

利用 Apriori 算法对医学生成绩进行课程关联性分析

牛猛

(皖南医学院 教务处, 安徽 芜湖 241002)

[摘要]在高校教学管理中, 成绩管理非常重要。通过分析成绩数据, 挖掘出隐含在成绩数据中的有价值的规律, 能辅助教学决策, 对专业建设、培养方案制定、课程设置、教学管理等方面具有重要的指导意义。主要运用 Apriori 算法对学生成绩进行分析, 挖掘出有关联性的课程成绩之间的相互规律。

[关键词]Apriori 算法; 成绩; 课程关联性; 分析

doi: 10.3969/j.issn.1673-9477.2015.02.032

[中图分类号] G642.47 **[文献标识码]** A **[文章编号]** 1673-9477(2015)02-115-04

高校医学生的培养方案, 其课程的设置, 是按照公共基础课、专业基础课、专业课程的先后顺序来安排的。从内容上来说, 公共基础课是专业基础课的基础, 专业基础课是专业课程的基础; 从学习效果上说, 公共基础课的学习效果对专业基础课的学习有很大的影响, 专业基础课的学习效果对专业课程的学习有很大的影响, 前期开设的专业课程的学习效果对与之有关联的后期开设的专业课程的学习也有很大的影响。

本文将从具体的课程关联上, 对医学生成绩进行分析, 利用 Apriori 算法对同一专业的课程成绩进行关联规则^[1]挖掘: 挖掘出有一定关联的公共基础课程成绩对专业基础课程成绩的影响; 有一定关联的专业基础课程成绩对专业课程成绩的影响; 有一定关联的专业课程成绩对后续专业课程成绩的影响。用挖掘出的规律反馈教学过程, 在辅助教学决策、提高教师教学质量和增加学生学习效果等方面都能给出指导性意见。

一、公共基础课成绩对专业基础课成绩的影响

以我校 2008 级临床医学专业学生学习基础化学和生物化学为例, 基础化学, 作为公共基础课, 是学习生物化学这门专业的基础。从教务系统中导出 2008 级临床医学专业的学生基本情况表, 共 821 人, 包含学号、姓名、性别、院系、当前所在级、学籍状态、专业、班级、身份证号等共 53 个字段; 导出 2008 级临床医学专业包含基础化学和生物化学的学生成绩表, 其中基础化学共 824 条记录, 生物化学共 824 条记录, 包含学年、学期、课程名称、考核方式、成绩、学分、当前所在级、专业、班级名称、学号、姓名、课程性质、课程类型等共 22 个字段。

然后, 对导出的原始数据进行修改, 如增加或删除字段, 增加或删除记录, 将不同数据

表连接至同一个数据挖掘库中^[2]。将学生基本情况与成绩表中的学号、专业、基础化学成绩和生物化学成绩这 4 个字段导出生成用于挖掘的学生成绩挖掘表 1, 其记录共有 820 条, 如表 1-1 所示。

表 1-1 完成数据预处理的学生成绩挖掘表 1

XH	ZYMC	JCHXCJ	SWHXCJ
200803010134	临床医学	94	80
200803010492	临床医学	81	88
200803010674	临床医学	78	75
200803010681	临床医学	90	85
200803010751	临床医学	87	53
200803010138	临床医学	71	76
200803010533	临床医学	62	75
200803010203	临床医学	63	64
200803010608	临床医学	58	57
200803010775	临床医学	66	70

之后, 根据 Apriori 算法的要求, 对学生成绩挖掘表 1 中的成绩进行离散化。将平均成绩按大于或等于 75 分(优良)、小于 75 分(一般)分别划分成“1”、“0”, 分别表示学习成绩为优良和学习成绩为一般及以下, 如表 1-2 所示。

表 1-2 完成数据离散化的学生成绩挖掘表 2

XH	ZYMC	JCHXCJ	SWHXCJ
200803010134	临床医学	1	1
200803010492	临床医学	1	1
200803010674	临床医学	1	1
200803010681	临床医学	1	1
200803010751	临床医学	1	0
200803010138	临床医学	0	1
200803010533	临床医学	0	1
200803010203	临床医学	0	0
200803010608	临床医学	0	0
200803010775	临床医学	0	0

之后, 给定 $\min_sup^{[3]}$ 为 0.2, $\min_conf^{[4]}$

[投稿日期] 2015-03-25

[作者简介] 牛猛(1982-), 男, 安徽淮北人, 助教, 硕士, 研究方向: 数据挖掘。

为0.5, 再通过Apriori算法挖掘所有的频繁项集^[5], 如表1-3所示。

表1-3 从学生成绩挖掘表2中挖掘出的频繁项集

频繁项集	支持度
基础化学成绩="1"	77.27%
基础化学成绩="0"	22.73%
生物化学成绩="1"	72.12%
生物化学成绩="0"	27.88%
基础化学成绩="1", 生物化学成绩="1"	67.91%
基础化学成绩="1", 生物化学成绩="0"	9.36%
基础化学成绩="0", 生物化学成绩="1"	4.21%
基础化学成绩="0", 生物化学成绩="0"	18.52%

最后, 计算最终频繁项集的所有非空子集的置信度, 删除小于min_conf的记录, 确定模型, 生成关联规则^[6], 如表1-4所示。

表1-4 从学生成绩挖掘表2中挖掘出的关联规则

频繁项集	置信度
基础化学成绩="优良" ⇒ 生物化学成绩="优良"	93.34%

通过挖掘的关联规则, 我们可以发现下述关系:

学生的基础化学成绩, 对学生的生物化学成绩有很重要的影响, 基础化学成绩是优良的, 其生物化学成绩在非常大程度上都是优良的。但基础化学成绩并非绝对因素, 基础化学成绩是一般的学生, 通过努力学习, 有一些都能够获得优良的生物化学成绩。

根据上述关系, 我们可以得出如下结论:

基础化学属于公共基础课, 生物化学属于专业基础课, 并且基础化学这门公共基础课是学习生物化学这门专业基础课的基础。对于同一专业, 有一定关联的公共基础课成绩对专业基础课成绩有非常大的影响。

二、专业基础课成绩对专业课成绩的影响

分析同一专业, 有一定关联的专业基础课成绩对专业课成绩的影响。以2008级临床医学专业学生学习生理学和内科学为例, 生理学, 作为专业基础课, 是学习内科学这门专业课的基础。

按照之前的方法, 将2008级临床医学专业的学生基本情况表与生理学成绩表和内科学成绩表连接至同一个数据挖掘库中, 并将学号、专业、生理学和内科学成绩这4个字段导出生成用于挖掘的学生成绩挖掘表3, 其记录共有820条, 如表2-1所示。

表2-1 完成数据预处理的学生成绩挖掘表3

XH	ZYMC	SLXCJ	NKXCJ
200803010196	临床医学	86	78
200803010392	临床医学	92	87
200803010773	临床医学	90	92
200803010476	临床医学	86	68
200803010542	临床医学	80	57
200803010693	临床医学	68	87
200803010778	临床医学	64	92
200803010035	临床医学	72	62
200803010037	临床医学	69	62
200803010493	临床医学	60	52

之后, 按照成绩是否大于或等于75分进行离散化, 如表2-2所示。

表2-2 完成数据离散化的学生成绩挖掘表4

XH	ZYMC	SLXCJ	NKXCJ
200803010196	临床医学	1	1
200803010392	临床医学	1	1
200803010773	临床医学	1	1
200803010476	临床医学	1	0
200803010542	临床医学	1	0
200803010693	临床医学	0	1
200803010778	临床医学	0	1
200803010035	临床医学	0	0
200803010037	临床医学	0	0
200803010493	临床医学	0	0

之后, 按照min_sup为0.2, min_conf为0.5挖掘所有的频繁项集, 如表2-3所示。

表2-3 从学生成绩挖掘表4中挖掘出的频繁项集

频繁项集	支持度
生理学成绩="1"	78.13%
生理学成绩="0"	21.87%
内科学成绩="1"	72.75%
内科学成绩="0"	27.25%
生理学成绩="1", 内科学成绩="1"	67.59%
生理学成绩="1", 内科学成绩="0"	10.54%
生理学成绩="0", 内科学成绩="1"	5.16%
生理学成绩="0", 内科学成绩="0"	16.71%

最后, 计算最终频繁项集的所有非空子集的置信度, 删除小于min_conf的记录, 确定模型, 生成关联规则, 如表2-4所示。

表2-4 从学生成绩挖掘表4中挖掘出的关联规则

频繁项集	置信度
生理学成绩="优良" ⇒ 内科学成绩="优良"	93.11%

通过挖掘的关联规则, 我们可以发现下述关系:

学生的生理学成绩, 对学生的内科学成绩

有很重要的影响，生理学成绩是优良的，其内科学成绩在非常大程度上都是优良的。但生理学成绩并非绝对因素，生理学成绩是一般的学生，通过努力学习，有一些都能够获得优良的内科学成绩。

根据上述关系，我们可以得出如下结论：

生理学属于专业基础课，内科学属于专业课，并且生理学这门专业基础课是学习内科学这门专业课的基础。对于同一专业，有一定关联的专业基础课成绩对专业课成绩有非常大的影响。

三、专业课成绩对专业课成绩的影响

分析同一专业，有一定关联的专业课成绩对专业课成绩的影响。以 2008 级临床医学专业学生学习手术学和外科学为例，虽同为专业课，但手术学是学习外科学的基础。

按照之前的方法，将 2008 级临床医学专业的学生基本情况表与手术学成绩表和外科学成绩表连接至同一个数据挖掘库中，并将学号、专业、生理学和内科学成绩这 4 个字段导出生成用于挖掘的学生成绩挖掘表 5，其记录共有 820 条，如表 3-1 所示。

表 3-1 完成数据预处理的学生成绩挖掘表 5

XH	ZYMC	SSXCJ	WKXCJ
200803010009	临床医学	87.5	88
200803010056	临床医学	69	67
200803010085	临床医学	84	73
200803010090	临床医学	79.5	83
200803010111	临床医学	63.5	57
200803010142	临床医学	64	51
200803010156	临床医学	79	73
200803010209	临床医学	89	93
200803010448	临床医学	69	77
200803010566	临床医学	71	82

之后，按照成绩是否大于或等于 75 分进行离散化，如表 3-2 所示。

表 3-2 完成数据离散化的学生成绩挖掘表 6

XH	ZYMC	SSXCJ	WKXCJ
200803010009	临床医学	1	1
200803010056	临床医学	0	0
200803010085	临床医学	1	0
200803010090	临床医学	1	1
200803010111	临床医学	0	0
200803010142	临床医学	0	0
200803010156	临床医学	1	0
200803010209	临床医学	1	1
200803010448	临床医学	0	1
200803010566	临床医学	0	1

之后，按照 min_sup 为 0.2，min_conf 为 0.5 挖掘所有的频繁项集，如表 3-3 所示。

表 3-3 从学生成绩挖掘表 6 中挖掘出的频繁项集

频繁项集	支持度
手术学成绩=“1”	79.84%
手术学成绩=“0”	20.16%
外科学成绩=“1”	74.08%
外科学成绩=“0”	25.92%
手术学成绩=“1”，外科学成绩=“1”	68.46%
手术学成绩=“1”，外科学成绩=“0”	11.38%
手术学成绩=“0”，外科学成绩=“1”	5.62%
手术学成绩=“0”，外科学成绩=“0”	14.54%

最后，计算最终频繁项集的所有非空子集的置信度，删除小于 min_conf 的记录，确定模型，生成关联规则，如表 3-4 所示。

表 3-4 从学生成绩挖掘表 6 中挖掘出的关联规则

频繁项集	置信度
手术学成绩=“优良” \Rightarrow 外科学成绩=“优良”	92.79%

通过挖掘的关联规则，我们可以发现下述关系

学生的手术学成绩，对学生的外科学成绩有很重要的影响，手术学成绩是优良的，其外科学成绩在非常大程度上都是优良的。但手术学成绩并非绝对因素，手术学成绩是一般的学生，通过努力学习，有一些都能够获得优良的外科学成绩。

根据上述关系，我们可以得出如下结论：

手术学、外科学均属于专业课，并且手术学是学习外科学的基础。对于同一专业，相互之间有关联的先授课的专业课成绩对后授课的专业课成绩有非常大的影响。

四、总结

将之前挖掘出的规律反馈给授课的教师和听课的学生，得到他们的认同。因此，教师和学生均要注意课程之间的关联，在讲授、学习课程的时候，要及时联系、复习与之关联的前期相关课程，这样互相促进，对提高教师的教学质量和增加学生的学习效果有重要的指导意义。这样的规律，对于教学管理部门来说，在专业建设、培养方案制定、课程设置、辅助教学决策等方面具有重要的指导意义。

[责任编辑 王云江]

(下转第 124 页)

3. 音节合并

白寨乡地名中存在着音节合并的现象,例如,后里僧寨[xou²¹²liŋ⁵³tʂai²¹²]中的第二个字“里”字和第三个“僧”字音节合并为[liŋ⁵³],取“里”字的整个音节和“僧”字的韵尾[ŋ]。

4. 其他

白寨乡方言地名中还有个别特殊的音变,例如舌尖中音变舌尖前音,艾台[ŋai²¹²ts^hai⁴²(tai⁴²)]中的“台”声母本应为舌尖中音[t],却变为了舌尖前音[ts^h]。这种音变不仅在白寨乡,乃至整个河北省也是极少的现象。

五、结束语

曲周县白寨乡的地名中的语音变化包括儿化、省读、语流音变,三种变化之间是相互联系的,有时会交织在一起,表现形式较为复杂,究其原因多是受口语发音的简单省力原则导致的。当然也有个别特殊音变暂时无法解释,因此本文对白寨乡地名

只是做了初步的分析研究,还不够成熟,有些问题有待进一步研究才能弄清楚。

注释:

①田恒金,李小平.河北方言地名中的一些音变[J].语文研究,2008(2).

②曲周县人民政府网http://Avw.v.qz.x.gov.cn/E_ReadNews.asp?NewsI

参考文献:

- [1]郭锡良.汉字古音手册[K].北京:北京大学出版社,1982.
- [2]田恒金,李小平.河北方言地名中的一些音变[J].语文研究,2008.
- [3]叶蜚声,徐通锵.语言学纲要(修订版)[M].北京:北京大学出版社,2011.
- [4]韩佳,怀仁县地名的语言学考察.硕士生论文,2014.5. D=6766

[责任编辑 王云江]

Phonetic research on places in dialect of Baizhai Country in Quzhou County

ZHANG Xiao-jia, LI Rui-huan

(School of Arts, Hebei Normal University, Shijiazhuang 050024, China)

Abstract: There are some special phonetic changes in the dialect of Baizhai Country in Quzhou County. The pronunciation of these dialect differs from the local pronunciation, which is the result of Sandhi. This paper elaborates on the pronunciation and phonetics change of place names in Baizhai Country.

Key words: Quzhou; dialect place names; phonetics change

(上接第 117 页)

参考文献:

- [1]杜孝平,马秀莉,唐世涓,等.快速关联规则挖掘算法[J].计算机工程与应用,2002(11):1-4,47.
- [2]谢琦,张振兴.基于 Apriori 算法和 OLAP 的关联规则挖掘模型设计[J].计算机应用,2007,27(6):4-5,9.
- [3]周虹,马丽丽.一种改进的 Apriori 算法[J].佳木斯大学学报(自然科学版),2007,25(4):492-494.

- [4]王杰,张静,张继生,等.数据挖掘中关联规则的研究与论证[J].鞍山科技大学学报,2005,28(2):123-126.
- [5]佟强,周园春,阎保平.关联规则挖掘算法.微电子学与计算机,2005,22(6):68-72.
- [6]陆丽娜,陈亚萍,魏恒义.挖掘关联规则中 Apriori 算法的研究[J].小型微型计算机系统,2000,21(9):940-943.

[责任编辑 王云江]

An analysis of curriculum relevance on medical students' achievements by apriori algorithm

NIU Meng

(Teaching Affairs Office, Wannan Medical College, Wuhu 241002, China)

Abstract: In teaching management of institutions of higher learning, achievement management is of great importance. Analyzing the achievement data to uncover the valuable laws hidden in achievement data can help teaching decisions and have important guiding significance for such aspects as major building, establishment of training program, curriculum setting, and teaching management. This paper mainly utilizes Apriori algorithm to analyze students' achievements and uncover the mutual laws among relevant courses' achievements.

Key words: Apriori algorithm; achievement; curriculum relevance; analysis