

参赛队号 #5225
第九届“认证杯”数学中国

数学建模网络挑战赛
承 诺 书

我们仔细阅读了第九届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

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（包括指导教师）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，我们接受相应处理结果。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：5225

参赛队员（签名）：

队员 1：蔡静

队员 2：党东

队员 3：冉维

参赛队教练员（签名）：

参赛队伍组别（例如本科组）：本科组

参赛队号 #5225

第九届“认证杯”数学中国

数学建模网络挑战赛

编号专用页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：

5225

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

参赛队号 #5225

2016 年第九届“认证杯”数学中国 数学建模网络挑战赛第二阶段论文

题 目 关于青少年存在心理问题的风险评估及预警机制的研究

关 键 词 TOPSIS 灰色关联分析 人工神经网络 风险指数 心理预警机制

摘 要

近年来,世界各地频繁地发生校园霸凌事件,给学生、家长和学校等各方面带来了极大的危害,也引起了社会广泛的关注.想要有效的抑制校园霸凌,就要从根本上去解决问题,也就是需要从学生的心理问题着手处理.但是由于并不是每一个家长、老师都精通心理学,所以建立一个具有普适性、科学性的预警机制就十分必要.

本文首先使用 EXCEL 和 SPSS 对附件所给数据进行了重组、格式转换、缺失值填充、标准化处理和剔除异常样本等处理,并运用 K-means 算法对各个年龄组的样本进行聚类,以检验数据预处理的结果是否可靠,为接下来的模型建立提供根本的数据保证.

针对问题一,本文建立了基于 TOPSIS 和灰色关联分析的风险指数模型,去评估附件所给案例的心理现状.该模型首先使用 R 型聚类分析确定了五个综合评价指标,通过熵权法客观地确定了各个指标的权重;然后分别通过 TOPSIS 和灰色关联分析两种方法求出风险指数 k_1 和 k_2 ,其中 k_1 为案例到正负理想解的距离之比, k_2 为案例对正理想解的灰色加权关联度的倒数与 0.05 的乘积;最后为了得到最终的风险指数模型,随机选取 10 个样本对 k_1 和 k_2 进行对比,得到了 k_1 和 k_2 两个风险指数的值基本一致、互不矛盾的结论,故可以定义最终的风险指数 $k = 0.5 * (k_1 + k_2)$. 根据案例的风险指数值,本文将案例分成了存在心理问题的风险高、较高、中、较低、低五种类别.

针对问题二,根据第一阶段模型得到的六个框架均有四个主要影响因素的结论,本文首先建立了基于 BP 神经网络模型和 RBF 神经网络模型的两种心理预警机制,将每个框架的值视作输出变量,其主要影响因素视作对应的输入变量,通过两种神经网络模型分别进行对输出变量的预测,若六个输出变量的平均值大于 0.17,则对样本的心理问题进行预警;然后将两种模型输出的预测值进行对比,得到了 BP 神经网络预测值的平均相对误差更小的结论,进而将基于 BP 神经网络的心理问题预警机制当作最终的心理问题预警机制;在此基础之上,本文生成了 10 组随机数据对最终的预警机制进行模型仿真,检验出该模型是可以在实际中进行应用的;最后利用该预警机制与实际案例相结合,从附件样本中随机选取三个预测样本作为案例,给出相应的心理问题的解决方案.

考虑到收集数据的误差,本文对风险指数模型作了灵敏度分析,以检验模型的实用性.本文客观公正地描述了模型的优缺点,并对部分缺点提出了改进建议,以期待它们能被完善,并得到推广和应用.

参赛队号: 5225

所选题目: C 题

参赛密码 _____
(由组委会填写)

Abstract

In recent years, campus bullying has occurred frequently all over the world, and do harm to many students, schools and parents, also caused widespread concern in the community. To be effective inhibition of bullying, go up from the fundamental solution to the problem, we need to solve the students' psychological problems. But due to the fact that not every parents, teachers are proficient in psychology, therefore, the establishment of a universal, scientific early warning mechanism is very necessary.

In this paper, we used Excel and SPSS to do the data reorganization, the format conversion fills in missing values, standardization and eliminating abnormal samples, and we used K-means to cluster each age group to make sure that data processing result is reliable.

For the first problem, this paper establishes the risk index model based on TOPSIS and grey relational analysis to assess attachment to the psychological status of the case. The model first uses R cluster analysis identified five comprehensive evaluation index, through the entropy weight method to objectively determine the weight of each index; then, according to TOPSIS and grey correlation analysis we get two risk indexes k_1 and k_2 , k_1 is the ratio of the distance between the positive and negative ideal solutions for the case, k_2 is the reciprocal of the grey weighted correlation degree of the positive ideal solution and the product of 0.05. Finally, in order to get the final risk index model, we randomly selected 10 samples to compare k_1 and k_2 , the get the conclusion that the two risk indexes is consistent and not conflicting. So we can define the final risk index $k = 0.5 * (k_1 + k_2)$. According to cases of risk index value, the case is divided into the existing psychological problems of high risk, little high, medium and little low, low five categories.

For the second problem, according to the conclusion of model of first phase that six framework has four main influence factors. In this paper, we first establish the BP neural network model and RBF neural network model. we regard each frame values regarded as output variables, and the main influencing factors are regarded as the corresponding input variables, by two kinds of neural network models were forecast on the output variable, if six output variable average value greater than 0.17, we will do early warning. Then the two output model of forecast value were compared, the conclusion of value, the average relative error of the BP neural network forecast is smaller, so we use the psychological problems early warning mechanism based on BP neural network as the ultimate psychological problems early warning mechanism. On this basis, we generate 10 random data sets the final early-warning mechanism of simulation model, verify the model can be applied in practice; Finally, using the early warning mechanism and actual case combination, from attached samples in three of the prediction samples were randomly selected as the case, gives the corresponding psychological problems solutions.

Considering the error of data collection, this paper made sensitivity analysis on the risk index model. This paper objectively describes the advantages and disadvantages of the model, and puts forward some advice to improve some of the disadvantages, to expect them to be perfect, and can be popularized and applied.

一、 引言	2
1.1 问题的背景	2
1.2 文献综述	2
1.3 问题重述	3
二、 问题分析	3
2.1 数据的分析	3
2.2 指标的分析	4
2.3 问题一的分析	5
2.4 问题二的分析	6
三、 模型假设	7
四、 符号说明	8
五、 数据预处理	9
5.1 数据重组	9
5.2 数据格式转换	9
5.3 缺失值填充	9
5.4 数据标准化	9
5.5 剔除异常样本	10
5.6 检验数据预处理结果	10
5.7 数据预处理结果概述	12
六、 模型的建立与求解	12
6.1 问题一的模型建立与求解	12
6.1.1 确立综合评价指标体系——R 型聚类分析	12
6.1.2 确定各指标权重——熵权法	15
6.1.3 基于 TOPSIS 方法的风险指数模型	16
6.1.4 基于灰色关联分析的风险指数模型	18
6.1.5 风险指数模型	20
6.2 问题二的模型建立与求解	21
6.2.1 基于 BP 神经网络的青少年心理问题预警机制	21
6.2.2 基于 RBF 神经网络的青少年心理问题预警机制	25
6.2.3 最终的青少年心理问题预警机制	28
6.2.4 青少年心理问题预警机制的模型仿真	30
6.2.5 三个实际案例的心理问题预警及心理问题解决方案	31
七、 模型的分析与应用	35
7.1 风险指数模型的灵敏度分析	35
7.1.1 风险指数 k 对样本到正理想解的距离 s^* 的灵敏度分析	35
7.1.2 风险指数 k 对样本到负理想解的距离 s^0 的灵敏度分析	36
7.1.3 风险指数 k 对样本对正理想解的灰色加权关联度 r 的灵敏度分析	37
7.2 风险指数模型的应用	38
八、 模型的评价与推广	39
8.1 问题一模型的评价与推广（基于 TOPSIS 和灰色关联分析的风险指数模型）	39
8.2 问题二模型的评价与推广（人工神经网络模型）	39
九、 参考文献	40
十、 附录	42
附录一	42
附录二	47

一、引言

1.1 问题的背景

近年来,世界各地频繁的发生校园霸凌事件,给学生、家长和学校等各方面都带来了极大的危害,也引起了社会广泛的关注.人们不由得开始思考——是什么导致了校园霸凌?据研究表明,霸凌行为是由霸凌者的心理问题直接导致的,并且这种行为会导致受害者和旁观者产生心理问题.

《关于加强中小学心理健康教育的若干意见》中指出:“良好的心理素质是人的全面素质中的重要组成部分,是未来人才素质中的一项十分重要的内容.当代学生是跨世纪的一代,他们正处在身心发展的重要时期,大多是独生子女,随着生理、心理的发育和发展,竞争压力的增长,社会阅历的扩展及思维方式的变化,在学习、生活、人际交往和自我意识等方面可能会遇到或产生各种心理问题.有些问题如不能及时解决,将会对学生的健康成长产生不良的影响,严重的会使学生出现行为障碍或人格缺陷.”从这项政策可以看出青少年的心理健康不仅对一个家庭有着重要的意义,更是与国家的兴衰密不可分.他们处在人生的特殊时期,心理正发生着重大的变化,外界的任何刺激都可能造成不可挽回的后果,为了青少年心理的健康发展,抑制校园霸凌迫在眉睫.

想要有效的抑制校园霸凌行为,就必须及时、准确地把握学生的心理现状,了解学生存在心理问题的风险.但是由于并不是每一个老师都精通心理学,所以建立一个具有普适性的模型让老师去正确的评估学生存在心理问题的风险就很有必要了.

了解学生的心理活动并不是最终的目的,预防学生产生心理问题才是学校、家庭和社会所希望的,也是抑制校园霸凌最根本、最有效的方法.所以想要有效的抑制校园霸凌,建立一套青少年心理问题的预警机制势在必行.

1.2 文献综述

(1) 心理问题的内涵

心理问题是指:个体不能够能够适应发展着的环境,不具有完善的个性特征;并且在认知、情绪反应、意志行为处于消极状态,不能保持正常的调控能力.在生活实践中,不能够正确认识自我、悦纳自己、控制自己,不能正确对待外界影响,不能使心理保持平衡协调

(2) 我国青少年心理问题的现状

2004年,谭晖等人的调查《上海市1036名中学生心理健康状况及人格类型》表明,中学生心理障碍发生率为24.9%,其中以预备班和高一年级发生率最高,达30%以上;中学生自杀意念存在率为20%~40%,高中女生最高^[1].

窥一斑而知全豹,仅从2004年上海学生存在心理问题的数据,不难发现当下我国青少年的心理问题形势险峻.2008年的调查表明,在我国青少年当中,有心理和行为问题的小学生约为13%、初中生约15%、高中生约为19%、大学生约25%,并且呈现

随着年级逐年上升的趋势.

(3) 已有预警机制

为了预防学生的心理问题,大多数学校都已经建立了自己的心理问题的预警机制,但是这些预警机制大多数都是单独的由学校内部完成,心理咨询室提供咨询和老师监控构成了预防机制的主体,社会和家庭并没有被纳入预警机制之内.

许志红在《大学生心理危机预警机制》中提出,学生心理问题预警机制的建立不单纯是学校内部的事情,它需要社会、家庭等各方面的支持和配合.形成完整的心理问题预警机制和信息传导流程,这样才能将风险降至最低限度.^[2]他构建了以学校为基础,以家庭、社会为辅助,以医疗机构为保证的预警机制系统.

1.3 问题重述

以学校为基础,以家庭、社会为辅助,以医疗机构为保证的预警机制在大体理论上已经足够完善.但是遗憾的是其中关于心理辅导站和老师具体如何对青少年的心理状况做出评价、心理问题做出预测,以及学校、家庭和社会需要做什么并没有给出详尽的解释.

本文将在此预警机制的基础之上,建立具体的模型.该模型可以根据学生的风险偏好、认知需要和决策风格评估青少年出现心理问题的风险指数以及对青少年的心理问题预测,并针对学生心理现状的不同,给出家庭、学校及社会应做的具体措施.

本文尝试解决以下问题:

(1) 问题一:建立合理的数学模型,评估附件中案例人物存在心理问题的风险;

(2) 问题二:结合第一阶段的模型,建立一套青少年心理问题的预警机制,并结合实际的案例,给出具体的解决其心理问题的方案.

二、 问题分析

2.1 数据的分析

首先,对附件中的五个组的原始数据进行整理可得以下表格.

表1 原始数据整理表

组名	组中年龄段(岁)	组中主要年龄段(岁)	位于主要年龄段中的样本比例	男女比
9岁组	8、9、10、11、12、13	9、10	92.86%	1: 1
11岁组	10、11、12、13、14	11、12	92.14%	1: 1
13岁组	12、13、14、15、16、17	13、14	93.57%	1: 1
15岁组	14、15、16、17、18	15、16	78.21%	1: 1
17岁组	16、17、18、19、20	17、18、19	97.14%	1: 1

从上表中可以看出,原始数据组中每个年龄组的男女比均为 1:1,但是会出现少量非主要年龄段的样本,故本文先将所有样本进行重新组合,将所有年龄介于 8-10 岁的样本组合成新的九岁组,将所有年龄介于 11-12 岁的样本组合成新的 11 岁组,将所有年龄介于 13-14 岁之间的样本组合成新的 13 岁组,将所有年龄介于 15-16 岁之间的样本组合成新的 15 岁组,将所有年龄介于 17-20 岁的样本组合成新的 17 岁组.

接着,考虑到题中有些数据是非数值型数据,为方便后续数据处理,本文进行数据格式转换,将非数值型数据转换成数值型数据.

其次,考虑到附件中缺失的数据可能给后续数据处理带来不便,本文通过 SPSS 对数据进行了替换.

由于数据量纲不同、自身变异或者数值相差较大可能带来较大误差,本文将指标分为成本型指标、效益型指标和其他指标.对于成本型指标和效益型指标,使用离差标准化方法对相关数据进行归一化处理,对于其他指标,使用 Z 标准化方法对相关数据进行归一化处理.

考虑到新的分组中,可能出现异常样本,本文根据拉依达准则对数据可靠性进行检验,并删除每组的异常样本.对于删除后的样本,本文根据 K-means 算法得到了每个青少年组的聚类图,验证了数据处理结果的有效性.

数据预处理流程图如下.

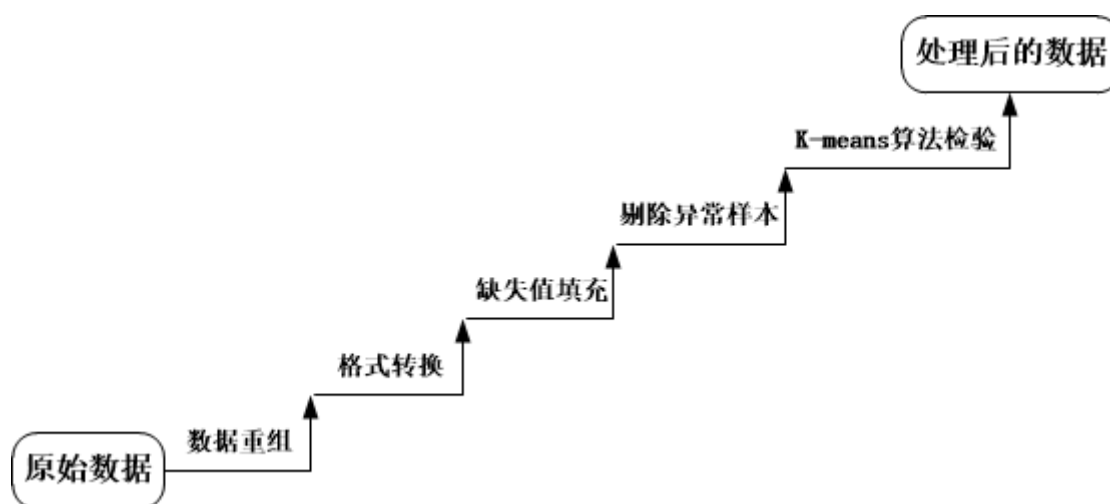


图 1 数据预处理流程图

2.2 指标的分析

为了方便后续数据处理和模型的建立,本文根据指标的定义将各个指标分为成本型指标和效益型指标.对于成本型指标来说,指标值越小越好;对于效益性指标来说,指标值越大越好.各个指标的定义和分类见下表.

指标名称	指标定义	指标分类
风险偏好得分	不同的行为者对风险的态度是存在差异的,一部分人可能喜欢大得大失的刺激,另一部分人则可能更愿意“求稳”.根据个体对风险的偏好将其分为风险回避者、风险追求者和风险中立者.	成本型指标
认知需要得分	认知需求是人类需要和动机之一.泛指个体对事物的追寻、认知、了解的内在动力,如求知欲、好奇心等.一旦认知需要受阻,人不但难以有所作为,而且会处于变态心理状态,失去健康幸福和人生价值.	效益型指标
理智型得分	理智型的决策风格者注重对所获得信息的精细加工,他们往往会对每一个备选行动方案进行认真调查,并给予富有逻辑性的评价,综合权衡每个方案的利弊之后做出决策. ^[11]	效益型指标
直觉型得分	直觉型决策风格者仅对搜集到的信息做粗略了解,不对信息进行加工,只靠感觉和预感做决策. ^[11]	成本型指标
依赖型得分	依赖型决策风格者在面对重大决策时往往不能靠自己做出决策,需要别人的指导和建议才能做决定. ^[11]	效益型指标
回避型得分	逃避型决策风格者通常只会一味的逃避,尽可能的逃避等待着事情自己解决. ^[11]	成本型指标
冲动型得分	冲动型决策风格者的行事风格是速战速决,仅凭一时冲动或者即时感觉在短时间做出决策. ^[11]	成本型指标

2.3 问题一的分析

为了进行综合评价,本文用 R 型聚类分析筛选出 5 个指标,并用熵权法计算每个指标的权重.为了评价附件中每个案例存在心理问题的风险大小,首先,本文尝试通过理想解法,构造心理状况最好的、存在心理问题的风险最小的假想案例 1 和心理状况最差的、存在心理问题的风险最大的假想案例 2,即正理想解和负理想解,并将每个案例与正负理想解的距离之比视作其存在心理问题的风险指数 k_1 .接着,本文尝试通过灰色关联分析法,计算出每个案例对正理想解的灰色加权关联度 r ,并将 $\frac{0.05}{r}$ 视作其存在心理问题的风险指数 k_2 .最后可得每个案例的风险指数为

$$k = \frac{1}{2}(k_1 + k_2) \dots \dots (1)$$

问题一模型流程图如下.

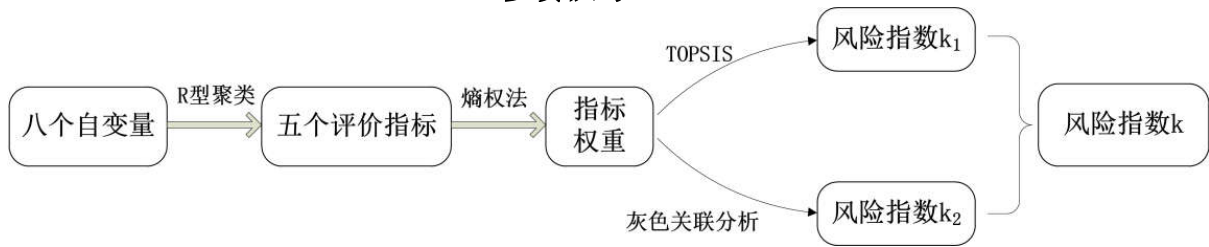


图2 问题一模型流程图

2.4 问题二的分析

通过查阅文献可以知道，针对同一个问题时个体会同时有正向和负向两种看法与心态，只是各自程度不同. 而对于同一问题，无论正负向框架，评价 A 和 B 都是指个体的心理状况的同一倾向.^[3] 将 A 当作个体心理状况正常的倾向, B 当作个体心理状况异常的倾向，由于框架值为 A 时的标准化后的数据值为-0.73，框架值为 B 时标准化后的数据值为 1.07，故当个体六个正负框架的得分的平均值大于 0.17 时，说明个体可能出现心理问题.

根据第一阶段的模型，影响生命教育正向框架的因素主要有风险偏好得分、直觉性得分、理智型得分和认知需要得分，影响生命教育负向框架的因素主要有风险偏好得分、直觉型得分、理智型得分和认知需要得分，影响生活方式正向框架的因素主要有风险偏好得分、理智型得分、依赖型得分和直觉型得分，影响生活方式负向框架的因素主要有风险偏好得分、回避型得分、冲动型得分和依赖型得分，影响娱乐正向框架的因素主要有风险偏好得分、回避型得分、冲动型得分和直觉型得分，影响娱乐负向框架的因素主要有风险偏好得分、依赖型得分、直觉型得分和冲动型得分.

为了建立青少年心理问题的预警机制，本文将六个正负框架值视为因变量，将其影响因素视为与其对应的自变量，构成六个网络，每个网络均有 4 个输入神经元和 1 个输出神经元. 将部分样本的数据作为训练样本集，其余样本的数据作为预测检验样本，建立 BP 神经网络和 RBF 神经网络模型，对比两个模型的预测结果误差，选取平均误差小的模型作为最后的预测模型. 对于一个新的青少年样本，可根据其输入变量的相关数据对输出变量进行预测，当六个预测结果的平均值大于 0.17 时，即提出预警，表示该青少年心理很可能出现问题.

问题二求解思路如下图所示.

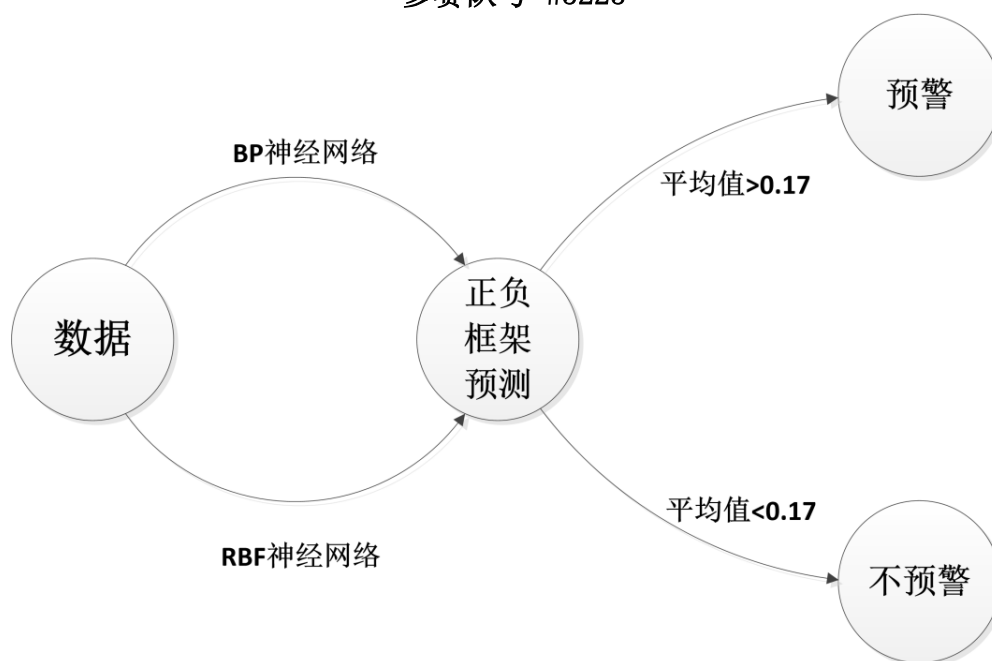


图 3 问题二求解思路图

同时，本文将附件中给出的样本视为实际案例，在预测样本中随机选取了 3 个样本作为实际案例，将数据带入问题二建立的模型中进行预测，得到 3 个样本中可能出现心理问题的样本及其出现心理问题的原因，并根据预测值给出解决心理问题的具体方案。

三、 模型假设

1. 假设附件中的数据真实可靠；
2. 假设除风险偏好、认知需求、决策风格外的其他因素对生命教育、生活方式、娱乐的正负框架的影响可以忽略；
3. 假设人在 20 岁之后心理已经稳定，不会发生较大突变；
4. 假设题中所给数据对应的青少年样本是真实存在的；
5. 假设 R 型聚类后，属于不同类别的指标之间相互独立；
6. 假设在进行青少年心理问题预警时，生命教育正负向框架、生活方式正负向框架和娱乐方式正向框架的重要性是相同的，即在这里不考虑各个框架的权重对预警结果的影响。

四、 符号说明

表 3 符号说明表

符号	解释
W_1	年龄介于 8~10 岁之间的青少年全体
W_2	年龄介于 11~12 岁之间的青少年全体
W_3	年龄介于 13~14 岁之间的青少年全体
W_4	年龄介于 15~16 岁之间的青少年全体
W_5	年龄介于 17~20 岁之间的青少年全体
Y_{11}	生命教育正向框架评价值
Y_{12}	生命教育负向框架评价值
Y_{21}	生活方式正向框架评价值
Y_{22}	生活方式负向框架评价值
Y_{31}	娱乐正向框架评价值
Y_{32}	娱乐负向框架评价值
x_0	年龄
x_1	性别
x_2	风险偏好得分
x_3	认知需要得分
x_4	理智型得分
x_5	直觉型得分
x_6	依赖型得分
x_7	回避型得分
x_8	冲动型得分
G_i	R 型聚类分析得到的第 i 类
C^*/C^0	正/负理想解
ω	权重向量
s_i^*/s_i^0	样本 i 到正理想解的距离/负理想解的距离
k_1	基于 TOPSIS 方法的风险指数
k_2	基于灰色关联分析的风险指数
k	风险指数
ρ	分辨系数
$\xi_i(k)$	灰色关联系数
r_i	样本 i 对正理想解的灰色加权关联度

五、 数据预处理

5.1 数据重组

考虑到原来的分组中存在各个年龄段所含年龄有交叉的现象，为方便后续问题的处理，首先将青少年重新分为 5 组. 即 9 岁组 W_1 （包含 8、9、10 岁的青少年），11 岁组 W_2 （包含 11、12 岁的青少年），13 岁组 W_3 （包含 13、14 岁的青少年），15 岁组 W_4 （包含 15、16 岁的青少年），和 17 岁组 W_5 （包含 17、18、19、20 岁的青少年），并将每个组的青少年重新进行编号.

5.2 数据格式转换

考虑到附件中有的数据为非数值数据，为方便后续处理，本文按照下表中的原则将非数值数据转换为数值数据.

表 4 数据格式转换表

变量名称	原始数据格式	转换后的数据格式
性别	女	1
	男	2
正负框架值	A	1
	B	2

5.3 缺失值填充

观察到原始数据中 15 岁组的序号为 188 的样本的娱乐正向框架值出现了数据缺失，为方便后续处理，考虑先对缺失数据进行填充. 由于该数据在转换格式前属于非数值型数据，故使用同一列的众数对其进行填充，也即该缺失值的填充值为 2.

5.4 数据标准化

考虑到数据量纲不同、自身变异或者数值相差较大可能给后续模型的求解结果造成误差，本文对数据进行标准化处理.

首先将所有变量分为三类：成本型变量，即取值越小越好的变量；效益型变量，即取值越大越好的变量；即非成本型又非效益型的变量，即不能明确判断取值是越大越好还是越小越好的变量.

对于成本型变量，其标准化公式为：

$$a'_{ij} = \frac{-a_{ij} + a_{j\max}}{a_{j\max} - a_{j\min}} \dots\dots(2);$$

对于效益型变量，其标准化公式为：

$$a'_{ij} = \frac{a_{ij} - a_{j\min}}{a_{j\max} - a_{j\min}} \dots\dots(3);$$

对于即非成本型又非效益型的指标，其标准化公式为：

$$a'_{ij} = \frac{a_{ij} - \mu}{\sigma} \quad (\text{其中 } \mu \text{ 为平均值, } \sigma \text{ 为标准差}) \dots\dots(4).$$

5.5 剔除异常样本

为了保证模型结果的准确性，本文将所给数据中出现异常的数据及其对应的样本剔除。

如果实验数据值的总体 a 是服从正态分布的，则由拉依达准则法，有：

$$P(|x - \mu| > 2\sigma) \leq 0.046$$

上式中， μ 与 σ 分别表示正态总体的数学期望和标准差。此时，在实验数据值中出现大于 $\mu + 2\sigma$ 或小于 $\mu - 2\sigma$ 数据值的概率是很小的。因此，根据“小概率事件实际不可能性原理”，将大于 $\mu + 2\sigma$ 或小于 $\mu - 2\sigma$ 的实验数据值视为异常值，予以剔除。

由于经过 Z 标准化方法处理后的数据应该满足正态分布，故将 Z 标准化后小于 -2 或大于 2 的样本数据视为异常值，将其和其对应的样本剔除。

5.6 检验数据预处理结果

为了检验数据预处理结果，判断预处理结果是否可靠，本文通过 K-means 算法对重组后的每个年龄段的样本进行聚类，若聚类后没有离群点，则说明数据预处理结果可靠。

9 岁组的聚类结果如下图所示.

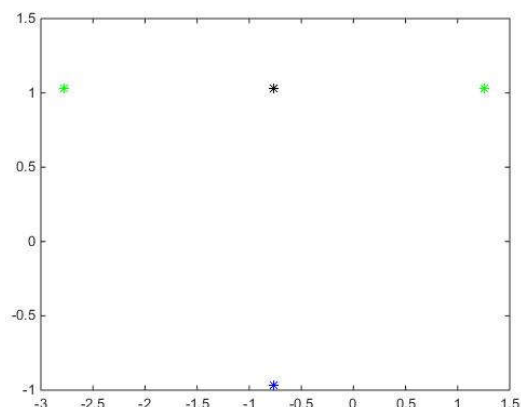


图 4 9 岁组聚类结果图

图中每一个点代表聚出的一个类别. 从聚类结果可以看出, 9 岁组中不存在离群点, 故对 9 岁组的数据预处理是可靠的、有效的.

11 岁组的聚类结果如下图所示.

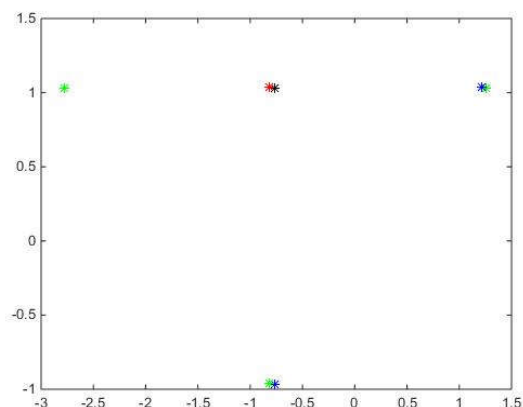


图 5 11 岁组聚类结果图

图中每一个点代表聚出的一个类别. 从聚类结果可以看出, 11 岁组中不存在离群点, 故对 11 岁组的数据预处理是可靠的、有效的.

13 岁组的聚类结果如下图所示.

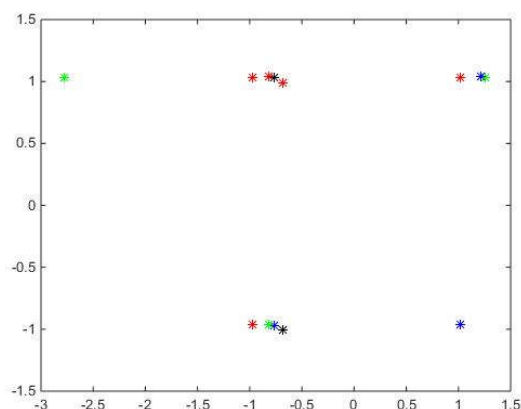


图 6 13 岁组聚类结果图

图中每一个点代表聚出的一个类别. 从聚类结果可以看出, 13 岁组中不存在离群点, 故对 13 组的数据预处理是可靠的、有效的.

15 岁组的聚类结果如下图所示.

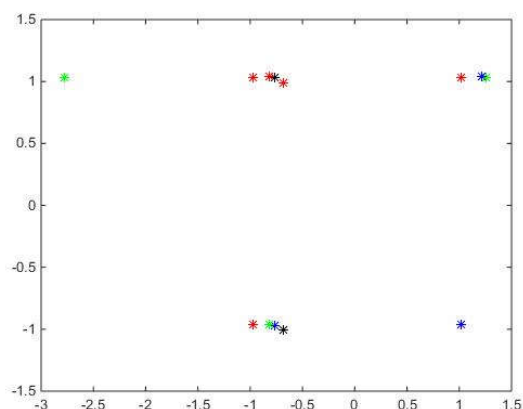


图 7 15 岁聚类结果图

图中每一个点代表聚出的一个类别. 从聚类结果可以看出, 15 岁组中不存在离群点, 故对 15 组的数据预处理是可靠的、有效的.

17 岁组的聚类结果如下图所示.

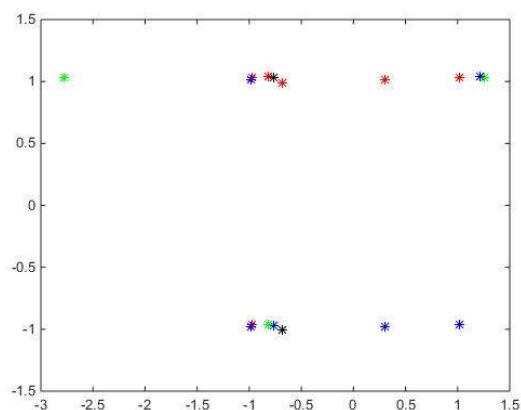


图 8 17 岁聚类结果图

图中每一个点代表聚出的一个类别.从聚类结果可以看出,17 岁组中不存在离群点,故对 17 组的数据预处理是可靠的、有效的.

综合上述聚类结果可知,数据预处理结果是可靠的,有效的.

5.7 数据预处理结果概述

经过数据预处理,一共得到有效样本 1077 个,其中 9 岁组的有效样本有 196 个,11 岁组的有效样本有 227 个,13 岁组的有效样本有 226 个,15 岁组的有效样本有 178 个,17 岁组的有效样本有 250 个.

六、模型的建立与求解

6.1 问题一的模型建立与求解

6.1.1 确立综合评价指标体系——R 型聚类分析

由于题中自变量个数过多,直接将每个自变量 ($x_0, x_1, \dots, x_7, x_8$) 当作综合评价模型中的指标可能造成计算的不便,并且某些自变量之间可能存在较强的相关性,故本文首先使用 R 型聚类分析,将所有自变量分为五类,在每一类中选取一个自变量作为最终进行综合评价时的指标.

将 x_0, x_1, \dots, x_8 视为待聚类变量,它们之间的 Pearson 相关系数计算公式为:

$$r = \frac{N \sum a_i b_i - \sum a_i \sum b_i}{\sqrt{N \sum a_i^2 - (\sum a_i)^2} \sqrt{N \sum b_i^2 - (\sum b_i)^2}} \dots\dots(5)$$

其中 N 为样本数量, a_i, b_i 分别为各个样本对应的这两个变量的取值.

利用五个年龄段的所有未被剔除的青少年的数据,通过 MATLAB 编程计算这 9 个变

量之间的相关系数得下表:

表 4 相关性系数表

	年龄	性别	风险偏好	认知需要	理智型	直觉型	依赖型	回避型	冲动型
年龄	1.00	0.00	-0.01	-0.01	-0.04	-0.01	-0.05	0.00	0.03
性别	0.00	1.00	-0.13	0.02	0.00	-0.04	0.02	-0.03	0.00
风险偏好	-0.01	-0.13	1.00	-0.09	0.07	-0.03	-0.06	0.07	-0.08
认知需要	-0.01	0.02	-0.09	1.00	-0.34	0.09	0.07	-0.28	0.14
理智型	-0.04	0.00	0.07	-0.34	1.00	-0.24	0.22	0.05	-0.16
直觉型	-0.01	-0.04	-0.03	0.09	-0.24	1.00	-0.19	0.21	-0.36
依赖型	-0.05	0.02	-0.06	0.07	0.22	-0.19	1.00	-0.34	0.18
回避型	0.00	-0.03	0.07	-0.28	0.05	0.21	-0.34	1.00	-0.33
冲动型	0.03	0.00	-0.08	0.14	-0.16	-0.36	0.18	-0.33	1.00

从上述相关系数矩阵中可以看出,某些自变量之间确实存在相关性,因此可以考虑从这些变量中选取少数几个具有代表性的变量作为指标进行后续的综合评价.

根据标准化后的数据进行 R 型聚类分析,其中变量间相似性度量采用相关系数,类间相似性度量的计算使用类平均法,公式为:

$$D(g_1, g_2) = \frac{1}{n_1 n_2} \sum_{a_i \in G_1} \sum_{a_j \in G_2} d(a_i, a_j) \dots \dots (6)$$

其中 g_1, g_2 为两个不同的变量, a_i, a_j 分别为各个样本对应的这两个变量的取值, n_1, n_2 分别为这两个变量所涉及的样本总数量.

通过 MATLAB 编程求解可得聚类树形图如下

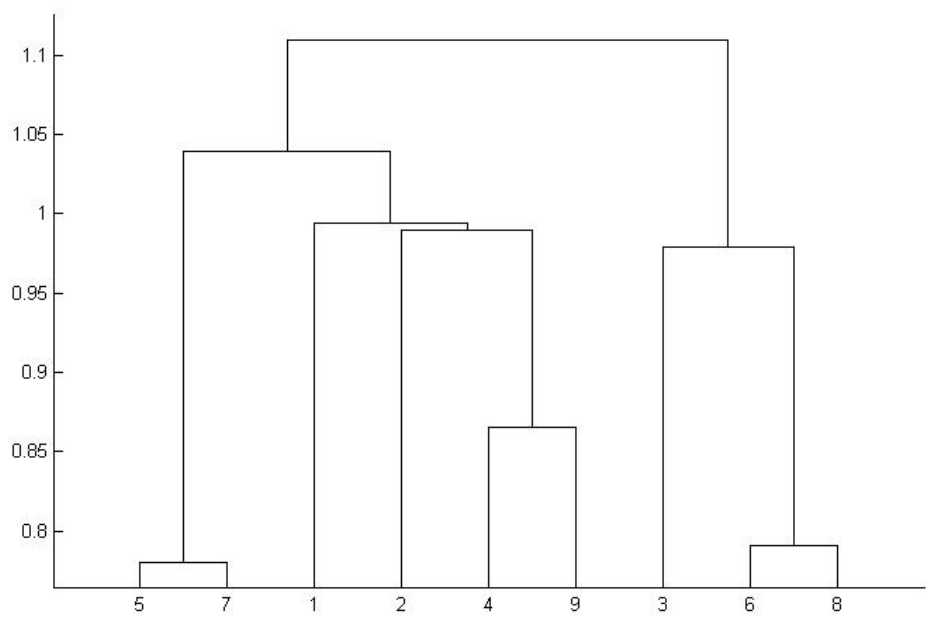


图 9 聚类树形图

距离值为 0.97 时，可分为 5 类，如下表所示：

表 5 聚类结果表

类别名	所含变量
G_1	x_4, x_6
G_2	x_0
G_3	x_1
G_4	x_3, x_8
G_5	x_2, x_5, x_7

在这六类中的每一类选取一个指标，这样就从 14 个指标中选定了 6 个可用作综合评价的指标，如下表所示.

参赛队号 #5225
表 6 综合评价指标表

指标符号	指标解释
x_0	年龄
x_1	性别
x_3	认知需要得分
x_4	理智型得分
x_2	风险偏好得分

根据聚类分析的结果，可以看出：

- (1) 理智型得分和依赖型得分的关联较大，在进行后续处理时可以只使用生理智型得分的信息，暂不使用依赖型得分的信息；
- (2) 风险偏好得分、认知需要得分、理智型得分这三个变量关联较小，属于不同的类别；
- (3) 性别、年龄与其他变量的关联较小，均可以单独成为一类；
- (4) 可以根据选出的这 5 个指标对附件中的青少年样本进行综合评价。

6.1.2 确定各指标权重——熵权法

为方便叙述，此节中均用 w 代替题中的自变量和因变量的取值。

步骤一：在不改变数值的范围的情况下对数据进行修正：

对于效益型指标 $w_i(j)$ ，令

$$w'_i(j) = \left[\frac{w_i(j) - \min_i w_i(j)}{\max_i w_i(j)} \right] * 0.9 + 0.1 \dots (7)$$

对于成本型指标 $w_i(j)$ ，令

$$w'_i(j) = \left[\frac{\max_i w_i(j) - w_i(j)}{\max_i w_i(j)} \right] * 0.9 + 0.1 \dots (8)$$

其中， $w_i(j)$ 为第 i 个青少年样本的第 j 个指标经过归一化后的标准化数值。

步骤二：计算第 j 个指标下第 i 个青少年样本占该指标的比重。

$$p_{ij} = \frac{w'_i(j)}{\sum_i w'_i(j)} \dots (9)$$

步骤三：计算第 j 项指标的熵值。

$$e_j = -k \sum_i p_{ij} \ln p_{ij} \dots (10) ,$$

其中调节系数 $k = 1/\ln(5) > 0$

步骤四：计算第 j 项指标的差异系数. 对第 j 项指标，指标的差异越大，对综合评价的作用就越大，熵值就越小. 定义差异系数：

$$g_j = \frac{1 - e_j}{5 - E_e} \dots\dots (11)$$

$$(\text{当 } 0 \leq g_j \leq 1, \sum_{j=1}^5 e_j = 1)$$

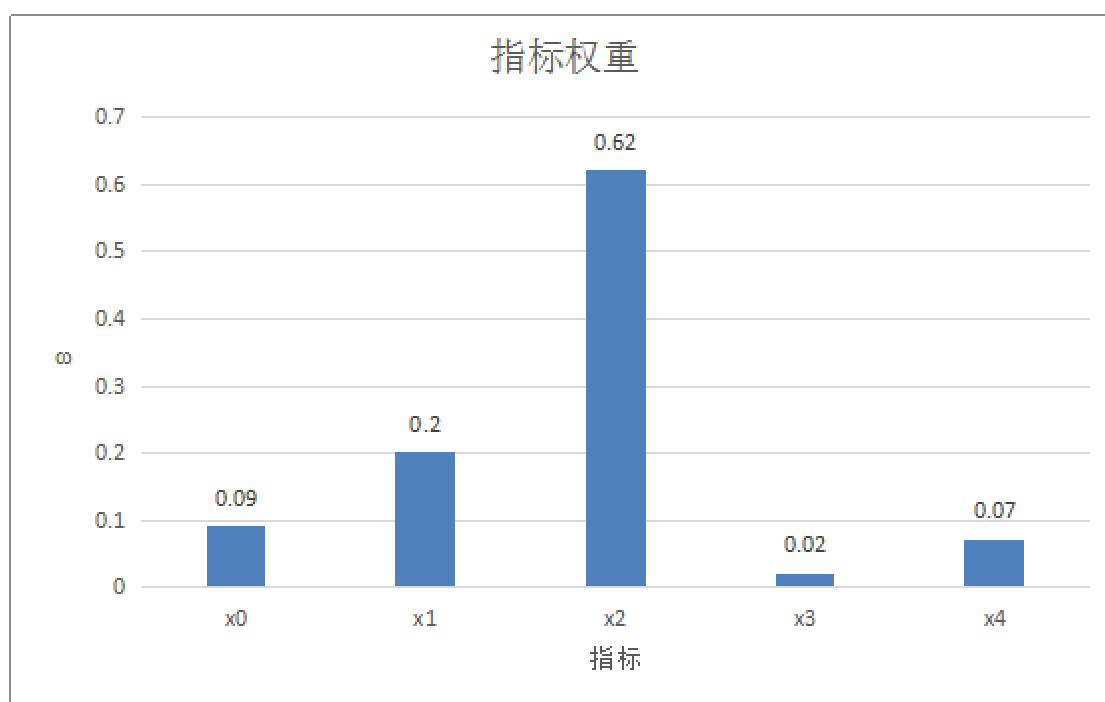
步骤五：求权值.

$$\omega_j = \frac{g_j}{\sum_{j=1}^5 g_j} = g_j \quad (j=1, 2, \dots, 5) \dots\dots (12)$$

由于 $\sum_{j=1}^5 g_j = 1$ ，故差异系数等于权值.

通过 MATLAB 编程计算可得各个组的五个指标的权重见下图.

图 9 指标权重图



从上面的图中可以看出，风险偏好得分在综合评价时占的比重最大，最重要；认知需要得分在综合评价时占的比重最小.

6.1.3 基于 TOPSIS 方法的风险指数模型

本文通过 TOPSIS 方法，评估附件中案例存在心理问题的风险，具体步骤如下.

步骤一：用标准 0-1 变换的方法求得规范决策矩阵.

在进行综合评价之前先对 R 型聚类分析得到的五个指标的值进行规范化处理，这一步主要有以下三个作用：

(1) 由于指标值有成本型和效益型两个类型，上述五个指标放在同一个表中不便于直接从数值大小判断方案的优劣，因此需要对数据进行预处理，使得表中任一指标下存在心理问题越严重的青少年样本变换后的指标值越大；

(2) 消除量纲；

(3) 五个指标值的数值差别有时很大，为了直观，更为了便于进行后续的综合评价，需要把这五个指标对应的数值归一化，即把表中数值均变换到[0, 1]区间上.

由于数据预处理时已经使用了标准 0-1 变换方法对数据进行了归一化处理，故这一步可以不用再重复进行.

步骤二：构成加权规范矩阵.

设权重向量为 $\omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5)^T$ ，则加权规范矩阵为 $C = (c_{ij})_{m \times n}$ ，其中：

$$c_{ij} = \omega_j \times b_{ij}, \quad i=1, 2, \dots, m, \quad j=1, 2, \dots, n.$$

在这一步中，取权向量为熵权法得到的权向量

$$\omega = (0.09, 0.20, 0.62, 0.02, 0.07)^T,$$

得到的部分样本的加权规范矩阵见附录二.

步骤三：确定正理想解和负理想解.

设正理想解 C^* 的第 j 个指标值为 c_j^* ，负理想解 C^0 的第 j 个指标值为 c_j^0 ，则

$$\text{正理想解 } c_j^* = \begin{cases} \max c_{ij}, & j \text{ 为效益型指标} \\ \min c_{ij}, & j \text{ 为成本型指标} \end{cases} \quad j=1, 2, \dots, n$$

$$\text{负理想解 } c_j^0 = \begin{cases} \min c_{ij}, & j \text{ 为效益型指标} \\ \max c_{ij}, & j \text{ 为成本型指标} \end{cases} \quad j=1, 2, \dots, n$$

通过 MATLAB 编程求解可得：

$$\text{正理想解 } C^* = [0.1409, 0.2043, 0.6200, 0.0245, 0.0698],$$

$$\text{负理想解 } C^0 = [-0.2467, -0.1981, 0.0, 0].$$

步骤四：计算各方案到正理想解和负理想解的距离.

各个样本 d_i 到正理想解的距离为

$$s_i^* = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^*)^2}, \quad i=1, 2, \dots, m \dots\dots (13)$$

各个样本 d_i 到负理想解的距离为

$$s_i^0 = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^0)^2}, i=1, 2, \dots, m \dots \dots (14)$$

通过 MATLAB 编程计算得到各个青少年样本到正负理想解的距离，部分距离数据见附录二。

步骤五：建立基于 TOPSIS 方法的风险指数模型。

根据上述步骤得到的每个样本到正负理想解的距离，定义样本存在心理问题的风险指数：

$$k_1 = \frac{s^*}{s^0} \dots \dots (15)$$

k_1 越大，样本存在心理问题的风险也就越大。

通过 MATLAB 编程求得每个样本存在心理问题的风险指数 k_1 ，所有样本的风险指数 k_1 结果见图，部分样本数据见附录二。

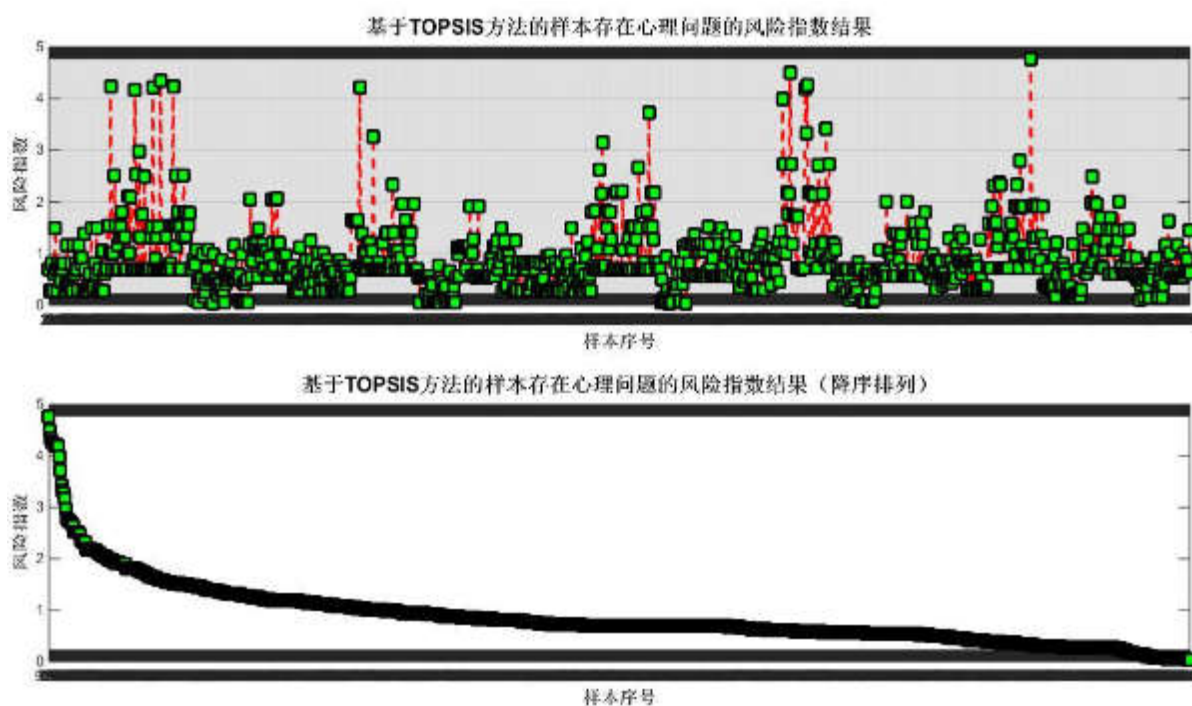


图 10 基于 TOPSIS 方法的样本存在心理问题的风险指数结果

从上图中可以看出：

- (1) 附件中的样本存在心理问题的风险指数 k_1 大多集中在 0-2 之间；
- (2) 风险指数值大于 2 的案例数目相对总体数目来说较少，及存在心理问题的风险高的案例数目相对总体数目来说较少。

6.1.4 基于灰色关联分析的风险指数模型

为了建立新的风险指数模型，便于与基于 TOPSIS 方法的风险指数模型作对比，增强风险指数模型的可靠性和有效性，本文使用灰色关联分析，得到了每个样本对正理想解 C^*

参赛队号 #5225

的灰色加权关联度，根据灰色加权关联度得到了第二个风险指数 k_2 。具体建模步骤如下：

步骤一：确定评价对象和评价标准。

将附件中所有未被剔除的样本作为评价对象，R 型聚类分析得到的指标作为评价指标。参考数列为：

$$C_0 = C^* = [0.1409, 0.2043, 0.6200, 0.0245, 0.0698]$$

比较数列为每个待评价样本的指标值组成的数列。

步骤二：确定各指标对应的权重值。

用熵权法得到的权重作为指标的权重值，即权重向量为：

$$\omega = [0.09, 0.20, 0.62, 0.02, 0.07]$$

步骤三：计算灰色关联系数：

$$\xi_i(k) = \frac{\min_s \min_t |C_0(t) - C_s(t)| + \rho \max_s \max_t |C_0(t) - C_s(t)|}{|C_0(k) - C_i(k)| + \rho \max_i \max_t |C_0(t) - C_s(t)|} \dots\dots (16)$$

为比较数列 C_i 对参考数列 C_0 在第 k 个指标上的关联系数，其中 $\rho = 0.5$ 为分辨系数。并称 $\min_s \min_t |C_0(t) - C_s(t)|$ 、 $\max_s \max_t |C_0(t) - C_s(t)|$ 分别为两级最小差及两级最大差。

步骤四：计算灰色加权关联度。

灰色加权关联度的计算公式为

$$r_i = \sum_{k=1}^n \omega_i \xi_i(k) \dots\dots (17),$$

其中： r_i 为第 i 个评价对象对正理想解的灰色加权关联度。

步骤五：建立基于灰色关联分析的风险指数模型。

根据上述每个样本对于正理想解的灰色加权关联度，定义样本存在心理问题的风险指数：

$$k_2 = \frac{0.05}{r} \dots\dots (18)$$

k_2 越大，样本存在心理问题的风险也就越大。

通过 MATLAB 编程求得每个样本存在心理问题的风险指数 k_2 ，所有样本的风险指数 k_2 结果见图，部分样本数据见附录二。

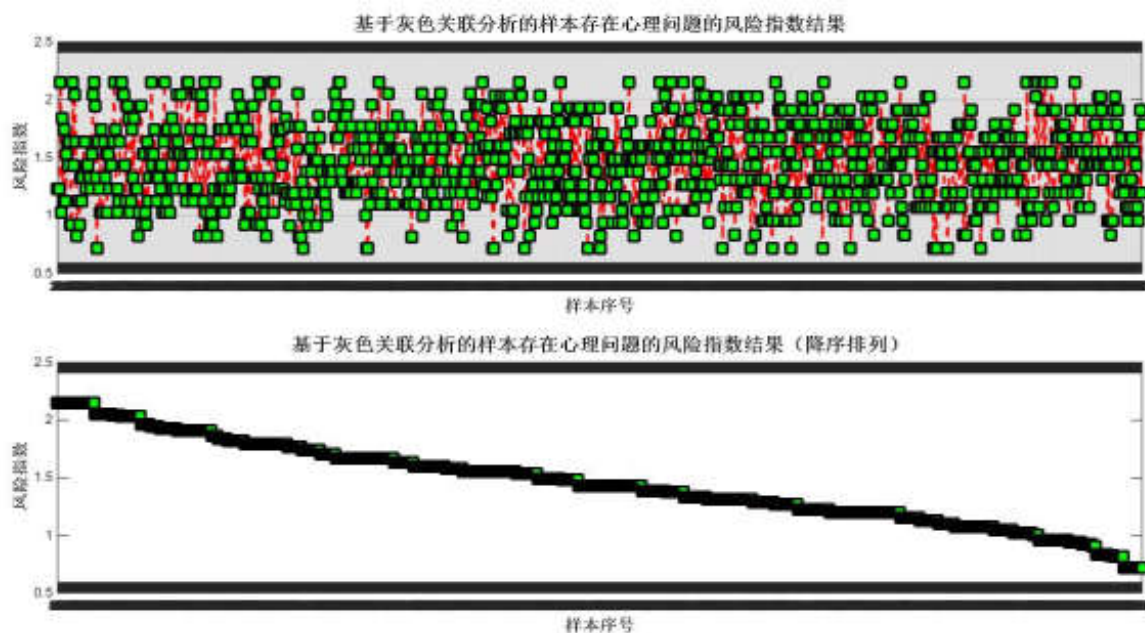


图 11 基于灰色关联分析的样本存在心理问题的风险指数结果

从上图中可以看出：

- （1）附件中的样本存在心理问题的风险指数 k_2 大多集中在 0.05-2 之间；
- （2）风险指数值大于 2 的案例数目相对总体数目来说较少，及存在心理问题的风险高的案例数目相对总体数目来说较少。

6.1.5 风险指数模型

为了得到最终的风险指数模型，首先将上述两个模型得到的风险指数 k_1 和 k_2 进行对比：随机选取 10 个样本，对比他们的风险指数 k_1 在所有的风险指数 k_1 中的排序和他们的风险指数 k_2 在所有的风险指数 k_2 的排序，可得以下对比结果图。

参赛队号 #5225

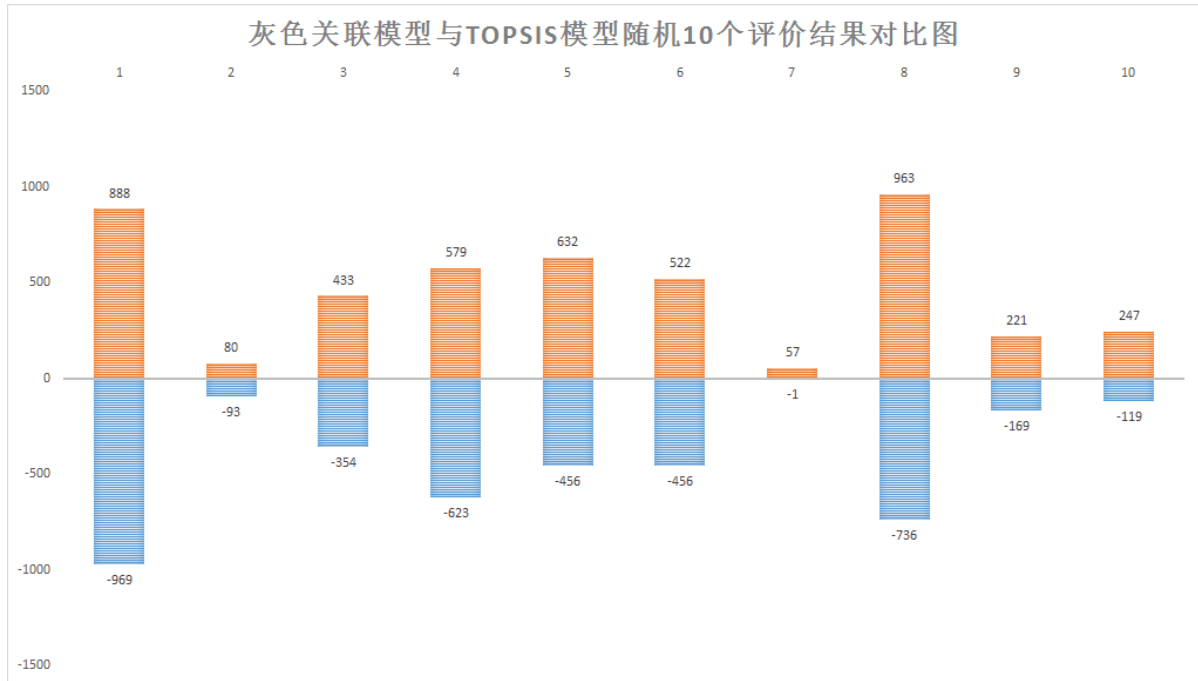


图 12 TOPSIS 和灰色关联分析结果对比图

从对比图中可以看出，两种模型得到的两个风险指数值基本是一致的、不矛盾的，故可以定义

$$k = 0.5 * (k_1 + k_2) \dots\dots (20)$$

作为最后的风险指数值。

k 值越大，案例人物存在心理问题的风险越大。

部分样本的风险指数 k 值的结果见附录二。

可以根据案例的风险指数值将所有案例分为存在心理问题的风险高、较高、中等、较低、低五种类别，具体分类原则如下：

- (1) 风险指数值大于 2 的案例属于存在心理问题的风险高的案例；
- (2) 风险指数值介于 1.5-2 的案例属于存在心理问题的风险较高的案例；
- (3) 风险指数值介于 0.8-1.5 的案例属于存在心理问题的风险中等的案例；
- (4) 风险指数介于 0.6-0.8 的案例属于存在心理问题的风险较低案例；
- (5) 风险指数小于 0.6 的案例属于存在心理问题的风险低的案例。

具体分类结果将在本文第七节、对风险指数模型进行灵敏性分析后给出。

6.2 问题二的模型建立与求解

6.2.1 基于 BP 神经网络的青少年心理问题预警机制

为了建立青少年心理问题的预警机制，首先考虑对青少年的生命教育正负框架值、生活方式正负框架值和娱乐正负框架值进行预测。

为了方便叙述, 本节以生命教育正向框架值的预测为例, 其余几个框架值的预测方式与生命教育正向框架值的预测方式相同, 可直接类推。

根据第一阶段的模型, 生命教育正向框架值的因素主要有风险偏好得分、直觉型得分、理智型得分和认知需求得分, 故可以将这四个影响因素当作自变量, 将生命教育正向框架值当作因变量, 通过自变量的取值预测因变量的取值. 四个自变量及一个因变量见下图。

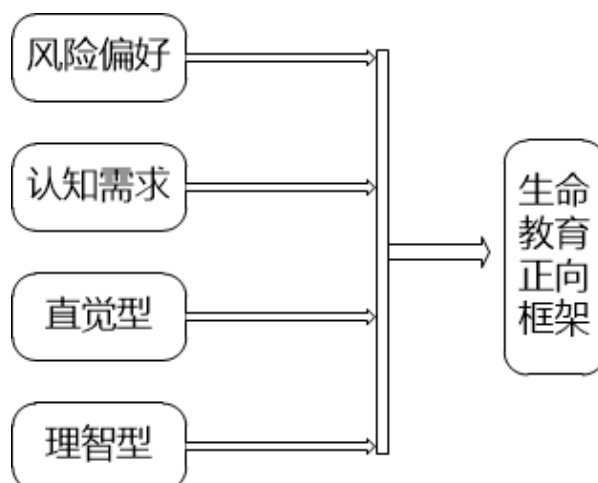


图 13 自变量及因变量展示图

本文首先尝试通过 BP 神经网络预测因变量的取值, 步骤如下:

步骤一: 建立 BP 神经网络, 将风险偏好得分、直觉型得分、理智型得分和认知需求得分当作输入变量, 生命教育正向框架值当作期待输出变量, 可得 BP 网络模型结构如下图。(以两个隐含层为例)

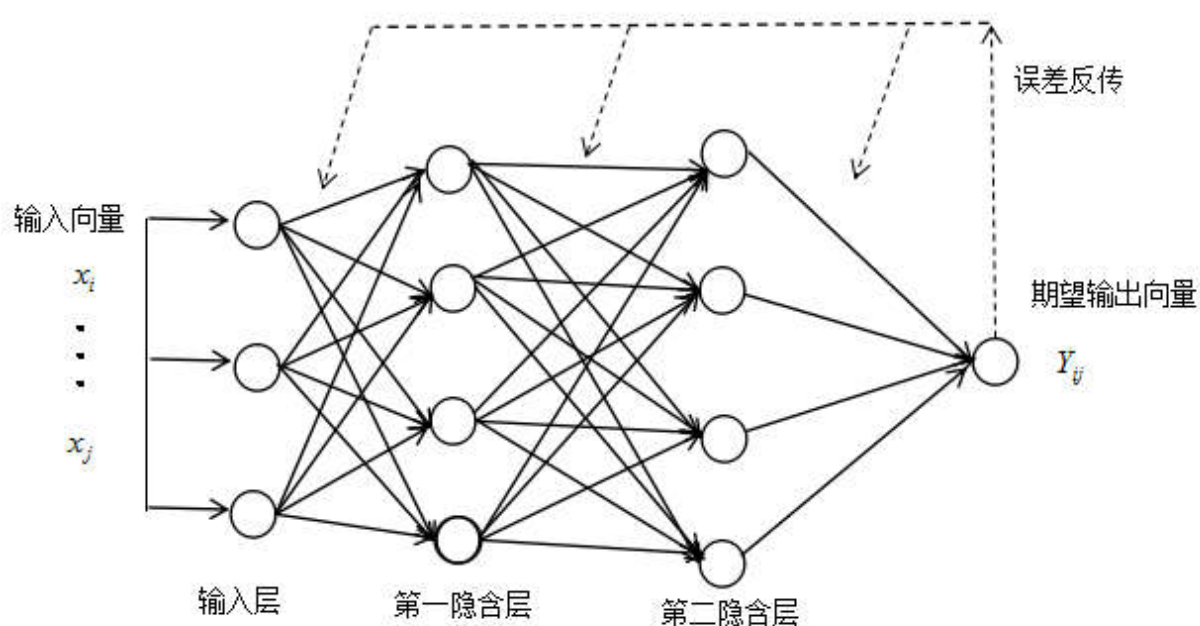


图 14 BP 网络模型结构图

步骤二: 对权系数 w_{ij} 置初值。

参赛队号 #5225

对各层的权系数 W_{ij} 置一个较小的非零随机数，但其中 $W_{i,n+1} = -\theta$ 。

步骤三：输入一个样本 $X = (X_1, X_2, \dots, X_4, 1)$ ，以及对应期望输出 Y 。

其中 X_1, X_2, X_3, X_4 为该样本对应的自变量的取值，期望输出 Y 为生命教育正向框架值的预测值。

步骤四：求各层的学习误差 d_i^k 。

对于输出层有 $k = m$ ，有

$$d_i^m = X_i^m(1 - X_i^m)(X_i^m - Y_i) \dots \dots (21) ;$$

对于其他各层，有

$$d_i^k = X_i^k(1 - X_i^k) \sum_1 W_{li} d_i^{k+1} \dots \dots (22) .$$

步骤五：修正权系数 W_{ij} 和阈值 θ 。

根据 $\Delta W_{ij} = -\eta \cdot d_i^k \cdot X_j^{k-1}$ 有：

$$W_{ij}(t+1) = W_{ij}(t) - \eta \cdot d_i^k \cdot X_j^{k-1} \dots \dots (23) ;$$

根据 $W_{ij}(t+1) = W_{ij}(t) - \eta \cdot d_i^k \cdot X_j^{k-1} + \alpha \Delta W_{ij}(t)$ 有

$$\begin{aligned} \Delta W_{ij}(t) &= -\eta \cdot d_i^k \cdot X_j^{k-1} + \alpha \Delta W_{ij}(t-1) \dots \dots (24) \\ &= W_{ij}(t) - W_{ij}(t-1) \end{aligned}$$

步骤六：当求出了各层各个权系数之后，可按给定品质指标判别是否满足要求。如果满足要求，则算法结束；如果未满足要求，则返回步骤三执行。

这个学习过程，对于任一给定的样本 $X_p = (X_{p1}, X_{p2}, \dots, X_{pn,1})$ 和期望输出 $Y_p = (Y_{p1}, Y_{p2}, \dots, Y_{pn,1})$ ， Y 都要执行，直到满足所有输入输出要求为止。

步骤七：将附件中的 1000 个样本作为训练样本，其余 77 个样本作为预测检验样本，将其通过 BP 神经网络得到的预测值与其实值作对比，可得 BP 神经网络在预测生命教育正向框架值时的相对误差和平均误差。

同理可得 BP 神经网络在预测其他五个框架值时的相对误差和平均误差。

取预测样本在六个框架的预测值的平均值，若平均值 > 0.17 ，则说明该样本可能出现心理问题，故进行预警；否则暂时不进行预警。

通过 MATLAB 编程求解可得六个框架值的 BP 人工神经网络的训练误差曲线图如下。

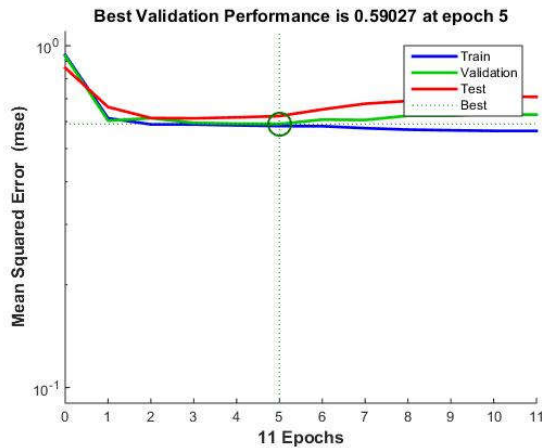


图 15 生命教育正向框架训练误差曲线图

从图中可以看出，在用 BP 神经网络模型对生命教育正向框架值进行预测时，误差曲线在连续 5 次检验后不下降，说明 training set 训练的误差已经不再减小，故模型的参数设置是合理的。

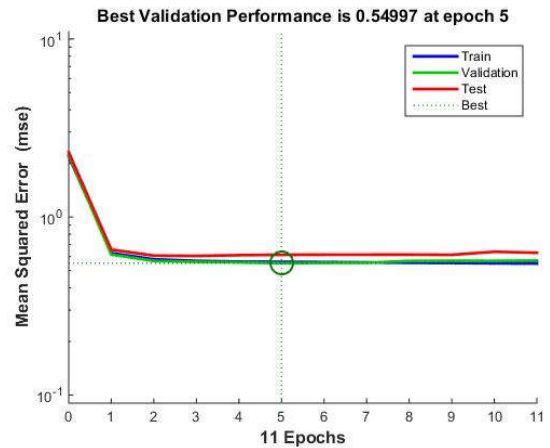


图 16 生命教育负向框架训练误差曲线图

从图中可以看出，在用 BP 神经网络模型对生命教育负向框架值进行预测时，误差曲线在连续 5 次检验后不下降，说明 training set 训练的误差已经不再减小，故模型的参数设置是合理的。

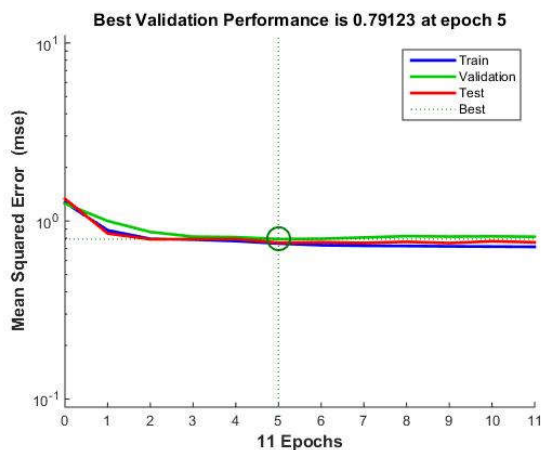


图 17 生活方式正向框架训练误差曲线图

从图中可以看出，在用 BP 神经网络模型对生活方式正向框架值进行预测时，误差曲线在连续 5 次检验后不下降，说明 training set 训练的误差已经不再减小，故模型的参数设置是合理的。

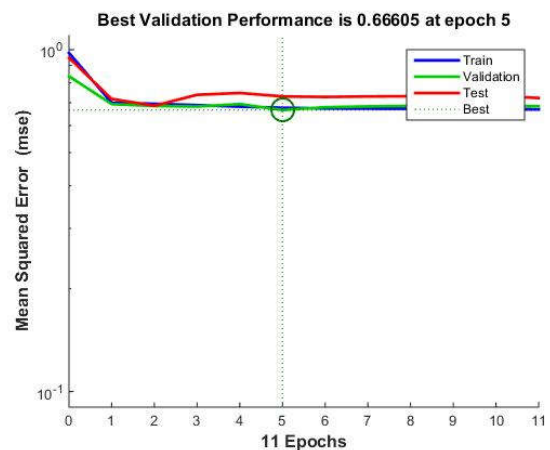


图 18 生活方式负向框架训练误差曲线图

从图中可以看出，在用 BP 神经网络模型对生活方式负向框架值进行预测时，误差曲线在连续 5 次检验后不下降，说明 training set 训练的误差已经不再减小，故模型的参数设置是合理的。

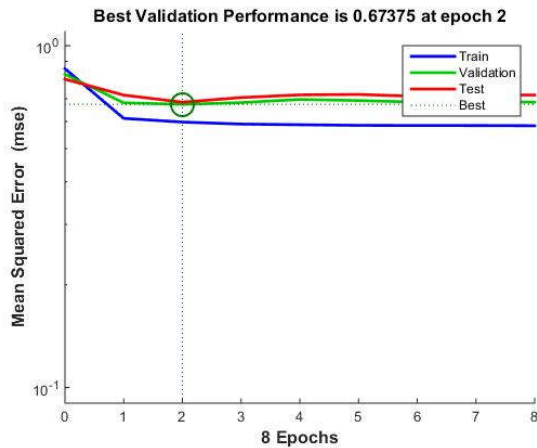


图 19 娱乐正向框架训练误差曲线图

从图中可以看出，在用 BP 神经网络模型对娱乐正向框架值进行预测时，误差曲线在连续 5 次检验后不下降，说明 training set 训练的误差已经不再减小，故模型的参数设置是合理的。

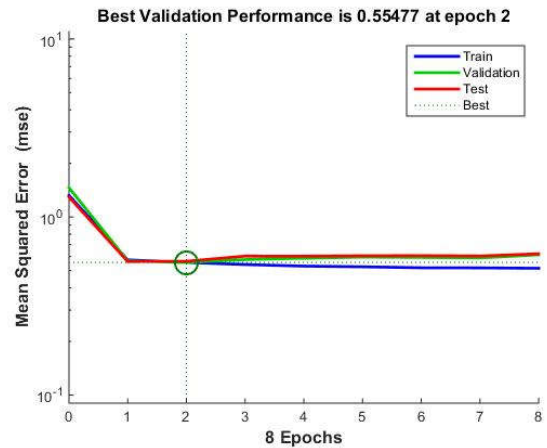


图 20 娱乐负向框架训练误差曲线图

从图中可以看出，在用 BP 神经网络模型对娱乐负向框架值进行预测时，误差曲线在连续 5 次检验后不下降，说明 training set 训练的误差已经不再减小，故模型的参数设置是合理的。

6.2.2 基于 RBF 神经网络的青少年心理问题预警机制

为了与 BP 神经网络的预测值形成对比，本文同时建立了基于 RBF 神经网络的青少年心理问题预警机制。同样以生命教育正向框架值的预测为例，基于 RBF 神经网络的青少年心理问题预警机制的建立步骤如下：

步骤一：建立 RBF 神经网络。

如图所示，RBF 神经网络通常是一种三层前向网络，第一层是输入层，由信号源节点组成；第二层为隐含层，其节点基函数是一种据部分布的、对中学径向对称衰减的非负非线性函数；第三层为输出层。

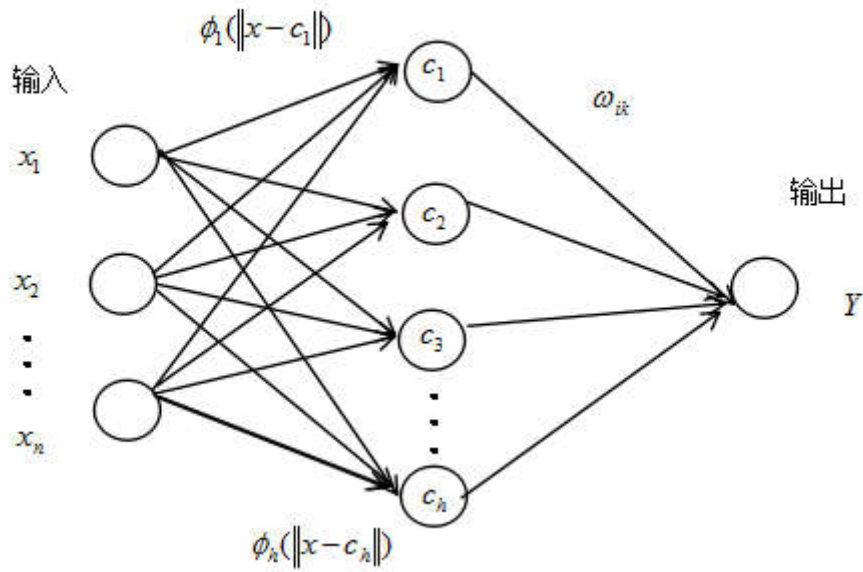


图 21 RBF 网络模型结构图

步骤二：初始化.

随机选择 h 个不同训练样本作为初始聚类中心 $\{c_i(0)\}_{i=1}^h$ ，并令 $n = 0$.

步骤三：抽取样本.

在输入空间中以某种概率抽取样本 x ，作为第 n 次迭代的输入向量.

步骤四：相似匹配.

准则确定 $k(x)$ 的值：

$$k(x) = \arg \min_i \|x(n) - c_i(n)\|, \quad i = 1, 2, \dots, h \quad c_i(n+1) = \frac{1}{N_i} \sum_{x \in W_i(n)} x, \quad i = 1, 2, \dots, h$$

其中， $c_i(n)$ 表示第 i 个径向基函数在第 n 次迭代时的中心.

步骤五：更新.

利用下述规则调整径向基函数的中心：

$$c_i(n+1) = \frac{1}{N_i} \sum_{x \in W_i(n)} x, \quad i = 1, 2, \dots, h$$

其中， N_i 为第 i 个聚类域 $W_i(n)$ 中包含的样本数

步骤六：继续.

如果 $c_i(n+1) \neq c_i(n)$ ，转步骤二，否则聚类结束，转步骤六.

步骤七：根据各聚类中心之间的距离确定各隐节点的宽度， $\sigma_i = kd_i$ ，其中 d 为第 i 个数据中心与其最近的数据中心之间的距离，即 $d_i = \min_j \|c_i - c_j(k)\|$ ， k 是重叠系数.

步骤八：确定输出权值 $\omega = [\omega_1, \omega_2, \dots, \omega_h]^T$.

使用最小均方算法直接计算：

$$\omega = G^+ d$$

$$G^+ = (G^T G)^{-1} G^T$$

步骤九：将附件中的 1000 个样本作为训练样本，其余 77 个样本作为预测检验样本，将其通过 RBF 神经网络得到的预测值与其实值作对比，可得 RBF 神经网络在预测生命教育正向框架值时的相对误差和平均误差。

同理可得 RBF 神经网络在预测其他五个框架值时的相对误差和平均误差。

取预测样本在六个框架的预测值的平均值，若平均值 > 0.17 ，则说明该样本可能出现心理问题，故进行预警；否则暂时不进行预警。

通过 MATLAB 编程求解可得六个框架值的 RBF 人工神经网络的性能图图如下。

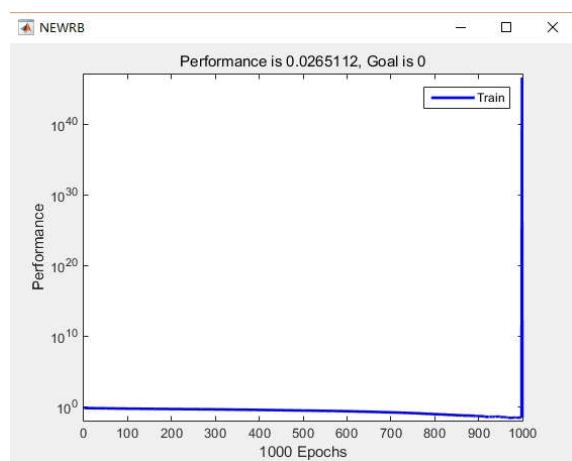


图 22 生命教育正向框架性能图

一般认为，最终结果在 $0-10^{-3}$ 之间时，RBF 神经网络模型性能较好.从图中可以看出，在对生命教育正向框架值进行预测时，RBF 神经网络的性能较差.

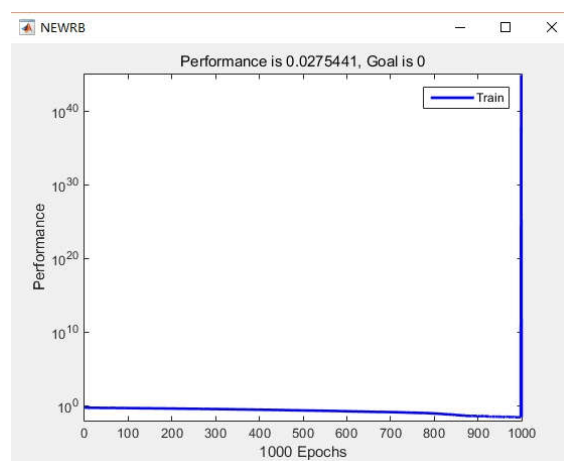


图 23 生命教育负向框架性能图

一般认为，最终结果在 $0-10^{-3}$ 之间时，RBF 神经网络模型性能较好.从图中可以看出，在对生命教育负向框架值进行预测时，RBF 神经网络的性能较差.

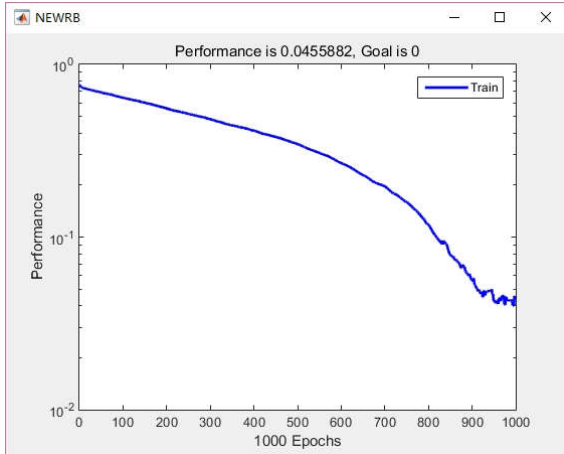


图 24 生活方式正向框架性能图

一般认为,最终结果在 $0-10^{-3}$ 之间时,RBF 神经网络模型性能较好.从图中可以看出,在对生活方式正向框架值进行预测时,RBF 神经网络的性能较差.

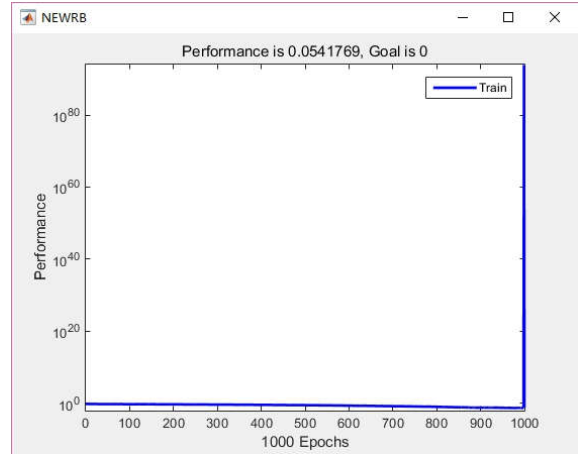


图 25 生活方式负向框架性能图

一般认为,最终结果在 $0-10^{-3}$ 之间时,RBF 神经网络模型性能较好.从图中可以看出,在对生活方式负向框架值进行预测时,RBF 神经网络的性能较差.

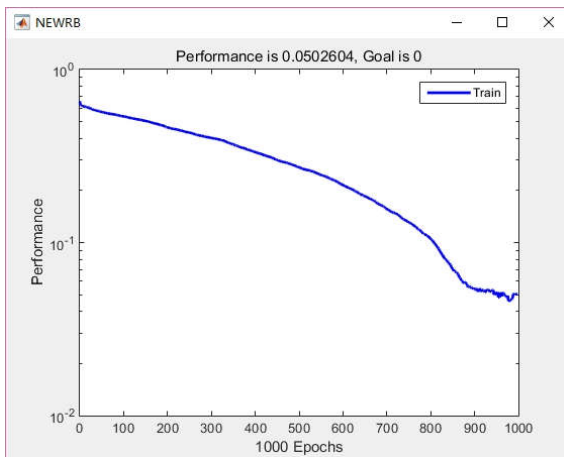


图 26 娱乐正向框架性能图

一般认为,最终结果在 $0-10^{-3}$ 之间时,RBF 神经网络模型性能较好.从图中可以看出,在对娱乐正向框架值进行预测时,RBF 神经网络的性能较差.

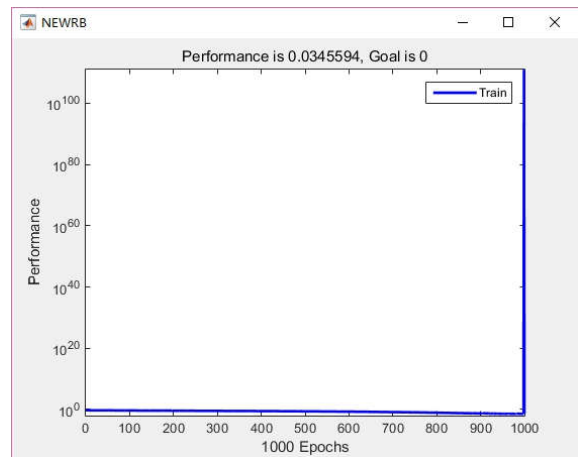


图 27 娱乐负向框架性能图

一般认为,最终结果在 $0-10^{-3}$ 之间时,RBF 神经网络模型性能较好.从图中可以看出,在对娱乐负向框架值进行预测时,RBF 神经网络的性能较差.

6.2.3 最终的青少年心理问题预警机制

为了得到相对误差尽量小的各个框架值的预测值,本文将 RBF 神经网络输出的预测值与 BP 神经网络输出的预测值进行对比,用平均误差小的模型作为最终的预测模型.

通过 MATLAB 编程求解可得 BP 神经网络对于 77 个预测样本的六个框架预测值的平均相对误差和 RBF 神经网络对于 77 个预测样本的六个框架预测值的平均相对误差如下表和图所示.

表 7 框架预测值平均相对误差表

框架名称	BP 神经网络预测值的平均相对误差	RBF 神经网络预测值的平均相对误差	最终预测模型
生命教育正向框架	10.75%	2305.67%	BP 人工神经网络模型
生命教育负向框架	30.82%	4712.86%	BP 人工神经网络模型
生活方式正向框架	-9.39%	4778.81%	BP 人工神经网络模型
生活方式负向框架	34.24%	8146.02%	BP 人工神经网络模型
娱乐正向框架	-9.85%	-4855.62%	BP 人工神经网络模型
娱乐负向框架	42.59%	5972.72%	BP 人工神经网络模型

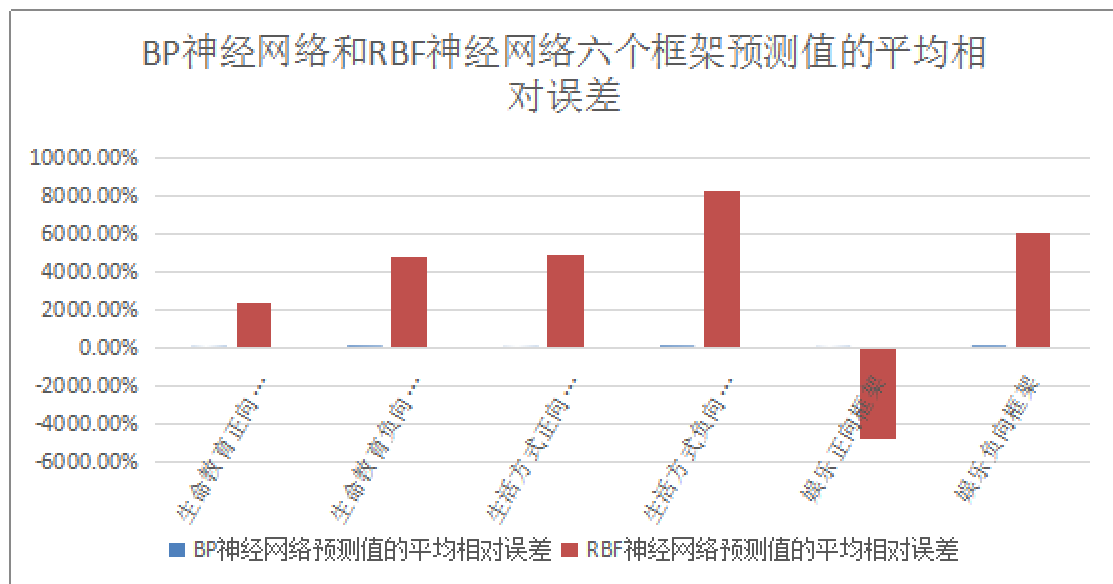


图 28 框架预测值平均相对误差图

通过上述表和图可以看出，可用 BP 人工神经网络模型预测的六个框架值，不使用 RBF 人工神经网络模型预测任何框架值。

故对于一个青少年样本，将他的风险偏好得分、直觉性得分、理智型得分和认知需要得分输入预测生命教育正向框架值的 BP 人工网络模型中，将他的风险偏好得分、直觉性得分、理智型得分和认知需要得分输入预测生命教育负向框架值的 BP 人工网络模型中，将他的风险偏好得分、理智型得分、依赖型得分和直觉型得分输入预测生活方式正向框架值的 BP 人工网络模型中，将他的风险偏好得分、回避型得分、冲动型得分和依赖型得分输入预测生活方式负向框架值的 BP 人工网络模型中，将他的风险偏好得分、回避型得分、冲动型得分和直觉型得分输入预测娱乐正向框架值的 BP 人工神经网络模型中，将他的风险偏好得分、依赖型得分、直觉型得分和冲动型得分输入预测娱乐负向框架值的 BP 人工神经网络模型中，可得该样本的六个框架值的预测值。计算六个预测值

参赛队号 #5225

的平均值，若平均值大于 0.17，则说明该样本可能出现心理问题，进行预警，否则不进行预警。

6.2.4 青少年心理问题预警机制的模型仿真

这里通过 MATLAB 生成了 10 组随机数，每组随机数中都包含了风险偏好得分、认知需要得分、理智型得分、直觉型得分、依赖型得分、回避型得分、冲动型得分的数值各一个。具体仿真数据见下表。

表 8 仿真数据表

样本序号	风险偏好得分	认知需要得分	理智型得分	直觉型得分	依赖型得分	回避型得分	冲动型得分
1	0.29	0.52	0.00	0.85	0.19	0.95	0.45
2	0.86	0.10	0.40	0.25	0.05	1.00	0.15
3	0.79	0.52	0.95	0.80	0.71	0.90	0.15
4	0.57	0.59	0.95	0.40	0.67	1.00	0.80
5	0.50	0.52	0.65	0.75	0.71	0.65	0.10
6	1.00	0.60	0.55	0.05	1.00	0.55	0.70
7	0.71	0.06	0.60	0.70	0.86	0.40	0.10
8	0.29	0.79	0.35	0.70	0.19	0.60	0.25
9	0.29	0.89	0.75	0.65	0.86	1.00	0.15
10	0.29	0.65	0.60	0.05	0.10	0.85	0.30

将每组随机数当作一个仿真样本，根据前文建立的青少年心理问题预警机制模型可得他们的六个框架值及其平均值以及是否进行预警的结果表和结果图如下。

表 9 仿真结果表

样本序号	生命教育正向框架预测值	生命教育负向框架预测值	生活方式正向框架预测值	生活方式负向框架预测值	娱乐正向框架预测值	娱乐负向框架预测值	平均值	是否进行预警
1	0.33	0.57	-0.17	0.21	-0.08	0.18	0.17	不预警
2	-0.29	-0.15	-0.19	0.01	-0.39	-0.18	-0.20	不预警
3	0.28	0.13	-0.02	0.06	-0.15	-0.32	0.00	不预警
4	0.20	0.37	0.03	0.57	-0.32	0.12	0.16	不预警
5	0.18	0.26	-0.05	0.19	-0.10	0.07	0.09	不预警
6	-0.30	-0.23	-0.16	0.15	-0.64	-0.14	-0.22	不预警
7	0.02	0.11	-0.13	0.12	-0.13	-0.23	-0.04	不预警
8	0.05	0.48	-0.07	0.25	-0.07	0.19	0.14	不预警
9	0.27	0.45	0.44	0.83	-0.09	0.34	0.37	预警
10	-0.15	0.37	0.04	0.22	-0.08	-0.14	0.04	不预警



图 29 仿真结果图

从仿真结果表和结果图中可以看出：

- （1）用于仿真的 10 个假想样本中，有 9 个样本不进行心理问题预警，1 个样本进行心理问题预警；
- （2）即使某些样本不需要进行心理问题预警，但该样本仍存在一定的心理问题；
- （3）本节建立的心理问题预警机制模型是实用的。

6.2.5 三个实际案例的心理问题预警及心理问题解决方案

通过以上模型的求解过程，本文得到了 77 个预测样本的六个框架的预测值，在这里选取其中的三个样本，将其作为实际案例，判断其是否需要心理问题预警及其出现心理问题的原因，并给出心理问题的解决方案。

具体每个案例的心理问题分析及解决方案见下表。

表 10 案例 A 心理问题分析及解决方案表

案例 A 心理问题分析及解决方案表	
案例 A 生命教育正向框架预测值	0.034
案例 A 生命教育负向框架预测值	0.287
案例 A 生活方式正向框架预测值	0.067
案例 A 生活方式负向框架预测值	0.257
案例 A 娱乐正向框架预测值	0.126
案例 A 娱乐负向框架预测值	0.050
案例 A 六个预测值的平均值及是否进行预警	平均值为 0.137，不进行预警
案例 A 心理问题解决方案	<p>从上述结果中可以看出，虽然该案例的六个预测值的平均值小于 0.17，不进行心理问题的预警，但是该案例的生命教育负向框架和生活方式负向框架的预测值大于 0.17，说明该案例的心理很可能在生命教育和生活方式上出现了问题，为了解决这两方面的问题，本文提出以下解决方案：</p> <ol style="list-style-type: none"> 1. 学校方面：定期开展与生命教育和健康生活相关的活动，提倡生命正向教育和健康的生活方式；教师及时对该生进行正确的引导，帮助纠正学生不健康的生命价值观和生活方式；班干部积极帮助该生改正不良生活习惯。 2. 家庭方面：对该生在健康生活和生命教育上加强重视，在健康生活方式方面父母做好表率作用，营造良好的家庭氛围有利于该生身心的健康发展； 3. 社会方面：营造积极健康的社会环境，让该生能够切身感受到社会的正能量，帮助学生树立正确的三观； 4. 教师和家长及时交流，共同监督，以免其受外界刺激心理产生突变；

表 11 案例 B 心理问题分析及解决方案表

案例 B 心理问题分析及解决方案表	
案例 B 生命教育正向框架预测值	0.050
案例 B 生命教育负向框架预测值	0.177
案例 B 生活方式正向框架预测值	0.049
案例 B 生活方式负向框架预测值	0.136
案例 B 娱乐正向框架预测值	0.219
案例 B 娱乐负向框架预测值	0.041
案例 B 六个预测值的平均值及是否进行预警	平均值为 0.112，不进行预警
案例 B 心理问题解决方案	<p>从上述结果中可以看出，虽然该案例的六个预测值的平均值小于 0.17，不进行心理问题的预警，但是该案例的生命教育负向框架和娱乐正向框架的预测值大于 0.17，说明该案例的心理很可能在生命教育和娱乐上出现了问题，为了解决这两方面的问题，本文提出以下解决方案：</p> <ol style="list-style-type: none"> 1. 学校方面：定期开展生命教育活动，加强对不健康的娱乐方式的管理；教师应当及时对学生引导，纠正不正确的观念和娱乐方式；班干部积极帮助该生改正不良娱乐方式。 2. 家庭方面：家长应当和学校形成统一战线，加强对不健康娱乐方式的监管，家长在监督其娱乐方式时要以身作则。对该生生命教育加以重视，为学生营造良好的家庭生活环境。 3. 社会方面：要加强对不适合学生的娱乐场所的监管，为学生营造积极健康的社会环境。

表 12 案例 C 心理问题分析及解决方案表

案例 C 心理问题分析及解决方案表	
案例 C 生命教育正向框架预测值	0.057
案例 C 生命教育负向框架预测值	0.622
案例 C 生活方式正向框架预测值	0.077
案例 C 生活方式负向框架预测值	0.212
案例 C 娱乐正向框架预测值	0.070
案例 A 娱乐负向框架预测值	0.092
案例 C 六个预测值的平均值及是否进行预警	平均值为 0.188，进行预警
案例 C 心理问题解决方案	<p>从上述结果中可以看出，该案例的六个预测值的平均值大于 0.17，进行心理问题的预警，并且该案例的生命教育负向框架和生活方式负向框架的预测值大于 0.17，说明该案例的心理很可能在生命教育和娱乐上出现了问题，为了解决这两方面的问题，本文提出以下解决方案：</p> <ol style="list-style-type: none"> 1. 学校方面：心理咨询室将其列为重点心理疏导对象，情况严重则应当向更加权威的社会医疗机构请求援助；教师要重点关注该生的生命教育和生活方式，及时把握该生的心理动态；班委帮助老师对该生进行监督。 2. 家庭方面：积极配合咨询室和老师的工作，不要给该生太大压力，在尊重该生为前提，主动向医疗机构寻求帮助。 3. 社会方面：设置心理咨询平台，为该生提供权威的心理治疗。

预警机制及解决方案流程图如下.

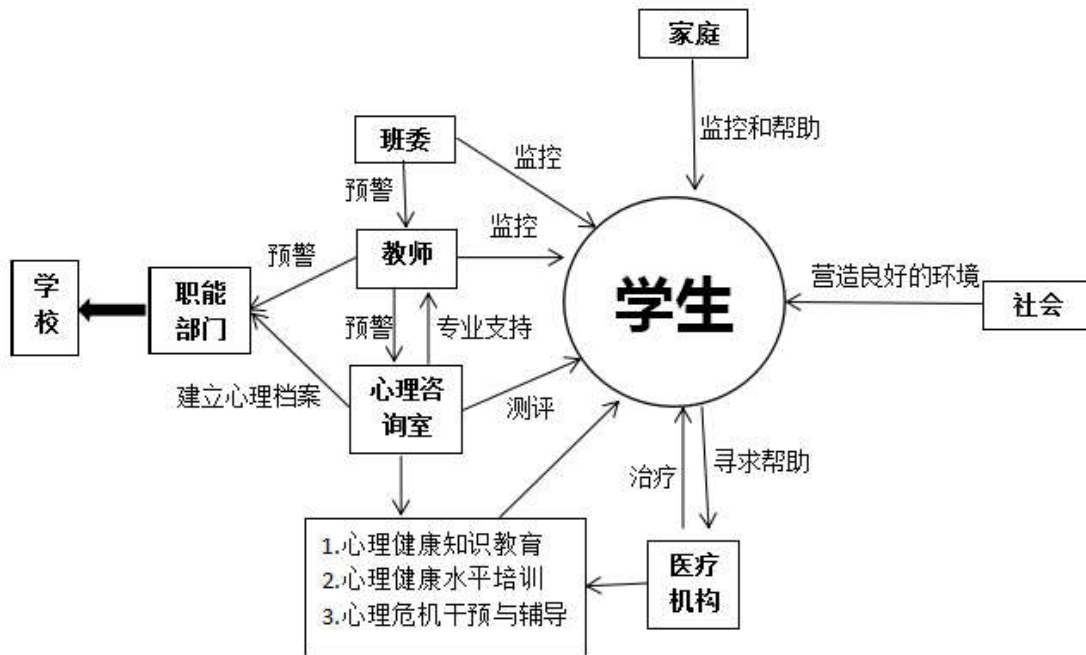


图 30 预警机制及解决方案流程图

七、模型的分析与应用

由于被试在填写问卷时的结果可能由于各种原因出现误差，题中数据很难完全与实际情况一致，故本节中首先对风险指数模型中得到的风险指数进行灵敏度分析，以检验和分析模型是否实用、可靠、有效. 其次将风险指数模型的结果应用于实际中，进一步对问题一给出具体的结果.

7.1 风险指数模型的灵敏度分析

根据之前建立的风险指数模型，可得如下公式：

$$k = 0.5 * (k_1 + k_2) = 0.5 * \left(\frac{s^*}{s^0} + \frac{0.05}{r} \right) \dots\dots (25)$$

由此可知，风险指数 k 与样本到正理想解的距离 s^* 、样本到负理想解的距离 s^0 、样本对正理想解的灰色加权关联度 r 有关. 故本文将研究 s^* ， s^0 ， r 发生微小变化时 k 值的变化情况，以检验风险指数模型的可靠性.

7.1.1 风险指数 k 对样本到正理想解的距离 s^* 的灵敏度分析

随机选取附件中的 3 个样本，固定其到负理想解的距离和其对正理想解的灰色加权关联度，将其到正理想解的距离依次增加 0.1，0.2， \dots ，1，计算出对应的风险指数值. 将 3

个样本的 k 值随 s^* 变化的曲线图放在同一个坐标系中, 得到结果图如下.

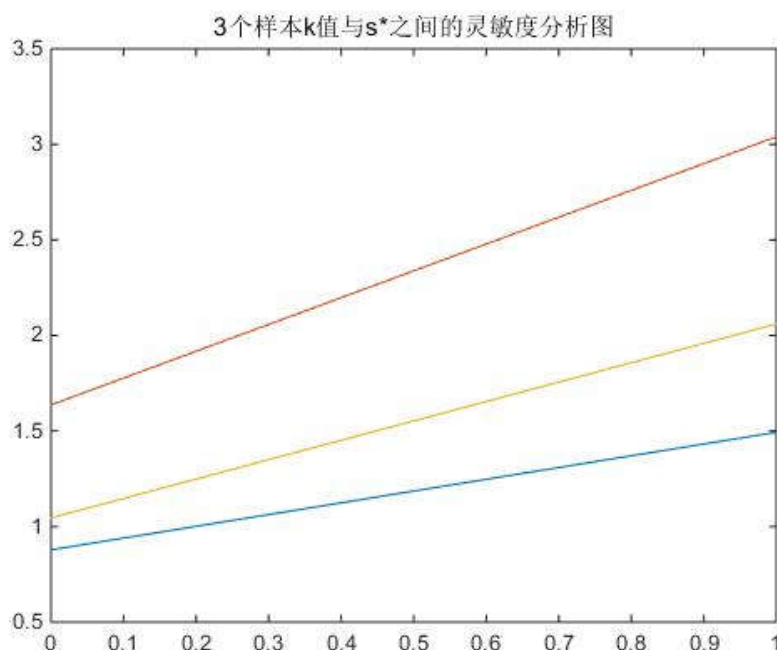


图 31 k 对 s^* 的灵敏度分析图

从上图中可以看出:

- (1) 风险指数值随样本到正理想解的距离增大而增大;
- (2) 三个样本的 $k-s^*$ 曲线的变化趋势均比较平缓, 故当 s^* 的取值发生微小变动时, 风险指数的值不会发生太大变动, 进而在误差范围内模型的结论的变动是可以接受的.

7.1.2 风险指数 k 对样本到负理想解的距离 s^0 的灵敏度分析

随机选取附件中的 3 个样本, 固定其到正理想解的距离和其对正理想解的灰色加权关联度, 将其到负理想解的距离依次增加 0.1, 0.2, \dots , 1, 计算出对应的风险指数值. 将 3 个样本的 k 值随 s^0 变化的曲线图放在同一个坐标系中, 得到结果图如下.

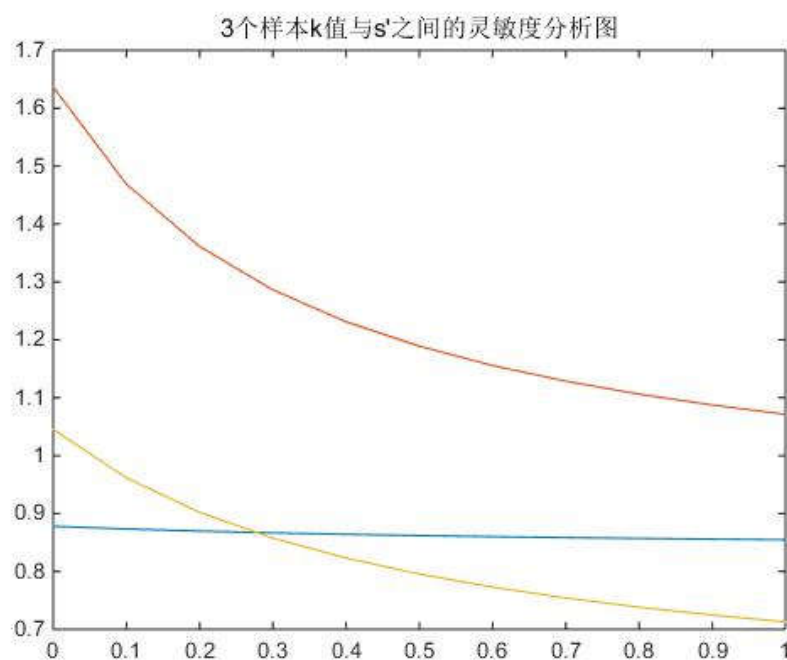


图 32 k 对 s^0 的灵敏度分析图

从上图中可以看出：

- (1) 风险指数值随样本到负理想解的距离的增大而减小；
- (2) 三个样本的 $k-s^0$ 曲线的变化趋势均比较平缓，故当 s^0 的取值发生微小变动时，风险指数的值不会发生太大变动，进而在误差范围内模型的结论的变动是可以接受的。

7.1.3 风险指数 k 对样本对正理想解的灰色加权关联度 r 的灵敏度分析

随机选取附件中的 3 个样本，固定其到正理想解的距离和其到负理想解的距离，将其对正理想解的灰色加权关联度依次增加 0.1, 0.2, ..., 1, 计算出对应的风险指数值. 将 3 个样本的 k 值随 r 变化的曲线图放在同一个坐标系中，得到结果图如下。

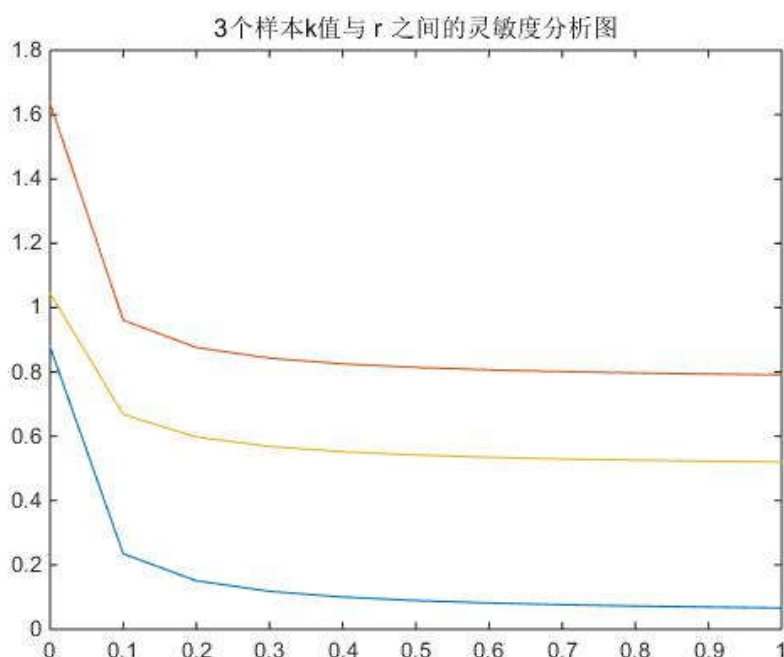


图 33 k 对 r 的灵敏度分析图

从上图中可以看出：

- (1) 风险指数值随样本对正理想解的灰色加权关联度的增大而减小；
- (2) 三个样本的 $k-r$ 曲线的变化趋势均比较平缓，故当 r 的取值发生微小变动时，风险指数的值不会发生太大变动，进而在误差范围内模型的结论的变动是可以接受的。

综合上述灵敏度分析，本文建立的风险指数模型得到的结果比较可靠，不会因为数据发生微小变化而发生较大的变化，故可以应用于实际中。

7.2 风险指数模型的应用

根据风险指数模型，本文对附件中的样本的存在心理问题的风险进行了评估，得到了如下结果：

- (1) 所有样本的风险指数的最大值为 3.15，平均值为 1.16；
- (2) 若将风险指数大于 2 的案例作为存在心理问题的风险高的案例，则有 30 个案例属于此类案例；
- (3) 若将风险指数介于 1.5-2 的案例作为存在心理问题的风险较高的案例，则有 120 个案例属于此类案例；
- (4) 若将风险指数介于 0.8-1.5 的案例作为存在心理问题的风险中等的案例，则有 792 个案例属于此类案例；
- (5) 若将风险指数介于 0.6-0.8 的案例作为存在心理问题的风险较低案例，则有 102 个案例属于此类案例；

(6) 若将风险指数小于 0.6 的案例作为存在心理问题的风险较高的案例, 则有 28 个案例属于此类案例;

(7) 所有样本的风险指数均大于 0, 也即所有样本都或多或少存在一定的存在心理问题的风险.

八、 模型的评价与推广

8.1 问题一模型的评价与推广(基于 TOPSIS 和灰色关联分析的风险指数模型)

■ 优点:

(1) 模型一建立在灰色关联度分析理论和理想解法的基础上, 充分利用了利用灰色理论和理想解法优缺点互补的特点, 建立一种新型的风险评价体系, 用于评价学生存在心理问题的风险;

(2) 模型一运用 R 型聚类将指标数目进行控制, 在保证模型精度的情况下简化了模型, 使得计算更加简便, 并因为指标之间的相互影响而造成的误差;

(3) 模型一使用熵权法确定各指标的权重, 相对于那些主观赋值法, 这种确定权重的方法更加客观、精确, 避免了人为因素带来的偏差;

(4) 模型一中的理想解法以方案与理想解、负理想解之间的距离作为判断方案优劣的标准, 直观而又易于判断, 更具有计算简便、速度快、精度高, 具有较强的实用性等优点;

(5) 模型一能够很好的处理多指标、多目标的综合评价问题, 评价模型原理简单易懂, 算法便捷, 科学实用.

■ 缺点:

模型一在运用理想解法求风险指数时, 理想解法以距离作为尺度仅仅能反映数据曲线之间的位置关系, 而不能体现数据序列的态势变化. 在指标值相差较大的情况下, 只要方案和理想方案之间的距离相近仍会得到方案优劣相近的结果.

■ 推广:

基于 TOPSIS 和灰色关联分析的风险指数模型可广泛应用于决策与综合评价问题.

8.2 问题二模型的评价与推广(人工神经网络模型)

■ 优点:

(1) 人工神经网络模型当中所有定性或定量的信息都等势分布贮存在网络内的各个神经元指责, 使得模型不仅能够同时处理定性、定量的研究, 还具有很强的鲁棒性和容错性;

(2) 人工神经网络模型可以充分逼近任意复杂的非线性关系, 使得它与各门现代科学

技术紧密相连;

(3) 人工神经网络模型采用并行分布处理方法, 能够实现高速度、高数量的运算;

(4) 人工神经网络模型具有学习和自适应不知道或者不确定系统的能力.

■ 缺点:

(1) 由于人工神经网络是高度非线性的系统, 其高度的复杂性使得不可能精确分析它的各项指标, 这限制了该模型的适用范围;

(2) 由于人工神经网络模型的工作原理注定, 它不宜用来求解必须得到答案和数字计算机解决得很好的问题.

■ 缺点改进建议:

在时间允许的情况下, 多次对 BP 人工神经网络进行重复学习, 以提高预测精度.

■ 推广:

人工神经网络模型适合于求解难于找到好的求解规则的问题, 同时也可以广泛应用于预测与管理、模式识别与图像处理、通信等方面.

九、 参考文献

[1]谭晖、储海宝、袁仁曦.《上海市 1036 名中学生心理健康状况及人格类型》.中国学校卫生. 2004 年 25 期

[2]许志红.《大学生心理危机预警机制研究》.教育探索. 2007 年第 12 期

[3]史燕伟、徐富明、李 燕 、刘程浩 、李 彬.《公共决策中的框架效应:理论解释及影响因素》.心理科学进展 2014 年 8 期

[3]蒋柯 .《从马家爵案透视学校心理安全保障机制》.专题研究:关注学生心理安全. 2004 年 05 期

[4]谢中清.《大学生心理健康预警机制研究》.武汉理工大学学报(社会科学版). 2008 年 10 月第 21 卷 第 5 期

[5]周孟英.《浅谈我国青少年心理健康的现状对策》.健康必读杂志. 2013 年 2 月第 2 期

[6]刘桂秀、冯银甫、杜艳芳、周宗斌、辜转荣、简峰.《青少年心理问题与行为偏差及其干预策略》.长江大学学报(自然科学版). 2011 年 4 月 第 8 卷 第 4 期

[7]池春姬、赵淑英、陈坪.《高校学生心理健康评估模型》.牡丹江师范学院学报(哲学社会科学版) 2004 年第 4 期

[8]李兴、梅景景、张成虎.《心理健康理论模型的整合及评估》. <http://wenku.baidu.com/view/855b875c767f5acfa1c7cd65.html>. 引用日期: 2016-5-13

[9]金佳、张武科.《框架效应影响因素及其认知机制研究综述》.西安电子科技大学学

参赛队号 #5225

报（社会科学版）.2015 年 3 月第 25 卷第 2 期

[10]季卫东、方文莉、郑云飞、高建军.《社区青少年心理问题干预进展》.临床心身疾病杂志.2007 年 7 月第 13 卷第 4 期

[11]雷强、杜国栋、杨世宏.《第九届“认证杯”数学中国数学建模挑战赛 2045 队 C 题》.<http://www.madio.net/forum.php?mod=viewthread&tid=263035&highlight=2045>.

引用日期：2016-5-13

十、 附录

附录一

K-means

```

k=4;
x=[]; %数据
[n,d] = size(x);
bn=round(n/k*rand);%第一个随机数在前 1/K 的范围内
%; 表示按列显示，都好表示按行显示
nc=[x(bn,:);x(2*bn,:);x(3*bn,:);x(4*bn,:)];%初始聚类中心
%x(bn,:) 选择某一行数据作为聚类中心，其列值为全部

%x 数据源，k 聚类数目，nc 表示 k 个初始化聚类中心
%cid 表示每个数据属于哪一类，nr 表示每一类的个数，centers 表示聚类中心
[cid,nr,centers] = kmeans(x,k,nc)%调用 kmeans 函数
%认为不该是 150，或者说不该是个确定值，该是 size(x,1)就是 x 行数
for i=1:150
    if cid(i)==1,
        plot(x(i,1),x(i,2),'r*') % 显示第一类
        %plot(x(i,2),'r*') % 显示第一类
        hold on
    else
        if cid(i)==2,
            plot(x(i,1),x(i,2),'b*') %显示第二类
            % plot(x(i,2),'b*') % 显示第一类
            hold on
        else
            if cid(i)==3,
                plot(x(i,1),x(i,2),'g*') %显示第三类
                % plot(x(i,2),'g*') % 显示第一类
                hold on
            else
                if cid(i)==4,
                    plot(x(i,1),x(i,2),'k*') %显示第四类
                    % plot(x(i,2),'k*') % 显示第一类
                    hold on
                end
            end
        end
    end
end
end
end
end

```

```

end
strt=['红色*为第一类；蓝色*为第二类；绿色*为第三类；黑色*为第四类' ];
text(-4,-3.6,strt);

%BasicKMeans.m 主类
%x 数据源，k 聚类数目，nc 表示 k 个初始化聚类中心
%cid 表示每个数据属于哪一类，nr 表示每一类的个数，centers 表示聚类中心
function [cid,nr,centers] = kmeans(x,k,nc)
    [n,d] = size(x);
    % 设置 cid 为分类结果显示矩阵
    cid = zeros(1,n);
    % Make this different to get the loop started.
    oldcid = ones(1,n);
    % The number in each cluster.
    nr = zeros(1,k);
    % Set up maximum number of iterations.
    maxgn= 100;
    iter = 1;
    %计算每个数据到聚类中心的距离，选择最小的值得位置到 cid

    while iter < maxgn
        for i = 1:n
            % repmat 即 Replicate Matrix，复制和平铺矩阵，是 MATLAB 里面的一个函数。
            %B = repmat(A,m,n)将矩阵 A 复制 m×n 块，即把 A 作为 B 的元素，B 由 m×n 个 A 平铺而成。B 的维数是 [size(A,1)*m, size(A,2)*n]。
            %点乘方 a.^b, 矩阵 a 中每个元素按 b 中对应元素乘方或者 b 是常数
            %sum(x,2)表示矩阵 x 的横向相加，求每行的和，结果是列向量。而缺省的 sum(x)就是竖向相加，求每列的和，结果是行向量。
            dist = sum((repmat(x(i,:),k,1)-nc).^2,2);
            [m,ind] = min(dist); % 将当前聚类结果存入 cid 中
            cid(i) = ind;
        end
        %找到每一类的所有数据，计算他们的平均值，作为下次计算的聚类中心
        for i = 1:k
            %find(A>m,4) 返回矩阵 A 中前四个数值大于 m 的元素所在位置
            ind = find(cid==i);
            %mean(a,1)=mean(a)纵向;mean(a,2)横向
            nc(i,:) = mean(x(ind,:));
            %统计每一类的数据个数
            nr(i) = length(ind);
        end
        iter = iter + 1;
    end

```

```

end

% Now check each observation to see if the error can be minimized some
more.
% Loop through all points.

maxiter = 2;
iter = 1;
move = 1;
%j~=k 这是一个逻辑表达式, j 不等于 k, 如果 j 不等于 k, 返回值为 1, 否则为 0
while iter < maxiter & move ~= 0
    move = 0;
    %对所有的数据进行再次判断, 寻求最佳聚类结果
    for i = 1:n
        dist = sum((repmat(x(i,:), k, 1) - nc).^2, 2);
        r = cid(i); % 将当前数据属于的类给 r
        %点除, a.\b 表示矩阵 b 的每个元素除以 a 中对应元素或者除以常数 a, a.
/b 表示常数 a 除以矩阵 b 中每个元素或者矩阵 a 除以矩阵 b 对应元素或者常数 b
        %nr 是每一类的个数
        %点乘(对应元素相乘), 必须同维或者其中一个是标量, a.*b
        dadj = nr./(nr+1).*dist'; % 计算调整后的距离

        [m, ind] = min(dadj); % 找到该数据距哪个聚类中心最近
        if ind ~= r % 如果不等则聚类中心移动
            cid(i) = ind; %将新的聚类结果送给 cid
            ic = find(cid == ind); %重新计算调整当前类别的聚类中心
            nc(ind, :) = mean(x(ic, :));
            move = 1;
        end
    end
    iter = iter+1;
end
centers = nc;
if move == 0
    disp('No points were moved after the initial clustering procedure.')

else
    disp('Some points were moved after the initial clustering procedure.
')
end

```

R 型聚类分析

```

load jl.txt
d=pdist(jl', 'correlation');

```



```
z=linkage(d,'average');
h=dendrogram(z);
set(h,'Color','k','LineWidth',1.3)
```

熵权法

```
function weights = EntropyWeight(R)
%% 熵权法求指标权重,R 为输入矩阵,返回权重向量 weights

[rows,cols]=size(R); % 输入矩阵的大小,rows 为对象个数,cols 为指标个数
k=1/log(rows); % 求 k

f=zeros(rows,cols); % 初始化 fij
sumBycols=sum(R,1); % 输入矩阵的每一列之和(结果为一个 1*cols 的行向量)
% 计算 fij
for i=1:rows
    for j=1:cols
        f(i,j)=R(i,j)./sumBycols(1,j);
    end
end

lnfij=zeros(rows,cols); % 初始化 lnfij
% 计算 lnfij
for i=1:rows
    for j=1:cols
        if f(i,j)==0
            lnfij(i,j)=0;
        else
            lnfij(i,j)=log(f(i,j));
        end
    end
end

Hj=-k*(sum(f.*lnfij,1)); % 计算熵值 Hj
weights=(1-Hj)/(cols-sum(Hj));
end
```

Topsis

```
clc,clear;
a=input('a='); %输入矩阵
Sstar=[];
S0=[];
[m,n]=size(a);
w=[0.0889,0.1968,0.6200,0.0245,0.0698];
c=a.*repmat(w,m,1);
Cstar=max(c); %求正理想解
```

参赛队号 #5225

```
C0=min(c); %求负理想解
for i=1:m
    Sstar(i)=norm(c(i,:)-Cstar); %求到正理想解的距离
    S0(i)=norm(c(i,:)-C0); %求到负理想解的距离
end
k1=Sstar./S0
%[sf,ind]=sort(k1,'descend') %求排序结果
```

灰色关联分析

```
clc,clear
a=[ ];
[m,n]=size(a);
cankao=[1.78,0.39,9.92,2.87,1.74]

t=repmat(cankao,[1,n])-a;
mmin=min(min(t));
mmax=max(max(t));
rho=0.5;
xishu=(mmin+rho*mmax)./(t+rho*mmax)
guanliandu=mean(xishu)
k2=guanliandu.\1
[gsort,ind]=sort(k2,'descend')
```

风险指数

```
k=0.5*(k1+k2)
```

人工神经网络

```
clc,clear;
a=load('jingliu.txt');%把自变量和因变量的数据保存在纯文本文件 jingliu.txt 中
a=a';%注意神经网络的数据格式，不要将矩阵转置
[m,n]=size(a);
P=a([1:m-1],[1:1000]);[PN,PS1]=mapminmax(P);% 自变量数据规格化到[-1,1]
T=a(m,[1:1000]);[TN,PS2]=mapminmax(T);% 因变量数据规格化到[-1,1]
net1=newrb(PN,TN) %训练 RBF 网络
x=a([1:m-1],[1001:end]); xn=mapminmax('apply',x,PS1); %预测样本点自变量规格化
yn1=sim(net1,xn);y1=mapminmax('reverse',yn1,PS2); %求预测值，并把数据还原
k1=a(m,[1001:end]); %实际值
delta11=abs(a(m,[1001:end])-y1)./a(m,[1001:end]); %计算 RBF 网络预测的相对误差
average_RBF=mean(delta11) %RBF 平均相对误差
```

参赛队号 #5225

```
net2=feedforwardnet(4); %初始化 BP 网络，隐含层的神经元取 4 个（多次试验）
net2=train(net2,PN,TN); %训练 BP 网络
yn2=net2(xn); y2=mapminmax('reverse',yn2,PS2); %求预测值，并把数据还原
delta12=abs(a(m,[1001:end])-y2)./a(m,[1001:end])); %计算 BP 网络预测的相对误差
average_BP=mean(delta12) %BP 平均相对误差
```

附录二

部分样本的加权规范矩阵

x_0	x_1	x_2	x_3	x_4
0.80	0.39	5.58	2.16	1.05
0.80	0.39	5.58	1.94	1.47
0.80	0.39	4.96	2.47	0.63
0.80	0.39	5.58	1.86	0.91
0.80	0.20	5.58	1.52	1.53
0.80	0.20	5.58	1.86	1.05
0.89	0.20	4.96	2.23	0.98
0.89	0.39	5.58	2.11	1.05
0.80	0.39	1.24	1.86	0.84
0.89	0.20	1.24	2.16	1.53

部分样本到正负理想解的距离及其 k_1 、 k_2 值

s^0	s^*	k_1	k_2	k
0.539	0.377	0.700	0.133	0.416
0.761	0.216	0.283	0.232	0.258
0.761	0.216	0.283	0.232	0.258
0.513	0.420	0.818	0.119	0.468
0.761	0.215	0.282	0.233	0.257
0.564	0.345	0.612	0.145	0.378
0.441	0.656	1.486	0.076	0.781
0.538	0.380	0.706	0.132	0.419
0.761	0.215	0.282	0.233	0.257
0.514	0.416	0.808	0.120	0.464
0.538	0.379	0.706	0.132	0.419
0.624	0.280	0.448	0.179	0.313
0.538	0.379	0.705	0.132	0.419
0.516	0.414	0.801	0.121	0.461
0.516	0.414	0.802	0.121	0.461

参赛队号 #5225

0.760	0.225	0.296	0.223	0.259
0.656	0.254	0.387	0.197	0.292
0.459	0.536	1.168	0.093	0.631
0.624	0.282	0.452	0.177	0.315
0.658699508	0.2518167	0.38229374 2	-0.061075	