# 北方工业大学本科毕业设计（论文）答辩报告

题　　目：基于图像识别技术的空气质量评估方法研究

指导教师：　　陈小光

专业班级：　　信息与计算科学11-1

学　　号：　　11105010129

姓　　名：　　李东月

# 一、课题研究的背景

工业污染

交通运输污染

污染指数“爆表”

# 二、课题研究的目的与意义

图像质量

空气质量

# 二、课题研究内容

- 本文通过搜集日常相机拍摄的天气数码图像，用图像处理技术对图像进行特征提取，并设计了基于BP人工神经网络的分类器，编写算法程序实现了对给定图像的空气质量指数（AQI）的预测，及空气质量类别的识别。

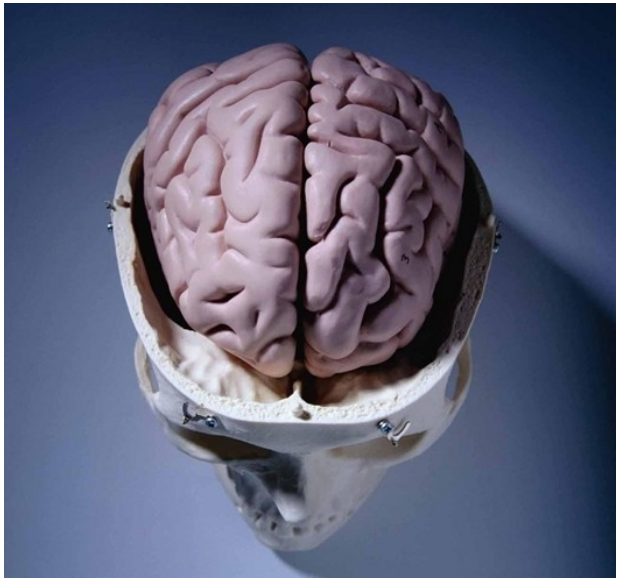
# 全文基本框架

- 第一章 绪论
- 第二章 图像处理及识别技术
- 第三章 图像分类器
- 第四章 基于BP神经网络分类器的空气质量评估算法
- 第五章 测试结果及分析
- 第六章 结论

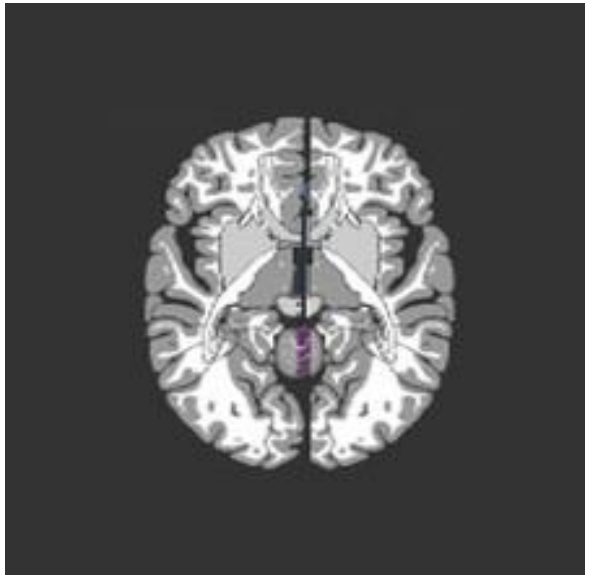
# 三、基于BP神经网络分类器的空气质量评估算法

- 为什么选用神经网络？

（1）生物学的启示

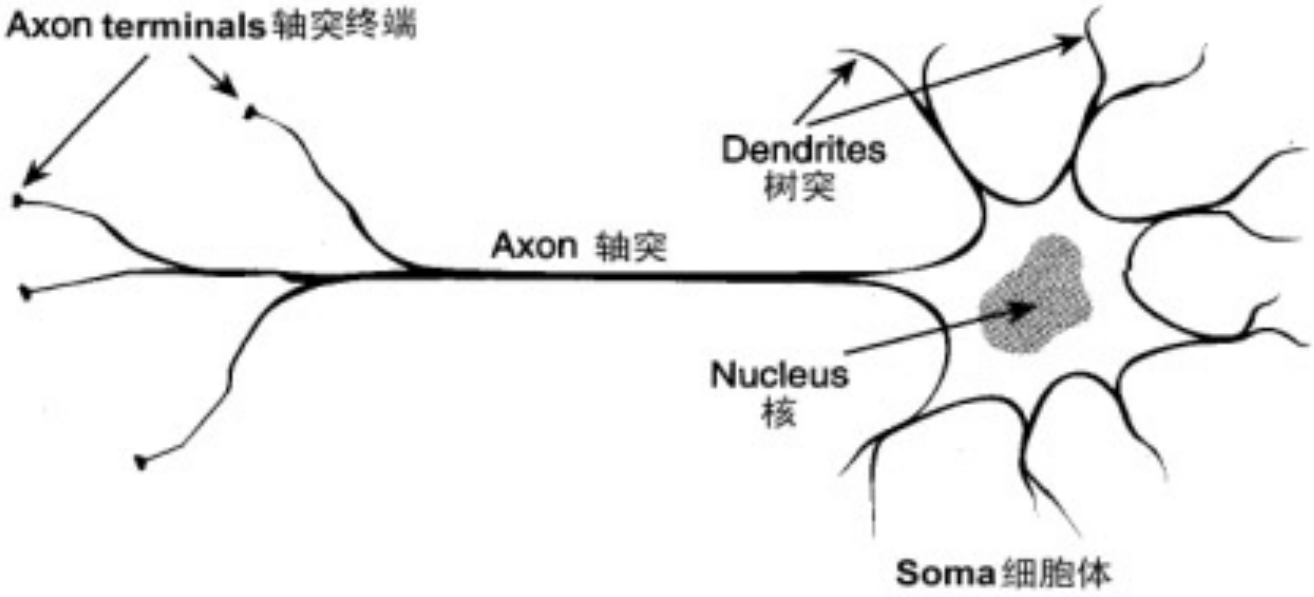


大脑皮层



外层：灰质　内层：白质

神经细胞利用电——化学过程交换信号。信号在大脑中实际怎样传输是一个相当复杂的过程，神经细胞利用一种我们还不知道的方法,把所有从树突突触上进来的信号进行相加，如果全部信号的总和超过某个阀值，就会激发神经细胞进入兴奋（fire）状态，这时就会有一个电信号通过轴突发送出去给其他神经细胞。如果信号总和没有达到阀值，神经细胞就不会兴奋起来。




神经细胞的结构

（2）数字版的神经网络

- 人工神经网络是由大量简单的神经元按照一定连接方式形成的仿生网，具有联想记忆和储存知识的性能。是从人脑的结构出发模拟人的智能行为。
- 人工神经网络与人脑在两方面具备相似性，即：

① 知识的获取依赖于对外界环境中学习。

② 通过神经元互相的连接强度，即突触权值，来对知识存储。

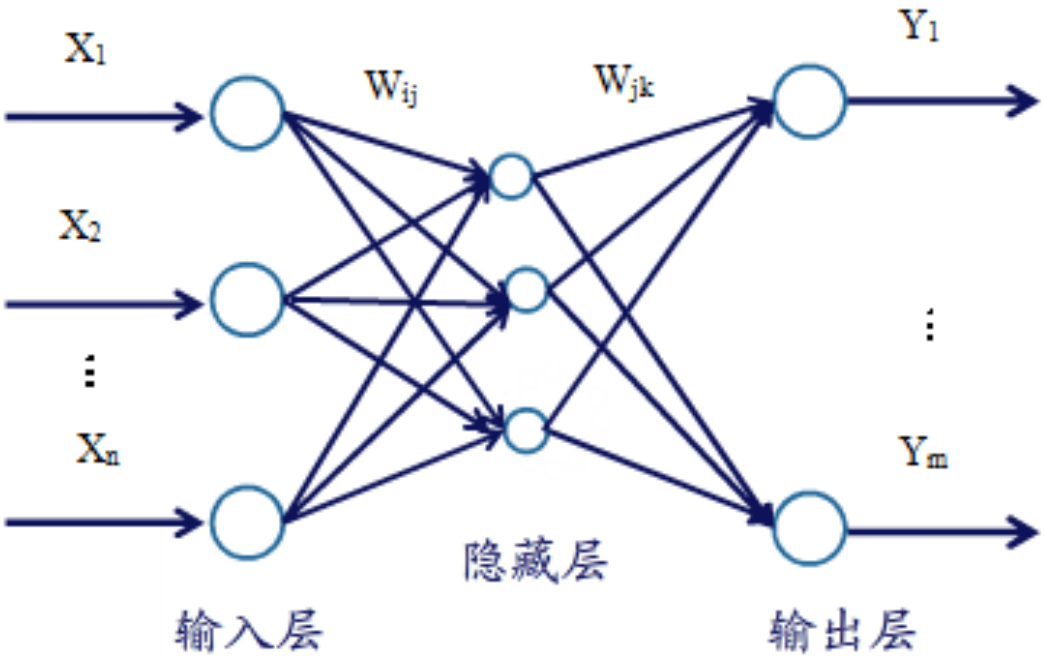
- 人工神经网络为逼近连续值、离散值及向量值的目标函数提供了鲁棒的学习方法，具有很强的自学能力，可以在学习过程中不断完善。

- 什么是BP神经网络

Back Propagation网络是1986年由Rumelhart和McCelland为首的科学家小组提出，是一种按误差反传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一.BP网络能学习和存贮大量的输入-输出模式映射关系,而无需事前揭示描述这种映射关系的数学方程.它的学习规则是使用最速下降法,通过反向传播来不断调整网络的权值和阈值,使网络的误差平方和最小.BP神经网络模型拓扑结构包括输入层（input）、隐层(hide layer)和输出层(output layer).

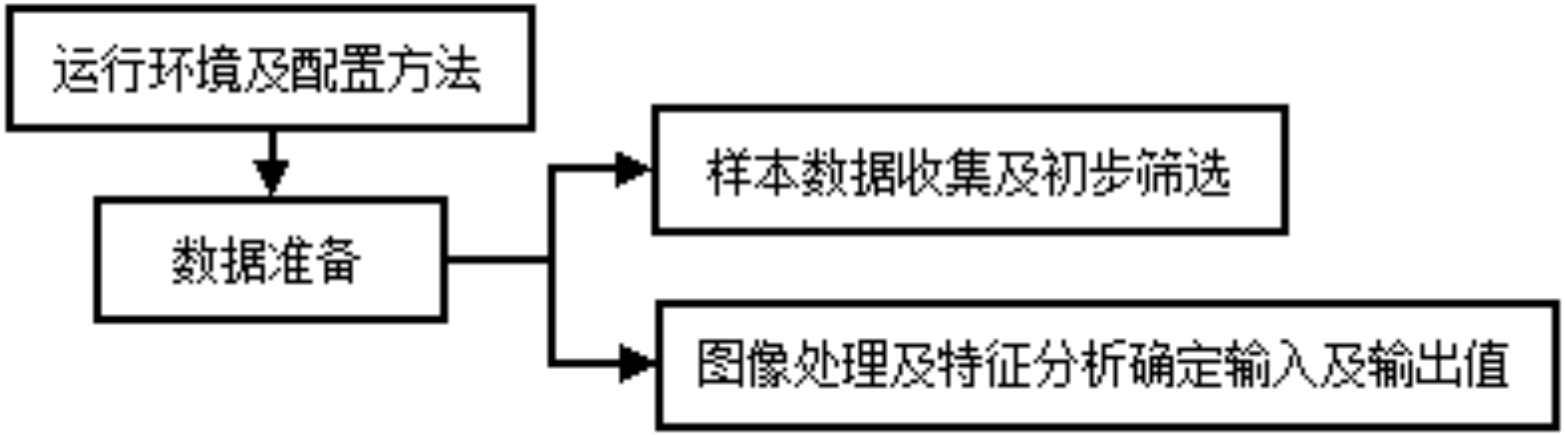
## ➢ BP神经网络工作原理

由信息的正向传播和误差的反向传播两个过程组成。

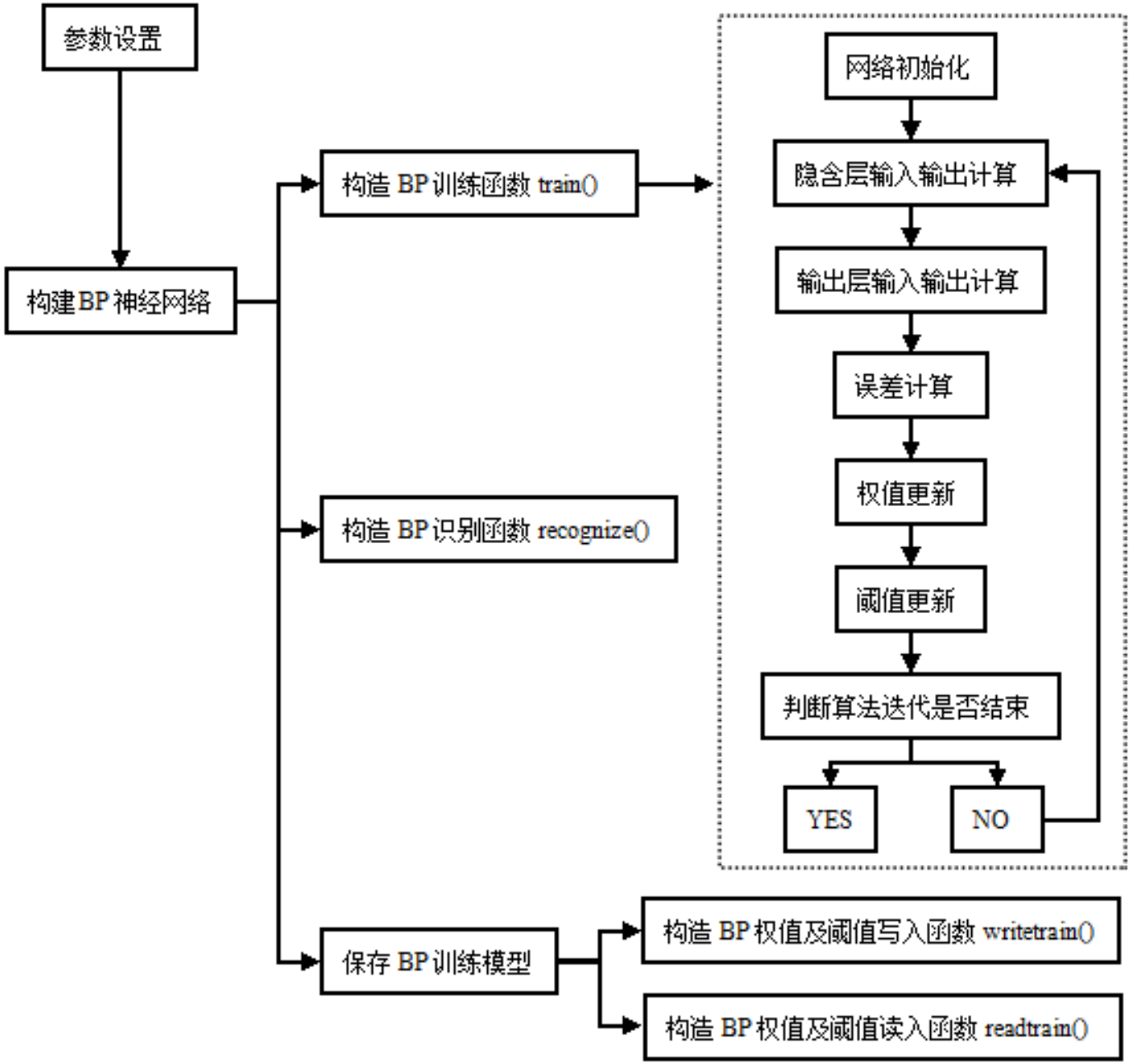


# （1）算法设计及算法流程

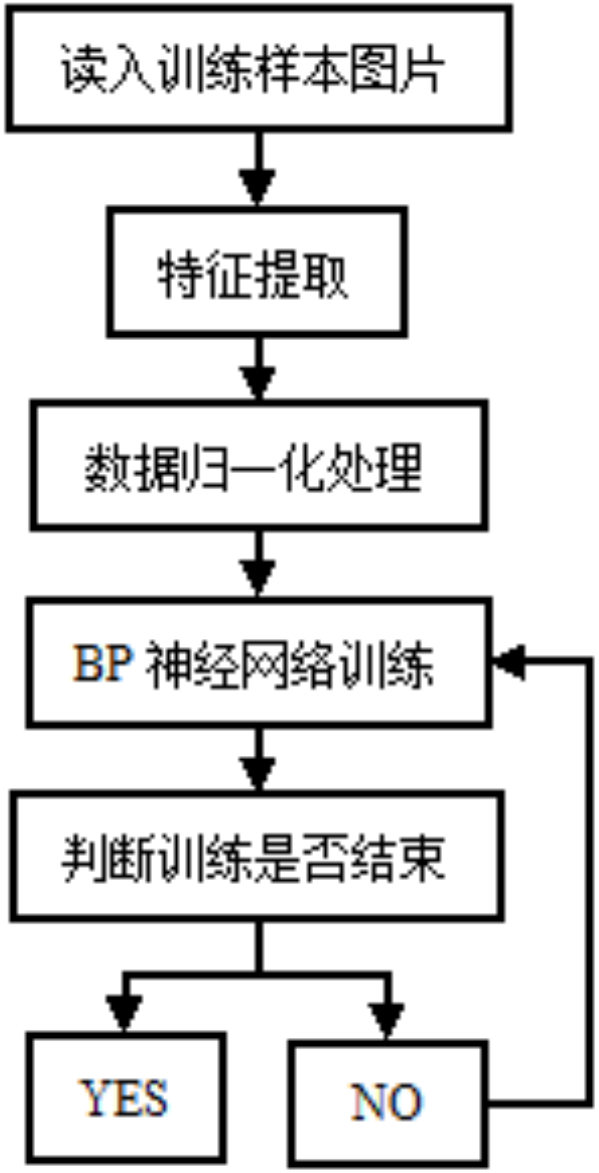
（1）前期准备：



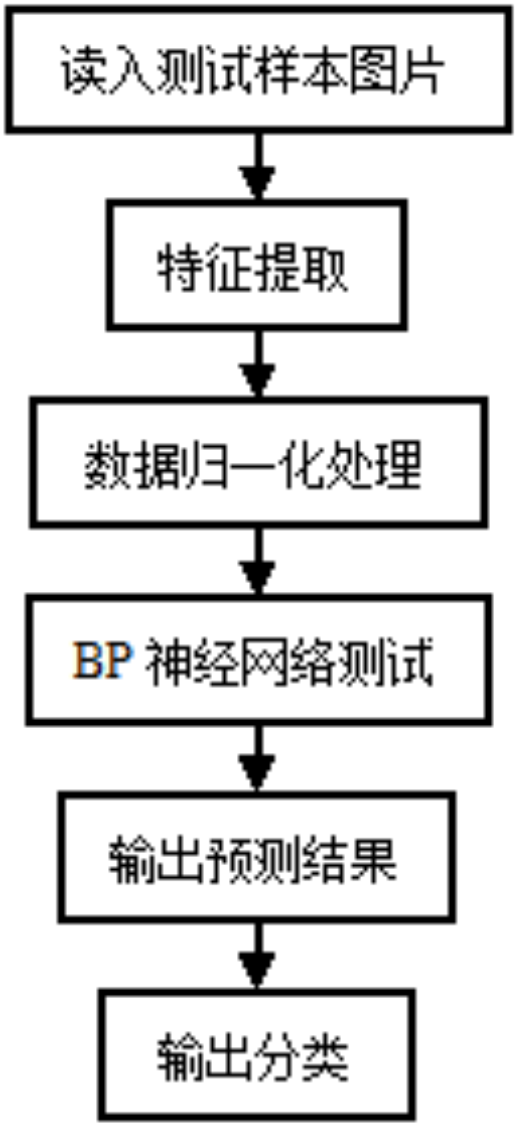
（2）BP 神经网络构建：


参数设置
构建BP神经网络
构造 BP 训练函数 train()
构造 BP 识别函数 recognize()
保存 BP 训练模型
构造 BP 权值及阈值写入函数 writetrain()
构造 BP 权值及阈值读入函数 readtrain()
网络初始化
隐含层输入输出计算
输出层输入输出计算
误差计算
权值更新
阈值更新
判断算法迭代是否结束
YES
NO

（3）BP 神经网络训练：

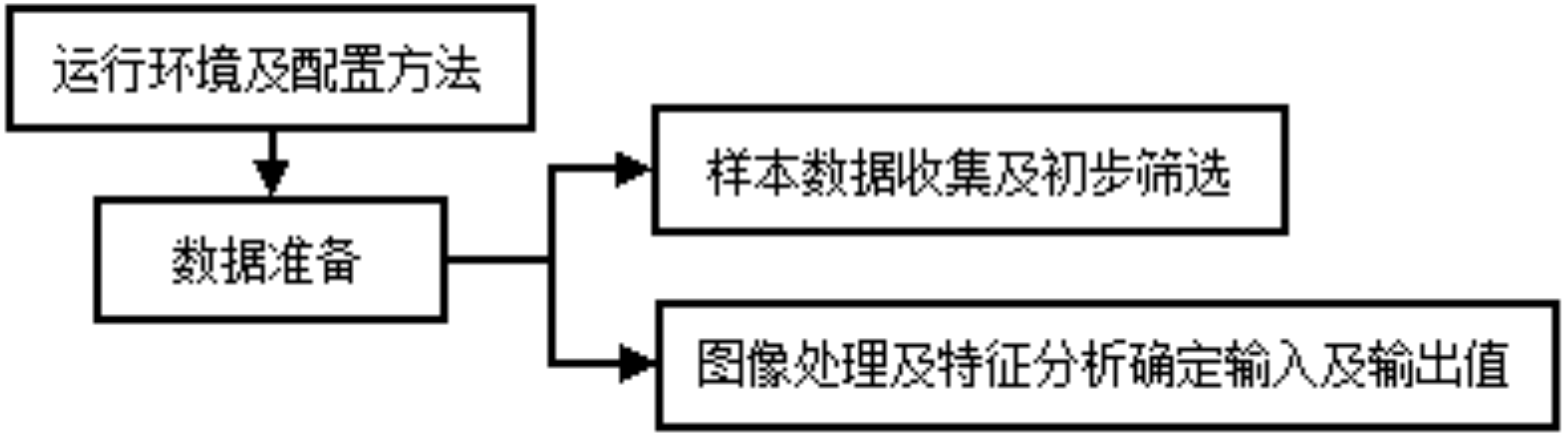



（4）BP 神经网络测试及分类：



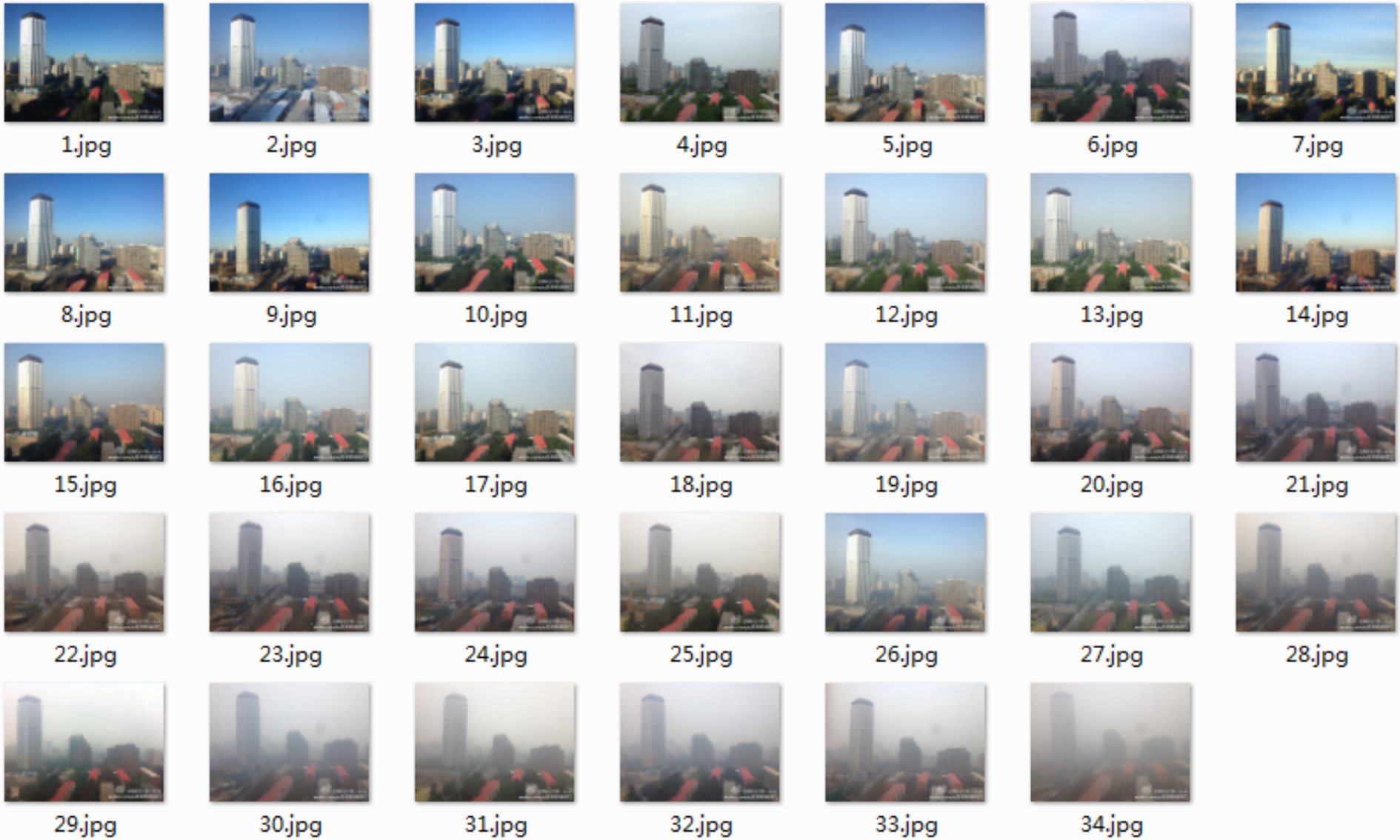
# （1）算法设计及算法流程

（1）前期准备：



# （1）数据准备

- ①　样本数据的收集及初步筛选




训练样本集train_samples

1.jpg

2.jpg

3.jpg

4.jpg

5.jpg

6.jpg

7.jpg

8.jpg

9.jpg

测试样本集test_samples

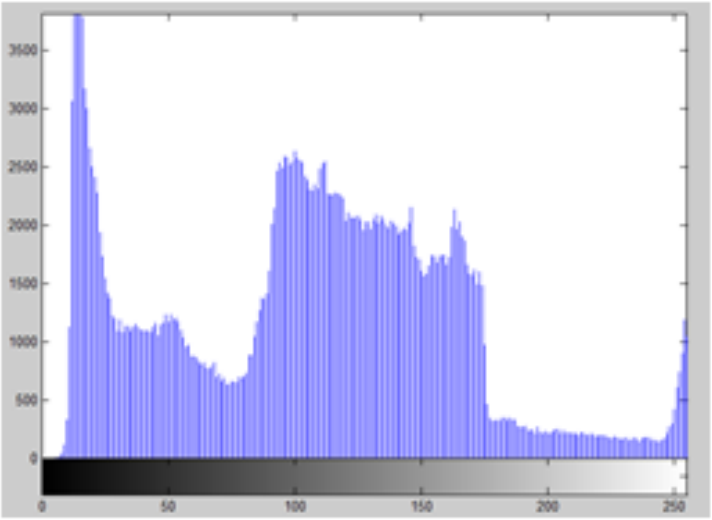
- ② 通过图像处理及特征分析确定输入及输出值

( a) AQI=10

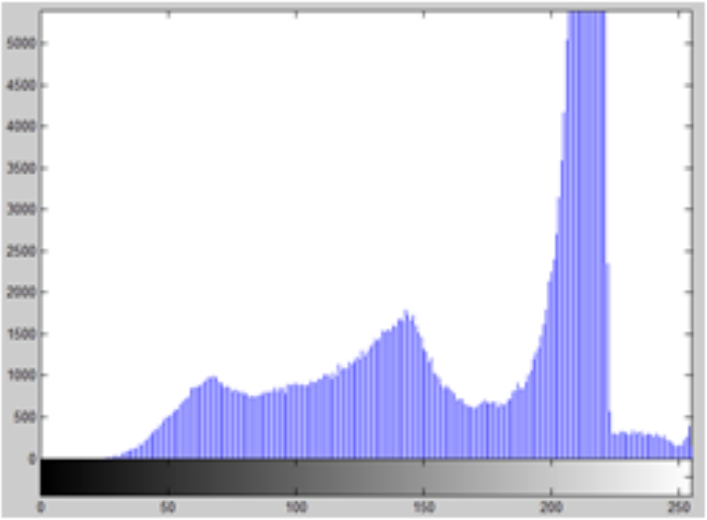
(b) AQI=125

( c) AQI=205

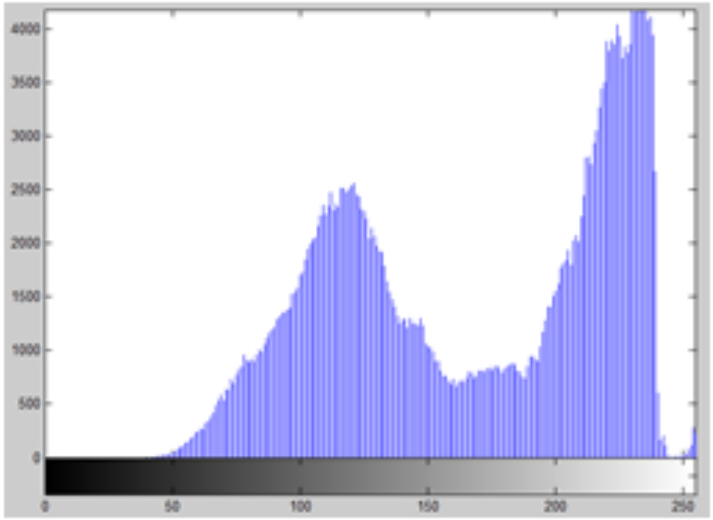
(d) AQI=280



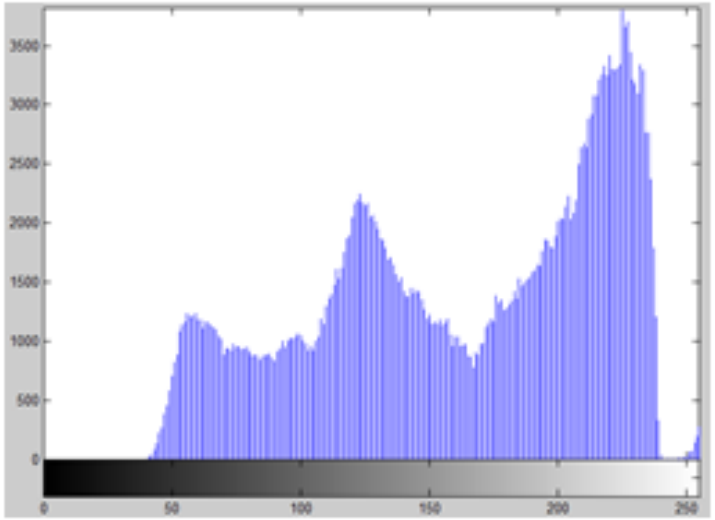

( a) AQI=10




(b) AQI=125

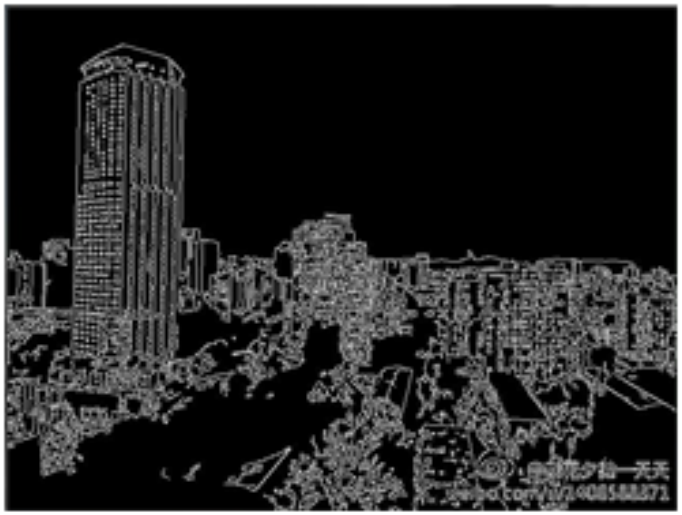



( c) AQI=205

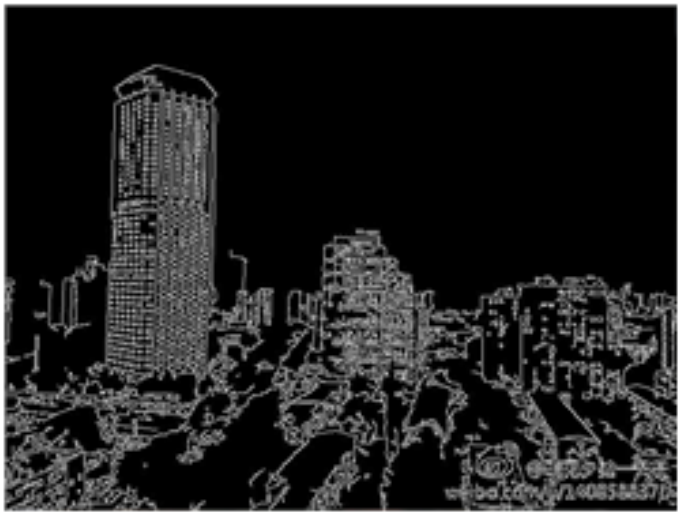



(d) AQI=280

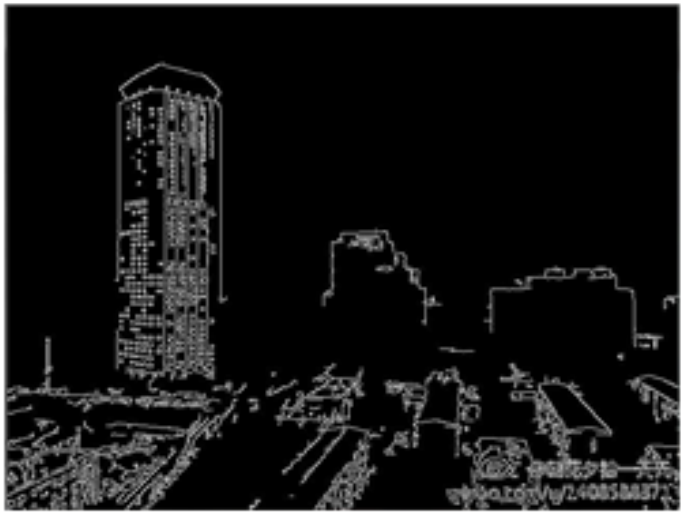
不同AQI值图像的灰度直方图



( a) AQI=10



(b) AQI=125

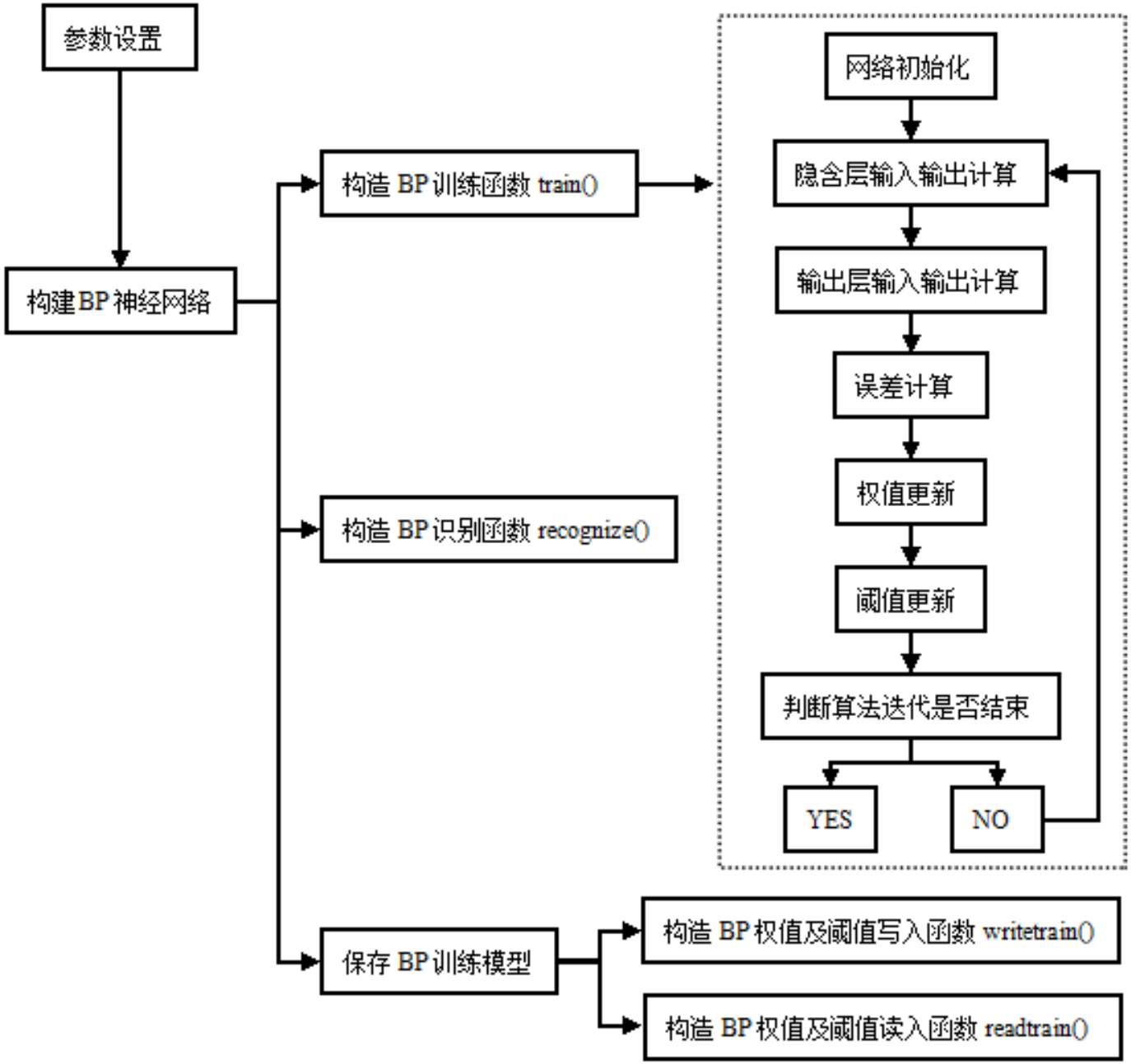


( c) AQI=205

(d) AQI=280

不同AQI值图像的canny边缘检测图

（2）BP 神经网络构建：


参数设置
构建BP神经网络
构造 BP 训练函数 train()
构造 BP 识别函数 recognize()
保存 BP 训练模型
构造 BP 权值及阈值写入函数 writetrain()
构造 BP 权值及阈值读入函数 readtrain()
网络初始化
隐含层输入输出计算
输出层输入输出计算
误差计算
权值更新
阈值更新
判断算法迭代是否结束
YES
NO

# （2）BP神经网络构建

➢ 1.参数设置

表 4.1　参数设置

| 参数 | 设置值 |
| --- | --- |
| 输入节点数（innode） | 2 |
| 隐含节点数（hidenode） | 10 |
| 输出节点数（outnode） | 1 |
| 训练样本数（trainsample） | 34 |
| 输入层和隐含层的权值学习率（rate_$w$） | 0.9 |
| 隐含层和输出层的权值学习率（rate_$w_1$） | 0.9 |
| 隐含层阀值学习率（rate_$b_1$） | 0.9 |
| 输出层阀值学习率（rate_$b_2$） | 0.9 |

# （2）BP神经网络构建

- 构造BP训练函数train()

① 网络初始化。初始化所有的网络权值为-1 和 1 之间的小随机数。初始化输入层，隐含层和输出层神经元之间的连接权值 $w$， $w_1$，初始化隐含层阈值 $b_1$，输出层阈值 $b_2$。初始化公式为：

$$w[i] = (2.0 * (double)rand() / RAND_MAX) - 1$$

# （2）BP神经网络构建

- 构造BP训练函数train()

② 隐含层输入输出计算。根据输入变量 $X$，输入层和隐含层间连接权值 $w_{ij}$ 以及隐含层阈值 $b_1$，计算隐含层输入 $O_1$ 及隐含层输出 $X_1$。

$$O_{1j} = \sum_{i=1}^{n} w_{ij} \cdot x_i \qquad j = 1,2\cdots,l$$

$$X_{1j} = f(O_{1j} + b_{1j})$$

# （2）BP神经网络构建

- 构造BP训练函数train()

式中，$l$ 为隐含层节点数，$f$ 为隐含层激励函数，该函数有多种表达形式，本文所选函数为 $S$ 型函数如图，其数值域是(0,1)。

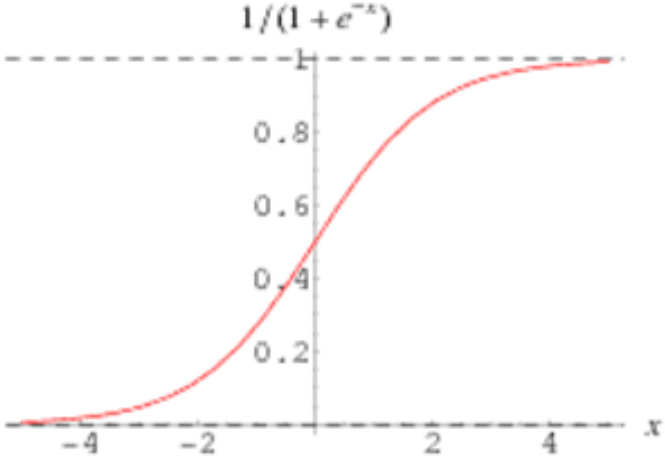



图 4.2　S 型函数

$$f(x)=\frac{1}{1+e^{-x}}$$

# （2）BP神经网络构建

- 构造BP训练函数train()

③ 输出层输入输出计算。根据隐含层输出 $X_1$，连接权值 $w_{jk}$ 和阈值 $b_2$，计算 BP 神经网络输出层输入 $O_2$ 及输出层输出 $X_2$。

$$O_{2k} = \sum_{j=1}^{l} X_{1j} \cdot w_{jk} \qquad k = 1,2,\ldots,m$$

$$X_{2k} = f(O_{2k} + b_{2k})$$

# （2）BP神经网络构建

- 构造BP训练函数train()

④ 误差计算。根据网络输出层输出 $X_2$ 期望输出 Y，计算网络误差 $E$ 。

1）隐含层节点和输出层节点之间连接权值修正量 $\delta_k$ 的计算

$$\delta_k = (Y_k - X_{2k}) \cdot X_{2k} \cdot (1 - X_{2k}) \qquad k = 1,2,\ldots,m$$

2）输入层节点和隐含层节点之间连接权值修正量 $\delta_j$ 的计算

$$\delta_j = X_{1j} \cdot (1 - X_{1j}) \cdot \sum_{k=1}^{m} \delta_k \cdot w_{jk} \qquad j = 1,2,\ldots,l$$

3）误差 $E$

$$E = \frac{1}{2} \sum_{k=1}^{m} (Y_k - X_{2k})^2 \qquad k = 1,2,\ldots,m$$

# （2）BP神经网络构建

- 构造BP训练函数train()

⑤ 权值更新。根据隐含层节点和输出层节点之间连接权值修正量$\delta_k$及输入层节点和隐含层节点之间连接权值修正量$\delta_j$及学习速率$\eta$，更新网络连接权值$w_{jk}$和$w_{ij}$。

$$w_{jk} = w_{jk} + \eta \cdot \delta_k \cdot X_{1j} \qquad j = 1,2,\cdots,l\,;\, k = 1,2,\ldots,m$$

$$w_{ij} = w_{ij} + \eta \cdot \delta_j \cdot x_i \qquad i = 1,2,\ldots,n\,;\, j = 1,2,\ldots,l$$

# （2）BP神经网络构建

- 构造BP训练函数train()

⑥ 阈值更新。根据隐含层节点和输出层节点之间连接权值修正量$\delta_k$及输入层节点和隐含层节点之间连接权值修正量$\delta_j$及学习速率$\eta$，更新网络节点阈值$b_2$和$b_1$。

$$b_{2k} = b_{2k} + \eta \cdot \delta_k \qquad k = 1,2,\ldots,m$$

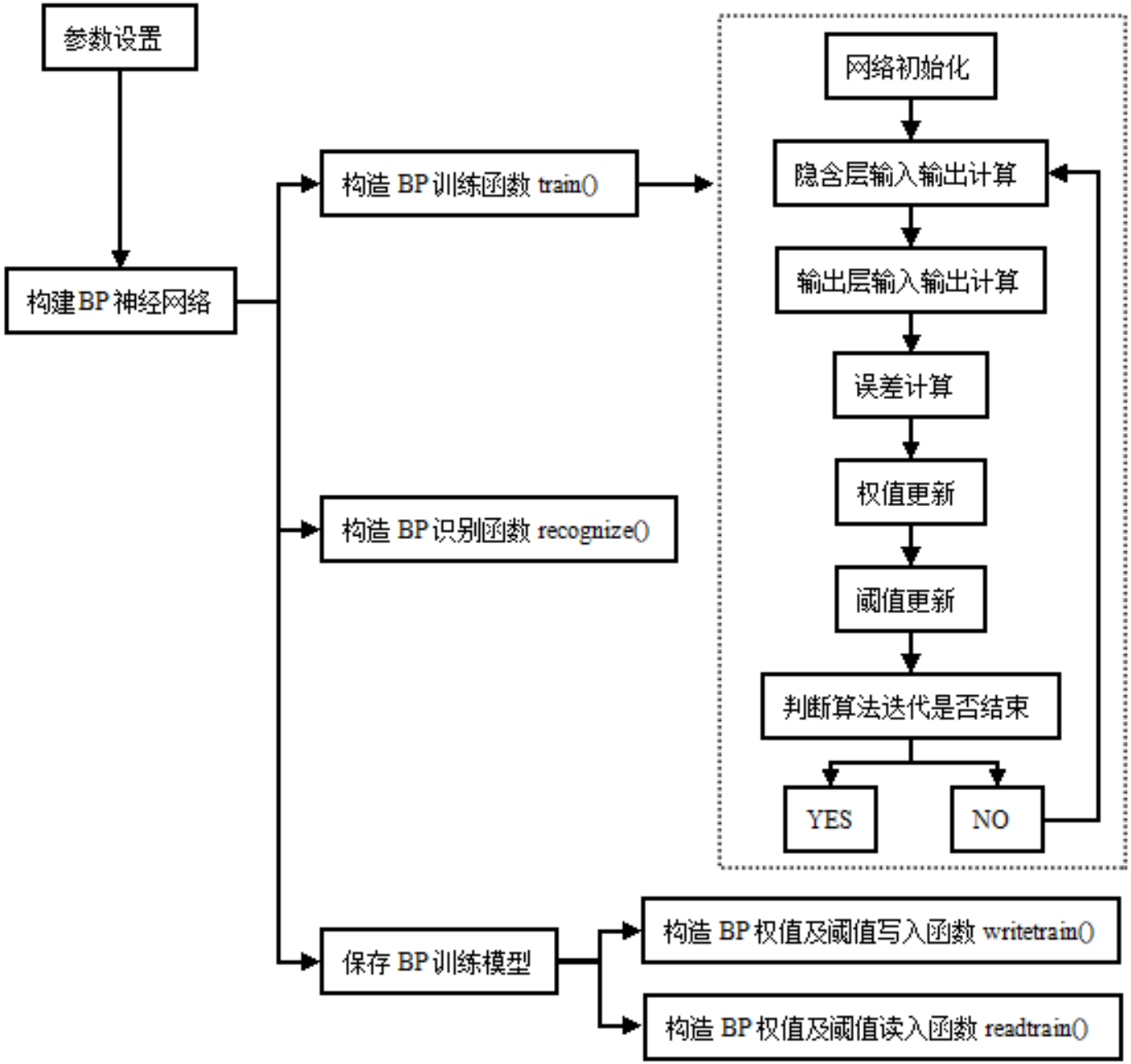
$$b_{1j} = b_{1j} + \eta \cdot \delta_j \qquad j = 1,2,\ldots,l$$

# （2）BP神经网络构建

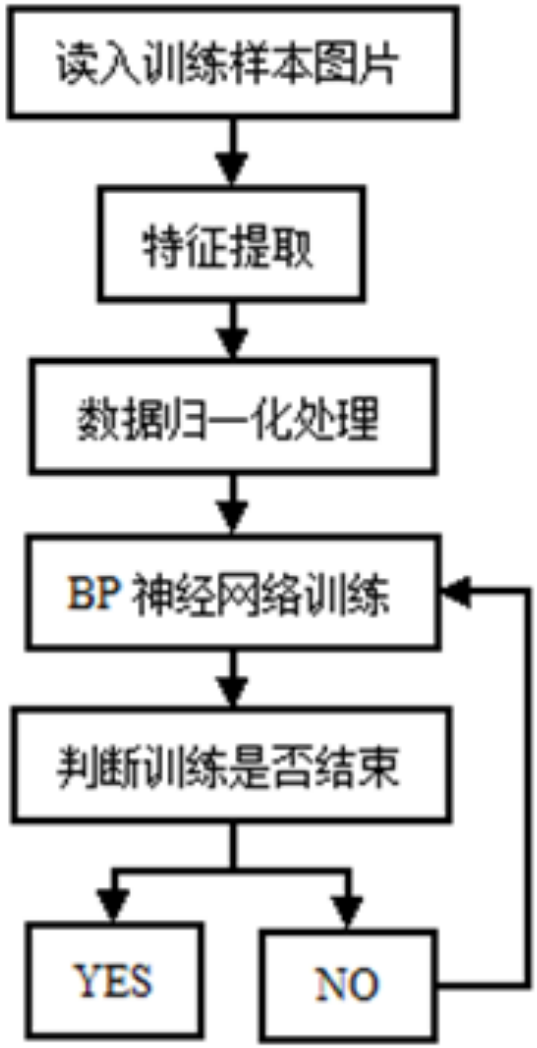
- 构造BP训练函数train()

⑦ 判断算法迭代是否结束。即判断误差 $E$ 是否满足要求。当误差达到预设精度或者学习次数大于设定的最大次数，则结束算法。否则选取下一个训练样本及对应的期望输出，返回到第②步，进入下一轮训练。

（2）BP 神经网络构建：


参数设置
构建BP神经网络
构造 BP 训练函数 train()
构造 BP 识别函数 recognize()
保存 BP 训练模型
构造 BP 权值及阈值写入函数 writetrain()
构造 BP 权值及阈值读入函数 readtrain()
网络初始化
隐含层输入输出计算
输出层输入输出计算
误差计算
权值更新
阈值更新
判断算法迭代是否结束
YES
NO

（3）BP 神经网络训练：



# （3）BP神经网络训练

- 构造BP训练阶段函数train_step()

1.对训练数据提取特征并归一化

数值归一化处理能够较好的改变网络的训练效率。因为$S$型函激励函数输出在（0,1）之间，所以要对数据进行归一化处理，在归一化处理的函数选择了线性归一化函数，归一化公式为：

$$x^* = \frac{x - \min}{\max - \min}$$

表 4.2　训练样本的特征数据归一化后的值

| 测试样本序号 | 白比 $result_1$ | 边缘统计 $result_2$ | 测试样本序号 | 白比 $result_1$ | 边缘统计 $result_2$ |
|---|---|---|---|---|---|
| 1 | 0.000000 | 0.831101 | 18 | 0.530612 | 0.683840 |
| 2 | 0.653061 | 1.000000 | 19 | 1.000000 | 0.629080 |
| 3 | 0.469388 | 0.823630 | 20 | 0.571429 | 0.658189 |
| 4 | 0.448980 | 0.845244 | 21 | 0.408163 | 0.644335 |
| 5 | 0.551020 | 0.813307 | 22 | 0.408163 | 0.497558 |
| 6 | 0.346939 | 0.789227 | 23 | 0.367347 | 0.483197 |
| 7 | 0.551020 | 0.791499 | 24 | 0.489796 | 0.552004 |
| 8 | 0.428571 | 0.772666 | 25 | 0.510204 | 0.453605 |
| 9 | 0.530612 | 0.761593 | 26 | 0.795918 | 0.486195 |
| 10 | 0.632653 | 0.919951 | 27 | 0.653061 | 0.367656 |
| 11 | 0.693878 | 0.806489 | 28 | 0.489796 | 0.211595 |
| 12 | 0.714286 | 0.823074 | 29 | 0.551020 | 0.206711 |
| 13 | 0.836735 | 0.914487 | 30 | 0.653061 | 0.118781 |
| 14 | 0.653061 | 0.680262 | 31 | 0.653061 | 0.123664 |
| 15 | 0.551020 | 0.695228 | 32 | 0.673469 | 0.149993 |
| 16 | 0.816327 | 0.695856 | 33 | 0.714286 | 0.176708 |
| 17 | 0.714286 | 0.772134 | 34 | 0.795918 | 0.000000 |

表 4.3　训练样本的期望输出 AQI 值归一化后的值

| 测试样本序号 | AQI 原始输出值 | AQI 归一化值 | 测试样本序号 | AQI 原始输出值 | AQI 归一化值 |
|---|---|---|---|---|---|
| 1 | 10 | 0.000000 | 18 | 140 | 0.382353 |
| 2 | 15 | 0.014706 | 19 | 150 | 0.411765 |
| 3 | 40 | 0.088235 | 20 | 160 | 0.441176 |
| 4 | 45 | 0.102941 | 21 | 165 | 0.455882 |
| 5 | 55 | 0.132353 | 22 | 170 | 0.470588 |
| 6 | 60 | 0.147059 | 23 | 175 | 0.485294 |
| 7 | 65 | 0.161765 | 24 | 185 | 0.514706 |
| 8 | 70 | 0.176471 | 25 | 190 | 0.529412 |
| 9 | 80 | 0.205882 | 26 | 200 | 0.558824 |
| 10 | 85 | 0.220588 | 27 | 205 | 0.573529 |
| 11 | 90 | 0.235294 | 28 | 210 | 0.588235 |
| 12 | 95 | 0.250000 | 29 | 220 | 0.617647 |
| 13 | 105 | 0.279412 | 30 | 240 | 0.676471 |
| 14 | 110 | 0.294118 | 31 | 260 | 0.735294 |
| 15 | 120 | 0.323529 | 32 | 280 | 0.794118 |
| 16 | 125 | 0.338235 | 33 | 300 | 0.852941 |
| 17 | 135 | 0.367647 | 34 | 350 | 1.000000 |

# （3）BP神经网络训练

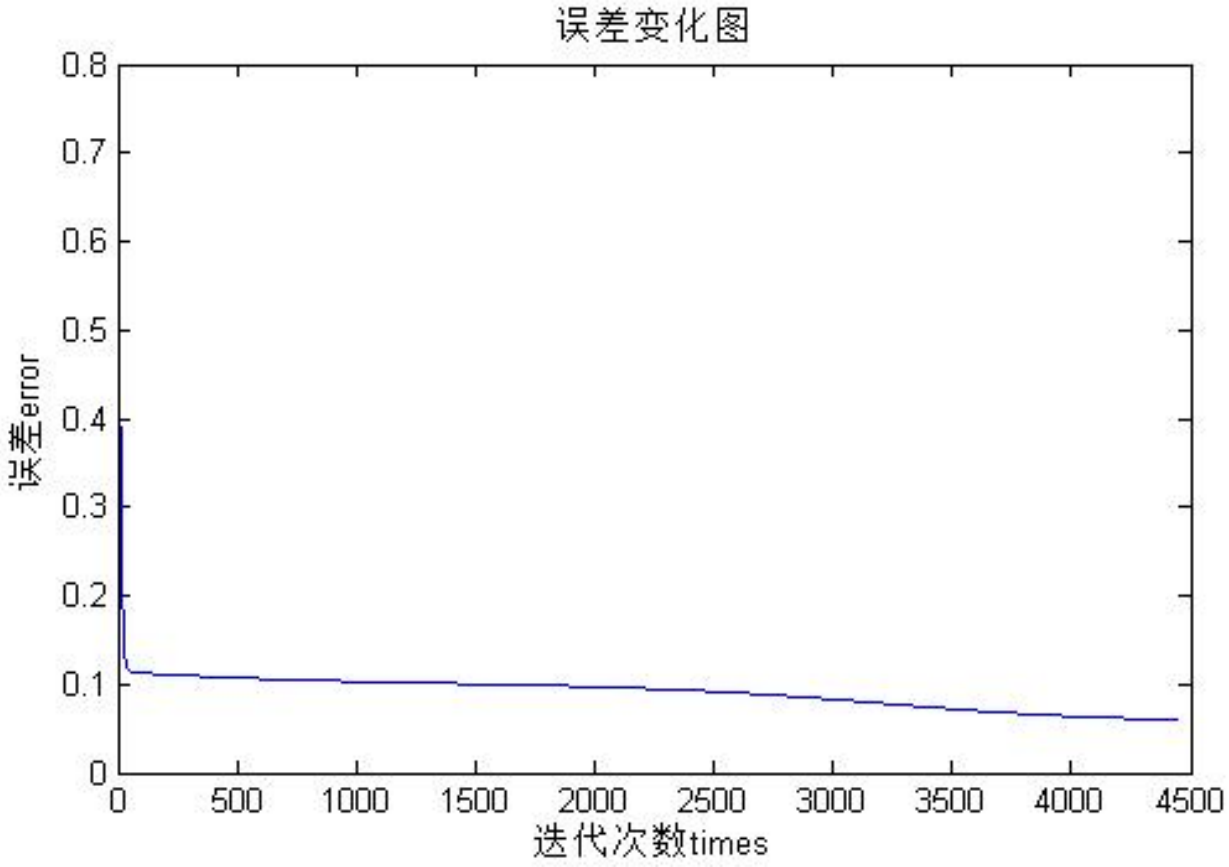
2.用归一化后的训练数据训练BP神经网络

设定允许的最大误差为0.06

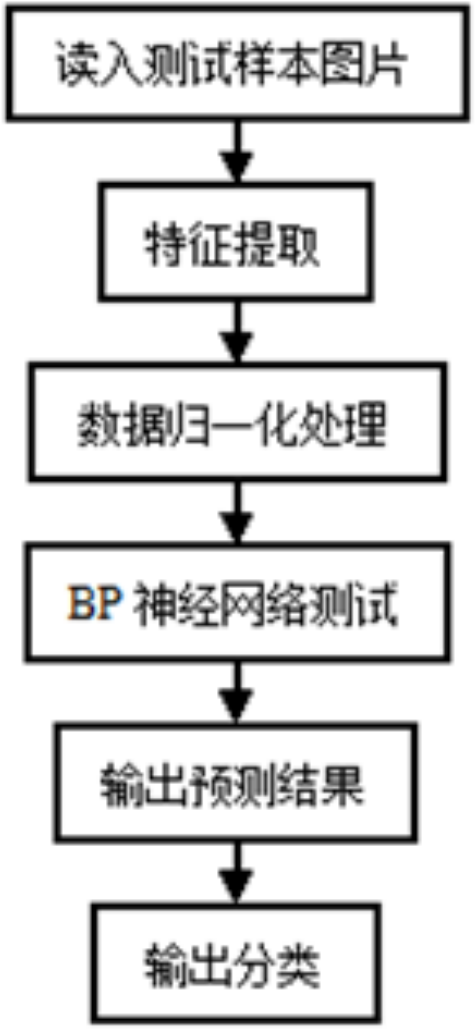
进行训练。用训练样本集中的数据对BP网络进行训练，在训练过程中根据网络误差调整网络的权值和阈值，直到迭代误差小于设定的最大误差，停止训练，训练阶段完成，进入测试阶段。

# （3）BP神经网络训练

2.用归一化后的训练数据训练BP神经网络，保存训练模型



（4）BP 神经网络测试及分类：



| AQI | AQI ° ⅝ | AQI ¿ ⅜ | —?…´ | ∂ Ω ϑ µ ƒ œ Ï | Ω ® È »..µ ¥ Î © |
|---|---|---|---|---|---|
| 0~50 | “ ª ∂ | ” ≈ | ¬ Ã.´ | ª È ⅝ Δ ÷ ¡ ø Ó À “ £” ± ; Œ6 Δ Œ6 ⅝ £ | ∏ ¿ ⅜ À ∫ ?..≥ £ Ó Ø £ |
| 51~100 | ∂ ° ¿ ∂ | ¡ ° | ª Δ.´ | ª È ⅝ Δ ÷ ¡ ø Ω” £µ ƒ ≥ Œ6 ⅝ Ô ..ƒ ∂ ° ?…Ÿ “ È ⅝ ∏ » À ∫ Ω ϑ µ –Ω œ ï’ œ Ï £ | ° ?…Ÿ “ È ⅝ ∏ » À ∫’ ¶ ı..Ÿ ª Ø ª Ó Ø £ |
| 101~150 | » º ∂ « ∂ Œ6 ∂ | » ≥ ».´ | “ ∏ » À ⅟ ø ¥ « ∂ ϑ ⅝ Ω ° ø µ À ⅟ ⅝ Ã § ø ¥ £ | ∂ Õ Ø ⅝ ⅝ À ∞ ƒ ⅜ ° ⅝ È œ Ø ≥ ≤ ° ° ⅜ ¶ ı..Ÿ § ± % ∏ ⅜ ∂ µ ƒ Ø ∂ Ø ∂ £ | ∂ Õ Ø ⅝ ⅝ À ∞ ƒ ⅜ ° ⅝ È œ Ø ≥ ≤ ° ° ⅜ ¶ ı..Ÿ § ± % |
| 151~200 | À ƒ ∂ ÷ ∂ Œ6 ∂ | ∫ Ï..´ | Ω “ ≤ Ω ⅝ Á ∏ » À ⅟ ø ¥ö ƒ ∂ ‘ Ω ϑ µ À ⅟ ƒ ⅜ ⅝ È ⅝ Ø Ø ≥ ” œ Ï £ | ∂ Õ Ø ⅝ ⅝ À ∞ ƒ ⅜ ° ⅝ È œ Ø ≥ ≤ ° ° ⅜ √ ≥ § ± % ∏ » ∂ µ ƒ Ø ∂ Ø ⅜ ∞ » À ∫ ¡ ϑ ı..Ÿ Ø ; À Ø £ | |
| 201~300 | Œ Â ∂ ÷ ∂ Œ6 ∂ | ◊ œ.´ | – ƒ ⅜ ∫ Ø ⅜ ° ⅜ ø ¥ ÷ – ° ⅝ Á À Ø Ø Ø ¡ ⅜ µ Ø ϑ µ » À ⅟ Δ ± È ⅝ ÷ ø ¥ £ | ∂ Õ Ø ⅝ ⅝ À Ø ƒ ⅜ ° ⅜ Œ ª ª ° ⅜ Ø ø ¥ £ Ù/ ƒ £ Ø £ ⅜ Ø ‘ À Ø ⅜ ∞ » À ∫ ı..Ÿ Ø ; À ∂ Ø £ | |
| >300 | ¡ º ∂ —œ ⅝ Œ6 ∂ | ∫ ⅟ Ï …´ | Ω ϑ µ À ƒ Ø ¡ ⅜ µ Ø √ œ ‘ « ø ⅜ ø ¥ Ã « ≥ ⅝ ƒ ≥ © ≤ ≤ ° £ | ∂ Õ Ø ⅝ ⅝ À Ø ϑ À ⅜ ± Ù/ ƒ £ ± √ Ã Â ⅜ ∫ ƒ ⅜ ∞ » » ∫’ ¶ √ ª ; Ø ª ; Ó Ø £ | |

# （4）BP神经网络测试及分类

输入已经归一化后的待测样本数据，调用bp.recognize()函数，进行测试并输出预测结果。

通过反归一化公式计算出相应的预测 AQI 值。反归一化公式为：

$$x = (\max - \min) \cdot x^* + \min$$

# 四、测试结果及分析

ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
1
输入
测试图像的原始AQI:
35
测试图像的原始空气级别:
优
测试数据
测试图像的预测AQI:
31.8329
测试图像的预测空气级别:
优
AQI误差:
9.04877
%
测试图像

ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
2
输入
测试图像的原始AQI:
85
测试图像的原始空气级别:
良
测试数据
测试图像的预测AQI:
91.4481
测试图像的预测空气级别:
良
AQI误差:
7.58599
%
测试图像
weibo.com/u/1408588371

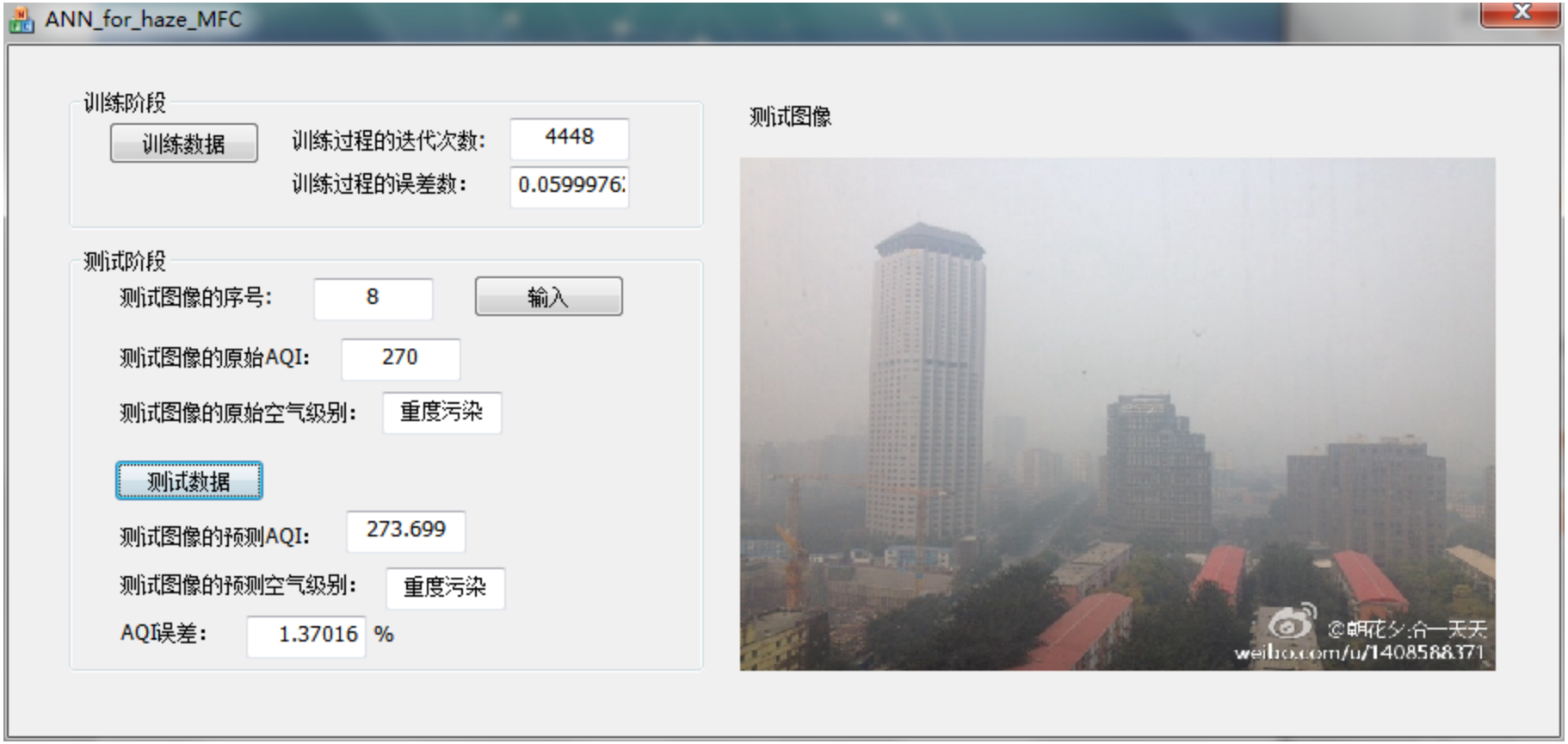
ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
3
输入
测试图像的原始AQI:
90
测试图像的原始空气级别:
良
测试数据
测试图像的预测AQI:
77.5557
测试图像的预测空气级别:
良
AQI误差:
13.827
%
测试图像
@朝花夕拾一天天
weibo.com/u/1408588371

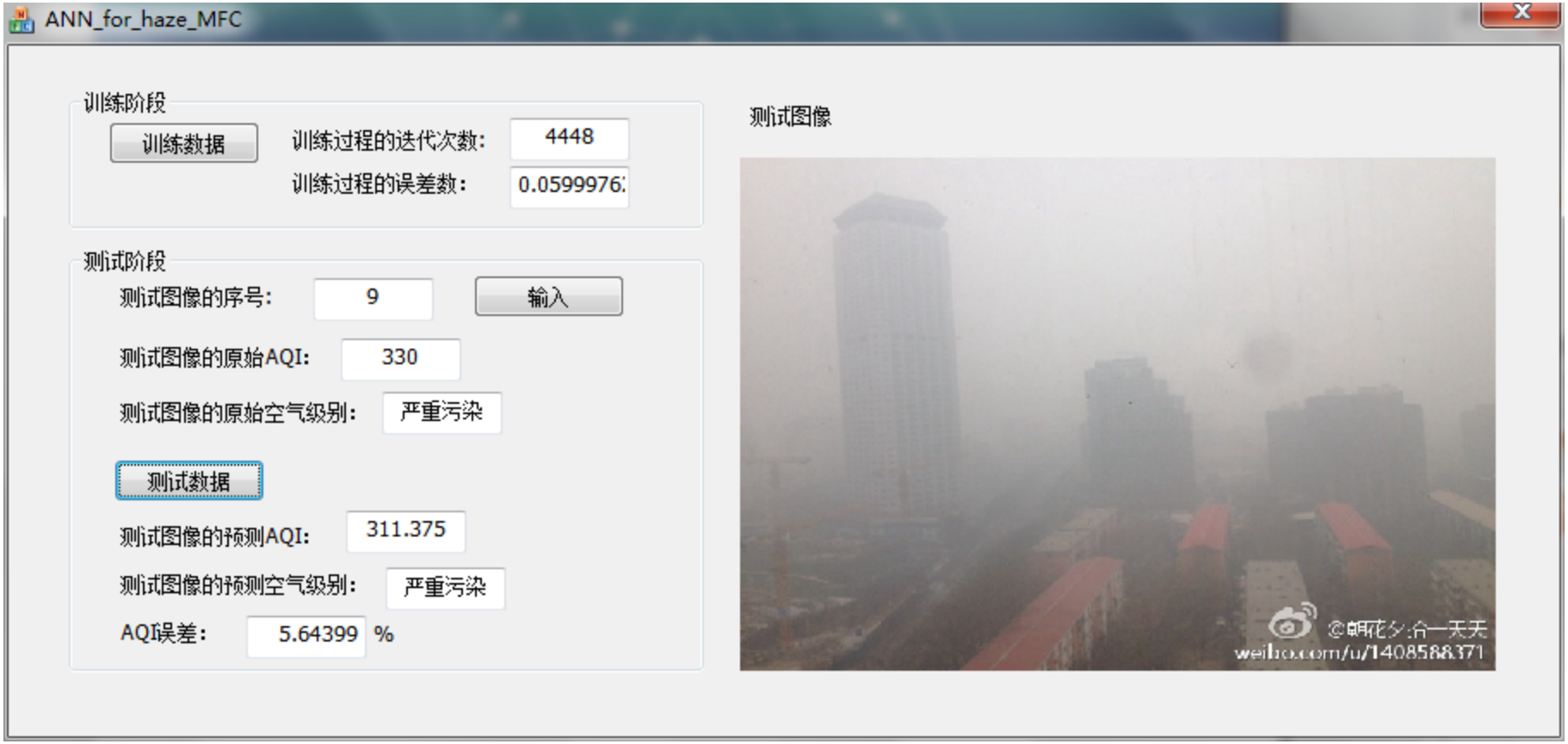
ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
4
输入
测试图像的原始AQI:
135
测试图像的原始空气级别:
轻度污染
测试数据
测试图像的预测AQI:
145.918
测试图像的预测空气级别:
轻度污染
AQI误差:
8.08736
%
测试图像
weibo.com/u/1408588371

ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
5
输入
测试图像的原始AQI:
185
测试图像的原始空气级别:
中度污染
测试数据
测试图像的预测AQI:
175.595
测试图像的预测空气级别:
中度污染
AQI误差:
5.08386
%
测试图像

ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
6
输入
测试图像的原始AQI:
195
测试图像的原始空气级别:
中度污染
测试数据
测试图像的预测AQI:
187.961
测试图像的预测空气级别:
中度污染
AQI误差:
3.60964
%
测试图像
@朝花夕拾一天天
weibo.com/u/1408588371

ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
7
输入
测试图像的原始AQI:
205
测试图像的原始空气级别:
重度污染
测试数据
测试图像的预测AQI:
208.485
测试图像的预测空气级别:
重度污染
AQI误差:
1.7002
%
测试图像
@朝花夕拾一天天
weibo.com/u/1408588371


ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
8
输入
测试图像的原始AQI:
270
测试图像的原始空气级别:
重度污染
测试数据
测试图像的预测AQI:
273.699
测试图像的预测空气级别:
重度污染
AQI误差:
1.37016
%
测试图像
@朝花夕拾一天天
weibo.com/u/1408588371


ANN_for_haze_MFC
训练阶段
训练数据
训练过程的迭代次数:
4448
训练过程的误差数:
0.0599976
测试阶段
测试图像的序号:
9
输入
测试图像的原始AQI:
330
测试图像的原始空气级别:
严重污染
测试数据
测试图像的预测AQI:
311.375
测试图像的预测空气级别:
严重污染
AQI误差:
5.64399
%
测试图像
@朝花夕拾一天天
weibo.com/u/1408588371

# 四、测试结果及分析

- 用测试样本集中的9组数据对BP神经网络进行测试，结果如下表所示

表 5.1　测试结果

| 测试样本序号 | 原始AQI | 原始空气质量级别 | 预测AQI | 预测空气质量级别 | AQI误差 |
|---|---|---|---|---|---|
| 1 | 35 | 优 | 31.8329 | 优 | 9.04877% |
| 2 | 85 | 良 | 91.4481 | 良 | 7.58599% |
| 3 | 90 | 良 | 77.5557 | 良 | 13.827% |
| 4 | 135 | 轻度污染 | 145.918 | 轻度污染 | 8.08736% |
| 5 | 185 | 中度污染 | 175.595 | 中度污染 | 5.08386% |
| 6 | 195 | 中度污染 | 187.961 | 中度污染 | 3.60964% |
| 7 | 205 | 重度污染 | 208.485 | 重度污染 | 1.7002% |
| 8 | 270 | 重度污染 | 273.699 | 重度污染 | 1.37016% |
| 9 | 330 | 严重污染 | 311.375 | 严重污染 | 5.64399% |

# 四、测试结果及分析

- 其中，位于中间部分的测试样本即5号，6号，7号，8号的测试结果比较好，预测误差都低于5.1%，平均测试误差在3.0%以下。可能的原因有1）在训练样本集中，中部的数据分布较多，对测试的准确度有一定影响。2）测试的样本数据中，对这四幅图片提取的白比和边缘统计特征比较有分析力，导致误差比较小，预测精度高。

表 5.1　测试结果

| 测试样本序号 | 原始AQI | 原始空气质量级别 | 预测AQI | 预测空气质量级别 | AQI误差 |
|---|---|---|---|---|---|
| 1 | 35 | 优 | 31.8329 | 优 | 9.04877% |
| 2 | 85 | 良 | 91.4481 | 良 | 7.58599% |
| 3 | 90 | 良 | 77.5557 | 良 | 13.827% |
| 4 | 135 | 轻度污染 | 145.918 | 轻度污染 | 8.08736% |
| 5 | 185 | 中度污染 | 175.595 | 中度污染 | 5.08386% |
| 6 | 195 | 中度污染 | 187.961 | 中度污染 | 3.60964% |
| 7 | 205 | 重度污染 | 208.485 | 重度污染 | 1.7002% |
| 8 | 270 | 重度污染 | 273.699 | 重度污染 | 1.37016% |
| 9 | 330 | 严重污染 | 311.375 | 严重污染 | 5.64399% |

# 四、测试结果及分析

- 其中，位于两边的测试样本即1号，2号，3号，4号，9号的测试结果相对较差，预测误差在5.5%以上，平均测试误差在8.8%左右。其中3号图片的预测误差达到了13.8%，出现了较大的误差。可能的原因有1）在训练样本集中，两边的数据分布较少，BP神经网络没有得到足够的训练，导致测试准确度下降。2）由于图片本身的质量问题，如图片拍摄的时间、地点和背景，天气原因如多云，阴天，雨雪等会影响拍摄效果及清晰度，导致了对五幅图片提取的白比和边缘统计特征不能够准确反映图片的特征，最终导致预测误差比较大，预测精度下降。

# 四、测试结果及分析

- 对测试样本集中的9幅图片进行测试，所设计的BP神经网络分类器都能够对空气质量级别做出准确的分类判断，而且BP神经网络总体的AQI预测精度达到了93.78%，平均的AQI测试误差为6.22%，总体测试结果比较理想。

# 误差来源

- 总体来说，影响BP神经网络预测准度的可能误差来源有以下四个。

（1）训练数据数目

神经网络预测的准确性和训练数据的多少有较大关系，尤其对于一个多输入多输出的网络，如果缺乏足够的网络训练数据，网络预测值可能存在较大的误差。本文为单隐层的BP神经网络，训练样本集中的数据不足，导致BP神经网络没有得到充分训练，BP神经网络预测值和期望值之间误差较大。

（2）特征的数目

更多的属性数目对于生成器而言意味着要计算更多的组合，使得生成器难度增大，需要的时间也更长。有时随机的关系会将生成器引入歧途，结果可能构造出不够准确的分类器（这在技术上被称为过分拟合）。因此，如果我们通过常识可以确认某个属性与目标无关，则将它从训练集中移走。本文选用的两个特征输入，特征数目偏少。

（3）特征中的信息

有时生成器不能从属性中获取足够的信息来正确、错误地预测标签（如试图根据某人眼睛的颜色来决定他的收入）。本文可以加入其他的特征属性，来降低错误率。

（4）待预测集中样本的分布

如果待预测样本来自不同于训练集中样本的分布，那么错误率有可能很高。因为数据属性值的分布可能是有很大差别的。本文中对测试集中的样本与训练集中的样本虽然来自同一数据源，但是也不能保证样本分布是无差别的。

# 改进方法

- 对于BP神经网络如何提高训练精度，降低预测误差，除了增加训练次数提高预设允许的最大误差外，有如下几个改进方法。

（1）调整网络结构

增加网络的层数可以进一步降低误差，提高精度，但会使网络复杂化，从而增加网络的训练时间。精度的提高实际上也可以通过增加隐层神经元的数目来获得，其效果更容易观察和掌握，所以应优先考虑。

（2）初始值选取

为了使误差尽可能小，需要合理选择初始权重和偏置，如果太大就容易陷入饱和区，导致停顿。一般应选为均匀分布的小数，介于 (-1，1) 。

（3）学习速率调整

学习速率的选取很重要，大了可能导致系统不稳定，小了会导致训练周期过长、收敛慢，达不到要求的误差。一般倾向于选取较小的学习速率以保持系统稳定，通过观察误差下降曲线来判断。下降较快说明学习率比较合适，若有较大振荡则说明学习率偏大。同时，由于网络规模大小的不同，学习率选择应当针对其进行调整。采用变学习速率的方案，令学习速率随学习进展而逐步减少，可收到良好的效果。

（4）输入特征值的选取

如果特征数目过多，则计算复杂度过大，所以只选择对训练数据有足够分类能力的特征。文本以灰度直方图中靠近白色的部分占整体的百分比即白比，和canny边缘检测图中的边缘统计作为输入特征，还可以选取其他更有足够分析力的特征作为BP神经网络的输入值。

# 五、总结

- 本文介绍了一种基于BP神经网络的图像分类器，对图像进行特征提取和分析，对训练样本集进行训练，自动学习出一个BP神经网络系统，可以对输入的测试样本集进测试，最终输出预测空气质量AQI和AQI的分类，BP神经网络的预测精度达到了93.78%，达到评估空气质量的目的。与传统技术相比，所采用的方法的有点在于其实现成本低，计算速度快；不足在于算法主要针对特定的训练样本集合和测试样本集进行训练和测试，不够通用，如需测试其他样本，还需要更多的训练样本和选取更多有效特征进行训练。可以得出一下结论：1）采用基于BP神将网络的空气质量评估方法是可行的；2）通过对训练样本集的训练和对测试样本集的测试实验，可见本文所设计和使用的算法模型能够被用来完成空气质量评估。

# 五、总结

- 未来展望

在未来的工作中，一方面需要进一步研究更有足够分析力的图像特征参数选取，以建立更为高效和高精确度的空气质量评估模型。另一方面，有关于神经网络的研究也将被研究，应用于对上述模型的自动分析与评估。

# 谢谢