS-5 红外 1 通道卫星云图, A 和 C 为温带气旋云系, B 为切变线云系, D 为季风云团, E 为热带云团. 云系由不同的云类组成, 如在切变线云系 B 中有高云、中云和低云等, 切变线 B 的北边界较光滑整齐, 而南边界却是模糊的. 温带气旋 A 也是由高中低云组成的, 边界模糊. 图 2.1~ 2.15 是在对云图进行平滑处理的基础上, 用 15 个不同的阈值分别对平滑后云图进行二值处理, 然后用跟踪算法得出云区的边界廓线. 从图 2 中可以看出, 切变线云系用第 8 个阈值分割最合适, 而温带气旋云系 A 用第 5 个阈值分割, 热带云团用第 4 个阈值分割而季风云团用第 10 个阈值分割, 即不同云系应采用不同温度阈值来进行分割.

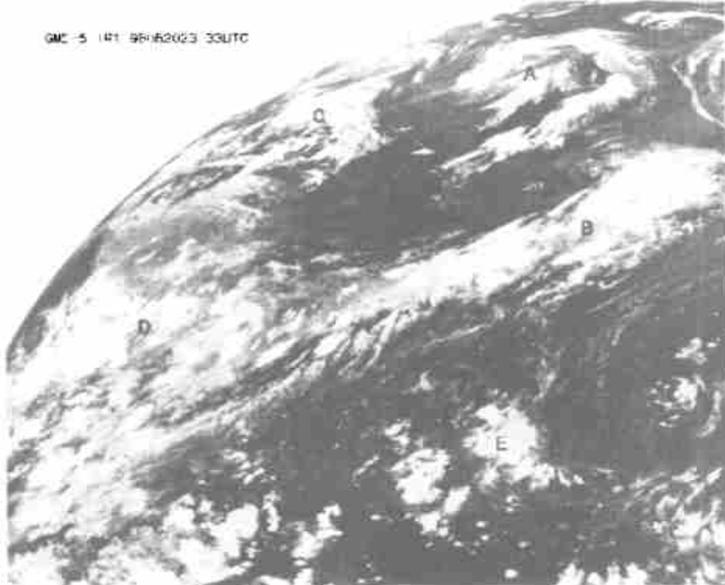


图 1 1998 年 6 月 20 日 23: 33(UTC), GMS-5 红外 1 通道卫星云图  
(A 和 C 为温带气旋云系, B 为切变线云系, D 为季风云团, E 为热带云团)

根据上述特点采用多阈值方法和人工神经网络方法相结合来分割卫星云图. 其基本原理是: 首先考虑卫星云图上可能存在的云系分割阈值, 然后用每一个阈值对卫星云图进行分割, 再将某个阈值包围的最能代表某种云系的分割区域挑选出来, 将挑选出来的区域综合在一起, 生成最后的分割图. 根据卫星红外云图探测云系的物理意义和经验, 在  $20 \sim 52^{\circ}\text{C}$  之间选择了 15 个阈值, 且阈值的选取不是等间隔选取, 如表 1 所示. 云系边界在  $0 \sim -40^{\circ}\text{C}$  区间出现的几率较大, 因而在这一区间选取较多的阈值. 图 2.1~ 2.15 是分别用这 15 个阈值对图 1 中的云图分割的中间结果, 图 2.16 把每一个云系用最适合的阈值分割的结果综合在一起构成的分割结果图.

表 1 分割阈值表

阈值	1	2	3	4	5	6	7	9	10	11	12	13	14	15
温度/°C	20	10	5	0	- 4	- 8	- 12	- 20	- 24	- 28	- 32	- 36	- 40	- 50

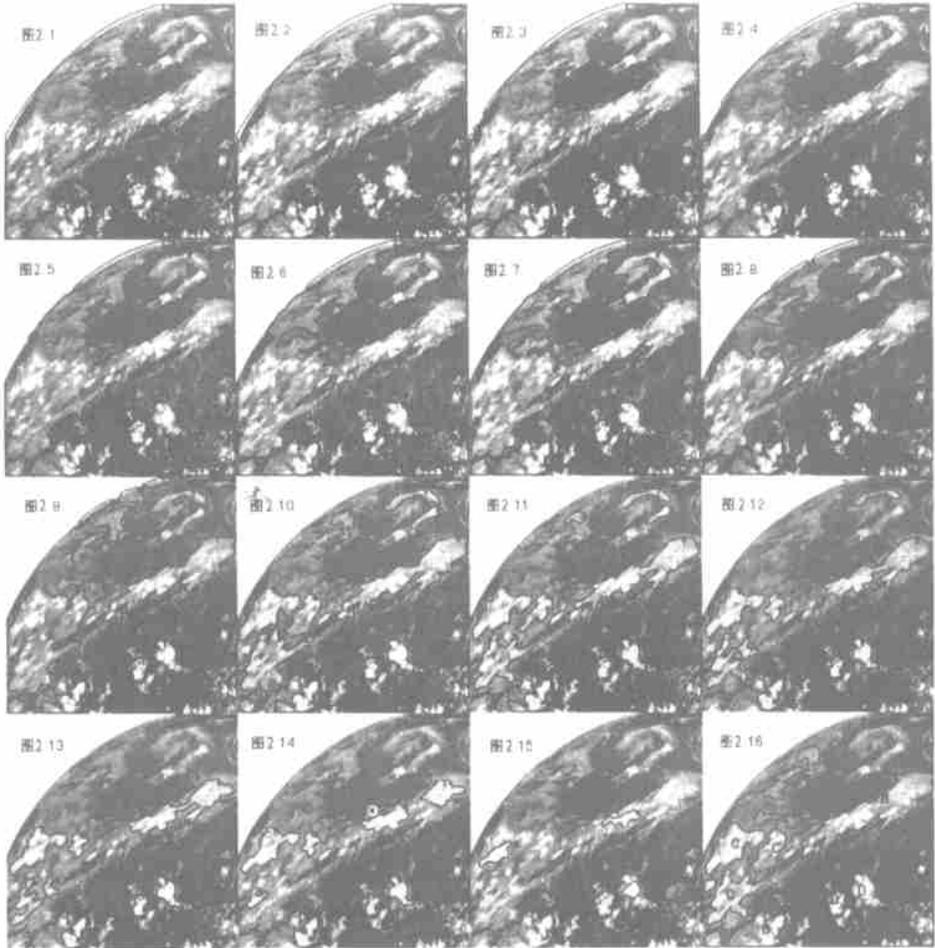


图 2 15 个单阈值分割图和最后分割结果图(图中的数字表示云区序号)

首先我们用 15 个阈值的分割结果构造出一个结构图如图 3, 它由图 2.1~ 2.15 导出, 在图 2 和图 3 中, 可以很清楚地看出卫星云图用单个不同的阈值进行分割的结果和它们之间的相互关系, 并可以方便地进行比较. 在温度较高阈值时, 云系联成一片, 在温度较低阈值时, 云系分割很碎, 我们称这一结构图为分割阈值树图, 从上向下看, 有两种类型的节点, 一种是不再分叉的节点, 另一种是在下一层次会分叉的节点. 等温线包围的云区在阈值树图上表示为节点, 节点号表示云区序号, 与图 2.1~ 2.15 中所标云区数字相对应; 分叉是指用某等温线包围的云块可用更低的等温线分割成独立的几块. 图 3 中用方框圈起来的节点是人工选择的要保留的节点, 用圆圈起来的节点是人工选择的. 该结构图对云图的分割情况进行了定量描述, 便于用计算机进行处理. 对这两种不同的节点, 采用不同的方法处理. 对于不再分叉的情况, 保留靠近树根的节点, 即最后一次分叉后的第一个节点.

## 2 卫星云图云系模型数据库

人工选择这些要保留的节点时, 综合考虑了很多因素, 较容易实现, 但计算机实现则比较困难, 本文通过训练人工神经网络模型解决这一问题. 对于那些分叉的情况, 由人工神经网络判断保留分叉前的节点还是分叉后的节点, 人工神经网络通过大量样本的学习, 可以记忆分割的特点, 并对新的类似情况做出正确判断, 这个过程称为修剪树.

用于训练和测试神经网络的样本数据来源于卫星云图云系模型库, 该云系模型库是针对我国陆地和海洋上夏季的天气尺度和中尺度云系建立的, 它包含了云系分割信息和云系分类信息. 本文使用 GMS 红外云图, 图像大小为 1024 × 1024, 采用原始分辨率, 这些图像是从不同年份 (1992 ~ 1994, 1997 ~ 1998 年 6 ~ 8 月) 选择的, 每幅云图上都有较典型的天气系统. 为了方便

地建立云系模型库, 我们设计编写了交互式卫星云图云系采样工具软件, 可自动生成 15 幅不同阈值分割图和未经修剪的阈值树, 然后气象专家根据经验判断出分割某一云系的最佳阈值, 交互输入要保留的节点号 (如图 3) 和它所属的类别, 最后该软件自动生成云系模型库中的样本. 用 177 幅云图采集了用于神经网络分割模型的有效样本数据 2764 个; 有效分类样本数据 3079 个. 其中包括与天气系统联系的典型云系类型 12 类, 云图上另外一些没有天气系统对应的云区, 按照它的云类型被归为积雨云、卷云、中云、低云等, 另有一些特别不明显、不重要的云区被定义为未知类型. 详细情况见表 2.

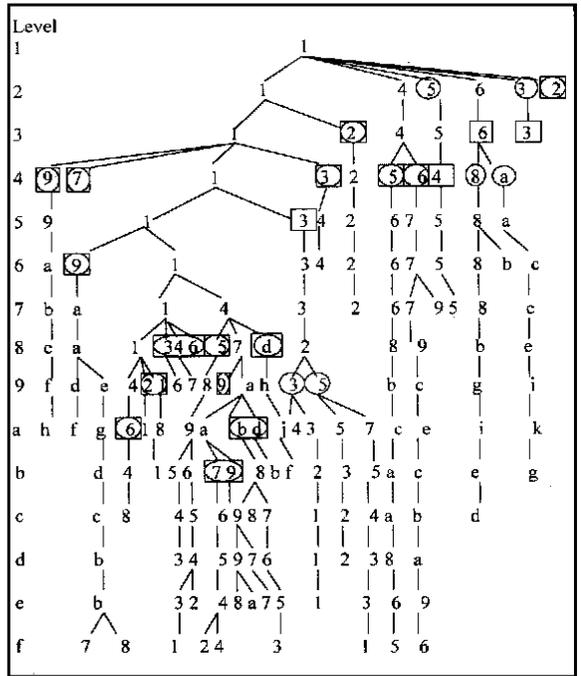


图 3 分割阈值树图  
(方框圈起来的节点对应人工选择的云区, 由圆圈起来的节点对应计算机选择的云区)

表 2 云系模型库中的云系类型

序号	云系类型	样本数	序号	云系类型	样本数	序号	云系类型	样本数
1	温带气旋	72	7	热带辐合带	521	13	积云	585
2	冷锋	75	8	云团	40	14	卷云	780
3	静止锋	35	9	MCC	5	15	中云	448
4	小槽	331	10	高空冷涡	22	16	低云	1266
5	深槽	14	11	大逗点云系	14			
6	热带气旋	67	12	东风波	2			

注: 1~16 类样本总数合计 3079, 1~12 类样本总数合计 1198.

### 3 人工神经网络试验

人工神经网络试验是在商用软件“神经网络计算机开发仿真系统 NeuralW orks11”环境下实现的, 硬件为 586PC 计算机. 采用多层前馈人工神经网络 BP 模型进行试验, 神经网络结构示意图如图 4, 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<sup>[5]</sup>. 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.

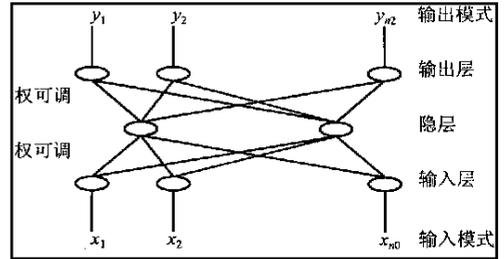


图 4 三层前馈神经网络结构

#### 3.1 神经网络输入特征

从云系模型库中选择了 32 幅云图, 它们包含了表 2 中所有的云系类型, 从构成的阈值树中抽取 484 个分叉的节点样本做为神经网络模型的训练集, 从云系模型库中其他云图构成的阈值树中抽取 2280 个分叉的节点样本做为测试集. 在图 3 中, 我们将上一个阈值包围的云区称之为“父区域”, 下一个阈值包围的云区则根据它们的面积大小, 顺序地称为“子区域 1”, “子区域 2”, “子区域 3”. 神经网络输入特征: ①父区域边界的周长; ②父区域的面积; ③父区域面积/周长; ④子区域 1 边界的周长; ⑤子区域 1 的面积; ⑥子区域 1 的面积/周长; ⑦子区域 2 边界的周长; ⑧子区域 2 的面积; ⑨子区域 2 的面积/周长; ⑩父区域对应阈值. 在构造神经网络结构时, 输出用两个节点表示, 一个表示保留父节点, 一个表示不保留父节点.

#### 3.2 神经网络结构试验

神经网络的输入层神经元数目一般与网络的输入特征相对应, 输出层神经元数目与待识别的类别数有关. 隐层层数和隐层神经元数目的选取目前还没有成熟的理论, 在大多数应用中, 依靠经验和试验决定. 通过调整神经网络模型的输入特征、隐层数、隐层节点数, 可改进测试正确率. 神经网络结构试验的部分结果见表 3. 综合考虑训练和测试正确率, 模型 2 效果最佳. 同时可以看出, 训练正确率高, 测试正确率不一定高, 即网络的推广性不一定好, 如模型 3. 这是因为在进行神经网络训练时会出现过量训练(over-training)的现象, 即网络结构相对于训练数据集太大. 表 3 的试验结果也说明了这一点, 网络模型 1, 2, 3 结构偏大, 虽然训练正确率较高, 但测试正确率却低于结构较小的网络模型 4 和 9 的测试正确率.

表 3 人工神经网络部分试验结果

序号	神经网络结构	特征数目	训练样本数	训练正确率/%	测试样本数	测试正确率/%	训练次数/万
1	10- 60- 20- 2	1~ 10	484	99.1	2280	83.2	883
2	10- 50- 10- 2	1~ 10	484	98.8	2280	86.4	738
3	10- 40- 8- 2	1~ 10	484	99.1	2280	82.7	1168
4	10- 18- 8- 2	1~ 10	484	90.6	2280	86.7	300
5	10- 99- 2	1~ 10	484	90.5	2280	85.8	125
6	9- 63- 18- 2	1~ 9	484	91.3	2280	84.5	1291
7	9- 45- 18- 2	1~ 9	484	95.9	2280	83.9	1244
8	9- 36- 8- 2	1~ 9	484	95.7	2280	83.6	4277
9	9- 81- 2	1~ 9	484	88.6	2280	86.7	300
10	9- 54- 2	1~ 9	484	89.2	2280	86.2	62

#### 4 综合测试

用样本集数据和实际卫星云图数据对多阈值和人工神经网络方法分割结果进行了综合测试, 结果如下:

(1) 单阈值样本集测试 用表 1 中的 15 个单阈值分别对云系模型库中的 177 幅云图进行分割, 结果见表 4. 总分割云区数为 4966. 从表 4 看出用单阈值分割云区, 正确率小于 15%, 在实际应用中是不可行的. 正确率是指某阈值下人工挑选出的云区数除以总云区数.

表 4 单阈值分割结果统计表

阈值	1	2	3	4	5	6	7	8	9	10	11	12	13	14
温度/°C	20	10	5	0	- 4	- 8	- 12	- 16	- 20	- 24	- 28	- 32	- 36	- 40
正确率/%	1	6	6	9	6	9	11	14	11	9	7	6	3	2
样本数	65	307	316	439	294	441	535	697	538	423	329	294	139	78

(2) 多阈值样本集测试 挑选了 7 个训练好的神经网络模型对云系模型库中的卫星云图做自动分割试验( 结果见表 5). 需要指出的是, 表 5 中的结果比表 3 的结果差, 这是因为其中包括神经网络模型造成的误差和由启发式搜索方法造成的误差. 启发式搜索是指计算机在确定某一云系的最佳阈值时, 经过多次从阈值树的树根到树梢以及从树梢到树根的搜索, 结合神经网络结果确定云系的最佳阈值. 分割正确率指启发式搜索与神经网络模型的分割正确率之和, 正确率由下式计算得出: 正确率= 分对的云区数/总云区数.

表 5 启发式搜索与神经网络模型自动分割结果

序号	神经网络结构	特征数	样本数	分割正确率/%	训练次数/万
1	10- 50- 10- 2	1~ 10	2280	61.1	738
2	10- 18- 8- 2	1~ 10	2280	56.9	300
3	9- 36- 8- 2	1~ 9	2280	41.2	4277
4	9- 81- 2	1~ 9	2280	61.1	300
5	9- 81- 2	1~ 9	2280	61.1	143
6	10- 40- 8- 2	1~ 10	2280	56.7	757
7	9- 54- 2	1~ 9	2280	56.0	1399

(3)实际云图测试 选择1997年7月18~21日和1998年6月15~17日的两组卫星云图做自动分割试验,这些云图不包含在云系模型库中,基本上每隔3h选择一张云图,首先由计算机自动分割这些云图,然后由气象专家评分,重点考虑了1~12类典型云系的正确分割率,对13~16类没有典型天气系统对应的云区未做评价,实际云图分割试验取得较好效果,云系分割正确率达到90%以上。

图5给出多阈值和神经网络相结合云图自动分割方法试验个例,使用了表3中的神经网络模型2,图5是1997年7月19日18:00(UTC)卫星原始云图叠加自动分割结果图。在这张卫星云图上的冷锋云系A,热带气旋云系B被较好地分割出来,图像左边的季风云团由于图像边界的影响,分割效果不太好。



图5 1997年7月19日18:00(UTC)卫星原始云图叠加自动分割结果图

## 5 分析与讨论

(1)在单阈值卫星云图自动分割试验结果(表4)中可以看出,如果用单阈值对云图分割,最好的分割正确率为14%,亮度温度为 $-16^{\circ}\text{C}$ ,它表明在实际应用中不可能用单阈值方法来实现云系自动分割。从表5可知,启发式搜索+神经网络模型自动分割测试正确率较好的结果是61.1%,在相同的评价标准下,多阈值和神经网络相结合云图分割方法明显优于单阈值法。通过进一步完善云系模型库和改进神经网络模型结构,可以提高云图自动分割正确率。

(2)由样本集测试结果可以看出:启发式搜索+神经网络模型自动分割正确率(表5)比实际云图分割试验结果低,这是因为卫星云图上云系的定义是一个比较模糊的概念,有时用不同的等温线均能正确描述该云系。比如在图2.5中分割出来的温带气旋是正确的,而图2.6中分割出来的区域3也是温带气旋的正确分割。但样本集是事先确定的,自动分

割结果与样本的选择完全一致时,才是正确的;而在实际云图分割试验中,分割结果符合实际情况时,就被认为是正确的。另一方面,对于不再分叉的云系区域,为了程序设计方便,启发式搜索选择了最靠近阈值树根部的云区,这与云系模型库中的人工采样也有差异;此外,对于云图上不重要的零散的云区是否用确切的等值线分割并不影响实际应用效果,但是在样本集测试中,正确率的计算是严格按照自动分割结果和人工分割结果是否一致计算出来的。因此从数学统计意义上看,启发式搜索+神经网络模型对样本库的自动分割结果正确率相对较低。实际云图测试实际上是云系自动分割方法简单的实际应用,正确率达90%以上,由于没有考虑属于13~16类的非重要云系的分割误差,因此统计出来的正确率相对偏高。但这种方法能很好地分割出具有天气学意义的云系,证明了多阈值和神经网络相结合的自动分割方法对天气尺度和中尺度云系的自动分割具有优势和潜力。

## 6 小 结

采用多阈值和人工神经网络相结合的方法对GMS红外云图进行自动分割试验,启发式搜索和人工神经网络方法分割正确率为61%。用484个样本对人工神经网络训练的正确率达到98.8%,用2280个独立样本测试的结果是86.4%。用1997年7月18~21日和1998年6月15~17日的两组卫星云图做自动分割模拟应用试验,正确率达到90%以上。

自动分割部分的输入是GMS红外云图,输出是分割出的云区及各云区的边界链码、起始点、周长、面积和云系类型,并保留了原始图像数据。在下一步云系识别过程中,可以在此基础上方便地建立云系特征库,并对云系进行分类识别试验。

通过本文的试验说明:多阈值和人工神经网络相结合的方法对于针对天气尺度和中尺度云系的卫星云图分割是可行的。

## 参 考 文 献

- 1 王耀生,王燕,冯晓娟,等.利用卫星多通道遥感资料进行云的识别试验.气象,1994,(6):24~32.
- 2 Brewer M, Malladi R. Methods for large-scale segmentation of cloud images. 8th Conference on Satellite Meteorology and Oceanography, 1998. 131~138.
- 3 Peak J E, Tag P M. Toward of automated interpretation of satellite imagery for navy shipboard applications. *Bulletin of the American Meteorological Society*, 1992, 73(7): 995~1008.
- 4 Peak J E, Tag P M. Segmentation of satellite imagery using hierarchical thresholding and neural networks. *Journal of Applied Meteorology*, 1994, 33: 605~616.
- 5 杨行峻,郑君里.神经网络.北京:高等教育出版社,1992.

## AUTOMATIC SEGMENTATION OF SATELLITE IMAGE USING HIERARCHICAL THRESHOLD AND NEURAL NETWORK

Shi Chunxiang   Wu Rongzhang   Xiang Xukang  
(*National Satellite Meteorological Center, Beijing 100081*)

### Abstract

Automatic segmentation of satellite image is the base of the automatic identification of cloud systems. This paper presents a combined method of hierarchical threshold segmentation and neural network to segment the image into separate synoptic systems for identification purpose. The processes include selecting all potential TBB thresholds of cloud segments and plotting temperature contours on satellite images for each threshold, then selecting cloud regions which is identified by a certain temperature contour and best representation of synoptic systems, and finally forming a segmented image by combining these regions. How to select these regions by computer is an experiential and uncertain problem. In this paper it is solved by the up-bottom and bottom-up heuristic and neural network method. A cloud pattern database is established. It includes 177 GMS satellite infrared images with 16 kinds of typical synoptic systems in the summer of 1992 - 1994, 1997- 1998. There are 484 training samples from 32 satellite images and 2280 testing samples from 145 satellite images. The neural network accuracy rate for these training samples is 98.8% and 86.4% for the testing samples. The experiment accuracy rate for the application test is above 90% using testing images of 18- 21 July 1997 and 15- 17 June 1998.

The input of the satellite image automatic segmentation system is GMS images, and the outputs include the boundary chain code, start location, perimeter and area of each region. These outputs are useful to further classification of cloud patterns.

**Key words:** Satellite image   Automatic segmentation   Artificial neural networks   Hierarchical threshold