

五一数学建模竞赛

承 诺 书

我们仔细阅读了五一数学建模竞赛的竞赛规则。

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）与本队以外的任何人（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它公开的资料（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，我们愿意承担由此引起的一切后果。

我们授权五一数学建模竞赛组委会，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）。

参赛题号（从 A/B/C 中选择一项填写）： C

参赛队号： T501034307

参赛组别（研究生、本科、专科、高中）： 本科

所属学校（学校全称）： 重庆邮电大学

参赛队员：队员 1 姓名： 童仕延

队员 2 姓名： 苏鑫阳

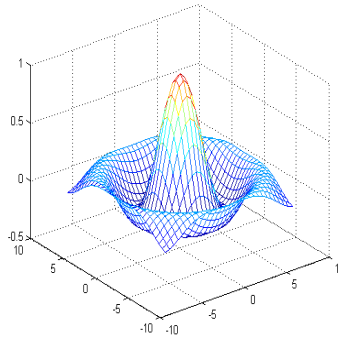
队员 3 姓名： 林培灿

联系方式：Email: 2386263505@qq.com 联系电话： 18757900117

日期： 2022 年 05 月 03 日

（除本页外不允许出现学校及个人信息）

五一数学建模竞赛



题目：利用投影寻踪法模型对消防设施及大队消防水平的综合评价与分析

关键词：TOPSIS 二元 logistics 回归 GA 遗传算法 PP 投影寻踪综合评价

摘要：消防安全关系民生，本文就防范火灾、加强火灾警报及消防管理等问题，运用 TOPSIS、二元 Logistics、投影寻踪综合评价等方法，为政府火警系统、消防建设献策。

针对问题一，我们首先对所给数据进行清洗，通过剔除探测器的误报数据、合并重复报警数据，筛选出 6 月 1 日至 18 日该城市发生的真实火灾起数为 343 起。然后，我们对不同类型部件以可靠性和故障率为指标建立 TOPSIS 评价模型 [1]，排序出最可靠的火灾探测器依次为：智能光电探头 (Intelligent photoelectric probe)、消火栓 (Hydrant) 等。

针对问题二，基于文献与问题一结论，我们选择部件种类、火灾报警次数、项目名称和辖区对应消防大队作为预测参数，依照二元 Logistics 算法 [2] 进行拟合预测，对辖区内建立区域报警部件类型智能研判模型，通过了适应度检验；并据此给出附件 3 各大队火警设备报警信号完成了真实性评价，多次拟合得出报警信号准确无误的平均概率为 92.91%。

针对问题三，基于问题一、二结论以及表 1 数据，我们选取部件可靠 α ，辖区火灾发生频率 γ ，部件故障率 β ，火灾报警设备分布密度 θ 作为综合管理水平技术指标，通过比较各评价模型，最终选择建立投影寻踪综合评价模型 [3]，分析了该市 18 支消防大队的总和管理水平，并且对管理最差的三支消防大队：H 大队、J 大队和 Q 大队提出了可行的改进方案。

针对问题四，基于先前研究结论，我们就火灾报警系统各部件维护管理这一实际问题，有针对性地对包括火灾控制器，智能光电探头等 7 个系统部件分别给出了维护意见，建议各部件应该按照可靠程度进行定期检修。

目录

一、 问题重述	4
1.1 问题背景	4
1.2 问题提出	4
二、 问题分析	5
三、 模型假设	6
四、 符号说明	7
五、 问题一的模型	7
5.1 真实火灾起数的筛选	7
5.2 火灾探测器可靠程度评价	7
六、 问题二的模型	10
6.1 研判模型参数选定	10
6.2 智能研判模型的建立	10
6.3 数据的真实性检验	11
七、 问题三的模型	12
7.1 技术指标的评价维度	12
7.2 模型的建立与求解	14
7.3 技术指标量化分析与改进意见	16
八、 问题四的刍议	17
九、 模型的评价及改进	18
9.1 模型的优点与不足	18
9.2 模型的改进	19
参考文献	20
A 附录 问题一的结论	22
B 附录 问题二的结论	23
C 附录 问题三的结论	24
D 附录 问题一的代码	25
E 附录 问题二的代码	26
F 附录 问题三的代码	30

一、问题重述

1.1 问题背景

火灾报警系统广泛运用在火灾初期的火源监测和楼层疏散场景。根据警报系统工作原理,探测器的灵敏度决定了对火灾特征响应的灵敏程度,但使用灵敏度更高的探测器会降低报警可靠性,实际运用中需权衡探测器的可靠性和灵敏度。此外,当故障时发生火灾,报警系统不会响应。据此,火警信号的准确性就有待考量。图 1 为全国 2021 年各月的火灾次数及变化趋势,表 1 为某城市 18 支消防大队的管辖面积。为了保障消防安全,需要为该市火灾报警系统及消防大队管理给出可行建议。

大队名称	管理面积 (平方公里)	大队名称	管理面积 (平方公里)
A 大队	1712	J 大队	58
B 大队	692	K 大队	1831
C 大队	1100	L 大队	1561
D 大队	1631	M 大队	1997
E 大队	412	N 大队	246
F 大队	1524	O 大队	483
G 大队	122	P 大队	24
H 大队	532	Q 大队	2151
I 大队	96	R 大队	13

表 1 各消防大队管辖面积

1.2 问题提出

问题一: 基于上述背景与附件,确定该城市 6 月 1 日至 18 日的真实火灾起数,并结合文献、图 1 和附件 2,筛选出附件 2 中在附件 1 出现的各类型部件,并依据可靠性和故障率建立模型评价分析各类型部件,指导政府挑选更可靠的探测器类型。

问题二: 基于问题一的结论及相关文献,建立可较好判定火情误报的区域报警部件类型智能研判模型,提高报警准确率;并评价附件 3 各大队辖区中各类部件发出报警信号的真实性,确定信号准确判定出真实火灾的概率。

问题三: 基于问题一、二中的辖区数据及结论,结合表 1 分析各消防大队的综合管理水平,并量化分析综合管理水平最低的三个辖区的技术指标,提出改进方案。

问题四: 基于相关文献及上述结论,有针对性地列述火灾报警系统各部件管理维护的意见建议。

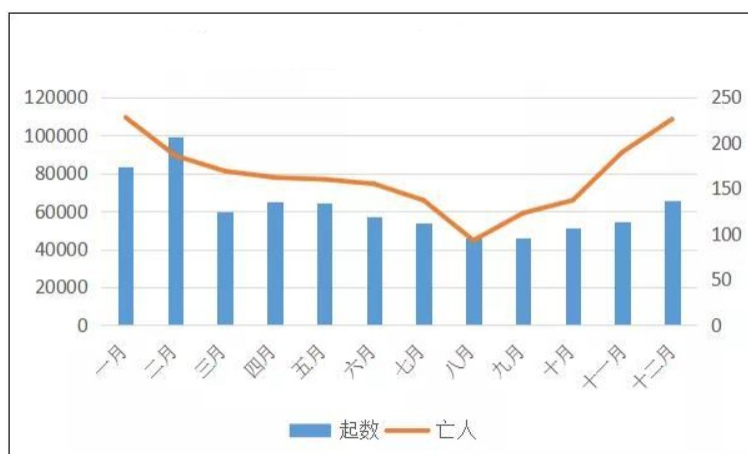


图 1 全国 2021 年各月的火灾次数及变化趋势

二、问题分析

2.1 考虑火警报警系统的工作原理

火警报警系统大致分为火灾探测器、报警控制器和火灾报警装置三部分：当探测器发现火情时，该探测器会发出警报并将信号传输给报警控制器；当报警控制器接收到探测器信号并判定为真实火灾时，会输出命令触发火灾报警装置。简言之，只有当火灾报警装置触发才认为发生火灾；当报警装置未触发但探测器报警，则可认为探测器误报；当探测器未报警但火灾真实发生，则说明探测器故障。

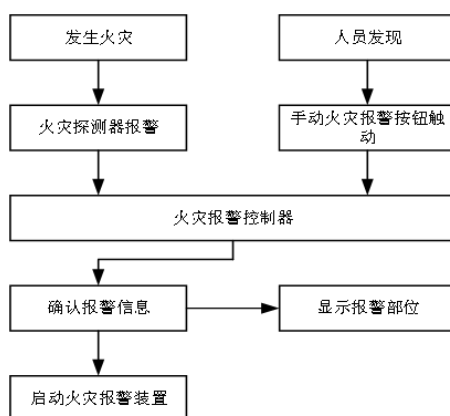


图 2 火灾报警系统工作原理流程

2.2 考虑各类型探测器的可靠程度差异

首先做数据清洗，在附件 1 列举出不同项目下探测器警报数据，为了统计真实火灾的起数，需要整理下述情况：1. 探测器误报数据，2. 同一场火灾的重复警报数据。因此我们考虑通过剔除了探测器的误报数据，合并同一场火灾警报数据得出真实火灾的起数。

不同类型的探测器对火情感知有相应的灵敏度,而灵敏度又与探测器的可靠性相关。探测器的可靠性直接地反映为:准确发现火情的频率,而准确率和误报率又是对立关系。通过查阅附件 4 等相关文献,借鉴误报率通常定义,考虑用将探测器的**误报率的正向指标**描述可靠性;用**部件每百万小时的故障次数**描述故障率,可靠性越强且故障率越低的部件则有更高的可靠程度。

2.3 考虑提高辖区火情警报的准确率

为了提高报警准确率,问题二要求建立区域报警部件类型智能研判模型;故需确定预测参数。查阅文献与附件后,我们考虑到:客观上,火警可靠性直接与探测器灵敏度关联;同时,附件 4 中介绍,可靠性还与探测器安装的环境适配相关。主观上,根据香农熵定律,当大量探测器高密度发出相同警报时,说明信息大,而信息量与不确定性有关;此外,各消防大队辖区发生火灾的概率也有差异。分主客观考虑建立模型更加符合实际,能较好地界定误报。基于此即可研判附件 3 中报警信号的真实性,并得出信息真实概率。

2.4 考虑辖区消防大队的综合管理水平

基于先前研究和已有文献,问题三要求对各辖区消防大队的综合管理水平做评价。我们考虑到:1. 消防辖区面积内探测器密度可以反映消防队的现代化管理水平;2. 面积越大的辖区通常来说都由综合管理能力更强的消防大队管辖,

2.5 考虑火灾报警系统部件的维护管理

火灾报警系统的维护也可以分两方面探讨:1. 软硬件层面,可以考虑为报警系统添加定期自检逻辑闭环,减少人工排查隐患的漏洞。2. 环境因素层面,可以合理的协调各消防大队的辖区分布,让综合管理能力强的大队负责真实火灾概率更低的辖区,具有对探测器更强的统筹管理能力。可以更具先前的研究结论和文献进行合理分析。

三、模型假设

1. 假设正常工作的火灾探测器检测到火灾时一定会报警,并忽略时间的影响。
2. 假设同一建筑内多个“机号、回路编号对于相同”的火灾探测器均发出火灾报警信号时,则认为是同一起火灾事故。
3. 假设该城市一年内的火灾起数的变化趋势与全国 2021 年各月份的火灾次数及变化趋势相似。
4. 假设在探测器安装得当,符合相关标准,不考虑因设备部署疏忽引起监测偏差。
5. 假设探测器安装时就与环境背景良好匹配,此后建筑用途没有发生大型调整。

四、符号说明

符号	意义
$X \quad Y$	当前输入数据或向量, 当前输出数据或向量
$\alpha \quad \beta$	探测器/部件可靠性、探测器/部件故障率
$\gamma \quad \theta$	辖区火灾发生频率、火警报警设备分布密度
$E_1 \quad E_2 \quad W_1 \quad W_2$	可靠性、损耗率的信息熵, 可靠性、损耗率的权重
Sco	探测器的可靠程度
$\sigma \quad p$	Logistic 模型的回归参数、火警信号真实的概率
$Ty \quad Na \quad Fn \quad sfs$	部件类型、火警次数、故障次数、消防大队编号
$U(d)$	消防大队综合管理水平指数

五、问题一的模型

5.1 真实火灾起数的筛选

在本实际问题中, 给出了探测器的火警警报数据包括报警准确情况, 我们对报警数据进行清洗筛选, 考虑到以下标准:

1. 原数据中存在一定的误报数据, 需要剔除误报的数据。

我们对附录 1 中“是否属于误报”字段的数据进行筛选, 将“是”真实火灾和“有 1 次是真实火灾”的数据保留。可以在 python 程序中遍历判断布尔值进行筛选。

2. 原数据中同一场火灾引起的探测器警报会重复地计入火灾起数, 需要进行合并。

我们依据给定的假设 2: “同一建筑内多个’机号、回路编号对于相同’的火灾探测器均发出火灾报警信号时, 认为是同一起火灾事故。”作为标准, 将符合假设的数据合并记为同一场火灾。我们认为该情形是由火情蔓延导致, 火灾蔓延规模扩大可以使多个探测器响应, 合并后计 1 次火灾起数; 对于回路中只有单个探测器响应且并非误判, 我们认为该情形, 探测器发现火灾后及时被扑灭, 应计 1 次火灾起数。根据附件 1 报警数据集, 我们在 python 中给出判断条件遍历的数据集, 最终得出该城市 6 月 1 日至 18 日真实火灾起数为 343 起。

5.2 火灾探测器可靠程度评价

不同类型的探测器因其采集特征对象、触发条件的不同, 具有灵敏度差异; 然而, 探测器的灵敏度与探测器的可靠性往往存在负相关的关系。为了指导政府更好地选择探测器的种类, 我们需要合理地评价不同类型探测器的可靠程度。在附录 2 中, 我们用 Excel

透视表筛选出在附件 1 中出现过的探测器, 含有**火焰探测器**, **点型点烟**等 10 类, 依据题目的要求, 我们将对这 10 类探测器进行可靠程度评价。而可靠程度又可以从以下两方面规定:

1. 探测器的可靠性 α :

在附录 1 中, 我们筛选出了该城市 6 月 1 日至 18 日真实火灾起数, 火灾探测器的可靠性一定程度上可以由误报率反映, 误报率越低说明探测器更加可靠。通过查阅附件 4 等相关文献, 我们发现误报率通常定义为: 每百万小时内各类部件的误报警次数。同时, 对于不同类型的探测器, 各自的警报情况也有波动, 需要对每类探测器的火警次数 Na_i 分类型计算; 而误报率和可靠程度又呈负相关, 需要对该指标取极差实现正向化。基于分析, 我们给出数学定义:

$$\begin{cases} Max = Max(Na_i) \\ M_i = \{(On_i, Ty_i, Na_i) | On, Ty, Na \in Attachment1 \&\& \\ Wm_i = "True" || "False, one of them was a real fire" \} \\ \alpha_j = \frac{Max - Na_i}{1 \times 10^6}, j = 10 \\ A = \{\alpha_1 \alpha_2 \dots \alpha_{10}\} \end{cases} \quad (5-1)$$

其中: M 为原始报警数据向量, On 为“项目名称”, Ty 为“部件名称”, Na 为“火警次数”, Wm 为“是否属于误报”, Max 为全部部件误报次数的最大值, α 为某类型探测器的可靠性, A 为对于类型探测器的可靠性向量。

最终我们将探测器的可靠性描述为: 每百万小时内某部件与最高次数的相对误报警次数, 当可靠性 α 越接近 1 时, 说明该类型的探测器更加可靠。

2. 探测器的故障率 β :

在附件 2 中我们对筛选出来的 10 类探测器的故障次数进行数学分析, 与可靠性的定义同理, 我们认为故障率可以由: 每百万小时内的部件故障次数反映。为了统一评价方向, 同样需要正向化指标, 我们给出数学定义:

$$\begin{cases} Max = Max(Fn_i) \\ M_i = \{(On_i, Ty_i, Fn_i) | On, Ty, Fn \in Attachment2 \} \\ \beta_j = \frac{Max - Fn}{1 \times 10^6}, j = 10 \\ B = \{\beta_1 \beta_2 \dots \beta_{10}\} \end{cases} \quad (5-2)$$

其中: M 为原始故障数据向量, On 为“项目名称”, Ty 为“部件名称”, Fn 为“故障次数”, Max 为全部元器件故障次数的最大值, β 为某类型某探测器故障率, B 为某类型探测器故障率的指标向量。我们将故障率表述为: 每百万小时探测器与最高次数的相对故障次数。

选中上述两个评价指标, 我们使用**熵权法**计算出可靠性和故障率的指标权重: 按照香农信息熵定义, 计算两项指标向量的信息熵, 记为 E_n 一般来说, 若某个特征的

信息熵越小, 表明该特征的值得变异 (对整体的影响) 程度越大, 提供的信息量越多, 在综合评价中所能起到的作用也越大, 其权重也就越大。故我们可以使用信息熵, 建立下述模型客观地计算各个指标的权重, 得出权重: 可靠性权重 $W_1 = 0.3735$ 、故障率权重 $W_1 = 0.6265$

$$W_i = \frac{1 - E_i}{k - (E_1 + E_2)} (i = 1, 2 \dots k) \quad (5-3)$$

我们按照 TOPSIS 综合评价法 [1], 建立对 10 类探测器可靠程度的评价模型, 得出探测器评分序列见表 2, 其中最可靠的探测器为: 智能光电探头 (Intelligent photoelectric probe), 其次为消火栓 (Hydrant)。

$$\begin{cases} Max_{1,4} = [\max(X_{i1}), \max(X_{i2}), \max(X_{i3}), \max(X_{i4})] \\ Min_{1,4} = [\min(X_{i1}), \min(X_{i2}), \min(X_{i3}), \min(X_{i4})] \\ D_{i=1,4}^+ = \sqrt{\sum_{j=1}^{j=4} W_j (B_{ij} - Max_{i-1})^2} \\ D_{i=1,4}^- = \sqrt{\sum_{j=1}^{j=4} W_j (B - Min_{i-1})^2} \\ Sco = \frac{D^+}{D^+ + D^-} \end{cases} \quad (5-4)$$

[1]

表 2 TOPSIS 评价得出的各类型部件的可靠程度指数 (Sco)

部件名称	可靠程度指数 (高优)
智能光电探头 (Intelligent photoelectric probe)	0.1531
消火栓 (Hydrant)	0.1217
压力开关 (Pressure switch)	0.1157
手动报警按钮 (Manual call point)	0.1120
点型感烟探测器 (Point smoke detector)	0.1093
复合探测器 (Multiple-sensor)	0.1061
点型感温探测器 (Point temperature detector)	0.1019
线型光束感烟探测器 (Linear beam smoke detector)	0.653
点型点烟 (Point type smoke)	0.0578
火焰探测器 (Flame detector)	0.0571

根据可靠程度指数, 我们推荐政府优先智能光电探头、消火栓作为火灾探测器。

六、问题二的模型

6.1 研判模型参数选定

问题二需要对建立可以较好判定火情误报的区域报警部件类型智能研判模型, 在本题实际背景下, 通过查阅有关火警警报算法的文献以及附件 2、4 所给出的信息, 我们认为该之恶能研判模型应该包括以下参数:

1. 探测器的部件类型 Ty :

探测器部件的可靠性直接与火警误报率相关, 依据问题研判火警信号的需要, 需要考虑探测器对火警信号真实性的影响。我们基于问题一的结论, 取各类探测器的可靠程度指标 D , 作为研判参数来代表探测器与火警真实概率间的关联性。所以**部件类型**与输入信号的真实性的直接关联;

2. 探测器火警次数 Na :

通过观察附件 2, 我们发现不同辖区下的火警探测器发出的火警次数存在较大的波动, 部分辖区的出现了密集的火警警报, 如图 X, 且多数警报为真实警报。根据香农信息熵原理, 当探测器火警密度越大时, 说明控制器接受的信息量密度也较大, **火警次数**可以反映火警警报的信息量, 而信息量又反映了不确定性; 所以可以将**火警次数**作为参数来描述与不确定性间的关联,

3. 辖区的消防大队: sfs

此外, 在表 1 中还给出了不同辖区消防大队的辖区面积, 消防队伍因为辖区规模、综合管理能力等的差异, 辖区发生火灾的概率一般都有差异。故考虑将**辖区的消防大队**也计入研判模型的参数。

6.2 智能研判模型的建立

对于火警信号的真实性的研判这一实际问题, 可以认为因变量只存在“真”和“假”两种情况, 因此可以用二分类变量“0”或“1”来描述; 因变量的值受到多个确定因素的影响, 且问题的目标函数是对因变量的定性分析, 同时需要让研判模型由较优的判定准确度。

基于上述给定参数以及问题的背景, 我们决定建立**二元 Logistic 回归模型** [2](简称“逻辑回归”)来研判区域部件类型火灾信号的真实性的原因包括:

1. 逻辑回归非常匹配用于预测对象非此即彼的定性问题。2. 逻辑回归对于多个相关因素的函数, 可以良好的拟合。

据上述分析, 我们规定目标函数——输出的火灾信号为真实时, $Y = 1$, 输出的火灾信号为虚假时, $Y = 0$, 我们可以建立多元线性回归方程:

$$Y = \sigma_0 + \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3$$

$$Y = \ln \frac{p}{1-p} \quad (6-1)$$

传统的 Logistic 回归应当满足以下形式:

$$f(x) = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}} \quad (6-2)$$

由 6-1 并根据 Logistic 函数形式 6-2 得到了火警信号为真实火灾的概率 p 的计算公式 6-3:

$$p = \frac{e^{\sigma_0 + \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3}}{1 + e^{\sigma_0 + \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3}} \quad (6-3)$$

[4] 基于上述的数学描述, 我们在 Matlab 中编写程序, 设定 80% 的数据为训练集, 20% 作为测试集, 拟合出二元 Logistics 回归模型的有关参数, 如下表 3:

参数名称	σ_0	σ_1	σ_2	σ_3
参数数值	-1.66509999508635	0.00229414695167035	0.193537198329226	-0.197881764495690

表 3 二元 Logistics 回归模型参数

综上得出对区域部件类型智能研判模型如 6-4,:

$$p = \frac{e^{-1.66509999508635 + 0.00229414695167035T_y + 0.193537198329226N_a - 0.197881764495690N_o}}{1 + e^{-1.66509999508635 + 0.00229414695167035T_y + 0.193537198329226N_a - 0.197881764495690N_o}} \quad (6-4)$$

由该逻辑回归模型可以计算出, 每一组报警信号参数对应使真实火灾的概率, 由于该公式的模型的拟合概率离散分布在 $p \in (0, 1)$ 间, 我们规定: 当概率 $p \in (0, 0.5)$ 时, 判定火警信号为假, 当概率 $p \in (0.5, 1)$ 时, 判定火警信号为真。然后, 我们对得出的 Logistic 回归模型的拟合结果作了对模型系数的 Omnibus 检验, 如表 4, 经过检验三项参数的都具有对逻辑回归拟合较好的显著性: 、

模型系数的 Omnibus 检验				
		卡方	自由度	显著性
步骤 1	步骤	448.629	3	0
	块	448.629	3	0
	模型	448.629	3	0

表 4 二元 Logistics 回归的 Omnibus 检验特征值

6.3 数据的真实性检验

我们基于上述中区域部件类型的智能研判模型, 将附件 3 的数据作为测试集, 对 15 组探测器的报警数据的真实性各测试了 10 次。经过测试, 我们得出项目名称为: ZDDXAZ、

1000QFBDZ、DHZYC17#,18#、HFHX1#3#5#CJJYLGQ、YYLNGY(YYXQ3#L) 共 5 组探测器真实地警报了火灾的发生, 如表 5, 详细数据见附录 B:

表 5 测试数据为真实火灾的概率情况				
部件名称	x2 火警次数	项目名称	x3 隶属消防机构	发生火灾概率
智能光电探头	13	ZDDXAZ	G 大队	65.92%
点型感烟探测器	13	1000QFBDZ	C 大队	51.48%
智能光电探头	13	DHZYC17#、18#L	A 大队	56.56%
线型光束感烟探测器	13	HFHX1#3#5#CJJYLGQ	B 大队	61.29%
智能光电探头	13	YYLNGY (YYXQ3#L)	I 大队	70.08%

之后, 我们对该模型的拟合优度进行适应度分析, 最终得出, 该模型**正确研判火警信号真伪的概率为 92.91%**

七、问题三的模型

7.1 技术指标的评价维度

问题三要求基于已有文献和结论, 结合表 1 分析消防大队的管理水平。本题属于典型的**综合评价类**问题, 需要结合题目要求设置相应的评价技术指标, 在题目和附件给出了四项评价指标:

1. 部件可靠性 α

部件可靠性受部件的灵敏度以及环境因素的影响, 在 5-2 中我们讨论了部件的可靠性; 基于结论, 部件可靠性是火灾报警系统可靠程度的衡量指标: 可靠的探测器可以及时的报告火情, 及时遏止火情蔓延; 反之, 不够可靠的探测器会引发火情的误报, 火警误报率越高反而说明了消防大队综合管理水平偏低, 将浪费大量的警力公共资源。故也需要把部件可靠性 α 纳入技术指标。

在本题中, 可靠性的指标主体由探测器改变为火警警报系统的所有部件, 故需要重新定义, 我们在 5-1 的基础上进行改进, 新定义部件可靠性 α 为: **某消防大队辖区内发生的真实火灾总数与所有辖区的总火警次数的比值。**

2. 部件故障率 β

火灾报警系统发生故障会导致火情漏报, 严重损害受灾民众的生命财产安全; 此外, 系统部件频繁的出现故障, 一定程度上反映了设备缺少管理维护, 或消防系统对防火隐患的轻视促使选择粗制滥造的防火部件, 没有把部件的可靠性放在首位。据此, 部件的故障率实为反映消防队伍综合管理水平的重要指标, 所以应该将部件故障率 β 作为综合评价的技术指标。

同理, 由于评价对象和指标主体的改变, 我们重新给部件故障率 β 的定义: 某大队辖区内设备故障次数占所有辖区的总故障次数的占比 (附件 2 中的占比) 之和。

3. 辖区火灾发生频率 γ

通过查询文献以及题目暗示, 研究显示, 辖区火灾的频率反映辖区内居民的消防隐患意识及辖区治安情况, 而且与消防队伍的消防监控体系相关 [5], 所以需要将火灾发生频率 γ 纳入考量。

根据附件 1 的有关数据, 我们定义辖区火灾发生频率 γ 为: 某消防大队辖区内总火警次数与所有辖区内的总火警次数的比例。

4. 火灾报警设备分布密度 θ

通过删选附件中和火灾数据、火警设备数量以及表 1 中的大队辖区面积等数据, 我们通过计算 Pearson 相关系数、绘制散点图 3 发现了区域监控探测器点数与火灾期数的关系, 多数火灾多发的地区会部署更多的火灾报警设备, 而火灾发生频率又与消防队伍的总和管理水平逻辑相关, 所以需要考虑一个与火灾探测器数量相关的技术指标。

在图 4 中我们用单位辖区面积的探测器数量与火灾数量做分析, 发现不同的消防大队辖区内探测器的数量与面积分布呈现 U 型, 存在两个峰值:1. 面积小但是分布部件数较多;2. 面积大但部件数也较多; 据此, 不能直接使用探测器数量的绝对值来衡量水平, 我们通过作通过远点的直线, 发现多数数据点的设备分布密度 θ (斜率) 都集中下半限, 而分布密度较高的数据零星分布, 我们认为密度大的地区为市中心或者卫星城等人口密度大的地区, 属于火灾严防严控区域, 而下半限的数据为一般重要的辖区。通常来说, 火灾的重点防控区域多派遣综合管理能力较强的消防大队负责, 也反映了消防队伍的现代化管控水平。图 5 的分布于图 4 分布相似也印证了该推断。

为了反映消防队现代化管控水平, 我们定义了火灾报警设备分布密度 θ 为: 大队辖区内单位面积下的火灾报警系统部件数量, 即火灾报警系统部件的密度。

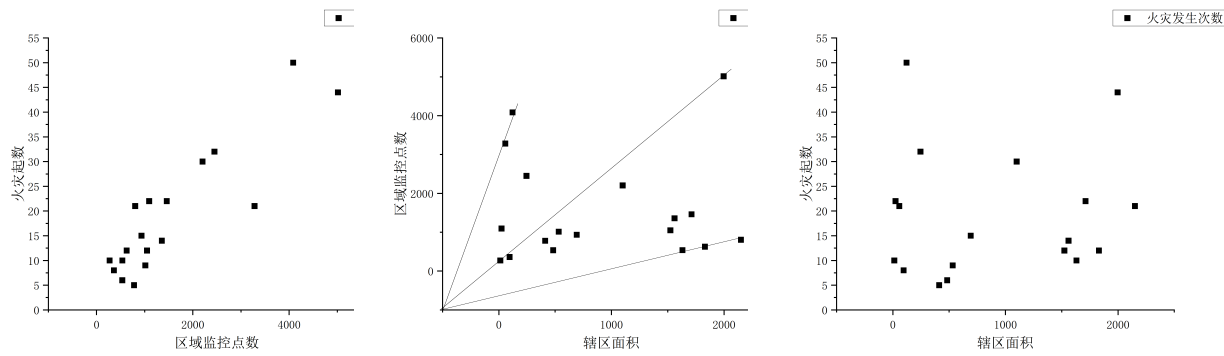


图 3 火灾起数与区域监控点数的关系 图 4 辖区面积与区域监控点数的关系 图 5 辖区面积与火灾起数的关系

表 6: 火灾起数、区域监控点数、辖区面积之间的 Pearson 相关度

		火灾起数	区域监控点数	辖区面积
Pearson 相关	火灾起数	1	0.88792	0.0619
p 值	火灾起数	—	8.76E-07	0.80723
Pearson 相关	区域监控点数	0.88792	1	0.02691
p 值	区域监控点数	8.76E-07	—	0.91558
Pearson 相关	辖区面积	0.0619	0.02691	1
p 值	辖区面积	0.80723	0.91558	—

综上所述, 我们给出了以下技术指标, α 表示部件可靠性, γ 表示辖区火灾发生频率
 β 表示部件故障率, θ 表示火灾报警设备分布密度:

$$\left\{ \begin{array}{l} M = \{(Na, PropN, sfs) | Ng, PropN, SFS \in Attachment 1\} \\ N = \{(Fn, PropF, sfs) | Fn, PropF, SFS \in Attachment 2\} \\ O = \{(S, Fire, sfs) | S, sfs \in Table1, S, Fire \in Attachment 1\} \\ \alpha_{sfs} = \frac{Fire_{sfs}}{\sum_1^i Na_i} \quad \beta_{sfs} = \sum_1^i PropF_{(i,sfs)} \\ \gamma_{sfs} = \frac{\sum_1^j Na_{(j,sfs)}}{\sum_1^i Na_i} \quad \theta_{sfs} = \frac{sum(M_{sfs})}{S_{sfs}} \end{array} \right. \quad (7-1)$$

在本技术指标模型中锁使用到的原始数据分别表示为向量 M、N、P——其中,M 属于来源于附件 1,Na 为表示火警次数,PropN 表示火警次数在附件 1 中火警次数总和中的占比;N 来源于附件 2,Fn 表示故障次数,PropN 表示故障次数在附件 1 中故障次数总和中的占比;O 来源于表 1,S 表示辖区面积;M、N、O 中的 sfs 都表示隶属的消防大队编号,Fire 表示在问题一中得出的各大队辖区内发生真实火灾的次数:

7.2 模型的建立与求解

基于上述四项技术指标, 我们对比了 TOPSIS、AHP 等常见的评价类模型, 最终选择建立**投影寻踪综合评价模型**来评估消防队伍的综合管理水平:

在本题的背景下, 对消防队伍综合管理水平受到多个因素综合影响, 且因素之间存在一定的关联性, 使用传统的 TOPSIS 或 AHP 分析因为过于数学化的描述, 无法很好的顾及变量之间的内联性, 而投影寻踪法可以很好地将多维问题降维量化出指标间的联系。由于本题的技术指标并不复杂, 维度适中, 可避免数据失真, 有良好的稳健性, 可以处理本题同类的简单非线性问题。故我们选择基于多元数据分析的投影寻踪算法来确定四项指标的最佳投影方向, 即可将**最佳投影指标** [6] 作为衡量消防大队综合管理水平的指数。

为了使得评价指标转为正向指数, 我们对四项指标分别作了正向化处理, 转为极大型指标统一量纲, 记录消防大队的四项指标为向量 X_{ij} :

指标序号	指标名称	指标优向
1	部件可靠性 α	正向
2	部件故障率 β	逆向
3	辖区火灾发生频率 γ	逆向
4	火灾报警设备分布密度 θ	逆向

表 7 各项指标的优化方向情况

$$X = [x_1, x_2, x_3, x_4] = [\alpha, \beta, \gamma, \theta] \quad (7-2)$$

$$\text{当 } x_{ij} \text{ 为正向指标时, } Y_{ij} = \frac{x_{ij} - \text{Min}(x_{ij})}{\text{Max}(x_{ij}) - \text{Min}(x_{ij})}$$

$$\text{当 } x_{ij} \text{ 为逆向指标时, } X_{ij} = \frac{\text{Max}(x_{ij}) - x_{ij}}{\text{Max}(x_{ij}) - \text{Min}(x_{ij})}$$

投影寻踪评价方法是通过降维度方式将多维非线性问题转变为一维线性问题, 最终用一维投影向量 $D = [d_1, d_2, d_3, d_4]$ 表征向量的某一方向, 在表征方向上多元技术指标可以投影为 7-3 的形式:

$$Q = \sum_{j=1}^m d_j \cdot Y_{ij} \quad i \in N^* \quad (7-3)$$

然而, 由于样本的多维度, 投影数值 Q 往往使离散分不的, 对于局部, 投影点不够集中, 但对于整体, 投影点又显得过于混乱。要使投影指标归一为一维投影值需要使得:1. 局部的投影点尽可能集中;2. 全局的投影点簇竟可能分散。所以可以根据样本的标准差 $|s(d)|$ 与, 投影局部密度 $\rho(d)$, 优化一维投影指标离散度, 得出投影寻踪的目标函数 U :

$$\begin{cases} S = \sqrt{\frac{\sum_{i=1}^n (Q_i - E_q)^2}{n-1}} \\ U = \rho(d) \cdot S \\ U = \sum_{i=1}^n \sum_{j=1}^n (R - r_{ij})(R - r_{ij}) \end{cases} \quad (7-4)$$

[3] 其中指数 R 为局部密度, 当 R 取较小值时, 大部分点间的距离都不在局部内, 反之则太过聚集。所以需要界定最优的局部密度来使 7-4 的模型可以最优地得出投影目标指标。对于本题的背景, 指标数量较少且较为简单, 可以理想化地认为四项指标呈现线性的形式, 其中 ω 为线性模型的评价系数, 见表 8:

$$U(X_i) = \omega_0 + \omega_1 d_1 + \omega_2 d_2 + \omega_3 d_3 + \omega_4 d_4 \quad (7-5)$$

对于最优局部密度的优化问题, 我们选用了常见的优化算法**遗传算法**, 来求取投影的局部最优密度以及相关的评分系数, 选取投影寻踪综合评价模型**拟合优度**为优化目标, 以模型中的评价系数为**染色体**, 经过对得出的四项技术指标 1000 次迭代后, 最终得出最优的模型评价参数分别为:

表 8 投影寻踪综合评价模型各参数值

参数名称	ω_0	ω_1	ω_2	ω_3	ω_4
参数数值	0	0.0035554	0.50453	0.28114	0.87314

将模型参数带入, 最终得出对消防大队综合管理水平的投影寻踪综合评价模型 7-6, 经过收敛校验, 该模型对于四项评价参数拟合的最优适应度为: 23.3905。

$$U(X_i) = 0.0035554d_1 + 0.50453d_2 + 0.28114d_3 + 0.87314d_4 \tag{7-6}$$

7.3 技术指标量化分析与改进意见

基于 7-6 的投影寻踪综合评价模型, 由于遗传算法的方案具有多解性, 我们对给定的 18 至消防大队的数据分别进行了 10 次评价, 取均值作为评价消防大队的综合管理能力的指数, 以减小因偶然极端变异情况引起的过大误差。最终得出 18 支消防队伍的综合管理水平指数如下表 9(篇幅原因此处只展示评分最低的前五支队伍, 详见附件 C):

消防大队编号	火灾报警设备分布密度	火灾发生频率 x2	部件故障率 x3	部件可靠性 x4	投影寻踪综合评分
H 大队	1.906015038	0.12	0.00	0.01	0.0761
J 大队	56.5862069	0.08	19.05	0.00	5.5973
Q 大队	0.373779637	0.08	30.16	0.02	8.5385
G 大队	33.47540984	0.00	48.54	0.00	13.7657
F 大队	0.68832021	0.11	49.49	0.01	13.9805

表 9 综合管理水平最低的前五支消防队的评分情况

根据投影寻踪综合评价模型得出的消防大队综合管理水平数据, 我们选取表现最差的前 3 支消防队给出可行的改进意见, 分别为:H 大队、J 大队和 Q 大队。

针对 H 大队, 经过对量化后的指标进行分析, 发现其单位面积监控点数较少, 意味着其消防监控系统及消防设施较低覆盖率低, 可能不能及时发现处理火情较小时期, 错过最好的扑灭时机, 其次该大队的设备故障率大、可靠性低, 分析可知 H 大队的设备可能较为陈旧未能及时更新。因此该大队应加大力度布设消防相关设施, 建立智能监控系统, 及时发现险情, 同时增加人员, 对设备进行定期检修维护、更新设施设备。

针对 I 大队, 其设备故障率高且设备可靠性低, 与 H 大队类似, 该大队的设备也存在陈旧, 监控能力薄弱的现象, 应采取类似措施, 提升设备的可靠性和减少故障率。此外, 该辖区的火灾发生频率高, 辖区内用户可能存在防火用火安全意识薄弱的问题, 因此辖区大队应积极做好用火安全宣传, 如: 开展消防安全知识培训, 宣传消防知识, 提升群众安全意识与自防自救能力等措施。

针对 Q 大队, 其火灾发生概率也位于高位, 应当采取上述同 I 大队的类似措施, 减少火种产生来源, 降低火灾发生的概率。同时 Q 大队辖区内的单位面积监控点数少, 未

能建立智能联动联防监控体系,因此加大对消防基础设施以及监控系统的资金投入,高效、及时、准确识别火灾可能性,及时灭火,减少对人民群众财产安全的损失。

八、 问题四的刍议

在问题一中我们挑选出了火灾报警器类型为 15 类探测器,查询资料,我们了解到火警系统部件一般分为**警报触发装置、火灾报警装置、联动输出装置以及其他辅助功能装置**,基于前三个问题的结论,我们分别给出四大类 7 种部件的维护方案。

1. 警报触发装置

火灾警报触发装置作为整个控制系统的核心,既可以接收探测器信号发出警报,又可以根据信号驱动联动灭火装置进行灭火。因此,对该控制器的维护保养就显得尤为重要。首先,要实时检查并确保控制器处于工作状态,连接线路正常。控制器由于裸露在墙壁外,要经常清理灰尘。此外,要及时更换控制器自带电池,防止因为失电造成的控制软件的丢失,必要时可对其进行备份。火灾事故广播作为及时扩散信息的设备也需要经常手动测试以确保正常工作。

以**智能光电探头、缆型感温探测器**为例:

智能光电探头:

智能光电探头是上述问题中最为可靠的火警系统部件,经过推荐,后续智能光电探头的数量会大幅增加,所以需要掌握智能光电探头的管理和维护方法:智能光电探头是一种点型声光探测器,其核心部件是调组会随光线调整变动的感光器件。因此需要定时抽检探头的工作状态,更换过曝过晦的光感胶片,保持智能光电探头的灵敏度适中;此外,不应将该器件安装在湿度过大或迎光的地方。

缆型感温探测器:

缆型期间由于电缆易老化破损的原因,容易因为点卡给短路引起故障:轻则终止火灾检测功能,重则因漏电火花诱发火灾。因此,需要高度重视感温电缆短路:定期对电缆线况进行检测,及时更换废旧电缆,给出一种通过测量电缆电阻的隐患排查方法如下:对非同一对线芯进行测量,若线间电阻小于 20MQ,则说明不同线芯间存在短路情况。卡件故障:如果感温电缆线芯及终端盒侧无故障,并且接线正确完好,有可能为微机调制器卡件故障,可以拆除其他未报故障的卡件进行对调,通过观察判断是否为卡件原因。

对于所有的火灾探测器都需要定期进行故障排查,检查设备电压与响应速度,组织模拟火灾检测设备可靠性,保障对火灾来袭的能有应变能力。

2. 火灾报警装置

以**火警控制器、声光报警器**为例:

控制器常见故障:

液晶屏损害,工控机电路过载工作烧毁电路,为了保障控制器的稳定性,通常控制器

会配备备用显示器,并且将控制器均衡负载到各个核心。火警控制器应该保证机壳接地,定期切断电源检修;控制器还应该安装自检程序,专员监控控制器的状态;同时还应保障检修控制器与探测器、报警器信号输入输出线路的导通性,保障控制器的稳定可靠。

声光报警器:

由于声光报警器接线直接接在设备上,与手动报警装置、点型探测器有所区别,一般故障大多数为接线脱落导致,同时检查声光报警器及地址拨码,若设备接线与拨码均无问题,则需更换声光报警器;有时声光报警器无故障,但试验时发现不能发出声光报警,则首先检查后部接线,测量 24V 电源是否正常,拆开声光报警器观察内部是否有断线故障,可对喇叭线进行重新焊接或重新更换声光报警器。

3. 联动输出装置 (以联动灭火装置的维护为例)

联动灭火装置是整个系统的执行机构。一般包括两种启动方式:自动启动和手动启动。自动启动可以及时察觉火情。而手动启动可以根据实际情况对相关设备手动消防,这是在控制器与联动灭火装置”失联情况下的”一种重要的补充。因此,在日常维护保养中应确保联动灭火装置中的众多设备是可以手动控制并操作的。此外,联动灭火装置通常包括室内室外消火栓系统、自动喷水系统、防火门防火卷帘、室内排风空调、排烟系统等。消火栓系统和自动喷水系统在日常保养中首先要确认能够手动启动。要确保每个设备处于待启动状态且通信正常。防火门、防火卷帘也应进行测试,确保其在动作时隔离效果良好,没有缝隙。室内排风空调要注意及时清理灰尘,防止通道堵塞。

4. 其他辅助功能装置

针对其它辅助功能,通常其由例如:火灾应急广播、火灾应急照明与疏散标志的控制装置组成。火灾应急广播利用公共广播系统和馈电系统在紧急情况下发送应急广播,应确保消防控制室能正常切换至该输出线路以及扬声器的正常使用。应急照明与疏散系统则可以保证在主电网停电时,提供照明,仍能保证民主安全疏散、逃生,因此应定时测试应急电源或应急供应电路的正常,对不正常设备进行及时的维修或更换。

九、模型的评价及改进

9.1 模型的优点与不足

1. 问题一模型的优点和不足

问题一使用到熵权法和优劣距离法,建立对探测器可靠程度的评价模型。

运用熵权法可以保障指标权重的客观性,相比 AHP 层次分析法更加精准可靠。同时,对于权重结果可以通过二次修正提高精度,具有更高的适应度、信度和效度。但是熵权法收到样本的数量限制较大,若本题的可靠程度模型的评价维度提高,得到的权重优度将会下降。

运用 TOPSIS 综合评价法,非常适用于本题给出的对客观数据的评价,对可靠性和

故障率的刻画效果好, 而且模型简洁易用, 对于本题只有两个维度的评价十分适用。但是 TOPSIS 法要求自变量和目标函数的间有较强的关联性, 当评价维度升高, 难以保证指标的关联性。

2. 问题二模型的优点和不足

问题二使用到二元 Logistic 回归模型, 二元 Logistic 回归模型是一种有高适应度二分类算法模型, 该模型非常切合本题中火警信号真实性研判的非正即误的伯努利分布特征; 逻辑简洁, 对火警信号真伪分类时效率较高, 占用计算机资源较少, 同时回归结果较容易观测到样本的概率分布。但是二元 Logistic 模型并不能用与判定高维度的非线性问题, 且对数据向量共线时非常敏感, 无法使用非平衡问题的判定。

3. 问题三模型的优点和不足

问题三使用到投影寻踪综合评价模型分析评价了消防队伍的综合管理水平。投影寻踪法是一种使用多元非线性问题的降维分析方法, 对于本题中四项联系不强的指标可以由较高的适应度; 同时, 该模型兼顾了纳入评价的所有技术指标, 可以避免数据的失真, 稳健性强。在选定模型评价参数时使用了遗传算法进行方案优化, 可以获得对于投影寻踪模型局部半径的局部最优解, 对非线性的指标优较高的适应度, 如图 7 参数确定是遗传算法的适应度曲线可以得到直观印证:

但是, 无论是投影寻踪 (PP) 还是遗传算法 (GA), 数据的运算量都非常庞大, 且数据处理范围优先。对于投影寻踪法, 但高维超体出现了极端的凹陷时, 拟合效果较差; 对于遗传算法, 无法得到全局最优的拟合模型参数, 具有多解性, 需要通过多组迭代获得数据取均值降低误差, 大幅增加的运算时间与资源占用。

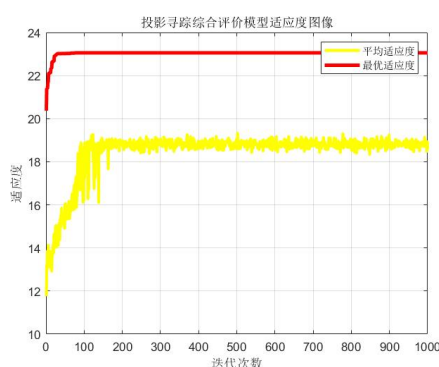


图 6 投影寻踪模型参数确定的遗传算法适应度

9.2 模型的改进

1. 问题一模型的改进方向

针对熵权法和 TOPSIS 组合的改进, 当问题维度升高后, 可以考虑使用正交试验等方法来结合熵权-TOPSIS 模型, 改进后可以正确反映多元问题的权重, 对于关联性不强

的指标,则需要修正模型优化评价算法。

2. 问题二模型的改进方向

针对 Logistic 回归模型对非线性多元问题中,共线向量的敏感性:查询了相关文献,该问题可以采用 UDEED 算法对非线性 Logistics 模型再做改进 [7];但对于本题目中较为单一的变量背景,该改进思路又过于臃肿。

3. 问题三模型的改进方向

针对遗传算法的多解性,可以使用浮点编码加入 Powell 方法建立混合遗传算法 [8]对模型进行改进,该方法可以让遗传算法区域全局最优;针对投影寻踪方法,由于计算量的猛增,难以处理高纬度复杂问题,查询有关文献,该问题可以引入改进蚁群算法增强维度投射的效率,适用于复杂问题以及高维数据处理问题 [9]。

经过上述分析,本题所给出对火灾防空以及消防建设的意见建议或有疏漏之处,对于不够严谨的细节,我们会尝试从上述的思路进行模型的进一步改进优化,追求严谨与科学。

参考文献

- [1] C. C.T., “Extensions of the topsis for group decision-making under fuzzy environment,” Fuzzy Sets and Systems, 2000.
- [2] 何晓群, 多元统计分析. 北京, 中国: 中国人民大学出版社, 2019.
- [3] 侯海龙, “江西制造业竞争力的动态投影寻踪评价——基于中部六省的对比分析,” 信息管理科技, 2021.
- [4] 王济川 and 郭志刚, Logistic 回归模型: 方法与应用. 北京, 中国: 高等教育出版社, 2001.
- [5] 林佳冬, “刍议高层建筑的消防隐患及防火监督,” 科技创新与应用, no. 34, p. 1, 2015.
- [6] 葛敏 and 许长新, “上市公司经营绩效的动态投影寻踪评价——基于中国汽车制造业上市公司数据的分析,” 科技管理研究, vol. 28, no. 6, 2008.
- [7] 庄立纯, 张正军, 张乃今, and 李君娣, “基于非线性 logistic 模型的改进 udeed 算法,” 计算机工程, vol. 45, no. 7, p. 4, 2019.
- [8] 陈静, 付超, and 杨楠, “一种改进的全局优化遗传算法,” 科技创新导报, no. 29, p. 1, 2010.
- [9] 王丽梅, 张宏磊, and 王育德, “基于改进 aco 的投影寻踪模型的研究,” 硅谷, no. 15, p. 3, 2010.

附录 A 问题一的结论

表 5 TOPSIS 评价得出的各类型部件的可靠程度指数 (Sco)

部件名称	可靠程度指数 (高优)
智能光电探头 (Intelligent photoelectric probe)	0.1531
消火栓 (Hydrant)	0.1217
压力开关 (Pressure switch)	0.1157
手动报警按钮 (Manual call point)	0.1120
点型感烟探测器 (Point smoke detector)	0.1093
复合探测器 (Multiple-sensor)	0.1061
点型感温探测器 (Point temperature detector)	0.1019
线型光束感烟探测器 (Linear beam smoke detector)	0.653
点型点烟 (Point type smoke)	0.0578
火焰探测器 (Flame detector)	0.0571

附录 B 问题二的结论

附表 3 测试数据 (Attachment 3 test data)				
x1 部件名称	x2 火警次数	项目名称	x3 隶属消防机构	发生火灾概率
智能光电探头	13	ZDDXAZ	G 大队	65.92%
点型感烟探测器	13	HYBL	J 大队	24.45%
点型感温探测器	13	ZJSJCXQ	N 大队	28.34%
点型感烟探测器	13	ZBDCF	H 大队	32.62%
手动报警按钮	13	CYZBGL、CK	Q 大队	37.17%
点型感烟探测器	13	1000QFBDZ	C 大队	51.48%
点型感烟探测器	13	QZYFSSZD	L 大队	20.99%
手动报警按钮	13	ATYYYYXGS	D 大队	12.89%
点型感烟探测器	13	SGWHJD、LSW	M 大队	46.54%
点型感温探测器	13	ZGYXFXBGL	P 大队	10.76%
智能光电探头	13	DHZYC17#、18#L	A 大队	56.56%
线型光束感烟探测器	13	HFHX1#3#5#CJJYLGQ	B 大队	61.29%
智能光电探头	13	ZLRS (SD) YXGSCQXFSS	E 大队	17.99%
点型感温探测器	13	YZJTGIFYXGSGMYZSC	F 大队	41.72%
点型感烟探测器	13	YYLNGY (YYXQ3#L)	I 大队	70.08%

附录 C 问题三结论

消防大队编号	火灾报警设备分布密度 x1	火灾发生频率 x2	部件故障率 x3	部件可靠性 x4	投影寻踪综合评分
H 大队	1.906015038	0.12	0.00	0.01	0.0761
J 大队	56.5862069	0.08	19.05	0.00	5.5973
Q 大队	0.373779637	0.08	30.16	0.02	8.5385
G 大队	33.47540984	0.00	48.54	0.00	13.7657
F 大队	0.68832021	0.11	49.49	0.01	13.9805
C 大队	2.002727273	0.06	50.00	0.01	14.1033
B 大队	1.349710983	0.10	52.85	0.00	14.9137
N 大队	9.959349593	0.05	53.28	0.01	15.0487
M 大队	2.510766149	0.02	55.43	0.00	15.6028
D 大队	0.329245861	0.12	56.87	0.01	16.0591
L 大队	0.869955157	0.10	60.65	0.01	17.1136
P 大队	45.58333333	0.08	71.15	0.02	20.2233
K 大队	0.342435827	0.11	74.95	0.01	21.1372
O 大队	1.109730849	0.13	75.30	0.01	21.2484
I 大队	3.770833333	0.12	78.03	0.00	22.0116
E 大队	1.893203883	0.13	79.67	0.01	22.4798
A 大队	0.852803738	0.08	81.62	0.01	22.9991
R 大队	20.92307692	0.12	84.32	0.05	23.8846

附录 D 问题一的代码

数据清洗的 python 代码

/ques1/PreSelect/RealFireTimes.py

```
1 import pandas as pd
2
3 # 读取数据，并去除开头标题
4 data = pd.read_excel("附件1 报警数据集 (Attachment 1 Alarm data set).xlsx", 1)
5
6 sum_dict = {} # 计同一个项目同一个机号同一回路发生的火灾次数
7
8 for index, row in data.iterrows():
9     if row["是否属于误报 (Whether a misstatement)"] != "是 (True)":
10         fire_times = int(row["火警次数 (Number of fire alarms)"])
11         dict_key = str(row["项目名称 (Object name)"]) + str(row["机号 (Machinery number)"]) + str(row["回路 (Loop)"])
12
13         # 如果属于这一类型，则认为只发生一次火灾
14         if row["是否属于误报 (Whether a misstatement)"] == "否，其中有1次为真实火灾 (False, one of them was a real fire)":
15             fire_times = 1
16
17         if dict_key not in sum_dict:
18             sum_dict[dict_key] = fire_times
19         elif fire_times > sum_dict[dict_key]:
20             sum_dict[dict_key] = fire_times
21
22 # 最终结果变量
23 total_fire_times = 0
24
25 for times in sum_dict.values():
26     total_fire_times += times
27
28 print(total_fire_times)
```

TOPSIS 的 Matlab 代码

/ques1/TOPSIS/sq.m

```
1 function [out]=sq(in)
2 [r,c]=size(in);
3 if sum(sum(in<0))==0
4     bin=in./sqrt(repmat(sum(in.*in),r,1));
5 else
6     bin=(in-repmat(min(in),r,1))./(repmat(max(in),r,1)-repmat(min(in),r,1));%标准化
    矩阵
7 end
8 p=bin./repmat(sum(bin),r,1);%得到概率矩阵
```



```

9 d=zeros(1,c);
10 for i=1:c
11     p1=p(:,i);
12     e=-sum(p1.*log1(p1))/log(r);
13     d(i)=1-e;
14 end
15 out=d./sum(d);%信息熵

```

/ques1/TOPSIS/mytop.m

```

1 function [out]=mytop(in)
2 a=sqrt(sum(in.*in));
3 b=in./a;%标准化矩阵
4 w=sq(in);
5 [m,n]=size(in);
6 w1=repmat(w,m,1);%权重矩阵
7 d1=sqrt(sum(w1.*(b-repmat(max(b),m,1)).^2,2));%得到每个样本的di+
8 d2=sqrt(sum(w1.*(b-repmat(min(b),m,1)).^2,2));%得到每个样本的di-
9 out=d2./(d1+d2);%计算每个样本的得分

```

/ques1/TOPSIS/log1.m

```

1 function [out]=log1(in)
2 n=length(in);
3 out=zeros(n,1);
4 for i=1:n
5     if in(i)==0
6         out(i)=0;
7     else
8         out(i)=log(in(i));
9     end
10 end
11 end

```

p/ques1/TOPSIS/code.m

```

1 clc;clear
2 load x.mat
3 x(:,1)=max(x(:,1))-x(:,1)
4 w=sq(x)
5 out=mytop(x)

```

附录 E 问题二的代码

数据清洗的 python 代码

/ques2/fileProcess.py

```

1 import pandas as pd

```

```

2 import copy
3 from pandas import DataFrame
4
5 # 读取数据，并忽略第一行标题行
6 data = pd.read_excel("附件1 报警数据集 (Attachment 1 Alarm data set).xlsx", header
    =1)
7
8 # 复制一份数据集，在复制数据集上进行操作
9 result_data = copy.deepcopy(data)
10
11 sub_dict = {} # 隶属消防机构编码记录
12 cm_dict = {} # 部件名称编码记录
13 on_dict = {} # 项目名称编码记录
14
15 sub_index = 0 # 隶属消防机构编码
16 cm_index = 0 # 部件名称编码
17 on_index = 0 # 项目名称编码
18
19 offset = 0 # 当对误判类型为 否（仅一次真实） 的类型，会把它拆分成两个记录，一个记
    录为原火灾发生次数减1，误判类型为是，一个记录为火灾发生次数为1，误判类型为否，因
    此每次执行次操作，为导致数据与原数据存在偏移
20
21 for index, row in data.iterrows():
22     index = int(str(index))
23     sub = row["隶属消防机构 (Subordinate to fire service)"] # excel中每行隶属消防
        机构 (Subordinate to fire service)记录值
24     cm = row["部件名称 (Component name)"] # excel中每行部件名称 (Component name)记
        录值
25     on = row["项目名称 (Object name)"] # excel中每行项目名称 (Object name)记录值
26
27     if sub not in sub_dict:
28         sub_dict[sub] = sub_index
29         sub_index += 1
30
31     result_data.loc[index + offset, "隶属消防机构 (Subordinate to fire service)"] =
        sub_dict[sub]
32
33     if cm not in cm_dict:
34         cm_dict[cm] = cm_index
35         cm_index += 1
36
37     result_data.loc[index + offset, "部件名称 (Component name)"] = cm_dict[cm]
38
39     if on not in on_dict:
40         on_dict[on] = on_index
41         on_index += 1
42

```

```

43 result_data.loc[index + offset, "项目名称 (Object name)"] = on_dict[on]
44
45 if row["是否属于误报 (Whether a misstatement)"] == "是 (True)":
46     result_data.loc[index + offset, "是否属于误报 (Whether a misstatement)"] =
47     0
48 elif row["是否属于误报 (Whether a misstatement)"] == "否 (False)":
49     result_data.loc[index + offset, "是否属于误报 (Whether a misstatement)"] =
50     1
51 else:
52     result_data.loc[index + offset, "火警次数 (Number of fire alarms)"] = data.
53     loc[
54                                     index
55     , "火警次数 (Number of fire alarms)"] - 1
56     result_data.loc[index + offset, "是否属于误报 (Whether a misstatement)"] =
57     0
58     values = copy.copy(data.iloc[index])
59     offset += 1
60     values[5] = 1
61     values[10] = 1
62     values[4] = cm_dict[cm]
63     values[8] = on_dict[on]
64     values[9] = cm_dict[cm]
65     result_data = result_data[:index + offset].append(values, ignore_index=True
66     ).append(
67         result_data[index + offset:], ignore_index=True)
68
69 # 输出成excel表格
70 sub_frame = DataFrame([sub_dict.keys(), sub_dict.values()])
71 DataFrame(sub_frame.values.T, index=sub_frame.columns, columns=sub_frame.index).
72     to_excel("隶属消防机构映射表格.xlsx")
73
74 cm_frame = DataFrame([cm_dict.keys(), cm_dict.values()])
75 DataFrame(cm_frame.values.T, index=cm_frame.columns, columns=cm_frame.index).
76     to_excel("部件名称映射表格.xlsx")
77
78 on_frame = DataFrame([on_dict.keys(), on_dict.values()])
79 DataFrame(on_frame.values.T, index=on_frame.columns, columns=on_frame.index).
80     to_excel("项目名称映射表格.xlsx")
81
82 result_data.to_excel("编号后数据.xlsx")

```

二元 Logistic 的 Matlab 代码

/ques2/code.m

```

1 %% 逻辑回归
2 clc;clear
3

```

```

4  load data.mat;%载入数据
5  [m,n]=size(data);
6  temp=randperm(size(data,1));%随机打乱已知样本数据
7  num=round(size(data,1)*0.8);%80%的数据分为训练组
8  train_x=data(temp(1:num),1:n-1);%训练组的x
9  train_y=data(temp(1:num),n);%训练组的y
10 test_x=data(temp(num+1:m),1:n-1);%测试组的x
11 test_y=data(temp(num+1:m),n);%测试组的y
12
13 chef=glmfit(train_x,train_y,'binomial','logit');%得到回归系数
14
15 %% 模型评价
16 test_y0=zeros(size(test_x,1),1);%初始化预测的概率
17 test=zeros(size(test_x,1),1);%初始化预测的结果
18 for i=1:size(test_x,1)
19     sum0=chef(1);
20     for j=2:n
21         sum0=sum0+chef(j)*test_x(i,j-1);
22     end
23     test_y0(i)=sigmoid(sum0);
24     if test_y0(i)>=0.5
25         test(i)=1;
26     else
27         test(i)=0;
28     end
29 end
30 correct=sum(test==test_y)/size(test_x,1);
31 disp('模型准确率为: ')
32 disp([num2str(correct*100) '%'])
33
34 disp('-----分割线-----')
35 %% 开始预测
36 load predict_x.mat;
37 predict_y=zeros(size(predict_x,1),1);%初始化预测概率
38 predict=zeros(size(predict_x,1),1);%初始化预测结果
39 for i=1:size(predict_x,1)
40     sum0=chef(1);
41     for j=2:n
42         sum0=sum0+chef(j)*predict_x(i,j-1);
43     end
44     predict_y(i)=sigmoid(sum0);
45     if predict_y(i)>=0.5
46         predict(i)=1;
47     else
48         predict(i)=0;
49     end
50 end

```

```

51 disp('预测概率为: ')
52 disp(predict_y)
53 disp('预测结果为')
54 disp(predict)

```

/ques2/sigmoid.m

```

1 %% sigmoid函数
2 function y=sigmoid(x)
3 y=1./(1+exp(-x));
4 end

```

附录 F 问题三的代码

投影寻踪综合评价的 Matlab 代码

/ques3/code.m

```

1 %% 遗传算法 GA 采用浮点数编码
2 %得到最好的a值
3 clc;clear
4 load data.mat;
5
6
7 %% 初始化参数
8 n=500;%种群数目
9 ger=1000;%迭代次数，总共有多少代
10 var=0.2;%变异概率
11 acr=0.4;%交叉概率
12
13 n_chr=4;%染色体个数，自变量个数
14 %length_chr=10;%精度，二进制编码长度，染色体内基因个数（这里不需要）
15 range_chr=[0 1];%自变量的取值范围***
16
17 best_fitness=zeros(1,ger);%存放每一代的最优适应度
18 ave_fitness=zeros(1,ger);%存放每一代的平均适应度
19
20 %% 初始化第一代solvevar(chr,var,n,n_chr,range_chr,i,ger)
21 chr=ys(n,n_chr);%存放个体的染色体
22 fitness=solvefit(chr,n,data);%计算适应度
23 best_fitness(1)=max(fitness);%将当前最优存入矩阵当中
24 ave_fitness(1)=mean(fitness);%将当前平均适应度存入矩阵当中
25 best_chr=findbest_chr(chr,fitness,n_chr);%存放最优染色体和适应度前n_chr个是染色体，
    n_chr+1是适应度
26 %% 开始遗传
27 for i=2:ger
28     chr=solveacr(chr,acr,n,n_chr);%交叉过程
29     chr=solvevar(chr,var,n,n_chr,range_chr,i,ger);%变异处理

```

```

30     fitness=solvefit(chr,n,data);%计算适应度
31     best=findbest_chr(chr,fitness,n_chr);
32     if best_chr(end)<best(end)
33         best_chr=findbest_chr(chr,fitness,n_chr);%存放当前的最优染色体和适应度
34     end
35     %替换劣等解
36     [chr,fitness]=solveworse(chr,best_chr,fitness);
37     best_fitness(i)=max(fitness); %将当前最优存入矩阵当中
38     ave_fitness(i)=mean(fitness);%将当前平均适应度存入矩阵当中
39     acr=acr*0.95;%自适应交叉算子
40     var=var*0.95;%自适应变异算子
41 end
42
43
44 %% 画出适应度图像
45 figure;
46 plot(1:ger,ave_fitness,'y','linewidth',2.5)
47 hold on
48 plot(1:ger,best_fitness,'r','linewidth',3);
49 grid on;
50 title('投影寻踪综合评价模型适应度图像')
51 legend('平均适应度','最优适应度');
52 xlabel('迭代次数')
53 ylabel('适应度')
54 disp(['最优染色体: ' 'a=' num2str(best_chr(1:end-1))])
55 disp(['最优适应度: ',num2str(best_fitness(end))])
56 disp('投影寻踪综合评价的评分为')
57 disp((best_chr(1:end-1)*data'))

```

/ques3/D.m

```

1 %% 求解Da
2 function da = D(z)
3 % R = r_m+3;
4 n = size(z,1);%默认为是一行n列
5 sum0 = 0;
6 r=zeros(n);
7 for i = 1:n
8     for k = 1:n
9         r(i,k) = abs(z(i)-z(k));
10        % sum0 = sum0+(R-r(i,k))*fun(R-r(i,k));
11    end
12 end
13
14 R = rand*(max(max(r))/3-max(max(r))/5)+max(max(r))/5;
15
16
17

```

```

18 for i = 1:n
19     for k = 1:n
20         sum0 = sum0+(R-r(i,k))*fun(R-r(i,k));
21     end
22 end
23
24 da = sum0;
25 end

```

/ques3/D.m

```

1 %% 求解Da
2 function da = D(z)
3 % R = r_m+3;
4 n = size(z,1);%默认为是一行n列
5 sum0 = 0;
6 r=zeros(n);
7 for i = 1:n
8     for k = 1:n
9         r(i,k) = abs(z(i)-z(k));
10        % sum0 = sum0+(R-r(i,k))*fun(R-r(i,k));
11    end
12 end
13
14 R = rand*(max(max(r))/3-max(max(r))/5)+max(max(r))/5;
15
16
17
18 for i = 1:n
19     for k = 1:n
20         sum0 = sum0+(R-r(i,k))*fun(R-r(i,k));
21     end
22 end
23
24 da = sum0;
25 end

```

/ques3/findbestchr.m

```

1 %% 寻找最优染色体和适应度
2 function best_chr = findbest_chr(chr,fitness,n_chr)
3 best_chr = zeros(1, n_chr+1);
4 [max0, max1] = max(fitness);
5 best_chr(1:n_chr) =chr(max1, :);
6 best_chr(end) = max0;
7 end

```

/ques3/fun.m

```

1 %% 判别函数,跃迁信号

```

```

2 function f = fun( x )
3 if x >= 0
4     f =1;
5 else
6     f = 0;
7 end
8 end

```

/ques3/code.m

```

1 %% 遗传算法 GA 采用浮点数编码
2 %得到最好的a值
3 clc;clear
4 load data.mat;
5
6
7 %% 初始化参数
8 n=500;%种群数目
9 ger=1000;%迭代次数，总共有多少代
10 var=0.2;%变异概率
11 acr=0.4;%交叉概率
12
13 n_chr=4;%染色体个数，自变量个数
14 %length_chr=10;%精度，二进制编码长度，染色体内基因个数（这里不需要）
15 range_chr=[0 1];%自变量的取值范围***
16
17 best_fitness=zeros(1,ger);%存放每一代的最优适应度
18 ave_fitness=zeros(1,ger);%存放每一代的平均适应度
19
20 %% 初始化第一代solvevar(chr,var,n,n_chr,range_chr,i,ger)
21 chr=ys(n,n_chr);%存放个体的染色体
22 fitness=solvefit(chr,n,data);%计算适应度
23 best_fitness(1)=max(fitness); %将当前最优存入矩阵当中
24 ave_fitness(1)=mean(fitness);%将当前平均适应度存入矩阵当中
25 best_chr=findbest_chr(chr,fitness,n_chr);%存放最优染色体和适应度前n_chr个是染色体，
    n_chr+1是适应度
26 %% 开始遗传
27 for i=2:ger
28     chr=solveacr(chr,acr,n,n_chr);%交叉过程
29     chr=solvevar(chr,var,n,n_chr,range_chr,i,ger);%变异处理
30     fitness=solvefit(chr,n,data);%计算适应度
31     best=findbest_chr(chr,fitness,n_chr);
32     if best_chr(end)<best(end)
33         best_chr=findbest_chr(chr,fitness,n_chr);%存放当前的最优染色体和适应度
34     end
35     %替换劣等解
36     [chr,fitness]=solveworse(chr,best_chr,fitness);
37     best_fitness(i)=max(fitness); %将当前最优存入矩阵当中

```



```

38     ave_fitness(i)=mean(fitness);%将当前平均适应度存入矩阵当中
39     acr=acr*0.95;%自适应交叉算子
40     var=var*0.95;%自适应变异算子
41 end
42
43
44 %% 画出适应度图像
45 figure;
46 plot(1:ger,ave_fitness','y','linewidth',2.5)
47 hold on
48 plot(1:ger,best_fitness','r','linewidth',3);
49 grid on;
50 title('投影寻踪综合评价模型适应度图像')
51 legend('平均适应度','最优适应度');
52 xlabel('迭代次数')
53 ylabel('适应度')
54 disp(['最优染色体: ' 'a=' num2str(best_chr(1:end-1))])
55 disp(['最优适应度: ',num2str(best_chr(end))])
56 disp('投影寻踪综合评价的评分为')
57 disp((best_chr(1:end-1)*data'))

```

/ques3/judge.m

```

1 %% 判断变异后是否超出边界
2 function out=judge(in,range)
3 if in<range(1)||in>range(2)
4     if abs(in-range(1))<abs(in-range(2))
5         out=range(1);
6     else
7         out=range(2);
8     end
9 else
10     out=in;
11 end

```

/ques3/S.m

```

1 %% 求解Sa
2 function S_a = S(z)
3 % 此求S(a)
4 n = size(z,1);%n行一列
5 z_mean = sum(z)/n;
6 S_a = sqrt( 1/n*(sum((z-z_mean).^2)) );
7
8 end

```

/ques3/solveacr.m

```

1 %% 交叉过程
2 function newchr=solveacr(chr,acr,n,n_chr)

```

```

3  for i=1:n
4      p=rand;
5      if p<acr %此处是将随机两个个体的随机两个染色体进行交换
6          acr_chr = floor((n-1)*rand+1); %要交叉的染色体
7          acr_n = floor((n_chr-1)*rand+1); %要交叉的节点
8          swag=chr(i,acr_n);
9          chr(i,acr_n)=chr(acr_chr,acr_n);
10         chr(acr_chr,acr_n)=swag;
11     end
12 end
13 newchr=chr;
14 for i=1:n
15     if sum(newchr(i,:).^2)>1.1 || sum(newchr(i,:).^2)<0.9
16         newchr(i,:)=ys(1,n_chr);
17     end
18 end
19 end

```

/ques3/solvefit.m

```

1 %% 适应度计算
2 function fitness=solvefit(chr,n,data)
3 fitness=zeros(n,1);%初始化适应度
4
5 for i=1:n
6     z=std(data,chr(i,:));
7     fitness(i)=S(z)*D(z);
8 end
9
10 end

```

/ques3/solvevar.m

```

1 %% 变异过程 对所有单个个体进行变异处理
2 function newchr=solvevar(chr,var,n,n_chr,range_chr,i,ger)
3 for j=1:n
4     for k=1:n_chr
5         p=rand;
6         if p<var
7             pm=rand;%浮点类型增加减少因子
8             change=rand*(1-i/ger)^2;%迭代次数越多，变化越少
9             if pm<=0.5
10                 chr(j,k)=chr(j,k)*(1-change);
11             else
12                 chr(j,k)=chr(j,k)*(1+change);
13             end
14             chr(j,k)=judge(chr(j,k),range_chr);%判断变异后是否超出边界
15         end
16     end

```

```

17 end
18 newchr=chr;
19 for i=1:n
20     if sum(newchr(i,:).^2)>1.1||sum(newchr(i,:).^2)<0.9
21         newchr(i,:)=ys(1,n_chr);
22     end
23 end
24 end

```

/ques3/solveworse.m

```

1 %% 替换劣等解
2 function [newchr,newfitness]=solveworse(chr,best_chr,fitness)
3 max0=max(fitness);
4 min0=min(fitness);
5 lim=(max0-min0)*0.2+min0;%限制适应度：淘汰最后20%的个体
6 replace=fitness<lim;
7 n=sum(replace);
8 chr(replace,:)=ones(n,1)*best_chr(1:end-1);
9 fitness(replace)=ones(n,1)*best_chr(end);
10 newchr=chr;
11 newfitness=fitness;
12 end

```

/ques3/std.m

```

1 %% 投影寻踪综合评价 优化GA
2 function [z,x]=std(data,a)
3
4 %% 初始化
5 [n,m]=size(data);%n为样本个数，m为指标个数
6
7 %% 数据归一化
8 x=zeros(n,m);%初始化x
9 for i=1:n
10     for j=1:m
11         x(i,j)=(data(i,j)-min(data(:,j)))/(max(data(:,j))-min(data(:,j)));
12     end
13 end
14
15 %% 计算z值
16 z=zeros(n,1);
17 for i = 1:n
18     sum0 = 0;
19     for j = 1:m
20         sum0 = sum0 + a(j)*x(i,j);
21     end
22     z(i) = sum0;
23 end

```

```

24 end

/ques3/ys.m

1 %% 满足a的约束条件
2 function out_a=ys(n,n_chr)
3 out_a=[];
4 k=0;
5     while k~=n
6         a=zeros(1,n_chr);
7         for i=1:n_chr-1
8             a(i)=rand;
9         end
10        sum0=sum(a(1:n_chr-1).^2);
11        if sum0>1
12
13            continue;
14        else
15            k=k+1;
16            a(n_chr)=sqrt(1-sum0);
17            out_a=cat(1,out_a,a);
18        end
19    end
20
21 end

```