

基于 Apriori 算法的高校学生 考试作弊动机分析与应对^①

冯 俊, 胥 莉, 闵 兰

成都理工大学 管理科学学院, 成都 610059

摘要: 依据对高校大学生考试作弊行为的调查数据, 并运用 Apriori 算法, 揭示了高校大学生作弊行为产生的因果关系. 相关实证分析结果显示, 主客观原因的相互交织造成了大学生的作弊行为, 其中学生自身的主观因素是主要原因. 可通过改革教育教学方法, 加强德育教育并树立诚信价值观, 严格考试管理, 制定恰当合理的考试法规等措施, 实现依法治考, 有效预防和惩治高校大学生的作弊行为.

关 键 词: 高校学生; 考试作弊; Apriori 算法; 应对策略; 依法治考

中图分类号: G642

文献标志码: A

文章编号: 1000-5471(2017)02-0174-07

考试, 是我国当前教育体制下检测学生学习情况的主要手段. 事实证明, 诚实应试条件下, 学生对知识的掌握程度与考试成绩呈正相关关系. 为了更好地激励学生获取知识, 大部分院校将学生成绩与各种利益(评优、评先进以及获得奖学金等荣誉)进行挂钩, 而出于对利益的追求或对惩罚的规避, 部分学生选择了考试作弊^[1]. 目前, 考试作弊已经成为高校不可回避的一个问题, 也是学者们研究的一个热点内容^[2-7].

1 研究现状

考试作弊在定义上存在广义和狭义两种界定. 广义的考试作弊是指在考试进行之前、考试进行之中和考试结束之后利用各种手段进行的舞弊行为. 而狭义的考试作弊则仅仅是指在考试进行中采取各种手段的舞弊行为^[8]. 考试作弊作为一种不公平的竞争行为, 一方面严重影响了考试衡量功能的正常发挥, 败坏了学校校风、学风; 另一方面极易造成学生冒险、侥幸心理的养成以及道德的滑坡^[2,6]. 早在 1997 年, 文献[8]在相关的调查研究中就发现, 考试管理方通过讲解和渲染考场规则, 以及加强巡查和督导等方式对考试作弊而言只能是治标不治本. 这也在一定程度上解释了为杜绝高校学生的作弊行为, 国家教育部门三令五申地强调要做好考试管理工作却收效甚微的原因. 另外, 文献[8]还认为, 对于学生作弊行为的研究要注重其心理和情感变化. 因此, 了解大学生作弊的实际情况对于预防考试作弊行为具有十分重要的意义. 为此, 本文专门对部分在校学生进行了随机抽样调查, 以期能够为作弊行为的研究提供真实客观的数据资料.

从大量的、不完全的、模糊的、随机的数据中, 提取出隐含其中的人们事先不知道但又潜在存在的信

① 收稿日期: 2016-04-19

基金项目: 成都理工大学中青年骨干教师培养计划项目(JXGG201412); 成都理工大学高等教育人才培养质量和教学改革项目(13JGY32, 13JGY30).

作者简介: 冯 俊(1979-), 男, 四川自贡人, 副教授, 博士, 主要从事应用数学、数据处理方面的研究.

息和知识的过程^[9-10]称为数据挖掘。作为数据挖掘的基本任务之一,关联规则挖掘就是从给定的数据集中发现频繁出现的项集以及数据中不同领域之间的联系,找出满足给定支持度和置信度阈值的多个域之间的依赖关系^[11-14]。Apriori 算法^[15-16]是各种相关算法中最经典、最著名的挖掘关联规则的算法。本文采用 Apriori 算法对调查问卷采集到的学生作弊行为的数据资料进行了处理,得到了学生进行作弊行为的一系列相关因素,为后续研究对策提供依据。

2 关联规则挖掘与 Apriori 算法

关联规则挖掘是指从大量数据的项集之间发现有趣的关联或相关联系。它在数据挖掘(KDD-Knowledge Discovery in Database)中是一个重要的课题,主要用于发现大型数据库中隐含的有价值的令人感兴趣的联系及规则,即在事务数据库中找出满足用户给定的最小支持度和最小置信度的关联规则。关联规则挖掘近几年已被广泛研究,应用于工程、金融、电信、教育、医学等领域。

令 $I = \{i_1, i_2, \dots, i_m\}$ 表示所有数据项目的集合,每个元素为一个项,共有 m 个数据项; $D = \{T_1, T_2, \dots, T_n\}$ 是所有事务的集合。以 T 表示事务(一些项目的集合), n 为所有事务的数目。 T 包含在 I 中,每个事务可用唯一的标志符 TID 来标识。

设 X 为某些项目的集合,如果 X 包含在 T 中,则称事务 T 包含 X 。关联规则“(X 包含在 T 中) $\Rightarrow$ (Y 包含在 T 中)”的意义为:在一个事务中某些项出现,可以推出另一些项在同一事务中也出现,记为 $X \Rightarrow Y$ (其中, X 包含在 I 中, Y 包含在 I 中,且 X 与 Y 不相交)。

事务集 D 中的规则 $X \Rightarrow Y$ 由支持度 s 和确信度 c 来表示。确信度 c 表示规则的强度,支持度 s 表示在规则中出现的频度。对于某一个具体的数据项 X ,其支持度 $s(X)$ 表示 D 中包含 X 的事务数量与 D 的总的事务数量之比。有时为了方便,数据集 X 的支持度常用数据库 D 中包含的 X 的数量来表示。规则 $X \Rightarrow Y$ 的支持度 s 定义为:在 D 中包含 $X \cup Y$ 的事务所占比例为 $s\%$,表示同时包含 X 和 Y 的事务数量与 D 的总事务量之比。规则 $X \Rightarrow Y$ 的确信度 c 定义为:在 D 中, $c\%$ 的事务包含 X 的同时也包含 Y ,表示 D 中包含 X 的事务中有多大可能包含 Y 。

最小支持度阈值(Mini support)表示数据项集在统计意义下的最低频繁度。最小确信度阈值(Mini confidence)表示规则的最低可靠度。一般有用户给定最小确信度阈值和最小支持度阈值。确信度和支持度大于相应阈值的规则称为强关联规则。而发现关联规则的任务就是从数据库中发现那些确信度、支持度大于等于给定阈值的强壮规则。

关联规则挖掘问题可以分解为两个子问题:(1)找出事务数据库 D 中所有大于等于用户指定的最小支持度的项目集。具有最小支持度的项目集称为最大项目集。(2)利用最大项目集生产所需要的关联规则。根据这两个问题,文献[16]提出了 Apriori 算法,主要用于挖掘交易数据库中项集间的关联规则。

Apriori 算法的基本步骤为:第一步,简单统计所有含有一个元素的项目集出现的频率,并找出那些不小于最小支持度的项目集,即一维最大项目集。第二步,循环。在第 k 步中,根据第 $k-1$ 步生成的 $k-1$ 维最大项目集产生 k 维候选项目集,对数据库进行搜索,得到候选项目集的项集支持度,与最小支持度进行比较,从而找到 k 维项目集。

3 问卷调查与分析

在本次调查问卷的设计上,参考了以往相关研究的内容,并重点考虑了本校学生的实际情况^[17]。问卷内容共涉及 13 个选项,涵盖了学生作弊的主观方面(学习状况、对待考试和考试作弊的态度等)、客观方面(监考老师、整体校风学风、学院对作弊的惩罚力度等)、学生自我认为的作弊原因以及常用的作弊手段,

并就一些消除作弊现象的方法征求了学生的建议.

问卷调查的对象以成都理工大学管理科学学院的学生为主体,涵盖了本科 2010 级至 2013 级 4 个年级.管理科学学院作为成都理工大学的一个大院,现具有 9 个本科专业,1 个专科专业,具有较强的代表性.在调查方法上主要采用简单随机抽样的方法随机抽取了 600 名学生进行问卷调查,其次辅以座谈与专访相结合.为确保本次调查信息的可靠性,在问卷调查中采取匿名填写的方式,因此也在调查过程中得到了学生的积极配合.本次调查共发放问卷 600 份,收回 588 份,有效问卷回收率 98%.问卷的结果对于学生作弊行为的分析具有很大的参考价值.

问卷中,学生作弊行为的形成被划分为主观和客观两个层面.主观方面,选取 3 项指标,包括:学生学习的理解程度、学生对考试的认识、学生对待作弊的态度 3 个方面(具体数据见表 1);客观方面,选取了 3 项指标,包括:学院对作弊的惩罚力度、学院的考风、考试诚信承诺书的有效性(具体数据见表 2).最后,我们就学生对作弊现象泛滥的责任方认识以及是否参与过作弊等要素进行了调查(具体数据见表 3).

表 1 作弊行为的主观因素

Q3: 学生对教师上课内容的理解程度			Q4: 学生对待考试的认识			Q5: 学生对待作弊的态度		
选项	选择人数	构成比例/%	选项	选择人数	构成比例/%	选项	选择人数	构成比例/%
A 完全理解	32	5.4	A 获得奖学金的途径	56	9.5	A 很正常	69	11.7
B 大部分能理解	291	49.5	B 对自己的阶段总结	449	76.4	B 不对,但可以理解	462	78.6
C 能理解小部分	234	39.8	C 不挂科就行	83	14.1	C 可耻行为	34	5.8
D 完全不理解	31	5.3				D 与我无关	23	3.9
合计	588	100.0	合计	588	100.0	合计	588	100.0

主观因素分析:(1)有接近 55%的学生能够理解上课教师所讲的大部分内容,而有约 40%的学生不能有效理解,只有约 5%的学生完全不能理解.(2)大部分学生将考试看作是一个阶段性总结而不是获得奖学金的途径;在对于考试成绩的追求上,有相当一部分学生仅仅停留在对自己的最低要求及格水平上.(3)有约 90%的学生表示能够理解学生的作弊行为,只有约 5.8%的学生认为考试作弊是一种可耻行为.(4)数据显示,考试作弊已经在广大学生中形成不以为耻的心理认可.

表 2 作弊行为的客观因素

Q10: 学生对学院作弊惩罚力度的感知			Q11: 学生对学院考风的感知			Q12: 学生对考试诚信承诺书有效性的认识		
选项	选择人数	构成比例/%	选项	选择人数	构成比例/%	选项	选择人数	构成比例/%
A 惩罚力度大	108	18.4	A 很好	117	19.9	A 有用,学生不再作弊	403	68.5
B 惩罚适度	391	66.5	B 一般	436	74.1	B 形同虚设,照常作弊	185	31.5
C 惩罚力度小	19	3.2	C 不好	21	3.6			
D 不了解惩罚措施	70	11.9	D 其他	14	2.4			
合计	588	100.0	合计	588	100.0	合计	588	100.0

客观因素分析:(1)大部分学生所感知到的学院考风整体上一般.(2)七成以上的学生认为学院对作弊行为的处罚适度.(3)学生考试诚信承诺书在学生群体中有一定的督促作用,但也有约 1/3 的学生认为这仅仅是一种形式主义.

显然,在学生考试作弊行为这一问题上,考试学生、监考老师和学校的相关要求 3 个方面构成了问题的主要方面.表 3 显示,在作弊现象普遍这一问题上,学生认为三者的综合作用导致了作弊的泛滥.这与以往的研究是吻合的,而我们的调查进一步发现:学生自身的不诚信是构成这 3 个主要方面的重要方面.

主客观两方面因素的作用结果就是学生作弊行为的产生与否. 基于此, 我们专门就学生是否在考试中参与过作弊这一问题进行了调查. 调查结果显示, 接近有 1/5 的同学曾经参与过考试作弊.

表 3 对考试作弊现象的认识

Q9: 作弊现象泛滥的责任方(含部分多选)			Q6: 学生是否参与过考试作弊		
选项	选择人数	构成比例/%	选项	选择人数	构成比例/%
A 老师监考不严	33	5.6	A 参与过	117	19.9
B 学生不诚信	381	64.8	B 没参与过	471	80.1
C 学校要求高	54	9.2			
D 都有	137	23.3			
合计	605	102.9	合计	588	100.0

其实, 在实际环境中, 两种影响因素并非相互独立而是相互影响的, 其影响的相互性可以从表 4 得到证实. 同时, 表 4 也证实了学生自身的不诚信行为是造成考试作弊的主要原因. 对于考试作弊的主要原因, 进行了详细调查并排序, 详见表 4.

表 4 学生作弊原因与处理手段

Q7: 学生自身考试作弊的具体原因			Q8: 学生常用的作弊手段			Q13: 减少大学生考试作弊现象的方法		
选项	选择人数	构成比例/%	选项	选择人数	构成比例/%	选项	选择人数	构成比例/%
B 不愿意补考(重修)	516	87.8	A 带小抄	348	59.2	D 改革考试方式	497	84.5
H 知识难,学不懂	360	61.2	E 都有	106	18.0	F 适当调整考试难度	439	74.7
C 家长期望	217	36.9	B 传纸条,打手势	89	15.1	B 注重实践能力的考察	435	74.0
D 不愿得罪同学,协调作弊,从众心理	214	36.4	C 找人代考	67	11.4	E 加强学生思想道德教育	116	19.7
G 因贪玩等不想学习	203	34.5	D 高科技	61	10.4	C 严格监考制度	80	13.6
F 拿学位证和找工作压力	194	33.0				A 加大处罚力度	77	13.1
E 各项工作导致没时间学习	101	17.2						
A 获得奖学金	73	12.4						

注: 此题为多项选择题.

表 4 数据显示出, 来自于补考或重修、学习困难、外部的压力以及作弊环境的影响, 不良学风的反作用以及获取学位和找工作的压力是导致学生作弊的几个主要因素.

4 基于 Apriori 算法的数据分析及应对策略

基于 Apriori 算法, 我们对调查问卷得到的数据进行了深入的分析. 在最小支持度取 0.4, 最小确信度为 0.5 的条件下, 获得符合条件的多条相关规则. 表 5 给出了挖掘出的部分结果, 以及相应的确信度. 如规则 1, 表示 10B 这个选项出现的情况下 11B5B 同时出现的确信度为 0.736 573.

从逻辑上来讲, 学生对作弊的理解和认可必然带来较高的作弊率, 而较高的作弊率又会反过来影响考风, 考风的形成受到影响, 良好的学风也就难以形成, 而统计数字(表 5)却显示了逻辑上的矛盾. 对此, 我们通过调查访谈发现, 学生感知到的学风中包含了很小比例的考风, 他们所认为的学风是指学院平时的整体学习氛围.

根据 Apriori 算法挖掘出的结果, 对于学生作弊这一现象的产生, 我们进行了简单的定性分析. 在分析之前, 我们首先假设影响考试作弊行为的因素为两种相互独立的因素, 即主观因素和客观环境因素. 在超

过一半以上的学生(54.9%)能够理解学习内容的前提下,对考试作弊起推动作用的主要因素为对考试作弊的理解认同(91.9%),这里向我们展示了至少3个严肃的问题:(1)在能较好理解所学内容(54.9%)的前提下,仍然有很多的学生(91.9%)认同了考试作弊这一违纪行为。(2)在普遍认为学校营造的整体考试环境、考风较好或一般(94.0%)的条件下,个体对于作弊行为的理解和认同(90.4%)在考试作弊问题上起了主要的作用(20.0%)。(3)目前防治作弊的办法,如考试承诺书,在大家觉得对考试作弊有防止作用的条件下(68.5%),仍有11.7%的人具有作弊行为。

表 5 Apriori 算法部分结果(最小支持度 0.4)

规则编号	项目 X	项目 Y	确信度	规则编号	项目 X	项目 Y	确信度
1	10B	11B5B	0.736 573	15	5B	10B11B6B	0.561 822
2	10B	11B5B9B	0.647 059	16	5B	10B12A6B	0.546 638
3	10B	11B6B9B	0.682 864	17	5B	10B6B	0.635 575
4	10B	12A6B	0.718 670	18	5B	10B6B9B	0.559 653
5	10B	12A	0.782 609	19	5B	6B	0.828 633
6	10B	12A5B6B	0.644 501	20	5B	6B9B	0.622 560
7	10B	12A5B9B	0.621 483	21	5B	9B	0.683 297
8	10B	12A6B	0.718 670	22	6B	12A	0.766 739
9	10B	12A6B9B	0.654 731	23	6B	12A5B	0.669 546
10	10B	6B	0.861 893	24	6B	12A5B9B	0.552 916
11	10B	6B9B	0.746 803	25	6B	13BDF	0.555 076
12	11B	10B	0.786 697	26	6B	5B	0.825 054
13	11B	10B12A	0.626 147	27	6B	5B9B	0.619 870
14	11B	10B12A6B	0.582 569	28	6B	9B	0.710 583

另外,Apriori 算法挖掘出的关联规则还显示出了一些问题:(1)在认为学院对作弊的惩罚力度适度(84.9%)的情况下,仍有一部分同学(17.2%)在考试中会有作弊行为。(2)在认为学院对作弊的惩罚力度适度的同学中,有64.7%的同学认为学院的考风一般,但对作弊行为都表示理解,认为学生对考试的不诚信是考试作弊行为的主要原因。(3)在学院对考试作弊惩罚力度适度的情况下,有24.8%的同学认为考试诚信承诺书对考试作弊来说形同虚设,只是一个形式。剩下的75.2%的同学认为对考试作弊行为起到了很好的作用;不过进一步挖掘可以发现,在这一部分认为有用的同学中,仍有10.9%的同学在考试中具有作弊行为;剩下的89.1%的同学表示没有作弊行为。

5 考试作弊的应对策略

若要有效地预防甚至杜绝考试作弊行为的发生,根据我们对以上各表调查数据的资料分析,和利用Apriori 算法挖掘出来的结果(表 5),我们认为应该从以下 4 个方面做起:

(一)推进教学体制改革,适当变换考试方式. 传统应试教育因为重知识轻能力,重笔试轻实践,重结论轻过程而越来越引起广大学生的反感. 而且,部分教师在授课方式上墨守陈规,教学内容枯燥乏味,不注重学生实践能力和创新能力的培养,不注重理论与实际的结合,造成学生学习兴趣的下降甚至产生厌学的情绪,考试方式基本上采取闭卷形式,而逢期末考试,学生因为平时所学不多就只能采取作弊的手段以期能够蒙混过关. 根据表 4 的结果显示,学生迫切希望学校能够改革考试方式,注重对于实践能力的考察,这也就要求我们的任课教师在授课过程中去多尝试一些新的教育教学方法,在考试方式上进行一些新的变革.

(二)加强学生德育教育,树立诚信应考的价值观念. 考试作弊不仅意味着学校思想道德教育在作弊者本人身上的失败,同时也极易使班上的其他同学面临道德上的巨大考验. 当某位同学因为作弊而逃避了处罚时,就会引起其他同学的效仿,而且一旦这种效仿开始就容易在学生中间形成恶性循环,最终引起整所

院校考风学风的败坏.与此相反,树立强烈的诚信应考的道德观念就能够有效地遏止不诚信考试行为的发生,而且一旦这种观念内化成一种文化就会极大地促进优良学风的建设和发展.

(三)积极采取切实有效的反作弊手段,严肃考试纪律.良好的考试纪律是学风优良的必要条件.预防治理考试作弊现象也不能忽视学生常用的作弊手段.问卷中我们专门调查了学生认为的常用的作弊手段,如表4所示.表4显示,学生的作弊手法呈现出越来越强的隐蔽性,因此,无论是学校还是监考老师,都要特别注意学生作弊手段的使用,特别是在重大的考试面前要更加留意高科技的作弊手段.另外,根据经验,一些诸如偷看、打手势等不易提供作弊证据的作弊手法逐渐在广大学生中间流行开来,监考老师要特别注意这一现象的存在.在考试纪律上,要加强考前考后的试卷管理工作,杜绝舞弊行为的发生.

(四)建立相关的考试法规,依法治考.强化理论认为,行为是结果的函数,当某一不期望出现的行为出现的时候给予一个负的强化物能够减少这种行为出现的可能.而且,从公平理论的角度来讲,如果作弊行为得不到应有的惩罚就会引起其他人心中的不满,这样不但破坏了考试的严肃性,而且极易引起其他同学的作弊效仿.因此,建立相关的考试法规,确定适当的惩罚措施能够在减小作弊行为的发生上起到很好的作用,而且合理的惩罚决定也常常能够达到以儆效尤的目的,对于以后学生的作弊念头也能起到很好的打压作用.而且,只是从道德方面来约束学生远离作弊也是不可能的,唯有建立相关的考试法规制定,将考试纳入法制轨道,将考试作弊行为从性质上上升到违纪违法的层次,真正实现依法治考,才能从根本上减少甚至杜绝考试作弊行为.

6 结 论

通过对调查数据的挖掘和分析,可以发现,在学校对考试作弊惩罚力度适度的情况下,仍有部分同学在考试中会有作弊行为.作弊行为的产生最主要的原因是同学们对考试作弊的理解甚至认同,以及学生的不诚信.

根据分析的结果,我们认为,对于大多数考试会作弊的人,考试承诺书并不能起到很好的作用,只有依靠外部约束才能阻止他们的考试作弊行为.由此,只有推进教学体制改革,适当变换考试方式,加强学生德育教育,树立诚信应考的价值观念,积极采取切实有效的反作弊手段,严肃考试纪律,制定严格的考试规范,将考试制度提升到法律方法法规的高度,依法治考,才能有效地保证考试的公平性,减少考试作弊行为的发生.

参考文献:

- [1] 段小磊.大学生考试作弊现象透视[J].高教发展与评估,2009,25(5):32—37.
- [2] 黄 璜.浅析考试作弊社会危害性及处罚[J].法制博览,2016(1):207—208.
- [3] 杨晓雯.美国高利益关联考试中的教师作弊研究[J].基础教育,2016,13(1):100—105.
- [4] 陈幼平.大学生考试作弊及其相关因素的研究[J].中国成人教育,2010(11):68—69.
- [5] 宗志翔,李国红.考试舞弊的法律分析[J].政法学刊,2005,22(4):65—67.
- [6] 张 京,庞引巧.大学生诚信教育的对策探究[J].社会科学论坛,2007(8):167—169.
- [7] 刘 军,李天舒.基于模糊综合评价的《机械创新设计》课程考核方式研究[J].西南师范大学学报(自然科学版),2015,40(8):177—180.
- [8] 兰 刚.大学生考试作弊的实证研究[J].青年研究,1997(4):37—42.
- [9] 崔 妍,包志强.关联规则挖掘综述[J].计算机应用研究,2016,33(2):330—334.
- [10] 江明华.关联规则挖掘在板球比赛技战术分析中的应用研究[J].西南师范大学学报(自然科学版),2014,39(4):190—194.

- [11] 陈 英, 何中市. 关联规则在高职学生综合素质测评分析中的应用 [J]. 西南师范大学学报(自然科学版), 2011, 36(4): 168—173.
- [12] 张宗郁, 张亚平, 张静远, 等. 改进关联规则算法在高校教学管理中的应用 [J]. 计算机工程, 2012, 38(2): 75—77, 81.
- [13] 薛福亮, 马 莉. 利用动态产品分类树改进的关联规则推荐方法 [J]. 计算机工程与应用, 2016, 52(4): 135—141.
- [14] 夏大飞, 张自力, 周 欢. 基于关联规则的高考数据隐私保护 [J]. 西南师范大学学报(自然科学版), 2013, 38(6): 129—135.
- [15] AGRAWAL R, IMIELINSKI T, SWAMI A N. Mining Association Rules Between Sets of Items in Large Databases [J]. Sigmod Record, 1993, 22(2): 207—216.
- [16] AGRAWAL R, SRIKANT R. Fast Algorithms For Mining Association Rules [C]//in Proc 20th Int Conf Very Large Data Bases Chile; Morgan Kaufmann, 1994: 487—499.
- [17] 余益兵. 社会适应问卷(简式版)在农村中小學生样本中的应用 [J]. 西南大学学报(自然科学版), 2015, 37(12): 140—145.

Apriori Algorithm Based Survey Analysis and Coping Strategies on Students' Examination Cheating in Higher University

FENG Jun, XU Li, MIN Lan

College of Management and Science, Chengdu University of Technology, Chengdu 610059, China

Abstract: Based on the researches in universities, the causal relationships of students' cheating can be revealed with the Apriori algorithm. The empirical analysis shows that the combination of the subjective and objective reasons causes students' cheating, and the students' own objective reasons are the main reasons. Reforming educational methods, enhancing moral education and setting up the value of integrity, managing examination strictly, creating strict law and legislation for the examination and managing examination affairs according to the law can reduce and even prevent the students' cheating.

Key words: college students; examination cheating; Apriori algorithm; coping strategies; managing examination affairs according to law

责任编辑 廖 坤