

西安电子科技大学2019年数学建模校内赛

承 诺 与 产 权 转 让 书

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们同意将参赛论文以及支撑材料中的所建模型、算法以及程序产权归属西安电子科技大学以及合作单位共有。特别的，B题参赛论文以及支撑材料中的相应产权西安电子科技大学拥有50%，合作单位享有50%。2019 年数学建模校内赛竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号是（从A/B中选择一项填写）：_____ A _____

参赛报名队号为：_____ 19B362 _____

报名时所属学院（请填写完整的全名）：_____ 微电子学院 _____

参赛队员姓名与学号(打印，用二号字，并签名)：

1. _____ 王立恒17140220086 _____

2. _____ 杨玉修17020130088 _____

3. _____ 陈俊鑫17020130006 _____

日期：2019年5 月3 日

西安电子科技大学2019年数学建模校内赛

评 阅 专 用 页

	评阅人1	评阅人2	评阅人3	总评
成绩				

基于熵权法和BP神经网络的视力预警与干预模型

摘要

随着现代信息化社会发展节奏加快，每人每天在接受更多有效信息的同时，用眼时间不断增加。我国近视人数逐年增加，且呈现低龄化趋势，这个问题已成为重大社会公共问题。本文针对儿童与青少年近视问题进行调查与分析，分析出眼睛视力预警机制模型的指标，引入AI 预警机制并提出了应用此模型的方案。

对于问题一，我们列出了影响视力的六个指标，并将包含这六个指标以及近视度数的调查问卷发布在网上，收集到较为广泛的真实数据。为了尽量消除计算的影响，我们将数据无量纲化，再建立并求解熵权法模型得到每一项指标的权重。

对于问题二，我们使用问题一解出的各项指标的权值，刻画近视强度 $f(i)$ 。进行无量纲化操作后，将每个调查对象对于每个指标的值带入 $f(i)$ ，并制成表格。以近视度作为自变量、近视度数作为因变量，建立近视度- 近视度数一元线性回归方程模型。并利用最小二乘法知识解得线性回归方程为： $\hat{y} = 105.8562 + 767.0195x$ 。用MATLAB做模型图并进行检验。与原始数据图进行比，发现趋势比较吻合。

对于问题三，我们使用问题一中各项指标的权值与问题二中最后得到的线性回归方程建立了一个视力预测模型公式 $\hat{y} = 105.8562 + 767.0195 \times (0.55x_{i1} + 0.04x_{i2} + 0.14x_{i3} + 0.16x_{i4} + 0.05x_{i5} + 0.05x_{i6})$ 。利用MATLAB建立可实现交互并拥有完整流程的预测界面，这个界面基于调查个体的回答可以给出预测视力，对当前用眼习惯的评价以及建议的做法，引导家长与孩子积极预防近视或防止度数加深。

对于问题四，我们建立了bp神经网络模型，利用MATLAB 实现此模型。我们将本次收集到的有关六个影响视力因素的回答输入此模型，继续输入所有调查对象对应的真实近视度数，使bp神经网络自己模拟出拟合曲线，使得下一次输入有关这六个影响因素的回答时，直接给出预测近视度数。

对于问题五，我们综合了上述问题，权衡每个影响视力因素的重要性，给出了一份有效可行的预防近视或防止度数加深的方案。

关键词：眼睛视力 熵权法模型 一次线性回归方程 最小二乘法 BP神经网络

1 问题重述

1.1 问题背景

近年来，我国近视低龄化的现象越来越严重，在经济发达地区尤为明显。有数据显示，我国小学生近视比例为45.7%，初中生近视比例为74.4%，高中生近视比例为83.3%，大学生近视比例则高达87.7%，我国人口近视发病率高达33%。除了先天因素的影响及户外运动的减少外，现在的儿童及青少年能更早更频繁地接触电子设备进行娱乐活动也是影响其视力的一个重要原因。此外，不正确的用眼方法以及对眼睛保护的不重视也造成了越来越多的小孩子患上近视。

近视不仅会给人带来不便，还会影响其特长与爱好的发展，甚至危及到国家安全。许多学校与专业都有严格的视力要求。近年来，军检中因为视力不合格而被淘汰的人很多，并且有增长趋势，各地方政府的征兵难度逐年增加。

所以分析不同影响因素对视力影响的程度，并做出预警机制模型与交互软件预测孩子成年后的视力情况，对家长与孩子更好的了解日常生活中这些因素对视力的影响，并且及早地预防近视的发生或是防止度数继续加深有深远的意义。

1.2 问题提出

本文将题述问题归结为如下五个问题：

问题一：

考虑注意信息化时代的优势，给出有关影响视力的关键因素的数据以及其获取途径。

问题二：

基于问题一，给出有高执行性的眼睛视力的演化机理模型。

问题三：

基于问题二，给出通俗易懂的眼睛视力的预警机制模型，其原则是在确定的时间做对应的预警，其面向的对象可以多样化。

问题四：

在上述基础上设计可供学习的信息系统（数据表），通过查阅资料和数据及仿真等手段进行眼睛视力的机器学习模型。

问题五：

写一份可以实现文中机理模型和学习模型的方案。

2 问题分析

2.1 问题一的分析

问题一要求设置合理的指标用来分析不同因素对于近视形成的影响，我们可以选用遗传因素，使用电子设备时长等指标。

首先建立初始矩阵，使得从行 x_i 来看是某个调查对象对所有影响视力指标的回答、从列 x_j 来看是某个指标对于不同调查对象所呈现的数据。利用极差公式将 x_j 无量纲化，并且根据每个调查对象关于同一指标的回答将初始矩阵划为比重矩阵。然后根据公式计算出第 j 项指标的信息熵 e_j 和差异系数 d_j ，进而用EXCEL求得每项指标的权重。从而得到各项指标对于近视的影响程度。

2.2 问题二的分析

问题二基于问题一给出的权重，定义某个调查对象的近视强度 $f(i)$ 为其对于某个影响因素的回答乘以此影响因素对应的权重。利用极差变换公式将 x_j 无量纲化。可以求出每个调查对象对应的近视度。将近视强度作为自变量、近视度数作为因变量，建立一个一元线性回归方程模型 $y = \theta_0 + \theta_1 x + \varepsilon$ 。并利用最小二乘法求出使得误差 $\sum_{i=1}^m \varepsilon^2$ 最小的 $\hat{\theta}_0$ 和 $\hat{\theta}_1$ 。最后使用MATLAB来进行数据的处理和图表输出。最后进行显著性检验。从而得到一个比较符合实际的、可以说明影响因素对视力影响程度的的线性方程。

2.3 问题三的分析

问题三中的预警机制模型需要基于问题一各个因素的权重和问题二中解出的一元线性方程，可以利用MATLAB做出一个可以进行交互的界面，引导家长或者学生自己对直系亲属近视率、日均使用电子产品时间、日均学习时间等的影响因素进行合理回答，并根据这些回答自动计算出这个人未来的视力情况（包括是否近视、近视度数以及近视程度所属范围）。

2.4 问题四的分析

问题四中的机器学习模型可以使系统自动对接收到的大量数据进行分析与学习，并总结规律，进而在输入新的数据时可以根据规律直接得出结果。可以基于前面收集到的数

据，利用MATLAB建立BP神经网络模型并进行求解，然后与问题二中解出的线性模型进行对比。

3 假设与符号

3.1 基本假设

- 假设各个影响视力的因素对于所有年龄段的人影响程度都相同；
- 假设近视度数的最高上限是1200 度；
- 假设我们所获得的数据能够代表广大儿童与青少年的视力数据；
- 假设近视的度数都是50的倍数。

3.2 符号说明

表 1: 本文符号说明

符号	代表意义
x_i	第i个调查对象在所有指标下的值
x_j	所有调查对象在第j个指标下的值
x_{ij}	第i个调查对象在第j个指标下的值
y_i	第i个调查对象在所有指标下的无量纲化后的值
y_j	所有调查对象在第j个指标下的无量纲化后的值
y_{ij}	第i个调查对象在第j个指标下的无量纲化后的值
e_j	第j项指标的信息熵
d_j	差异系数
K	玻尔兹曼常数
w_j	第j项指标的权重
$f(i)$	近视的强度
θ	最小二乘法线性方程中的系数
ε	最小二乘法线性方程中的误差

4 模型建立与求解

4.1 问题一的建模与求解

4.1.1 模型的准备

根据研究表明，视力受到多方面因素影响^[1]，我们选取几种主要因素，包括遗传因素（以直系亲属中近视率刻画），户外活动时间，电子产品使用时间，看书学习时间，不正当用眼姿势，眼部保护等方面。以调查问卷形式，我们抽样调查了多组数据，并制成以下表格。我们希望通过数据分析，量化影响因素并得到相应权重。

表 2: 部分近视数据表

直系三代亲属中近视率	日均户外活动时间	日均电子产品使用时间	日均看书学习时间	是否有不正确用眼姿势	是否有眼部保护
0.33	1	7	2	1	2
0.5	2	3	3	1	0
0	2	4	5	1	1
0	1	2	3	1	1
0.25	1	6	1	1	2
0	2	2	3	1	1
0	1	6	2	1	2
0.33	1	6	1	1	1
0	3	4	4	1	2
0	1	3	2	0	0

数据解释：直系三代亲属中近视率=近视人数/总人数；户外运动时间等数据单位均为小时；有不正确用眼姿势习惯数据为1，没有则为0；无眼部保护习惯（包括眼保健操，护眼产品使用等）数据为2，偶尔有数据为1，经常性数据为0。

4.1.2 熵权法模型的建立

对权重量化，多采用层次分析法、模糊综合评价法、主成分分析法等方法来评价资源安全；然而，层次分析法和模糊综合评价法在计算指标权重时的主观性较强，主成分分析不能反映全部指标的信息。鉴于此，本文选取一种依据历史客观数据反映实际情况的评价方法——熵权法^[2]，从而有效避免指标权重计算的主观随意性和模糊性。

熵权法是通过计算指标的信息熵，利用指标的差异程度来度量已知数据中包含的有效信息和指标权重。熵值越小，说明指标的离散程度越大，表明其信息的有效价值越大，该指标在综合评价中对目标的影响也就越大。若某列元素数值都相同，则熵最大值为1，熵权为0。这表明在某指标上若各评价对象的数值相同，则该指标未包含任何有价值的信息。若某列元素数值相差越大，则熵值就越小，熵权就越大，则说明该指标包含的有价值的信息越多，指标越重要。应用熵权法进行评价的操作步骤如下：

(1) 初始数据矩阵的建立及标准化

设有m个调查对象，n个评价指标，则形成评价系统的初始数据矩阵

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$$

其中 x_{ij} 表示第i个调查对象在j项指标中的值 ($i=1,2,3\cdots m, j=1,2,3\cdots n$) ; x_j 表示第j项指标下，各个调查对象值形成的列向量。

由于各个指标之间存在量纲差异，量纲的引入会对计算造成影响。为消除量纲带来的影响，常采取极差变换法^[3]对数据进行无量纲化。

我们将造成或加重近视的影响因素归为正向指标，则遗传因素，电资产品使用时间，看书学习时间，不正当用眼姿势均为正向指标。户外活动时间，眼部保护均为负向指标。

极差变换公式如下：

$$X'_j = \frac{x_j - \min\{x_j\}}{\max\{x_j\} - \min\{x_j\}} \quad (\text{适用于正向指标})$$

$$X'_j = \frac{\max\{x_j\} - x_j}{\max\{x_j\} - \min\{x_j\}} \quad (\text{适用于负向指标})$$

(2) x'_{ij} 所占比重 y'_{ij}

$$y'_{ij} = \frac{x'_{ij}}{\sum_{i=1}^m x'_{ij}}$$

由此得到比重矩阵 $Y = (y_{ij})_{m \times n}$

(3) 第j项指标的信息熵 e_j 和差异系数 d_j

$e_j = -K \sum_{i=1}^m y_{ij} \ln y_{ij}$ ，K为玻尔兹曼常数， $0 \leq e_j \leq 1$ ，且规定 $y_{ij} = 0$ 时， $y_{ij} \ln y_{ij} = 0$ ， $d_j = 1 - e_j$ 。

(4) 第j项指标的权重 w_j

$$w_j = \frac{d_j}{\sum_{j=1}^n d_j} = \frac{1 - e_j}{n - \sum_{j=1}^n e_j}$$

4.1.3 模型的求解

根据理论和公式我们计算出各项指标对近视影响权重如下表:

表 3: 各项指标对近视影响的权重

指标 x_j	直系亲属近视率	户外运动时间	电子产品使用时间	看书学习时间	是否有不正确用眼姿势	是否有眼部保护
权重 w_j	0.547351289	0.043075302	0.143779778	0.162440807	0.048426796	0.054926028

4.2 问题二的建模与求解

4.2.1 问题二模型的准备

问题二要求建立近视演化模型, 根据问题一的求解, 我们可以得到一个影响因素的加权和公式, 可基于此公式联系近视情况以及近视者近视程度进行分析。

4.2.2 问题二模型的建立

我们引入近视度的概念刻画近视程度 $f(i)$ 。结合问题一计算出的权重, 我们可以做如下定义:

$$f(i) = \sum_{j=1}^n x_{ij} \times w_j$$

即 $f(i) = 0.55x_{i1} + 0.04x_{i2} + 0.14x_{i3} + 0.16x_{i4} + 0.05x_{i5} + 0.05x_{i6}$ 。

最终我们所呈现的近视度将以数据的形式呈现, 因此我们需要考虑到将影响因素 x_{ij} 量化, 并消除量纲带来的影响。这里我们仍然采用极差变换法, 极差变换公式如下:

$$X'_j = \frac{x_j - \min\{x_j\}}{\max\{x_j\} - \min\{x_j\}} \quad (\text{适用于正向指标})$$
$$X'_j = \frac{\max\{x_j\} - x_j}{\max\{x_j\} - \min\{x_j\}} \quad (\text{适用于负向指标})$$

作出相应变化以及公式代入后, 我们能够得到每个评价对象相应的近视度, 借此我们可以衡量其是否有近视风险以及近视的严重程度。

4.2.3 问题二模型的求解

根据公式，我们计算出每个评价对象对应的近视度，并制成表格。部分表格如下，具体表格见附录。

对照这些评价对象的近视情况（包括是否近视，近视度数，散光等），我们不难发现，近视度与近视情况成正相关，这是符合常识的。我们所得到的图大致符合这一结果：

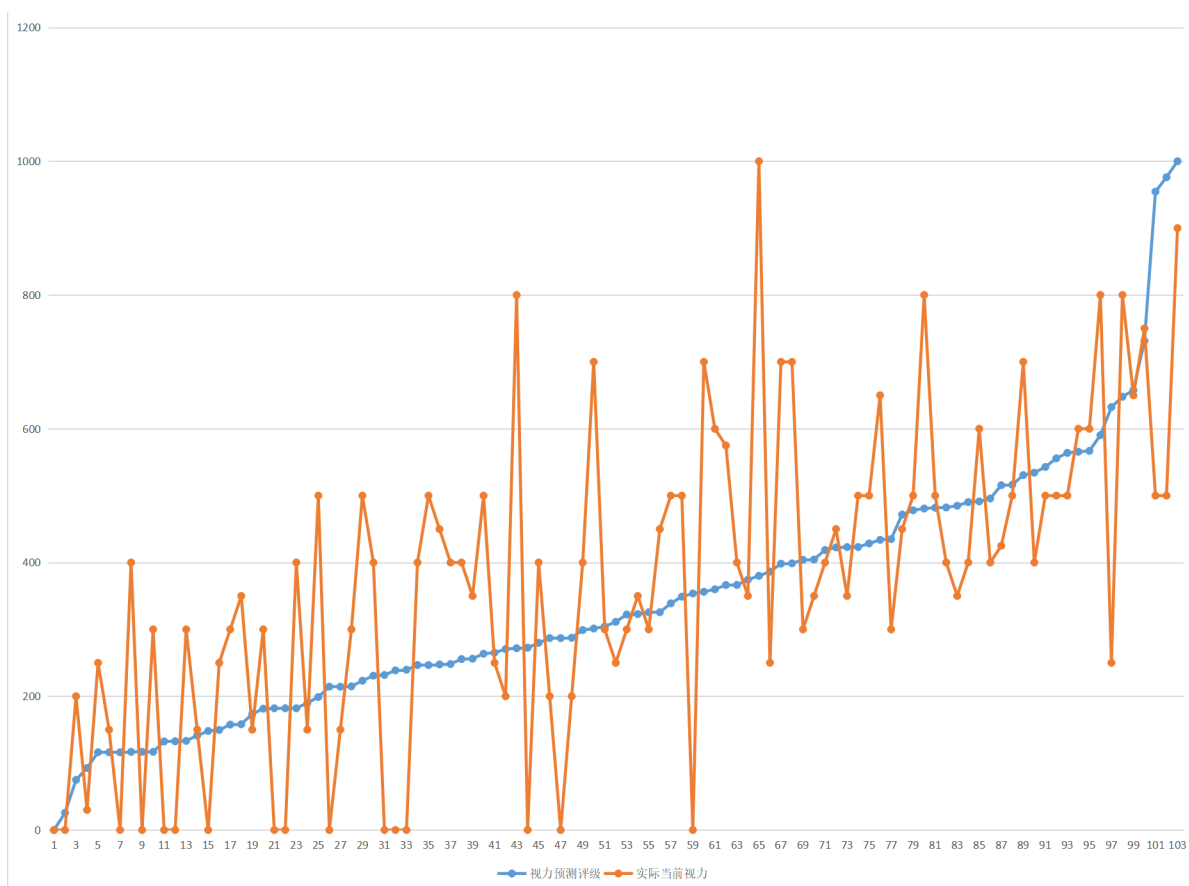


图 1：预测视力与实际视力比较折线图

当近视度大于某个值时，出现近视，且之后的近视度数随着近视度的增加而升高。因此，我们希望建立近视度-近视度数的一元线性回归模型，具体模拟和研究近视度与近视度数的关系，以便更好地预防和干预近视。

4.2.4 近视度-近视度数一元线性回归方程模型的建立

我们以近视度作为自变量 x ,以近视度数作为因变量 y , 建立如下一元线性回归方程:

$$\begin{cases} y = \theta_0 + \theta_1 x + \varepsilon \\ E\varepsilon = 0, D\varepsilon = \sigma^2 \end{cases}$$

关于模型参数估计: 有 m 组观测值: $(x_1, y_1) \quad (x_2, y_2) \cdots (x_m, y_m)$

$$\text{设} \begin{cases} y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, & i = 1, 2, 3 \dots m \\ E\varepsilon_i = 0, D\varepsilon_i = \sigma^2, & \text{且}\varepsilon_i\text{相互独立} \end{cases}$$

记 $Q = Q(\theta_0, \theta_1) = \sum_{i=1}^m \varepsilon_i^2 = \sum_{i=1}^m (y_i - \theta_0 - \theta_1 x_i)^2$,

最小二乘法就是选择 θ_0 和 θ_1 估计 $\hat{\theta}_0$ 和 $\hat{\theta}_1$ 使得 $Q(\hat{\theta}_0, \hat{\theta}_1) = \min Q(\theta_0, \theta_1)$ 。

$$\text{解得} \begin{cases} \hat{\theta}_0 = \bar{y} - \hat{\theta}_1 \bar{x} \\ \hat{\theta}_1 = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \end{cases} \text{可得出线性回归方程: } \hat{y} = \hat{\theta}_0 + \hat{\theta}_1 x = \bar{y} + \hat{\theta}_1 (x - \bar{x}).$$

模型的样本容量 $m=103$, 以左眼右眼度数的平均值代替近视度数, 以近视度 $f(i)$ 为横坐标, 近视度数 $g(i)$ 为纵坐标在平面直角坐标系上标出, 我们利用Matlab来进行数据的处理和图表输出。主要代码如下:

bint 是回归系数的区间估计, r 是残差, rint 是置信区间, stats 是用于检验回归模型的统计量值, 相关系数 r^2 。相关系数 r^2 越大, 说明相关性越强。

4.2.5 一元线性回归方程模型的求解与检验

(1) 利用MATLAB处理数据

```
x = xlsread('C:\Users\12086\Desktop\001.xlsx',1,'A1:A103');
y1 = xlsread('C:\Users\12086\Desktop\001.xlsx',1,'B1:B103');
X = [ones(103,1),x];
[b,bint,r,rint,stats] = regress(y1,X);
rcoplot(r,rint)
z1 = b(1) + b(2) * x;
plot(x,y1,'bx',x,z1,'r')
```

(2) MATLAB计算结果分析及检验

计算得出相关系数 $r = 0.53925 > 0$, 为正相关, 符合建模结果。

显著性检验， $t = r\sqrt{\frac{m-2}{1-r^2}} = 0.53925 \times \sqrt{\frac{103-2}{1-0.53925^2}} = 6.471$ ，根据显著水平 $\alpha = 0.05$ ，查t分布表得 $t_{\frac{\alpha}{2}}(m-2) \approx 2.9$ ，显然 $t > t_{\frac{\alpha}{2}}(m-2)$ ，故两者有相关性。

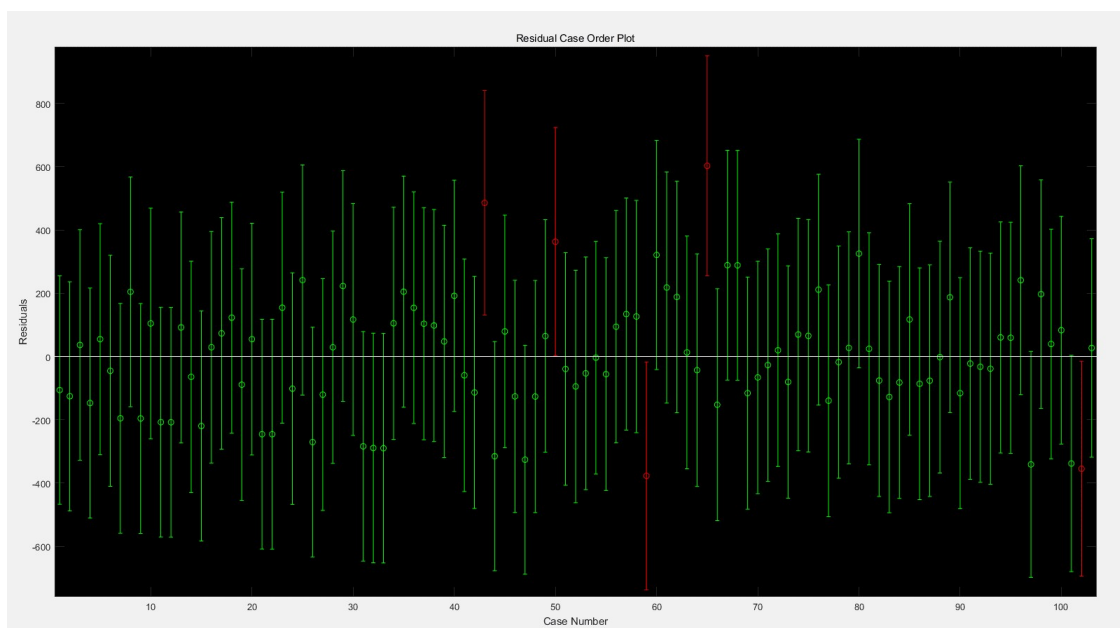


图 2：残差图

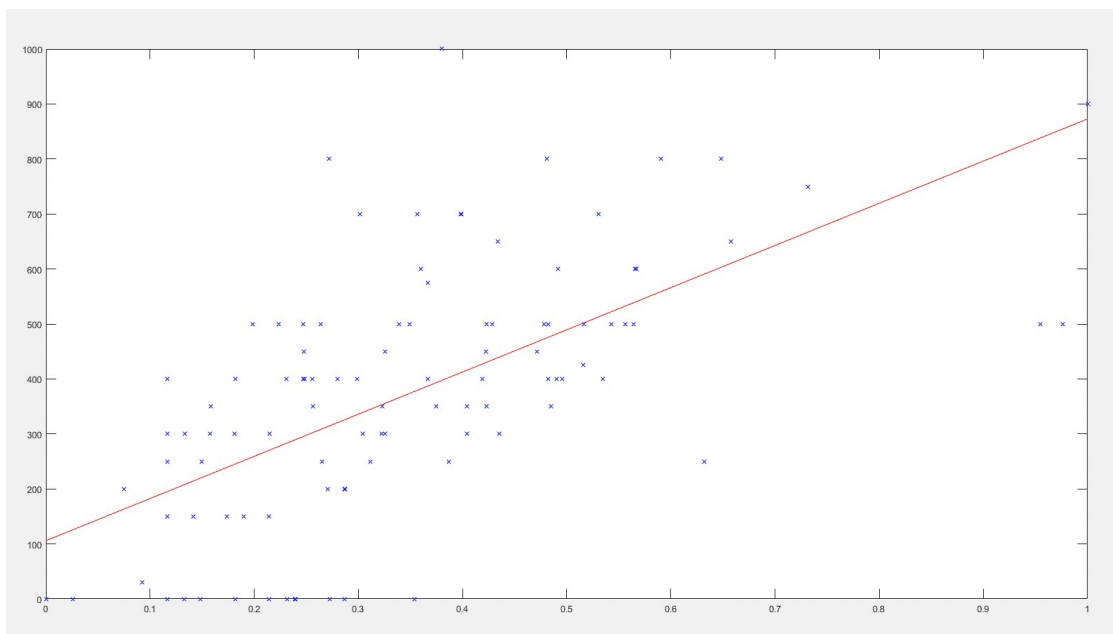


图 3：散点图

MATLAB计算得出 $b=105.8462 \quad 767.0195$; $bint=30.7752 \quad 180.9172 \quad 575.1794 \quad 958.8596$ 。由此可得 $\hat{\theta}_0 = 105.8562$ 置信区间为 $(30.7752, 180.9172)$ $\hat{\theta}_1 = 767.0195$ 置信区间为 $(575.1794, 958.8596)$ 分析残差图可以看出, 除5个数据点外, 其余数据的残差离零点均较近, 且残差的置信区间包含零点。再分析散点图发现数据点基本上均匀分布在回归方程两侧。这说明回归模型 $\hat{y} = 105.8562 + 767.0195x$ 能较好地符合原始数据, 而5个数据点可以视为坏值。

4.3 问题三的建模与求解

4.3.1 问题三模型的准备

问题三要求建立一个预警模型, 我们首先建立一个视力预测模型, 根据预测结果结合资料选取临界值讨论, 以达到预警的效果。

4.3.2 问题三模型的建立

每个调查对象 i 都对应着有六个评价指标以及六组数据。根据第一问和第二问的求解基础, 我们知道, 可以通过极差变换法将六组数据无量纲化得到标准数据 x_{ij} ($j=1, 2, \dots, 6$)。

根据第一问计算出的权重以及第二问引入的近视度概念, 我们可以得到以下公式:

个体 i 的近视度: $f(i) = 0.55x_{i1} + 0.04x_{i2} + 0.14x_{i3} + 0.16x_{i4} + 0.05x_{i5} + 0.05x_{i6}$

在第二问的求解中, 我们得到了近视度-近视度数的一元线性回归方程

$$\hat{y} = 105.8562 + 767.0195x$$

将 $x = f(i)$ 带入回归线性方程可得

$$\hat{y} = 105.8562 + 767.0195 \times (0.55x_{i1} + 0.04x_{i2} + 0.14x_{i3} + 0.16x_{i4} + 0.05x_{i5} + 0.05x_{i6})$$

此即视力预测模型公式, 将评价对象的各项数据处理后代入能得到一个预测视力。

4.3.3 问题三模型的求解

用matlab进行编程 (详细源代码见附录), 流程图如下:

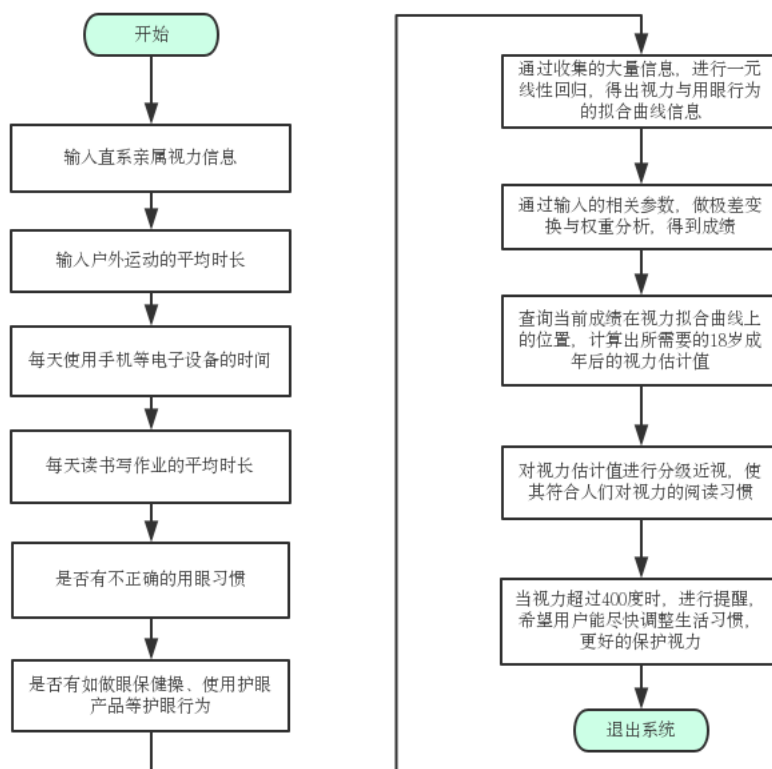


图 4：问题四的流程图

对于每一个评价个体，最终都能输出一个预测视力值。我们随机从调查表中选取两个个体，求得其相应的预测视力度数如下图：

命令行窗口

请按实情输入您或您的孩子的相关情况，我们将为您预测您孩子成年后的视力水平。

请输入待测试者的直系三代中的近视比例：
 比例为：近视人数/直系三代亲属总数。
 比例为：1/3

请输入待测试者现在每天进行户外运动的平均时长（0至1小时按1小时计算，输入1即可，1至2小时按2小时计算，输入2即可，以此类推）：1

请输入待测试者现在每天使用电脑、手机等电子产品的平均时长（0至1小时按1小时计算，输入1即可，1至2小时按2小时计算，输入2即可，以此类推）：7

请输入待测试者现在平均每天看书、写作业的总时长（0至1小时按1小时计算，输入1即可，1至2小时按2小时计算，输入2即可，以此类推）：2

请输入待测试者当前是否存在近距离或不当姿势使用手机、电脑或看书写作业等的行为（如果有请输入1，无请输入0）：1

请输入待测试者当前是否坚持每天做眼保健操或使用护眼产品（如果均有请输入0，有一项请输入1，均没有请输入2）：2

请确认数据：
 1. 待测者的三代直系亲属的近视比例为0.33333
 2. 待测者每天平均花费约1小时在运动上
 3. 待测者每天平均花费约7小时在电子设备上
 4. 待测者每天平均花费约2小时在阅读与作业上
 5. 存在不合适的阅读或使用电子设备的习惯。
 6. 没有做眼保健操或使用护眼产品的习惯
 请确认数据是否正确，正在为您计算，请稍后……。

预计在18岁成年后视力度数为500度
 当前的习惯对眼睛较为不好，请尽快予以矫正，\n建议减少电子产品的使用时间，阅读与写作业时多进行适当的放松，多进行一些户外运动

fx >>

图 5：预测视力例1图

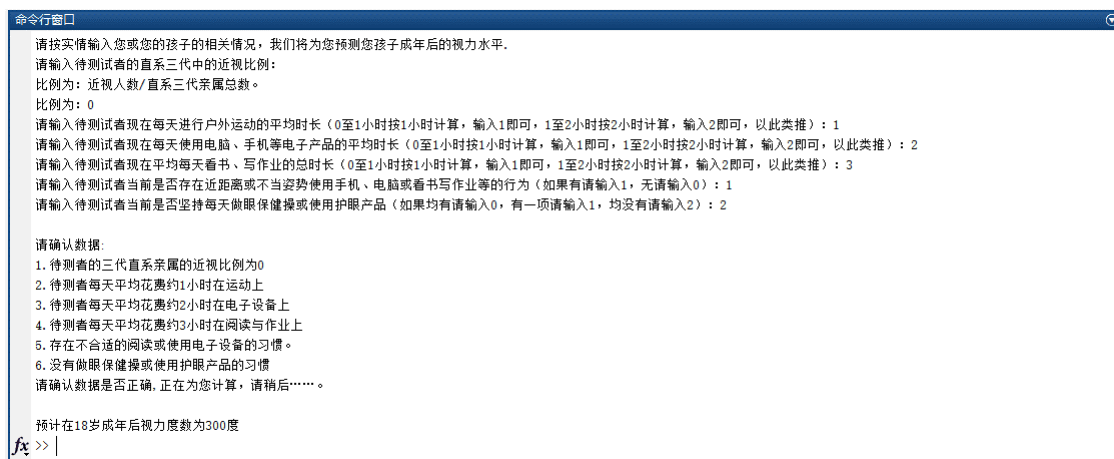


图 6: 预测视力例2图

对照数据里他们的近视度数，发现实际值和预测值相近（前者实际度数450，预测度数500；后者实际度数280，预测度数为300），这表明我们的预测模型是具有较高准确性的。

根据相关资料显示^[4]，当近视度数 $\in (0, 300)$ 时为轻度近视，近视度数 $\in (300, 600)$ 时为中度近视，近视度数 $\in (600, 900)$ 时为高度近视，近视度数大于900 时，为超高度近视。

因此，我们在视力预测模型中，分别以300, 600, 900 为临界值，并设置相应预警提示。

4.4 问题四的建模与求解

4.4.1 问题四模型的准备

BP神经网络是在输入层与输出层之间增加若干层神经元，这些神经元称为隐单元，他们与外界没有直接联系，但其状态的改变能影响输入与输出之间的关系。BP神经网络的计算过程由正向计算过程和反向计算过程组成。正向传播过程，输入模式从输入层经隐单元逐层处理，并转向输出层，每层神经元的状态只影响下一层神经元的状态。如果在输出层不能得到期望的输出，则转入反向传播，将误差信号沿原来的连接通路返回，通过修改各神经元的权值，使得误差信号最小。

4.4.2 问题四模型的建立

(1) 生成BP网络

$Net = newff(PR, [S1S2...SN], \{TF1TF2...TFN\}, BTF, BLF, PF)$ PR:由R维的输入样本最小最大值构成的 $R \times 2$ 维矩阵。 $[S1 S2 ...SN]$:各层的神经元个数。 $\{TF1 TF2 ... TFN\}$: 各层的神经元传递函数。BTF: 训练用函数的名称。

(2) 网络训练

$[net, tr, Y, E, Pf, Af] = train(net, P, T, Pi, Ai, VV, TV)$ 。

(3) 网络仿真

$[Y, Pf, Af, E, perf] = sim(net, P, Pi, Ai, T)$ 。

在影响视力的方向上，我们收集了遗传因素、阅读习惯、电子产品的使用习惯、运动因素等六个数据，因此，在隐含层中共有六个神经元。

4.4.3 问题四模型的求解

我们借助MATLAB的神经网络算法进行定义了两个数据矩阵，P矩阵包含所有已经获得的103份调查到的日常生活数据，T矩阵包含了所有的被调查者的当前实际视力数。在导入数据后，先进行数据的归一化，利用mapminmax函数，使所有输入的数据归一化到 $[-1,1]$ 之间。

接下来对数据进行分组，随机抽取70%的数据作为训练数据，15%的数据作为测试数据。经过多次尝试，我们使用了'logsig'作为隐层的传输函数，'purelin'作为输出层的传输函数。

运行后的结果为:

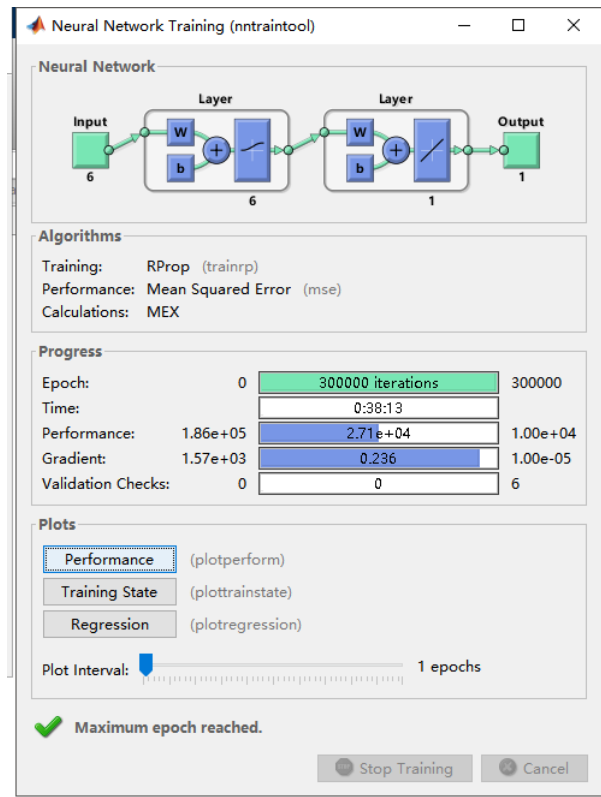


图 7: BP神经网络的结果分析图

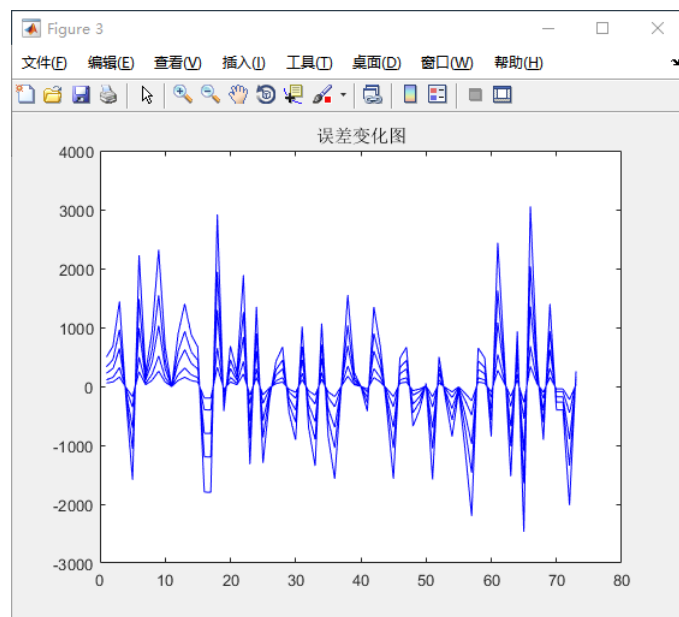


图 8: 误差图

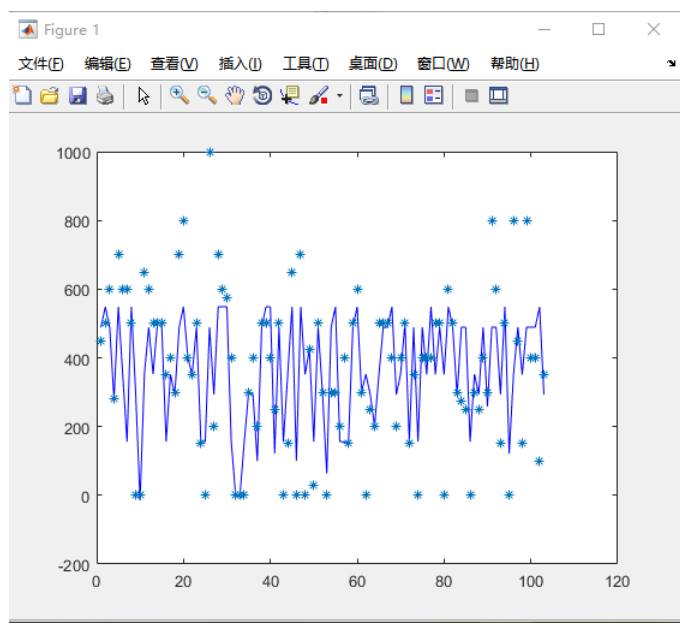


图 9: 实际点与预测曲线图

实线为预测曲线，“*”为实际的点，可以看到基本保持一致。（由于当前可供进行机器学习的数据较少，因此预测曲线波动较大，如果数据集足够多，波动应当会大幅度减少，逐渐趋于稳定。）

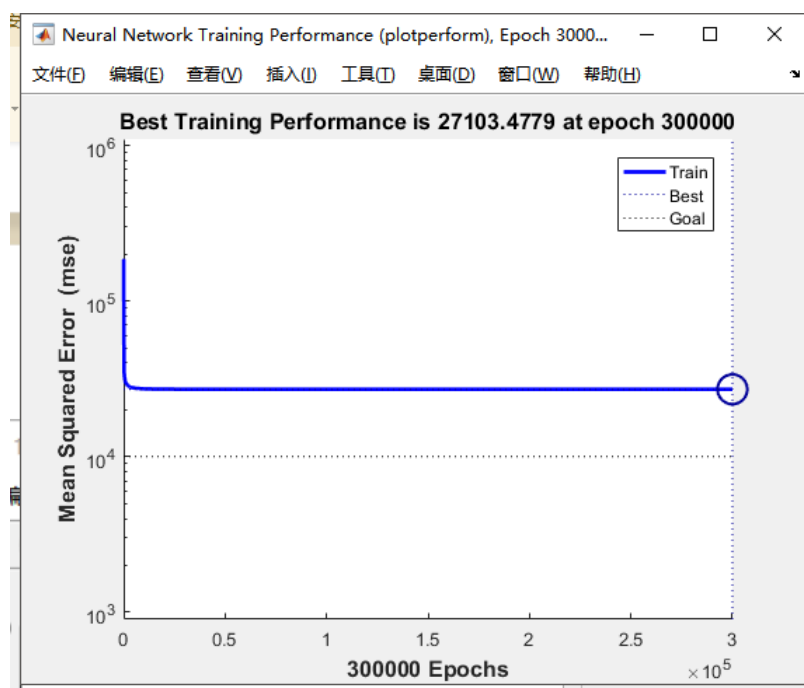


图 10: 训练数据的梯度和均方误差之间的关系图

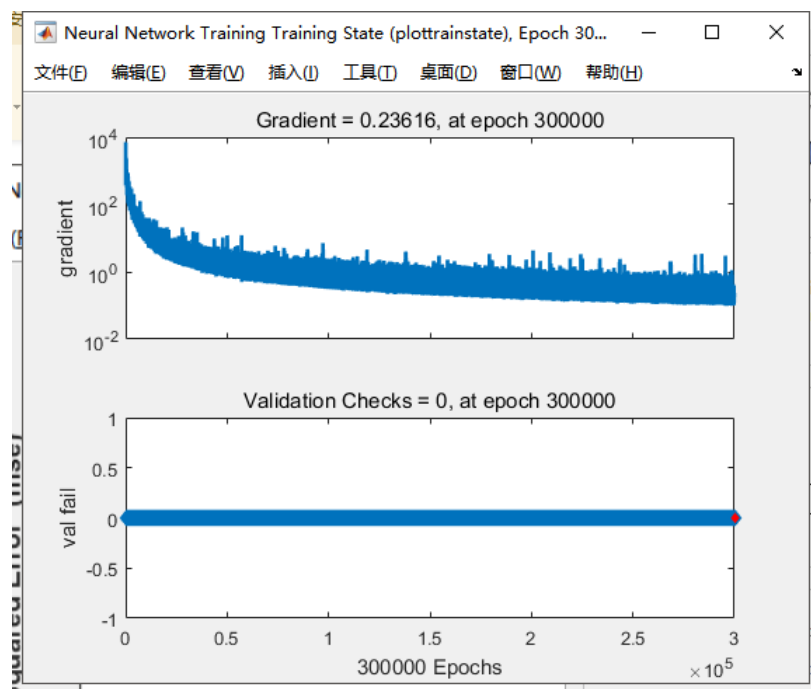


图 11: 学习梯度与学习次数图

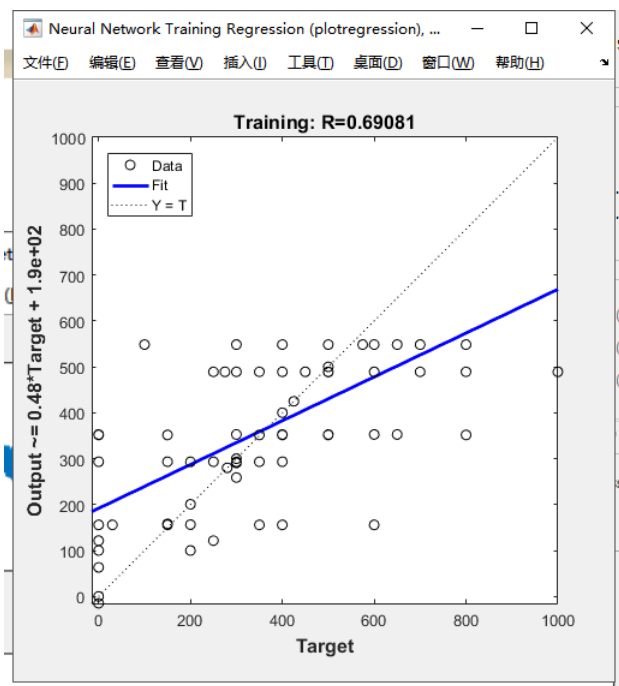


图 12: 残差的正态的检验图 (Q-Q图)

4.5 问题五的解答

问题五要求我们写一份面向大众，可以高效实行上述模型的方案。下面是我们以信的形式给出的方案：

亲爱的家长朋友与同学们：

你们好！

随着科学技术的发展和电子设备的普及，我国近视发生率逐年增加，在儿童与青少年阶段尤为明显。所以预防近视要从小做起，对于已经近视的同学来说，防止度数的加深刻不容缓。我们给出的可行方案根据权重排序如下：

第一点，在日常生活中**尽量减少电子设备的使用**。利用电子设备搜集信息十分方便，但我们花费在电子产品上的大多数时间都是为了娱乐。所以适当地减少这部分时间，并且每看电子设备半个小时都要向远处张望5分钟来缓解眼睛的疲劳。

第二点，在学习与看书时，**眼睛与书的距离大约为30cm**。每学习或读书半个小时都要向远处眺望5分钟。

第三点，**增强对眼睛的保护**。比如每天坚持做眼保健操、经常双手搓眼、经常转动眼球、多吃含有叶黄素和花青素的食物。

第四点，无论是看带电子设备还是书籍，都要**注意姿势端正、用眼卫生**。躺下来看书、在强光或者弱光环境下看书、在颠簸的汽车上看书等行为都会导致眼睛疲劳、视力下降，甚至造成散光与斜视。以上这几种行为都要尽量避免，注意用眼卫生。

第五点，**增加户外活动，减少近距离用眼时间**。参加户外运动，可以增加晒太阳的几率，给眼球补钙，从而预防近视；适当的户外活动还可以降低眼压，减少近视发生的机会；球类运动还可以增加眼球的弹性来预防近视。

第六点，遗传因素对于儿童及青少年近视来说有一定的影响，但是做好以上六点可以有效地防止近视或度数加深。

总的来说，预防近视或者防止近视度数加深需要**培养自己良好的自控能力**，在眼球发育期间注意培养良好的卫生习惯和建立合格的环境条件。建议家长与孩子对未来孩子是否近视以及近视的度数**做出预测**，并根据生活习惯**做出相应改变**来预防近视。

数学建模小组

2019年5月3日

5 模型的优缺点

5.1 模型的优点

- 运用EXCEL处理大量的数据，使得数据处理过程简洁高效，也使得最后的结果直接明了；
- 模型建立恰当，最后的图表中只出现了很少数的坏值；
- 运用MATLAB做成的交互界面指示清晰，可以很便捷地引导家长或学生输入自己的答案并很快地得出未来视力的预测结果；

5.2 模型的缺点

真实的数据难以全面地收集，大部分的数据来源于大学生，数据的缺乏导致我们无法得出最真实的结果，只能通过模拟计算出比较符合事实的结果。

6 模型的改进与推广

6.1 模型的改进

在视力的预测模型和预警模型中，我们都只着重考虑了近视下的情况，而没有过多讨论不近视情况。在模型的改进方面，我们可以将不近视考虑在内，调查更多的数据，结合评价和预测模型，综合得出不近视与近视的临界指标值。

在视力的影响因素中，我们可以引入更多的指标和数据，例如年龄因素，生活作息习惯等，将相应的数据量化求权重，使得评价模型更加准确。

除此之外，还可以将后天行为（例如眼部保护习惯，不良用眼习惯等）对近视情况的影响更加具体化，例如引入衰减系数和增加系数，通过微分方程模型模拟其对近视度数的缓减和加剧。

6.2 模型的推广

对于这种以将影响因素量化通过熵权法求权为基础的建模思路，我们不仅可以用在视力的预警和干预模型上，对于很多其他情景也同样适用。例如病人的康复过程模拟，景区人流量的变化模拟等。

参考文献

- [1] 林斌, 廖翔, 我国青少年近视影响因素与防治措施研究[J], 中国卫生标准管理, 2017年第20 期: P1-2, 2017。
- [2] 刘冲昊, 柳群义, 基于熵权法的中国铁矿石安全评价[J], 矿产保护与利用, 2018年10月第5 期: P86-93, 2018。
- [3] 慈铁军, 刘晓瑜, 基于决策者偏好的区间数多属性决策属性值规范化方法[J], 统计与决策, 2015年第3期: P36-38, 2015。
- [4] 李文静, 青少年近视程度与眼压的关系探讨[J], 中国实用乡村医生杂志, 2019年第2期: P24-25, 2019。

附录

1.问题二MATLAB源代码

```
x = xlsread('C : \Users\12086\Desktop\001.xlsx', 1, 'A1 : A103');
y1 = xlsread('C : \Users\12086\Desktop\001.xlsx', 1, 'B1 : B103');
X = [ones(103, 1), x]; %输入数据:
[b, bint, r, rint, statas] = regress(y1, X); %回归分析及检验:
rcoplot(r, rint) %残差分析: 作残差图
z1 = b(1) + b(2) * x;
plot(x, y1, 'bx', x, z1, 'r') % 预测及作图
```

2.问题三MATLAB源代码

```
function mutual
clc;
disp('请按实情输入您或您的孩子的相关情况，我们将为您预测您孩子成年后的视力水平。');
disp('请输入待测试者的直系三代中的近视比例: ');
disp('比例为: 近视人数/直系三代亲属总数。');
x1=input('比例为: ');
x2=input('请输入待测试者现在每天进行户外运动的平均时长（0至1小时按1小时计算，输入1即可，1至2小时按2小时计算，输入2即可，以此类推）: ');
x3=input('请输入待测试者现在每天使用电脑、手机等电子产品的平均时长（0至1小时按1小时计算，输入1即可，1至2小时按2小时计算，输入2即可，以此类推）: ');
x4=input('请输入待测试者现在平均每天看书、写作业的总时长（0至1小时按1小时计算，输入1即可，1至2小时按2小时计算，输入2即可，以此类推）: ');
while 1 x5=input('请输入待测试者当前是否存在近距离或不当姿势使用手机、电脑或看书写作业等的行为（如果有请输入1，无请输入0）: ');
switch x5
case 1
break;
case 0
```

```

break;
otherwise
disp('输入错误，请按照要求重新输入! ');
end
end
while 1
x6=input('请输入待测试者当前是否坚持每天做眼保健操或使用护眼产品（如果均
有请输入0，有一项请输入1，均没有请输入2）: ');
switch x6
case 2
break;
case 1
break;
case 0
break;
otherwise
disp('输入错误，请按照要求重新输入! ');
end
end
disp(' ');
disp('请确认数据:');
X = ['1.待测者的三代直系亲属的近视比例为', num2str(x1)];
disp(X)
X = ['2.待测者每天平均花费约', num2str(x2), ' 3$'];
disp(X)
X = ['3.待测者每天平均花费约', num2str(x3), ' 3>f ° °'];
disp(X)
X = ['4.待测者每天平均花费约', num2str(x4), ' 3Æ'];
disp(X)
if(x5 == 1)

```



```

disp('5.存在不合适的阅读或使用电子设备的习惯。')
else
disp('5.阅读与使用电子设备习惯良好。')
end
if(x6==0)
disp('6.认真做眼保健操并使用护眼产品。')
elseif x6 == 1
disp('6.认真做眼保健操或使用护眼产品。')
else
disp('6.没有做眼保健操或使用护眼产品的习惯')
end
disp('请确认数据是否正确,正在为您计算,请稍后……。')
pause(3);
xs = xlsread('C : \Users\12086\Desktop\001.xlsx',1,' A1 : A103');
ys = xlsread('C : \Users\12086\Desktop\001.xlsx',1,' B1 : B103');
XS = [ones(103,1),xs]; %输入数据:
[b, ,r,rint, ] = regress(ys,XS);%回归分析及检验:
rcoplot(r,rint) %残差分析: 作残差图
z1 = b(1) + b(2) * xs;
plot(xs,ys,'bx',xs,z1,'r')%预测及作图
disp(' ')
x1 = (x1 - 0)/1;
x2 = (5 - x2)/4;
x3 = (x3 - 1)/9;
x4 = (x4 - 1)/6;
x5 = x5/1;
x6 = x6/2;
score = 0.547351289*x1+0.043075302*x2+0.143779778*x3+0.162440807*
x4 + 0.048426796 * x5 + 0.054926028 * x6;
grade = score * b(2) + b(1);

```

```

if grade < 50
X='预计在18岁成年后视力度数为0-50度'; elseif grade < 100
X=' 预计在18岁成年后视力度数为100度';
elseif grade < 150
X=' 预计在18岁成年后视力度数为150度';
elseif grade < 200
X = ['预计在18岁成年后视力度数为200度'];
elseif grade < 250
X=' 预计在18岁成年后视力度数为250度';
elseif grade < 300
X=' 预计在18岁成年后视力度数为300度';
elseif grade < 350
X=' 预计在18岁成年后视力度数为350度';
elseif grade < 400
X=' 预计在18岁成年后视力度数为400度';
elseif grade < 450
X=' 预计在18岁成年后视力度数为450度';
elseif grade < 500
X=' 预计在18岁成年后视力度数为500度';
elseif grade < 550
X=' 预计在18岁成年后视力度数为550度';
elseif grade < 600
X=' 预计在18岁成年后视力度数为600度';
elseif grade < 650
X=' 预计在18岁成年后视力度数为650度';
elseif grade < 700
X=' 预计在18岁成年后视力度数为700度';
elseif grade < 750
X=' 预计在18岁成年后视力度数为750度';
elseif grade < 800

```

```

X = ' 预计在18岁成年后视力度数为800度';
elseif grade < 850
X = ' 预计在18岁成年后视力度数为850度';
elseif grade < 900
X = ' 预计在18岁成年后视力度数为900度';
elseif grade < 950
X = ' 预计在18岁成年后视力度数为950度';
elseif grade < 1000
X = ' 预计在18岁成年后视力度数为1000度';
elseif grade < 1050
X = ' 预计在18岁成年后视力度数为1050度';
elseif grade < 1100
X = ' 预计在18岁成年后视力度数为1100度';
elseif grade < 1150
X = ' 预计在18岁成年后视力度数为1150度';
elseif grade < 1200
X = ' 预计在18岁成年后视力度数为1200度';
else
X = ' 注意!! 预计在18岁成年后视力度数为1200度以上';
end
disp(X)
if grade > 400
disp('当前的习惯对眼睛较为不好, 请尽快予以矫正, \n建议减少电子产品的使用时间,
阅读与写作业时多进行适当的放松, 多进行一些户外运动')
end

```

3. 问题四MATLAB源代码

```

clear;
clc;
%导入数据P = xlsread('C:\Users\12086\Desktop\001.xlsx', 2, 'A1 : G103');
% 输入数据集信息 (为减小篇幅, 此处改成从xls导入)

```

```

T = xlsread('C : \Users\12086\Desktop\001.xlsx', 2, 'I1 : I103');
% （在实际中使用的是直接输入的数据集，以提升神经网络的运算速度）
P=P';
T=T';
%数据的归一化处理
[p1,ps]=mapminmax(P);
[t1,ts]=mapminmax(T);
%选取训练集和测试集
[trainsample.P, valsample.P, testsample.P] =dividerand(P,0.7,0.15,0.15);
%取70%为训练集，15% 为测试集
[trainsample.T, valsample.T, testsample.T] =dividerand(T,0.7,0.15,0.15);
net = newff(minmax(P),[6,1], 'logsig', 'purelin', 'trainrp'); % 创建网络
%设置网络参数
net.trainParam.show=500;
net.trainParam.lr=0.01;
net.trainParam.epochs=300000;
net.trainParam.goal=1e4;
[net,tr] = train(net,P,T);
W1 = net.iw1,1 %输入层到中间层的权值
B1 = net.b1 %中间各层神经元的阈值
W2 = net.lw2,1 %中间层到输出层的权值
B2 = net.b2 %输出层各神经元的阈值
X = sim(net,P)
plot(X,'b')
hold on
plot(T,'*')
%计算仿真，其一般用sim函数
[normtrainoutput,trainPerf]=sim(net,trainsample.P,[],[],trainsample.T);
% 训练的数据，根据BP得到的结果
[normvalidateoutput,validatePerf]=sim(net, valsample.P,[],[], valsample.T);

```

```

% 验证的数据，经BP得到的结果
[normtestoutput,testPerf]=sim(net,testsample.P,[],[],testsample.T);
% 测试数据，经BP得到的结果
%将所得的结果进行反归一化，得到其拟合的数据
trainoutput=mapminmax('reverse',normtrainoutput,ts);
validateoutput=mapminmax('reverse',normvalidateoutput,ts);
testoutput=mapminmax('reverse',normtestoutput,ts);
%正常输入的数据的反归一化的处理，得到其正式值
trainvalue=mapminmax('reverse',trainsample.T,ts);
%正常的验证数据validatevalue=mapminmax('reverse',valsample.T,ts);
% 正常的验证的数据testvalue=mapminmax('reverse',testsample.T,ts);
%正常的测试数据
%做预测，输入要预测的数据pnew
pnew = [1/2, 2, 3, 4, 1, 1]';
%此处可以随时更改
%第一个数据为三代近亲近视比例;
%第二个数据为每天户外运动时长
%第三个数据为平均每天使用电子产品时间
%第四个数据为平均每天阅读或看书的时间
%第五个数据为是否有不正确的用眼习惯
%第六个数据为是否有使用护眼产品或者做眼部保养
pnewn = mapminmax(pnew);
anewn = sim(net,pnewn);
anew = mapminmax('reverse',anewn,ts); %绝对误差的计算
errors = trainvalue - trainoutput;
%plotregression拟合图
figure,plotregression(trainvalue,trainoutput) %误差图
figure,plot(1:length(errors),errors,'-b')
title('误差变化图')

```

4.原始数据及计算出的近视程度

表 4: 附录数据表

5、直系亲属近视率	6、日均户外运动时间	7、日均电子产品使用时间	8、日均看书学习时间	9、有无不合理用眼习惯	10、是否做眼部保养	近视程度 $f(i)$
0.33	1	7	2	1	2	0.449980705
0.5	2	3	3	1	0	0.440506914
0	2	4	5	1	1	0.26441675
0	1	2	3	1	1	0.189087579
0.25	1	6	1	1	2	0.363143603
0	2	2	3	1	1	0.178318753
0	1	6	2	1	2	0.253379249
0.33	1	6	1	1	1	0.379468692
0	3	4	4	1	2	0.254037471
0	1	3	2	0	0	0.102099831
0	2	2	2	1	2	0.1787083
0	3	5	4	1	2	0.270013002
0	2	2	3	1	1	0.178318753
0.33	1	5	2	1	2	0.418029643
0	2	3	5	1	1	0.24844122
0	2	6	2	1	2	0.242610423
0	3	8	2	1	2	0.263792659
0	1	4	3	1	1	0.22103864
0.25	2	5	2	1	2	0.363472715
0.5	3	5	3	1	1	0.489152165
0	3	6	5	1	1	0.285598987
0.111	2	10	2	1	2	0.36726854

0.75	1	10	2	1	2	0.727794839
0	3	5	2	1	2	0.215866067
0	2	7	2	1	2	0.258585954
0	3	5	7	1	2	0.351233406
0	2	1	7	1	2	0.298100108
0.25	2	1	2	1	2	0.299570591
0.4	3	2	3	0	1	0.338063647
0.25	2	2	4	1	1	0.342230043
0	2	6	3	1	2	0.269683891
0	1	1	1	1	1	0.118965112
0.25	2	10	3	1	0	0.415497808
0	1	4	2	1	2	0.221428187
0	2	2	2	1	1	0.151245285
0	2	4	4	1	2	0.264806297
0	1	10	1	1	2	0.290207904
0	5	5	3	1	2	0.221401883
1	3	2	4	1	1	0.741974685
0.25	1	3	2	1	2	0.342290478
0	2	10	3	1	1	0.306123
0.167	1	5	4	1	2	0.382958319
0	2	5	4	1	2	0.280781828
0	5	10	5	1	2	0.355426473
0.4	2	3	1	1	2	0.386550878
0	1	10	1	1	2	0.290207904
0.25	2	5	3	1	0	0.335620154
0	3	9	4	1	2	0.333915126
0.333	2	7	4	1	0	0.44007484
0	4	2	4	0	2	0.162890788
0	5	4	3	1	2	0.205426353
0	2	7	6	1	2	0.366879825
0	1	3	5	1	1	0.259210045
0.5	5	3	3	0	1	0.387236656
0.25	2	2	2	1	2	0.315546122
0	2	10	1	1	2	0.279439079
0	1	4	2	1	2	0.221428187
0	2	5	2	1	2	0.226634892

0.2	2	3	4	1	1	0.33083801
0.5	2	5	2	1	1	0.472847523
0	1	2	2	1	2	0.189477125
0	2	2	3	1	1	0.178318753
0	4	3	3	1	2	0.200219647
0	2	9	2	1	2	0.290537016
0	3	7	3	1	2	0.274890596
0	2	2	5	1	1	0.232465689
0.167	2	2	4	1	2	0.3242629
0.5	2	2	2	1	2	0.452383944
0	2	2	3	1	2	0.205781767
0	2	2	2	1	2	0.1787083
0.5	4	2	3	1	2	0.457919761
0	2	6	2	1	2	0.242610423
0.25	2	4	2	1	2	0.347497184
0	2	5	2	1	1	0.199171878
0.5	2	5	5	1	2	0.58153094
0.25	1	10	1	1	2	0.427045727
0.333	2	3	3	1	1	0.376562263
0	3	8	2	1	2	0.263792659
0.25	2	5	6	1	2	0.471766586
0.25	1	7	1	1	2	0.379119134
0.333	1	7	2	1	1	0.424159745
0.4	2	8	1	1	2	0.466428533
0	3	5	3	1	2	0.242939534
0.25	2	6	2	1	2	0.379448245
0.5	3	5	3	1	2	0.516615179
0	1	5	2	0	2	0.188976922
0	2	2	2	1	2	0.1787083
0	2	3	5	1	2	0.275904234
0.25	1	1	7	1	1	0.418243741
0	1	3	7	1	1	0.313356981
0.25	2	5	5	1	1	0.417230104
0.5	2	3	5	0	1	0.473690068
0	2	3	2	1	2	0.194683831
0.25	2	2	2	1	2	0.315546122

0	3	7	5	1	1	0.301574518
0	2	8	4	0	2	0.280281624
0.25	3	7	4	1	1	0.411338872
0.5	2	3	2	0	2	0.419932679
0.4	2	5	5	1	2	0.526795811
0.25	1	3	5	1	2	0.423510882
0.5	1	3	4	1	2	0.533275236
1	2	4	3	1	1	0.757621104
0	1	3	6	1	2	0.313746527