

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_____19B314_____

报名时所属学院（请填写完整的全名）：_____计算机科学与技术学院_____

参赛队员姓名与学号（打印，用二号字，并签名）：

1._____沈子卓_____17160110002_____

2._____孙燮阳_____17130120193_____

3._____孙博远_____17130120188_____

日期：_____2019_____年_____5_____月_____3_____日

西安电子科技大学 2019 年大学生数学建模校内赛

评 阅 专 用 页

	评阅人 1	评阅人 2	评阅人 3	总评
成绩				

智能化多因素动态近视预警模型

摘要

近年来,我国青少年的近视率逐年提高并向低龄化发展,引起了社会和政府的极大关注。近视率的逐年提高严重影响青少年的身体健康,甚至影响到了我国的征兵工作。本文针对不同年龄段学生的视力情况,综合分析影响视力的关键因素、不同影响因素权重分配、近视学生比例等因素,建立了四个模型,对学生当前的视力状况进行判断、预测和个性化预警,并提出合理建议。

模型 I 利用类层次分析原理,建立了得分制量化模型。将影响视力的 12 种关键因素量化为 12 种指标,并配以相应的分值和权重以衡量不同指标对学生视力的影响程度。提出“综合得分”概念来衡量学生的综合视力情况。

模型 II 利用概率分布原理、微分方程原理和统计特征,建立了视力演化机理模型。构建了两个基本函数,线性化的近视发生概率-综合得分函数($P - G_i$)和符合正态分布的学生人数-综合得分函数($M - G_i$)。根据实际规律,提出三种情况下的 $P - G_i$ 函数的简化方式,建立三组微分方程解出标准临界综合得分 G_{is} 、危险临界综合得分 G'_{is} 和安全综合得分 G''_{is} 。从而将综合得分范围划分为四个区域:极度危险、危险、相对安全、安全。

模型 III 利用模型 I、II 得到的结论,建立了面向大众的智能化动态 AI 预警机制。考虑到用户更加关心个体的近视发生率的预测和具体调整建议,利用指标权重和三个临界综合得分确定学生的近视风险等级,根据其相对偏离标准分数的指标,智能化的提出个性化建议和视力提升方案。

模型 IV 针对模型 I 中各项指标权重进行计算,并设计了一种改进 BP 神经网络算法。引入神经网络训练出 12 个指标的最优权重分配。设计了可供学习的信息系统;确定了神经网络的基本结构。在传统神经网络的基础上加以改进,创新性地将最后一层隐含层暴露作为额外输出层,使权重分配向量能够显式的被提取,同时使得该模型在理论上能够以更快的速度收敛。

最后基于上述模型,我们向政府和学校提出了一套应用推广方案。针对原始数据采集、基础模型训练、试点测试、全面推广等 4 个阶段分别提出可执行方案,该方案具有较强的可执行性和可操作性。

本文提出的模型具有灵活性强、可执行性强、灵敏度高、可推广移植性强、模型动态变化的特点,理论上可针对具有多因素影响、统计规律近似正态分布特点的各类问题进行拓展延伸。

关键字: 得分制 正态分布 概率统计 微分方程 AI 反馈机制 神经网络

一、问题重述

近年来，我国青少年的近视率逐年提升且低龄化趋势显著，大学生近视比例高达87.7%，引起社会 and 政府的极大关注。近视率的逐年提高不仅严重影响了青少年的身体成长，还造成了我国“征兵难”的局面。因此，如何降低青少年近视患病率、如何预防青少年近视、如何及时预警较高近视患病风险的学生的家长和老师，及时纠正孩子们的错误用眼习惯成为急需解决的难题。为解决这个问题，我们需要建立合理有效的数学模型对青少年的视力状况进行评估，预测青少年的近视风险，并对学生家长和老师提出合理化建议。

待解决的问题如下：

- 确定影响视力的关键因素并建立量化模型，给出需要获取的数据以及获取数据的可能途径。
- 设计一个具有较高可执行性的眼睛视力的演化机理模型。
- 设计一种面向学生父母和老师的 AI 预警机制，针对学生的视力情况给出对应的个性化预警建议。
- 设计一个可供机器学习的信息系统，建立针对眼睛视力的机器学习模型。
- 为实现和推广模型的应用所需的数据收集和大范围推广制定一份方案。

二、模型假设

- 假设问题的研究对象为 0 ~ 22 岁儿童及学生。
- 仅考虑正常因素引起的近视，不考虑突发事件和突发重大疾病等极端情况引发的近视。
- 假设每个视力影响因素对视力的影响相互独立。
- 调查对象均为正常个体，不考虑由于少数特殊人员引起的差异。
- 假设调研结果数据真实并具有一定的参考价值。

三、符号说明

在这里，我们列出了本文中使用的符号和标记，如表 1 所示。部分变量和符号会在后面的章节中给出具体定义。

符号	含义
$\Omega = \{A, B\}$	总体样本空间，包含事件 A 与事件 B
$P(x)$	事件 x 发生的概率
A	$A = \{S_i \mid i = 0, 1, 2, 3, 4, 5\}$ ，表示大学毕业 (22 岁) 前已近视
B	表示大学毕业 (22 岁) 前未近视
S_i	当前学生在第 i 个时期发生近视
a_n	第 n 种影响指标
Q_{in}	在第 i 个时期，第 n 种指标的得分， $0 \leq Q_{in} \leq 10$
q_{in}	第 i 个时期，第 n 种因素对近视情况影响的权重， $0 < q < 1$
G_i	当前学生在第 i 个时期的近视综合得分
G_{is}	第 i 个时期的标准临界综合得分
G'_{is}	第 i 个时期的危险临界综合得分
G''_{is}	第 i 个时期的安全临界综合得分

表 1 符号说明

四、量化模型及数据来源

4.1 问题分析

首先我们需要确定影响学生视力的关键因素。近年来，有诸多有关近视成因与分析的研究，我们通过收集、查阅相关研究资料和论文，总结出 12 种影响视力相关因素。基于每种因素对于学生视力的影响程度的不同，我们有针对性地建立了量化模型。

为衡量每种因素对于学生视力的影响程度，我们提出“得分制模型”，即对每种因素赋相同的分数额度，通过调整它们的加权系数，反映各种因素对学生视力的影响程度。

对于得分值模型的系数配比和量化分析，我们需要大量的数据作为支撑，使结论更加真实、可靠。除了通过教育部统计、卫生部门调研等传统手段，在信息化时代获取信息的途径广泛，可以通过网络调查、电子产品的“家长模式”，甚至可穿戴电子设备等获取大量真实可靠的数据信息。

4.2 前提声明

为了更好的研究不同指标对不同时期的学生近视情况的影响,我们首先将并将学生的成长阶段分为 5 个时期进行考虑,对于每一个时期,我们总结了 12 个不同的考量指标,用于评估当前年龄学生的近视情况。

学生成长阶段的 5 个时期如表 2 所示:

符号	含义
S_0	先天性近视
S_1	上小学前 (0 ~ 6 岁) 近视
S_2	上小学后到上初中前 (6 ~ 12 岁) 近视
S_3	上初中后到上高中前 (12 ~ 15 岁) 近视
S_4	上高中后到上大学前 (15 ~ 18 岁) 近视
S_5	大学期间 (18 ~ 22 岁) 近视

表 2 学生成长阶段

12 种影响近视指标如表 3 所示:

符号	含义	符号	含义
a_1	读书时眼睛与书本的距离	a_7	饮食习惯
a_2	握笔姿势	a_8	做眼保健操、眼球操情况
a_3	每日读书或写字的平均时间	a_9	用眼卫生知识情况
a_4	每日电子产品 (手机、电脑、电视等) 平均使用时间	a_{10}	每天户外运动时间
a_5	学校课桌椅的高矮及舒适度	a_{11}	用眼时环境光线情况
a_6	每日平均睡眠时间	a_{12}	心理因素与身体状态

表 3 影响近视的指标

4.3 得分制模型

由假设和已知前提可知, 事件 A(大学毕业 (22 岁) 前已近视) 与事件 B(大学毕业 (22 岁) 前未近视) 共同构成样本空间, 即事件 A 发生的概率与事件 B 发生的概率和为 1:

$$P(A) + P(B) = 1 \quad (1)$$

其中, 由于 $A = \{S_i \mid i = 0, 1, 2, 3, 4, 5\}$, 即 A 事件由 $S_0 \sim S_5$ 事件构成, 即 A 发生的概率为由 $S_0 \sim S_5$ 事件的概率之和, 即:

$$P(A) = \sum_{i=0}^5 P(S_i) \quad (2)$$

特别的, $P(S_0)$ 为先天性近视的概率, 与后天因素无关, 为一个由统计学规律确定的定值。下面讨论各个因素对不同时期学生近视发生率的影响:

引入得分制的方法, 对于每一种可能影响近视发生率的指标, 我们规定每种指标最高得分为 10, 用 Q_{in} 表示第 i 个时期, 第 n 种因素的得分, 且 Q_{in} 满足 $0 \leq Q_{in} \leq 10$ 。

接下来考虑第 i 个时期第 n 种因素对近视发生率影响的权重 q_{in} , q_{in} 满足 $0 < q_{in} < 1$, 且所有权重的加和为 1。

$$\sum_{n=1}^{12} q_{in} = 1, (i = 1, 2, 3, 4, 5) \quad (3)$$

由上述既得公式, 我们可以计算每一个学生在第 i 个时期近视发生的综合得分 G_i :

$$G_i = \sum_{n=1}^{12} q_{in} \times Q_{in} \quad (4)$$

因此, 我们需要确定一个视力的演化机理模型和一系列临界综合得分, 并利用这一系列临界综合得分的值划分学生的得分区域, 从而确定当前学生在当前阶段发生近视的概率, 进而保证该量化模型的可执行性和准确性。

4.4 信息来源

按教育部规定, 学校每年会对学生定期进行体检和视力检查。因此通过对比每年的检查结果统计获得不同时期的学生的基本身体素质情况和全体学生的近视发生率。对于我们所需要的 12 项指标数据, 可以通过线下、线上问卷调查的形式进行数据采集。但调查范围不局限于学生, 我们更倾向于面向家长和老师进行较为正式严肃的调查, 以便得出更加真实、可靠的数据。

针对上述可能存在的影响近视发生的指标, 我们给出如下可能的较为可靠的具体信息来源:

- 通过各教育部门 (学校、教育局等) 获取学生的统计数据。
- 通过各卫生部门 (各级医院、卫生所等) 获取相关年龄对象的体检数据。

- 不定期抽查中小学的护眼操落实情况，获得相关统计数据。
- 通过相关的运动 app 获取对象的户外运动数据。
- 通过现有各类电子产品的“家长模式”获取对象每日使用电子产品的时长数据，统计得出学生针对电子产品的使用情况。
- 通过面向相关对象进行周期性的问卷调查获取相关数据（如每月对各年龄段学生进班进行调研）。
- 通过相关对象的可穿戴设备获取相关数据（如运动手环可记录每日睡眠时长）。

五、视力演化机理模型

5.1 问题分析

基于问题一的研究, 我们确定了 12 种影响学生视力的关键因素, 建立了得分制模型。据统计, 2018 年我国在校生人数高达 2.76 亿, 全国共有义务教育阶段学校 21.38 万所。针对我国庞大的学生数量, 我们采用概率论与数理统计的思想, 将不同地区、不同年龄段和不同等级学校的学生进行分类研究, 建立研究适合某一特征群体的演化机理模型。

我们首先建立学生的近视发生概率-综合得分函数 $(P - G_i)$ 和学生人数-综合得分函数 $(M - G_i)$ 。在两个函数的基础上进行推广, 建立一系列微分方程, 分别求解确定第 i 个时期的标准临界综合得分 G_{is} 、危险临界综合得分 G'_{is} 和安全综合得分 G''_{is} 。依据一系列的临界综合得分划分危险等级区, 预测具体的衡量学生近视发生的风险。

5.2 近视发生概率-综合得分函数 $(P - G_i)$

根据得分制模型, 通常情况下, 学生近视发生概率会随综合得分的升高而降低。为了简化分析, 将该函数做线性化处理。

$$P(G_i) = -k \times (G_i - G_{imax}) \quad (5)$$

上式中 G_i 满足 $G_{imin} \leq G_i \leq G_{imax}$ 。其中 k 为近视风险变化率, 表示随综合得分变化时, 学生患近视的概率变化情况, 该值由 G_{imax} (最高综合得分) 和 G_{imin} (最低综合得分) 共同确定, 如下式所示:

$$k = \frac{1}{G_{imax} - G_{imin}} \quad (6)$$

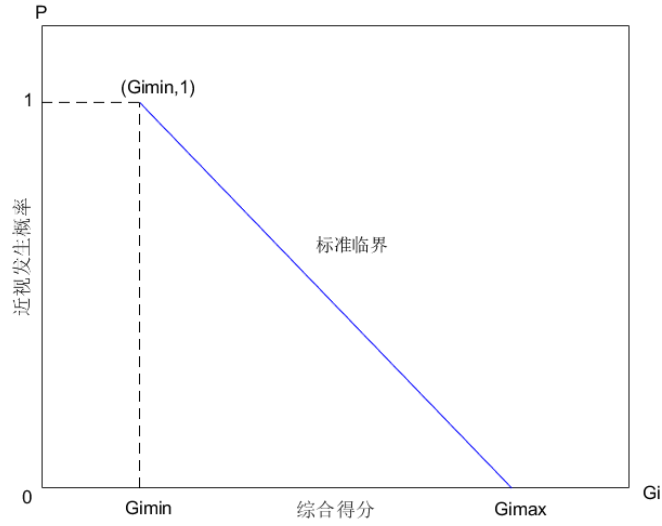


图 1 近视发生概率-综合得分函数 ($P - G_i$)

5.3 学生人数-综合得分函数

在现实统计中，学生人数-综合得分函数 ($M - G_i$) 近似服从正态分布：

$$M(G_i) = \frac{1}{\sqrt{2\pi} \times \sigma} \times e^{-\frac{(G_i - \mu)^2}{2 \times \sigma^2}} \quad (7)$$

其中 G_i 满足条件 $G_{imin} \leq G_i \leq G_{imax}$

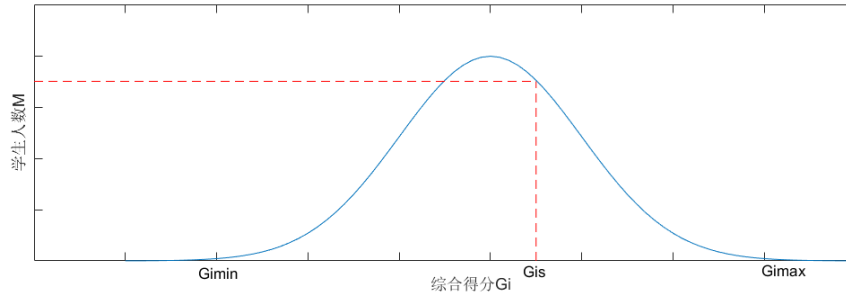


图 2 学生人数-综合得分函数 ($M - G_i$)

5.4 三个临界得分的确定

5.4.1 确定标准临界综合得分 G_{is}

在实际情况中，得分较高的学生仍有发生近视的概率。记在实际调查中所获得的 i 时期的总学生人数为 M_s ，该时期新增的患近视比例为 $m\%$ ，则存在一个标准临界综合得分 G_{is} 应使得下式成立：

$$\int_{G_{imin}}^{G_{is}} M(G_i) \times P(G_i) dG_i = M_s \times m\% \quad (8)$$

即在已知权重参数的条件下，通过近视概率-综合得分函数 ($P - G_i$) 和学生人数-综合得分函数 ($M - G_i$) 可以确定 i 时期内近视发生的总人数，当其与实际调查统计得到的学生总数中该时期的实际近视发生人数相同时可以确定最佳的标准临界综合得分。

5.4.2 确定危险临界综合得分 G'_{is}

一般情况下，当学生的综合得分低于某个分值时，近视发生的概率可视为 1，即一定会发生近视，因此，针对近视概率-综合得分函数 ($P - G_i$)，引入危险临界综合得分 G'_{is} 的概念，则近视发生概率-综合得分函数 ($P - G_i$) 可演变为如下形式：

$$P(G_i) = \begin{cases} 1 & G_{imin} \leq G_i < G'_{is} \\ -k' \times (G_i - G_{imax}) & G'_{is} \leq G_i \leq G_{imax} \end{cases} \quad (9)$$

上式中 k' 为当前条件下的近视风险变化率，由下式决定：

$$k' = \frac{1}{G_{imax} - G'_{is}} \quad (10)$$

图像如下所示：

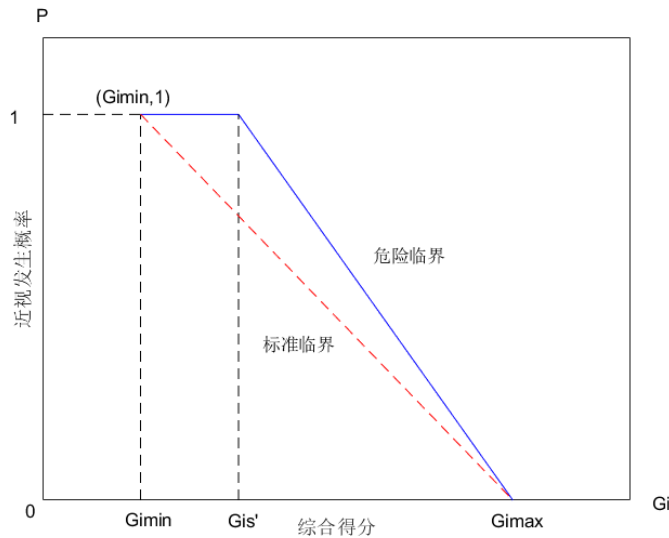


图 3 危险临界综合得分的近视发生概率-综合得分函数

与标准临界综合得分的确定方式类似，我们利用上述函数和近似服从正态分布的学生人数-综合得分函数 ($M - G_i$) 确定 i 时期内近视不发生的总人数，当其与实际调查统计得到的学生总数中该时期的实际近视不发生人数相同时可以确定最佳的危险临界综

合得分，即如下式所示：

$$\int_{G'_{is}}^{G_{imax}} (1 - P(G_i)) \times M(G_i) dG_i = M_s \times (1 - m\%) \quad (11)$$

求解上述微分方程可以就得到最佳的危险临界综合得分 G'_{is}

5.4.3 确定安全临界综合得分 G''_{is}

在一般情况下，当学生的综合得分超过某个分值时，近视发生的概率可视为 0，即认为几乎不会发生近视，因此，针对近视概率-综合得分函数 $(P - G_i)$ ，引入安全临界综合得分 G''_{is} 的概念，则近视发生概率-综合得分函数 $(P - G_i)$ 可演变成如下形式：

$$P(G_i) = \begin{cases} -k'' \times (G_i - G''_{is}) & G_{imin} \leq G_i < G''_{is} \\ 0 & G''_{is} \leq G_i \leq G_{imax} \end{cases} \quad (12)$$

上式中 k' 为当前条件下的风险变化率，由下式决定：

$$k'' = \frac{1}{G''_{is} - G_{imin}} \quad (13)$$

图像如下所示：

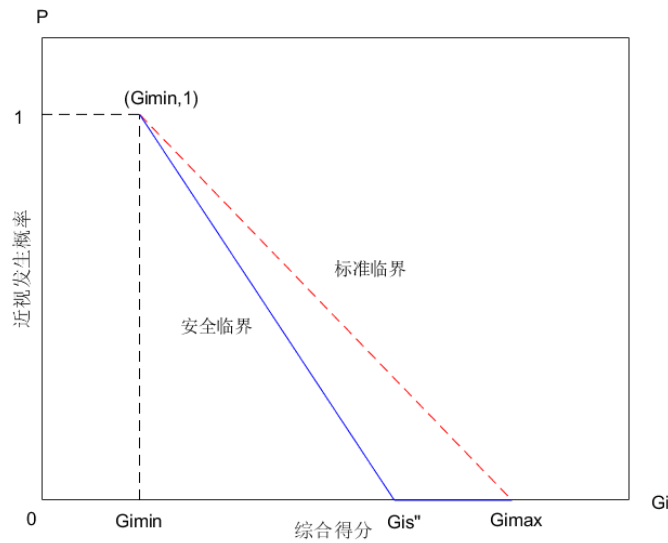


图 4 安全临界综合得分的近视发生概率-综合得分函数

与上述两种临界综合得分的确定方式类似，在已知合理权重的条件下，我们利用上述函数和近似服从正态分布的学生人数-综合得分函数 $(M - G_i)$ 确定 i 时期内近视发生的总人数，当其与实际调查统计得到的学生总数中该时期的实际近视发生人数相同时可以确定最佳的安全临界综合得分，即如下式所示：

$$\int_{G_{imin}}^{G''_{is}} P(G_i) \times M(G_i) dG_i = M_s \times m\% \quad (14)$$

求解上述微分方程可以就得到最佳的安全临界综合得分 G''_{is}

5.5 得出结论

经过上述过程，在已知权重合理的条件下可以确定第 i 个时期的标准临界综合得分 G_{is} 、危险临界综合得分 G'_{is} 和安全临界综合得分 G''_{is} ，这三个参数满足如下不等关系

$$G_{imin} < G'_{is} < G_{is} < G''_{is} < G_{imax} \quad (15)$$

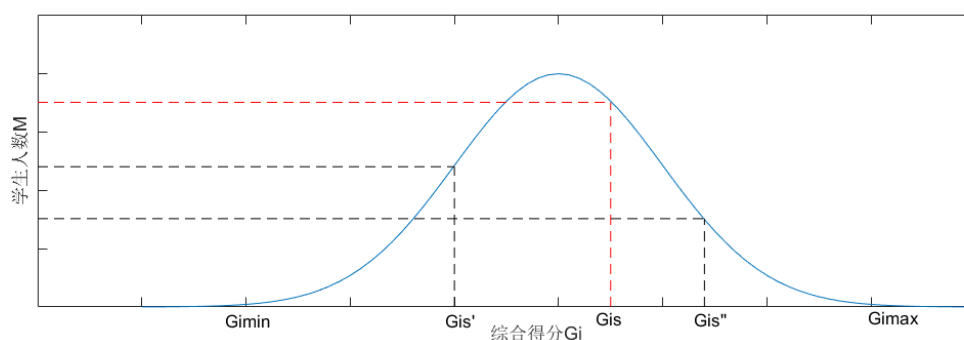


图 5 学生人数-综合得分函数 ($M - G_i$)

依照上述不等关系把得分区间划分为极度危险、危险、相对安全、安全等 4 个等级，因此对于得分为 G_i 的学生，有如下结论：

- 当 $G_{imin} \leq G_i < G'_{is}$ 时，说明该学生得分处于极度危险区域，近视发生的可能性极大，接近 100%
- 当 $G'_{is} \leq G_i < G_{is}$ 时，说明该学生得分处于危险区域，近视发生的可能性较大，容易发生近视。
- 当 $G_{is} \leq G_i < G''_{is}$ 时，说明该学生得分处于相对安全区域，近视发生的可能性较低，不易发生近视。
- 当 $G''_{is} \leq G_i \leq G_{imax}$ 时，说明该学生得分处于安全区域。近视发生的可能性极小，接近 0%

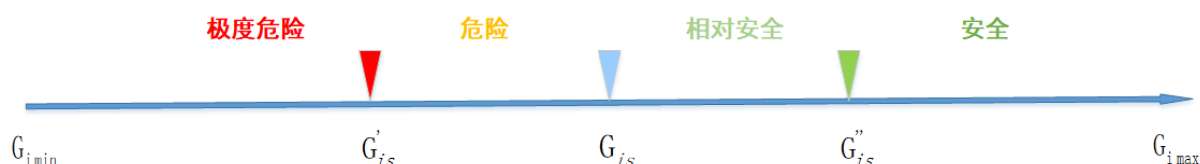


图 6 近视风险区域划分

六、智能化 AI 预警机制

6.1 问题分析

在得分制模型和视力演化模型的基础上，我们已经可以初步分析学生的视力情况并给出预测分级的结构。基于在视力演化模型中，我们将近视风险划分为四个区域：极度危险、危险、相对安全和安全。为此我们建立了一个面向大众使用的 AI 预警机制。

在构建 AI 预警模型的过程中，引入神经网络的方法，根据学生提供的现有准确的测试数据，通过神经网络模型训练各项指标的权重，并利用权重计算其视力综合得分，锁定分数所在近视风险等级。在确定学生视力等级后，根据用户相对偏离合适范围的指标，智能化给出合理化建议和视力提升方案。此预警模型不仅对未近视学生、老师和家长有预警作用，同时可以对已近视学生的近视发展情况做出反馈和预测。

6.2 AI 预警机制设计

考虑到绝大多数参与调查和使用该预警机制及其衍生产品的用户对相关专业知识了解不深，因此我们面向绝大多数对专业知识不了解的用户设计了一个基于简单的选择问卷的形式，将用户的调查问卷结果转换成相应输入分数，利用现有的神经网络模型对当前学生的近视发生率进行预测，并针对其主要影响因素进行智能化预警的 AI 预警系统。

另外，对于每一份调查问卷和实际调研的结果，除参与评测外，仍具有作为该模型的训练集的价值。可以利用这种方式不断进行模型的优化，即每一份调查问卷既可以为该用户进行近视预测和智能化预警，同时将该用户提供的数据添加进现有模型的训练集，使得模型逐渐趋于稳定和准确，从而实现一个自适应自优化的 AI 预警机制神经网络学习系统。

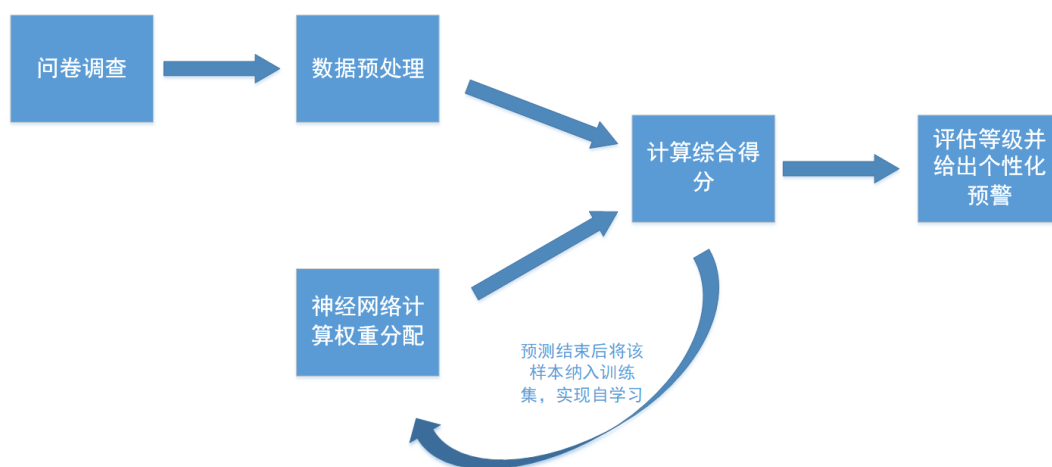


图 7 自学习 AI 预警机制流程

6.3 面向大众的选择调查问卷设计

下表是参考相关资料和统计数据得知的针对不同影响近视发生指标的等级对应表:

指标	$a_1(\text{cm})$	$a_2(\text{cm})$	$a_3(\text{h})$	$a_4(\text{h})$	a_5	$a_6(\text{h})$
A	≤ 25	≥ 3.3	≤ 3	≤ 1.5	合适	≥ 8
B	$20 \sim 25$	$2.5 \sim 3.3$	$3 \sim 4$	$1.5 \sim 2.5$	偏高/偏低	$7 \sim 8$
C	$15 \sim 20$	$1.8 \sim 2.5$	$4 \sim 5$	$2.5 \sim 3.5$	不合适	$6 \sim 7$
D	$10 \sim 15$	$1 \sim 1.8$	$5 \sim 6$	$3.5 \sim 4.5$		$5 \sim 6$
E	$5 \sim 10$	< 1	$6 \sim 7$	$4.5 \sim 5.5$		$4 \sim 5$
F	$0 \sim 5$		> 7	> 5.5		< 4

指标	a_7	$a_8(\text{次})$	a_9	$a_{10}(\text{h})$	a_{11}	$a_{12}(\text{h})$
A	优	$6 \sim 7$	熟练掌握	≥ 1.5	适宜	优秀
B	良	$4 \sim 5$	非常了解	$1 \sim 1.5$	较亮/较暗	良好
C	差		基本了解	$0.5 \sim 1$		
D	极差	$2 \sim 3$	经常接触	$0.25 \sim 0.5$		
E		1	偶尔接触	$0 \sim 0.25$	亮/暗	合格
F		0	从未接触	0	极良/极暗	不合格

表中等级 A,B,C,D,E,F 分别代表该学生该项得分 10,8,6,4,2,0 分。

表 4 等级评分标准

以上表内指标为内容设计选择题形式的调查问卷，通过用户反馈的各指标等级数据，按照上表转换成对应分数，并通过现有预警机制求得用户的综合得分和等级评定，再将预警机制的反馈结果呈现给用户。整个预警机制的使用流程对用户而言十分简便，适应性和可操作性极强，其低用户成本的特性使得该预警机制可以大规模推广。

6.4 AI 预警系统基本工作原理

通过以上图表以及问题二中第 i 个时期的标准临界综合得分 G_{is} 、危险临界综合得分 G'_{is} 和安全综合得分 G''_{is} 的计算，我们将预测对象分为四个等级：极度危险，危险，相

对安全，安全。

在给出评级后，系统会根据用户的选择调查问卷数据，选取其中相对偏离合适范围的数据对用户生成智能化、个性化的预警，告知用户应该在哪些方面多加注意以防加大近视的可能性。

6.5 预警反馈内容设计

- 对于被判定为极度危险的对象，预警机制会生成针对该用户的所有指标的建议，并表示希望其家长、老师或学生应持续关注对象的日常学习、生活中的用眼习惯及卫生情况，尽早干预以防加重近视的可能性。
- 对于被判定各个等级的对象，预警机制会生成针对该用户相对偏离合适范围的指标的建议，并表示其家长、老师或学生应多关注对象的日常学习、生活中的用眼习惯及卫生情况，有效预防近视发生或防止近视加重。
- 对于被判定为相对安全的对象，预警机制会生成针对该用户相对偏离合适范围的指标的建议，加强对象对用眼习惯和用眼卫生的意识，促使其改正不合适的学习、生活方式，预防对象变成近视。
- 对于被判定为安全的对象，预警机制会生成针对该用户相对偏离合适范围的指标的建议，并告知用户应保持良好的学习、生活习惯，注意纠正不良习惯。如若被预测对象的各项指标均在合理范围内，则告知对象应继续保持当前状态。

七、信息系统设计与机器学习模型

7.1 设计信息系统

根据选取的量化后的指标以及演化机理模型，我们选择使用 BP 神经网络设计了一种自学习、自适应的模型。由于模型采用的 12 种影响近视的指标已经实现量化，因此我们依据量化指标和学习目标设计了可供学习的信息系统，如下表所示：

特征属性标签	数据类型	特征属性标签	数据类型
a_{i1}	int	a_{i7}	int
a_{i2}	int	a_{i8}	int
a_3	int	a_{i9}	int
a_{i4}	int	a_{i10}	int
a_{i5}	int	a_{i11}	int
a_{i6}	int	a_{i12}	int

表中 $a_{i1} \sim a_{i12}$ 的取值范围是 $[1, 10]$

对于需要预测的标签，对应以下两种数据形式：

对于传统的神经网络模型，预测标签应是当前学生是否会发生近视，即

预测属性标签	数据类型
是否发生近视	bool

对于改进后的神经网络模型，预测标签转化为

预测属性标签	数据类型	预测属性标签	数据类型
p_{i1}	int	p_{i7}	int
p_{i2}	int	p_{i8}	int
p_3	int	p_{i9}	int
p_{i4}	int	p_{i10}	int
p_{i5}	int	p_{i11}	int
p_{i6}	int	p_{i12}	int

表中 $p_{i1} \sim p_{i12}$ 的取值范围是 $(0, 1)$
 $p_{i1} \sim p_{i12}$ 满足关系 $\sum_{n=1}^{12} p_{in} = 1$

根据我们选取的指标特征和设计的 AI 的预警机制，我们创造性地将传统 BP 神经网络进行变形，以适应我们建立的演化机理模型。

7.2 基本神经网络结构

模型采用 3 层 BP 神经网络，输入层以 12 个神经元作为特征向量，对应 $a_1 \sim a_{12}$ 这 12 个影响近视的指标，根据经验和通过训练学习后，考虑网络的学习次数和后续计算出的标准临界综合得分综合比较后选定合适的隐含层节点数。这里以选择隐含层节点数为 12 为例，输出层以 12 个神经元做为输出向量，对应于 $a_1 \sim a_{12}$ 12 个指标对近视影响程度的权重。再通过演化机理模型确定出标准临界综合得分 G_{is} 、危险临界综合得分 G'_{is} 和安全综合得分 G''_{is} 对测试集进行分类，通过其预测的正确率对神经网络参数进行反馈修正。

7.3 改进神经网络

我们创新地将传统的神经网络改进，将预测分类的任务从简单的二分类近视和不近视的问题，转换成寻找最优的权重向量的问题。

实质上是将隐含层的最后一层暴露出来，使其成为输出层，通过得到的 12 个权值（在原网络中即隐含层的 12 个各个神经元到输出层神经元之间的连接权的平均值）作为输出向量，结合演化机理模型，计算出当前数据集下训练出的标准临界综合得分 G_{is} 、危险临界综合得分 G'_{is} 和安全综合得分 G''_{is} ，通过计算真实的数据的近视与不近视的比例 k_1 ，以及标准综合得分评判的近视与不近视的比例 k_2 ，将 $(k_1 - k_2)^2$ 视作该神经网络的损失函数，使得改进后的神经网络能够迅速收敛，从而能够训练出最优的指标权重向量。

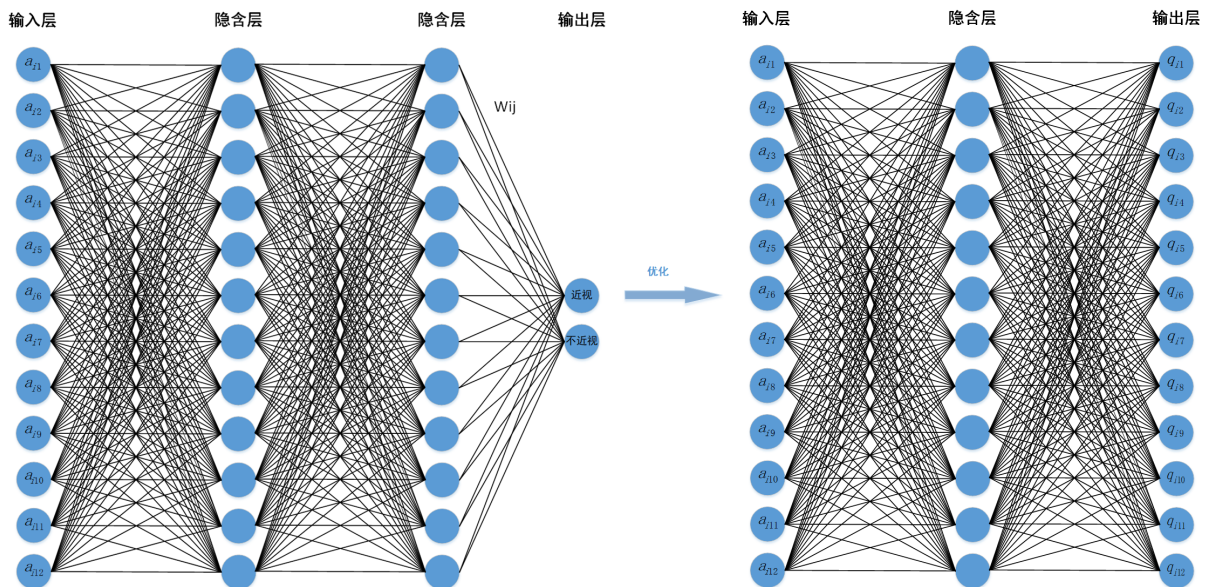


图 8 神经网络改进示意图

7.4 神经网络训练方法

用于神经网络训练的样本集必须至少满足以下两个要求：

- 训练样本集要足够大，样本集的大小对神经网络的预测能力有很大影响，样本集越大，预测的准确率越高，即标准综合得分的选取越合适，权重分配越准确，分类效果越好。
- 样本集尽量包含学生的各项指标，即要求训练样本集基本完整。
- 从训练样本集中选出 10% 的数据作为改进神经网络的验证集，使其预测结果更加精准。

设置基本训练参数为：

- 学习率 $\alpha = 0.01$
- 训练次数 $t = 50000$
- 训练目标误差 $MSE = (k_1 - k_2)^2 \leq 0.0001$

在对 BP 神经网络进行训练时，由于我们人为地设置 $(k_1 - k_2)^2$ 作为损失函数，在理论上该模型的训练过程能够以更快的速度收敛，能够有效地减小训练时间。

7.5 改进神经网络评估

近视的评估过程较为复杂，有诸多影响视力的指标，且各指标对近视的影响程度以及各个指标之间的相互影响程度难以明确。而 BP 神经网络模型具有较强的自学习、自适应、联想记忆以及非线性模式识别的能力，适用多因素、多指标的识别。除此之外，由于人为地优化了网络结构以及创新式的损失函数的引入，使得原本的 BP 神经网络从随机的、漫无目的的寻找最优输出结果到能够很快的收敛，大大减小了训练时间和成本。而且此神经网络模型可推广到其他类似的多因素、多指标的分类预测问题，具有极强的可移植性和可执行性。

八、模型应用和实现方案

在模型应用实现和推广的过程中，应至少包含以下四个阶段：

8.1 原始数据采集阶段

首先，我们需要对同一地区的 5000 ~ 10000 名高中应届毕业生进行线上或线下调查。其次，通过调取该这批学生的入学体检表，统计近视人数为 m_i ；调取高三毕业体检表，统计近视人数为 m_o 。得到高中阶段新增近视人数为 $m = m_o - m_i$ ，作为真实数据的样本。

根据收集的真实可信的问卷调查，我们可以得到学生人数-综合得分函数 $(M - G_i)$ ；根据近视新增人数，可以建立微分方程，通过神经网络模型，计算出合理的标准临界综合得分 G_{is} 、危险临界综合得分 G'_{is} 和安全综合得分 G''_{is} ，由此形成基础预测模型，可用于该地区普通高等学校学生的近视预测。

按照上述的数据采集方法，我们对根据我国教育国情划分的四级教育地区、八类教育等级学校，三十二种模型进行数据统计，得到更加可靠的结果。

8.2 基础模型训练阶段

获取原始数据后，需要对数据进行持久化储存，备份，预处理等过程，然后对改进的神经网络模型进行训练，再根据模型的输出结果调整模型，形成模型训练闭环。以下是模型训练的步骤：

- 数据集选取，即通过各个渠道收集有用信息，并将信息持久化储存。为防止因未知因素丢失、损坏数据，应将数据进行备份，包括多重备份，异地备份等。
- 数据预处理，即对已有数据进行处理，如填充缺失值（使用属性的中心度量填充等方法）、数据集成，数据归约等。
- 训练模型，即使用设计的改进神经网络模型对处理好的数据进行训练。
- 分析改进，即根据模型的输出结果调整神经网络的网络结构，参数设置等，以达到更加准确稳定的输出结果和更快的收敛速度。

8.3 试点测试阶段

试点学校的选择应包括四个等级的教育地区、八个类型的学校。即至少应选取三十二种年龄段不同、地区不同的典型学校的学生进行模型试点投放。进而得到一个完整的适用于多种复杂情况的动态模型数据库。

四级教育地区分别为：

- 一级地区：一线城市和考生大省部分地区
- 二级地区：新一线城市
- 三级地区：二、三线城市
- 四级地区：四级以下城市和农村

八类教育等级学校分别为：

- 学龄前教育
- 小学
- 普通初中
- 重点初中
- 普通高中

- 重点高中
- 一本大学
- 非一本大学

为使模型更加可靠，建议调查人员在每种类型学校和阶段，都进行多样化的试点运行，进行模型具体参数的验证和调整。

8.4 全面推广阶段

经过试点学校的实验和调整，由于模型的自学习和自适应性，模型的训练集也相对应的大幅扩大，此时的神经网络 + 演化机理模型应该已经基本趋于稳定，预测结果和AI 预警机制的建议应更加准确，即模型在试点过程中已经实现了不断自我学习和自我完善。

因此，在小范围的试点结束后，应收集试点学校的汇报信息，了解模型在该地区的该类学校中的表现情况，通过适当的方式大规模向全国范围内推广。在推广过程中，应至少注意以下几点：

- 对于特殊少见。即统计规律与同层次学校大幅不同且数据特征与模型契合度极差的学校，应针对该种学校采用特殊化的参数处理，并且避免其数据作为模型的训练集，以防大量不合理数据对训练结果产生负面影响。
- 建立长久有效的反馈机制，以便模型推广后及时收集特殊数据和错误分析。
- 通过长期的新的实验和调研，可在有大量数据基础的条件下向训练模型中添加新的指标进行训练。

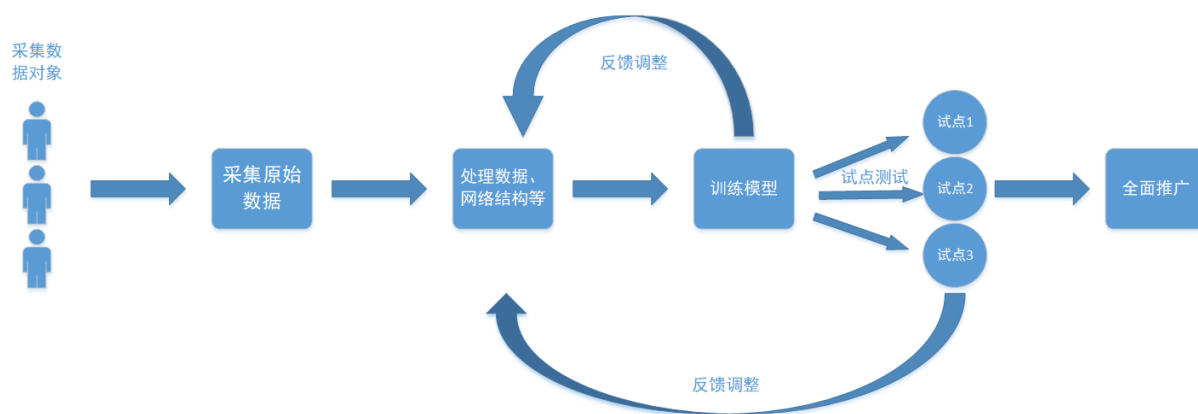


图9 方案实施流程图

九、模型验证和结果分析

9.1 模型验证

对于本模型的验证,需要大量真实数据作为基础支撑来训练神经网络模型以计算权重分配,但即使在处于现在的信息化时代也很难在短时间内收集到大量真实可靠的数据作为训练集参与运算。

因此,我们从模型推广应用的角度出发。通过分析本模型的使用特点,从模型的推广适应性角度来进行模型验证。

9.2 结果分析

本模型得到的结果,包括利用神经网络模型得到 12 种指标的权重和利用权重得到的一个合理的、适用于某类学生的一个视力综合得分的计算方法。

我们利用得到的近视发生概率-综合得分函数 $(P - G_i)$ 和学生人数-综合得分函数 $(M - G_i)$,利用一定量的数据,建立微分方程,计算出三个临界视力综合得分,即标准临界综合得分 G_{is} 、危险临界综合得分 G'_{is} 和安全综合得分 G''_{is} ,将视力综合得分划分为四种近视危险区域。当我们需要预测一名学生的视力情况时,我们首先将针对该学生的调查结果带入到视力演化机理模型,得出该学生视力综合分数,评判该学生当前视力状态,根据其视力等级及近视发生概率,给出合理的个性化的智能预警方案。

由于学生的年龄在不断增长,每一类型学生的生活习惯有所不同,因此视力演化机理模型也应动态随时间变化,实时更新适用于多种复杂情况的动态模型数据库。权重比例也在随着时间推移和训练集的不断增加而波动变化,因此同一类型的学生应进行定期测试分析,动态调整模型参数,确保模型的可靠性、稳定性。

十、模型的改进与推广

10.1 模型改进

在本题中,所提出的模型是用于解决针对某地区,处于第 i 个时期的学生的视力情况进行的分析和预测。学生的年龄段不同、学生所在学校的竞争压力和学习压力的不同、所在地区的经济水平、教育程度的不同均会影响模型的参数。因此,对于处在不同的环境的学生的视力情况进行预测评估时,需要多样化数据库信息,找到与之环境情况相似的地区模型,进行预测分析。

为使模型的使用者能够更加方便的使用我们的模型,我们从我国的教育国情出发,建立相应的动态子模型库。根据我国地区的经济发展状况不同,将教育地区划分为 8 个等级;根据学校的教育层次和学习压力,将学校划分为 4 个等级。因而形成 32 种动态子模型。此外,随着国家对青少年近视的日益重视,青少年的近视率和用眼习惯会不断

的变化，信息系统内的数据具有较强的时效性，因此模型的信息数据库要进行时事数据更新和替换以保证预测结果更加真实、可靠。

10.2 模型推广

本模型在理论上可以对具有多因素影响、统计规律近似正态分布特点的问题进行分析。因此针对的问题范围无需限定在学生近视的小范围内，可以解决有如上数据特征的诸多问题，如，人类各类疾病的分析，工业级产品质量预测分析，大规模的农产品质量预测分析等。

十一、模型优缺点分析

11.1 模型优点

- 本模型的服务对象是从老师、家长和学生等对专业知识了解较少的群体，使用选择调查问卷的形式，具有低用户成本的特性。
- 本模型在本质上可以对具有多因素影响、统计规律近似正态分布特点的问题进行分析。具有很好的可执行性和可移植性。
- 本模型通过引入神经网络智能化动态地处理数据，又将数据结果重新作为训练集参与神经网络的构建，具有极强的自学习性和自适应性。
- 对于神经网络的设计具有创新性，人为增加一个损失函数，使得该神经网络能够以较快的速度收敛。
- 本模型并没有停留在概率预测层面，而是能够对学生的视力情况进行合理性评估，给出个性化较强预警和合理性建议。

11.2 模型缺点

- 对于近视演化机理模型的建立，我们在建立近视发生概率-综合得分函数时依据其变化趋势采用了线性化的思想，与实际函数间会产生一定的误差。
- 在针对权重训练神经网络模型时，若想得到真实、准确的预测结果，需要大量的基础数据作为支撑。

参考文献

- [1] 李静. 中小学生视力不良的影响因素分析 [J]. 中国校医,2014,28(04):316-317.
- [2] 李圣琦, 甘娜.2394 名高校新生近视与体质健康相关因素分析 [J]. 中国校医,2018,32(11):801-802+818.

- [3] 赵爱华, 马德珍. 泰安市农村中小学生视力不良影响因素分析 [J]. 中国校医, 2018, 32(09): 676-678+681.
- [4] 张彩虹. 临安市 12780 名中小学生近视调查 [J]. 中国校医, 2015, 29(03): 178-179.
- [5] 王赞, 谌丁艳, 熊华威, 张浩, 周丽. 我国青少年近视影响因素与防治措施研究 [J]. 实用预防医学, 2016, 23(03): 380-381+385.
- [6] 陈夏明, 王艳, 陈惠芳. 2012—2016 年深圳市宝安区某街道入园儿童视力筛查结果分析 [J]. 中国校医, 2018, 32(05): 321-322+324.
- [7] 刘冰, 程兰芳, 张凤琴. 977 例大学生近视影响因素分析 [J]. 中国校医, 2018, 32(11): 803-805+808.

附录 A 计算综合得分-MATLAB 源程序

```
function [Gi] = score(qi, Qi)
Gi = 0;
for i = 1 : 12
Gi = Gi + qi(i) * Qi(i);
end

%测试样例:
>> qi = [0.1,0.08,0.02,0.05,0.05,0.12,0.08,0.2,0.1,0.1,0.1,0.1];
>> Qi = [2,8,6,8,4,2,10,6,8,4,2,8];
>> k = score(qi, Qi)

k =
6.0000
```

附录 B 预测近视发生概率-MATLAB 源程序

P_score.m

```
%参数说明: Gi是被预测对象的综合得分, Gimin是训练集样本中最低得分, Gimax是训练集样本中最高得分
%在确定的权重系数条件下, 计算被预测对象该得分的近视发生概率
%
function [Pgi] = P_score(Gi,Gimin,Gimax)
k = Gimax - Gimin;
Pgi = -k * (Gi - Gimax);
Pgi

%测试样例:
```

```
>> P_s = P_score(10,1,12)
```

```
P_s =  
    22
```

%程序说明：在样本集中，最低得分为1分，最高得分为12分，该对象的得分为10分。

%输出结果表示该对象发生近视的概率为22%

附录 C 求解标准临界综合得分–MATLAB 源程序

```
function [Gis] = standard_critical(Gimin,Gimax,Ms,m)
```

%根据近视发生概率-综合得分函数Pgi、得Gi分人数函数Mgi、样本集最低分Gimin、

%样本集最高分Gimax、样本集总人数Ms、样本集近视比例m%计算划分近视与非近视的标准临界。

```
syms Mgi Pgi sig miu Gi
```

```
Mgi = (1 / (sqrt(2*pi)*sig)) * exp(-((Gi-miu)^2)/(2*sig^2));
```

```
Pgi = (Gi - Gimax) / (Gimax - Gimin);
```

```
f = Mgi * Pgi;
```

```
Gis = int(f,Gi,Gimin,Gi)
```

测试样例：

```
>> standard_critical(1,12,120,70)
```

```
Gis =
```

```
(2251799813685248*(sig^2*exp(miu/sig^2)*exp(-miu^2/(2*sig^2))*exp(-1/(2*sig^2))  
-sig^2*exp(-Gi^2/(2*sig^2))*exp((Gi-miu)/sig^2)*exp(-miu^2/(2*sig^2)) +  
(2^(1/2)*miu*pi^(1/2)*erfi((2^(1/2)*(Gi-miu)*(-1/sig^2)^(1/2))/2))/(2*(-1/sig^2)^(1/2)) +  
(2^(1/2)*miu*pi^(1/2)*erfi((2^(1/2)*(miu-  
1)*(-1/sig^2)^(1/2))/2))/(2*(-1/sig^2)^(1/2)))/(62088675899714871*sig) -  
(4503599627370496*2^(1/2)*pi^(1/2)*(erf((2^(1/2)*(miu-1)*(1/sig^2)^(1/2))/2) +  
erf((2^(1/2)*(Gi-miu)*(1/sig^2)^(1/2))/2)))/(20696225299904957*sig*(1/sig^2)^(1/2))
```

%因未代入全部数值型数据，故这里输出的是标准临界的符号表达式。

danger_critical.m

% 计算危险临界综合分数

safe_critical.m

% 计算安全临界综合分数

因计算危险临界综合分数和安全临界综合分数的程序与计算标准临界综合分数的程序重复度较高，故在这里不做展示，仅需改变 standard_critical.m 中的个别参数即可。