

数据挖掘在体育院校计算机等级考试 成绩分析中的应用研究

刘钟情

(成都体育学院 信息技术中心, 四川 成都 610000)

摘要: 为了提高体育院校学生计算机等级考试过级率,使用 Weka 平台,采用 Apriori 关联规则对成都体育学院 13 级部分班级的计算机基础一和基础二课程学生平时测试成绩及平时测试次数、期末卷面成绩、计算机等级考试成绩进行数据挖掘分析,获取了一些有益的分析结果,对提高计算机基础课程的教学效果起到一定的指导作用。

关键词: 数据挖掘; Apriori 算法; Weka; 体育院校; 关联规则; 成绩分析

中图分类号: TP311 **文献标志码:** A **文章编号:** 1673-0143(2016)-0377-05

DOI: 10.16389/j.cnki.cn42-1737/n.2016.04.015

Application of Data Mining in Computer Grade Examinations in Sports Colleges and Universities

LIU Zhongqing

(Information Technology Center, Chengdu Sport Institute, Chengdu 610000, Sichuan, China)

Abstract: In order to improve the level of computer grade examinations of students in sports colleges and universities, Weka software and Apriori are used to do data mining on the data including the courses of computer basic one and basic two with students' usually test scores and usually test times, the results in the final exam, computer grade test scores of some classes in Chengdu Sport Institute, thus some beneficial analysis results are obtained to improve the teaching effect of computer basic course.

Keywords: data mining; Apriori algorithm; Weka; sports colleges and universities; association rules; performance analysis

0 引言

近年来,计算机技能已经成为衡量体育院校学生素质的重要方面,对于学生而言,计算机一级和二级证书也得到了很多用人单位的认可。因此,计算机基础课程也日益受到体育院校的重视。由于各方面的原因,学生的等级考试过级率普遍不高。因此,各种针对提高体育院校学生计算机等级考试的研究也随之而来。以笔者所在的成都体育学院为例,为了提高学生计算机基础课程的学习积极性,教研室主任沈乐君副教授开发了大规模随机化考试系统,系统可按章节知识点内容为计算机基础课程随机生成试题,便于任课教师随堂考试,从而提高教学效果。近两年来,教研室老师已形成计算机

基础一和基础二平时测试5次的惯例。目前对于学生计算机基础课程成绩的计算方式是:该门课总评成绩为期末卷面分的60%+学生平时分的40%,学生平时5次测试作为计算机基础课程平时分的主要打分依据,学期末分析总评成绩的最低分、最高分、平均分、及格人数、各分数段的人数。而对于学生计算机等级考试成绩则只简单地统计各班及各任课教师的学生通过人数。这种成绩分析方法虽然能一定程度上纵向了解各班级学生对课程的掌握程度。但使用随机化考试系统有没有提高学生的等级考试过级率,平时测试的成绩、平时测试的次数、学生期末卷面成绩与学生计算机等级考试成绩之间有哪些必然联系?本研究旨在使用数据挖掘技术找出学生计算机平时测试成绩及次数、期末卷面成绩与计算机等级考试成绩之间的潜在关系,从而提高学生的计算机等级考试过级率提供依据。

1 数据挖掘与Apriori算法

数据挖掘是指从大量的数据中通过算法搜索隐藏于其中信息的过程。将先进的数据挖掘技术引入教育考试数据分析领域,可以获得潜在规律,为教育管理和决策提供参考^[1]。关联规则数据挖掘是数据挖掘领域的热点之一。关联规则反映海量数据集中数据项集之间的关联关系。Apriori算法是第一个关联规则挖掘算法,本研究采用Apriori算法对计算机基础课程的平时测试成绩及次数与计算机等级考试成绩之间进行关联规则挖掘。Apriori算法采用逐层搜索的迭代方法,利用先验知识进行候选数据项集剪枝,缩小搜索范围,以下是Apriori算法产生频繁项集部分的伪代码^[2]。

```

k=1
Fk={i| $i \in I \wedge \sigma(\{i\}) \geq N \times \text{minsup}$ } //生成1阶频繁项集
repeat
    k=k+1
    Ck=apriori-gen(Fk-1) //生成候选数据项集
for 每个事务 t ∈ T do
    Ct=subset(Ck, t) //生成事务 t 中包含的候选数据项集
for 每个候选项集 c ∈ Ct do
    σ(c)=σ(c)+1 //累加候选数据项集的支持数
end for
end for
Fk={c|c ∈ Ck ∧ σ(c) ≥ N × minsup} //生成 k 阶频繁项集
until Fk=∅
result=∪ Fk

```

2 数据挖掘

2.1 数据分析与数据准备

为了挖掘学生计算机基础一和计算机基础二平时测试成绩及次数、期末考试卷面分、计算机等级考试成绩之间的关联规则,需要计算机基础一和计算机基础二学生的平时测试成绩及次数,相应学生的期末考试卷面分,计算机等级考试成绩。由于计算机基础课的平时测试成绩许多老师在教务管理系统中登录完期末考试成绩之后就删掉了,只能从少数任课老师那里取得还存留的部分班级平时测试成绩的数据和期末卷面成绩,相应班级的计算机等级考试成绩从教务处获得。从成都体育学院13级旅游管理和13级经济管理班提取54位学生计算机基础二平时测试和期末卷面成绩及计算机二级等级考试数据,又从13级体育教育4个班级提取105位学生计算机基础一平时测试和期末卷面成绩及计算机一级等级考试数据。这54位计算机基础二和105位计算机基础一的学生包括等级考试报名且参加考试和报名未考试的学生,且所有被抽选的学生计算机基础一和基础二的任课教师均为同一个教学经验丰富的老师。

2.2 数据预处理

将学生姓名、计算机等级考试笔试、机试成绩、实际参加平时测试的次数、平时测试的5次成绩、期末卷面成绩合并至同一数据表中^[3]。由于计算机基础一等级考试只有机试没有笔试,且计算机基础一和计算机基础二的平时5次测验的评分细则不同,所以将计算机基础一和计算机基础二的学生记录分别放在不同的数据表中。图1和图2分别为计算机基础一和计算机基础二处理后的数据表结构及小部分表记录。

姓名	computer	times	first	second	third	fourth	fifth	end
胡杨	69	5	34	21	35	42	16	68
赖欢	64	5	30	11	34	36	7	48
杨业宝	65	5	41	18	37	42	20	77
李凯	51	4	0	10	23	36	5	63

图1 计算机基础一数据表

Fig. 1 Data of course of computer basic one

姓名	pen	computer	times	first	second	third	fourth	fifth	end
王哲	75	61	5	17	17	13	18	18	84
李艳荣	67	64	5	16	16	12	16	17	74
陆佳瑶	68	61	5	16	16	14	14	16	74
厉玉倩	67	62	5	17	17	14	16	16	80

图2 计算机基础二数据表

Fig. 2 Data of course of computer basic two

2.3 数据转换

为适应 Apriori 算法的要求,需要对计算机基础一和基础二数据表数据进行数据转换。将姓名字段转换为 ID 字段,对两表的机试与笔试成绩处理为:通过为 t,不通过设置为 f,再分别对两表的成绩字段 first、second、third、fourth、fifth、end 进行转换,按照设置的相应转换规则进行转换,n 表示优秀,g 表示良好,m 表示中等,b 表示差,o 表示缺考。比如,对于计算机基础二表的 first 字段,20 分以上为 n,15~20 为 g,10~14 为 m,10 以下为 b,缺考为 o,其他字段转换规则依此类推。计算机基础一数据表的转换方法与计算机基础二一致,只是评分细则不完全相同。转换后的数据表如图3和图4所示。

ID	computer	times	first	second	third	fourth	fifth	end
1 t			5 m	g	g	n	m	m
2 t			5 m	b	m	g	b	f
3 t			5 g	m	g	n	g	m
4 f			4 o	b	m	g	b	m

图3 转换后的计算机基础一数据表

Fig. 3 Conversion data of course of computer basic one

ID	level	pen	computer	times	first	second	third	fourth	fifth	end
1	2 t	t			5 g	g	m	g	g	g
2	2 t	t			5 g	g	m	g	g	m
3	2 t	t			5 g	g	m	m	g	m
4	2 t	t			5 g	g	m	g	g	g

图4 转换后的计算机基础二数据表

Fig. 4 Conversion data of course of computer basic two

2.4 数据挖掘结果分析

数据挖掘算法在 Weka 平台上实现。Weka 是由新西兰怀卡托大学开发的用于机器学习和数据挖掘研究的开源工具,能够运行于几乎所有的操作系统平台^[4-5]。Weka 平台汇集了很多当今最前沿的数据挖掘算法,能处理分类、聚类、关联规则挖掘等数据挖掘^[6]。挖掘后的结果见图5和图6。

对计算机基础一关联分析规则解读为:第1条规则,学生第5次平时测试成绩为良好,则其平时测试次数为5,其置信度为0.98;第2条规则,学生第2次平时测试成绩为中等,则其平时测试次数为5,其置信度为0.97;第3条规则,学生第4次平时测试成绩为良好,则其平时测试次数为5,其置信度为0.95;第4条规则,学生一级等级考试机试通过,则其平时测试次数为5,其置信度为0.94;第5条规则,学生第3次平时测试成绩为良好,则其平时测试次数为5,其置信度为0.93;第6条规则,学生第2次平时测试成绩为良好,则其平时测试次数为5,其置信度为0.93;第7条规则,学生期末考试卷面成绩为中等,则其平时测试次数为5,其置信度为0.92;第8条规则,学生一级等级考试通过且第3次平时测试成绩为良好,则其平时测试次数为5,其置信度为0.92;第9条规则,学生第5次平时测试成绩为良好,则其平时测试次数为5,其置信度为0.92;第10条规则,学生第4次平时测试成绩为优秀,则其平时测试次数为5,其置信度为0.91;以上关联规则对计算机基础一任课老师的启示是:要让学生通过计算机基础一,必须让其参加5次平时测试,这也体现出教研室的考试系统对提高学生的等级考试过级率有很大帮助,且从平时各次成绩的档次大致可以预测其等级考试的通过情况。

```

1. fifth=g 42 ==> times=5 41 <conf:(0.98)> lift:(1.06) lev:(0.02) [2] conv:(1.6)
2. second=m 38 ==> times=5 37 <conf:(0.97)> lift:(1.05) lev:(0.02) [1] conv:(1.45)
3. fourth=g 42 ==> times=5 40 <conf:(0.95)> lift:(1.03) lev:(0.01) [1] conv:(1.07)
4. computer=t 63 ==> times=5 59 <conf:(0.94)> lift:(1.01) lev:(0.01) [0] conv:(0.96)
5. third=g 56 ==> times=5 52 <conf:(0.93)> lift:(1.01) lev:(0) [0] conv:(0.85)
6. second=g 40 ==> times=5 37 <conf:(0.93)> lift:(1) lev:(0) [0] conv:(0.76)
7. end=m 52 ==> times=5 48 <conf:(0.92)> lift:(1) lev:(-0) [0] conv:(0.79)
8. computer=t third=g 38 ==> times=5 35 <conf:(0.92)> lift:(1) lev:(-0) [0] conv:(0.72)
9. first=g 60 ==> times=5 55 <conf:(0.92)> lift:(0.99) lev:(-0) [0] conv:(0.76)
10. fourth=n 44 ==> times=5 40 <conf:(0.91)> lift:(0.98) lev:(-0.01) [0] conv:(0.67)

```

图5 计算机基础一关联分析结果图

Fig. 5 Associative analysis of course of computer basic one

```

1. pen =t 30 ==> computer=t 30 <conf:(1)> lift:(1.2) lev:(0.09) [4] conv:(5)
2. first=g 29 ==> computer=t 29 <conf:(1)> lift:(1.2) lev:(0.09) [4] conv:(4.83)
3. pen =t times=5 28 ==> computer=t 28 <conf:(1)> lift:(1.2) lev:(0.09) [4] conv:(4.67)
4. times=5 first=g 28 ==> computer=t 28 <conf:(1)> lift:(1.2) lev:(0.09) [4] conv:(4.67)
5. first=g 29 ==> times=5 28 <conf:(0.97)> lift:(1.24) lev:(0.1) [5] conv:(3.22)
6. computer=t first=g 29 ==> times=5 28 <conf:(0.97)> lift:(1.24) lev:(0.1) [5] conv:(3.22)
7. first=g 29 ==> computer=t times=5 28 <conf:(0.97)> lift:(1.3) lev:(0.12) [6] conv:(3.76)
8. times=5 42 ==> computer=t 40 <conf:(0.95)> lift:(1.14) lev:(0.09) [4] conv:(2.33)
9. pen =t 30 ==> times=5 28 <conf:(0.93)> lift:(1.2) lev:(0.09) [4] conv:(2.22)
10. pen =t computer=t 30 ==> times=5 28 <conf:(0.93)> lift:(1.2) lev:(0.09) [4] conv:(2.22)

```

图6 计算机基础二关联分析结果图

Fig. 6 Associative analysis of course of computer basic two

对计算机基础二关联分析规则解读为:第1条规则,计算机二级笔试通过了,则计算机二级机试一定通过了;第2条规则,第一次平时测试结果为良好,则计算机二级机试一定通过了;第3条规则,计算机二级笔试通过了并且平时测试了5次,则计算机二级机试一定通过了;第4条规则,第1次平时测试为良好并且测试了5次,则计算机二级机试一定通过了。前4条规则都说明:凡是计算机二级笔试通过了的学生,则机试一定通过了,且机试通过与否与第1次平时测试成绩和平时测试次数直接有关。第5条规则,第1次平时测试成绩为良好,则参加了5次测试,其置信度为0.97;第6条规则,计算机二级机试通过并且第1次测试通过,则参加了5次测试,其置信度为0.97;第7条规则,第1次平时测试为良好,则计算机二级机试必通过并且测试了5次,其置信度为0.97;第8条规则,参加了5次平时测试的学生,计算机二级机试都通过了,其置信度为0.95;第9条规则,学生计算机二级笔试通过了,则必参加了5次平时测试,其置信度为0.93;第10条规则,学生计算机二级笔试通过了并且机试通过了,则必参

加了5次平时测试,其置信度为0.93。以上规则对计算机基础二任课老师的启示是:笔试通过比机试通过率低,平时的第1次测试成绩可直接预测学生的机试通过与否,要想确保学生通过二级,则学生必须要参加5次平时测试,这也体现出教研室的考试系统对提高学生的等级考试过级率有很大帮助。

3 结语

使用数据挖掘技术对考试数据进行分析,可以提高考试数据的利用水平,获得潜在的规律和趋势,为提高教学质量提供参考依据。本文使用Weka平台,采用Apriori关联规则对成都体育学院13级部分班级的计算机基础一和基础二课程学生平时测试成绩及平时测试次数、期末卷面成绩、计算机等级考试成绩进行数据挖掘分析,得出了平时测试成绩及次数、期末卷面成绩、计算机等级考试成绩之间的关联规则,各规则的置信度均达到0.91以上。

参考文献(References)

- [1] 罗美淑,刘世勇,夏春艳.数据挖掘技术在教学评价中的应用研究[J].教育探索,2013(2):81-82.
- [2] 袁梅宇.数据挖掘与机器学习:WEKA应用技术与实践[M].北京:清华大学出版社,2014:104-105.
- [3] 陈平,王利钢.Apriori算法在高校成绩分析中的应用[J].信息化研究,2013,39(5):69-70.
- [4] WITTEN I H, FRANK E.数据挖掘:实用机器学习技术[M].北京:机械工业出版社,2006.
- [5] HALL M, FRANK E, HOLMES G, et al. The WEKA data mining software: an update[J]. SIGKDD Explorations, 2009, 11(1):10-18.
- [6] 姜凯,左风朝.Weka平台上解决聚类的改进差分进化算法[J].计算机工程与设计,2012,33(2):591-593.

(责任编辑:范建凤)