

2008年 第一届“数学中国杯” 数学建模网络挑战赛

题 目 C题 沙尘暴与各项气象指标的关系分析

关 键 词 BP神经网络、基于沙尘动力学的等级评定模型、搜索算法

摘 要：

沙尘暴是我国西北地区的灾害性天气之一,近年来沙尘暴出现的频率不断增加,影响范围不断扩大,破坏程度逐年加剧,已成为我国西北地区环境恶化的重要标志。为了使人们更好的对沙尘暴进行预防,本文分别从数学角度的数据分析方面和物理角度的动力学机理方面建立模型对沙尘暴的形成进行了研究,得到了沙尘暴与气象条件之间的关系,提出了预防沙尘暴的方法。

在问题一中,本文从统计学理论角度出发,建立基于BP神经网络的沙尘暴分析预测模型,提供了一种通过训练历史数据,得到沙尘暴发生次数与各气象指标之间的关系,从而根据相关气象数据的监测结果,对沙尘暴进行预测的方法;并通过对陕西57036号观测站1961年到2000年这40年内有记录的338组数据进行训练和预测,对该模型进行了精度分析,其精度检验结果为:

误差指标	最大偏差 $\delta \max$	二维范数距离 $W = \sqrt{\frac{\sum_i (t_i - o_i)^2}{n}}$	成功界限指数 CSI
误差数值	1.0072	0.0447	98.22%

表明该模型具有很高的精度。同时本文中还对该模型进行了灵敏度分析,对预测结果也做了合理性说明,通过对结果的分析,发现该模型存在预测保守这一问题,为了解决这一问题,本文还提出了适当进行风险估计的方法。

在问题二中,本文从沙尘暴的形成机理出发,建立了基于沙尘动力学的沙尘暴等级评定和预测模型,并提出了在不同的条件下运用该模型的方法。尤其针对在数据不足的情况下,提出了通过建立各气象指标的概率分布模型和随时间的变化规律,利用空间搜索算法来确定模型参数的方法。该模型不仅能够评定沙尘天气等级,还能预测发生沙尘暴的概率。本文给出了如何评定沙尘天气等级和预测沙尘暴概率的方法。本文还给出了风险的定义及预测方法,为人们有效预防沙尘暴提供了帮助。在该问题中,基于物理机理建模和通过建立各气象指标的概率分布模型和随时间的变化规律,利用空间搜索算法来确定模型参数是本文的最大亮点。

最后,本文还对模型进行了深入讨论和评价,提出了改进模型的想法和思路,对从事沙尘暴预测这一研究的工作人员提供了新的思路和启发。

参赛队号 1070

所选题目 C

参赛密码 _____
(由组委会填写)

目录

一、问题重述与分析.....	4
1.1问题重述.....	4
1.2问题分析.....	5
1.2.1问题背景分析.....	5
1.2.2对问题一的分析.....	5
1.2.3对问题二的分析.....	5
二、模型假设.....	6
三、符号说明.....	7
3.1模型一中符号说明.....	7
3.2模型二中符号说明.....	7
四、模型建立与求解.....	8
(一)，对问题一研究.....	8
4.1 基于BP神经网络的沙尘暴分析预测模型的建立.....	8
4.1.1 BP神经网络的基本理论.....	8
4.1.2 BP神经网络的建立.....	10
4.1.3 利用MATLAB软件对BP神经网络的学习训练和预测.....	13
4.1.4 预测结果分析.....	16
4.1.5 模型的灵敏度分析.....	16
(二)，对问题二研究.....	18
4.2 基于沙尘动力学机理的沙尘天气等级评估模型的建立.....	18
4.2.1，沙尘暴形成条件的分析.....	18
4.2.2，沙尘暴形成机理的分析.....	19
4.2.3建立沙尘暴等级评估模型.....	19
4.2.4 对沙尘暴等级评定模型参数的确定.....	22
五、模型检验与结果分析.....	24
5.1 模型一的检验.....	24
5.1.1模型的精度分析.....	24
5.1.2 模型的稳定性分析.....	25
六、模型的进一步讨论.....	26
6.1 模型一的进一步讨论.....	26
6.2 模型二的进一步讨论.....	26
七、模型优缺点.....	27
7.1 模型一的优缺点.....	27
7.1.1 模型的优点.....	27
7.1.2 模型的缺点.....	27
7.2 模型二的优缺点.....	27
7.2.1 模型二的优点.....	27
7.2.2模型二的缺点.....	27
八、参考文献.....	28

一、问题重述与分析

1.1 问题重述

沙尘暴是一种灾害性天气，近年来频繁发生，对环境和人们的生产、生活都造成了影响甚至危害。对沙尘暴的发生条件需要加以研究，以便人们进行有效的预测和预防。请为沙尘暴的发生建立数学模型，具体问题如下：

问题一：对沙尘暴的发生次数进行分析，建立相应的数学模型，使人们可以根据相关气象数据的监测结果，估计出沙尘暴在某地区某段时间内发生的次数。

问题二：评估沙尘天气气象等级，预测在近期内观测站所在地发生沙尘暴的风险，并在高风险时发出警报信息。

1.2 问题分析

1.2.1 问题背景分析

沙尘暴是我国西北方的自然现象，但近几年发生频率不断增加，影响范围不断增大，人类生存环境受到威胁，成为我国西北地区春季频繁发生的一种自然灾害性天气，对人类的危害也越来越大。对沙尘暴做出较为准确的预报对防灾减灾具有十分重要的实际意义。我国气象工作者针对沙尘暴已有很多研究工作。目前对沙尘暴的统计预报，主要方法有K-最近邻法、人工神经网络法、相似离度方法等。但大多数是侧重于对产生沙尘暴天气的天气气候学特征进行研究和诊断分析，总结归纳其大气环流特征，或者是对个别沙尘暴个例的特殊成因分析找出沙尘暴预报指标[1]。从沙尘暴的发生条件进行研究，对沙尘暴天气进行可靠的预测，以便人们对沙尘暴天气进行有效的预防，从而尽可能减少沙尘暴给人们带来的灾害，是当今研究的热点。

1.2.2 对问题一的分析

问题一要求我们对沙尘暴的发生次数进行分析，建立相应的数学模型，使人们可以根据相关气象数据的监测结果，估计出沙尘暴在某地区某段时间内发生的次数。

解读该问题要求，我们得到本题主要有以下几个要点：

1，本问题研究的对象为我国西北地区各个观测站点在过去四十多年的气象指标和沙尘暴次数的观测数据；

2，建模的关键是如何实现使人们根据相关气象数据的监测结果就能够估计出沙尘暴在某地区某段时间内发生的次数的目的。显然，需要得到沙尘暴的发生次数与相关气象指标的关系。

经过上述分析，我们得到我们的主要工作就是如何找出沙尘暴的发生次数与各种气象指标如月平均风速、大风发生日数、月平均气温、总辐射、月平均降水、月蒸发量、空气相对湿度之间的关系。附表中已经给出了过去五十多年我国西北地区各观测站点沙尘暴的发生次数与各种气象指标的数据。我们可以从数学角度出发，通过对这些数据进行统计分析，从现有的各种观测数据中找到沙尘暴的发生次数与各种气象指标之间

在数据上的关系。找出为解决这一问题，我们有很多种选择，如聚类分析法，相关分析法等。考虑到数据的随机性和非线性，且样本数较大，我们选择在处理非线性问题上具有很大优势的BP神经网络进行分析。只要我们对某一个站点过去的统计数据进行了训练，我们就可以利用训练好的网络对未来的数据进行分析判断，从而达到与预测的目的。

1.2.3对问题二的分析

问题二要求我们评估沙尘天气气象等级，预测在近期内观测站所在地发生沙尘暴的风险，并在高风险时发出警报信息。我们从以下几个方面分析题意。

1.2.3.1沙尘天气气象等级如何确定？

根据所查资料，沙尘天气等级一般分为以下四个等级：

(1)、浮尘：尘土、细沙、均匀地浮游在空中，使水平能见度小于10公里的天气现象；

(2)、扬沙：风将地面尘沙吹起，使空气相当混浊，水平能见度在1~10公里以内的天气现象；

(3)、沙尘暴：强风将地面大量尘沙吹起，使空气很混浊，水平能见度小于1公里的天气现象；

(4)、强沙尘暴：大风将地面尘沙吹起，使空气非常混浊，水平能见度小于500米的天气现象[2]。

可见，沙尘天气等级是以能见度为标准划分的。于是在该问题中，我们从能见度入手分析。考虑到能见度主要由沙尘的浓度决定，而沙尘的浓度又与各气象指标密切相关，于是我们可以从各气象指标引起沙尘天气的物理机理进行建模，从而得到沙尘天气等级与各气象指标的联系，这样我们就可以通过气象指标的数据对沙尘天气等级进行评估。

1.2.3.2发生沙尘暴的风险如何定义？

经过分析，我们认为发生沙尘暴的风险由沙尘天气的等级和沙尘暴的发生概率这两个因素综合决定。沙尘天气的等级越强，发生沙尘暴的概率越大，则风险越高。

1.2.3.3如何预测发生沙尘暴的风险？

首先，我们可以预测出沙尘天气的等级和发生沙尘暴的相对概率，然后将沙尘天气的等级和发生沙尘暴的相对概率都在一定规则下进行量化，于是可以得到风险的一个量化值，根据该值大小，用气象局经常用以评定灾害天气强度的“蓝，黄，橙，红”四个等级层次评定风险等级。

报名号#1070

二，模型假设

- 1, 假设沙尘暴各影响因素间相互独立;
- 2, 假设风速, 湿度, 辐射量、降雨量和蒸发量服从正态分布;
- 3, 假设沙粒为球状;
- 4, 假设不考虑沙粒与周围沙粒间的热传递;
- 5, 假设各气象指标的值在一个小时内不发生变化。

三，符号说明

3.1模型一中符号说明

t_i	-----	表示神经网络第i个样本的期望输出值
o_i	-----	表示神经网络第i个样本的计算输出值
W	-----	二维范数距离
cf	-----	正确报出沙尘暴的次数
wf	-----	试漏报和错报次数
CSI	-----	成功界限指数

3.2模型二中符号说明

f	-----	沙尘暴等级
s	-----	能见度
β_{ext}	-----	为消光系数
β_{sg}	-----	气体分子瑞利散射系数
β_{sp}	-----	颗粒物的散射系数
β_{ap}	-----	颗粒物的吸收系数
β_{ag}	-----	气体的吸收系数
N_i	-----	i粒径段颗粒物的浓度数
r_i	-----	i粒径段颗粒物的对数中位径
$Q_{ext,i}$	-----	颗粒物的消光效率
Ri	-----	梯度Richardson数,用来表征大气的稳定度的一个参数

报名号#1070

V_a	-----	高度 h_a 处的风速
V_b	-----	对流速度
θ_a	-----	高度 h_a 处的位温
θ_g	-----	地表的位温
θ_{va}	-----	高度 h_a 处的虚位温
θ_{vg}	-----	地表的虚位温
G	-----	表示地表沙粒的热导率
$\Delta\theta$	-----	表示沙粒间的传导温差
C	-----	大地比热容
q_g	-----	为地表湿度
l_{down}	-----	月为降雨量
l_{up}	-----	月蒸发量
k_q	-----	地表湿度系数
k_j	-----	第j种地貌对消光系数的影响系数为
θ_0	-----	大地温度
l_0	-----	地表原有水量

四、模型建立与求解

(一)，对问题一研究

首先，从数学角度出发，从现有的各种观测数据中找到沙尘暴的发生次数与各种气象指标在数据上的关系。鉴于各种气象指标和沙尘暴发生次数的观测数据具有很强的随机性和非线性，考虑到BP神经网络具有较强的学习能力和非线性特性，使其能够对非线性函数进行很好的逼近。于是我们建立基于BP神经网络的沙尘暴分析预测模型对数据进行分析，得到沙尘暴的发生次数与各种气象指标在数据相互关联的神经网络，从而可以根据已经训练好的神经网络，在得到各种气象指标后，对某地某时间内的沙尘暴次数进行估计。

4.1 基于BP神经网络的沙尘暴分析预测模型的建立

4.1.1 BP神经网络的基本理论

4.1.1.1 BP神经网络的结构

BP神经网络是一个具有三层或三层以上的阶层神经网络，上下层之间各种神经

报名号#1070

元实行权连接,而各层之间神经元无连接。最基本的BP网络是三层前馈网络,即输入层,隐含层,输出层。各层之间实行全连接,典型的三层前馈型网络拓扑结构形式如图 1

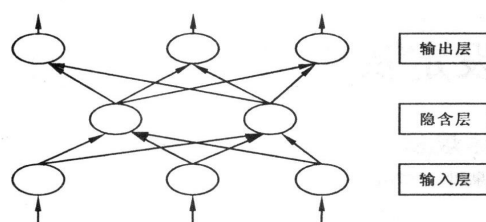


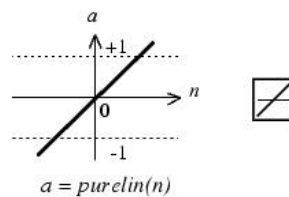
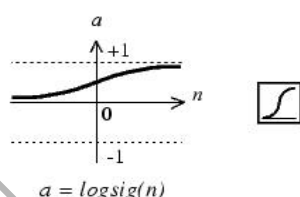
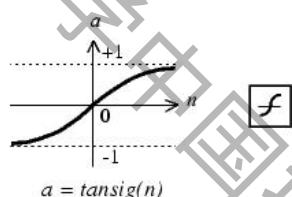
图 1

层与层之间的传递函数一般为值域是(0,1)函数,常用传递函数有以下几种:

$$\text{tansig} = \frac{1 - e^{-2x}}{1 + e^{-2x}}$$

$$\text{logsig} = \frac{1}{1 + e^{-x}}$$

purelin () 函数



4.1.1.2 BP 神经网络的学习规则

- 1) 模式顺传播过程(神经元的激活值从输入层经中间层向输出层)；
- 2) 误差逆传播过程(网络的希望输出与实际计算输出之差的误差信号由输出层经中间层到输入层逐层修正连接权)；
- 3) 记忆训练过程(在模式顺传播与误差逆传播过程中反复交替进行)；
- 4) 学习收敛过程(网络趋向收敛也即网络的全局误差趋向极小值)。

4.1.1.3, BP神经网络学习公式推导

BP 神经网络的学习公式推导的思想是对各层之间的权值与阈值进行修正,使得误差函数E沿梯度方向下降。误差的小与大能够反应计算准确度的高与低,BP 神经网络有自己的误差计算公式。对于第k 个样本,希望输出与计算输出的偏差是:

$$\delta_i = t_{ki} - o_{ki} \quad \backslash * \text{MERGEFORMAT} <1>$$

网络的误差计算公式是:

$$E_k = \frac{\sum_i (t_{ki} - o_{ki})^2}{2} \quad \backslash * \text{MERGEFORMAT} <2>$$

其中 o_{ki} 、 t_{ki} 分别是网络的计算输出与期望输出, i 表示样本中的第 i 个输入量。

假设BP神经网络只有一个隐含层,输入层为 x_i , 隐含层为 c_j , 输出层为 y_k 输出层的期望输出为 p_k , 输入层与隐含层间、隐含层与输出层间的网络权值与阈值分别为 w_{ji} 、

q_{kj} 与 b_j 、 b_k ，则BP网络模型的计算公式为：

1) 隐含层的输出：

$$C_j = f\left(\sum_i \omega_{ji}x_i - b_j\right) = f(D_j) \quad \backslash * \text{MERGEFORMAT} <3>$$

其中 $D_j = \sum_i \omega_{ji}x_i - b_j$ 。

2) 输出层的计算输出：

$$y_k = f\left(\sum_j q_{kj}c_j - b_k\right) = f(D_k) \quad \backslash * \text{MERGEFORMAT} <4>$$

其中 $D_k = \sum_j q_{kj}c_j - b_k$ 。

3) 网络权值与阈值修正：输出层的误差公式为：

$$\begin{aligned} E &= \frac{1}{2} \sum_j (p_k - y_k)^2 = \frac{1}{2} \sum_j \left(p_k - f\left(\sum_j q_{kj}c_j - b_k\right) \right)^2 \\ &= \frac{1}{2} \sum_j \left(p_k - f\left(\sum_j q_{kj}f\left(\sum_i \omega_{ji}x_i - b_j\right) - b_k\right) \right)^2 \end{aligned} \quad \backslash * \text{MERGEFORMAT} <5>$$

而

$$\frac{\partial E}{\partial q_{kj}} = \sum_m \frac{\partial E}{\partial y_m} \frac{\partial y_m}{\partial q_{kj}} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial q_{kj}} \quad \backslash * \text{MERGEFORMAT} <6>$$

其中，

$$\frac{\partial E}{\partial y_k} = \frac{1}{2} \sum_m [(-2)(p_m - y_k)] \frac{\partial y_m}{\partial y_k} = -(p_m - y_k) \frac{\partial y_k}{\partial q_{kj}} = f'(D_k)c_j \quad \backslash * \text{MERGEFORMAT} <7>$$

令 $\delta_k = (p_m - y_k)f'(D_k)$ ，则有

$$\frac{\partial E}{\partial q_{kj}} = \eta \delta_k c_j \quad \backslash * \text{MERGEFORMAT} <8>$$

由于权值的修正 Δq_{kj} 正比于误差函数沿梯度下降，则有

$$\Delta q_{kj} = \frac{\partial E}{\partial q_{kj}} = \eta \delta_k c_j \quad \backslash * \text{MERGEFORMAT} <9>$$

由此得到权值的修正公式为：

$$q_{kj}(m+1) = q_{kj}(m) + \Delta q_{kj} = q_{kj}(m) + \eta \delta_k c_j \quad \backslash * \text{MERGEFORMAT} <10>$$

同理可以得到阈值修正公式为：

$$\theta_k(m+1) = \theta_k(m) + \Delta \theta_k = \theta_k(m) + \eta \delta_k \quad \backslash * \text{MERGEFORMAT} <11>$$

对于隐含层的权值与阈值修正公式与输出层的公式类同，下标做相应的变换即可[3]。

4.1.2 BP神经网络的建立

我们以对陕西省57036号观测站（西安）这一站点的数据进行训练为例，建立基于BP神经网络的分析和预测模型，提出如何利用神经网络对样本进行学习以及训练结束后如何进行预测的方法，利用此方法就可以对所有观测站点的数据进行学习，从而可以对所有观测站点进行预测。建立BP神经网络的过程如下：

4.1.1.2 输入输出数据预处理

由于BP神经网络训练样本集中输入、输出（目标）样本参数的绝对值离散性有时太大或过于集中，在网络权值矩阵误差函数逼近过程中，易产生局部误差最优或误差震荡等缺陷（如右图 2误差与权值矩阵函数关系图示）。必须对输入、输出进行归一化处理，防止在网络权值矩阵误差函数训练时不会进入局部误差最优或误差震荡等缺陷区域。此外，归一化处理可以使得具有不同物理意义和量纲的输入变量赋以同等重要的地位，避免数值大的变量掩盖数值小的变量，避免神经元饱和，保证网络的收敛性，提高网络的收敛速度。

归一化处理：我们可采用下列公式将数据压缩在（0，1）的范围内，

$$X_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad \backslash * \text{MERGEFORMAT} <12>$$

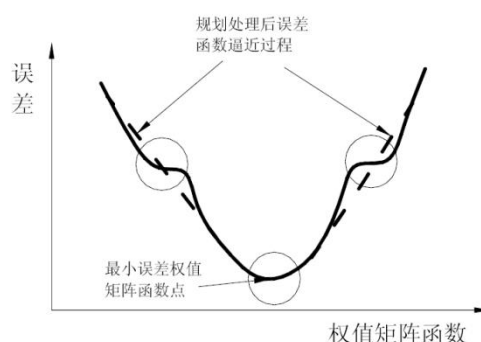


图 2误差与权值矩阵函数关系图

4.1.2.1 BP神经网络层数的选取

具有偏差和至少一个Sigmoid型隐含层加上一个线性输出层的网络，能够逼近任何有理函数。增加层数主要可以更进一步的降低误差，提高精度，但是同时也会增加网络的复杂性，从而增加网络权值的训练时间。增加隐含层中的神经元节点数目也可以提高误差精度，其训练效果也比增加层数更容易观察和调整，所以通常状况下，应该优先考虑增加隐含层中的神经元节点的数目。我们选用三层BP网络进行分析预测。

4.1.2.2 网络输入层和输出层的设计

输入层起缓冲器的作用，把数据源加到网络上，其节点数目取决于数据源的维数。基于BP算法的神经元网络输出层神经元可以根据要求解的问题和数据表示的方式而定。在设计输入层和输出层时，应该尽可能的减小系统规模，使系统的学习时间和复杂性减小。

由于沙尘暴预测的气象因素指标为7维的输入向量，输入层一共设7个神经元。输出量为沙尘暴月发生次数，输出层有1个神经元。

4.1.2.3 BP神经网络隐含层节点的选择

基于BP算法的神经元网络中各层节点数目的选择对于网络性能的影响很大,过多的隐含节点对网络的概括能力产生不利影响,既影响网络对于新输入的适应性。而过少的隐含层节点数目会影响网络学习的精度,并且是网络学习出现局部极小的情况增多,所以节点的数目要选择恰当。常用的解决办法就是使隐含层数目可变。一种是开始放入足够的隐含单元,然后把学习后的那些不起作用的隐含层单元逐步去掉,一直减少到不可收缩为止。另一种是开始放入比较少的隐含层单元,学习一些次数后,还不成功就要增加隐含单元个数,一直达到比较合理的隐含单元数目为止。

实际操作中,根据经验,可以参照以下公式进行设计:

$$n = \sqrt{n_i + n_o} + a \quad \backslash * \text{ MERGEFORMAT } <13>$$

式中: n 为隐层节点数, n_i 为输入节点数, n_o 为输出节点数, a 为1~10 之间的常数。

在本题中,我们对西安观测站进行分析预测,选取数据满足 $n_i + n_o = 338$, 取 $a=10$, 于是我们最终选取 $n=28$ 。

4.1.2.4 训练样本的采集

要建立一个沙尘暴预测模型,就需要采集若干组的实验观测样本。可将已知的数据分为两组,其中,将气象指标的数据用来构成训练样本,沙尘暴的次数作为检验样本。样本由输入输出数据构成。设网络的输入节点数是 n ,输出节点数是1,则 $n+1$ 个数据组成一个样本。其前 n 个值作为输入数据,后1个值是期望输出值。在本题中我们对陕西省57036号观测站(西安)进行分析,考虑到附表中所给数据存在一些缺失(1968、1969、1971年等数据缺失)和错误,我们选取57036号观测站1961年到2000年这40年内有记录的338组数据作为训练的样本集(见附件中附表一)。于是,我们得到每次训练的输入节点数是7,输出节点数是1。

4.1.2.4 BP神经网络学习速率的选择

学习速率决定每一次循环训练所产生的权值的变化量。过大的学习速率可能导致系统的不稳定,但是过小的学习速率将导致训练时间较长,收敛速度很慢,不过能保证网络的误差值不跳出误差表面的低谷而最终趋于最小的误差值,所以在一般情况下,倾向于选取较小的学习速率以保证系统的稳定性,学习速率的选取范围一般在0.01~0.9之间。我们选择网络学习速率为0.05。

4.1.2.5 传输函数的选定。

前面已经介绍了BP网络结构常用的几种传递函数，考虑到数据样本的随机性和非线性性，我们采用对非线性问题解决较好的S型正切激活函数 $\text{tansig} = \frac{1-e^{-2x}}{1+e^{-2x}}$

作为隐含层的传递函数 $\text{tansig} = \frac{1-e^{-2x}}{1+e^{-2x}}$ ，基于相同的考虑，输出层的传递函数我

们采用S型对数激活函数 $\text{logsig} = \frac{1}{1+e^{-x}}$ 。

4.1.2.6 训练方法的选取。

采用附加动量法使反向传播减少了网络在误差表面陷入低谷的可能性有助于减少训练时间。学习速率太大导致学习的不稳定，太小又导致极长的训练时间。自适应学习速率通过保证稳定训练的前提下，达到了合理的高速率，可以减少训练时间。在网络训练中，可以采用 `traingdx` 梯度下降w / 动量和自适应 `lr`的 BP 算法训练函数；`traingda` 梯度下降w / 自适应 `lr`的 BP 算法训练函数；`trainlm` Levenberg-Marquardt的BP算法训练函数；`traingdm` 梯度下降w / 动量的BP算法训练函数。

在本题中，我们采用 `traingdx`训练方法。

4.1.2.7 期望误差的选取。

一般情况下，作为对比，可以同时两个不同的期望误差值的网络进行训练，最后通过综合因素的考虑来确定其中一个网络。期望误差的控制，通过对网络训练参数的设定实现。考虑到具体网络训练时的训练精度和训练时间，同时确保能够对系统进行较好的预测，保证预测的实际精度，我们在网络中使`net.trainParam.epochs = 10000`，`net.trainParam.goal = 0.002`，即确定训练次数极值为10000次，网络训练的误差允许值为0.002。

综上所述：我们在本题中以西安为例利用BP神经网络对沙尘暴的次数进行分析预测。其BP网络拓扑结构为 $7 \times 28 \times 1$ 的结构。中间层神经元的传递函数为S型正切函数`tansig`。由于输出已被归一化到区间 $[0, 1]$ 中，输出层神经元的传递函数可以设定为S型对数函数`logsig`。网络的训练函数采用学习率可变的动量BP算法修正神经网络的权值和阈值函数`traingdx`。保证神经网络的期望误差达到0.002，能够很好的进行实际预测。

4.1.3 利用MATLAB软件对BP神经网络的学习训练和预测

MATLAB的BP网络工具箱中包含了进行BP网络分析和设计的许多函数，为BP神经网络的应用研究提供了强有力的工具。本文用到的主要工具箱函数有：

4.1.3.1 建立神经网络的函数

newff() ,用于建立一个BP神经网络;

```
net= newff ( minmax(pp), [28,1], {'tansig','logsig'}, 'traingdx');
```

建立一个 $7 \times 28 \times 1$ 的神经网络。中间层神经元的传递函数为S型正切函数 **tansig**。输出层神经元的传递函数可以设定为S型对数函数 **logsig**。网络的训练函数采用学习率可变的动量BP 算法修正神经网络的权值和阈值函数 **traingdx**。

4.1.3.2 神经单元的传递函数

tansig() 函数, **purelin()** 函数和 **logsig()** 函数;

4.1.3.3 网络训练参数的设定

```
net.trainParam.epochs = 20000; 训练次数
net.trainParam.goal = 0.002;    期望误差
net.trainParam.lr= 0.05;        学习率
net.trainParam.show = 50;       输出控制
```

4.1.3.4 神经网络的训练函数

train() , 用于训练建立的BP神经网络;

```
net=train(net, pp, tt) 用已有数据进行网络训练。
```

4.1.3.5 神经网络的仿真函数

sim() 用于仿真训练前后的BP神经网络。

A=sim(net,pp) 输出网络模型的预测结果, 进行有效的检验; 通过改变系统影响参数 **pp**, 对系统进行精确预测。

通过对训练好神经网络模型参数的获取, 得到可以随时进行系统预测的神经网络。

具体参数为: $7 \times 28 \times 1$ 的三层 BP 神经网络, 其中输入层与隐含层的传递函数为 S 型正切函数 **tansig()**, 隐含层与输出层的传递函数为 S 型对数函数 **logsig()**。各层间的权重和阈值分别为 $IW\{1,1\}$, $b\{1\}$, $IW\{2,1\}$, $b\{2\}$ 。

当应用网络是只需按照获得的网络参数进行初始化 **net=init(net)**, 便可得到需要的预测网络, 进而通过仿真函数 **sim(net,x)**, 对输入参数为 **x** 的系统进行预测。

具体的网络模型可以表示为:

隐含层的输出:

$$C_j = f\left(\sum_i IW\{1,1\}x_i - b\{1\}\right) = \tan sig\left(\sum_i IW\{1,1\}x_i - b\{1\}\right) \quad \text{MERGEFORMAT} \quad (14)$$

输出层的输出:

报名号#1070

$$y_k = f\left(\sum_j LW\{2,1\}c_j - b\{2\}\right) = \log \operatorname{sig}\left(\sum_j LW\{2,1\}c_j - b\{2\}\right) \quad \text{\textbackslash* MERGEFORMAT <15>}$$

用图 3直观表示如下：

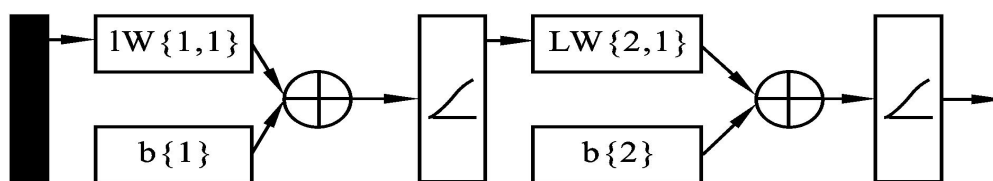


图 3

4.1.3.6 利用BP神经网络进行训练

对陕西省57036号观测站（西安）的数据利用BP神经网络模型进行训练，训练过程可从下图中看出：

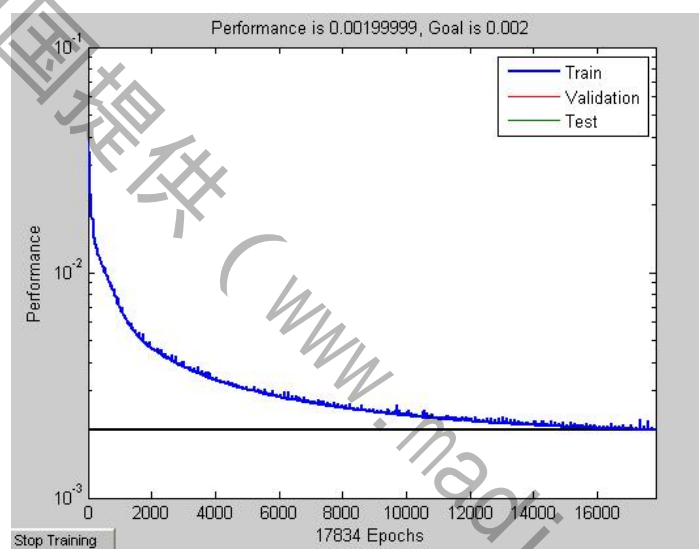


图 4

4.1.3.7 对训练后网络进行测试

我们利用已经训练好的神经网络对训练的数据进行预测，得到预测结果如下（限于篇幅，这里只列出1961年到1999年部分数据，其余数据见附表）：

年	月	沙尘暴天数	预测结果
1961	3	1	1
1961	5	2	2
1961	6	1	1
1961	7	4	3
1961	8	1	1

报名号#1070

1962	5	3	3
1962	6	1	1
1963	1	2	2
1963	7	4	4
1964	2	2	2
1964	5	1	1
1964	6	2	2
1964	7	1	1
1965	4	1	0
1965	6	2	2
1965	10	1	0
1966	9	1	0
1978	1	1	1
1978	3	1	1
1979	3	3	3
1980	3	2	2
1980	4	1	0
1980	5	1	0
1981	5	2	2
1983	3	1	1
1984	4	2	2

4.1.4 预测结果分析

上表（详见附表）列出了陕西省57036号观测站1961年到2000年每个月的沙尘暴发生次数预测情况，可以得出在338次中出现5次预测错误、1次预测次数不准确，错误率

$$p_{\text{wrong}} = \frac{n_{\text{wrong}}}{n_{\text{whole}}} = \frac{6}{338} = 1.7\% \quad \backslash * \text{MERGEFORMAT} <16>$$

说明BP神经网络对沙尘暴次数的预测具有很高的准确率。

同时，对预测结果分析我们还能得出以下结论：

1，从所有预测结果看出：该网络对沙尘暴次数为0的情况，预测很准；对沙尘暴次数为1的情况，会误判为不会发生；对沙尘暴次数为多次的情况，会有很小几率误认为是少于实际次数发生。说明该网络模型会保守预测沙尘暴发生情况。对发生几率很小的情况，不会予以告警。针对这种情况，在实际预测过程中，我们对归一化数据还原时，可以基于风险估计的考虑，适当增大预测值（例如将四舍五入的原则变为三舍四入原则，如若计算结果为发生1.41次，则认为发生了两次），这样便能够在一定程度上提高预测的准确性。另一方面，即使对沙尘天气的发生次数进行了高于实际次数的风险估计，也符合防患于未然的大众心理。

2，预测的数据中许多气象指标值均是一个月的平均值，这会在一定程度上掩盖峰

报名号#1070

值。即平均值不满足发生沙尘暴的条件，而实际在某天会达到较高的沙尘暴发生风险。这样就使得预测值与实际值存在一定的偏差。

4.1.5 模型的灵敏度分析

为了分析模型的灵敏度，我们选取任意几个月进行分析，将归一化后的各气象指标从0~0.9变动，分析预测结论的变化。

选择观测站中，两月份的参数值：

月平均风速	大风天数	月平均气温	总辐射
0.296296	0.466667	0.73871	0.904096
0.592593	0.6	0.325806	0.451276
月平均降水	月蒸发量	相对湿度	沙尘暴天数
0.158958	0.733177	0.27451	2
0.171531	0.32824	0.392157	1

将各影响因素从0~0.9变动，保持其它因素不变，进行分析，得到预测结果列表如下：

气象指标	参考值	影响因素变动时预测值的变动情况				
		0	0.1	0.2	0.3	0.4
月平均风速	2	0	0	1	2	2
	1	0	0	0	0	0
大风天数	2	1	1	1	1	2
	1	0	0	0	0	0
月平均气温	2	4	4	4	4	4
	1	3	3	2	1	0
总辐射	2	0	0	0	0	0
	1	1	1	1	1	1
月平均降水	2	0	0	3	3	1
	1	0	0	1	4	4
月蒸发量	2	0	0	0	0	0
	1	2	1	1	1	1
相对湿度	2	0	1	1	2	2
	1	0	0	0	0	1
气象指标	参考值	影响因素变动时预测值的变动情况				
		0.5	0.6	0.7	0.8	0.9
月平均风速	2	2	2	2	3	3
	1	0	1	1	2	2
大风天数	2	2	1	0	0	0
	1	1	1	0	0	0
月平均气温	2	4	4	3	0	0

报名号#1070

	1	0	0	0	0	0
总辐射	2	0	0	0	1	2
	1	1	1	1	3	4
月平均降水	2	0	0	0	0	0
	1	4	4	4	4	4
月蒸发量	2	0	1	2	2	1
	1	1	1	2	3	4
相对湿度	2	1	0	0	0	0
	1	0	0	0	0	0

通过分析表格数据，我们可以分析灵敏度如下：

A，在灵敏度测试列表中，参考值的位置表示在表格中。

B，通过分析可以得到：各影响因素对沙尘暴产生的总体影响情况，月平均风速越高，沙尘暴产生的可能性越大；同样，大风天数、总辐射、月蒸发量数值增大，沙尘暴产生的可能性提高，月平均气温、相对湿度数值减小，使得沙尘暴产生可能增大。

这与实际是相符的，影响沙尘暴的因素主要是大风、沙尘源、不稳定的大气等：风速越高，大风次数越多，越易于发生沙尘暴；辐射强度越大则地表温度越高，加之气温越低，则大气情况越不稳定；其余指标则是通过影响沙源的力学状态而作用的。

C，同时，各影响因素与沙尘暴产生可能性之间没有确切的关系，而是在参考值的一定波动范围内具有一定的预测结果。具体说来，测试中发现预测结果并不是随相对湿度的减小，沙尘暴产生的可能性增大（从表格中可以看出），相反，当相对湿度减小到最小值时，反而没有发生沙尘暴。

该结果的出现与神经网络学习训练预测的数据分析方法本身有密切的联系。神经网络通过学习训练实现数据的聚类，注重将样本进行学习训练，以实现高精度拟合。也就是说，当样本中不存在一些极端数据情况时，这样的情况被训练为不会发生沙尘暴。从而，得到了只有在参考值附近变动才会出现符合实际的预测结果的结论。这是神经网络方法本身仅仅考虑数据间的数学关系导致的情况。

以上情况的发生，虽然不能很好的表征实际意义，但是在实际分析和预测中，这些极端数据是不会出现的，通过对大批量样本数据的高精度学习训练神经网络可以实现对预测的全局最优化。正如，在上面结果分析中发现，预测结论很贴合实际，并且，期望误差在训练中能够控制在一个很小的数量级（本网络期望误差为0.002）。

（二），对问题二研究

鉴于单纯的从数据分析的角度评定沙尘暴的等级和预测沙尘暴，只能用数据规律反应出沙尘暴发生的规律，而无法反应沙尘暴产生的机理，其分析预测结果很可能与实际偏差较大，或出现实际情况无法解释的现象。于是我们从物理的角度分析沙尘暴产生的机理，建立基于沙尘动力学机理的沙尘天气等级评估和预测模型。

4.2 基于沙尘动力学机理的沙尘天气等级评估模型的建立

要从物理机理角度分析沙尘暴的形成过程，首先要确定沙尘暴的形成必须具备的条件。

4.2.1, 沙尘暴形成条件的分析

有关研究指出沙尘暴的形成必须具备四个条件：一是地面上的沙尘源，它是形成沙尘暴的物质基础；二是大风，这是沙尘暴形成的动力基础，也是沙尘暴能够长距离输送的动力保证；三是不稳定的空气状态，这是重要的局地热力条件，沙尘暴多发生于午后傍晚说明了局地热力条件的重要性；四是干旱的气候环境，沙尘暴多发生于北方的春季，而且降雨后一段时间内不会发生沙尘暴是很好的证据[4]。对上述条件进行剖析，我们得出：

1, 沙尘源是个地理条件问题，相对来说比较固定，对同一地点来说，沙尘源相当于一个定值。

2, 气候的干旱程度也是个地理条件问题，对同一地点来说，气候的干旱程度也是固定的。

3, 大风也是描述空气的稳定性的一个指标，故可以将大风这一条件归为空气稳定状态中去。

4, 虽然资料中并没有提及地理位置对沙尘暴的影响，但根据生活常识，我们知道不同的地理位置产生沙尘暴的条件也是不一样的，故地理位置也是影响沙尘暴形成的一个条件。

于是，经过上述分析，我们可以得到，为了提高精度、简化模型，我们认为以观测站点为对象建立评估预测模型是最好的选择。这样既能简化对沙尘源和气候干旱程度以及地理位置等这些指标的分析，还能提高评估预测的精度。

这样我们就只需对空气的稳定性进行分析了。很显然，各种气象指标如月平均风速、大风发生日数、月平均气温、总辐射、月平均降水、月蒸发量、相对湿度等均是影响空气稳定性的因素。下面，我们就需要进一步研究这些气象指标是怎样通过空气稳定性影响沙尘暴的形成过程的，即研究它们在物理机理上的联系。

4.2.2, 沙尘暴形成机理的分析

通过了解，沙尘暴是冷空气在移动过程中出现的一种强烈锋生过程，并且沙尘暴产生的大气边界层是极其不稳定的（包括静力和动力不稳定），涡流发展旺盛。根据产生沙尘暴的天气学动力学概念，我们可以用起沙和扬沙两种机制来概括其产生。

起沙，说明边界层内必有较强的风速垂直切变存在（动力不稳定），扬沙说明边界层必然存在浮力上升运动（静力不稳定），可见，空气的稳定度能直接反应产生沙尘暴的可能性和强度。于是我们得到这样的沙尘暴形成机理：由某些气象条件下，大气极不稳定，于是产生沙尘暴。气象条件越恶劣，导致大气越不稳定，从而产生沙尘暴的可能性和强度就越大。所以沙尘暴的等级与气象条件就可以通过大气的稳定性联系起来。

4.2.3建立沙尘暴等级评估模型

4.2.3.1, 对能见度的分析

在该问题的分析中我们已经讨论得到沙尘暴的等级评定是以能见度为依据的，能见度越大，沙尘暴的等级越小。所以，我们可以建立沙尘暴等级与能见度的关系式为

$$f = \begin{cases} red, s \leq 500m \\ orange, 500m < s \leq 1000m \\ yellow, 1000m < s \leq 10000m \\ blue, s > 10000m \end{cases} \quad \backslash * \text{ MERGEFORMAT } \langle 17 \rangle$$

其中， f 为沙尘暴等级， s 为能见度。

根据气象知识，一般能见度定义如下：

$$s = 3.912 / \beta_{ext} \quad \backslash * \text{ MERGEFORMAT } \langle 18 \rangle$$

式中， β_{ext} 为消光系数。它由四部分组成：

$$\beta_{ext} = \beta_{sg} + \beta_{sp} + \beta_{ap} + \beta_{ag} \quad \backslash * \text{ MERGEFORMAT } \langle 19 \rangle$$

其中， β_{sp} 对总消光系数的贡献最大，故只考虑 β_{sp} 。其中， β_{sg} 为气体分子瑞利散射系数； β_{sp} 为颗粒物的散射系数； β_{ap} 为颗粒物的吸收系数； β_{ag} 为气体的吸收系数；它们的单位皆为 km^{-1} 。考虑到在沙尘暴天气中，气体的散射系数和吸收系数相对于颗粒物来说可以忽略，于是，消光系数主要由颗粒物影响。于是可得到消光系数的计算公式为：

$$\beta_{ext} = \sum_{i=1}^m \beta_{ext,i} = \sum_{i=1}^m N_i Q_{ext,i} \pi r_i^2 \quad \backslash * \text{ MERGEFORMAT } \langle 20 \rangle$$

其中， N_i 表示 i 粒径段颗粒物的浓度数； r_i 表示 i 粒径段颗粒物的对数中位径； $Q_{ext,i}$ 表示颗粒物的消光效率[5]。

4.2.3.2, 对空气稳定度的分析

首先，我们用物理学中常用来判定一个行星边界层稳定性大小的一个重要参数-----梯度Richardson数（简称为 Ri ）来表征大气的稳定度。

Ri 定义如下：

$$Ri = \frac{\frac{g}{\theta_v} \frac{\partial \theta_v}{\partial h}}{\left(\frac{\partial u}{\partial h}\right)^2 + \left(\frac{\partial v}{\partial h}\right)^2} \quad \backslash * \text{ MERGEFORMAT } \langle 21 \rangle$$

式中， $\theta_v = (1 + 0.608q)\theta$ 为虚位温，为方便计算，我们用 $\frac{\Delta \theta_v}{\Delta h}$ 近似代替 $\frac{\partial \theta_v}{\partial h}$ ，用 $\frac{\Delta u}{\Delta h}$ 近似代替 $\frac{\partial u}{\partial h}$ ，那么就能确定一个新的总体Richardson数 Ri_b 为：

$$Ri_b = \frac{g \Delta \theta_v \Delta h}{\theta_v [(\Delta u)^2 + (\Delta v)^2]} \quad \backslash * \text{ MERGEFORMAT } \langle 22 \rangle$$

$$Ri_b = \frac{g h_a (\theta_{va} - \theta_{vg})}{\theta_a V^2} \quad \backslash * \text{ MERGEFORMAT } \langle 23 \rangle$$

其中， $V = (V_a^2 + V_b^2)^{1/2}$ ， V_a 是高度 h 上的风速， V_b 是一个对流速度，并由下式确定

报名号#1070

$$V_b = 2(\theta_g - \theta_a)^{1/2} \quad \backslash * \text{MERGEFORMAT} \quad \langle 24 \rangle$$

θ_a 和 θ_g 分别为 h 和地面上的位温, θ_{va} 和 θ_{vg} 分别为 h 和地面上的虚位温[6]。

4.2.3.3, 建立消光系数与大气稳定度的关系

根据式<23>可以看出, $Ri_b < 0$ 表示静力和动力不稳定, 这种情况肯定有湍流发生, 于是有利于起沙和扬沙, 而且 Ri_b 负得越多, 起沙和扬沙的可能性与强度就越大, 空气中的沙粒浓度就越大。即 $-Ri_b \propto N_i$, 又由 $\sum_{i=1}^m N_i Q_{ext,i} \pi r_i^2 \propto N_i$ 得到

$$\beta_{ext} \propto -Ri_b \quad \backslash * \text{MERGEFORMAT} \quad \langle 25 \rangle$$

因此, 消光系数与 Ri_b 的大小成负的线性关系, 即有:

$$\beta_{ext} = -p Ri_b \quad \backslash * \text{MERGEFORMAT} \quad \langle 26 \rangle$$

4.2.3.4, 建立地表温度与辐射的关系

由能量守恒, 我们得到地表温度 θ 与辐射效率 φ 关系的确定:

$$G \Delta \theta + C \frac{d\theta}{dt} = a\varphi \quad \backslash * \text{MERGEFORMAT} \quad \langle 27 \rangle$$

其中, G 表示地表沙粒的热导率, $\Delta \theta$ 表示沙粒间的传导温差, C 为大地比热容[7]。

在假设中我们已经讨论得到在同样的环境中, 沙粒与周围沙粒的温度相同, 故沙粒相互之间传热 $G \Delta \theta = 0$ 。于是, 我们可以得到地表温度

$$C \frac{d\theta}{dt} = a\varphi \Rightarrow d\theta = \frac{a}{C} \varphi dt \Rightarrow \int d\theta = \frac{a}{C} \int \varphi dt \Rightarrow \theta_g = k_\theta \Phi + \theta_0 \quad \backslash * \text{MERGEFORMAT} \quad \langle 28 \rangle$$

其中, Φ 为月总辐射热量, θ_0 为大地温度, θ_g 为地表温度, k_θ 为地表温度系数, θ_0 为大地温度。

4.2.3.5, 建立地表湿度与降雨量和蒸发量的关系

考虑到地表原有水量+降雨量-蒸发量的值能反应地表湿度的大小。例如, 该值越大表明地表湿度越大。于是, 我们建立地表湿度与降雨量和蒸发量之间的联系如下:

$$q_g = (l_{down} - l_{up} + l_0) \times k_q \quad \backslash * \text{MERGEFORMAT} \quad \langle 29 \rangle$$

其中, q_g 为地表湿度, l_{down} 为月降雨量, l_{up} 为月蒸发量, k_q 为地表湿度系数, l_0 为地表原有水量。

4.2.3.6, 对沙尘源因子的分析

考虑到不同的地貌地表的沙尘量不一, 将对沙尘浓度大小产生不同的影响, 我们根据含沙量大小将地貌分为不同类型如下: 1 戈壁地、2 沙漠地、3 荒漠地、4 耕地。设每j种地貌对消光系数的影响系数为 k_j ($j=1,2,3,4$), 则可得到:

$$\beta_{ext} = -p \times k_j \times Ri_b \quad \backslash * \text{MERGEFORMAT} <30>$$

4.2.3.7 建立沙尘暴等级评估模型

综合上述分析, 我们得到沙尘暴等级评估模型为:

$$s = \frac{3.912}{\beta_{ext}} = \frac{3.912}{(-p \times k_j \times Ri_b)} = \frac{3.912\theta_a[V_a^2 + 4(\theta_g - \theta_a)]}{gh_a[\theta_a - \theta_g + 0.608(q_a\theta_a - q_g\theta_g)]k} \quad \backslash * \text{MERGEFORMAT} <31>$$

其中, 系数 $k = -k_j \times p$ 为常数, 地表湿度 $q_g = (l_{down} - l_{up} + l_0) \times k_q$, 地表温度 $\theta_g = k_\theta \Phi + \theta_0$ 。将其带入上式中得到:

$$s = \frac{3.912\theta_a(V_a^2 + 4k_\theta\Phi + 4\theta_0 - 4\theta_a)}{kgh_a\left\{\theta_a - k_\theta\Phi - \theta_0 + 0.608k\left[q_a\theta_a - k_qk_\theta\Phi(l_{down} - l_{up} - l_0) - k_q\theta_0(l_{down} - l_{up} - l_0)\right]\right\}} \quad \backslash * \text{MERGEFORMAT} <32>$$

4.2.4 对沙尘暴等级评定模型参数的确定

经过分析, 该模型中的参数分以下三类:

- (1) 待定系数 k_θ 、 k_q 、 k ;
- (2) 未知气象指标数据 Φ 、 l_{down} 、 l_{up} 、 q_a 、 V_a : 由于我们建立的机理模型中气象指标均为瞬时值, 而我们得到的观测数据为平均值, 故所有气象指标的瞬时数据未知;
- (3) 已知常数: 重力加速度 g (一般取 $g = 9.8m/s^2$), 观测站测量高度 h_a (一般为 $h_a = 1.56m$), 大地温度 θ_0 (由气象资料得 θ_0 一般为 $10-12^\circ C$), 地表原有水量 l_0 (由气象资料可得)。

可见要确定上述参数, 我们有以下几种方法:

- (1) 若能得到 Φ 、 l_{down} 、 l_{up} 、 q_a 、 V_a 这些气象指标在每一时刻的值, 带入模型中即可以确定出待定系数 k_θ 、 k_q 、 k , 从而可以利用该模型算出任意时刻的能见度, 从而确定任意时刻的沙尘天气等级。同时, 若还能直接得到每一时刻的地表温度 θ_g 和地表湿度 q_g 的数据, 那么就只需确定待定系数 k 即可。可见, 该模型对于数据的要求较高, 拥有的数据越多, 模型就越适用。当然, 要想获得这些气象指标的瞬时值, 并不是一件容易的事情, 只有气象观测人员才最有可能获得这些实时数据。所以, 该模型最适合于从事气象观测的人员使用。

(2) 一般情况下, 鉴于这些瞬时数据很难记录, 加上很少公布, 我们只能得到上述气象指标在一段时间(如一个月)的平均值。在这样的情况下, 我们可以利用搜索算法对上述模型的参数进行确定。其具体思想是: 首先, 我们建立各气象指标的概率分布模型及随时间的变化规律, 利用我们所能得到的平均值数据得到为到每个时刻的数据, 采用空间搜索算法, 利用该模型对沙尘暴次数进行预测, 使得预测发生次数与实际发生次数误差最小, 进而确定参数。

4.2.4.1 建立各气象指标的概率分布模型

由于 Φ 、 l_{down} 、 l_{up} 、 q_a 、 V_a 这些气候指标的影响因素很多, 且每个因素均不在影响中起主导作用, 对于多因素、且影响均不突出的实际指标, 一般认为其服从正态分布, 即不会出现偏离正常值很远的数。故我们认为这些指标都是服从正态分布, 如下所示:

$$\Phi \sim N(\Phi^0, \sigma_\Phi^2)$$

$$V_a \sim N(V_a^0, \sigma_{V_a}^2)$$

$$\theta_a \sim N(\theta_a^0, \sigma_{\theta_a}^2)$$

$$l_{down} - l_{up} \sim N(l^0, \sigma_l^2)$$

$$q_a \sim N(q_a^0, \sigma_{q_a}^2)$$

式中 Φ^0 和 σ_Φ^2 分别表示指标 Φ 的期望和方差, 其余符号类似。

由概率论知识可知, 两个服从正态分布的变量的和依然服从正态分布, 则有

$$X_1 = 4k_\theta \Phi + 4\theta_0 - 4\theta_a \sim N(\mu_1, \sigma_1) \quad \backslash * \text{MERGEFORMAT} <33>$$

$$X_2 = \theta_a - k_\theta \Phi - \theta_0 - 0.608k_q k_\theta k \Phi l_0 - 0.608k_q k \theta_0 (l_{down} - l_{up} - l_0) \sim N(\mu_2, \sigma_2) \quad \backslash * \text{MERGEFORMAT} <34>$$

$$\text{其中, } \mu_1 = 4k_\theta \Phi^0 + 4\theta_0 - 4\theta_a^0$$

$$\sigma_1 = 4k_\theta \sigma_\Phi^2 + 4\sigma_{\theta_a}^2$$

$$\mu_2 = \theta_a^0 - k_\theta \Phi^0 - \theta_0 - 0.608k_q k_\theta k \Phi^0 l_0 - 0.608k_q k \theta_0 (l^0 - l_0)$$

$$\sigma_2^2 = \sigma_{\theta_a}^2 - k_\theta \sigma_\Phi^2 - 0.608k_q k_\theta k l_0 \sigma_\Phi^2 - 0.608k_q k \theta_0 \sigma_l^2$$

则可见度可化为

$$s = \frac{3.912\theta_a(V_a^2 + X_1)}{kgh_a(X_2 + 0.608q_a\theta_a k - 0.608k_q k_\theta k \Phi l_{down} + 0.608k k_\theta k_q \Phi l_{up})} \quad \backslash * \text{MERGEFORMAT} <35>$$

4.2.4.2 建立各气象指标随时间的分布规律

我们仍以西安为例, 阐述我们的思想。

我们建立陕西西安各年一月份一天中各时段可见度的模型, 考虑对气象来说, 在一小时内变化不明显, 精度足够, 以小时为间隔建立模型。

尽可能确定各因素分布函数中的参数, 以减少搜索空间

利用题中所给月平均数据可近似一天的日平均数据, 再根据客观规律来确定一天内各指标随时间的分布, 进而确定出某一时刻各数据的值, 以之作为该时刻相应数据的分布的均值。

(1) 建立温度随时间的变化规律

我们分析认为，一天中的气温在中午14时最高，夜里达到最低，因此可认为温度在一天内呈正弦规律变化，只是幅度不同而已。于是我们建立气温随时间的变化规律为：

$$\theta_a(t) = \theta_{\max} \sin(\pi t / 12 - 2\pi / 3) + \theta_a, t \in [0, 24] \text{ 且 } t \in \mathbb{Z} \quad \text{MERGEFORMAT} \langle 36 \rangle$$

据考察，该地区一月份通常最高温度 $\theta_{\max} + \theta_a$ 在 $10 \sim 20^\circ\text{C}$ 之间，我们取幅值在 $10 \sim 20^\circ\text{C}$ 之间等间隔变化，据此得到所有均值，

(2) 建立总辐射随时间的变化规律

考虑到总辐射强度在夜晚为0，早晚弱，中午强，故认为在白天服从正弦分布，在14时达到峰值。即总辐射强度随时间的变化规律为：

$$\Phi(t) = \begin{cases} \Phi_{\max} \sin(\frac{\pi}{12}t - \frac{2}{3}\pi), & 6 \leq t \leq 18 \\ 0, & \text{其它} \end{cases}, t \in \mathbb{Z} \quad \text{MERGEFORMAT} \langle 37 \rangle$$

同时满足一天内辐射总强度为该日均辐射强度，即归一化后，乘以该日均辐射强度（月辐射强度除以天数31） $\Phi^0(t) = \frac{\Phi(t)}{\sum \Phi(t)} \times \frac{\Phi_{\text{月}}}{31}$.

(3) 建立总蒸发量随时间的变化规律

考虑到蒸发强度白天强、晚上弱，于是我们认为白天的蒸发强度为一定值，夜晚为另一一定值，只是不同的日期比值不同，于是可得：

$$l_{up}(t) = \begin{cases} l_{day}, & t \in [6, 18] \\ l_{night}, & t \in [0, 6] \cup [19, 24] \end{cases}, t \in \mathbb{Z} \quad \text{MERGEFORMAT} \langle 38 \rangle$$

风速，降雨量，空气相对湿度与时间关系变化不明显，故可认为在所有时刻，取某一数值的概率相同，认为月平均值即可以近似分布中的均值。这时我们就得到了所有因素在任意时刻（以小时为间隔）的分布模型，其中所有的均值已经确认，方差未知。

4.2.4.3 利用空间搜索算法确定参数

在给定方差、待定系数的情况下，利用概率论知识，求出各时刻不发生沙尘暴的概率，在概率较大时认为其发生，统计一个月内发生次数，与实际发生次数进行比较，当误差最小时停止搜索，此时得到的系数即是我们所希望得到的数据。根据上述分析，我们用 MATLAB 软件编写出了实现程序，但是由于时间原因，我们在本文中没有得出最优解。具体实现程序见附录。

五，模型检验与结果分析

5.1 模型一的检验

5.1.1 模型的精度分析

为了检验模型的精度，我们从以下几个指标进行讨论。

绝对误差 δ ：预测值和实际值的最大偏差，这一指标主要是从个体考虑所产生的最大误差。在现实预测中的实际意义就是反映在特定的某一次预测中，最大所能出现的预测误差为多大。

根据计算，我们得到 $\delta_{\max}=0.2693$ $\Delta_{\min}=-0.2905$ ，将归一化值，恢复后得到 $\Delta_{\max}=0.2693 \times 4=1.0772$ $\Delta_{\min}=-1.162$ 。

分析表明在该模型在实际预测中，最大可能多预测一次，或少预测1次。实际结果中，可以总结得到：最大可能出错为少预测一次。这种预测精度能够满足人们的要求。

(2) 二维范数距离 $W = \sqrt{\frac{\sum_i (t_i - o_i)^2}{n}}$ (其中： t_i 表示神经网络第*i*个样本的期望输出值， o_i 表示神经网络第*i*个样本的计算输出值， n 表示进行*n*次预测)：表示在多次预测中，所有预测值与实际值的偏差距离。这一指标主要是从总体上考虑误差。在现实预测中的实际意义就是反映在多次预测中，总体误差的大小。根据计算我们得到：

$$W = \sqrt{\frac{\sum_i (t_i - o_i)^2}{n}} = 0.0447 \quad \backslash * \text{MERGEFORMAT} <39>$$

可见总体误差很小，说明BP神经网络方法有非常好的收敛性，能够很好的贴合实际。同时，从上述分析中我们还可以得出结论：预测次数越多，该神经网络模型的精度越高。

(3) 成功界限指数

我们定义成功界限指数为：

$$CSI = \frac{cf}{cf + wf} \times 100\% \quad \backslash * \text{MERGEFORMAT} <40>$$

其中 cf 为正确报出沙尘暴的次数， wf 为漏报和错报次数。

陕西省57036号观测站1961年到2000年每个月份得沙尘暴日发生次数预测情况：一共预测338次，其中出现5次错报、一次预测次数不准确。

即 $cf=332$ 、 $wf=6$ ，于是成功界限指数为：

$$CSI = \frac{cf}{cf + wf} \times 100\% = \frac{332}{338} \times 100\% = 98.22\% \quad \backslash * \text{MERGEFORMAT} <41>$$

可见成功界限指数也很高，说明该模型精度较高。

综合上述分析，我们可以得出结论，基于BP神经网络的分析预测模型在预测沙尘暴问题中精度较高，能够很好的运用到沙尘暴预测的实际问题中去。

得神经网络预测的准确率很高，说明神经网络方法在沙尘暴预测过程中可以得以很好的运用，并得到精确结果。

5.1.2 模型的稳定性分析

为了检验模型的稳定性，我们可以从以下几个方面进行分析。

(1) 模型对各个因子的稳定性分析，即灵敏度分析，这在第四部分我们已经详细讨论过，这里不再分析。

(2) 模型对样本大小的稳定性分析，我们可以通过减少训练样本的数据量来分析对模型精度的影响，如我们采用陕西省57036号观测站1981年到2000年的观测数据作为样本再次进行训练，然后进行预测，预测结果表明，样本减少后的神经网络预测精度降低，可见模型对样本大小的稳定性较低。

六，模型的进一步讨论

6.1 模型一的进一步讨论

在问题一中，我们建立的基于BP神经网络的沙尘暴分析预测模型在预测沙尘暴发生次数中，达到了较高的精度，能够得到很好的应用，但是若能在以下方面对该模型进一步的深入研究和讨论，则能更好的改进该模型。

1, BP神经网络主要是单纯从数据的角度分析得出各气象指标与沙尘暴发生次数之间的关系，不能很好的反映事物的物理机理，必然不能解释一些物理事实。若能通过改进算法将物理机理和数据关系联系起来综合分析，则能达到更理想的效果。

2, 由于神经网络是以传统统计学理论为基础的，而传统统计学的前提是要有足够大的样本，在本问题中，我们采用的BP神经网络的训练样本较大，所以能达到较高的精度，若在实际问题中我们能够采集的样本很小，该模型未必还能得到这么高的精度。若能通过改进算法，使得BP神经网络在小样本条件下也能达到很高的精度，则将使该模型在实际中应用更加广泛。

3, 我们在该模型中的训练时间较长，在实际生活中，进行预测当然是越快越好，而BP神经网络的精度是跟训练时间成正比的，若能对BP神经网络进行改进，使得在保证精度的前提下缩短训练时间，将大大提高该模型的实用性。目前，对这方面的研究也有很多，如S函数输出限幅算法和累积误差校正算法以及学习系数 η 的自应调整算法[8]等都较好的解决了训练时间过长的问题，但各自依据的公式是有所不同的。这对我们提出更好的改进方法也是很大的启发。

6.2 模型二的进一步讨论

在问题二中，我们建立的基于沙尘动力学的机理模型从物理意义角度建立起沙

报名号#1070

尘暴与各气象指标之间的关系，能够客观的反应事物的物理机理，理论依据坚实，而且在该模型中通过建立各气象指标的概率分布模型和随时间的变化规律，利用空间搜索算法来确定模型参数的方法思路开阔。但有以下方面需要改进：

(1) 该模型中有些概率分布模型的数据处理较为粗糙，主观性太强，若能在处理这些问题时更加客观，更加细致，则能提高模型精度；

(2) 该模型中在确定参数时搜索空间太大，使得计算过于复杂，若能在算法上加以改进，使得计算简单，则能增加模型的实用性。

七，模型优缺点

7.1 模型一的优缺点

7.1.1 模型的优点

- (1) 模型的精度较高，通过预测验证，能够达到较高的精度；
- (2) 模型的通用性强，适用于其它分析预测问题；
- (3) 模型简单，使用方便。

7.1.2 模型的缺点

(1) BP神经网络模型是基于传统统计学理论从数据角度分析问题的，不能很好的解释事物的物理机理，在有些时候与实际情况不符；

- (2) 该模型对数据样本要求较高，且训练时间长。

7.2 模型二的优缺点

7.2.1 模型二的优点

(1) 模型基于沙尘动力学机理建立，能很好的反映事物的物理机理，理论依据坚实，推导过程严谨，物理概念清晰；

(2) 模型考虑的因素全面，机理分析透彻，可以对将来一段时间内各因素指标进行预测，从而既可以进行短期预测，也可以进行中长期预测，既可以预测沙尘暴的强度又可以预测发生概率，全面具体。

(3) 模型中确定参数时，运用分布机理，对月均数据进行分析，得到日均数据，进而得到各时刻数据，在分布参数无法确定的情况下，将其与待定系数同等看待，采用空间搜索法，使预测结果与实际发生次数差别最小，进而获得所有待定数据，不仅解决了各因素影响沙尘暴的具体关系，还得到了各因素的分布情况。

7.2.2模型二的缺点

- (1) 在确定模型参数时，搜索空间大，耗费时间长；
- (2) 各概率分布模型的数据处理较为粗糙，使得模型精度下降。

八、参考文献

【1】赵翠光，人工神经网络方法在沙尘暴短期预报中的应用，《气象》，第30卷第4期，2002年9月。

【2】碧涵，沙尘天气分为四个等级，问天网<http://www.tq121.com.cn/>，2008年4月11日。

【3】刘良林、王全凤、林煌斌，BP神经网络参数设定及应用，《基建优化》，第28卷第2期，2007年四月。

【4】吕峥，沙尘暴是怎样形成的，千龙网<http://www.qianlong.com/>，2008年4月11日。

【5】刘新民、邵敏，北京市夏季大气消光系数的来源分析，环境科学学报，第24卷第2期，2004年3月。

【6】杨德保、尚可政、王氏功，《沙尘暴》，北京，气象出版社，2003年3月。

【7】金伟其，《辐射度、光度与色度及其测量》，北京，北京理工大学出版社，2006年6月。

【8】陈郁、秦奋、余明全，BP神经网络改进及其在Matlab中应用程序发布的研究，河南教育学院学报（自然科学版），第16卷第2期，2007年6月。