

# 第三届数学中国数学建模网络挑战赛

地址：内蒙古数学会

电话：0471-4343756

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email：2010@tzmcm.cn

## 第三届“ScienceWord 杯”数学中国

### 数学建模网络挑战赛

#### 承 诺 书

我们仔细阅读了第三届“ScienceWord 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，我们将受到严肃处理。

我们允许数学中国网站([www.madio.net](http://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。

我们的参赛报名号为：1257

参赛队员（签名）：

队员 1：梁燕

队员 2：马剑

队员 3：瞿光祥

参赛队教练员（签名）：梁燕

参赛队伍组别：本科组

# 第三届数学中国数学建模网络挑战赛

地址：内蒙古数学会

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

电话：0471-4343756

邮编：010021

Email：2010@tzmcm.cn

## 第三届“ScienceWord 杯”数学中国

### 数学建模网络挑战赛

#### 编号专用页

参赛队伍的参赛号码：（请各个参赛队提前填写好）：

1257

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）：

# 第三届数学中国数学建模网络挑战赛

地址：内蒙古数学会

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

电话：0471-4343756

邮编：010021

Email：2010@tzmcm.cn

## 2010 年第三届“ScienceWord 杯”数学中国 数学建模网络挑战赛

题 目 高校图书馆的智能服务

关 键 词 超期易度 粗糙集 系统聚类 熵值法 诚信度 双闭环反馈模

### 摘 要：

高校图书馆超期现象普遍存在，能否合理的分析超期现象的原因及制定更加完善的借阅规则关系到高校图书馆资源的利用以及发展。

本文首先分析了影响书籍超期易度指标的主要因素，即超期天数和人均超期书数，然后对这两个主要因素进行数据收集。由于各类高校的情况有所差异，我们选取了郑州作为研究对象。

问题1中，本文分别基于书均超期天数和人均超期书数对各类图书的超期易度排序，结果十分吻合，相互印证。得到p（天文、地球科学）类图书的超期度最高。为了使模型更加的完善，本文基于粗糙集理论对前面得出的超期易度排序进行合理的调整。

接着，根据影响图书超期现象的三种主要原因及其指标特点，分析数据，得出指标值，应用系统聚类法、熵值法基于指标值将超期图书分为三类，由于相同的数据来源，分析得到三种主要原因与三种图书类别存在对应关系，本文便巧妙应用统计聚类模型。结合超期天数分布条形图特点，确定这种对应关系，从而得出不同类别图书的超期原因。

问题2中，我们主要从规则的针对性强弱，对提高图书利用率的贡献度两方面评价借阅规则的优劣。发现其忽略了读者的差别，书籍的类型、将图书规则定为“静态”。据自动化控制理论中的开环控制系统和闭环控制系统原理，将其视为开环控制系统。

针对开环控制系统的弊端，首先根据类别确定借阅规则，接着建立了双闭环反馈模型对其完善。引入了读者诚信度的概念。第一，对借阅天数与借阅时限进行比较，反馈调节超期天数。超期天数通过反馈环节对读者诚信度进行更新。第二，超期天数通过系统中的调解元件得到初级罚款指数，再结合读者诚信度，得出综合借阅指数。接着本文利用博弈论的方法对罚款金额、图书馆催还制度、续借天数与超期天数进行分析。通过综合借阅指数对罚款金额、图书馆催还制度、续借天数的幅度合理的控制，可以控制超期问题，制定出基于差异读者、差异图书的较合理的借阅方案。

参赛队号 1257

参 赛 密 码

所选题目 C

## 1 问题总述

高校图书馆图书超期现象十分普遍，要求重点是寻找解决办法，提高图书利用率。为了解决此问题我们应该完成以下目标：

1. 确定便于分析超期问题的图书的分类方法
2. 搜集便于使用的实例数据，使下面的处理“有数可依”、有针对性
3. 确定评价超期易度的指标、求出哪些类书容易发生超期现象
4. 找到分析给定超期图书的主要超期原因的方法以便“对症下药”
5. 评价给定借阅规则的优劣、设计优化方案

## 2 问题的分析与建模思路

超期现象普遍存在，各高校图书馆对此问题采取了一定的措施，但效果不明显。研究图书超期的文献大都停滞在简单调查总结，定性描述阶段。我们建立模型讨论研究了该问题，充分挖掘数据的信息，给出了定量计算与定性分析结合的方案。

首先，我们对常见催还清单的数据选取了书均超期天数、人均超期书数等指标来刻画图书的超期易度并对郑州大学的催还清单数据，编程统计指标值。借用《图书分类法》对图书分成 22 大类，下面的工作都是针对各类分别讨论。

问题 1，基于书均超期天数、人均超期书数确定的各类图书的超期易度排序，十分吻合，相互印证。结果显示郑大的 p（天文、地球科学）类图书的超期度最高。然后我们引进粗糙集理论对前面得出的超期易度排序进行调整。

问题 2，影响图书超期的因素较多。原因大致可分为无意型、故意型、确实急需型三类。据统计数据显示，无意型占 50%，读书习惯影响占 15%，教学安排占 10%，借阅制度占 15%，图书互借约占 10%。可见无意型极具普遍性。

我们根据多方面的文献总结出影响图书超期的三个主要原因类型及其指标特点，然后对数据统计其指标值，应用系统聚类法、熵值法仅基于数据把超期图书分为三类，我们认为，既然三个原因类型与仅以数据统计算出的类别针对相同的对象的不同研究属性，那么这两个三类之间一定存在一一对应关系，这样寻找原因问题就巧妙转化为求适当的统计聚类方法。结合超期天数分布条形图特点，很容易确定这种对应关系，从而给出了不同类别图书的不同超期原因。

问题 3，我们主要从规则的针对性强弱，对提高图书利用率的贡献度两方面评价借阅规则的优劣。发现郑大的借阅规则存在很多不足。由于借阅规则“一刀切”，没有考虑读者的差别、不同类书的差别，据自动化控制理论中的开环控制系统和闭环控制系统原理，改系统属于开环控制系统。

针对开环控制系统无针对性的弊端，我们建立了双闭环反馈模型。引入了读者诚信度的概念。第一，对借阅天数与借阅时限进行比较，反馈调节超期天数。超期天数通过反馈环节对读者诚信度进行更新。第二，超期天数通过系统中的调解元件得到初级罚款指数，再结合读者诚信度，得出综合借阅指数。接着本文利用博弈论的方法对罚款金额、

图书馆催还制度、续借天数与超期天数进行分析。通过综合借阅指数对罚款金额、图书馆催还制度、续借天数的幅度合理的控制，可以控制超期问题，制定出基于差异读者、差异图书的较合理的借阅方案。

### 3 模型假设

1. 认为图书馆系统经过长时间的运作，已经是稳定系统，存书比例与需求比例大体吻合。
2. 影响超期原因因素很多，但数据难于收集，故我们只选取一些方面研究，虽然不全面但认为合理。
3. 假设诚信度可以从读者的超期图书数量和超期天数反映。
4. 假设可以从获取读者的借阅历史情况，包括读者的超期图书数量和超期天数。
5. 假设查到的全国平均罚款金额准确。
6. 假设高校的图书馆有能力采取本文提出的催还方法

### 4 符号及变量说明

为了便于描述问题，我们用一些符号来代替问题中涉及的一些基本变量，如表 4—1 所示，其它一些变量将在文中陆续说明。

符号系统一览表

主要符号	符号意义
$A(x)$	初步罚款指标
$\lambda_j$	读者诚信度
$\gamma_j$	综合借阅指标
$e$	超期易度
$D$	总超期天数
$d_s$	书均超期天数
$N$	总超期人数
$n_{s/r}$	人均超期书数

### 5 图书分类确定

按照分类号对图书分类，有据可依，全国统一，标准合理，层次分明、科学合理；便于通过数据直观得到类别；收集带有分类号的超期图书统计数据较简单；这种分类简

明适合对超期图书统计研究。既然分类法标准全国通用，那么基于这种分类建模求解，就使解决问题的方法具有通用性、一般意义。

据中国图书分类法我们将图书分为 5 个部分，22 类，具体如下：

五大部类	22 个大类
马克思主义、 列宁主义 毛泽东思想、 邓小平理论	A 马克思主义、列宁主义 毛泽东思想、邓小平理论
哲学	B 哲学、宗教
社会科学	C 社会科学总论
	D 政治法律
	E 军事
	F 经济
	G 文化、科学、教育、体育
	H 语言、文字
	I 文学
	J 艺术
	K 历史、地理
自然科学	N 自然科学总论
	O 数理科学和化学：O1 数学 O3 力学 O4 物理学 O6 化学 O7 晶体学
	P 天文学、地球科学
	Q 生物科学
	R 医药、卫生
	S 农业科学
	T 工业技术：
	U 交通运输
	V 航空、航天：
	X 环境科学安全科学
综合性图书	Z 综合性图书

## 6 实例数据搜集

题目所附数据不知来源，相关情况无法考证搜集，而且条形码不能反映图书的类别，仅依靠书名判别书的类别，不利于处理大量数据。我们经过一番努力，搜集到一些数据，

最后选取了郑州大学图书馆主页发布的 3 月 30 日统计的超期图书催还清单作为主要数据样本进行分析、讨论、建模，求解。

具体数据见附见《郑大数据.xls》

## 7 选取数据指标

对于每一类书我们选取以下指标描述其特点：

- (1) 超期易度  $e$ ：指定图书被超期借阅的概率。
- (2) 逾期率：指定书目超期数量占总总借阅数量的比重。
- (3) 总超期天数  $D$ ：每本书的超期天数累计总合
- (4) 总超期人数  $N$ ：超期人数总和
- (5) 书均超期天数  $d_s$

书均超期天数是指每一种类的书籍平均超期的天数。由于各种书籍都存在很多书籍，学生对每本书籍的喜欢程度存在一定的偏差，但是通过平均概念就能很好的反映出本类书籍在同学之间的欢迎程度。

$$d_s = \frac{\sum x_i}{\sum i}$$

$x_i$  第  $i$  个人借此类书超期的天数

- (6) 人均超期天数  $d_r$

人均超期人数是指每一类书籍平均超期的人数。通过分析，可以知道不同的书籍都存在一定的超期情况，根据统计规律，通常一种书籍借阅的次数越多，他发生超期的可能性就越大。所以统计一类书的书均超期人数能够反映出书籍超期的可能。

$$d_r = \frac{\sum y_j}{\sum j}$$

$y_j$  第  $j$  天超期某类书的人数

- (7) 人均超期书数：

$$n_{s/r} = N_{\text{书}} / N_{\text{人}}$$

从数据可以看出，有些学生有很多书未还，并且未还的期限很长；而有的学生仅有 1 本书未还，且欠的日期很短。这些都能从侧面反映出借书人的个人信息。为了在以后几问中应用相关数据，我们建立了读者信用度的概念。

- (8) 读者信用度：

当然读者累计超期天数越多，信用度越差，但相同的累计超期天数条件下，借阅次数越多者，信用度相对较高，基于以上考虑，我们定义信用度为借阅次数越多

历史超期时间累计与借阅次数累计的比值。

注意到我们没有搜集到历史累计借阅次数、历史累计超期天数数据，我们从充分挖掘已有数据的信息，定义超期书目的相似系数来反映读者信用度。

(9) 同类比重： $r$

一般认为某人欠书未还的数量越多，时间越长，表明此人的信用度越低，但可能此人的确有科研需要，这时他一般会同时超期几本同类的书，从这点出发我们可以认为超期书的种似相同程度越高，其信用度与其他人相比也越高。

(10) 是否续借后超期： $flag$  布尔量

针对郑州大学的催还清单，我们列出了以下三点作为第一问的相关因素。

● 各类图书的超期天数： $d_i$

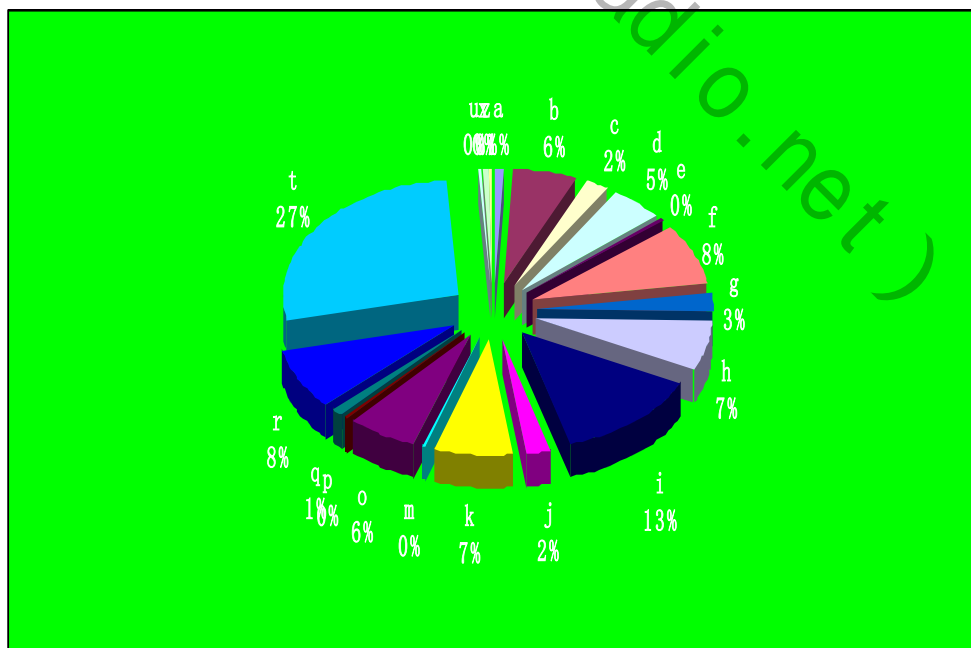
需要说明是，我们要选定一个合适的日期来统计各图书超期情况，使得大部分超期图书已归

● 书均超期天数 $d_s$

● 人均超期书数： $n_{s/r} = N_{\text{书}} / N_{\text{人}}$

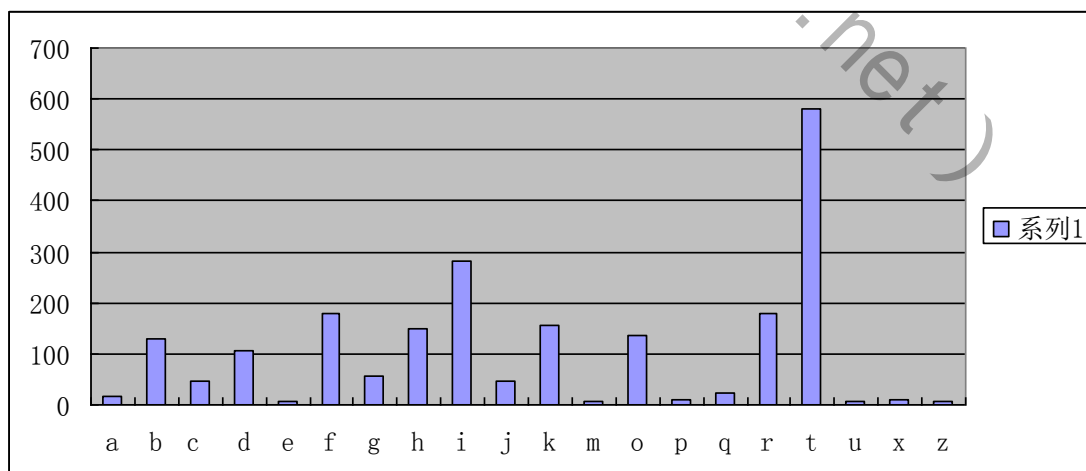
## 8 数据统计结果

具体表格见附表《整理后数据.xls》、《读者编号与超期天数.xls》，通过编写 matlab 程序对郑大数据进行分类处理，程序见附录—程序—程序 2，将以上重要指标分类统计如下：





超期 人数	人均超期 天数	书均超期 天数
15	127	119
128	112	75
48	134	99
105	220	98
8	63	63
178	76	37
57	82	42
148	90	63
281	118	80
47	110	61
156	93	69
5	120	120
135	158	107
9	480	332
24	80	52
180	176	83
581	173	73
6	40	24
9	99	81
7	53	53



据图初步直观分析得：两张表格从不同角度反映出书籍类型的特点。表 1 表明各类

书所占的比重，表 2 说明各类书超期数量的大体范围。综合两张表格的相关数据特这，本文认为，i, p 是最容易超期的书籍，b、d、h、k、o、r 相对容易超期，其他则超期的数量较少。

## 9 求书籍易超期类型

### 9.1 基于书均超期的图书易超期度评价

应用极差法归一化对书均超期天数及人均超期天数排序，

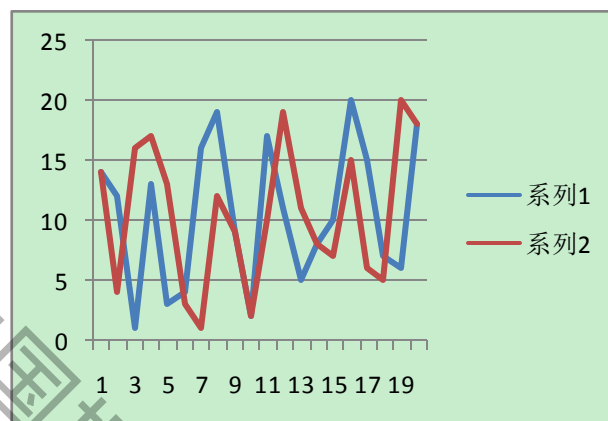
$$\text{计算公式: } x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

图书类别	书均超期天数	序号 1	超期易度 排序 1	人均超期 天数	序号 2	超期易度 排序 2
a	119	14	p	127	14	p
b	75	12	m	112	4	d
c	99	1	a	134	16	r
d	98	13	o	220	17	t
e	63	3	c	63	13	o
f	37	4	d	76	3	c
g	42	16	r	82	1	a
h	63	19	x	90	12	m
i	80	9	i	118	9	i
j	61	2	b	110	2	b
k	69	17	t	93	10	j
m	120	11	k	120	19	u
o	107	5	e	158	11	k
p	332	8	h	480	8	h
q	52	10	j	80	7	g
r	83	20	z	176	15	q
t	73	15	q	173	6	f
u	24	7	g	40	5	e
x	81	6	f	99	20	z
z	53	18	u	53	18	u

求序号 1，序号 2 的相关系数

得  $r=1$ ,

序号 1 与序号 2 的折线图



两只曲线变化趋势很相似,可见超期基于书均超期与人均超期书数的超期度评价结果很接近,相互应正了我们建立评价指标的合理性。

我们认为各类书的人均超期书数主要反映了书的易超期度,下面确定各类图上的超期易度。

## 9.2 确定超期易度

由以上数据求的超期易度排序为:

最易超期书类为 P (天文学地球科学), 接下来是 O 类, C 类。

## 9.3 粗糙集理论修正权重值

以上超期易度排序肯定与实际真实有偏差,我们引进粗糙及理论修正权重,重新排序。

首先介绍粗糙集理论:

定义 1: 信息系统: 信息系统是一个四元组:  $S = (U, A, V, f)$ . 其中:  $U$  表示所要讨论对象的非空有限集合, 也称论域,  $A$  为属性非空有限集合,  $V = \bigcup_{a \in A} V_a$  是属性值的集合,

$f$  是信息函数:  $f: U \times A \rightarrow V$ .  $A = C \cup D$  且  $C \cap D = \emptyset$ , 子集  $C$  和  $D$  分别称为条件属性集和决策属性集。

定义 2: 等价关系: 信息系统  $S = (U, A, V, f)$ , 设  $B \subseteq A$ , 等价关系  $IND(B)$  定义为:

$$IND(B) = \{(x, y) \in U \times U : f_a(x) = f_b(y), \text{ for } a \in B\}.$$

$IND(B)$  将  $U$  划分为若干类  $E_i$ , 式中,  $1 \leq i \leq U/IND(B)$ ,  $|\cdot|$  表示集合中元素的个数.

定义 3: 等价类: 设  $B \subseteq A$ , 对所有  $x \in U$ , 关于  $B$  的等价类  $B(x)$  定义为:

$$B(x) = \{y \in U : (x, y) \in IND(B)\}.$$

定义 4: 上、下近似及边界: 设  $X \subseteq U, B \subseteq A$ , 则  $X$  对于  $B$  的下近似  $B_-(X)$  和上近似  $B^+(X)$  及边界  $BND(X)$  定义为:

$$B_-(X) = \{x \in U : B(x) \subseteq X\}$$

$$B^+(X) = \{x \in U : B(x) \cap X \neq \emptyset\}$$

$$BND(X) = B^+(X) - B_-(X).$$

定义 5: 粗糙集:  $B_-(x)$ ,  $B^+(x)$  称作  $X \subseteq U$  的粗糙集: 给定论域  $U$ ,  $B$  是  $U$  上定义的等价关系:  $[x]_B$  是  $U$  上的等价关系  $B$  生成的  $B$ -等价关系;  $X$  是  $U$  上的一个子集,  $X \subseteq U$ ;  $B_-(x)$ ,  $B^+(x)$  分别是  $X \subseteq U$  的下近似, 上近似, 而且

$$B_-(x) = \{x \in U \mid [x]_B \subseteq X\} \quad (1)$$

$$B^+(x) = \{x \in U \mid [x]_B \cap X \neq \emptyset\} \quad (2)$$

定义 6: 知识的依赖性: 对于给定的近似空间  $K = (U, R)$ , 并且  $P, Q \subseteq R$ , 当  $ind(P) = ind(Q)$ , 知识  $Q$  依赖于知识  $P$ . 知识  $Q$  对  $P$  的依赖度定义为:

$$k = r_p(Q) = Card(pos_p(Q)) / Card(U)$$

其中  $Card$  表示集合的基数.

## 9.4 粗糙集理论调整排序的实现:

为了确定各类图书超期系数的排序,, 则应根据粗糙集理论对表中属性进行约简, 即确定表中各属性的重要度. 然后再根据各类书的重要度来确定各属性的权重. 因此, 建立基于粗糙集理论的权重确定模型应分以下两步进行:

图书馆通过问卷调查,统计各先验数据位超期借阅者的超期原因,作为先验数据库。

第一步:利用粗糙集理论对表中属性进行约简.即从排序对象中去掉一部分序次特异值(如某理工科学图书馆的文史科书籍的逾期率明显大于立功科的逾期率,而这所图书馆的馆藏书籍供给比例与需求比例在长期磨合下趋于平稳,这时我们认为此值特异,应该去掉)再来考虑没有该属性后分类怎样变化.若去掉该属性后,其分类情况变化很大,则说明该属性的强度大,即超期易度高;反之说明该属性的超期易度小,即重要性低,超期易度低。

对于各类图书,论域  $U = \{u_1, u_2, \dots, u_i, \dots, u_m\}$ , 属性集合  $C = \{c_1, c_2, \dots, c_j, \dots, c_n\}$ ,  $i=1,2,\dots,22,$  决策属性集合  $D = \{a_i\}$ , 此时根据粗糙集理论中依赖度的定义公式,计算依次去掉  $c_j$  后,决策属性对条件属性的依赖度  $r_c(D)$ , 观察每个属性将  $U/D$  的正域改变的大小,从而得到每个属性的重要程度。

第二步:确定各属性的权重.粗糙集理论中的属性重要度表达了当前数据环境下属性对决策的影响,不能反映决策者的先验知识.因此,在此将二者结合起来确定属性的权重。

假设对属性集  $C = \{c_1, c_2, \dots, c_n\}$  中  $n$  个属性,由决策者先验知识给定的权重分别为  $p(a_1), p(a_2), \dots, p(a_n)$ . 然后,收集大量的评价样本,由这些样本数据来考察所有条件属性将样本划分为决策类的分类能力,并由决策属性对条件属性的依赖度  $r_c(D)$ , 然后我们去掉属性  $c_j$  后,再重新考察分类情况,得到属性  $c_j$  的重要度  $r_c(D) - r_{c-c_j}(D)$ , 再此基础上将各个属性的重要度进行归一化处理并把处理后的结果  $q(a_j)$  作为属性  $j$  的客观权重. 即

$$q(a_j) = \frac{r_c(D) - r_{c-c_j}(D)}{\sum_{j=1}^n r_c(D) - r_{c-c_j}(D)}$$

显然  $q(a_j)$  在 0 和 1 之间取值. 这样分别得到各个属性主观和客观的权重,为了充分考虑主观和客观两个方面的影响,运用加权计算确定各个属性的综合权重如下:

$$w(c_j) = \alpha p(c_j) + \beta q(c_j)$$

其中  $\alpha$ ,  $\beta$  分别是主观权重和客观权重的权重。

由于我们没有得到基于郑州大学图书馆的调查数据,所以我们拟用模型一中的两个

排序的算术平均值作为客观调查数据的代表，进行了修正。

由于客观权重是由历史数据处理后得到的，在此我们赋予较大的权重，而主观确定的权重带有较强的认为因素，在此赋予较小的权重。所以在计算过程中取  $\alpha = 0.6$ ， $\beta = 0.4$ 。

修正结果如下：P、d、u、c、o、其他类别。

## 10 探究各类图书的超期原因

### 10.1 郑州大学图书馆借阅规则介绍

#### 10.1.1 借书权限

读者类型	借书限量	文艺书限量	借阅天数	可续借次数	续借天数
教师	12 册	12 册	30	1	15
教辅干部	10 册	10 册	30	1	15
工人	5 册	5 册	30	1	15
研究生	12 册	5 册	30	1	15
本、专科生	5 册	2 册	30	1	15
其他	3 册	2 册	30	1	15

注：借书到期前 10 天内才可以办理续借手续。可选择拿书或直接用借阅证来馆续借，也可按有关规定在网上进行续借。续借日期从该书应还日期算起。

#### 逾期费规定

- 1、为提高图书利用率，确保大多数读者的权益，请读者在借阅期限内归还所借图书，对超过借阅期限者收取逾期费。
- 2、每册书逾期一天逾期费 0.1 元。
- 3、寒暑假中到期的书，可至开学后头两周内还清，超过第二周的，按实际借书时间（包括假期）计算逾期费。

4、读者所借图书中有逾期未还的，不得再借其他的书，也不得办理续借手续。

5、读者还书时如没带钱，逾期费可以先行记帐，但总数不得超过五元，否则不再办理借书和续借手续。

## 丢失书刊赔偿规定

读者所借书刊丢失以后，应购买相同版本的，到还书处办理赔偿手续。如无法赔偿原版本图书，则按以下规定赔偿：

普通图书：按出版时间计算赔偿倍率。五年以内按图书原价的3倍赔偿；六至十年按原价的5倍赔偿；十一至十五年按原价的8倍赔偿；十五至二十年按原价的10倍赔偿；二十一年以上的图书按20倍赔偿。

## 10.2 建立超期类型特点分析表

查阅资料文献，我们大致定性地掌握了主要的超期原因及其特点：

超期类型特点分析表

超期类型	原因简析	超期时间一般特点	其他特点
故意超期	急需：如考级考研科研用到的热门图书	短/中等长	超期书的类同比重大，主要易超期书
	认为0.1元的罚款不算什么	很长	
	市场紧俏，脱销书，谎称丢失，长期占用	很长	珍贵、旧、热门
无意超期	忙于学习、工作完全忘记还书	短，不会超过催还缓还期	主要是易超期书
	自以为记着但其实记错了	短，不会超过催还缓还期	主要是易超期书
意外超期	图书被盗、遗失后来不及及时处理	中等长	

另外，据官方统计总结得出：高校图书馆图书借阅超期情况：

无意型占50%，读书习惯影响占15%，教学安排占10%，借阅制度占15%，图书互借约占10%

由于以上三类超期原因具有彼此不同的鲜明的特征，因此在图书馆系统采取措施时

可有针对性地选择不同的方案。

但是，在获得更加有说服性的证据之前（比如说借阅者的先前），图书管理员往往是通过经验来对犯罪者进行经验性的归类，这样往往存在较大的主观性，对“科学性”这一理念造成一定的冲击。我们在这里提出一种方法，根据超期时间列表资料，就可以较科学的对图书超期原因类型做出判断。

### 10.3 熵原理对超期原因的分类判别

首先引入一下熵模型（模型具体介绍见附录一模型介绍 **1 熵模型**）

1948 年，贝尔实验室的学者香农（Shannon）应用概率论知识和逻辑方法推导出了信息量的计算公式。这个反映信息量的统计量。

假设某一实验可能有  $N$  种结果，它们出现的概率分别为  $p_1, \dots, p_N$ ，则事先告诉你将会出现第  $i$  种结果的信息，其信息量为  $-\log_2 p_i$ ，而该实验的不确定性则可用这组信息的平均信息量

$$H = -\sum_{i=1}^N p_i \log_2 p_i$$

来表示。

由于上式与物理学中的“熵”仅相差一个负号，因此也把平均信息量  $H$ （即期望值）称为熵。

实际上，熵反映的是问题的“模糊度”，熵为 0 时问题是完全清楚的，熵值越大则问题的模糊程度也越大。

然后，我们就将对三种超期原因特征进行描述了。

#### 10.3.1.1 无意超期

我们考察的是稳定系统情况，默认无意超期者会得到及时提醒，故此类借阅者的超期时间在一定范围内随机波动（一般为 7 天的催还缓冲时间），在所有可能超期时间段内，出现超期的概率是相等的。设这类超期时间的最大值为  $D$ ，则此类超期时间的概率密度  $f(x)$  为：

$$f(x) = \begin{cases} 1/D, & 0 \leq x \leq D; \\ 0, & \text{others.} \end{cases}$$

其概率密度分布图如图 3：



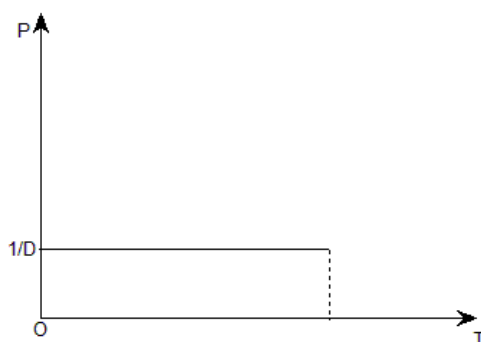


图 3 无意超期时间概率分布图

根据熵原理，此类犯人的熵值  $H$  满足：

$$H(D) = -\int_0^D \frac{1}{D} \cdot \log_2(1/D) dD$$

### 10.3.1.2 科研，考级等急需

由于此类借阅者既要考虑用书需求又考虑罚钱损失，这两种矛盾的心态，使借阅者超期时间不会太长（以尽量减少罚款），也不会太短（研讨需要看书期较稳定），根据的研究，这类超期时间天数的概率密度近似服从正态分布。

概率密度  $f(x)$  为：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

其概率密度分布图如图 2：

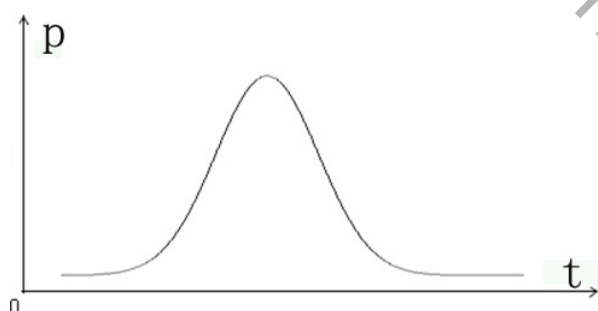


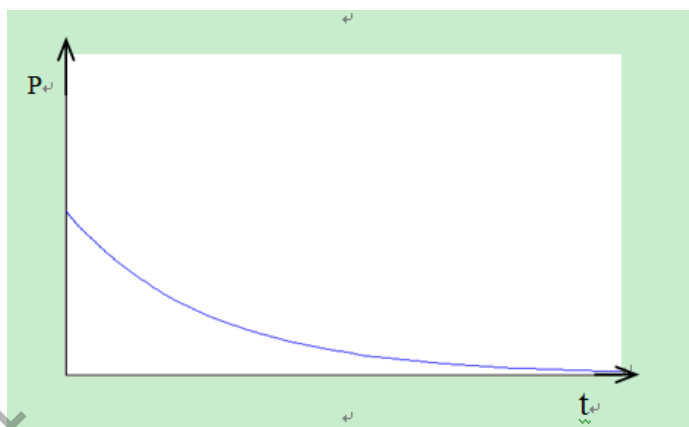
图 2 科研需要概率分布图

根据熵原理，此类犯人的熵值  $H$  满足：

$$H(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-\mu)^2}{2\sigma^2}} \cdot \log_2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-\mu)^2}{2\sigma^2}}$$

### 10.3.1.3 故意拖延

此类读者仅指在催还情况下，依然不还。这类读者超期时间一般中长，且随着超期时间增加，此类超期发生的概率减少。



概率密度函数为：

$$f(x) = e^{-kx}$$

根据熵原理，此类犯人的熵值  $H$  满足：

$$H(k) = -\int e^{-kx} \log_2 e^{-kx} \quad H(k) = -\int e^{-kx} \log_2 e^{-kx}$$

于是，对于一组新的数据，我们可以根据熵原理求出其熵值  $H'$ ，根据  $H'$  与  $H(k)$ ， $H(\mu, \sigma)$ ， $H(D)$  三者之间的欧氏距离，选择距离最近的那个类型，则可认为此属于该种类型超期。

## 10.4 基于聚类方法的原因判断

据先验规律吧原因分为三类，然后据统计数据指标值，应用一定的数值分析聚类方法把指标值分为三类，既然我们选取这些数据指标时是定性分析得出他们相关度很高，那么我们有理由认为这两个三类能对应起来，分析聚类法得到的三类数据的共同特点，我们很容易与特点鲜明的预定分类一一对应。这时分析原因的问题就转为寻求合适的聚类方法与统计理论分析数据。

由模型 1 中，图书超期类型分为三类：无意超期型，故意超期型，科研需要型。三类的特点已分析清楚。下面就进行聚类分析工作。

### 10.4.1 系统聚类法

系统聚类分析是一门多元统计分类法，它事先无须知道分类对象的分类结构，而只需要一批地理数据；然后选好分类统计量，并按一定的方法步骤进行计算；最后便能自然地、客观地得到一张完整的分类系统图，又称谱系图。它的优点在于可以由粗到细的分类。

我们选取书均超期天数和人均超期天数两项聚类要素对这二十类书的超期情况进行

行聚类分析。各类书的书均超期天数和人均超期天数如下表所示：

图书类别	a	b	c	d	e	f	g	h	i	j
对应类别编号	1	2	3	4	5	6	7	8	9	10
书均超期天数	119	75	99	98	63	37	42	63	80	61
人均超期天数	127	112	134	220	63	76	82	90	118	110

图书类别	k	m	o	p	q	r	t	x	z
对应类别编号	11	12	13	14	15	16	17	19	20
书均超期天数	69	120	107	332	52	83	73	81	53
人均超期天数	93	120	158	480	80	176	173	99	53

设第  $i$  类图书  $\omega_i$  ( $i=1,2,3,\dots,20$ ) 的书均超期天数和人均超期天数为  $(v_{i1}, v_{i2})$ ,

则  $\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_{19}, \omega_{20}\}$ 。系统聚类分析的步骤如下：

1) 计算20个样本点两两之间的距离  $\{d_{ij}\}$ ，记为矩阵  $D = \{d_{ij}\}_{n \times n}$ 。如果使用绝对值距离来测量点与点之间的距离，使用最短距离法来测量类与类之间的距离，即：

$$d(\omega_i, \omega_j) = \sum_{k=1}^2 |v_{ik} - v_{jk}|, \quad D(G_p, G_q) = \min_{\substack{\omega_i \in G_p \\ \omega_j \in G_q}} \{d(\omega_i, \omega_j)\}$$

由以上距离公式，可以算出距离矩阵：

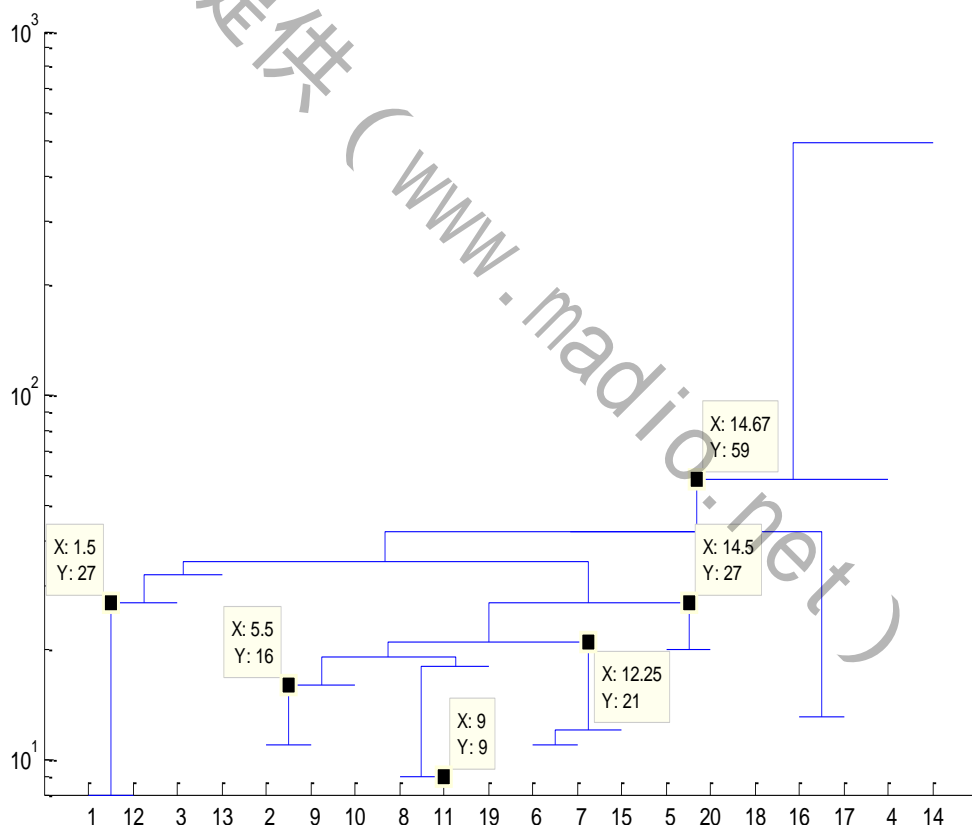
$$\begin{matrix} & \omega_1 & \omega_2 & \omega_3 & \cdots & \omega_{19} & \omega_{20} \\ \omega_1 & 0 & 59 & 27 & \cdots & 66 & 140 \\ \omega_2 & & 0 & 46 & \cdots & 19 & 81 \\ \omega_3 & & & 0 & \cdots & 53 & 127 \\ \vdots & & & & \ddots & \vdots & \vdots \\ \omega_{19} & & & & & 0 & 74 \\ \omega_{20} & & & & & & 0 \end{matrix}$$

2) 首先构造20个类，每一个类中只包含一个样本点，每一类的平台高度均为零；

- 3) 合并距离最近的两类为新类，并且以这两类间的距离值作为聚类图中的平台高度；
- 4) 计算新类与当前各类的距离，若类的个数已经等于1，转入步骤5)，否则，回到步骤3)；
- 5) 由MATLAB画出聚类（程序见附录——程序1）图如下图所示：
- 6) 确定类的个数和各类情况

由以上据类图可大致将图书超期情况分为三类：

超期现象	图书类别
第一类	p类
第二类	d类 c类 o类 u类
第三类	a类 m类 b类 i类 j类 h类 k类 x类 f类 g类 q类 e类 z类 r类 t类



### 10.4.2 分析原因类型与聚类分类对应关系

P类：此类为天文，地球科学，但对照郑州大学的院系专业设置表，没有发现直接

与天文，地球科学相关的科系，故我们有理由排除科研、考级、教学安排等急需原因，结合 P 类图书的超期天数分布图，看到少于一个月的超期分布率大，而超过一个月时间的却几乎没有，还有少一部分超期天数超过几年，这样我们认为这是由于罚款较少或意外丢失造成的。

总之，应该是学生丰富课余生活，自学借用。主要由于普遍无意错过还书期，另外这类书自学需要时间较长，难免造成一些遗失现象。

**Dcou 类：**此四类分别为政治法律、社会科学总论、数学科学与化学、交通运输。对照郑大院系与专业设置表，我们发现：数学系，化学系，文学院，历史学院，外文学院，商学院，教育学院，管理工程，物流管理，公共管理学院等由此可知学生专业知识辅导借书，急需之用借书多。

结合此思类的超期时间分布图，看到此四类的分布图大致相同，以超期时间短为主要特点，课认定为无意超期类型+科研需要类型。

**其他类：**以 T 类作为代表的图书。综合性超期时间除了一般规律的以超期时间少为主外，几乎没有长期超期现象，求中长时间均匀分布，集中体现了无意超期的普遍性，由于罚款交少而中长期超期。

## 11 总体评价借阅规则

在深入分析过图书馆哪些类别的书容易发生超期现象和发生超期现象的原因之后，我们知道并不是所有的书都存在同等程度的超期现象，而是只有个别类型的书存在严重的超期现象。而且不同类别的书，导致出现超期现象的原因也不尽相同。那么该高校的借书规则（假设某高校的借书规则为：借书时限为1个月，可以续借一次，续借时限为15天，超期一天罚款0.1元。）显然存在着一些不合理的地方：

首先，该规则定的过于呆板，没有根据不同类型的书的实际特点制定相应的规则，把规则制定得细致具体。

其次，该规则对于超期者的处罚法措施也不够具体，既达不到处罚有恶劣超期现象的读者和保障图书馆藏书高效利用的目的，也没有达到酌情处罚由于特殊的可以理解的原因而出现超期现象的读者。

再次，官方调查显示：50%以上超期现象是由于读者的疏忽大意造成的。而该规则也不能以一种积极主动的态度去避免这种由于读者的疏忽大意导致的超期现象的发生。

最后，如果能建立读者信誉度积累，反馈到该读者以后的借阅资格权限上，以更积极的态度去超期现象的发生。

## 12 给定借阅方案的评价与新方案提出

### 12.1 自动控制理论介绍

首先，我们先对自动控制的基本原理进行简单介绍

**自动控制系统：**为了实现各种复杂的控制的任务，首先要将被控对象和控制装置按照一定的方式连接起来，组成的一个有机总体

自动控制系统的控制方法有开环控制系统和闭环控制系统。

### 开环控制系统

#### Open-loop Control System

◆定义：系统的控制输入量不受输出量影响的控制系统

◆特点：系统的输出量与输入量间不存在反馈的通道，这种控制方式称为开环控制。

在开环控制系统中，不需要对输出量进行测量，也不需要将输出量反馈到系统输入端与输入量进行比较。因此，开环控制系统，又称顺序控制系统。



### 闭环（反馈）控制系统

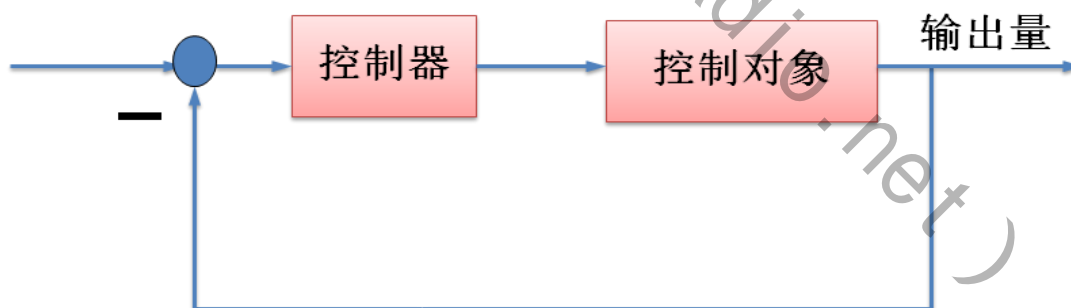
#### Closed-loop Control System

闭环控制系统又称反馈控制系统。

反馈——把取出的输出量送回输入端，并与输入信号相比较产生偏差信号的过程，称为反馈。

反馈控制——采用负反馈(正反馈较少)并利用比较后的偏差进行控制的过程（利用偏差修正偏差）。

由于引入了被反馈量的反馈信息，整个控制过程成为闭合系统，因此反馈控制也称为闭环控制。



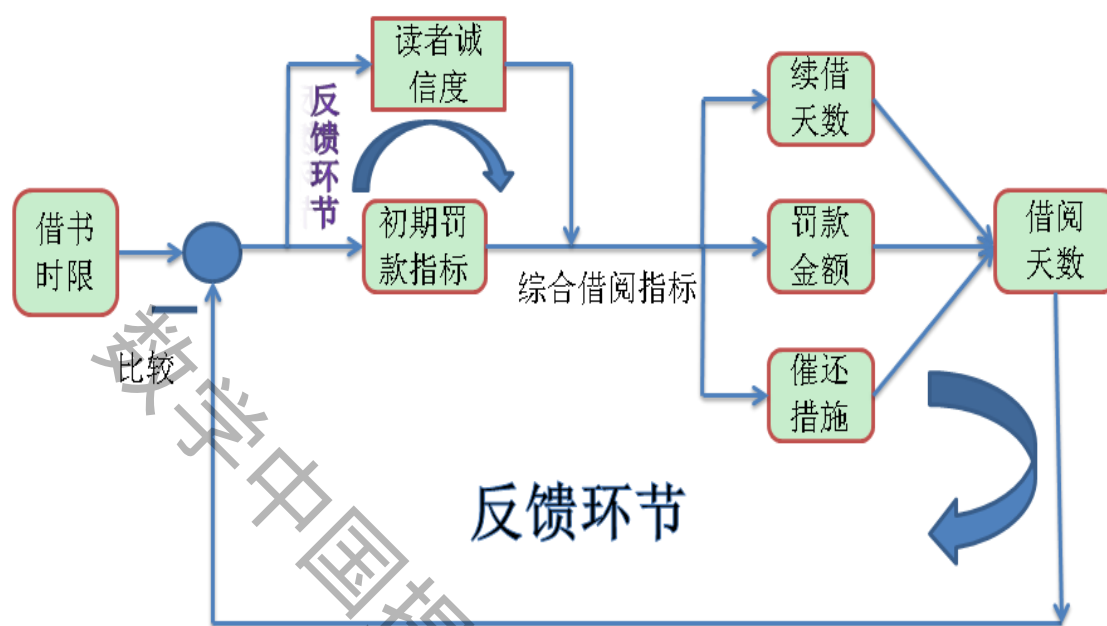
### 闭环与开环控制系统的比较

开环控制：顺向作用，没有反向的联系，没有修正偏差能力

闭环控制：又称偏差控制，可以对不同的输出采取不同的控制。因此，控制精度高，合理性强。

问题二中假设的某高校的借书规则，我们应用了自动控制理论对其进行分析，借书规则的合适与否会影响到图书馆资源的利用。会对本题的超期现象起到一定的缓解。可将其视为开环控制系统。

根据开环控制系统与闭环系统的比较，我们为了弥补开环的不足，制定出更加合理的借阅规则，在问题二中我们建立了反馈控制模型。



双闭环反馈模型

### 双闭环反馈环节的作用

**诚信度的反馈环节：**因为每次借阅读书都有可能会影响到读者的诚信度，所以在这里我们引入了反馈环节，将读者此次的借阅超期天数反馈回读者的个人信息中，对读者诚信度进行实时的更新。

**借阅天数反馈环节：**根据此反馈环节我们得到了此次的借阅超期天数。

整个反馈系统的作用。

首先根据反馈环节得到借阅天数，接着通过比较环节可以得出借阅超期天数，超期天数又分别通过诚信度反馈模型对读者信息进行更新，同时通过调节环节得到初步罚款指数，初步罚款指数与诚信度结合考虑通过修正环节输出综合借阅指数，凭此来对影响超期问题的三因素罚款金额、图书馆催还制、续借天数进行了控制。

### 下面介绍的是反馈系统的组成部分

**给定元件——**每类书籍规定的借阅时限

**比较元件——**其职能是把被控量实际值与给定量进行比较，求出它们的偏差。

**补偿元件——**其职能是利用反馈的方式使输出更合理，更趋向与理想化。这里，我们引进了读者的诚信度作为系统的补偿环节。

**触发元件——**其职能就是通过改进后的读者综合借阅指数来控制执行元件。

**调整元件**

**触发元件——**其职能就是通过改进后的读者综合借阅指数来控制执行元件。

**执行元件——**其职能是直接推动被控对象，使其被控量发生变化。这里，我们考虑了图书可以续借的天数，图书馆催还，以及超期后的罚款金额。

## 12.2 每个环节的功能，及其建立的数学模型。

### 1、给定元件

对问题二中所提出的借阅规则分析器优劣后，我们对其进行改进，及根据问题一中建立的合理模型，对所有类型的书籍进行分类。分类后根据不同的类别给出不同的借阅时限，续借标准天数。

书籍类别	借阅时限	续借标准天数
其他	15	8
d c o u	30	15
p	45	20

### 2、比较元件

其职能是把被控量实际值与给定量进行比较，求出它们的偏差。这里，我们就是简单的根据图书借阅天数与借阅时限进行做差，即超期天数。

超期天数=图书借阅天数-借阅时限

### 3、调节元件

调节环节的输入是超期天数，根据超期天数可以给出初步罚款指标

这里采用模糊数学中的隶属函数来处理所得数据。在隶属函数的选取上，为了使处理后的数据既能适当的区分罚款类别，又能给出最高上限，控制了罚款金额。

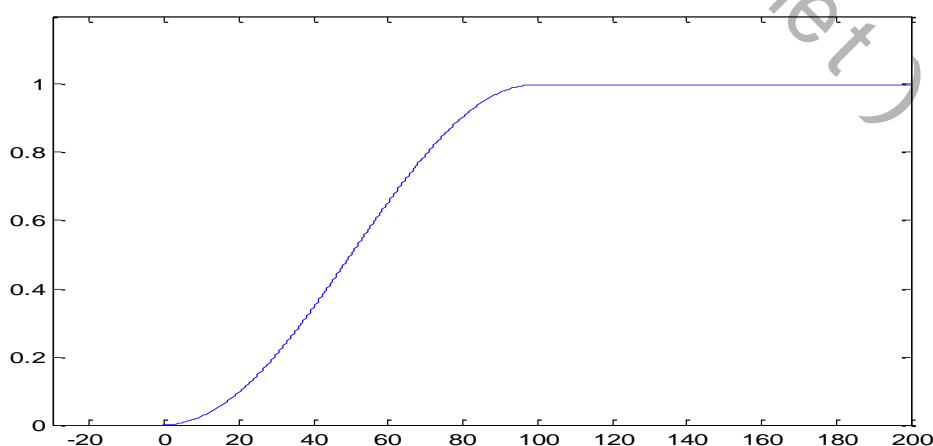
对于没有超期的同学，其罚款数目理应为 0；对于超期天数过长，给出罚款金额上限。对于中间一部分，占超期图书的大部分。为了合理的区分类别，在这里我们选用了偏大型中升岭形分布函数来处理。升岭形分布函数的公式为

$$A(x) = \begin{cases} 0 & x \leq a_1 \\ \frac{1}{2} + \frac{1}{2} \sin \frac{\pi}{a_2 - a_1} \left( x - \frac{a_1 + a_2}{2} \right) & a_1 \leq x \leq a_2 \\ 1 & a_2 < x \end{cases}$$

得到读者此次借阅的初步罚款指标的 $A(x)$ 。

其中参数 $a_1 = 0$ ， $a_2 = 100$ ；

升岭形分布函数的图形如下：





#### 4、补偿元件——读者诚信度

因为考虑到读者的差异，不同读者的借阅超期原因不同。根据读者的借阅历史记录，可以观察出读者借阅图书中超期图书的数量，以及超期天数。

首先对数据进行采样，因为我们查找到的郑州大学大约 3000 多组催还数据中，全部拿来分析较为繁琐。因此我们对这些数据进行随机抽取。

将所有数据按顺序分为 10 类；

每一类中根据随机抽样原理抽出一组样本作为研究对象。

将十个样本对应的读者，将其所借阅图书中超期图书的数量进行统计，并将超期图书的平均天数进行计算，绘出下列表格：

读者编号	超期图书数量	平均超期天数	读者编号	超期图书数量	平均超期天数
X0605273	5	67	X0703104	4	694
X0823197	10	8	Z0942231	4	8
Q0541031	3	56	Y0900418	2	11
Y0702922	8	27	X0708022	1	8
Y0802916	8	52	Y0901510	1	35

注：平均超期天数 =  $\frac{\text{每本图书的超期天数之和}}{\text{超期图书数量}}$

评价读者的诚信度应该考虑该读者的借阅历史情况，所以我们可以通过以上两个因素来确定每位的诚信度。在此，为了对每一位读者给出合理的评价，需要对两个因素进行数据的标准化处理。

超期图书数量的评价

这里，读者的超期图书数量是可以查找到的，是以固定的分数给出的，采用的数据处理方法——极差变换法，则有公式

$$r_j = \frac{C_j - \min\{C_j\}}{\max\{C_j\} - \min\{C_j\}}$$

得到对各位读者的超期图书数量的评价  $r_j (j = 1, 2, \dots, 10)$ 。

超期天数的评价

a、数据分类处理

虽然超期天数的差距过大，如果只是简单的应用极差变换法会对目标评价造成一定的影响，先对超期天数进行类别划分，根据样本数据可分为三类。分别为 A, B, C。划分标准如下：

超期天数类别	超期类别标准	超期天数下限	超期天数上限	超期标准
A	1	100	$\infty$	500

B	2	30	99	65
C	3	0	29	15

对所有数据进行类别划分后，根据以下公式：

$$F_j = S \times \frac{P_j - T_{\min}}{\bar{T}}$$

得到了超期天数的评价  $F_j$ 。 $S$  是超期类别标准， $T_{\min}$  是每类超期天数下限， $\bar{T}$  是每类的超期标准。

b、平移——极差变换法

$$f_j = \frac{F_j - \min\{F_j\}}{\max\{F_j\} - \min\{F_j\}}$$

得到了处与[0,1]之间量化标准的超期天数评价  $f_j$  ( $j = 1, 2, \dots, 10$ )。

对读者信用度的综合评价

读者的信用度对其所遵循的借阅规则造成了影响，所以我们在此特意对其进行综合评价，这样不仅有利于图书馆资源的合理利用，同时更合适的借阅规则对于每位读者来说，也可以获得较大的收益。

读者的信用度是通过读者的超期图书数量和超期天数两个方面因素决定的，读者超期图书数量过多会影响到多种类的图书资源的利用，超期天数过多会影响到这类书长期无法使需要它的读者借阅。根据经验，在此我们把超期图书数量和超期天数的权重分配确定为

$$W = (\omega_c, \omega_f) = (0.4, 0.6)$$

在这里，认为对于具有图书超期天数的读者，都是已经有了超期图书数量的读者，故超期天数的权重应大于超期图书数量的权重。则由加权求和的方法，利用公式

$$\lambda_j = (\omega_c, \omega_f)(r_j, f_j)^T$$

得到对每个读者的诚信度综合评价  $\lambda_j$ 。于是，确定读者诚信度的数学模型归纳为

$$\begin{cases} \lambda_j = (\omega_c, \omega_f)(r_j, f_j)^T \\ r_j = \frac{C_j - \min\{C_j\}}{\max\{C_j\} - \min\{C_j\}} \\ F_j = S \times \frac{P_j - T_{\min}}{\bar{T}} \\ f_j = \frac{F_j - \min\{F_j\}}{\max\{F_j\} - \min\{F_j\}} \end{cases}$$

根据此模型，我们便能在掌握读者历史借阅情况的基础上，给每位读者一个合理的诚信度评价。

## 5、修正元件

### 改进后的综合借阅指标

为了给出更加合理的借阅规则，我们仅考虑读者此次的借阅情况有所不足，所以我们在原有的指数判断基础上加入了读者诚信度。

在这里我们需要对诚信度进行一下处理，因为诚信度处于 0—1 之间，所以我们可以用 1 减去诚信度，然后作为一个求解因素，我们称之为反诚信度。

根据以上分析得到的初步罚款指标和反诚信度指标，我们可以得出综合借阅指标。根据我们对一些高校图书馆的借阅规则了解，在此我们把初步罚款指标和反诚信度指标权重分配确定为

$$U = (\omega_A, \omega_\lambda) = (0.6, 0.4)$$

在这里，认为初步罚款指标应该占有较大的权重，因为这个指标是读者此次借阅的情况，反诚信度是作为一个辅助因素考虑。则由加权求和的方法，利用公式

$$\gamma_j = (\omega_A, \omega_\lambda) (1 - A_j, \lambda_j)^T$$

根据上述模型，便可以求出每位读者此次借阅的综合借阅指标。

通过修正后，得到的综合借阅指标我们把他划分为三个层次，高，中，低。

层次高反应了读者的诚意度好，及超期天数相对较短。

综合借阅指标值	指标层次
$[0, 0.5]$	高
$[0.5, 0.8]$	中
$[0.8, 1]$	低

## 6、触发元件

触发元件的输入是此次借阅的综合借阅指标，经过触发元件对后一级——执行元件进行控制。

### 1、罚款金额

根据官方调查结果，文献借阅超期罚款的多少没有统一标准。以普通图书为例，国内大学图书馆罚金最高为 0.2 元，最低为 0.05 元，平均为 0.1 元。

所以我们将罚款金额设定为三个层次：

甲类：每天罚金为 0.2 元

乙类：每天罚金为 0.1 元

丙类：每天罚金为 0.05 元

综合借阅指数	罚款金额
高	丙
中	乙
低	甲

## 2、图书馆催还制度

借鉴国内外图书馆界的催还方法，在这里，我们将根据综合借阅指数而采取相应的催还方法。

综合借阅指数	催还方法
高	发 Email 催还、打电话
中	发信函
低	在校报或校园网进行公示，以示警告

## 3、续借天数

根据郑州大学已有的借阅规则中可以得到续借的天数为 30 天。考虑书籍类别的差异，我们已经定出了各类的续借标准天数。

考虑了读者的综合借阅指数，我们在原来的续借标准天数上进行增减幅度的调整。

综合借阅指数	续借天数幅度
高	2%
中	0
低	-2%

通过运用 matlab 编程将其视为分段函数，不同阶段的输入，对应不同的输出。

## 7、执行元件

### 1、博弈论与纳什均衡的概述

高等图书馆超期现象受诸多因素的影响，是多种要素相互作用的结果。只看到其中一个方面很难得出令人信服的结论。在这里，我们尝试用一种新的视角来分析高校图书馆超期现象进而提出合理的借阅标准。这就是博弈分析方法。

在现实生活中，人们的许多决策之间是相互依赖的。你的最优决策依赖于别人的决策，别人的最优决策依赖于你的决策；你的决策还依赖于别人认为你怎么决策，你怎么认为别人认为你怎么决策，如此等等。研究这种互动环境下的理性选择的理论叫博弈论 (theory of rational choices under interactions)。

在博弈理论中包含一个重要的概念和思想，这就是“纳什均衡”。纳什均衡(Nash equilibrium)是由美国数学家纳什首先提出并证明的。所谓纳什均衡，指的是所有参与人最优选择的一种组合。在这种组合下给定其他人的选择，没有任何人有积极性做出新的选择。纳什均衡也可以从另一个角度来理解：假设博弈中的所有参与人事先达成一项协议，等于说这个协议构成一个纳什均衡：在给定遵守协议的情况下，没有人有积极性偏离协议规定的自己的行为规则。这就是纳什均衡的理论思想。

### 2、高校图书馆超期问题的纳什均衡理论

用纳什均衡理论来分析高校图书馆超期问题，我们就会发现，在一定条件下，高校超期问题是可以续借的天数，图书馆催还，以及超期后的罚款金额之间的一种纳什均衡。这时，高校图书馆不可能大幅度增加可以续借的天数，来减少读者的超期天数。也不能耗费大量的人力资源对每位借阅超期的读者采取催还制度，但是过多的罚款金额不但不能避免超期问题，反而可能影响图书馆与读者之间的关系，从而影响图书馆的声誉，以及馆藏资源的利用率，最终影响图书馆功能的正常发挥。所以可以续借的天数，图书馆

催还制度，已经罚款金额需要满足纳什均衡要求的“协议”。在这种情况下，无论对图书馆还是读者都是有利的。

声明：通过博弈论和纳什均衡理论的分析，我们可以确定这三个执行元件确实可以控制我们的目标超期天数。由于时间的不足，我们没有根据纳什均衡理论确定出具体合适的“协议”，只是进行了定性的分析

## 13 模型推广模型改进

### 13.1 模型推广

若需把模型应用与实践，可以编出具有以上功能的软件，有利于图书馆超期现象实行有利的控制。

### 13.2 模型改进

引起图书超期的原因很多，我们只能对主要的原因进行分析。如果说原因是一组随机变量，那么我们的理想是找到他们的数学期望，任何以为某一模型求出的原因判断只是他的一个变量值，

所以我们提出任何非调查统计的原因类型分析结果只能是某一模型的结果，期合理性本身就是被用来检验的对象，而不是别的结果的标准。所以我们给出基于实际调查数据的残差修正模型，经过 1、2 次修正，从粗到精，逐步逼近那个期望值。

应用神经网络理论对原因类型分析残差修正：原因类别与超期图书的特点指标之间有很复杂的非线性关系，要能真正给出原因的很好地逼近期望值必须找到他们之间的复杂映射关系，而神经网络通过训练样本可以映射非常复杂的非线性关系，尤其有数学软件的帮助，我们可以方便地做到。

#### 参考文献：

- [1] 杨启帆 方道元，数学建模，杭州：浙江大学出版社，1999。
- [2] 杨桂元 黄己立，数学建模，合肥：中国科学技术大学，2008。
- [3] 周圣武 李金玉 周长新，概率论与数理统计，北京：煤炭工业出版社，2007。
- [4] 姜启元 谢金星，数学建模，北京：高等教育出版社，2006。
- [5] 雷攻炎，数学建模讲义，北京：北京大学出版社，1999。
- [6] 陈 丽，读者借书逾期不还的原因分析，科技情报开发与经济，2007 年第 17 卷，第 22 期。
- [7] 涂海青，章志勇高校图书馆借阅情况的统计与分析，河南图书馆学刊，2006 年 4 月。
- [8] 胡寿松，自动控制原理简明教程，科学教育出版社，2008。
- [9] 王秀霞，从图书清欠看高校图书馆执行借阅制度的人为因素，现代情报，第 7 期，2002 年 7 月。
- [10] mhtml:file:///C:/Users/hp/Desktop/图书超期情况的读者名单，请及时归还.mht

## 附录:

### ◆ 模型原理介绍:

#### 1 熵模型:

首先引入一下“熵”<sup>[1]</sup>的概念。

熵原是统计物理和热力学中的一各物理概念. 熵是分子运动无序度的度量, 熵值大, 表示系统分子运动的无序度越高. 设系统内有两种物质(二原系统), 物质 1 有  $n_1$  个分子, 物质 2 有  $n_2$  个分子, 该系统的熵值可由波尔兹曼公式计算:

$$E = k \ln Q \quad (3)$$

$Q$  是系统中两种物质分子的微观排列方式, 其值为  $Q = \frac{(n_1 + n_2)!}{n_1! n_2!}$ . 根据斯梯林公式

$\ln n! = n \ln n - n$ , 则:

$$\begin{aligned} E &= k \ln \left[ \frac{(n_1 + n_2)!}{n_1! n_2!} \right] = k(n_1 + n_2) \ln(n_1 + n_2) - k(n_1 \ln n_1 + n_2 \ln n_2) \\ &= -k \left[ n_1 \ln \frac{n_1}{n_1 + n_2} + n_2 \ln \frac{n_2}{n_1 + n_2} \right] \end{aligned} \quad (4)$$

$E$  是系统  $(n_1 + n_2)$  个分子的总熵值, 除以分子总数, 便得到系统的单位熵值:

$$e = \frac{E}{n_1 + n_2} = -k \left[ \frac{n_1}{n_1 + n_2} \ln \frac{n_1}{n_1 + n_2} + \frac{n_2}{n_1 + n_2} \ln \frac{n_2}{n_1 + n_2} \right] \quad (5)$$

令  $y_1 = \frac{n_1}{n_1 + n_2}$ ,  $y_2 = \frac{n_2}{n_1 + n_2}$  分别为系统中物质 1 和物质 2 的占有率, 则系统的单位

熵值为:

$$e = -k(y_1 \ln y_1 + y_2 \ln y_2) \quad (6)$$

扩展到多元 ( $m$  元) 系统, 其单位熵值函数为:

$$e = -k \sum_{i=1}^m y_i \ln y_i$$

以下附上基于熵理论的相关定理:

**定理 1** 若实验仅有有限种结果  $S_1, \dots, S_n$ , 其发生的概率分别为  $p_1, \dots, p_n$ , 则当

$p_1 = p_2 = \dots = p_n = \frac{1}{n}$  时, 即事件呈均匀分布时, 此实验具有最大熵。

**定理 2** 若实验为必然事件，其熵为 0。

◆ **程序：**

%%%

**程序 1：系统聚类**

```
clc,clear
a=[119,127;75,112;99,134;98,220;63,63;37,76;42,82;63,90;80,118;61,110
    69,93;120,120;107,158;332,480;52,80;83,176;73,173;24,40;81,99;53,53];
y=pdist(a,'cityblock');yc=squareform(y)
z=linkage(y);
[h,t]=dendrogram(z)
set(gca,'yscale','log','ylim',[0 1000])
```

%%%

**程序 2：求超期天数等指标**

```
%%首先导入数据文件matlabsuju.Mat
%下面编写函数
function majian (a_x)
y=a_x(1,:);
z=a_x(2,:);
y1=floor(y/10000);
y2=floor(mod(y,10000)/100);
y3=mod(y,100);
riqi=[y1;y2;y3];%借书日期
chaoqi=datetime(2010,3,30)-datetime(y1,y2,y3) ;%%每本书超期时间列表
shujuncqts=sum(chaoqi)/length(chaoqi);%平均每本书超期天数
shujuncqts=floor(shujuncqts);%取整
bar(chaoqi)%画条形图
chaoqi1=sort(chaoqi);
plot(chaoqi1)
hist(chaoqi1,50)%画直方图
%%%%%%%%%下面计算该类借书人数 n+1
n=0;
z=sort(z);
for i=1:length(y)-1
    if z(i+1)~=z(i)
        n=n+1;
    end
end
renshu= n+1;
```

```
renjuntianshu=floor(sum(chaoqi)/(n+1));
[renshu;renjuntianshu;shujuncqts]'%给出输出参数矩阵
end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

### 程序 3：求超期易度排序

```
x = [ 119    75    99    98    63    37    42    63    80    61    69
120   107   332   52   83   73   24   81   53
      127   112   134   220   63   76   82   90   118   110   93
120   158   480   80   176   173   40   99   53]

x1=x(1,:)
x2=x(2,:)
for i=1:length(x)
    x11(i)=(x1(i)-min(x1))/(max(x1)-min(x1));
end
[x111,xuhao1]=sort(x1)
%[x111,xuhao1]=sort(x11,'descend')
%%同理求的序数2
xuhao=[xuhao1;xuhao2]'
r=corrcoef([xuhao1',xuhao2']) %求相关系数
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

### ◆ 其他

郑州大学院系设置情况：

- 1 环境与水利学院  
环境工程、水利水电工程、给水排水工程、地理信息系统
- 2 物理工程学院  
物理学、应用物理学、测控技术与仪器、电子科学与技术、电子信息科学与技术
- 3 机械工程学院  
工业设计、机械工程及其自动化、机械设计制造及其自动化
- 4 电气工程学院  
自动化、电气工程及其自动化、生物医学工程
- 5 信息工程学院  
电子信息工程、通信工程、计算机科学与技术
- 6 材料科学与工程学院  
材料科学与工程、材料成型及控制工程、高分子材料与工程、包装工程
- 7 化学工程学院  
化学工程与工艺、制药工程、过程装备与控制工程、热能与动力工程、环境科学、



安全工程

8 土木工程学院

土木工程、建筑环境与设备工程

9 建筑学院

建筑学、城市规划、艺术设计(环境艺术设计方向)

10 数学系

数学与应用数学、信息与计算科学

11 化学系

化学、应用化学

12 生物工程系

生物技术

13 管理工程系

工业工程、工程管理、电子商务、物流管理

14 工程力学系

工程力学、安全工程(设备与结构安全方向)

15 文学院

汉语言文学、对外汉语

16 新闻与传播学院

新闻学、广告学、广播电视新闻学(播音与主持艺术方向)

17 历史学院

考古学、历史学(人文科学实验班)

18 外语学院

英语、日语、俄语、德语

19 公共管理学院

哲学、社会工作、行政管理、公共事业管理

20 法学院

法学

21 商学院

工商管理、经济学、国际经济与贸易、金融学、统计学、会计学

22 旅游管理学院

旅游管理

23 教育学院

思想政治教育、应用心理学、教育学

24 体育系

体育教育

25 音乐系

音乐表演、音乐学

26 美术系

绘画、艺术设计、雕塑

27 信息管理系

信息管理与信息系统、图书馆学、档案学

28 基础医学院

- 基础医学
- 29 公共卫生学院
- 预防医学
- 30 第一临床学院
- 31 第二临床学院
- 32 第三临床学院
- 33 药学院
- 药学、药物制剂
- 34 护理学院
- 护理学、护理学(专)、医学检验(专科)、药学(专科)
- 35 口腔医学院
- 口腔医学
- 36 临床医学系
- 临床医学(五年制)、临床医学(七年制)、医学影像学、麻醉学
- 37 医学检验系
- 医学检验
- 38 远程教育学院
- 39 应用科技学院
- 电子信息工程、临床医学、人力资源管理、教育学(现代教育技术方向)、计算机应用和维护(专)、应用电子技术(专)、会计电算化(专)、中英文秘(专)、社区医疗(专)、医学影像(专) 67763121
- 40 软件技术学院
- 软件技术(专)、计算机网络技术(专)、计算机多媒体技术(专)、计算机应用技术(专)、计算机信息管理(专)、会计电算化(专)
- 41 国际教育学院
- 42 成人教育学院
- 43 城市学院 郑州大学升达经贸管理学院（合作办学）
- 44 郑州大学西亚斯国际学院（合作办学）
- 45 郑州大学体育学院（合作办学）

◆ 超期天数分布图：

