



项目编号: 2018039

CALIS 全国农学文献信息中心研究项目 结题报告

项目名称: 基于数据挖掘的高校图书馆阅读推广工作研究

项目关键词: 数据挖掘、借阅统计、高校图书馆、阅读推广

项目单位(盖章): 沈阳农业大学图书馆

通信地址:(详细地
址含邮编) 辽宁省沈阳市沈河区东陵路 120 号
沈阳农业大学图书馆 邮编 110866

项目主持人: 李志勇


联系电话: 13940428449

电子邮件: 121632296@qq.com

提交日期: 2019.4.20

项目结题验收单

1 专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	基于数据挖掘的高校图书馆阅读推广工作研究				
主持人	李志勇	 职务/职称	馆员		
所在单位	（加盖公章）沈阳农业大学图书馆				
专家意见	<p>该课题选题明确，具有较强的实践指导意义，通过调查、分析并利用沈阳农业大学（以下简称我校）图书馆汇文自动化系统检索、搜集、整理本馆读者以及借阅信息，应用数据挖掘技术，对提升高校图书馆阅读推广服务工作的方法进行了全面深入的研究。从选题、立项、实施、研究、到结题，该项目主持人积极主动查找资料、走访调研、咨询专家，进行了多维度、多渠道的探索性研究。该课题的研究成果已完成预期目标要求，并取得一定成绩，对高校图书馆阅读推广工作提供决策具有重要的现实意义。</p> <p style="text-align: center;">专家组一致同意，通过对该课题的鉴定。</p> <p style="text-align: center;">（如需要可增加页数）</p>				
专家签字	刘国英	赵洁	王光		
职务/职称	副研	副研	副研		

题目：基于数据挖掘的高校图书馆阅读推广工作研究

关键词：数据挖掘、借阅统计、高校图书馆、阅读推广

1 研究背景、目的及意义

1.1 研究背景

2012年，党的十八大工作报告提出“开展全民阅读活动”，2015年政府工作报告中再次提出让人民群众享有更多文化发展成果，提倡全民阅读，建设书香社会。面对全民阅读推广活动的深入开展，形式多样的高校图书馆阅读推广活动也蓬勃开展起来。就在高校图书馆阅读推广工作开展的如火如荼之际，一个令人担忧的不争事实却摆在了面前，高校图书馆读者到馆率与书籍借阅率呈持续下滑态势。据《中国青年报》对全国部分高校的调查显示，17%的学生一年中未购过一本课外读物，8%的学生从未走进过图书馆。高校图书馆虽然拥有大量的图书资源，大学阶段历来又被普遍认为是博览群书的黄金时期，但是当今时代我们所面临的现实状况是高校图书馆阅读人数减少，当代大学生摆脱了应付高考的沉重负担，开始热衷于上网、旅游、看电视等消遣方式，少有人热衷于读书^[1]。当代大学生阅读状况令人担忧，高校图书馆读者到馆率、书籍借阅率持续下滑已经成为一个不争的事实。不仅大学生的阅读频率和人数比例在下降，而且大学生的阅读态度和阅读习惯也存在诸多问题——比如阅读心理浮躁，表现出较强的功利主义色彩，仅仅为自己过级、就业、学技术等即时需要而阅读，或者将阅读作为一种消遣娱乐的方式，习惯于快餐式随意的阅读，缺乏阅读经典名著的能力和意识，并没有将阅读看成是一种自我的提升，还有不少大学生在阅时没有明确的计划性和系统性。为什么会如此现象，已有研究表明，一方面现代信息、网络技术的高速发展为读者提供了多样便捷的电子阅读途径，某种程度冲击了传统阅读行为，但更为重要的原因是高校图书馆阅读推广工作自身存在着不足与问题。首先，高校图书馆阅读推广普遍流于形式，浮于表面。现代图书馆应用自动化管理系统，后台存储了海量的有关读者入馆与借阅的信息，这些信息正是读者阅读行为的体现，包括读者的阅读兴趣、习惯与需求。而研究表明高校图书馆在推行阅读推广活动时，多是注重形式以及推广技术、平台的多样化、现代化却很少挖掘揭示阅读主体——读者的个性化需求与习惯的数据来指导工作。其次，高校图书馆阅读推广缺乏长效性与持续性。虽然每年高校图书馆都会精心筹备几次大型的阅读推广活动，但这些活动由于缺乏持续性，延续周期过长，在短时间内存在一定感召力却很难起到持续激励的效果。

1.2 目的及意义

1.2.1 研究目的

面对当代师生阅读率下降及阅读倾向发生偏差的严峻形势,高校图书馆发挥自身的资源和人才优势,对广大师生进行切实可行的阅读指导,是一个越来越值得研究的课题。建立长效的阅读推广机制,能够建立和稳定高校的阅读推广队伍,并能彻底扭转学生阅读“退化”的不利局面^[2]。高校图书馆要通过多种方式和多种渠道主动对师生进行阅读指导,让大学生具备为社会服务的知识、技能和文化修养,让教师更好地为教学科研服务。数据挖掘技术就是通过对数据库中的信息进行分析来找出其中隐藏的能够反映不同项目之间的关联性以及类别特征等知识,而在高校图书馆数据库中的各种读者信息、图书信息以及读者关于图书资源的借阅记录与浏览日志等,都隐含着大量能够反映读者兴趣特征、图书关联性等潜在的、有价值的知识,应用数据挖掘技术把这些潜在的知识信息都找出来,对于指导图书馆开展个性化信息推荐阅读推广服务具有很重要的参考价值。对于图书馆海量借阅信息的利用,不是简单的对数据的统计分析,而是要应用数据挖掘技术,从馆藏文献借阅信息中找出隐藏在图书借阅数据中的知识,特别是读者的阅读兴趣和图书的利用情况。本项目将应用数据挖掘技术对高校图书馆流通系统借阅数据以及读者阅读行为信息进行数据挖掘,找出隐藏在庞大数据中的有价值信息。将这些有价值的信息,试着运用到高校图书馆阅读推广服务工作中,为读者提供个性化阅读推广服务,力争在读者阅读推广服务工作中起到很好的导航作用,为高校图书馆读者个性化阅读推广服务提供依据和保障,并为高校阅读推广工作建立长效的阅读推广机制,向纵深发展,具有可持续效应。

1.2.2 研究意义

1. 适应高校图书馆发展需求,改变高校图书馆原有的被动服务模式。在信息时代,高校图书馆作为知识信息服务中心,存储了大量的信息资源,通过数据挖掘分析,从海量信息中挖掘出读者服务信息,为读者有针对性开展信息服务,这样可以使图书馆的资料得到充分利用,同时也为现代高校图书馆开展阅读推广服务提供可行性依据^[3]。

2. 优化馆藏。通过对高校图书馆读者借阅结果的分析与高校专业学科设置情况对比分析,可以为图书馆文献的采购、上架、排架及剔除的合理化提供依据。

3. 可以提高读者阅读推广服务工作的分析决策能力。

4. 提升高校图书馆的地位。在大多数读者眼里,高校图书馆只是一个文献资源存放地,都是坐等读者来浏览图书。如现在高校图书馆已经逐步改变被动服务模式,开展个性阅读推广服务,可以使读者不只是在有需求时才会记起,图书馆可以定期或随时向读者推荐他们感兴趣的书籍或其他信息,这样使得图书馆将自身推销出去,让读者感受到图书馆随时都存在,这样就能在不知不觉中提高了高校图书馆的地位^[4]。

2 研究内容及方法

2.1 研究内容

通过调查总结目前高校图书馆阅读推广工作普遍存在的不足与问题，利用沈阳农业大学图书馆汇文自动化系统检索、搜集、整理本馆读者以及借阅信息，利用数据挖掘技术，挖掘读者借阅行为习惯、需求和兴趣，并将由数据挖掘分析形成的知识应用到高校图书馆阅读推广服务中，为本校图书馆开展阅读推广服务提供决策依据，满足不同读者群体的个性化服务需求，并为不同类型的读者群体提供个性化的阅读服务。

1. 阐明阅读推广工作的重要性以及深远意义，总结目前高校图书馆阅读推广工作普遍存在的不足与问题。

2. 分析了解当前我国高校图书馆应用数据挖掘技术开展服务状况，再深入剖析数据挖掘技术对图书馆馆藏资源利用情况以及开展个性化读者服务过程中存在的问题，并对国内外关于数据挖掘在图书馆中应用研究现状进行分析。

3. 对数据挖掘技术的相关理论进行了解掌握，同时对数据挖掘的几种方法、步骤以及规则、算法进行研究。

4. 利用数据挖掘的聚类法设计有关高校图书馆读者借阅与阅读推广活动的调查问卷。

5. 整理调查问卷信息与数据，应用数据挖掘方法进行分析研究。

6. 通过定性与定量相结合的方法，提出高校图书馆个性化阅读推广策略，建立长效阅读及激励机制。

2.2 研究方法

2.2.1 数据挖掘方法。对我馆读者与借阅信息利用数据挖掘，将借阅信息转化为可为阅读推广工作提供决策依据的“知识”。通过购买教材、网络检索以及向学科专家学习掌握数据挖掘的各种方法，熟练数据挖掘应用条件与算法。

2.2.2 定性与定量相结合的研究方法，提出高校图书馆阅读推广服务的策略。

2.2.3 文献综述法。从以往的各种文献资料分析、归纳总结出高校图书馆阅读推广工作不足与问题。

2.3 研究具体内容

2.3.1 数据挖掘

(1) 数据挖掘基本概念数据挖掘(Data Mining, DM)最早出现于1989年，就是从大量的、不完全的、有噪声的、模糊的、随机的数据库中提取隐性信息并且加工成可理解的信息和知识

的过程。数据挖掘常用的方法有关联规则、决策树及统计分析等。数据挖掘的任务主要是能分类、回归、时间序列分析、预测、聚类、汇总、关联规则、序列发现等^[5]。

(2)数据挖掘步骤 数据挖掘在实际应用中，数据挖掘步骤一般主要由确定数据挖掘对象、数据准备、数据预处理、数据挖掘、结果解释与评估、知识应用等六个部分组成。如下图 2.1

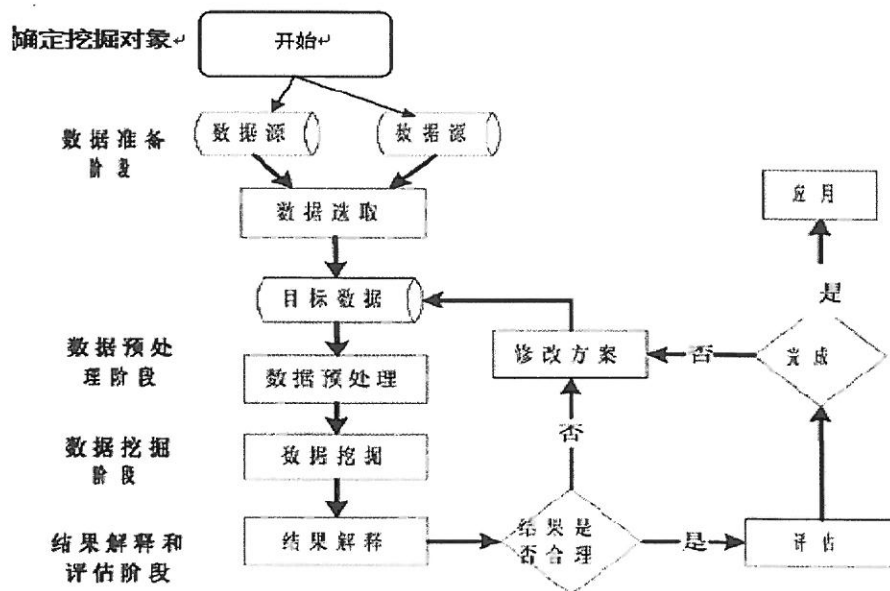


图 2.1 数据挖掘的步骤

(3)数据挖掘方法 目前最流行的几种数据挖掘方法有概念描述、关联分析、分类和预测、聚类分析、孤立点分析以及时间序列分析等。本课题主要采用聚类分析法与关联规则法。聚类分析是数据挖掘研究领域中的一种重要算法，它通过将物理的或抽象的对象集合分组，找出属性相近的对象。通过聚类分析得出的分组被称为簇，它是多个数据对象的集合，这些对象与同一个簇中的对象具有很大的相似性，而与不同簇中的对象具有很大的相异性。聚类分析算法可以获得数据的分布情况，通过观察和分析不同簇的相似性和相异性，可以得出数据的隐藏特点和规律。关联规则也是数据挖掘研究领域非常活跃而且成果丰硕的研究方向，它反映了一个对象与其他对象之间的相互依存性和关联性，帮助人们发现数据库中截然不同的对象之间暗藏的联系。当多个对象之间存在着一定的关联关系时，我们可以通过其中一个对象来推测其他的对象^[6]。

本文选取 K-Means 算法做为聚类分析算法，它是一种很典型的基于划分的聚类算法，能够很好地解决数值型属性的数据对象的聚类问题，对于处理中小型结构的数据库，该算法具有较高的可伸缩性和很高的算法效率，虽然其抗干扰性偏低。对于噪声数据的处理能力不高，但由于图书馆中数据的完整性问题已经在数据预处理过程中得到了解决，因此选择 K-Means 算法可

以有效地进行数据挖掘。K-Means 算法，也叫 K-均值算法，采用数据之间的距离作为相似性的评价指标，即当两个对象的差距越小，认定其相似性越大，当两个对象的差距越大，认定其相似性越小。其具体的聚类分析过程如图 2.2 所示。

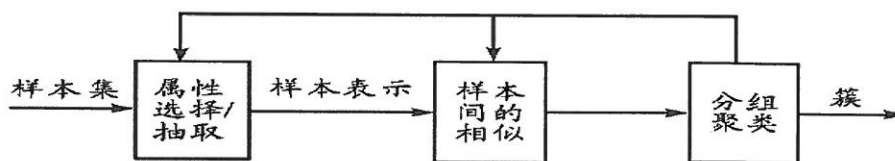


图 2.2 K-Means 算法的聚类分析过程

2.3.2 图书馆汇文系统流通统计数据聚类分析

图书流通情况指标包括借次率、阅读率、利用率和周转率等借阅流通数据，这些比率说明图书馆和读者的复杂关系，映射了阅读推广对象的阅读状况、需求、期望与趋向。

(1) 读者的聚类分析

目前我校图书馆面对学生读者采用的是分类借书规则，本科生每名学生每人最多借四本书，借期是 60 天；研究生每名学生每人最多借九本书，借期是 90 天。而在图书馆的日常工作中，我们往往发现有些同学来图书馆的频率比较高，几乎每周都来借书和还书，而有些同学则一年也难得来图书馆一趟。因此，通过对学生读者借阅次数的聚类分析，我们可以更好地了解读者的阅读行为，并将读者细分成不同的群体，制定人性化的借阅规则，从而更好地为读者提供个性化的服务。

为保证数据挖掘的有效性，本文选取了 2018 年一年的借阅记录，由于本次数据挖掘采用 Spss Statistics 软件，因此需要将原始数据转换成 Excel 格式。首先通过 SQL SERVER 2000 的查询分析器将相关的数据转换出来，最后得到的有效流通数据共有 7698 条，如图 2.3 所示。

这里我们使用 Spss Statistics 软件进行数据挖掘，首先将得到的学生读者借阅次数表导入到 Spss 软件，然后选择 K-均值聚类分析，如图 2.4 所示。我们知道，K-均值聚类分析的核心问题是确定 K 取值的大小，根据经验，我将 K 值设定为 3，即将读者分为“活跃读者”、“一般读者”、“不活跃读者”。

	B	C	D	E	F	G	H
1							
2							
3	证件号	读者类型	读者单位	年级组	借阅册次		
4	2017108161	本科生	2017级林学院林学类	2017	15		
5	2018151013	本科生	2018级信息与电气工程学院计算机科学与技	2018	17		
6	2018166057	本科生	2018级类别招生建筑类	2018	6		
7	2016101284	本科生	2016级类别招生植物生产类	2016	9		
8	190540007	本科生	2015级-水利学院	2015	79		
9	2018166031	本科生	2018级-拔类招生	2018	2		
10	2018121084	本科生	2018级-类别招生	2018	9		
11	20171187006	本科生	2017级-食品学院	2017	6		
12	2017111016	本科生	2017级农学院大气科学类	2017	20		
13	198040006	本科生	2015级-林学院	2015	14		
14	2016134045	本科生	2016级-类别招生	2016	18		
15	2016101493	本科生	2016级-类别招生	2016	12		
16	193340006	本科生	2015级-食品学院	2015	9		
17	190740008	本科生	0116010500	2015	36		
18	2017109062	本科生	2017级类别招生动物生产类	2017	2		
19	194040018	本科生	2015级-园艺学院	2015	1		
20	2018136010	本科生	2018级按类招生工商管理类	2018	5		
21	2017169012	本科生	2017级-园艺学院	2017	56		
22	2017198014	本科生	2017级畜牧兽医学院水产养殖学	2017	12		
23	2017101256	本科生	2017级类别招生植物生产类	2017	5		
24	2016136125	本科生	2016级-类别招生	2016	15		
25	2017161065	本科生	2017级-土地与环境学院	2017	11		
26	196140021	本科生	2015级-土地与环境学院	2015	45		
27	2016160104	本科生	2016级-类别招生	2016	5		
28	2017130043	本科生	2017级-经济管理学院	2017	12		
29	2016168079	本科生	2016级类别招生电子信息类	2016	8		
30	190947002	本科生	2015级-畜牧兽医学院	2015	4		
31	2017101334	本科生	2017级-类别招生	2017	37		
32	2017130019	本科生	2017级-经济管理学院	2017	4		
33	2017149066	本科生	2017级-食品学院	2017	4		
34	2016101313	本科生	2016级-类别招生	2016	3		
35	2016136111	本科生	2016级-类别招生	2016	3		
36	2017101211	本科生	2017级类别招生植物生产类	2017	2		
37	191140022	本科生	2015级-农学院	2015	1		
38	2016166042	本科生	2016级-类别招生	2016	3		
39	190940007	本科生	2015级-畜牧兽医学院	2015	2		
40	2016125020	本科生	2016级-类别招生	2016	1		

图 2.3 学生读者借阅次数表

K-均值聚类分析

变量(V): 借阅次数

聚类数(U): 3

方法: 迭代与分类 仅分类

聚类中心: 读取初始聚类中心(E)
 打开数据源(D) 文件(F)
 外部数据文件(X) 文件(F)
 导入最终聚类中心(W)
 初始聚类(D)
 数据文件(S) 文件(F)

确定 帮助(P) 重置(R) 取消 帮助

图 2.4 K-均值聚类分析的属性设置

通过软件分析得出的结果如图 2.5 所示。

最终聚类中心

	聚类		
		2	3
借阅 次数	4	3	1
			4

每个聚类中的案例数

聚类	1	123.000
	2	1285.000
	3	6291.000
有效	7698.000	
缺失	.000	

图 2.5 读者聚类结果

在上图中，我们可以看出聚类-1 借阅次数最多，平均借书 34 次，可以代表“活跃读者”这一组，但人数只有 123 人，仅占总人数的 1.6%；聚类-2 平均借书 13 次，可以代表“一般读者”这一组，人数有 1285 人，占总人数的 16.7%；聚类-3 借阅次数最少，平均借书 4 次，可以代表“不活跃读者”这一组，人数有 6291 人，占总人数的 81.7%。由此可以看出，有很大一部分读者对图书馆的利用率非常低，更何况目前全校本科生人数在 20000 多，而有借书记录的同学仅有 7698 人，说明一大半同学一年内都没有到图书馆来过。因此，图书馆非常有必要积极主动地开展相应的宣传工作，吸引广大学生来馆看书，提高图书馆的文献利用率。最后，我们可以将聚类的分组结果返回到原始的学生读者借阅次数表中，得到每位同学具体的分组情况，如图 2.6 所示。

证件号	读者类型	读者单位	年级组	借阅册次	聚类分组
2017108161	本科生	2017级林学院林学类	2017	15	2
2018151013	本科生	2018级信息与电气工程学院计算机科学与技术	2018	17	2
2018166057	本科生	2016级类别招生建筑类	2016	6	3
2016101284	本科生	2016级类别招生植物生产类	2016	9	3
190540007	本科生	2015级-水利学院	2015	79	1
2018166031	本科生	2016级-按类招生	2016	2	3
2016121084	本科生	2016级-类别招生	2016	9	3
2017187006	本科生	2017级-食品学院	2017	6	3
2017111016	本科生	2017级农学院大气科学类	2017	20	2
198040006	本科生	2015级-林学院	2015	14	2
2018134045	本科生	2016级-类别招生	2016	18	2
2018101493	本科生	2016级-类别招生	2016	12	2
193340006	本科生	2015级-食品学院	2015	9	3
190740008	本科生	0116010500	2015	36	1
2017109062	本科生	2017级类别招生动物生产类	2017	2	3
194040016	本科生	2015级-园艺学院	2015	1	3
2018136010	本科生	2018级按类招生工商管理类	2018	5	3
2017169012	本科生	2017级-园艺学院	2017	58	1
2017199014	本科生	2017级畜牧兽医学院水产养殖学	2017	12	2
2017101256	本科生	2017级类别招生植物生产类	2017	5	3
2018136125	本科生	2016级-类别招生	2016	15	3
2017161065	本科生	2017级-土地与环境学院	2017	11	2
196140021	本科生	2015级-土地与环境学院	2015	45	1
2016160104	本科生	2016级-类别招生	2016	5	3
2017130043	本科生	2017级-经济管理学院	2017	12	2
2016188079	本科生	2016级类别招生电子信息类	2016	8	3
190947002	本科生	2015级-畜牧兽医学院	2015	4	3
2017101334	本科生	2017级-类别招生	2017	37	1
2017130019	本科生	2017级-经济管理学院	2017	4	3
2017149086	本科生	2017级-食品学院	2017	4	3
2018101313	本科生	2016级-类别招生	2016	3	3
2018136111	本科生	2016级-类别招生	2016	3	3
2017101211	本科生	2017级类别招生植物生产类	2017	2	3
191140022	本科生	2015级-农学院	2015	1	3
2016166042	本科生	2016级-类别招生	2016	3	3
190940007	本科生	2015级-畜牧兽医学院	2015	2	3
2018125020	本科生	2016级-类别招生	2016	1	3

图 2.6 读者聚类分组表

(2) 图书借阅的聚类分析

图书馆最基础和最核心的职能是提供文献资料的借阅服务，因此我们可以对图书的借阅情况进行聚类分析，来获取有益的信息，以此指导我们不断改进服务方式，提高服务水平。

图书书名的聚类分析 我校图书馆的图书复本量目前一般为 4 本，但在实际的流通过程中，有些热门的图书如畅销小说以及部分专业书籍供不应求，往往需要多次预定才能够借到，而部分冷门图书则无人问津，长时间地空置在书架上，造成了资源的浪费。因此，我们可以根据图书的书名进行聚类分析，找出其中的热门图书和冷门图书。第一步：先从流通数据表中提取正题名、索取号、责任者、出版者、借出次数等相关字段，在使用数据筛选法选出不重复的数据，组成新的图书借阅次数表，如图 2.7 所示。

A	B	C	D	E
1	索取号	正题名	责任者	出版者
2	H36/5:1	自学日语	日本阿曾克出	外语教学与研究出版社
3	H319.9/6	商务英语实用会话	浩瀚英语研究	中国石化出版社
4	H019/42	中外妙答精粹	江左浩主编	石油工业出版社
5	I312.645	狼的诱惑	[韩]可爱尚著	世界知识出版社
6	I312.645	失去的你	金润姬著	海峡文艺出版社
7	I14/79	侦探小说	汤学智,吴岳	中国和平出版社
8	K25.34/1	岁月沧桑	赵清宇著	百花文艺出版社
9	H313/84	三选记大学英语四级词汇	任德宇主编	东方出版社
10	H319.9/55	英文妙语妙用	汪福祥著	石油工业出版社
11	H319.6/9	英语专业四级词汇与语法2000题	余高峰主编	华东理工大学出版社
12	H313/85	四级英语单词红秘笈	邱勇著	气象出版社
13	H313/86	四级英语单词红秘笈	袁昊主编	机械工业出版社
14	H319.4/36	四级词汇短语题全面突破	袁昊主编,刘峰	机械工业出版社
15	I247.5/1	狼图腾	姜戎著	长江文艺出版社
16	O13/50	高等数学辅导	邹本腾等主编	机械工业出版社
17	H313/91	开心记单词	弘杨,李克军	机械工业出版社
18	B222.4/1	易经	宋浩编著,王	北京出版社
19	H313/92	英语名言佳句	朱先德,朱胜	中国科学技术出版社
20	H310.4/1	星火·贯通英语·15篇文章贯通4级词汇	马德高,王西	新华出版社
21	H315/49	手把手教你四六级作文	包凡一编著	世界图书出版
22	O3/44	高等数学典型题精解	陈兰祥主编	学苑出版社
23	H313/89	公共英语等级考试(4-5)级词汇逆序突破	文武主编	水利水电出版社
24	H319.4/55	大学英语四级分类阅读200篇	章小捷,顾乾	重庆出版社
25	TB52.1/1	轻松数码相片制作	王永辉编著	人民邮电出版社
26	H033/2	一句话改变一生	(美)约翰·巴	中国发展出版社
27	I14/72	哲理小说	汤学智,吴岳	中国和平出版社
28	G841/5	篮球	(日)手塚政则	湖南科学技术出版社

图 2.7 图书借阅次数表

第二步：将图书馆借阅次数表导入到 SPSS 软件中，选择 K-Means 聚类算法进行分析，这里仍然选择聚类数 k=3，分别代表“热门图书”、“一般图书”、“冷门图书”，最后得到的结果如图 2.8 所示。

最终聚类中心

	聚类		
	1	2	3
借阅	2	8	1
次数	1	3	3

每个聚类中的案例数

聚类	1	2	3
有效	61467.000	7410.000	1693.000
缺失	.000	.000	.000

图 2.8 图书书名的聚类结果

由上图可以看出，聚类-3 代表热门图书，平均借出次数为 13 次，这样的图书共有 1693 本，占总借出图书的 2.4%。聚类-2 代表一般图书，平均借出次数为 8 次，共有图书 7410 本，占总数的 10.5%，聚类-1 代表冷门图书，平均借出次数为 2 次，共有图书 61467 本，占总数的 87.1%。第三步：将聚类分组结果返回到开始的图书借阅次数表中，得到每本图书的分组情况，如图 2.9 所示。

序号	书号	书名	译者	出版社	借出次数	关联分值
1	H0645.1	自学日语	日本图书出版社, 于日平编译	外语教学与研究出版社	1	1
2	H319.0/6	国际英语交际会话	潘琦英等编著	中国石化出版社	1	1
3	H019/42	中外名著译林	江有忠主编	石油工业出版社	2	1
4	Q321.645/1	爱的秘密	傅利明等译, 曹金译	世界图书出版	2	1
5	Q32.645/6	失去的你	金莉娅著	海味文艺出版社	2	1
6	H47/9	侦探小说	陈学智, 吴岳主编	中国和平出版社	1	1
7	K825.34/1	岁月沧桑	赵清宇著	百花文艺出版社	1	1
8	H313/64	三三语记大学英语四级词汇	盛维宇主编	东方出版社	3	1
9	H019.8/65	英文抄写材料	汪福祥著	石油工业出版社	1	1
10	H319.6/9	汉语专业四级: 听力与语法200题	俞高峡主编, 华海, 刘洪, 李川编著	华东理工大学出版社	6	2
11	H313/66	四级英语单词汇编	顾勇著	气象出版社	3	1
12	H313/68	四级词汇与语法全国竞赛	顾勇主编	机械工业出版社	5	2
13	H319.4/35	四级阅读8120例	曹真主编, 刘洪等编	机械工业出版社	1	1
14	Q47.5/1753	爱斯基	董淑娟	长江文艺出版社	8	2
15	Q13/61	高等数学辅导	李书敏, 王玉清	机械工业出版社	3	1
16	H313/91	开心记单词	范磊, 李永军主编	机械工业出版社	3	1
17	B222.4/1	易经	荣清编著, 王清松, 李可讲读	北京出版社	1	1
18	H313/92-3	英语听力入门	李光, 李且, 屈主编	中国科学技术大学出版社	1	1
19	H319.4/1	星火: 英语听力15篇文章听力468例	马德高, 王森, 曹主编	群学出版社	3	1
20	H315/49	手拉手英语四六级作文	倪凡一编著	世界图书出版公司北京公司	1	1
21	Q13/44	高等数学典型题精解	陈兰兰主编	学苑出版社	3	1
22	H313/69	公共英语等级考试(4-6级)词汇	文武主编	水利电力出版社	1	1

图 2.9 图书聚类分组表

(3) 图书借阅的关联规则

Apriori 算法是关联规则分析中最核心的算法, 它的原理是根据前一次找到的频繁项集来生成新一次的频繁项集^[6]。采用 Apriori 算法进行关联规则分析, 必须保证数据格式为交易格式, 即对于任何一项借阅行为采用一个单独的记录, 如图 2.10 所示。

客户	采购
1	jam
2	milk
3	jam
3	bread
4	jam
4	bread
4	milk

图 2.10 交易格式

因此, 可以从数据流通表中提取出读者证号和索取号两个字段, 将其转换成交易格式的图书借阅表, 如图 2.11 所示。

	A	B
1	读者证号	索取号
2	313010800254	H36/5:1
3	311010800546	H319.9/6
4	311010810915	H019/42
5	311040100103	I321.645/1
6	313010801141	I321.645/1
7	313010810516	I312.645/8
8	313010810518	I14/79
9	313030100231	K825.34/1
10	313010801431	H313/84
11	311010801654	H319.9/55
12	311030900123	H319.6/9
13	313010810507	H313/85
14	311070501460	H313/86
15	312070600136	H319.4/36
16	313030810113	I247.5/1753
17	313030100216	O13/50
18	313070600101	H313/91
19	312030800110	H313/92-3
20	313040100115	H310.4/1
21	313010801057	H315/49
22	313021400119	O13/44
23	312010800224	H313/89
24	312030900131	H319.4/52
25	312030900130	H319.4/52
26	312051000112	TB852.1/1-3

图 2.11 交易格式的图书借阅表

图书索取号的关联规则挖掘 将转换好的图书借阅表导入到 SPSS 软件中，选择 Apriori 算法进行关联规则分析，如图 2.12 所示。在最低条件支持度和最小规则置信度的文本框中可以选择支持度和置信度的阈值，为了保证挖掘出有价值的关联规则，这里将支持度定为 5%，置信度定为 20%

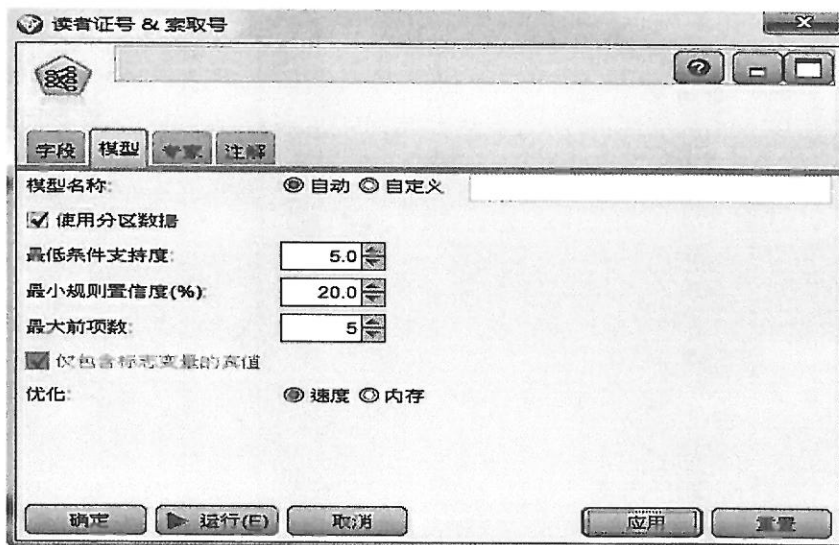


图 2.12 Apriori 属性设置

最后得到相关的挖掘结果 86 条，如图 2.13 所示。

O13/336	O171/7	24.324	44.444
H315/22	TN710/51	13.514	40.0
TP391.41/131	H315/211		
	TN710/51	13.514	40.0
	H315/211		
H315/210	TH126-44/21	21.622	37.6
H315/22	H315/211	24.324	33.333
H315/210	O171/7	24.324	33.333
TP391.41/131	TP391.41/192	16.216	33.333
O13/336	TP391.41/192	16.216	33.333
TN710/50	H315/211	24.324	33.333
TP391.72/48	H319.4/430	10.811	25.0
H315/22	TH126-44/21	21.622	25.0
TP391.72/48	H319.4/430	10.811	25.0
	H315/211		
H315/22	TP391.41/146	10.811	25.0
	TN710/51		
TP391.41/131	TP391.41/146	10.811	25.0
	TN710/51		
TN710/50	TP391.41/146	10.811	25.0
	TN710/51		

图 2.13 关联规则挖掘结果

由上图可以看出，有 44.44% 的同学在借阅了 O13/336（《高等数学学、练、考》）后，还借阅了 O171/7（《简明微积分研究》）；有 40% 的同学在借阅了 H315/22（《高级英语写作技巧》）或 TP391.41/131（《计算机平面图形设计》）后，还借阅了 TN710/51（《模拟电子技术基础教程》）和 H315/211（《实用商务英语写作指导》）；有 37.5% 的同学在借阅了 H315/210（《英语写作要点与策略》）后，还借阅了 TH126-44/21（《Auto CAD 机械绘图案例教程》）；有 33.3% 的同学在借阅了 H315/22（《高级英语写作技巧》）后，还借阅了 H315/211（《实用商务英语写作指导》）。

图书分类号的关联规则挖掘 在图书馆的日常工作中，往往可以发现：虽然读者所借阅的图书并不一定完全相同，但所借图书的分类号之间却隐含着一定的规律性，如借阅“A 马列主义”类的书籍可能还会借阅“B 哲学”类的书籍。通过图书分类号的关联规则挖掘，可以找出这些隐藏的规律，便于为读者的借阅提供参考，也可以为图书馆的馆藏布局和书刊排架提供有益的信息。第一步：对图书借阅表中的索取号字段进行数据转换，提取出其中的分类号，形成交易格式的分类号借阅表，如图 2.14 所示。

A	B
读者证号	分类号
313010800254	分类号H
311010800546	分类号H
311010810915	分类号H
311040100103	分类号I
313010801141	分类号I
313010810516	分类号I
313010810518	分类号I
313030100231	分类号K
313010801431	分类号H
311010801654	分类号H
311030900123	分类号H
313010810507	分类号H
311070501460	分类号H
312070600136	分类号H
313030810113	分类号I
313030100216	分类号O
313070600101	分类号H
312030800110	分类号H
313040100115	分类号H
313010801057	分类号H
313021400119	分类号O
312010800224	分类号H
312030900131	分类号H
312030900130	分类号H
312051000112	分类号T
311010810626	分类号H

图 2.14 分类号借阅表

第二步：将数据导入到 SPSS 软件中，选择 Apriori 算法进行关联规则分析，设置支持度为 10%，置信度定为 60%，最后得到相关的挖掘结果 33 条，如图 2.15 所示。

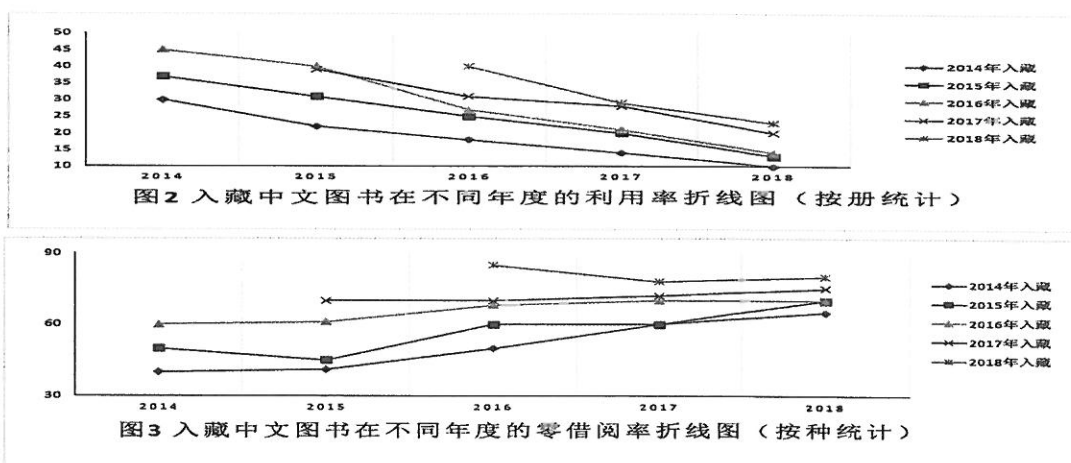
分类号D	分类号I	72.857	86.275
分类号I	分类号G	48.571	82.353
分类号D	分类号J	44.286	83.871
分类号D	分类号I		
分类号D	分类号G	40.0	82.143
分类号I	分类号I		
分类号D	分类号G	38.571	85.185
分类号D	分类号D		
分类号D	分类号H	35.714	84.0
分类号J	分类号I		
分类号D	分类号F	21.429	86.667
分类号D	分类号H		
分类号I	分类号G	18.571	84.615
分类号I	分类号I		
分类号I	分类号G	18.571	84.615
分类号I	分类号H		
分类号I	分类号O		
分类号I	分类号G		
分类号I	分类号J	17.143	83.333

图 3.23 分类号挖掘结果

由上图可以看出，有 86.275%借阅了“D 政治、法律”类图书的同学，还借阅了“I 文学”类图书；有 82.353%借阅了“I 文学”类图书的同学，还借阅了“G 文化、科学、教育、体育”类图书；有 86.667%借阅了“J 艺术”类图书的同学，还借阅了“F 经济”类图书。

(4) 图书利用率与零借阅率数据纵向维度分析

图书利用率和零借阅率是揭示馆藏利用的重要指标。通过对我校图书馆 2014—2018 年入藏中文图书的利用率（册）和零借阅率（种）的统计分析发现：（1）当年入藏图书的利用率在第 1~2 年呈上升趋势，第 2 年达 40%左右的最高点；从第 3 年开始逐年下降，第 5 年下降到 20% 以下。同时，在年借阅量中，当年入藏图书借阅总量占该年的总借阅量的 18.29%，近 3 年入藏图书占 52.42%。由此可见，图书利用的高峰期在图书入藏 1~3 年内。（2）零借阅率最低值在图书入藏第 2 年，这和利用率最高值在第 2 年相对应。第 3 年开始 60%的图书处于零借阅状态，然后逐年以 5%~10%的比例递增。第 5 年开始 80%图书是零借阅。（3）入藏图书利用率下降、零借阅率增长的趋势逐年明显。2015—2018 年，入藏第 1 年图书的利用率分别为 38.1%、31.0%、27.0%、24.9%，4 年下降了 13.1%。零借阅率分别为 53.2%、66.4%、71.4%、73.3%，4 年增加了 20.11%（见图 2、图 3）。



（5）图书利用率、零借阅率数据横向维度分析

笔者随机抽取了 1 个批次新书 75 种 188 册，其中相同复本 57 种 170 册，分别陈放在 2 个新书展示区，一个新书展示区是电梯前展架平板式展示，另一个新书区是检索机边新书架竖式排列。另外 18 种（册）新书直接入开架书库上架。借阅时间 1 个月。数据统计显示：新书平板展览式宣传的借阅率为 24.24%，新书竖式排列的借阅率为 12.12%，而同期直接入库的新书借阅率为零。由此可见，新书宣传对提高新书借阅率有较大作用，而且不同的宣传方式、地点对新书借阅率的影响也较大。

图书荐购是读者直接参与图书馆藏书建设、增加图书采访针对性、提高图书利用率的重要举措。我馆于 2012 年始推出了“你选书、我买单”师生荐购图书活动。任选 1 个批次荐购入藏图书，将同期入藏新书和同期馆藏图书比较，借阅时间 1 年，按种统计结果如下：荐

购图书的利用率为 75%，零借阅率为 25%；同期入库新书的利用率为 36%，零借阅率为 64%，同期馆藏图书的利用率为 20%，零借阅率为 80%。即荐购图书的利用率比同期入库新书高 39%，同期馆藏图书高 50%，零借阅率则分别低 39%和 55%。由此可见，读者直接参与藏书建设对提高图书的利用率和降低零借阅率具有非常重要的作用（见图 2.18）。

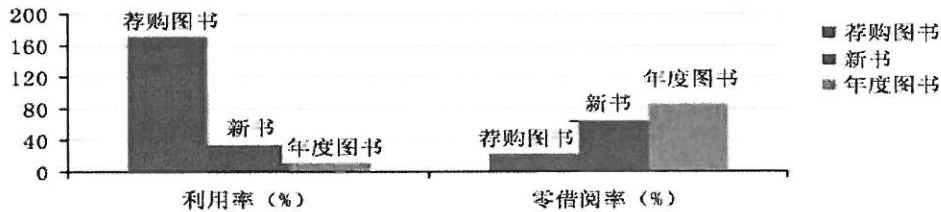


图 2.18 荐购图书、新书、年度图书同期利用率、零借阅率

2.3.2 读者阅读情况问卷调查数据挖掘

(1) 调查问卷的设计 采用七级等距离结构来设计调查问卷的选项，“1”代表强烈不影响，“7”代表强烈影响，中间值“4”代表基本不受到影响，“>4”代表用户受到影响，“<4”代表用户没有受到影响。问卷包括三部分内容：用户基本情况调查、影响用户借阅行为的因素调查和用户借阅后的评价调查，其中第二部分影响因素的自变量包括 4 类共 27 个。通过文献调研，结合高校图书馆学生和教师科研人员的特殊性，把影响用户借阅行为的因素归纳成四部分：外部因素、推荐因素、书本身的因素、个人偏好因素，并基于以上影响因素设计问卷的变量^[7]。影响借阅的外部因素有：是否有检索系统、检索系统中是否有被查阅的书籍、书架上是否可以找到、书籍是否被借出、借阅次数、是否全部在馆、摆放位置、馆员服务、课堂必备、书籍分类准确性；影响借阅的推荐因素有：导师推荐、专家推荐、网络评分；影响借阅的书本身因素有：题目、封皮设计、封面评论、摘要导言、目录、内容、知名度、实用性；影响借阅的个人偏好因素：作者、出版社、版本、新旧、纸张手感、中英版。

调查问卷统计结果本次调查共发放问卷 80 份，涵盖了我校 17 个不同的学院，问卷回收率为 95%，因此，这次的问卷可以综合的反映我校学生的借阅行为以及图书馆的使用情况。

(2) 调查结果分析 影响用户借阅一本图书的原始因素有 27 个，但真正具有继续研究价值的，方差累计值超过了 80%的影响因子有 15 个：作者、出版社、出版日期、权威专家推荐、有无检索系统、课堂必备、中英文版不同、目录、借阅次数、馆员服务态度影响、图书摆放区域、图书内容、图书分类的准确性、实用性、题目。媒体的成熟，以及电子阅读的便捷性、丰富性和多样性等特点，传统的纸质阅读受到严重冲击，数字化阅读方式越来越受到学生青睐；网络已成为大多数学生课余生活的主要内容，使得他们在有限的课余时间里对纸媒的阅读时间

减少；当前严峻的就业形势、激烈的就业竞争也使得大学生的读书目的越来越实际，与择业、社会生存相关的实用技能、外语、计算机等类图书成为大学生阅读的重点，功利性阅读明显；面对纷繁复杂的网络信息以及快节奏的学习生活，大学生没有精力或能力从中去粗取精，找到适合自己的读物，因此浅阅读、快餐式阅读充斥着大学生的阅读生活；大学生阅读习惯和需求的变化对图书馆的管理和服务提出了更高要求。

2.3.4 建立长效阅读推广机制的建议

(1) 专题书目推荐

各个高校图书馆可以单独设置空间，来进行某个主题的书目推荐，定期或不定期的进行图书展览，介绍主题的选题背景，及主题介绍，有哪些作者哪些名家，作品介绍等等，来供读者方便选书阅读。另外，每期新书到馆，及时推荐给大家，做好到室新书的宣传工作，设立专一的新书架及新书宣传栏，对重点推荐新书做好作者、书目、出版社索书号等信息介绍，尤其对于新书内容也要做简洁的提要，让读者对到室重点图书有个大致的了解。各室还可以设立每月好书榜，按本室借阅率排名，把排名高的图书推荐给读者，还可以把豆瓣月评分高的好书来推荐给读者。要定期更新，把真正的好书推荐给读者，而不只是流于形式。

(2) 图书采购从馆员转向用户决策

用户参与荐购图书的借阅率远高于同期入库新书和同期馆藏。长期以来，图书馆资源采购的模式是馆员选择在前、用户利用在后，采访馆员个人的价值判断成为采购依据。当图书馆进入以用户需求为中心时，采访的模式应转变为用户使用在先、图书馆选择保存在后。近年来，读者决策采购（Patron-driven Acquisitions, PAD）成为许多大学图书馆购买图书特别是电子书的重要模式。PAD先根据馆藏政策，参照传统的纲目购书或阅选计划设定购书范围及每本书的预设文档，将书商目录导入图书馆的馆藏检索系统，这时图书馆只是将这些书目展示给读者，但尚未向书商支付费用，只有当读者真正开始通过点击链接直接阅读该书，并达到图书馆预设的次数、时间、数等指标门槛后，才会自动触发图书馆向书商租用或购买的指示^[8]。因此，不管是网上的“你点我购”，还是书店现场的“你选我买”，读者决策采购形式，将成为图书馆采访的新模式。

(3) 打造多媒体阅读空间

阅读推广中，为读者打造多媒体阅读空间，也是很受欢迎的，通过音乐和电影推荐图书也受到读者欢的方式之一，尤其是女性读者更偏爱电影和音乐^[9]。音乐空间，在该空间放置一些上榜好书，听音乐读好书，营造温馨艺术氛围的音乐空间，让学生既可以专心欣赏音乐同时，又

可以阅读。影视空间，可以播放些经典名著电影，让读者从电影中读到好书。建立听书源平台，收藏一些需要推荐的好书，有正能量的图书来供读者听书。读者可以从轻松的听书中，读到好书，获得知识。

（4）与各社团合作

与高校各个社团合作，鼓励并参与组织阅读社团。例如读协，红学会，诗歌协会等社团，大力开展各种活动，来进行图书推广，例如续写，话剧，舞蹈，音乐大赛，演讲比赛等各种形式，来吸引读者，共同分享好书内容，以多种形式促进读者阅读关注图书，提高图书借阅率。也可以与校外各阅读社团合作，做好阅读推广工作。

（5）研究读者的阅读心理和阅读行为

为了能更好地满足读者对阅读的需求，我们必须长期地研究读者的阅读心理和阅读行为，这样才能更加深入地推广阅读服务。一方面要营造良好的阅读环境和氛围。作为高等学府的文化空间，图书馆的阅读环境非常重要，所以应加大高校图书馆的建设力度，创造良好的环境氛围；另一方面，图书馆还应通过推荐图书目录等方法推广阅读服务，以培养读者的阅读能力，同时增强其识别文献好坏的能力。

（6）将阅读推广渗透到学校教学环节中

图书馆根据学校的办学方针、教师的研究方向、学生的学习生活等不同层次的需求，积极参与教学活动，对学校的教学改革和发展将起到促进作用。开展阅读课程教育，能有针对性地解决一部分大学生如何选择读物、如何有效阅读的困惑，阅读课程采取的学分制也可以保障阅读推广的长效性。通识教育现已发展成为现代大学的一种教育理念，各个高校都在不同程度地实施通识教育课程体系。通识教育理念下的高校图书馆应积极加入通识教育体系，按通识教育的需要开展资源建设和阅读指导工作，以文献检索课教学为基础，加强学生的信息素养教育，使学生在学术探究过程中获得专业信息人员的引导。

（7）阅读推广要充分利用先进的传媒手段

网上、网下结合，与读者互动，与社会合作。高校图书馆要善用社会资源，包括人力与财力资源和“国家数字图书馆推广工程”等其他数字图书馆系统的平台接口，在条件成熟的情况下，与其他数字图书馆系统实现全面联通，建立一个超大型、跨行业、跨地域的全国数字图书馆网络体系，通过互联网为读者提供阅读服务，实现资源共享^[10]。

3 结论与建议

图书馆借阅统计工作数据挖掘为阅读推广工作提供了多维化参考信息,通过对近5年图书馆期刊借阅流通数据的统计、分析、了解了阅读流通趋势。为提高高校图书馆管理实效性,并考虑读者服务体系的现实需求,图书馆、信息科和教学办等各方应携手努力,制定读者服务体系建设相应措施,使阅读推广活动由“被借阅”走向“主动推荐”,加强与各高校、教育机构、出版商等多方面的合作,深入研究读者的阅读心理和阅读行为,营造融洽的阅读环境,做好阅读指导服务,以满足广大师生的阅读需求。

4 项目成果

4.1 建立主题书展广角

4.2 建立图书馆阅读宣传社团

5 参考文献

- [1] 王艳. 高校图书馆促进大学生成才的实证分析 [J]. 大学图书馆学报, 2013(1):33-37.
- [2] 乔慧君, 周筠珺. 高校图书馆借阅率与学生学习状态关系研究 [J]. 大学图书馆学报, 2015(1):55-60.
- [3] 谢春枝. Lib QUAL + (R) 图书馆服务质量调查的实证分析——以武汉大学图书馆为例 [J]. 大学图书馆学报, 2009(5) : 24—28.
- [4] 华东师范大学图书馆馆藏文献利用调查与分析 [EB/OL].
- [5] 巫红霞. 基于数据仓库和数据挖掘的高校教学管理决策支持系统 [D]. 南京: 南京航空航天大学, 2009: 5.
- [6] 杨光, 张学潮. 数据挖掘在高校图书馆用户行为分析中的应用——以山西大学图书馆为例 (J). 晋图学刊, 2011, 02:19-28.
- [7] 郑怀远. 数据挖掘在高校图书馆管理系统中的应用研究——以福建工程学院图书馆为例 (J). 情报探索, 2013, 01:97-99.
- [8] 刘华. “读者决策采购”在美国大学图书馆的实践及其对我国的启示 [J]. 大学图书馆学报, 2012(1):45-46.
- [9] 杨 婵. 图书馆阅读推广活动的反思与重构 [J]. 四川图书馆学报, 2011(2) : 58—61.
- [10] 王 波. 图书馆阅读推广亟待研究的若干问题 [J]. 图书与情报, 2011(5) : 32—35, 45.