

# 基于人工神经网络的 GMS 云图四类云系的识别\*

白慧卿 方宗义 吴蓉璋

(国家卫星气象中心,北京 100081)

郑君里

(清华大学电子工程系,北京 100084)

## 提 要

应用人工神经网络方法较成功地识别了 GMS 云图中的四类云系,并与传统的统计识别方法进行了对比,得出结论:神经网络方法更适合于云系的特征识别.还研究了神经网络本身的问题,取得一些网络设计的经验.

**关键词:**神经网络 云系 识别

## 引 言

卫星云图识别是一个非常重要的研究课题,国内外学者在这一领域已探索了几十年<sup>[1]</sup>. 计算机技术的发展使云图的判识由早期的目视判读向自动化发展,传统的计算机识别方法基于统计决策理论,但当假设的统计概率模型不成立时,这种方法的识别性能就会大大下降.

神经网络理论的发展解决了这一问题<sup>[2]</sup>. 这种理论模拟人脑神经网络的工作原理,以智能思想解决问题,具有很强的自组织、自适应的学习能力,不需要获取待识别模式的先验概率.它是通过对多个样本的学习,获得样本的知识并将知识分布存贮于网络中,从而达到对此类事物的理解和识别.

本文用神经网络方法判识了 GMS 云图中的四类云系:冷锋、静止锋、雹暴云团、强对流复合体(MCC). 与统计识别方法比较,神经网络的识别率高于统计识别方法.

## 1 神经网络模型及算法

人工神经网络目前已有几十种网络模型,根据不同网络的特点和本文实际问题,这里选择多层前馈网络,模型如图 1 所示. 图中输出层  $K$  有  $m$  个神经元. 隐层  $J$  有  $n_1$  个神经元,输入层  $I$  有  $n$  个神经元, $I$  层与  $J$  层以及  $J$  层与  $K$  层之间的权重分别用  $W_{ji}$  和  $W_{kj}$  表示.

网络分层排列,每层由若干神经元组成,每个神经元只接受低层各神经元的输入,经

\* 1997-04-18 收到,1997-07-25 收到修改稿.

过神经元运算后,通过连接权重  $W$  传到更高一层的各个神经元中.本文中神经元的功能函数选择 *Sigmoid* 函数.

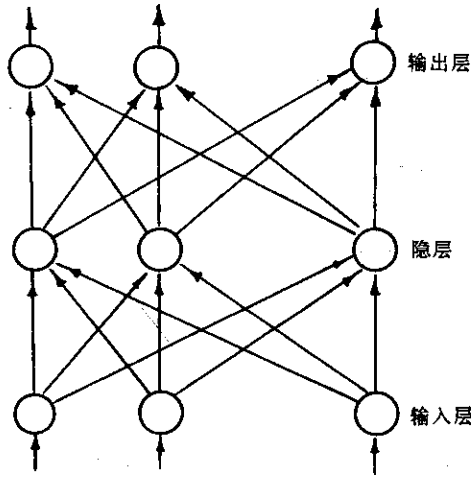


图 1 多层前馈网络模型

神经网络通过学习获得知识,网络学习的过程就是各个权和神经元的自动调节过程.调节的准则使网络的输出与实际的模式尽可能地接近,如果有误差,则自动调节权使误差减小直至满足要求.

对每一个(输入、输出)模式对,其误差平方为

$$E = \frac{1}{2} \left[ \sum_k (t_k - O_k)^2 \right]$$

下标  $k$  代表第  $k$  个神经元,  $t_k$  代表理想模式,  $O_k$  代表网络的实际输出.

利用梯度下降法,输出层权的调节量

$$\Delta W_{kj} = -\alpha \frac{\partial E}{\partial W_{kj}} = \alpha \cdot \delta_k \cdot O_j \quad (1)$$

其中

$$\delta_k = O_k(1 - O_k)(t_k - O_k) \quad (2)$$

式中  $\alpha$  是学习的步长,称为学习率,  $O_j$  是与输出层相连的第  $J$  隐层的各个神经元的输出,  $\delta_k$  为输出层的误差项.

对于隐层,权的调节公式为

$$\Delta W_{ji} = \alpha \cdot \delta_j \cdot O_i \quad (3)$$

其中

$$\delta_j = O_j(1 - O_j) \sum_k \delta_k \cdot W_{kj} \quad (4)$$

式中  $O_i$  是比  $J$  隐层低一层的神经元的输出值,  $\delta_j$  为  $J$  隐层的误差项,  $W_{kj}$  为  $J$  层与  $K$  层之间的连接权重.

由式(1)~(4)可以看出,权  $W$  的调节是通过将误差项反向传播完成的.通过调节  $W$  使输出的误差  $E$  逐渐减小直到满足要求.该算法也叫 Back Propagation 算法,简称 B-P 算法.

在 B-P 算法中,  $\alpha$  是一个很重要的参数.  $\alpha$  太大时,在训练开始阶段,系统误差下降很

快,但当接近最佳点时,网络的权将产生反复振荡而不能收敛.如果 $\alpha$ 太小,则训练时间很长,为了既加快学习收敛又避免这种振荡,可以采用“惯性”调整.即:

$$\Delta W_{(t+1)} = \alpha \delta_{(t+1)} \cdot O_{(t+1)} + \eta \cdot \Delta W_t \quad (5)$$

与式(1)及式(3)相比,式(5)增加了一项 $\eta \cdot \Delta W_t$ ,式中的 $t, t+1$ 表示学习的两个连续时序编号.增加的项 $\eta \cdot \Delta W_t$ 叫做惯性量,记忆上一时刻的修改方向, $\eta$ 称为惯性系数, $\eta$ 越大,则每一次权调整量与上一次权调整量更密切相关.

B-P算法是多层神经网络的一个非常有效的算法,它完成了一个从输入 $n$ 维到输出 $m$ 维的非线性映射.用这种方法解决云系的特征判识问题,只要提取出云系的 $n$ 维特征,不管这些特征的统计分布如何,将其输入给神经网络,则网络通过学习便能自动将其判识成 $M$ 类云系.

## 2 云系特征提取

本文使用的卫星云图数据集取自85906-01-06课题组对我国1992年和1993年四川、河南、湖北3省GMS云图中降雨云系的调查统计.本文选择了100多张图,从中挑选了162个典型云系,其中47个冷锋、55个静止锋、25个雹暴云团和35个MCC.每类各取15个样本组成训练集,其余102个独立样本为测试集.

### 2.1 原始特征

根据四类云系的特点,本文选取了下列6个特征.

(1)体态比 $P$   $P = \text{长轴尺度} / \text{短轴尺度}$ .

(2)边界亮温 $T$   $T$ 是指云系边界轮廓线的亮温值的绝对值,  $T = \left| \frac{\text{边界亮温}}{10} \right|$ .

(3)不同亮温段云面积与整个云系面积比值取4个亮温段:边界亮温 $\sim -32^\circ\text{C}$ 、 $-32 \sim -52^\circ\text{C}$ 、 $-52 \sim -65^\circ\text{C}$ 及 $-65^\circ\text{C}$ 以上,用 $t_1, t_2, t_3$ 和 $t_4$ 分别表示4个亮温段的像素数,用 $s$ 表示云系整体面积内的像素数,则 $T_1 = t_1/s, T_2 = t_2/s, T_3 = t_3/s, T_4 = t_4/s$ 分别表示4个比值.

(4)地理位置特征 以云系北边界的经纬度值描述云系的位置,经度特征 $X = \text{实际经度值} / 100$ ,纬度特征 $Y = \text{实际纬度值} / 100$ .

(5)走向特征 以云系长轴与东—西向之间的夹角的正切值描述云系的走向,  $D = \text{tga} = b/a$ .

(6)尺度 尺度是指云系长轴所达的范围(用 $L$ 表示,单位:km),  $L = \text{实际尺度} / 1000$ .

经过上述描述,将每种矢量都描述成一个10维的矢量 $[P, T, T_1, T_2, T_3, T_4, X, Y, D, L]$ ,那么如何从这10维原始特征中提取和选择出对识别分类最有意义的特征呢?这是模式识别问题中一个很重要的环节——特征提取和变换.另外从神经网络的设计来考虑,特征维数越多,网络输入层神经元势必增加,致使网络结构增大,所以把高维特征空间压缩到低维特征空间是设计网络的一个重要环节.本文通过K-L变换将高维特征空间压缩到低维特征空间.

## 2.2 基于 K-L 变换的特征提取

K-L 变换是一种线性变换:

$$\vec{Y} = A \cdot \vec{X}$$

$\vec{X}$ 、 $\vec{Y}$  分别为变换前后的  $n$  维矢量,  $A$  是  $n \times n$  的变换阵, 设变换前后的期望矢量分别为  $M_X$  和  $M_Y$ , 协方差阵分别为  $\sum_X$  和  $\sum_Y$ , 则

$$M_Y = A \cdot M_X$$

$$\sum_Y = A \sum_X A^T$$

K-L 变换就是选择  $\sum_X$  的特征矢量矩阵  $\Phi$  的转置阵  $\Phi^T$  为变换阵, 此时

$$\vec{Y} = \Phi^T \vec{X} \quad (6)$$

式(6)称为 K-L 变换.

通过 K-L 变换后

$$\sum_Y = \Phi^T \sum_X \Phi = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = \Lambda$$

即 K-L 变换使  $\sum_Y$  对角化为  $\Lambda$  (对角阵).  $\sum_Y$  的对角化表示变换后的各个特征彼此不相关, 是彼此独立的新变量. K-L 坐标系中的  $y_1, y_2, \dots, y_n$  为  $\sum_X$  的各特征矢量方向,  $n$  个变量中对应特征值大的变量对模式贡献大, 按特征值大小顺序排列成  $\lambda_1, \lambda_2, \dots, \lambda_m, \dots, \lambda_n$ . 可以选取  $m$  个变量表示  $\vec{X}$ , 而对分类不会带来太大的影响, 这样就实现了变量压缩.

将原始 10 维数据经过 K-L 变换后得到:

$$\begin{array}{ll} \lambda_1 = 14.1257 & \lambda_2 = 0.400305 \\ \lambda_3 = 0.00422639 & \lambda_4 = 0.0274866 \\ \lambda_5 = 0.000426519 & \lambda_6 = 0.0167331 \\ \lambda_7 = 0.0208997 & \lambda_8 = 0.00574661 \\ \lambda_9 = 1.611724 & \lambda_{10} = 0.166071 \end{array}$$

$\lambda$  值由大到小排列为:  $\lambda_1 > \lambda_9 > \lambda_2 > \lambda_{10} > \lambda_4 > \lambda_7 > \lambda_6 > \lambda_8 > \lambda_3 > \lambda_5$ . 依次去掉  $\lambda$  值小的对应变量, 直到分类器的分类性能显著下降为止.

在特征提取过程中使用了神经网络和最大似然(bayes)两种分类器. 训练集由四类云系中各随机抽取 15 个样本组成, 测试集由剩余的 102 个样本组成, 试验结果见表 1, 其中网络结构是指各层的神经元数, 网络参数指学习率  $\alpha$  和惯性系数  $\eta$ . 设  $t_1, t_2$  为网络理想输出,  $y_1, y_2$  是实际输出, 若  $|t_1 - y_1| > 0.2$  或  $|t_2 - y_2| > 0.2$ , 则认为网络判错.

由上述实验看出, 在 8 维特征时, 神经网络的分类效果是最好的, 而 Bayes 分类性能保持原来水平, 当维数进一步降低后, 神经网络仍然保持很高的识别率, 而 Bayes 分类器的分类性能却大大下降, 这说明神经网络可以从比 Bayes 方法更少的信息量中作出对

事物的判断.从上述的特征提取过程可以看出,4维数据保留了主要的信息,神经网络仍能对其识别.神经网络的平均识别率对训练集是99.58%,对测试集是96.58%,而统计方法的平均识别率对训练集是92.08%,对测试集是82.58%.很明显,神经网络比统计方法识别率高.

表1 特征提取实验结果

		特征维数						
		10	8	7	6	5	4	3
神经网络分类	错判数	1	0	0	0	1	0	3
	网络结构	10-7-4-2	8-5-3-2	7-5-3-2	6-4-2	5-4-2	4-3-2	3-3-2
	网络参数	0.9,0.7	0.8,0.6	0.8,0.6	0.8,0.6	0.8,0.5	0.8,0.5	0.6,0.5
	系统误差	0.005431	0.000015	0.000014	0.000428	0.007035	0.005731	0.004746
	识别率(%)	98.33	100	100	100	98.33	100	95
	错判数	3	2	4	3	4	3	8
测试	识别率(%)	97.06	98.03	96.08	97.06	96.08	97.06	92.16
	错判数	3	4	5	4	4	6	5
统计分类	识别率(%)	95.00	93.33	91.67	93.33	93.33	90.00	91.67
	错判数	2	2	18	20	18	15	10
	识别率(%)	98.03	98.03	82.35	80.39	82.35	82.29	90.20

### 3 神经网络设计过程

#### 3.1 结构设计

基于B-P算法的多层前馈网络分类器的输入层、输出层的设计完全根据问题的要求而定.

输入层神经元数目等于样本特征数据的维数,例如在特征提取过程中,输入层神经元可以是10,8,7等.输出层根据所要识别的模式类别数进行设计.本文要识别的模式是四类云系,输出层可设4个神经元,每个神经元代表一类模式.为了减小网络结构,本文用类别数的编码确定四类云系,因 $2^2=4$ ,故输出层用2个神经元.

关于隐层的设计,对于输入为模拟量的网络来说,神经元数目的确定还没有理论上的指导<sup>[3][4]</sup>,这要靠具体实验确定.实验结果如表2所示,实验中网络参数 $\alpha=0.8$ 、 $\eta=0.5$ ,迭代次数500次.理论上讲<sup>[4]</sup>,单隐层的网络能完成任何复杂的 $m$ 维 $\rightarrow n$ 维映射,至于选单隐层还是多隐层,这与问题的复杂性有关.有的问题,隐层数增多,收敛加快<sup>[5]</sup>,但本文的实验却相反,即:单隐层比双隐层效果好,而且按 $n_1 > n_2$  ( $n_1$ 、 $n_2$ 分别是第一、第二隐层的神经元数).通过多次实验,单隐层8个神经元的网络其分类效果最好,表3是4-8-2网络的测试效果(仅列出部分样本的测试结果).

#### 3.2 参数选择

在式(5)中,学习率 $\alpha$ 和惯性系数 $\eta$ 的选择对网络的收敛效果影响很大. $\alpha$ 、 $\eta$ 选得合适时,网络能很好地收敛,否则会进入误差很大的局部极小点.本文将 $\alpha$ 、 $\eta$ 简化成 $0 < \alpha < 1$ 、 $0 < \eta < 1$ 的某一常数,通过多次实验得知,当 $\alpha=0.8$ 、 $\eta=0.5$ 时,网络的收敛效果最好.当取 $\alpha$ 、 $\eta$ 为常数时,一般来讲,网络处理的问题不同, $\alpha$ 、 $\eta$ 的最优取值也是不同的.在后续工

作中,作者将讨论 $\alpha$ 、 $\eta$ 随学习的过程而改变的自动优化算法.

表2 神经网络结构实验结果

网络结构	系统误差	训练集		测试集	
		错判数	识别率(%)	错判数	识别率(%)
4-2-2	0.054249	34	43.33	74	27.45
4-3-2	0.005782	1	98.33	3	97.06
4-4-2	0.010223	3	95.00	5	95.01
4-2-2-2	0.035875	23	61.67	29	71.57
4-5-2	0.12778	2	96.67	5	95.10
4-3-2-2	0.015533	4	93.33	6	94.12
4-8-2	0.005269	1	98.33	1	99.02
8-8-2	0.000384	0	100	1	99.02
8-5-3-2	0.000067	0	100	4	96.08
8-3-5-2	0.048067	32	46.67	74	27.45

表3 4-8-2网络的测试结果

样本 序号	神经元 $n_1$		神经元 $n_2$		实际 模式	测试 结果	样本 序号	神经元 $n_1$		神经元 $n_2$		实际 模式	测试 结果
	实际值	理想值	实际值	理想值				实际值	理想值	实际值	理想值		
0	0.90	0.90	0.91	0.90	1	1	51	0.89	0.90	0.08	0.10	4	4
2	0.06	0.10	0.00	0.10	2	2	53	0.89	0.90	0.08	0.10	4	4
4	0.90	0.90	0.91	0.90	1	1	55	0.89	0.90	0.08	0.10	4	4
6	0.90	0.90	0.91	0.90	1	1	57	0.89	0.90	0.08	0.10	4	4
8	0.05	0.10	0.00	0.10	2	2	59	0.04	0.10	0.93	0.90	3	3
10	0.89	0.90	0.91	0.90	1	1	61	0.24	0.10	0.06	0.10	2	2
12	0.89	0.90	0.91	0.90	1	1	63	0.06	0.10	0.88	0.90	3	3
14	0.88	0.90	0.91	0.90	1	1	65	0.02	0.10	0.39	0.90	3	×
16	0.89	0.90	0.91	0.90	1	1	67	0.07	0.10	0.09	0.10	2	2
18	0.88	0.90	0.90	0.90	1	1	69	0.04	0.10	0.01	0.10	2	2
20	0.85	0.90	0.88	0.90	1	1	71	0.05	0.10	0.00	0.10	2	2
22	0.94	0.90	0.94	0.90	1	1	73	0.89	0.90	0.92	0.90	1	1
24	0.93	0.90	0.93	0.90	1	1	75	0.89	0.90	0.91	0.90	1	1
26	0.88	0.90	0.90	0.90	1	1	77	0.89	0.90	0.91	0.90	1	1
28	0.06	0.10	0.00	0.10	2	2	79	0.89	0.90	0.08	0.10	4	4
30	0.08	0.10	0.00	0.10	2	2	81	0.89	0.90	0.08	0.10	4	4
32	0.12	0.10	0.00	0.10	2	2	83	0.89	0.90	0.08	0.10	4	4
34	0.91	0.90	0.84	0.90	1	1	85	0.89	0.90	0.08	0.10	4	4
36	0.92	0.90	0.87	0.90	1	1	87	0.89	0.90	0.09	0.10	4	4
38	0.91	0.90	0.78	0.90	1	1	89	0.90	0.90	0.11	0.10	4	4
40	0.92	0.90	0.91	0.90	1	1	91	0.89	0.90	0.08	0.10	4	4
42	0.92	0.90	0.91	0.90	1	1	93	0.89	0.90	0.08	0.10	4	4
44	0.12	0.10	0.92	0.90	3	3	95	0.90	0.90	0.92	0.90	1	1
46	0.19	0.10	0.07	0.10	2	2	97	0.90	0.90	0.12	0.10	4	4
48	0.13	0.10	0.99	0.90	3	3	99	0.89	0.90	0.16	0.10	4	4
50	0.88	0.90	0.08	0.10	4	4	101	0.89	0.90	0.08	0.10	4	4

### 3.3 训练集样本的顺序与数量的选择

训练网络时,样本的输入顺序影响网络的学习效果. 设  $A_i, B_i, C_i, D_i$  分别表示四类云系的第  $i$  个样本, 当将四类样本均匀地送入网络时, 即: 按  $A_1, B_1, C_1, D_1, A_2, B_2, C_2, D_2, \dots, A_{15}, B_{15}, C_{15}, D_{15}$  顺序输入, 则网络能很好地记忆这些样本. 若按类别顺序输入, 即: 按  $A_1, A_2, \dots, A_{15}, B_1, B_2, \dots, B_{15}, C_1, C_2, \dots, C_{15}, D_1, D_2, \dots, D_{15}$  顺序输入, 则网络能很好地记忆后输入的样本, 先输入的类别信息被后输入的类别信息所掩盖, 出现先学后忘的现象, 特别对 8 维特征的样本, 网络几乎完全忘记了第一类和第二类信息. 表 4 是实现结果. 实验中网络参数  $\alpha=0.8, \eta=0.6$ , 学习迭代 1000 次.

表 4 样本顺序实验结果

样本输入方式	网络结构	训练错判数				测试错判数			
		A	B	C	D	A	B	C	D
按类别顺序输入	8-5-3-2	15	15	0	0	30	10	0	0
	4-3-2	0	4	0	0	5	4	0	1
四类均匀等间隔输入	8-5-3-2	0	0	0	0	0	2	0	2
	4-3-2	0	1	0	0	0	1	0	1

在神经网络的识别系统中, 训练集的大小也是一个关键问题, 这直接影响着网络的推广能力, 即对测试集的识别能力. 在选择训练集为 20、38、36、44、52 时, 在训练阶段, 网络识别效果很好, 但对测试集识别效果较差, 这是由于训练集样本没有代表四类云系中的各种形态, 如锋面有强弱之分, 如果训练集中不包括强锋面, 则网络就不能识别测试集中的相应模式, 致使推广能力降低, 所以在训练集的生成时, 训练样本一定要有代表性. 当训练集为 60 时, 4-8-2 结构网络的识别率达 99.02%, 当样本数继续增加时, 网络也相应地增大, 并且学习速度减慢, 计算复杂性增加.

### 3.4 迭代次数对网络性能的影响

迭代次数不仅反映网络学习的快慢, 而且也影响着网络的推广能力. 例如: 网络识别 10 维特征、系统迭代 3000 次时, 误差  $E=0.00343$ , 测试精度 96.08%; 继续迭代 5000 次时, 系统误差  $E=0.00007$ , 但网络的识别率不但没有提高, 反而下降到 91.18%, 由此可见, 并非是网络的系统误差越小其测试识别率越高, 而是存在一个最佳点.

## 4 结 论

人工神经网络在云系特征识别中比传统的统计方法更具有优越性. 神经网络在训练时能达到 100% 的识别, 在测试时最高达 99.02%, 而 Bayes 方法的最高识别率对训练集是 95%, 对测试集是 98.03%. 当特征维数减少后, 神经网络的性能更加优于 Bayes 方法, 神经网络在特征提取过程中的平均识别率对训练集是 99.58%, 对测试集是 96.58%, 而 Bayes 的平均识别率对训练集是 92.08%, 对测试集是 82.58%. 由此可见, 神经网络方法可采用更少的信息识别模式, 而且可以不考虑这些信息的分布情况.

人工神经网络的性能受其结构、参数的影响. 如果结构和参数选得不合适, 则网络便会陷入局部极小点, 本文 4-8-2, 8-8-2 的结构优于其它结构的网络, 其参数  $\alpha$  为 0.8 左右,

$\eta$  为 0.5 左右,网络的推广能力在系统误差下降到一定程度后有一个最佳点.另外训练集的形成对网络的推广能力也有很大影响,如果训练集能代表模式的各种形态,则网络的推广能力强.样本的学习次序影响网络的识别效果,如果按类别次序学习,则会出现遗忘现象,将学习样本均匀地送入网络,网络的学习效果好.

B-P 算法存在两个明显的缺点:①存在局部极小点;②学习算法收敛速度慢.在后续工作中,将在这两方面做些改进工作.

致谢:本文云图的选择得到叶惠明同志的热情指导,特此致谢.

### 参 考 文 献

- 1 白慧卿.用人工神经网络识别 GMS 云图中的几种云系.硕士论文.1995.5.
- 2 Peak J E. Application of Neural Networks to Large-Scale Cloud Pattern Recognition. Ad-238986. 1991. 1~25.
- 3 杨行俊,郑君里.人工神经网络原理.北京:高等教育出版社,1992. 26~89.
- 4 张立明.人工神经网络的模型及其应用.上海:复旦大学出版社,1993. 13~47.
- 5 徐雷,迟惠生.静态前馈网络的监督学习方法研究进展.电子学报,1992,20:106~109.

## IDENTIFICATION ABOUT FOUR KINDS OF CLOUD SYSTEMS IN GMS IMAGES BASED ON NEURAL NETWORK

Bai Huiqing    Fang Zongyi    Wu Rongzhang  
(National Satellite Meteorological Center, Beijing 100081)

Zheng Junli

(Department of Electronic Engineering, Qinghua University, Beijing 100084)

### Abstract

Based on the theory of neural network, four kinds of cloud systems in GMS images are successfully identified. Comparing neural network with traditional maximum likelihood method, we can draw the conclusion that the neural network is even better for identifying cloud systems. Some problems about neural network itself are also studied, and some experiences for designing neural network are obtained.

**Key words:** Neural network    Cloud system    Identification