

# 基于数据挖掘技术的高校成绩 信息管理的应用与研究

朱丽丽

(金陵科技学院 教务处,江苏 南京 211169)

**摘 要:**目前高校普遍采用学生成绩管理系统,系统更多的是对学生成绩进行简单存储满足查询和统计的需求。实行学分制以及学生选课的自由度放开,对高校的成绩管理提出了更多的需求和更高的要求。一直以来受各方面因素影响的学习成绩,它为提高教学质量研究提供可靠依据。通过建立数据仓库,利用数据挖掘工具,挖掘分析课程成绩之间的潜在关系和影响成绩的主要因素,对做出相适应的教学管理决策提供重要依据,对管理决策的合理性和科学性具有一定的推进作用,同时还能更好的为学生、教师、教学服务,最终提高高校的整体竞争力。

**关键词:**数据挖掘;学生成绩;关联分析;频繁项集

中图分类号: G642.47

文献标识码: A

文章编号: 1671-931X (2015) 05-0075-04

## 一、引言

随着我国高等教育迅猛发展,高校的竞争日益激烈,教学质量才是高等院校生存和发展的根本。衡量教学质量的重要指标之一就是学生学习成绩。现阶段,各大高校的学生成绩管理系统大多处于简单的信息管理阶段,如何发掘课程成绩之间的潜在关系,找到成绩的影响因子,是高校提高教学管理决策的重要依据,所以提高高校教学质量的重要工作就是研究学生成绩信息管理中潜在的知识。近些年研究学生成绩信息管理工作的角度也发生了根本的变化,从传统查询统计管理方式到以挖掘成绩数据的隐含信息的管理模式。高校成绩信息管理应用数据挖掘技术是教育改革创新的需要也是自身发展的必然要求。现代化的高校成绩信息管理能够更好的服务于师生、服务于教学,指导教师合理安排教学

计划,有针对性地提高学生的学习能力,最终提高高校的整体竞争力。

## 二、数据挖掘

数据挖掘是指从实际应用数据中提取有用的信息和知识的过程,这些应用数据的特点:不完整、海量、模糊。挖掘出的信息和知识都是隐含在数据内部,事先不知道的但又很重要的。通俗地讲,数据挖掘就是利用各种分析工具和相应方法,对数据库中海量数据进行研究与分析并进行归纳与整合,挖掘数据内部的隐含信息和关系,为教学管理人员在进行教育改革决策时提供参考依据<sup>[1,3]</sup>。

### (一)数据挖掘常用方法

数据挖掘主要的研究方法:关联分析方法、统计分析方法、聚类分析方法、人工神经网络分析方法、决策树方法、遗传分析方法。

收稿日期:2015-06-29

作者简介:朱丽丽(1980-),女,江苏淮安人,工学硕士,江苏南京金陵科技学院教务处讲师,研究方向:数据挖掘、教育管理、计算机。

## (二)高校成绩信息管理应用数据挖掘的方法

关联分析方法：关联分析用来发掘数据库中项集之间的关联，即查找出数据库中隐藏的联系，即多个变量取值中存在某些规律性，比如在成绩数据库中《C 语言程序设计》和《数据库系统原理》两门课程成绩存在一定的关联，这种关联在制定教学计划进行课程设置时具有高效地指导作用。

关联分析中比较经典的关联算法是 Apriori 算法。Apriori 性质为：“频繁项集的所有非空子集都必须是频繁的”。该性质属于一种特殊的分类，称作反单调，意指如果一个集合不能通过测试，则它的所有超集也都不能通过相同的测试。Apriori 算法的执行过程分为两个步骤：首先是连接步，即为了寻找  $L_k$ ，通过  $L_{k-1}$  与自己连接产生候选  $k$ -项集的集合，记为  $C_k$ ；其次是剪枝步，即由于  $C_k$  是  $L_k$  的超集，扫描数据库，确定  $C_k$  中每个候选的计数，删除不满足计数要求的项集，从而确定  $L_k$ <sup>[2,3]</sup>。

在数据挖掘过程第一步也是最重要的一步：数据离散化处理，它直接影响后继数据挖掘的精度和质量。数据挖掘技术具体采用：熵值离散法与关联规则集成的算法对连续数据进行离散化再进行关联规则挖掘，即先将连续数据转化为布尔数据，再进行布尔关联规则挖掘。

## 三、成绩信息管理中数据挖掘技术的应用

成绩数据库中哪些课程之间联系密切，课程之间的影响因素、影响程度，一直没有比较准确的量化关系。通过学生的成绩分布情况，教师和教学管理人员很难直接找出先行课程和后继课程以及它们之间的联系，限制了因材施教。因此对这些海量的成绩数据进行课程相关性关联规则挖掘，预测课程间存在的内部联系，挖掘出影响学生成绩的因子。

### (一)熵值离散法与关联规则集成算法主要编码及算法

```
1.//生成频繁项集的操作
fAprioriSupp=supp; //支持度
clock_t start, stop, tick; //时钟变量
double timeused; //挖掘时间
start=clock(); //开始计时
FindLargeItem(); //挖掘频繁 1-项集
for(int i=0; i<LargeItemCount[0]; i++)
    fprintf (fp, "% s, % .2f\t", LargeItem [0] [i],
fSuppFreItemSet[0][i]);
for(int k=1; LargeItemCount[k-1]!=0; k++) //逐级
产生频繁 k-项集
{LargeItemCount[k]=AprioriGen(k); //由频繁 k-1
项集生成频繁 k-项集
if(LargeItemCount[k]!=0)
{for(int i=0; i<LargeItemCount[k]; i++)
```

```
    fprintf (fp, "% s, % .2f\t", LargeItem [k] [i],
fSuppFreItemSet[k][i]);
    fprintf(fp, "\n\n");
} //逐级保存频繁 k-项集
}
nLargeItemCount=k-1; //保存最大频繁项集的级
数
CString strValue;
2.//生成关联规则
fAprioriConf=conf;
float fSupp=0.0;
float fSubSet=0.0;
float tmp=0.0;
int nLine=0;
CString Rule, strTemp, strTemp1, strValue;
CString str;
nCountRule=0;
for(int ii=0; ii<LargeItemCount[nLargeItemCount-1]; ii++)
{strTemp="";
strTemp=LargeItem [nLargeItemCount-1][ii]; //获
得当前频繁项
fSupp=fSuppFreItemSet[nLargeItemCount-1][ii];
FreSubSet(strTemp); //求当前频繁项的真子集
for(int jj=0; jj<nCountSubSet; jj++)
{for(int kk=0; kk<nLargeItemCount-1; kk++)
{for(int mm=0; mm<LargeItemCount[kk]; mm++)
{if(strcmp(SubSet[jj], LargeItem[kk][mm])==0)
{fSubSet=fSuppFreItemSet[kk][mm]; //求当前真子
集的事务计数
tmp=fSupp/fSubSet;
if(tmp>=fAprioriConf)
{Rule="";
strTemp1=SuppleSet(SubSet[jj], strTemp); //求当
前真子集在频繁项中补集
Rule=SubSet [jj]+"==>" +strTemp1; //生成关联规
则
```

```
Rules[nCountRule]=Rule;
fSuppRules[nCountRule]=tmp;
nCountRule++;
nLine++;
}}}}
(二)数据预处理
```

本文使用的数据源：南京某高校 2012 级、2013 级、2014 级学生成绩信息，因为原始的数据都是连续值，系统不能识别，它只能识别运行离散值，所以要将连续数据进行离散化。原始数据库中有三个数据表(2012, 2013, 2014)，用年份来代表某年的学生成绩信息。由于全部数据量较大，现只将其中一个数

表 1 学生成绩信息

学号	姓名	C 语言程 序设计	高等数 学 A1	离散结构	数字逻辑	汇编语言 程序设计	计算机组成原 理与系统结构	面向对象程 序设计	面向对象程 序设计实验
1205104002	周培培	77	87	87	89	84	74	87	83
1205104003	吴恩慧	74	87	76	89	85	79	91	87
1205104005	沈一鸣	81	80	74	82	68	78	91	90
1205104006	王佳炜	47/\$60	74	84	85	92	80	84	83
1205104007	韩雪	76	82	75	72	82	55/\$48	70	85
1205104008	瞿振宇	88	80	91	88	84	89	82	80
1205104009	杨小芳	72	81	88	88	81	68	75	80
1205104010	李琳	87	87	91	92	82	87	82	80
1205104011	陈佳怡	82	83	94	89	85	88	69	77
1205104012	胡桃平	84	91	89	92	84	99	91	90
1205104013	钱栢霆	85	82	89	90	82	89	85	90
1205104014	刘湘	68	70	81	86	83	79	93	90
1205104015	沈莹	69	85	89	92	84	92	89	86

表 2 连续数据离散化结果

学号	姓名	高等数学 A1	C 语言程 序设计	操作系统
1205104002	周培培	87	77	75
1205104003	吴恩慧	87	74	82
1205104005	沈一鸣	80	81	74
1205104006	王佳炜	74	47/\$60	79
1205104007	韩雪	82	76	68
1205104008	瞿振宇	80	88	89
1205104009	杨小芳	81	72	88
1205104010	李琳	87	87	96
1205104011	陈佳怡	83	82	92
1205104012	胡桃平	91	84	94
1205104013	钱栢霆	82	85	81
1205104014	刘湘	70	68	72
1205104015	沈莹	85	69	77
1205104016	邢婷	52/\$37	60	71

学号	姓名	高等数学 A1	C 语言程 序设计	操作系统
1205104002	周培培	1	2	2
1205104003	吴恩慧	1	2	1
1205104005	沈一鸣	1	1	2
1205104006	王佳炜	2	4	2
1205104007	韩雪	1	2	3
1205104008	瞿振宇	1	1	1
1205104009	杨小芳	1	2	1
1205104010	李琳	1	1	0
1205104011	陈佳怡	1	1	0
1205104012	胡桃平	0	1	0
1205104013	钱栢霆	1	1	1
1205104014	刘湘	2	3	2
1205104015	沈莹	1	3	2
1205104016	邢婷	4	3	2

据表 2012 的部分原始数据显示如表 1:

本文数据预处理主要进行数值规约中数据泛化和连续数据离散化。熵值离散法与关联规则集成的算法:首先划分连续数据的离散区间,其次属性值根据离散区间的位置进行分类标号,最后将类别标号转化为布尔标号。成绩数据库中 2012 级学生高等数学、C 语言程序设计、操作系统三门课程的成绩,因为成绩数据有较多的冗余值,只能部分呈现实验结果。下面右边表格表示左边表格中某门课程某个学生成绩分数属于某个区间,区间标识分别为:0、1、2、3、4,对应的布尔属性为:优秀,良好,中等,及格,不及格共五类(表 2)。

成绩数据离散化后,采用 a 表示课程《高等数学

A1》,b 表示课程《C 语言程序设计》,c 表示课程《操作系统》等等,其学生成绩分析逻辑数据表部分如表 3 所示。

(三)关联规则挖掘

本文选取支持度为 0.2,置信度为 0.5。通过关联规则挖掘得到的关联规则结果如表 4 所示。

(四)数据挖掘结果反馈

1. 高等数学 优秀 =>C 语言 优秀 0.436 0.764
- 2.高等数学 良好=>C 语言 良好 0.548 0.801
3. C 语言 良好=>操作系统 良好 0.513 0.681
4. C 语言 优秀=>操作系统 优秀 0.467 0.746
- 5.高等数学 及格=>C 语言 及格 0.621 0.811
- 6.高等数学 不及格=>C 语言 不及格 0.375 0.876

表 3 学生成绩分析逻辑数据表

学号	a	b	c
1205104002	1	2	2
1205104003	1	2	1
1205104005	1	1	2
1205104006	2	4	2
1205104007	1	2	3
1205104008	1	1	1
1205104009	1	2	1
1205104010	1	1	0
1205104011	1	1	0
1205104012	0	1	0
1205104013	1	1	1
1205104014	2	3	2
1205104015	1	3	2
1205104016	4	3	2

表 4 关联规则结果

规则	支持度	置信度
a0=>b0	0.436	0.764
a1=>b1	0.548	0.801
b1=>c1	0.513	0.681
b0=>c0	0.467	0.746
a3=>b3	0.621	0.811
a4=>b4	0.375	0.876
a0,b0=>c0	0.331	0.748
a1,b1=>c1	0.402	0.697
a3,b3=>c3	0.336	0.764
a4,b4=>c4	0.348	0.849
a2,b2=>c2	0.313	0.681

- 7.高等数学 优秀,C 语言 优秀=>操作系统 优秀 0.331 0.748
- 8.高等数学 良好,C 语言 良好=>操作系统 良好 0.402 0.697
- 9.高等数学 及格,C 语言 及格=>操作系统 及格 0.336 0.764
- 10.高等数学 不及格,C 语言 不及格=>操作系统 不及格 0.348 0.849
- 11.高等数学 中等,C 语言 中等=>操作系统 中等 0.313 0.681

上述的关规则可以看出：学生高等数学的成绩好坏,直接影响 C 语言的成绩。C 语言的成绩好坏,直接影响操作系统的性能。如果高等数学和 C 语言

成绩都优秀，操作系统成绩也优秀的概率达到 74.8%。本次实验结果只是针对某三门课程,但是在课程库中海量的课程信息之间还存在许多隐含规则,比如:确定哪些课属于基础课、专业基础课、专业课;基础课和专业课之间的先后关系;专业基础课和专业课之间的先后关系;专业课之间还存在一些特定的先后关系等等，对这些规则的挖掘将是下一步要探讨的工作，这些重要规则的挖掘对制定教学计划进行课程设置，确定前续和后继课程有一定的帮助。循序渐进的学习过程,需要把基础打好打扎实。如果学生对前继课程掌握得不好，那么对学习后继课程也会受到一定的影响。课程之间存在一定的内在关联和先后顺序。对课程的相关性挖掘来预测课程之间的关联，还可以根据关联结果制定该专业的教学计划,使教学效果达到最佳,提高高校的教学水平。

四、结束语

数据挖掘技术应用在学生成绩管理的决策上，并发掘影响学生成绩的重要因素和课程之间潜在的规律知识。通过挖掘的结果进行科学分析,为教学管理人员提供科学的决策信息，并合理的安排课程进行课程建设和专业建设,提高学生的学习兴趣,这对高校提高教学质量具有重要的现实意义。

参考文献：

[1] 袁小玲.数据挖掘在学生成绩管理中的应用研究[D].上海:华东师范大学,2009.

[2] 周峰. 基于数据挖掘技术在商业银行服务渠道优化策略研究[D].江苏:南京大学,2013.

[3] 朱丽丽. 基于数据挖掘技术的高校教务管理的应用与研究[J].价值工程,2014,(11).

[4] 廖敬. 数据挖掘技术在学生成绩管理系统的应用研究[D].广东:中山大学,2014.

[5] 闫敏娟.基于数据挖掘技术的学生成绩管理研究[D].武汉:武汉理工大学,2011.

[6] 张艳玲. 数据挖掘在高校学生成绩管理系统的应用[D].河北:河北科技大学,2014.

[7] 刘小桃,林延庆.数据挖掘在高校学生学习成绩分析中的应用[J].考试周刊,2014,(41).

[8] 樊同科,孙姜燕.基于数据挖掘的高校学生学习成绩分析应用研究[J].计算机与现代化,2013,(3).

[责任编辑：刘 骋]  
(下转第 96 页)

(上接第 78 页)

## The Research and Application on the Score Management at Universities Based on Data Mining Technology

ZHU Li-li

(Teaching Affairs Office, Jingling Institute of Technology, Nanjing 211169, China)

**Abstract:** At present, the system of the student score management in most college can only store students' scores for inquiries and statistics. For the credit system and selecting courses, the system of the score management is need to be improved. After the construction of data warehouse, data mining tools can be utilized to analyze the potential relationship between courses and scores and the major factors, which supplies an important foundation to suitable teaching management, helps to make the reasonable and scientific management decision, Meanwhile it provides better services to students and teachers and enhances the overall competitive power of a college.

**Key words:** data mining; student score; relevant analysis; frequent item sets

(上接第 80 页)

- [J].材料科技与设备,2014,(4):76-78. [3] 朱茂桃,夏长高,高翔.膜片弹簧疲劳断裂的试验分析[J].汽车工程,2001,(2):139-142.
- [2] 叶玉春. 喷丸强化对汽车离合器膜片弹簧疲劳寿命影响的实验研究[J].机械工程师,2009,(07):46-47. [责任编辑:詹华西]

## Cause analysis of Diaphragm Spring

CHENG Ji-hao LIU Wen-bin WAMG Jun-lin LIU-Min

(Research and Development Center of Wuhan Iron and Steel(Group)Steel Co., Wuhan 430080, China)

**Abstract:** According to the phenomenon of 50CrVA diaphragm spring cracking, this paper tries to find out the reasons that observing the carking sample by metallographic microscope and SEM. It shows that, at the time of working the outer circular of the diaphragm spring, micro-cracks appeared on the edge, then during heat treatment, early cracking site was oxidized, at last it was instability cracking at the time of elastic detection.

**Key words:** diaphragm spring; oxid; crack