

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 19B367

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. 孙毓寅 17140288021

2. 杨昌帅 17130110001

3. 张 琦 17130110044

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				

基于 BP 神经网络的马尔可夫近视预警优化模型

一、 摘要

本文建立了基于 BP 神经网络的灰色马尔可夫优化模型来预测儿童青少年近视度数的增长趋势。机器学习是 AI 时代解决问题的热点方法，就机器学习，本文提出基于随机森林的机器学习模型，客观地给出对于儿童青少年近视度数的建议。

针对问题一，本文从外部因素、内部因素和其它因素 3 个角度考虑。设立用眼环境、近距离工作量、户外活动量、遗传、个体参数和生活习惯 6 个关键因素。然后根据这 6 项关键因素的特性，将关键因素的影响分解到对象亮度、距对象距离、光的颜色、近距离工作时间、近距离阅读页数、户外活动时间、睡眠时间和父母患高度近视情况 12 个可测量的参量上，结合与近视度数的指数函数和偏微分函数关系，建立了 6 项关键因素的量化模型。由这些模型我们可以得出以下结论：户外运动量越大，生活习惯越好，用眼环境越柔和，近距离工作量越小，遗传情况越好，个体参数越均衡，则近视度数增长越慢。

针对问题二，本文提出基于灰色马尔可夫预测的视力演化机理模型。马尔可夫预测模型的实现需要基于近视度数初始状态的概率，本文使用灰聚类分析方法来获得此目标量。将灰色聚类分析得到的近视度数初始状态代入基于逆矩阵的马尔可夫预测模型中，能够得到相同时间间隔中近视度数改变量在某一范围中的概率，即近视度数的演化机制。

针对问题三，本文采用 BP 神经网络模型替换问题二中的灰聚类分析法来计算各项关键因素的权重。将 BP 神经网络所得权重代入问题二中提出的灰色马尔可夫模型，进而计算出近视度数增加达到某一度数的概率。通过将近视度数与相应概率的权值乘积累加，计算得到标准概率系数 ω ，通过将 ω 与本文设立的最高风险系数 ω_0 进行对比，对照预警级别判定表，判断是否对于对象的实力情况做出预警。当 ω 小于 4.9 时，近视度数增长状况良好；当 ω 在 4.9 和 5.6 之间时，近视度数增长状况需要引起注意，当 ω 大于 5.6 时，近视度数增长过快，需要采取措施。针对非学者群体，本文采用简单的待测量和易于理解的输出量实现预警机制。

针对问题四，本文选取泛化性强的随机森林模型，结合本文所提出的 12 项待测参量，设计出可执行数据库的具体格式，建立相应的基于随机森林的机器学习模型，并生成人机对话窗口。在建立决策树时，每一节点的分支方向以各项指标的基尼系数为参考，最终生成的决策树经过随机森林模型的演化，生成近视度数 300 度以下、300 至 600 度、600 至 900 度和 900 度以上四个视力范围的权重。将所有决策树相同级别的权重求平均，得到四个视力范围的最终权重。这四项中权重最大的一项即为对研究对象下一阶段近视度数的预测。因此当用户将可执行数据库中的参量按要求输入到此窗口中，将会得到此模型对数据对象近视度数增长情况的建议。针对非学者群体，本文将采用亲切简单的语言实现人机交互。

针对问题五，本文提出通过访问教务管理系统和各软件厂商提供数据接口的方式实现数据的收集，通过数据的汇聚实现数据库开放，通过底层数据交换实现数据实时采集。同时本文提出通过普及视力预警程序和加强屈光率的监测与矫正，来更好地实现本文模型。

关键字：马尔可夫链 灰聚类 BP 神经网络 机器学习 随机森林 基尼系数

二、问题重述

2.1 问题的背景

目前,我国儿童青少年的近视率已超 7 成,青少年近视率排世界第一。2019 年 4 月 29 日,国家卫健委召开记者发布会,发布了我国儿童青少年近视率的最新数据,并提出三点问题^[1]。

第一,我国儿童青少年总体近视发病形势严峻。2018 年全国儿童青少年总体近视比例达到 53.6%,其中高中生近视比例达到 80%以上。

第二,低年龄段近视问题比较突出。小学阶段从一年级的 15.7%增长到六年级的 59.0%,初中阶段从初一年级的 64.9%增长到初三年级的 77.0%,小学和初中阶段成为我国近视防控的重点年龄阶段。

第三,高度近视问题不容忽视。近视度数高于 600 度即为高度近视,高度近视是致盲性眼病之一,容易导致一系列严重的并发症。我国高三学生患高度近视的人数占总人数的 21%以上。

高近视率在影响下一代的身心健康和前途发展的同时,还大大增加了国家征兵的难度。因此,儿童青少年近视眼防控是一项全社会视力健康促进活动,关系着下一代身心健康成长。

2.2 问题的提出

建立数学模型,就近视现象的影响因素及预警机制进行研究,结合信息化时代优势,为科学预测提供定量数据,解决以下问题:

1. 结合信息化时代的优势,提出影响视力的关键因素并对其进行量化,要求明确给出需要获得的对应数据及可以获得的可能途径。
2. 基于问题 1,提出可执行的眼睛视力演化机理模型。
3. 基于问题 2,提出面向非学者群体的眼睛视力 AI 预警机制模型。
4. 基于以上问题,设计可供学习的信息系统,以及通过查阅资料和数据以及仿真等手段进行眼睛视力的机器学习模型。
5. 提出一份实现本文所述模型的具体方案。

三、问题的分析

3.1 问题一的分析

要提出影响视力的关键因素,可以从外部因素、内部因素和其它因素三个角度考虑。设立用眼环境、近距离工作量、户外活动量、遗传、个体参数和生活习惯六个关键因素。为了量化这些关键因素,本文将关键因素分解为可测量的参量,具体提出得到这些可测量参量的途径,从而计算出关键因素的数据。这样就得到了影响视力的关键因素的系数。最终通过将各个系数整合到近视度数上来实现对近视度数的预测。

3.2 问题二的分析

本文选取基于灰色马尔可夫预测的视力演化机理模型。选取灰色马尔可夫预测模型的原因是其能够以概率的形式反映由数据不足及其它因素带来的不确定性,使得预测的结果更为客观可靠。马尔可夫预测模型的实现需要基于近视度数初始状态的概率,由于灰聚类分析方法能够在样本数目较小时规避结果的偶然性,故本文使用灰聚类分析方法来获得此目标量。将灰聚类分析得到的近视度数初始状态代入马尔可夫预测模型中,就能够得到相同时间间隔中近视度数改变量在某一范围中的概率,即近视度数的演化机制。

3.3 问题三的分析

为实现视力预警机制模型,本文采用 BP 神经网络模型代替问题二中的灰聚类分析法来计算各项关键因素的权重。将 BP 神经网络所得权重带入问题二中提出的马尔可夫模型,进而计算出近视度数增加达到某一速度的概率。通过将近视度数与相应概率的权值乘积累加,计算得到标准概率系数 ω ,通过将 ω 与本文设立的最高风险系数 ω_0 进行对比,对照预警级别判定表,判断是否对于对象的实力情况做出预警。当 ω 小于 4.9 时,近视度数增长状况良好;当 ω 在 4.9 和 5.6 之间时,近视度数增长状况需要引起注意,当 ω 大于 5.6 时,近视度数增长过快,需要采取措施。为了照顾非学者群体,本文采用简单的待测量和易于理解的输出量实现预警机制。

3.4 问题四的分析

为实现机器学习,本文选取泛化性强的随机森林模型,结合本文所提出的 12 项待测参量,设计出可执行数据库的具体格式,建立相应的基于随机森林的机器学习模型,并生成人机对话窗口。在建立决策树时,每一节点的分支方向以各项指标的基尼系数为参考,最终生成的决策树经过随机森林模型的演化,生成近视度数 300 度以下、300 至 600 度、600 至 900 度和 900 度以上四个视力范围的权重。将所有决策树相同级别的权重求平均,得到四个视力范围的最终权重。这四项权重中最大的一项即为本模型预测针对此时研究对象的下一阶段近视度数预测。因此当用户将可执行数据库中的参量按要求输入到此窗口中,将会得到此模型对数据对象近视度数增长情况的建议。针对非学者群体,本文将采用亲切简单的语言实现人机交互。

3.5 问题五的分析

为了实现本文所述模型,需要相对较完善的数据库、较成熟的视力预警程序和矫正屈光率措施,本文提出通过访问教务管理系统和各软件厂商提供数据接口的方式实现数据的收集,通过数据的汇聚实现数据库开放,通过底层数据交换实现数据实时采集。同时提出通过普及视力预警程序和实施矫正屈光率措施,来更好地实现本文模型。

四、模型假设

假设 1: 假设视力的增长是自然连续的,没有突然下降。

假设 2: 假设本文得到的数据都是真实的,不存在偶然性。

假设 3: 假设本文所提到的近视度数均非假性近视。

五、符号说明

符号	说明
θ	用眼环境系数
Q	近距离工作量系数
τ	户外活动量系数
p	个体参数系数
ξ	遗传系数
S	生活习惯系数
P	马尔可夫条件概率
η_j^k	灰聚类分析权重
σ_i^k	系数矩阵
l	中间层神经元数
μ	样本估计量
$gini$	基尼系数

六、模型的建立

6.1 问题一：影响视力关键因素的量化模型

研究影响儿童青少年近视的因素是一个多方面的问题。本文将对视力的影响理解为对于近视度数的影响，最终通过将各个系数整合到近视度数上来实现对近视度数的预测。同时，通过将较宽泛的物理量细化为可测量的物理量，获得题目所要求的影响视力的关键因素的具体数值。

6.1.1 眼球工作及近视原理概述

眼球工作的原理是光线通过晶状体时。通过睫状肌的调整使像清晰的出现在视网膜上，调整的原则是在看近物时睫状肌紧张晶状体变厚，看远物时睫状肌放松晶状体变薄。形成的光学像清晰的呈现在视网膜上，视网膜上丰富的感光细胞就会将呈在视网膜上的光影转换成电信号，传递到大脑成像。



图 1. 眼球成像(左)与近视(右)原理示意图^[2]

而近视的原理则是由于眼球长时间看近物，睫状肌处于持续紧张的状态，使得肌肉无法正常放松。于是再看近物时，晶状体无法变回原来的厚度，呈现的影像无法清晰的呈现在视网膜上，就会看东西模糊，形成近视现象。

6.1.2 视力的影响体系建立

在研究对于视力的影响因素时，本文从内部因素、外部因素和其他因素三个方面切入。其中，外部因素的影响主要体现在用眼环境、近距离工作量和户外活动量；内部因素的影响主要体现在遗传和个体参数；其它因素主要体现在生活习惯。经过研究与分析，本文将这六项影响因素设置为关键因素。

表 1. 影响视力的关键因素

影响方面	关键因素
外部因素	用眼环境
	近距离工作量
	户外活动量
内部因素	个体参数
	遗传
其它因素	生活习惯

6.1.3 关键因素的定义

本文提出六个关键因素，并将在下文阐述如何量化这些关键因素。由于本文的最终目的是通过将与各关键因素相对应的数据输入模型，经过模型演算预测对象未来的近视度数并给出相应预警，所以本文将所有影响视力的关键因素，转化成对于近视度数的影响系数。最终通过将各个系数整合到近视度数上来实现对近视度数的预测。

(1) 用眼环境

良好的用眼环境会在一定程度上缓解眼球疲劳，眼球与外部环境的交互是建立在接收信息的基础上的。为便于计算，本文将眼球接收到的信息进行分析与筛选，将用眼环境定义为眼球接收到的光强和环境中的不同色光。

眼球接收到的光强难以测量，为了得到关键因素的具体数值，我们将眼球接收到的光强 $U(t)$ 定义如下

$$U(t) = \frac{\alpha |l(t) - l_0|}{|d(t) - d_0|} \#(1)$$

式中 α 为光强常数， l_0 为观察对象初始亮度， $l(t)$ 为 t 时刻观察对象的亮度， d_0 为眼球距观察对象的初始距离， $d(t)$ 为 t 时刻眼球距观察对象的距离。

环境中的不同色光对于近视度数的影响系数记作 $C(t)$ ，称为色光系数。色光系数通过图 2 中的比对色光系数卡直接读出。

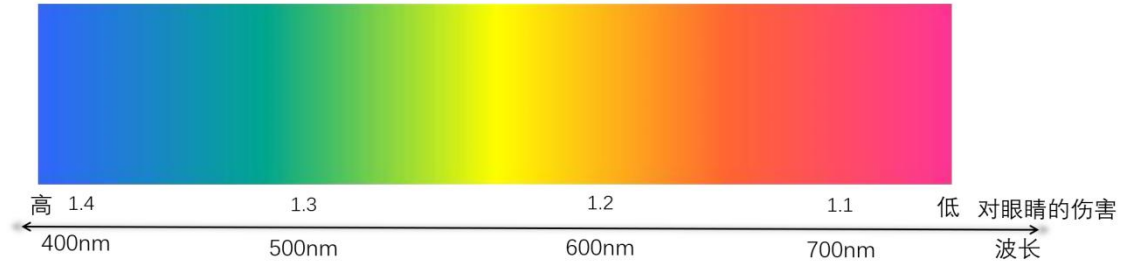


图 2. 色光系数卡

故用眼环境系数 θ 为

$$\theta = U(t) \cdot C(t) \quad (2)$$

(2) 近距离工作量

眼球疲劳的主要原因就是长时间近距离的不规律用眼，因此本文将近距离工作量列入影响视力的关键因素中。规定 33 厘米以内的工作为近距离工作。由于工作量这一概念本身较为宽泛，难以测量。经过分析与筛选，本文将近距离工作量系数 Q 分解为了近距离时间系数和近距离强度系数两方面。

本文定义 t 时刻工作时间对近视度数的影响系数为 $T(t)$ ，称作近距离时间系数。

$$\frac{\partial T(t)}{\partial t} > 0 \quad (3)$$

$$\frac{\partial^2 T(t)}{\partial t^2} > 0 \quad (4)$$

经过分析与推导，本文用指数函数模拟时间系数与近视度数的关系，即

$$T(t) = \gamma e^{\varphi t} \quad (5)$$

其中 γ, φ 均为常数。

本文定义工作强度为阅读数量带来的压力，可知近距离强度系数 y 与阅读页数 x 成线性增长关系。即

$$y = kx + b \quad (6)$$

其中 k, b 均为常数。

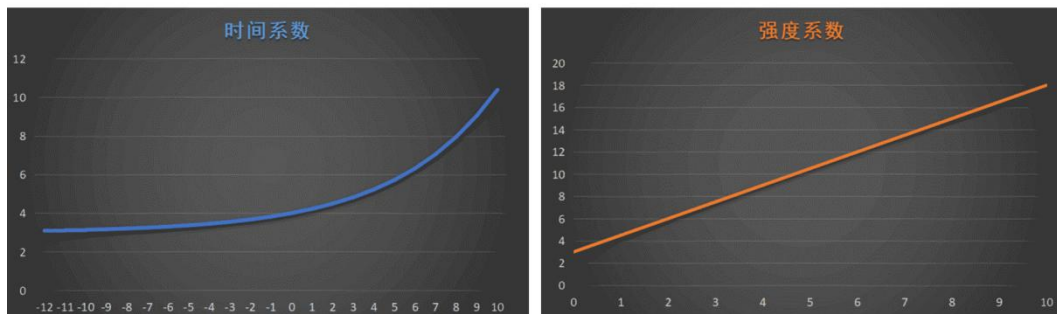


图 3. 时间系数与强度系数示意图

故近距离工作量系数 Q 有表达式如下

$$Q = \frac{y}{T(t)} \#(7)$$

(3) 户外活动量

研究表明，合理的户外活动是预防近视度数增长的有效措施，户外活动时间与近视的发展率和进展率呈负相关。本文将户外活动量简化为户外活动时间系数 τ ，则 τ 反比于户外活动时间长度 t ，即

$$\tau = \frac{\beta}{t} \#(8)$$

其中 β 为常数。

(4) 个体参数

近视度数的生长速度视个体不同而变化。其主要影响因素为年龄、性别、身高和体重。本文用体重指数 BMI 来衡量身高和体重对于近视度数的影响。即

$$BMI(n) = \frac{weight(n)}{hight^2(n)} \#(9)$$

其中， $weight(n)$ 为 n 岁时对象的质量， $hight(n)$ 为 n 岁时对象的高度。

$$B(n) = \frac{m}{BMI + p} + p \#(10)$$

$B(n)$ 为 n 岁时身高体重对于近视度数的影响系数，称体型系数。

同时，近视度数的自然增长随年龄的分布函数如下

$$f(n) = \frac{1}{\sqrt{2\pi}\sigma n} e^{-\frac{(\ln n - \mu)^2}{2\sigma^2}} \#(11)$$

其中 n 为年龄，取值为整数。 σ, μ 为统计学参量， $f(n)$ 为 n 岁时的年龄系数。故个体参数系数 p 的计算方式为

$$p(n) = B(n) \cdot f(n) \#(12)$$

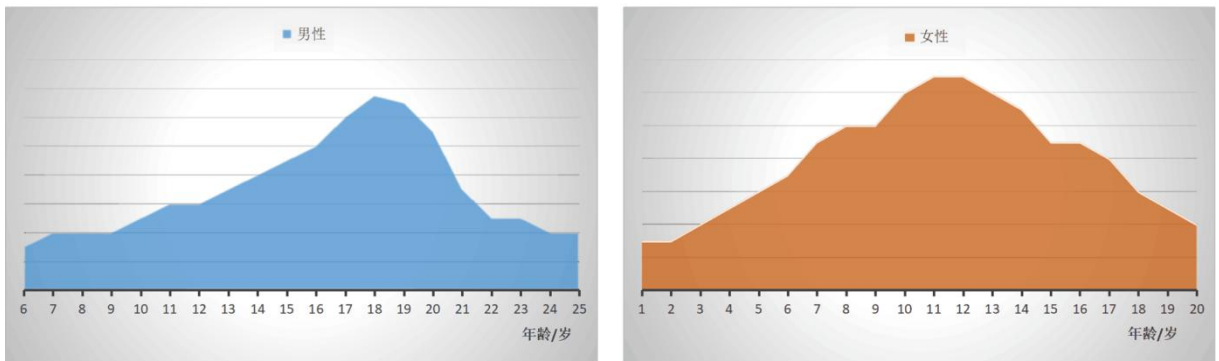


图 4. 不同性别对于近视度数随年龄增长情况的影响

(5) 遗传

根据现有研究^[4]，高度近视占有近视情况的 30%，而遗传对于高度近视患病率的影响举足轻重。因此本文将遗传作为影响视力的关键因素之一。为便于计算，本文仅考

考虑父母对子女的影响。将子女患高度近视的概率定义为遗传系数 ξ 。

表 2. 父母患高度近视情况对子女视力的影响

双亲 A	双亲 B	遗传系数 ξ
患高度近视	患高度近视	$\geq 98\%$
患高度近视	近视基因携带	50%
近视基因携带	近视基因携带	25%
患高度近视	正常	10%-15%

(6) 生活习惯

在生活习惯方面本文主要考虑睡眠时间，称生活习惯系数 S 。根据关于学生近视状况调查结果，可知睡眠时间与近视度数有较为密切的关系且作用方向与近视度数增加方向相反，如图 5。根据调查结果本文将 9 小时作为生活习惯系数 S 的评判标准。

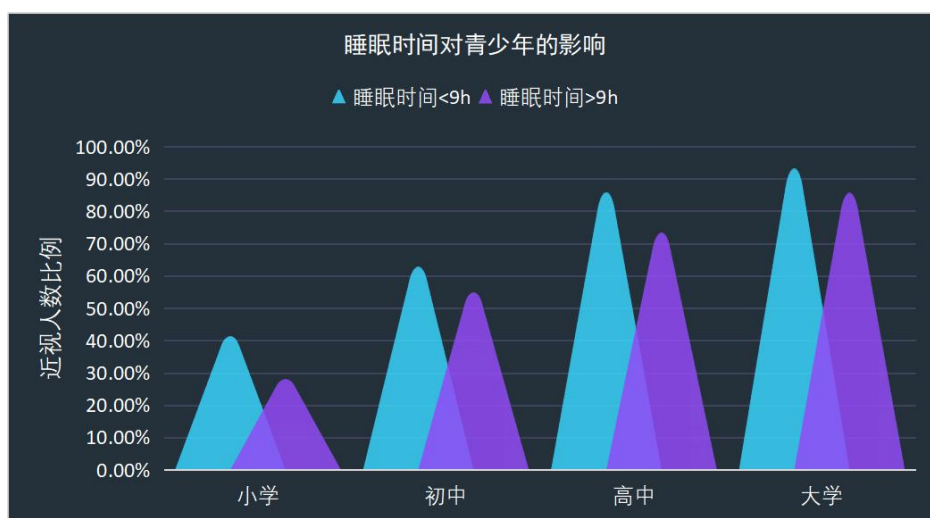


图 5. 睡眠时间对于不同年龄段近视人数比例的影响

另外，本题为了便于计算将生活习惯简化为睡眠时间，但实际生活习惯还有许多其它因素，如饮食习惯。建议饮食上多摄入含 Ca^{2+} 、 Fe^{2+} 、 Zn^{2+} 和维生素 A、C 等微量元素的食物，如多吃胡萝卜、黄豆、蔬菜、牛奶、花生等，对此本文不多做展开。

6.1.4 参量的获得途径

在互联网时代，信息化的共享已经成为不可忽视的主题之一。这也为获得本文模型中所需的数据提供了便利。结合信息化共享的思想，本文给出了各项参量的几种可行的获得途径。

表 3. 各参量获得途径

关键因素	计算所需的参量	获得参量的途径
用眼环境	对象亮度	通过现代科技手段测量
	与对象的距离	直接测量
	色光	观测或通过网络查询
近距离工作量	近距离工作时间	记录工作时间长度
	近距离工作强度	记录工作时阅读的大致页数
户外活动量	户外活动时间	记录户外活动时间长度
个体参数	身高	通过教育管理系统及相关软件得到
	体重	
	性别	
	年龄	
遗传	父母患病及携带情况	通过基因检测与网络共享查询
生活习惯	睡眠时间	利用智能手环或检测软件记录

6.2 问题二： 基于灰色马尔可夫预测的优化视力演化机理模型

视力在时间推移中会随着年龄的增长而产生变化,当近视度数上升到某一临界点时,需要通过人为干预来阻止其继续上升。因此,合理地记录和预测近视度数,是视力保护的关键之一。由于本题数据不足,具有很大不确定性,本文选用马尔可夫预测模型进行分析。

马尔可夫预测模型能够以概率的形式反映由数据不足及其它因素带来的不确定性,使得预测的结果更为客观可靠。

6.2.1 马尔可夫预测模型的介绍

具有 m 个状态的系统,其状态的变化只发生在参数的离散值上,一般说来,系统将来处于状态 i 的概率与它全部历史有关,所以应该用条件概率^[5]:

$$P(X_{n+1} = i | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \quad (13)$$

其中 $X_0 = x_0, X_1 = x_1, \dots, X_n = x_n$ 代表系统以前的所有状态,如果将来的状态只与现在的状态有关,条件概率变为:

$$P(X_{n+1} = i | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = i | X_n = x_n) \quad (14)$$

此过程称为马尔可夫过程。

对于马尔可夫链,能够自时刻 t_m 的状态 i 变为时刻 t_n 的状态 j 的条件概率为:

$$p_{ij}(m, n) \stackrel{\text{def}}{=} P(X_n = j | X_m = i), n > m \quad (15)$$

如果 p_{ij} 只与时间差 $t_n - t_m$ 有关,而与时间起点 t_m 无关,则称 p_{ij} 为齐次马尔可夫链。

对于一个具有 m 个状态的系统，转移概率矩阵是指一个可以将各种情况的转移概率归纳的矩阵。

$$P \stackrel{\text{def}}{=} \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1m} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2m} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & p_{m3} & \cdots & p_{mm} \end{bmatrix} \quad \#(16)$$

$$P(n) = P(n-1) \times P = P(n-2) \times P \times P = \cdots P(0) \times P^n \quad \#(17)$$

初始状态概率和转移概率决定系统在某时刻的状态概率，马尔可夫链的核心思想就是求解初始状态概率和转移概率矩阵。

6.2.2 灰聚类分析方法

要运用马尔可夫预测模型，需要知道近视度初始状态的概率。一般的计算方式是根
据大量观测数据计算出某一度数出现的频率，用这一频率作为近视度数处在这一度数的
概率。但是这种计算方法需要大量数据，本题已有的条件很难满足这一要求，如果使用
这种方式，本题的结果将缺乏可靠性。因此本文通过研究与分析，采用灰色聚类分析方
法，可以在数据样本数较少的情况下得出近视度数的初始量。

灰聚类分析法能够在灰色关联矩阵或灰数的白化权函数的学术基础上，将观测指标
或观测对象聚集成为若干个可定义类别。

设样本中有 n 个聚类对象， m 个聚类指标和 s 个不同灰类，参照第 i 个对象($i = 1, 2, \dots, n$)关于指标 j ($j = 1, 2, \dots, m$)的样本值，将第 i 个对象归于第 k ($k = 1, 2, \dots, s$)个灰类
之中。将 j 指标 k 子类的白化权函数记为 $f_j^k(*)$ ，典型的白化权函数如图 6 所示，其中
 $x_j^k(1), x_j^k(2), x_j^k(3), x_j^k(4)$ 为白化权函数的转折点。

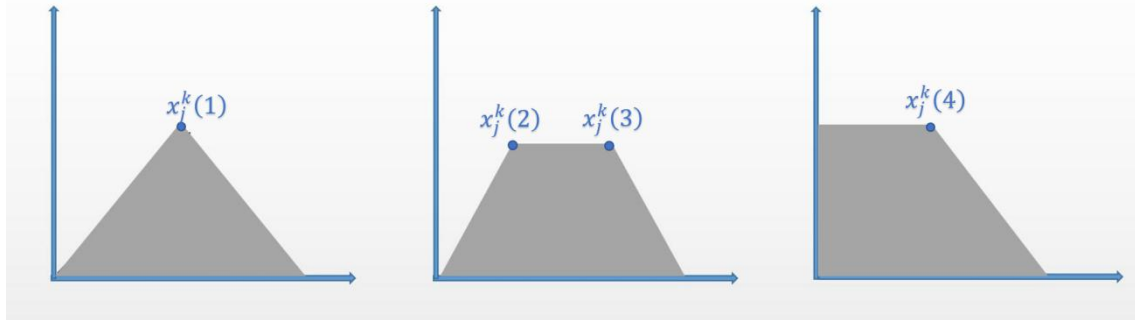


图 6. 白化权函数转折点示意图

灰色聚类的计算方法：

- (1) 求解各 j 指标 k 子类白化权函数的临界值 λ_j^k ；
- (2) 计算各 j 指标 k 子类的权重

$$\eta_j^k = \frac{\lambda_j^k}{\sum_{j=1}^m \lambda_j^k} \quad \#(17)$$

- (3) 计算聚类系数矩阵

$$\sigma_i^k = \begin{bmatrix} \sigma_1^1 & \sigma_1^2 & \sigma_1^3 & \cdots & \sigma_1^s \\ \sigma_2^1 & \sigma_2^2 & \sigma_2^3 & \cdots & \sigma_2^s \\ \sigma_3^1 & \sigma_3^2 & \sigma_3^3 & \cdots & \sigma_3^s \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_n^1 & \sigma_n^2 & \sigma_n^3 & \cdots & \sigma_n^s \end{bmatrix} \#(18)$$

其中

$$\sigma_i^k = \sum_{j=1}^m f_j^k(x_{ij}) \times \eta_j^k \#(19)$$

(4) 如果

$$P = \max_{1 \leq k \leq s} \{\sigma_i^k\} \#(20)$$

则将对象 i 归于灰类 P 。

6.2.3 灰色聚类在近视度数中的应用

在实际计算中, 根据问题一所得出的结果, 本文将近视度数分为六个单项, 分别是用眼环境、近距离工作量、户外活动量、个体参数、遗传和生活习惯。同时设立四个指标, 设立依据见表 4。

表 4. 近视度数评判指标

指标	低度近视	中度近视	高度近视	超高度近视
近视度数	≤ 300	300~600	600~900	≥ 900

由于各单项的数值可能差别较大, 如果直接进行灰聚类分析, 会引起较大的误差。所以在进行计算前, 应该先进行相应的无量纲和归一化处理, 用求得的各个单项指标值除以该项指标的最大值, 进而列出四个指标的中间转折点, 见表 5。

表 5. 各白化权函数中间转折点值

单项项目	低度近视(a)	中度近视(b)	高度近视(c)	超高度近视(d)
用眼环境	0.93	0.81	0.76	0.62
近距离工作量	0.89	0.79	0.63	0.49
户外活动量	0.85	0.76	0.64	0.49
个体参数	0.86	0.75	0.65	0.48
遗传	0.90	0.84	0.76	0.62
生活习惯	0.83	0.74	0.62	0.50

按灰聚类分析法, 我们得到的结果实际上是研究对象在该状态下, 处于何种近视度

数水平的概率，即处于低度近视、中度近视、高度近视或超高度近视的概率，也称近视度数的初始情况。这一结论在对视力的初始状态进行评价的基础上，为下一步用马尔可夫链预测近视度数奠定了基础。

6.2.4 概率转移矩阵与马尔可夫链

设系统的初始向量为 $A_1 = (a_1, b_1, c_1, d_1)$ ，下一年的预测值为 $A_2 = (a_2, b_2, c_2, d_2)$ 。转移概率矩阵 P 为：

$$\begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix}$$

根据马尔可夫链，则下一年的预测可用如下方程表示：

$$a_1 p_{11} + b_1 p_{21} + c_1 p_{31} + d_1 p_{41} = a_2 \#(21)$$

$$a_1 p_{12} + b_1 p_{22} + c_1 p_{32} + d_1 p_{42} = b_2 \#(22)$$

$$a_1 p_{13} + b_1 p_{23} + c_1 p_{33} + d_1 p_{43} = c_2 \#(23)$$

$$a_1 p_{14} + b_1 p_{24} + c_1 p_{34} + d_1 p_{44} = d_2 \#(24)$$

$$p_{11} + p_{12} + p_{13} + p_{14} = 1 \#(25)$$

第3年的预测可用如下方程表示：

$$a_2 p_{11} + b_2 p_{21} + c_2 p_{31} + d_2 p_{41} = a_3 \#(26)$$

$$a_2 p_{12} + b_2 p_{22} + c_2 p_{32} + d_2 p_{42} = b_3 \#(27)$$

$$a_2 p_{13} + b_2 p_{23} + c_2 p_{33} + d_2 p_{43} = c_3 \#(28)$$

$$a_2 p_{14} + b_2 p_{24} + c_2 p_{34} + d_2 p_{44} = d_3 \#(29)$$

$$p_{21} + p_{22} + p_{23} + p_{24} = 1 \#(30)$$

从方程组系数和方程式的结果看，有一定规律性，由此考虑能否通过解方程的形式来求转移概率。列出第四、五、六年的四个方程式。

$$a_3 p_{11} + b_3 p_{21} + c_3 p_{31} + d_3 p_{41} = a_4 \#(31)$$

$$a_3 p_{12} + b_3 p_{22} + c_3 p_{32} + d_3 p_{42} = b_4 \#(32)$$

$$a_3 p_{13} + b_3 p_{23} + c_3 p_{33} + d_3 p_{43} = c_4 \#(33)$$

$$a_3 p_{14} + b_3 p_{24} + c_3 p_{34} + d_3 p_{44} = d_4 \#(34)$$

$$p_{31} + p_{32} + p_{33} + p_{34} = 1 \#(35)$$

$$a_4 p_{11} + b_4 p_{21} + c_4 p_{31} + d_4 p_{41} = a_5 \#(36)$$

$$a_4 p_{12} + b_4 p_{22} + c_4 p_{32} + d_4 p_{42} = b_5 \#(37)$$

$$a_4 p_{13} + b_4 p_{23} + c_4 p_{33} + d_4 p_{43} = c_5 \#(38)$$

$$a_4 p_{14} + b_4 p_{24} + c_4 p_{34} + d_4 p_{44} = d_5 \#(39)$$

$$p_{41} + p_{42} + p_{43} + p_{44} = 1 \#(40)$$

根据方程的特点，将 (21)、(26)、(31)、(36) 组合得到以下方程组：

$$a_1 p_{11} + b_1 p_{21} + c_1 p_{31} + d_1 p_{41} = a_2 \#(41)$$

$$a_2 p_{11} + b_2 p_{21} + c_2 p_{31} + d_2 p_{41} = a_3 \#(42)$$

$$a_3 p_{11} + b_3 p_{21} + c_3 p_{31} + d_3 p_{41} = a_4 \#(43)$$

$$a_4 p_{11} + b_4 p_{21} + c_4 p_{31} + d_4 p_{41} = a_5 \#(44)$$

可用向量写为

$$A \times P_1 = B_1 \#(45)$$

其中

$$P_1 = (p_{11}, p_{12}, p_{13}, p_{14}), B_1 = (a_2, a_3, a_4, a_5)^T$$

同理，(22)、(27)、(32)、(37) 可写为

$$A \times P_2 = B_2 \#(46)$$

其他方程可依次写为

$$A \times P_3 = B_3 \#(47)$$

$$A \times P_4 = B_4 \#(48)$$

将这四个向量写成矩阵形式，即

$$A \times P = B \#(49)$$

则

$$P = A^{-1} \times B$$

这样，转移概率矩阵 P 就可通过逆矩阵求得。

6.2.5 灰色马尔可夫模型的可执行性

根据上文进行总结，可将应用灰色马尔可夫预测近视度数发展情况归结为如下计算过程：

- (1) 根据灰聚类分析，求得近视度数的初始情况。
- (2) 利用 n 年的近视度数发展数据根据上述方法求出转移概率矩阵。
- (3) 运用马尔可夫链预测近视度数未来的发展情况。

故本文提出的灰聚类分析优化后的马尔可夫演化机理模型是可执行的。

6.3 问题三：基于 BP 神经网络的马尔可夫视力预警机制优化模型

6.3.1 BP 网络模型的选取

由于各个影响因素与近视的关系没有很明确的定量权重，因此本文采用 BP 神经网络和模糊评价算法取代灰聚类分析来计算权重。

在本题缺乏数据的背景环境下，此方法将使本文模型能够广泛应用于多样的数据环境。此外，BP 神经网络的优势也十分突出，现阐述如下：

非线性映射能力强。对于任何复杂非线性映射，BP 神经网络均能够较好地拟合。这个特征恰好符合本文所设立的关键因素。由于影响视力的因素有多个，相互之间关系不甚紧密，且每个因素与视力之间的关系十分复杂，难以用纯粹的线性关系刻画影响体系，

故选取 BP 神经网络来实现这些复杂的非线性映射。

泛化能力强。这是指即使对于网络中从未出现过的样本数据，BP 神经网络模型也能够完成从输入到输出的正确映射。

客观性强。区别于层次分析法等需要大量主观确定的数据计算的模型，BP 神经网络模型仅需要对输入的样本进行学习，其客观性较强。

BP 网络也称反向传播神经网络，它能够通过样本数据的训练，不断修正网络的权值和阈值，使误差函数逐渐逼近期望输出^[7]。

6.3.2 BP 神经网络模型的组成

BP 神经网络模型是由输入层、中间层和输出层三部分构成的，其中中间层可以有一层或多层，例如一个 $m \times k \times n$ 的三层 BP 神经网络结构，本文选用 S 型传递函数

$$f(x) = \frac{1}{1 + e^{-x}} \quad \#(50)$$

根据反传误差函数

$$E = \frac{\sum_i (t_i - O_i)^2}{2} \quad \#(51)$$

（其中 t_i 为期望输出， O_i 为网络的计算输出），不断调节网络的权值和阈值使函数 E 达到极小。

6.3.3 本题模型求解

（1）输入输出层的设计

由本文可知，影响近视度数的主要因素为用眼环境系数 θ 、近距离工作量系数 Q 、户外活动时间系数 τ 、个体参数系数 p 、遗传系数 ξ 和健康系数 S 。这 6 个关键因素可通过 12 个待测量获得，分别是光源强度、距光源的距离、光的颜色、近距离工作时间、近距离阅读页数、户外活动时间、身高、体重、性别、年龄、父母患病及携带情况和睡眠时间。由于此预警机制是面向非学者人群，故本题模型所用的遗传算法优化后的 BP 网络结构的输入输出层均为如表 3 中所列的、非学者人群能够理解且易于测量的量。这些量通过计算与推导能够较为准确地得到对近视度数的影响，故近视度数作为输出，所以输入层的节点数为 12 个，输出层的节点数为 1。

（2）中间层设计

有关研究表明，有一个中间层的神经网络，只要中间层节点足够多，就可以以任意精度逼近一个非线性函数。因此，本文采用含一个中间层的三层多输入单输出的 BP 网络建立预测模型。在网络设计过程中，中间层神经元数目的确定十分重要。中间层神经元个数过多，会加大网络计算量并容易产生过度拟合问题；神经元个数过少，则会影响网络性能，达不到预期效果。网络中中间层神经元的数目与实际问题的复杂程度、输入和输出层的神经元数以及对期望误差的设定有着直接的联系。目前，对于中间层中神经元数目的确定并没有明确的公式，只有一些经验公式，神经元个数的最终确定还是需要根据经验和多次实验来确定。本文在选取中间层神经元个数的问题上参照了以下公式：

$$l = \sqrt{n + m} + a \quad \#(52)$$

其中， n 为输入层神经元个数， m 为输出层神经元个数， a 为 $[1, 10]$ 之间的常数。

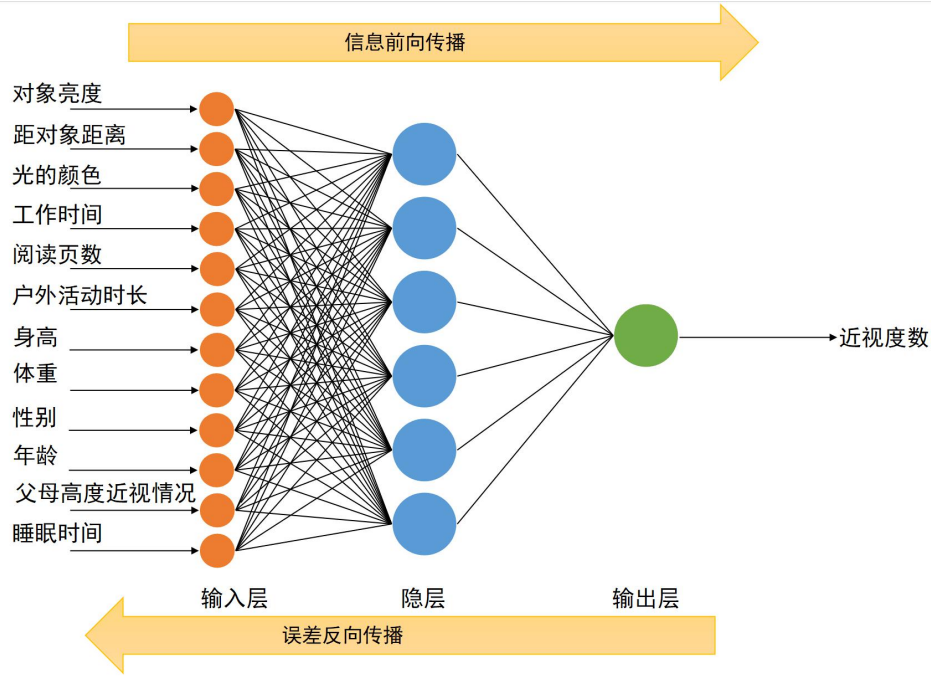


图 7. 近视度数关键影响因素的 BP 网络结构

(3) BP 神经网络模型权重的计算

为了计算各关键因素的权重，BP 神经网络模型首先要求求出输出层与中间层之间计算可得各关键因素的权重，即

$$\Delta v_{kj} = \eta \delta_k a_j \quad (53)$$

其中，

$$\delta_k = y_k(1 - y_k)(t_k - y_k) \quad (54)$$

对于中间层和输入值之间的权值 w ：

$$\Delta w_{ji} = \eta \delta_j a_i \quad (55)$$

$$\delta_j = a_j(1 - a_j) \sum_k \delta_k v_{kj} \quad (56)$$

由此可以得出各参量的权重量。

6.3.4 预警机制说明

由于本题所述模型中输入层与输出层均为易于理解且便于测得的物理量，故此预警机制完全可面向非学者人群。

在输入 12 个具体参量后，模型运用 BP 神经网络模型代替问题二中的灰聚类分析法来计算各项关键因素的权重，将所得权重带入问题二中提出的灰色马尔可夫模型，进而计算出近视度数增加达到某一速度的概率。

通过将近视度数与相应概率的权值乘积并累加，得到近视度数增长量在某一区间的概率系数，为了便于运算，将此系数除以 100，得到标准概率系数 ω 。本文设立最高风

险系数 ω_0 ，经过计算得

$$\omega_0 \in [4.9, 5.6]$$

本文通过对比 ω 和 ω_0 ，设立预警机制，见表 7。

表 6. 预警级别判定表

标准概率系数值	反馈
≤ 4.9	需要继续保持
$4.9 \sim 5.6$	需要稍加留意
≥ 5.6	需要采取措施

6.4 问题四：基于随机森林算法的眼睛视力机器学习模型

鉴于影响视力因素的多样性，经过分析比对，本文选择泛化性强的随机森林算法作为机器学习的基本模型。

6.4.1 随机森林算法的选取

选取随机森林的原因是其模型较简单，适合机器学习。BP 神经网络的训练可能具有一定计算难度，需要设置学习率并进行超参数的微调。而随机森林在训练过程中建立的决策边界很容易理解，更适合推广。

随机森林是一个用随机方法建立的，包含多个决策树的分类器。其输出的类别式由各个树输出的类别的众数而定。随机性主要体现在样本的随机性和分支的随机性两方面。

与其他模型相比，随机森林的优势在于：

- (1) 不需要对关键因素进行筛选，所以能够处理高维度、多因素的数据。
- (2) 由于特征子集可随机选择，结果的可靠性得到了保证。
- (3) 模型运算的结果能够指出主要因素。
- (4) 对于不平衡的数据集而言，随机森林算法能够平衡误差。

6.4.2 可执行信息系统的建立

为了使 AI 预警机制上升到机器学习层面，需要提供一个数据库的设计方案。我们根据上文中所提出的 12 项相关参量及其相关权重，设计出所需的信息系统样式。本文选取对象亮度、与对象的距离、光的颜色、近距离工作时间、近距离阅读页数、户外活动时间、身高、体重、年龄、性别、父母是否患高度近视或携带近视基因和睡眠时间为数据库所含的参量，见图 8。

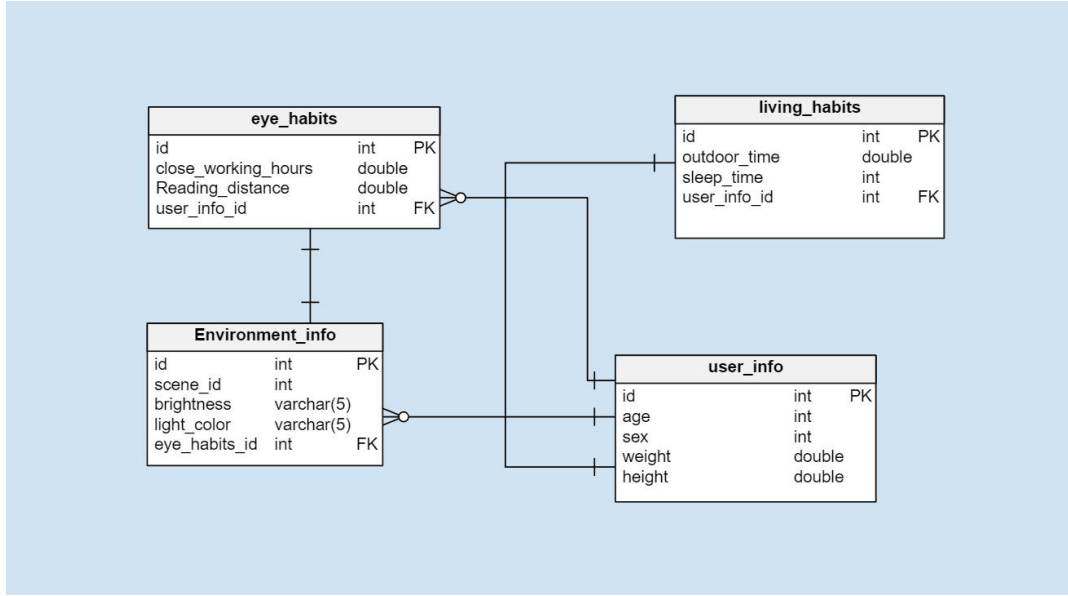


图 8. 数据库设计

6.4.3 随机森林算法

(1) 随机森林算法中样本的随机性

假设我们的估计量为 μ ，样本大小为 N ，从样本中有放回地再抽样 N 个样本。原来每个样本被抽中的概率相同，均为 $1/N$ ，得到新的样本我们称为 *Bootstrap* 样本，重复 B 次之后我们得到 B 个 *Bootstrap* 样本集，在每一个样本集上都有对应的估计量 $\mu^{(b)}$ ，对于 B 个 $\mu^{(b)}$ ，我们可以直接得到标准差、置信区间和偏置等。

(a) 其中对 *Bootstrap* 重复的第 b 次($b = 1, \dots, B$),

步骤一：通过有放回的从 x_1, \dots, x_n 中抽样得到抽样样本 $x^{*(b)} = x_1^*, \dots, x_n^*$ 。

步骤二：根据 $x^{*(b)}$ 计算 $\hat{\mu}^{(b)}$ 。

(b) $F_{\hat{\mu}}(\cdot)$ 的 *Bootstrap* 估计为 $\hat{\mu}^{(1)}, \dots, \hat{\mu}^{(B)}$ 的经验分布函数。

(2) 随机森林算法中分支的随机性

随机森林算法中分支的随机性是指从所有属性里随机选取 K 个属性，选择最佳分割属性作为节点建立 *CART* 决策树。在每个节点决策树分支的方向由基尼系数决定。

基尼系数的计算方法为

$$gini(T) = 1 - \sum p_j^2 = 1 - \sum \left(\frac{n_j}{S}\right)^2 \quad \#(57)$$

$$gini_{split}(T) = \frac{S_1}{S_1 + S_2} gini(T_1) + \frac{S_2}{S_1 + S_2} gini(T_2) \quad \#(58)$$

p_j ——类别 j 在样本 T 中出现的概率

n_j ——样本 T 中类别 j 的个数

S —— T 中样本的个数

式中, $S_1.S_1$ 与 $T_1.T_2$ 关系类似。

(3) 建立决策树系统

重复以上两步 z 次, 建立出 z 棵 $CART$ 决策树。

(4) 计算结果

建立出的 z 个 $CART$ 决策树形成随机森林, 通过投票表决机制, 对数据进行分类。

6.4.4 随机森林算法可行性

随机森林中的每一棵树都可看作是一棵 $CART$ 树, 输入总体数据 n 个, 即构造 n 棵 $CART$ 树, 即输入各种因素的数据量。由于在本题中只需考虑近视度数这一输出特征, 所以运用随机森林算法, 能够得到输出数据的类型, 即近视度数的可能值。故基于随机森林算法的机器学习模型是可行的。

Input Field	Form 1 (Left)	Form 2 (Middle)	Form 3 (Right)
体重/kg	49	49	60
身高/cm	165	175	168
近距离工作时间/h	2	4	5
年龄	18	20	19
看物体距离/cm	30	24	20
性别	女	男	女
户外活动时间/h	2	1	0.5
父母是否高度近视	否	一方	一方
经常接受的亮度	中等	中等	暗
当前近视度数	100	200	400
风险提示	近视加重风险低, 请继续保持哦	近视加重风险中等, 请合理安排作息哦	近视加重风险高, 请及时矫正

图 9. 程序输入界面

6.5 问题五: 模型的实现

要实现本文中的模型, 需要建立完善的数据体系, 实现视力预警程序的普及, 并需要保证屈光率矫正工作的实施。

6.5.1 数据库的建立

要真正意义上实现视力预警机制模型和机器学习模型, 首先需要建立一个较为完善的数据库。

数据库要求至少覆盖所有在读中小學生, 数据内容包括对象亮度、与对象的距离、光的颜色、近距离工作时间、近距离阅读页数、户外活动时间、身高、体重、年龄、性别、父母患高度近视或携带近视基因情况和睡眠时间。

(1) 数据的收集

通过录入教务管理系统信息、各软件厂商提供数据接口的方式实现数据的收集。对于身高、体重、年龄、性别、父母患病或携带近视基因等参量, 进行系统的收集与整理。

教务管理系统通过组织大中小学校学生体检、信息录入等手段实现数据的收集, 这也是较为普及的一种信息收集方式。本文建议在未来将基因检测结果列入教务系统收集

数据中，来完善数据的收集。

软件接口方式则需要以下步骤^[6]：

① 协调多方软件厂商工程师到场，了解所有系统业务流程以及数据库相关的表结构设计等，细节推敲，确定可行性方案

② 编码

③ 测试、调试阶段

④ 交付使用

接口对接方式的数据可靠性与价值较高，一般不存在数据重复的情况；数据通过接口实时传输，满足数据实时性的要求。

（2）数据的开放

实现数据的采集汇聚，开放数据库是最直接的一种方式。两个系统分别有各自的数据库，同类型的数据库之间的数据开放是比较方便的：

① 如果两个数据在同一个服务器上，只要用户名设置的没有问题，就可以直接相互访问。

② 如果两个系统的数据库不在一个服务器上，建议采用连接服务器的形式处理。

开放数据库方式可以直接从目标数据库中获取需要的数据，准确性高，实时性也能得到保证，是最直接便捷的一种方式。

（3）底层数据交换

通过获取软件系统的底层数据交换，基于底层 IO 请求与网络分析等技术，采集目标软件产生的所有数据，将数据转换与重新结构化，输出到新的数据库，供相应软件调用。

基于底层数据交换的数据直接采集方式，直接从各式各样的软件系统中开采数据，获取精确、实时的数据，自动建立数据关联，输出利用率极高的结构化数据，让不同系统的数据源有序、安全、可控的联动流通，提供决策支持，提高运营效率，产生经济价值。

6.5.2 程序的普及

为了实现本文中模型的实际应用，大众对本模型的熟悉程度也十分重要，这就要求提高本文所给程序的普及度。程序应通过医院、学校对亲属、家长的介绍，近视眼镜配置机构及体检机构的宣传来提高普及度。

6.5.3 矫正措施的实施

由于小升初、初升高是屈光不正发生率迅猛上升的关键时期，本文建议有关部门对于小学五六年级和初中三年级的学生进行重点监测，建议对学生的近距离工作量和户外活动时间进行着重监测与控制，避免过度用眼。

同时建议各中小学医务系统配合矫正，对于屈光不正的学生根据预警系统的提示即视进行矫正，并定期复查。

七、模型分析与评价

7.1 模型的优点

(1) 本文采用基于 BP 神经网络的灰色马尔可夫预测模型来解决问题，模型推广性强，在数据较少时能够较为准确地给出预测结果，并为使用者提供相应建议，使得计算简便、与其他方法融合性强。

(2) 本文选取随机森林算法构造机器学习模型，在保留权值计算的同时使用了对于机器语言而言更加友好的随机森林，使得模型便于理解且能够更加客观详细地分析近视度数演化可能。

(3) 本文模型的产品面向非学者人群，机器学习模型的输入输出界面简单易懂。结合教务管理系统和监测政策考虑，贴合实际生活。

7.2 模型的缺点

(1) 本文没有考虑假性近视对于系统的影响，这与实际情况有所出入，不够完善。

(2) 本文在建模过程中缺乏大量数据做基础，在推广到过程中可能会出现某一参数需要稍加修改的情况，在今后的应用中改进。

八、模型的推广

本文采用的基于 BP 神经网络的马尔可夫预测模型可应用于医疗系统、教育系统、征兵选拔系统。

九、参考文献

- [1] 央视，《国家卫健委：全国儿童青少年近视率 53.6% 高中生高达 81%》，<https://www.cn-healthcare.com/article/20190429/content-518441.html>，2019 年 5 月 1 日
- [2] 张国庆，《2018 年中国近视患者人数已经超过 4.5 亿人》，<http://www.199it.com/archives/759520.html>，2019 年 5 月 1 日
- [3] 李景恒 黄萍 易细香，《儿童高度近视与微量元素关系的探讨》，中国斜视与小儿眼科杂志 2006 年第 14 卷第 2 期 47 页
- [4] 东莞华夏眼科，《父母后天性高度近视会遗传？》，https://jingyan.baidu.com/article/86fae346_e56b323c48121a73.html，2019 年 5 月 1 日
- [5] 张晓华 邱延峻，《基于逆阵的路面综合性能马尔可夫预测》，《东北公路》，第 26 卷第 3 期第 7 页
- [6] 百度经验，《如何收集大数据》，<https://jingyan.baidu.com/article/f25ef254bc16b2482c1b821f.html>，2019 年 5 月 3 日
- [7] 佚名，《数据预测之 BP 神经网络具体应用以及 matlab 代码》，<https://blog.csdn.net/u013713294/article/details/53482693>，2019 年 5 月 3 日
- [8] 郭萌，《道路交通事故灰色马尔可夫预测模型》，中国安全科学学报，16 卷第 2 期，第 17 页，2006 年
- [9] 爱尔眼科，《中国青少年用眼行为大数据报告》，http://www.sohu.com/a/276506077_271282，2019 年 5 月 2 日

十、附录

10.1 问卷调查表

序号	户外活动时间/h	性别	身高/cm	体重/kg	当前近视度数	近距离工作时间/h	环境亮度	睡眠时间	父母患高度近视情况 (0/1/2)
1	3	男	181	59	500	4	中等	8	1
2	3	女	159	56	300	2	暗	8	1
3	1	男	169	68	400	2	暗	10	1
4	1	男	174	76	350	2	中等	9	0
5	2	男	166	55	550	4	亮	6	1
6	2	女	171	43	550	4	中等	5	2
7	2	男	163	65	50	1	暗	6	1
8	1	男	175	67	150	3	中等	6	1
9	2	男	179	57	500	1	中等	7	0
10	0	男	170	57	250	3	中等	9	0
11	3	男	177	71	400	2	中等	7	2
12	1	男	173	69	400	4	中等	7	2
13	3	男	183	71	550	4	亮	8	2
14	4	女	155	52	150	4	暗	7	1
15	4	男	175	71	50	3	暗	9	0
16	1	女	157	51	500	1	暗	5	0
17	5	男	184	76	550	3	亮	9	2
18	3	女	171	43	450	4	中等	6	2
19	2	女	155	54	150	3	中等	7	2
20	2	男	165	60	50	2	亮	6	1
21	0	女	166	57	300	4	亮	6	1
22	1	男	174	75	100	4	亮	10	2
23	2	女	160	53	250	3	中等	6	1
24	2	女	165	49	400	1	亮	6	1
25	3	男	175	78	500	4	暗	7	1
26	1	女	158	41	550	1	中等	7	1
27	4	男	184	56	100	3	亮	8	0
28	4	男	176	73	300	3	中等	7	0
29	2	男	171	56	100	3	中等	8	0
30	1	男	166	56	450	1	亮	10	0
31	0	男	175	60	500	2	中等	9	1
32	1	男	168	73	550	2	中等	6	1
33	1	男	175	63	400	2	暗	7	1
34	3	女	174	54	400	2	亮	7	1
35	4	男	177	78	450	1	中等	8	2
36	2	女	168	45	200	2	暗	8	0
37	2	男	175	67	450	3	中等	8	0
38	4	女	175	60	150	3	亮	10	1
39	3	男	179	61	350	1	亮	9	1
40	5	男	170	59	400	3	中等	8	1
41	2	男	172	67	350	3	中等	5	0
42	1	男	180	64	100	1	中等	7	2
43	3	男	164	65	350	4	中等	10	0
44	1	男	167	56	450	1	中等	7	2
45	3	男	175	57	250	2	暗	7	1
46	2	男	180	65	100	3	中等	8	1
47	2	女	173	63	500	3	中等	9	2
48	1	女	167	54	250	1	中等	9	2
49	2	女	172	55	200	4	暗	9	1
50	2	男	180	56	400	4	中等	8	1
51	4	男	179	71	550	3	暗	10	2
52	3	男	183	74	200	3	亮	8	1
53	1	男	174	66	500	3	暗	9	1
54	1	女	171	47	400	2	中等	7	0
55	3	女	161	49	550	4	暗	9	1
56	1	男	165	71	450	3	中等	5	1
57	4	男	171	69	450	1	中等	9	1
58	3	男	174	55	250	1	暗	8	1
59	2	女	157	47	500	2	中等	7	2
60	2	男	183	67	150	1	暗	9	1
61	2	男	178	63	150	4	暗	6	1
62	3	女	163	61	100	3	中等	7	1
63	1	女	159	49	200	4	中等	5	1
64	4	女	157	49	50	1	中等	5	0
65	1	女	165	58	500	4	亮	6	0
66	4	女	172	43	150	3	中等	8	2
67	2	女	161	41	150	4	中等	10	0
68	2	男	172	74	350	2	亮	5	1
69	1	男	169	64	250	2	亮	5	1
70	3	女	156	56	300	4	亮	8	0
71	5	女	177	70	150	3	中等	10	1
72	5	女	172	43	500	2	亮	10	1
73	3	男	166	79	150	1	中等	8	1
74	1	男	174	67	500	4	暗	8	0
75	5	男	180	64	350	3	暗	7	1
76	4	男	166	75	550	3	暗	9	1
77	5	男	179	77	100	3	中等	10	2
78	2	男	183	61	250	4	暗	10	1
79	2	男	168	58	200	2	中等	9	0
80	3	男	180	67	500	4	暗	8	0
81	1	男	183	66	500	2	中等	7	1
82	2	女	173	50	150	4	中等	8	1
83	3	女	158	64	300	4	亮	6	2
84	4	女	156	52	250	4	暗	7	2
85	1	男	178	67	250	2	暗	7	1
86	4	男	165	78	250	4	中等	8	1
87	2	女	163	44	550	2	中等	6	0
88	4	男	164	57	450	3	亮	10	0
89	2	男	173	65	150	3	中等	6	2
90	4	男	174	76	450	3	暗	9	1
91	4	女	174	49	250	1	亮	9	0
92	1	女	155	61	500	1	暗	7	2
93	4	男	175	61	200	1	亮	6	2