

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)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：

参赛队员（签名）：

队员 1：

队员 2：

队员 3：

参赛队教练员（签名）：

参赛队伍组别：

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）：

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）：

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

题 目 基于神经网络的高校图书馆服务系统

关 键 词 借阅超期，神经网络，RFM，借阅规则

摘 要：

本文基于高校图书馆智能服务，以图书馆借阅超期现象为研究主体，建立数学模型将借阅超期书籍进行分类，分析每种类型可能的超期原因，并根据借阅超期的原因设计合理的借阅规则，最终达到减少借阅超期现象，提高图书馆资源利用率的目标。

高校图书馆的借阅超期问题要从读者主客观两方面进行考虑。首先，超期图书自身具有以下属性：图书类型，应还时间，超期天数，读者类型等。同时，借阅超期现象也受读者心理以及对借阅制度的态度的影响。模型的建立过程中，我们假设所有借阅超期读者都已经进行过续借，且图书馆馆藏资源保持相对稳定。在模型中，我们利用自组织竞争(SOM)神经网络模型从大量数据中寻找超期图书之间所存在的共性，并将其分类，然后利用改进的 RFM 模型分析借阅超期的原因。在对图书馆借阅规则的评价中，我们使用了基于径向基函数(RBF)的神经网络得到借阅规则改变后的新数据，并对其优劣进行评价。最后，我们通过分析超期原因，制定出新的借阅规则。

本文中我们解决了以下两个关键性问题。

第一，超期图书分类及原因分析模型。首先，我们获得了某大学图书馆的借阅超期催还清单。然后，我们使用 Microsoft Visual Studio 2008 提取我们需要的数据：读者证号，图书索取号，超期天数，罚款数额以及最长借阅时限等。提取了这些数据之后，我们将这些数据形成的五维矩阵输入到 Matlab7.8.0 中，使用 SOM 神经网络对其进行分析，找到数据的聚类中心，从而获得具有相似性的几类图书，达到了对图书分类的目的。接下来，我们对美国数据库营销研究所的 RFM 模型进行了改进，通过 R、F、M 三个指标分析超期原因，最终得出哪些图书容易发生超期现象，及其可能的超期原因。

第二，图书馆借阅规则评价及设计模型。为了能够评价新规则的优劣，我们首先使用 RBF 神经网络模型对已有数据进行分析，得到借阅规则与超期天数之间的潜在关系。然后，我们通过这种关系计算出已有数据在新规则下借阅超期的天数。接着我们使用第一部分中的模型对新旧两组数据进行分析比较，进而评价借阅规则的优劣。我们还通过分析超期原因，以及各种超期原因最有可能的超期天数区间，对借阅规则进行重新设计，以便达到减少借阅超期现象，提高图书馆资源利用率的目的。

最后，我们对模型进行了分析和改进。我们提出了如何将模型中没有涉及的因素考虑进来以增加模型的准确性和可信程度。同时我们还根据我们的分析，对图书馆的管理提出了可行性建议，使高校图书馆能够更好地为师生提供服务。

参赛队号 1539

所选题目 C

参赛密码 _____
(由组委会填写)

Abstract

Based on library intelligent services of colleges, our paper mainly studies the phenomenon of library overdue borrowing. We build a mathematical model to make the classification of overdue books and analyze possible overdue causes for each type. And we design a reasonable borrowing rule according to overdue causes, which ultimately reduces the phenomenon of overdue borrowing and enhance utilization of library resource.

The overdue borrowing of college libraries should be considered from both subjective and objective aspects. First, the overdue books have their own properties, such as book category, deadline of borrowing, overdue days, reader category and so on. Meanwhile, it is also affected by the readers' psychology and their attitude towards borrowing rules. We assume that all borrowings has been renewed, and the library resource remains relatively stable. In the model, we use Self-Organizing Mapping (SOM) Neural Network model to extract common characters from the large amount of data and classify into different types. Then, use the improved RFM model to analyze the causes of overdue borrowing. When evaluating library borrowing rules, we make use of the Radial Basis Function (RBF) Neural Networks to predict the new data, and evaluate its quality. Finally, we design new rules according to the overdue causes.

In this paper, we solve two critical problems as follows.

Firstly, the classification of overdue books and causes analysis model. First of all, we get access to the list of overdue books. Then, we use Microsoft Visual Studio 2008 to extract data we need, such as reader's number, call number of books, amount of fine and longest borrowing time. After extracting the data, we input the five-dimension matrix made by the data to the Matlab7.8.0. Using SOM Neural Networks to analyze the data, we get some cluster centers. In this way, we get the classification of the books with similarity and get to our aim. Next, we improve the RFM model of American Database Marketing Institute. We use the three standards R, F, M to analyze overdue causes. Finally, we get which kind of books becomes overdue easily and the causes of it.

Secondly, evaluation and design model of borrowing rules. In order to evaluate the quality of the new borrowing rules, we use RBF model to analyze the present data first. And we get the potential relations between borrowing rules and overdue days. Then, we work out the overdue days under the new borrowing rules according to the relation. Next, we using the model of Part 1 to compare and analyze the two groups of data. Thus, we can evaluate the quality. We also redesign the borrowing rules by analyzing overdue causes and the possible overdue intervals of different causes. In this way, we reduce the overdue phenomenon and improve the utilization of library resource.

Finally, we analyze and improve the model. We propose how to put the factors not involved before into the model, in order to increase the accuracy and credibility. Meanwhile, we put forward some practical proposals to the management of college library. Therefore, it can provide better services for students and teachers.

一、问题重述

图书馆源于保存记事的习惯，是为读者使用文献而提供的专门场所。高等院校图书馆是一所院校的“心脏”，它为教学科研而服务，是全校师生的共享资源。高等院校图书馆具有以下特性：服务对象的稳定性、信息资源的专业性、文献利用的集中性及读者需要的阶段性[1]。随着科技的发展，信息时代的到来，用户对信息资源的需求量越来越大，图书馆中的各类信息资源的利用率也越来越高。高校图书馆的一些不良现象也随之而来，比如对书籍的损坏、借阅超期等。

借阅超期行为是指用户在图书馆借阅图书资源过程中，由于主客观原因，超过图书馆规定的此类图书资源借阅期限的行为。读者享有使用文献资料的权利，同时也负有不得侵害其他读者正当使用图书资源的义务[1]。读者的超期行为使图书被少数用户所占有，影响了这部分图书的正常流通，进而影响了整个图书馆的资源利用率。

在搜索引擎下输入“图书超期”等关键字，成千上万的网页被检索出，大多都是超期读者的信息列表。这说明超期现象比较严重，需要得到适当的解决。处理超期罚款是现在普遍采用的一种处理方法，但在处理过程中，处理不当会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，最终影响读者对图书馆功能的正常使用。

随着信息化的发展，图书的预约服务也逐渐普及。图书的预约与借阅超期的现象也有很大的内在联系。试通过数学模型来研究以下的问题：

- (1) 对于借阅图书超期，原因众多。考虑到高校的特点，大体上可以把超期原因分为三类：1、无意错过还书期限；2、由于教学科研需要，长期使用；3、由于罚款数目较少，所以故意拖延。请通过给出的数据或自行查找相关数据，建立合理的模型，分析哪些书容易发生超期现象，以及当发生超期时，分析其可能的原因。以便图书馆采取相应的措施改进服务。
- (2) 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。试根据问题(1)中的模型评价这种借书规则的优劣。请设计一种更加合理的借阅规则。

二、模型假设

- 由于原有数据(包括题目提供的数据与我们搜索的数据)中，没有提供超期图书是否已经被续借，所以我们假设在原有数据中，所有超期借阅读者都已经进行了续借。
- 假设高校图书馆的服务对象只有老师与学生，且教师与学生的借阅习惯大致相同，即数据分析中忽略教师与学生的图书借阅超期现象差异。
- 高校图书馆的服务对象和图书馆的馆藏图书资源保持相对稳定。

三、符号说明

i 表示图书类别

k 表示使用神经网络划分出的类别总数

a_{ij} 表示神经网络划分出的第 j 个类别中 i 类图书的数量

w_j 表示第 j 个类在总图书数中所占比例

$sum(j)$ 表示第 j 个类别中的图书数量

x 表示超期天数

四、问题分析

本题所要解决的问题对象是高校图书馆，而高校图书馆与其他公共图书馆有所区别，具有以下特性：

- 服务对象的稳定性
- 信息资源的专业性
- 文献利用的集中性
- 读者需要的阶段性

首先，高校图书馆的服务对象主要是教师和学生，图书资源由教师和学生共享。高校的规模在一定时间内是保持不变的，所以图书馆的服务对象保持相对稳定。教师和学生借阅书籍的范围主要与其学习、研究的方向有关，按照中图法对图书进行分类，图书可分为22大类，每一类针对不同的科学领域，也就对应相应教师或学生不同的专业方向和研究方向等。相关领域的书籍对某一知识或者技术的介绍、讲解程度不一，初学者和研究人员可以根据需要选择借阅不同程度的书籍。

另外，在高校中随着课程的进行，学生所学科目不断变化，所学知识也不断地深入，因此，教师和学生对于书籍的借阅具有阶段性的特点。由于开设不同课程的原因，某一图书需求量可能在某一阶段内增加或减少。同时，由于学生年级的变化，同一类图书的主要借阅人群可能会在不同年级间变换。由于考试的需要，在每学期期末的时候，图书借阅量也可能因进入复习阶段而有所增加。

借阅超期的原因要从读者主客观两方面进行考虑。首先，每本超期图书都有各自的图书类型，应还时间，超期天数，读者类型等信息。同时，借阅超期也受读者心理以及读者对借阅制度的态度的影响。

解决这一问题的核心在于，我们要根据已知的超期图书借阅信息，包括：图书类别、超期时间、应还日期，读者类型等，从大量数据中寻找超期图书之间所存在的共性，将具有共性的图书分成一类，并通过这种共性总结出此类图书借阅超期可能的原因。如果我们能够分析得到容易超期的图书类别以及其超期的原因，进而改善图书馆借阅制度，那么就可以使图书资源得到最大效率的利用，从而维护和改善图书馆与读者之间的关系以及图书馆的声誉。

五、模型建立与求解

5.1 超期图书分类与原因分析模型

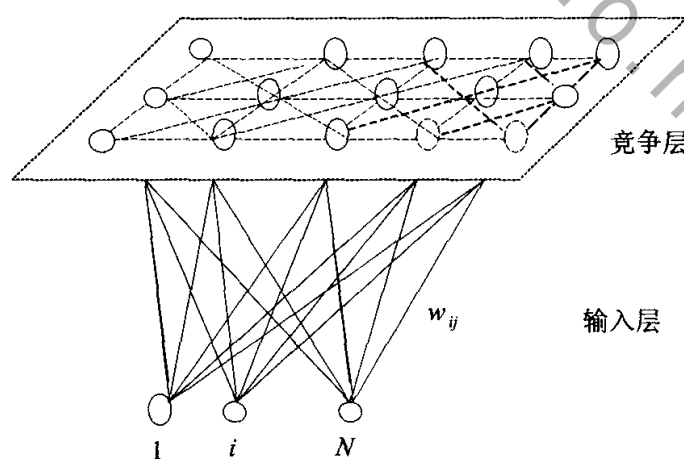
5.1.1 理论的选择与介绍

图书超期的原因很多，但是考虑到高校的特点，原因大体上分为三大类：1、无意错过还书期限；2、由于教学科研需要，长期使用；3、由于罚款数目较少，所以故意拖延。这些原因的分类都是人脑根据外界情况分析而得到的结果。每位读者借阅图书具有不同的目的，对图书馆的借阅规则认识也不尽相同，对于图书超期后果抱有不同的态度等等一些原因造成了最后的超期。超期是读者在外界客观因素的影响下主观造成的，我们可以假设如果每位读者对借阅图书足够重视，并且遵守借阅规则，那么“图书超期”现象将很少发生。

对于这种人的主观因素占主导作用的问题中，为了让我们模型更好的作用于实际情况，解决实际问题，我们选择基于模拟人脑思维方式的神经网络建立模型。

大量数据中寻找超期图书之间所存在的共性，分析它们的超期原因；我们选择具有较强聚类 and 容错能力的自组织特征映射（Self-Organizing Mapping, SOM）神经网络理论进行图书超期原因分类。SOM 网络是一种具有无监督学习功能的神经网络，SOM 网络可以采用各神经元（特征参数）之间的自动组织去寻找各类型间固有的、内在的特征。从而可以根据人的思维方式特点对超期图书原因进行分类，寻找图书种类与图书超期原因之间的内在联系。

SOM 的网络结构如图【1】所示，它由输入层和输出层组成。输入层神经元数为 n ，竞争层由 m 个神经元组成的一维或者二维平面阵列，网络是全连接的，即每个输入节点都同所有的输出节点相连接。应用于我们模型中的 SOM 网络由 4 部分组成：（1）用于接受的输入单元，形成对输入矩阵的判别函数；（2）比较选择机制，通过判别函数选择一个具有较大函数输出值的处理单元；（3）局部互连作用，用于同时激励被选择的处理单元及其最邻近的处理单元；（4）自适应过程，用于修正被激励的处理单元的函数，以增加其对应于特定输入的判别函数的输入值[5]。



图【1】

SOM 神经网络理论应用于本问题，能够将超期图书信息输入矩阵，在竞争层映射成一维图形，并保持其拓扑结构不变；该神经网络通过对输入矩阵的反复学习可以使权重向量空间与输入矩阵的概率分布趋于一致（概率保持性）。SOM 网络的竞争层各神经

元竞争对输入矩阵的响应机会，获胜神经元有关的各权重朝着更有利于它竞争的方向调整“即以获胜神经元为圆心，对近邻的神经元表现出兴奋性侧反馈，而对远邻的神经元表现出抑制性侧反馈，近邻者相互激励，远邻者相互抑制”。从而可以将具有共性的超期图书聚类到一起，每本超期图书属于确定的一种类别，该类别中潜藏着超期原因的共性和相似性。

5.1.2 图书超期原因分类模型

第一步：建立输入矩阵

根据本问题中提供的数据我们提取其中的重要信息，建立输入矩阵。分析可知，借书证号、索书号(对应中国图书分类法分析了每本书的所属类别)、超期天数、罚款金额、规定借书时长(包括续借时间)是我们从已有数据中直接获得的，能够反映图书超期内在原因的重要信息。我们提取这些已知的重要信息得到模型的输入矩阵：

输入矩阵= $\{\text{借书证号 } J, \text{索书号 } S, \text{超期天数 } C, \text{罚款金额 } F, \text{规定借书时长 } G\}$
即：

$$Input_i = \{J_i, S_i, C_i, F_i, G_i\}$$

我们将搜集到的数据(郑州大学图书馆借阅超期数据)提取出 4×3940 大小的输入矩阵(见附件)，作为 SOM 神经网络的输入，进行图书馆超期图书分类。

第二步：SOM 算法处理

- (1) 将权值 W_{ij} 赋予小的随机初始值；

设置一个较大的初始邻域 N_c ，并设置 SOM 网络的循环次数 T ；

- (2) 给出一个新的输入模式 $X_k: X_k = \{X_{1k}, X_{2k}, \dots, X_{nk}\}$ ，输入到网络上；

- (3) 计算新输入模式 X_k 和所有的输出神经元的距离 d_{jk} ，并选择和 X_k 距离最小的神经元 C ，即

$$\|X_k - W_c\| = \min_j \{d_{jk}\}$$

则 C 即为获胜神经元；

- (4) 更新结点 C 及其领域结点的连接权值

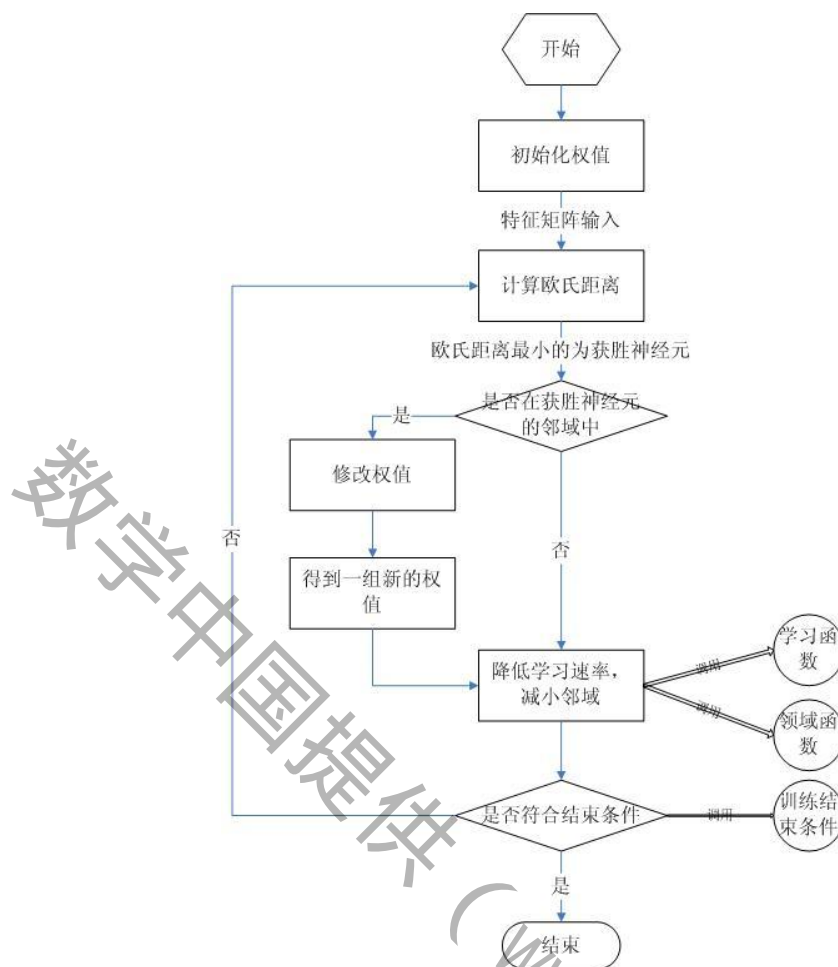
$$W_{ij}(t+1) = W_{ij}(t) + \eta(t)(X_i - W_{ij}(t))$$

其中 $0 < \eta(t) < 1$ 为增益函数，随着时间逐渐减小；

- (5) 选取另一个学习模式提供给网络的输入层，返回步骤(3)，直到输入模式全部提供给网络；

- (6) 令 $t = t + 1$ ，返回步骤(2)，直至 $t = T$ 为止。

SOM 算法流程图如图【2】所示：



图【2】

在自组织映射模型的学习中，通常取网络循环次数为 $500 \leq T \leq 10000$ 。 N_c 随着学习次数的增加逐渐减小。增益函数 $\eta(t)$ 也即是学习率。由于学习率 $\eta(t)$ 随时间的增加而渐渐趋向零，因此，保证了学习过程必然是收敛的。一般要求：

$$\sum_{k=n}^{\infty} \eta(t+k) = \infty, \quad \sum_{k=0}^{\infty} \eta^2(t+k) < \infty$$

其中： $0 < \eta(t+k) < 1$ ， $k=1, 2, \dots, \infty$ 。

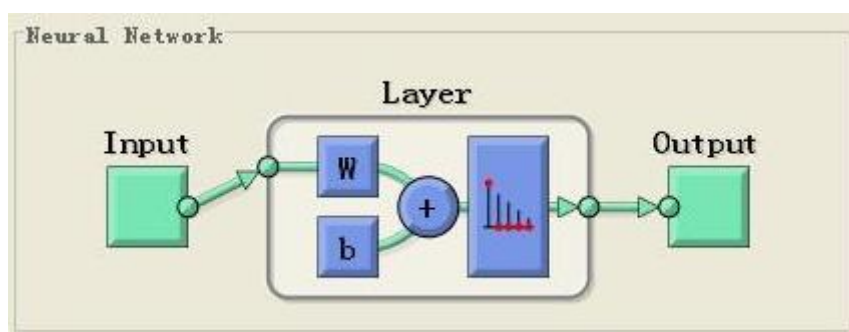
在实际的权系数自组织过程中，一般，对于连续系统取：

$$\eta(t) = \frac{1}{t}$$

对于离散系统，则取：

$$\eta(t+k) = \frac{1}{t+k}。$$

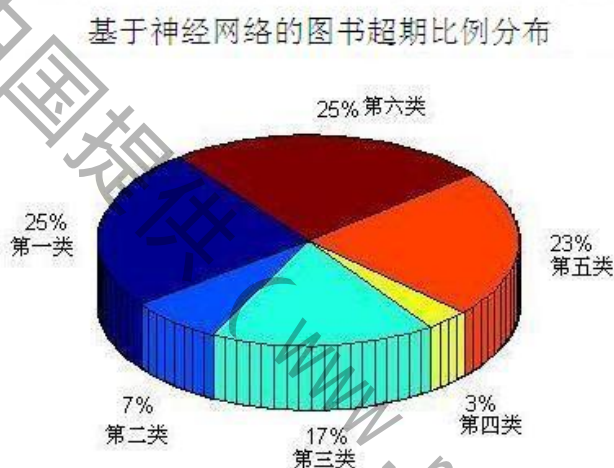
我们运用SOM算法将第一步得到的输入矩阵进行处理，通过MATLAB自组织映射神经网络模拟，如图【3】所示。



图【3】

第三步：获得分类结果

通过以上SOM神经网络对输入矩阵的处理，得到超期图书原因的分类结果。我们通过MATLAB模拟将输入矩阵（即图书超期数据）划分为六大类，如图【4】所示，在每一个类内部，超期数据之间都具有有一些相似性。对这种相似性的研究我们将在RFM模型中详细讨论。



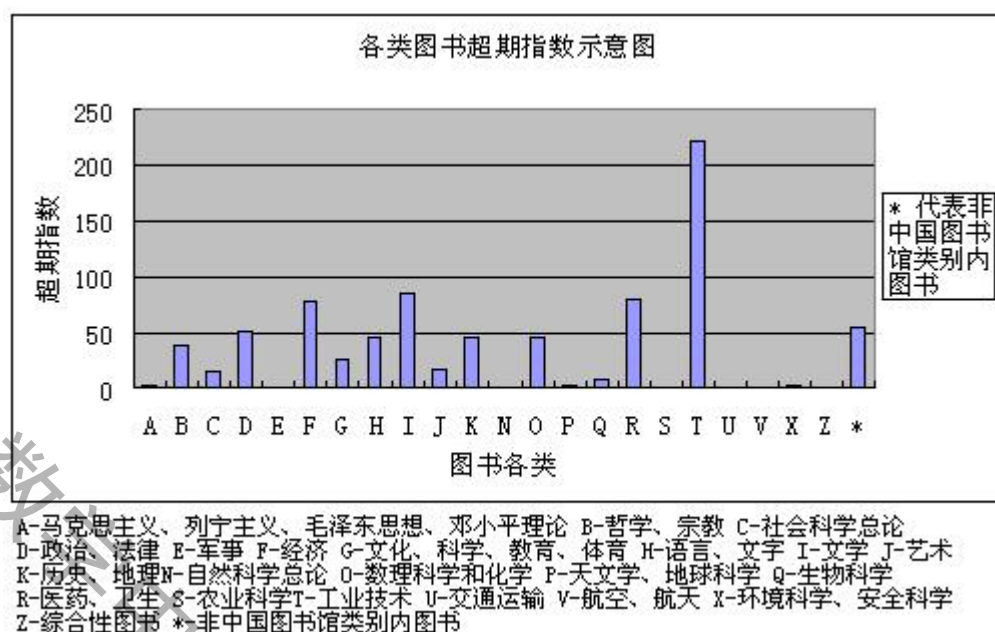
图【4】

在计算图书超期比例分布的基础上，我们进一步计算得到如图【5】所示的图书超期指数示意图。我们按照中国图书馆分类法将全部图书分为23类，其中非中国图书馆类别的图书被划分为一类。利用在SOM神经网络分类中得到的数据，我们用下公式分别计算出这23类图书的超期指数 $E(i)$ 。

$$E(i) = \sum_{j=1}^k a_{ij} w_j$$

$$w_j = \frac{\text{sum}(j)}{n}$$

其中， i 表示图书类别， k 表示使用神经网络划分出的类别总数， a_{ij} 表示神经网络划分出的第 j 个类别中 i 类图书的数量， w_j 表示第 j 个类在总图书数中所占比例， $\text{sum}(j)$ 表示第 j 个类别中的图书数量。



图【5】

观察分析图【5】可以发现，T（工业技术）、I（文学）、F（经济）、R（医药、卫生）等四类图书的超期指数较高，也就是说这四类图书与其他类别图书相比较更容易出现超期现象。

5.1.2 RFM 模型

在通过神经网络模型得到对数据的分类之后，我们使用RFM模型来分析引起各类借阅超期现象的原因。

RFM模型是美国数据库营销研究所Arthur Hughes一项研究，它通过三个要素：最近一次消费（Recency），消费频率（Frequency），消费金额（Monetary），对客户数据库进行分析，这三个要素构成了数据分析最好的指标。

我们对RFM模型进行了改进，将其应用在图书超期原因分析中，采用RFM指标作为评价指标，其中近度(Recency)衡量读者最近一次超期距现在的时间，近度时间太长被认为是图书超期行为已经发生变化的信号；频度(Frequency)衡量在一段给定的时间内读者出现借阅超期的次数，即读者借阅超期频率，这一指标用以识别是否存在频繁超期现象，它可以作为图书归还制度的主要依据并从某种程度上评测读者的责任心；值度(Monetary)衡量每个类别（由神经网络算法划分出的类别）中所有图书平均的借阅超期天数，这一指标用来评估图书的受欢迎程度。

本文中的超期类型识别分别计算每个超期类型的RFM 的均值和所有超期记录的RFM 总平均值，然后分别使用每个超期类型的RFM 均值与的RFM 总均值进行比较。如果单个类型的均值大于总均值，便给该值一个向上的箭头标记“↑”，反之则标记“↓”。

对于R 值来说，低于总均值意味着该类型读者近期有图书超期记录；而对于F 值而言，高于平均值表示该类型的读者经常拖延归还图书；对于M 值来说，高于总均值意味着该类型读者借阅的图书都是比较受欢迎的图书，我们可以认为这种类型的图书比较容易超期。RFM 指标一共存在8 种可能的组合，根据其每种组合的表现，可以划分为七种超期原因：

1. 第一类—— $R \downarrow F \uparrow M \uparrow$ ：这种类型可被认为是近期有超期记录且频繁拖延，累计

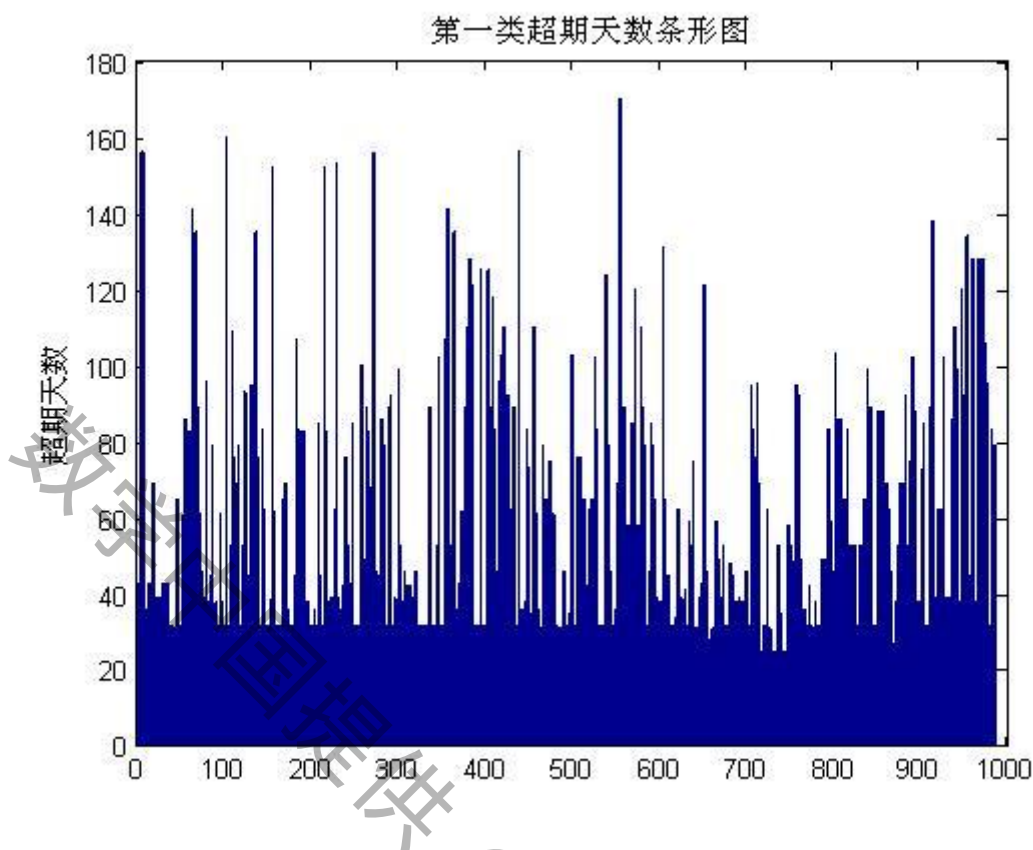
借阅量较大的读者,更多借阅的是热门图书。综合以上因素,其超期原因可视为由于个人兴趣或教学长期使用,或因为罚款数额较小故意拖延造成借书超期。

2. 第二类—— $R \downarrow F \uparrow M \downarrow$ ：这种类型可被认为是近期有超期记录且频繁拖延,累计借阅量较大的读者,更多借阅的是比较冷门的图书。综合以上因素,其超期原因可视为由于罚款数额较小,故意拖延造成借书超期。
3. 第三类—— $R \downarrow F \downarrow M \uparrow$ ：这种类型可被认为是偶有拖延的读者,更多借阅的是比较热门的图书。综合以上因素,其超期原因可视为由于个人兴趣或教学长期使用或无意错过还书时间造成借书超期。
4. 第四类—— $R \downarrow F \downarrow M \downarrow$ ：这种类型可被认为是近期有超期记录但偶有拖延的读者,更多借阅的是比较冷门的图书。综合以上因素,其超期原因可视为由于无意错过还书期限造成借书超期。
5. 第五类—— $R \uparrow F \uparrow M \uparrow$ ：这种类型可被认为是近期无超期记录但以往频繁拖延的读者,更多借阅的是比较热门的图书。综合以上因素,其超期原因可视为图书归还的特定困难(例如读者到异地进行实习学习)造成借书超期。
6. 第六类—— $R \uparrow F \uparrow M \downarrow$ ：这种类型可被认为是近期无超期记录但以往频繁拖延的读者,更多借阅的是比较冷门的图书。综合以上因素,其超期原因可视为不熟悉图书借阅归还制度造成借书超期。
7. 第七类—— $R \uparrow F \downarrow M \uparrow$ 或 $R \uparrow F \downarrow M \downarrow$ ：这两种类型可被认为是近期无超期记录且偶有拖延记录的读者。综合以上因素,其超期原因可视为由于无意错过还书期限造或者不熟悉图书借阅归还制度成借书超期。

我们通过对实际数据的计算,得到以下表格中每一类图书RFM值与总体RFM值的比较,这样我们就可以利用RFM模型对实际数据中各个类别的超期原因进行分析:

类别	图书数量	近度 (R)	频度 (F)	值度 (M)	比较结果
1	984	29.0306	0.0059	51.8171	$R \downarrow F \uparrow M \downarrow$
2	281	31.3574	0.0024	93.1993	$R \uparrow F \uparrow M \downarrow$
3	682	25.4164	0.0020	75.2287	$R \downarrow F \downarrow M \downarrow$
4	133	41.0602	0.0012	216.8872	$R \uparrow F \downarrow M \uparrow$
5	889	27.2694	0.0008	164.1957	$R \downarrow F \downarrow M \uparrow$
6	971	24.1977	0.0009	125.7734	$R \downarrow F \downarrow M \uparrow$
	3940	29.7220	0.0022	121.1836	

从比较结果上看,类别1属于超期原因的第二类,即由于罚款数额较小,故意拖延造成借书超期;类别2属于超期原因的第六类,即不熟悉图书借阅归还制度造成借书超期;类别3属于超期原因的第四类,即无意错过还书期限造成借书超期;类别4属于超期原因的第七类,即无意错过还书期限造或者不熟悉图书借阅归还制度成借书超期;类别5、6属于超期原因的第三类,即由于教学科研原因长期使用或无意错过还书时间造成借书超期。



图【6】

通过对超期原因的分析，我们已经知道第一类图书的超期原因主要是罚款数额较小，故意拖延造成借阅超期。图【6】为第一类超期天数条形图，观察图中超期天数的分布情况，综合考虑超期原因，我们可以得到结论：由于罚款额较小，故意超期最有可能的超期天数区间为30-100天。通过同样的方法我们分别观察分析六个类别的超期天数条形图，得到以下数据：

表【1】

超期原因	类别	最有可能的超期天数区间（天）
故意超期（罚款额较小）	1	30-100
无意超期	3,4,5,6	0-30
教学原因长期使用	5,6	100以上
不熟悉图书馆制度	2,4	50-350

5.2 借阅规则的设计模型

5.2.1 对借阅规则的评价

● 评价方法

我们认为，图书馆的借阅规则和借阅者的潜意识在一定程度上存在相互作用，图书馆的借阅规则通过潜意识过程作用于读者，决定读者是否拖延还书时间以及超期天数。

读者的思维意识及上文提到的潜意识是一个非线性的过程，我们可以将它理解为一

个非线性的函数。所以，我们可以利用基于径向基函数(RBF)的神经网络的特性完成该非线性函数的逼近和内插。RBF 神经网络的理论认为，在一个物理过程的输入和输出之间，客观上存在着一个固定但是未知的映射关系。而 RBF 神经网络的学习过程相当于输入至输出映射的重建过程，训练的目的在于使其能够拟合成一个超曲面，根据该超曲面对输入模式给出相应的输出模式，从而实现基于一组给定稀疏数据点的超曲面重建[6]。

下面介绍我们的评价方法：

第一步：我们使用 RBF 神经网络来模拟获得读者的这种潜意识，根据高校图书馆的制度（制度参数：借阅天数、罚款数额等）来训练 RBF 神经网络，其输出的目标是已知的图书超期天数，训练的结果是获得从输入到输出的映射关系，即读者潜意识。

第二步：我们使用新的图书馆制度作为已训练后的 RBF 神经网络的输入，利用获得的映射关系得到新规则下可能出现的超期天数。

第三步：利用上述方法得到的超期天数，并根据超期天数和图书馆制度计算得到罚款数额，从而组成图书超期原因分类模型的输入矩阵，用来模拟新规则下的数据分类。

● 优劣分析

根据上述建立的模型，采用“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”的借书规则，通过以上评价方法，最终可以得到图【7】中所示的各类图书超期原因比例分布图：



图【7】

通过对第二部分题目中图书借阅规则的分析，发现有如下特点：

1. 该借书规则易于理解，计算简单，便于宣传与实施。
2. 该借书规则没有考虑到不同类型的图书超期情况的特点。
3. 该借书规则没有考虑到各种超期原因对罚款金额的敏感度差异。

结合图【7】，可以发现：

1. 在该规则下，类型一占 25%，即有故意拖延或教学任务造成的图书超期占有所有图书超期的 25%，可以认为此规则导致过多的读者故意拖延归还图书。
2. 在该规则下，类型三、四、五、六占 70%，可估计无意超期造成的图书超期占有所有图书超期的 35%，可以认为这样的规则导致无意超期占有有一定比例。

综上所述，我们可以得到借阅规则——“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”的优劣如下：

优点：易于理解，计算简单，便于宣传与实施；

缺点：故意拖延占图书超期的比重较大；
没有考虑到各种超期原因对罚款金额的敏感度差异；
图书资源利用率不高。

5.2.2 新借阅规则的设计

对于不同的超期原因我们可以采用不同的管理措施：

对于超期指数较高的图书类型，我们可以适当延长借阅时限以及续借时限，在上面的分析中我们已经分析出 T、I、F、R 等四类图书的超期指数较其他类型高，因此，而我们将这四类图书的借阅时时限和续借期略作调整，具体调整如表【2】所示。

表【2】

图书类型	借阅时限	续借次数	续借期
T、I、F、R 四类	60 天	1 次	30 天
其余类型	30 天	1 次	15 天

我们可以根据不同超期原因的最有可能超期天数区间来设置不同的罚款额，根据表【2】，对于无意超期的读者我们可以采取超期后每本每天0.1元的罚款额，而对于超期区间在30-100天的故意超期读者我们可以将加大处罚力度以示警告，把罚款额提升到每天每本0.2元，而对于超期超过100天的读者，为了鼓励他们还书，我们将罚款额再次降到每本每天0.1元。同时，我们还可以定期开设“免责日”，通过这种方式鼓励长期拖欠的读者还书。以上对罚款额的设置，我们用一下函数 $F(x)$ 表示：

$$F(x) = \begin{cases} 0.1x & (0 < x \leq 30) \\ 0.3x - 6 & (30 < x \leq 100) \\ 14 + 0.1x & (x > 100) \end{cases}$$

其中， x 表示超期天数

新规则测试：

根据以上新规则，我们修改原有数据中的罚款数和拖延天数，再次利用自组织竞争神经网络模型对数据进行分析得到的类别信息，并与原有类别信息进行对比，从而可以比较改进前后的优劣变化。



图【8】

通过图8与图7的对比，可以得到以下相关信息：

1. 相对于原规则，新规则中类型一比例降低了 3%（由 25%降低到 22%），即故意拖延或教学任务造成图书超期占有所有超期的比重下降了 3%，可以认为新规则减少了读者故意拖延图书归还的现象。
2. 相对于原规则，新规则中类型三、四、五、六占 72%，可估计无意超期造成图书超期占有所有超期的比重为 40%，可以认为新规则使得无意超期的比重略有增加。
3. 另外，根据计算数据显示，约 30%超期记录在新规则下没有超期，可以认为新规则有效的减少了超期现象的发生。

综合以上因素，我们有理由相信，新规则使得由于故意拖延而造成的超期现象比例下降，无意拖延超期现象比例上升，且有效减少超期现象发生，是更加合理的借阅规则。

为了图书馆的工作取得效果，在上述的借阅规则的基础上，我们还可以增加以下辅助措施：

1. 做好新生的入馆教育工作。
强调图书馆的各项规章制度，让读者了解它的重要性，在以后利用图书馆的过程中避免违规行为的发生，形成良好的借阅习惯。
2. 定期做好图书的催还工作[4]，可以采用电话催还，邮件催还等方式。
3. 完善罚款制度。
 - a) 放宽宽限期。宽限期就是借期过了几天之内如果还了书，可以不罚款。
 - b) 除了罚款外，对于超期图书读者，暂时限制其继续借阅的权限。
4. 建立奖惩制度。
图书馆可以从罚款中挪出一部分，作为读者的奖励基金。第一，奖励借阅专业书比较多的学生；第二，奖励借阅图书较多而又不超期的读者，每年根据记录，出一张排行榜，给予精神上和物质上奖励；对于借阅图书较多，而又不超期的读者，可以增加他们借阅册数。
5. 公布超期罚款的收支
超期罚款的收入应专款专用，并且每年一次应将罚款的收支情况公布出来。这是消除读者对图书馆误解的最好方法，建议主要用于购买新书和用于对读者的奖励。
6. 所借图书在寒暑假、法定节假日及学生外出实习期间（实习前办理相关手续）达到借阅期限者，不做逾期处理，在返校后7日内归还。
7. 可以根据读者类型（如：教职工、本科生、研究生等），对借阅书籍最大册数、借阅时限、续借次数、续借期等进行合理

六、模型的评价与改进

【图书超期原因分类模型的分评价与改进】

该模型首先根据中图法将所有超期图书进行分类，然后对各类超期图书进行统计与分析，得出各类超期图书的总数，进而得知容易超期图书的类别。这样的分析方法，得出的结果比较客观；但是由于分类只有23类，这些结果不免有些粗糙；所以可以进一步对图书进行分类。例如我们可以在维度中添加日期的月份。因为高校有寒暑假以及期中期末考试等，在这两类特殊时期比平时更加容易出现借阅超期现象。加入这一维度，我们在分析超期原因时，就将这一因素也考虑在内，这回事我们的模型更加准确，可信度更高。

在对超期原因分类进行分类时，我们使用了 SOM 神经网络，这种神经网络的特性使需要在学习速度和最终权值向量的稳定性间进行折中。SOM 神经网络的神经元数目，如果选择不当，可能使有的神经元的初始权值向量离输入向量太远以至于它从未在竞争中获胜，也从未得到过学习，这将导致“死”神经元的产生；对神经元数目的选择和如何避免死神经元的产生是进一步改进方向。

【借阅规则的评价方法评价与改进】

借阅规则的评价方法能够良好的模拟出在新借阅规则下的超期天数，从而为新借阅规则的评估提供较准确的数据依据，使得我们可以利用 SOM 神经网络对超期图书原因进行分类解释。但其中使用的 RBF 神经网络进行了模型验证时，由于需要模拟的数据量较大，导致 RBF 隐层的神经元数目较多，计算量庞大。同时 RBF 神经网络进行函数逼近时，如果 RBF 函数的扩展速度越大，函数的拟合就会越平滑，但是过大的扩展速度就意味着需要非常多的神经元来适应函数的快速变化，而如果扩张速度设定过小，则需要很多神经元来适应函数的缓慢变化，网络的性能就不会很好。所以我们在实验时需要对扩展速度进行调优，需要我们找到更合适的方法进行调优选择。

七、参考文献

- [1] 鲍婷婷,涂娟,刍议高校图书馆超期借阅处罚制度,农业图书情报学刊[J],2008,(02),172-173+188
- [2] 刘静春,余玲,数字图书馆中读者偏好智能分析系统,情报杂志[J],2008,(05),107-109
- [3] 陈伯成,梁冰,周越博,林析泉,赵延,自组织映射神经网络(SOM)在客户分类中的一种应用,系统工程理论与实践[J],2004,(03),8-14
- [4] 陈竞,高校图书馆图书超期借阅超期罚款中存在的矛盾及对策——以上海城市管理职业技术学院图书馆为例,农业图书情报学刊[J],2009,(08),89-91
- [5] 李春华,李宁,史培军,自组织特征映射神经网络原理和应用研究,北京师范大学学报(自然科学版),2006,(05),543-547
- [6] 张德丰,MATLAB神经网络应用设计,机械工业出版社,2009,163-165

附件：

附件一：

```
function net = myNetWork(data)
% 本函数用来生成一个 SOM 神经网络，并用于数据的分类
% data 是输入的特征矩阵；net 是分类后的神经网络
```

```
b = [0 999999;1 23;0 1500;0 150;40 100];
```

```
net = newsom(b,[2,3]);
```

```
w = net.IW{1,1};
```

```
plot(data(2,:), data(3,:),'+r');
```

```
title('原始数据分布及聚类中心');
```

```
xlabel('图书分类');
```

```
ylabel('超期天数');
```

```
hold on;
```

```
circle = plot(w(:,1),w(:,2),'ob');
```

```
figure,plotsom(net.IW{1,1},net.layers{1}.distances);
```

```
title('初始网络权值');
```

```
axis([0 1 0 1]);
```

```
xlabel('w(i,1)');
```

```
ylabel('w(i,2)');
```

```
net.trainParam.epochs = 100;
```

```
net = train(net, data);
```

```
w = net.IW{1,1};
```

```
figure,plot(w(:,1),w(:,2),'ob');
```

附件二：

```
function day = mynewgrnn()
```

```
% 本函数使用 RBF 神经网络模拟新规则下图书的可能超期天数
```

```
% day 是神经网络模拟出的新规定下的超期天数
```

```
% 3.txt 文件和 4.txt 是本地磁盘中的文件
```

```
% p 是旧规则下的相应参数，target 是旧规则下的相应超期天数，m 是新规则下的参数
```

```
fid1 = fopen('3.txt', 'r');
```

```
p = fscanf(fid1, '%f',[7,3940]);
```

```
fclose(fid1);
```

```
fid2 = fopen('8.txt', 'r');
```

```
m = fscanf(fid2, '%f',[7,3940]);
fclose(fid2);
```

```
fid3 = fopen('5.txt','r');
target = fscanf(fid3, '%f',[1,3940]);
fclose(fid3);
```

```
net = newrb(p, target, 1);
day = sim(net,m);
```

附件三：部分原始数据——数据来源：郑州大学图书馆

读者证号	读者姓名	题名	索取号	文献条码号	应还日期	馆藏地点
C0501012	常欢欢	PHP 网络编程技术与实例	TP312PH/C163	4010281274	20061218	中心馆自科书库
C0501012	常欢欢	ASP.NET 精彩编程百例	TP393.092/Y375	4010143069	20061218	中心馆自科书库
C0501012	常欢欢	宇宙奥秘探索	P159-49/L262	4010356435	20061218	中心馆自科书库
C0757015	徐熠旋	研究生英语精读教程. 上: 教师参考书	41.68054/892+2	1010267549	20080606	南区社科
C0825401	郭伟廷	惠特曼名作欣赏	47.63023/288	4010032931	20100315	中心馆社科书库
C0825401	郭伟廷	消费经济学	F014.5/W652	4010361257	20100315	中心馆社科书库
C0825401	郭伟廷	消费经济学	F014.5/T623.02	4010374320	20100315	中心馆社科书库
C0825401	郭伟廷	私有化的局限	F014.1/W579	4010435852	20100315	中心馆社科书库
C0825401	郭伟廷	消费伦理研究	B82-053/H215	4010601564	20100315	中心馆社科书库
C0874035	丁治云	中西医结合变态反应病学	R593.105/C591	4010170374	20100311	中心馆自科书库
C0874035	丁治云	皮肤病治疗经验病例荟萃	R751.05/T126	4010438029	20100311	中心馆自科书库
C0874035	丁治云	中国皮肤病性病图鉴	R751-64/Z882.02	4010444749	20100311	中心馆自科书库
C0874035	丁治云	皮肤组织病理学入门: 皮肤科医师的必备知识	R751.02/G187-2	4010539581	20100311	中心馆自科书库
Q0241009	丁攀峰	一生	I565.44/M846/=2	3010000650	20081229	东区流通库

Q0341004	殷璐	内分泌学	R58/B715	30100002 70	2010030 8	东区流通库
Q0341004	殷璐	糖尿病学	R587.1/X888	40101595 81	2010030 8	东区流通库
Q0441008	邓晓辉	神经外科手术 精要与并发症	R651.06/Z309	30100173 87	2010030 8	东区流通库
Q0441008	邓晓辉	手术学全集. 神经外科卷	R651/D923	30100239 54	2010030 8	东区流通库
Q0541025	夏杰		R5-43/=058.02	40200152 05	2010032 3	东区流通库
Q0541025	夏杰		R322.3/=M113. 02	30200205 20	2010032 3	东区流通库
Q0541025	夏杰		R322.1/=F151. 02	30200205 70	2010032 3	东区流通库
Q0541025	夏杰		R322.5/=S315. 02	30200205 75	2010032 3	东区流通库
Q0541031	臧素华	心脏外科学	R654/W087/2	30101505 25	2010011 8	东区流通库
Q0541031	臧素华	心脏外科基础 图解	R654.2-64/Z88 0/=2	30100028 37	2010020 8	东区流通库
Q0541031	臧素华	实验室诊断速 查	R446/Z431=2	40105832 82	2010021 1	东区流通库
Q0541055	沈培培	儿科速查	R72-62/W447	40101594 81	2010031 6	东区流通库
Q0541123	方慧	临床脑血管解 剖手册	R322.81-62/Z4 30	40104690 32	2010031 5	东区流通库
Q0541163	曾庆虎	英语修辞简明 教程	H315/L260	40100598 13	2009090 5	中心馆社科书库
Q0541165	罗成汉	心电图精粹	R540.4/Z081	C049108	2008042 5	东区流通库
Q0541165	罗成汉	现代心功能学	R540.4/S422	30100003 97	2008042 5	东区流通库
Q0541165	罗成汉	实验诊断学彩 色图谱	R446-64/Z137	30100021 27	2008042 5	东区流通库
Q0541174	张继要	毕飞宇文集	I247/B758	40102302 57	2007020 5	中心馆文艺书库
Q0541179	杨政	神经外科临床 解剖学图谱	R322.8-64/Y70 1	40104061 82	2010031 5	中心馆自科书库
Q0641097	李振涛	浮世绘	I267/L602	30100008 58	2010022 0	东区流通库
Q0641097	李振涛	中国人：全译 本	C955.2/L526/2 =2	30100112 81	2010022 0	东区流通库