

# 基于改进 Apriori 算法的高校课程预警规则库构建<sup>①</sup>



任 鹤, 吴 猛, 汗古丽·力提甫, 杨 勇

(新疆师范大学 计算机科学技术学院, 乌鲁木齐 830054)

通讯作者: 杨 勇, E-mail: 68523593@qq.com

**摘 要:** 在如何构建完善的课程预警规则库是高校成绩预警研究中的一个重点问题, 本文对高校学生成绩进行清洗、离散化后, 利用 Apriori 算法挖掘不及格课程之间的相关关联, 构建基础预警规则库, 在此基础上进一步挖掘“及格”, “良好”等级课程对其他课程的影响, 从而进一步扩充预警规则库. 针对大量冗余规则的情况, 在传统的支持度-置信度框架下利用提升度、兴趣度等方法筛选出强关联规则, 提高规则库的准确度, 并对挖掘出的规则进行了针对性的分析, 研究方法和结论可为教学管理提供决策支持.

**关键词:** 数据挖掘; 成绩预警; 关联规则; Apriori 算法

引用格式: 任鹤, 吴猛, 汗古丽·力提甫, 杨勇. 基于改进 Apriori 算法的高校课程预警规则库构建. 计算机系统应用, 2021, 30(7): 290-295. <http://www.c-s-a.org.cn/1003-3254/8040.html>

## Early Warning Rule Repository of University Student Achievement Based on Improved Apriori Algorithm

REN Ge, WU Meng, HANGVL Litip, YANG Yong

(School of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, China)

**Abstract:** How to build a complete early warning rule repository is a key issue in the research on the early warning of university student achievements. In this study, after cleaning and discretizing the data on university student achievements, we use the Apriori algorithm to mine the correlation between failed courses and construct the basic early warning rule repository. On this basis, the influence of courses with “pass” and “good” grades are explored to further expand the early warning rule repository. In the case of copious redundant rules, strong association rules are filtered out by lift and interest in the traditional support-confidence framework to improve the accuracy of the repository and specifically analyze the mined rules. Our methods and conclusions can support the decision-making of teaching management.

**Key words:** data mining; early warning of students' achievement; association rules; Apriori algorithm

高校学生成绩一直是直接评价学生学业水平的重要指标, 也是间接评价高校教师教学效果及高校管理水平的指标之一<sup>[1]</sup>. 随着高校学生成绩信息量的不断增加, 对学生成绩进行简单的统计、对比等信息

处理方式难以满足学生和教育工作者的实际需求, 学生成绩预警研究已成为当今高校信息管理领域研究的热点<sup>[2-5]</sup>.

成绩预警是指通过对学生课程成绩进行分析, 挖

① 基金项目: 新疆维吾尔自治区教学改革与研究项目 (ZJG2019-11); 新疆维吾尔自治区高等学校科研计划 (XJEDU2016S066); 新疆师范大学博士科研启动基金 (XJNUBS1609)

Foundation item: Teaching Reform and Research Projects in Xinjiang Uygur Autonomous Region (ZJG2019-11); Scientific Research Plan of Colleges and Universities in Xinjiang Uygur Autonomous Region (XJEDU2016S066); Start up Fund for Doctoral Research of Xinjiang Normal University (XJNUBS1609)

收稿时间: 2020-11-19; 修改时间: 2020-12-21; 采用时间: 2021-01-07; csa 在线出版时间: 2021-06-30

掘各门课之间的联系,基于已有的成绩数据,对学生未来的成绩走向进行预警。挖掘课程之间的相互联系,构建完善的预警规则库是学生成绩预警研究中的关键点<sup>[6-9]</sup>。高校课程之间具有很强的关联性<sup>[10]</sup>,这种关联性体现在专业知识体系构建以及知识传授的前后次序。挖掘深层次的课程关联关系,可以为学生成绩的预警分析提供决策支持,对高校提高教学管理水平具有重要应用价值。

关联规则挖掘<sup>[11]</sup>是数据挖掘中一种重要的数据分析方法,这其中经典的 Apriori 算法<sup>[12]</sup>可有效地从某一任务的事务集中挖掘出客户感兴趣且富有价值的隐藏规则。近年来,将 Apriori 算法应用在高校成绩预警分析中成为一个研究热点,何楚等<sup>[13]</sup>采用 Apriori 算法和 FP-growth 算法挖掘学生成绩的频繁规则并结合基于图论的谱聚类算法对规则中涉及的公共课与专业进行分类,提出基于课程关联分类模型的学生成绩预测,研究方法新颖,但仅采用支持度与置信度进行规则挖掘,没有剔除某些有较强干扰性的规则;郭鹏等<sup>[14]</sup>在对学生成绩进行聚类离散化后,再引入兴趣度挖掘课程间的关联规则,其侧重点在于不同等级成绩之间的课程规则挖掘,而非针对不及格课程之间的规则挖掘;王华等<sup>[15]</sup>为减少 Apriori 算法的 I/O 负载,利用交集操作记录频繁项集从而降低时间代价,不足点在于受限 Matlab 平台;刘华婷等<sup>[16]</sup>针对 Apriori 算法的效率问题提出采用线性表进行数据存储,利用 Apriori 算法中的超集定理判断非频繁项集,从而在扫描速度上进行优化,加快算法的时间效率,遗憾的是该研究仅在单一且规模较小的数据集上进行了效率测试;袁路妍等<sup>[17]</sup>通过减少 Apriori 算法中挖掘频繁项集的比较次数,以提升算法的效率,该法所挖掘出的规则在度量上仍只采用了支持度与置信度;赵峰等<sup>[18]</sup>借鉴分治算法的思想,提出将事务集中的课程进行预分类,再得到各自类别的频繁项集进行规则挖掘,虽避免了扫描整个事务集,但仍无法保证规则的有效性。

综上,Apriori 算法在高校成绩预警应用中的基本思想:一是对课程或学生成绩本身进行类别处理,再通过引入一种新的参数对挖掘出的规则进行有效性度量;二是针对 Apriori 算法自身存在的运行效率缺陷提出改进,加强 Apriori 算法的适用性。对此,本研究在传统的 Apriori 算法之上提出利用提升度筛选出彼此关联程度较高的课程规则,再引入兴趣度针对预警规则库

所需求的课程规则进行筛选,最终得到兴趣度较高的关联规则构建课程预警规则库,相比较传统 Apriori 算法,本研究提出的算法能大大降低了冗余规则数量,缩短算法运行时间,提高规则挖掘的效率。

## 1 高校课程预警规则分析方法

### 1.1 相关定义

定义 1. 项。

定义集合  $I = \{i_1, i_2, \dots, i_m\}$  表示由  $m$  个元素组成的一个有限集合。设相关问题数据集  $D$  是本问题的事务集合:  $D = \{T_1, T_2, \dots, T_n\}$ , 它是由  $n$  事务  $T$  组成。现规定  $D$  集中某一元素  $T_i$  由一或多个  $I$  集中的元素  $i_j$  构成, 即  $T \subseteq I$  且  $T \neq \emptyset, i = 1, 2, \dots, n; j = 1, 2, \dots, m$ 。

定义 2. 候选项集。

由一或多个项组成的集合  $\{i_1, i_2, \dots, i_k\}$  称为候选项集, 包含  $k$  个项的候选项集称为  $k$  候选项集, 如  $A = \{\text{面向对象程序设计 (及格)}, \text{软件工程 (及格)}\}$  是一个 2—候选项集, 且  $A \subseteq T$ 。

定义 3. 频繁项集  $L_k$ 。

若该候选项集  $A$  的绝对支持度  $support(A)$  满足算法中预设定的最小支持度阈值  $sup_{min}$ , 也即  $support(A) \geq sup_{min}$  时, 则称该项集  $A$  为频繁项集, 包若  $A$  中含  $k$  项则称为频繁  $k$  项集  $L_k$ ; 当时  $k \geq 2$ , 可挖掘  $L_k$  中不同项之间的关联规则。

定义 4. 关联规则。

本算法中所有的关联规则都是由频繁  $k$ -项集  $L_k$  得到的, 例如频繁 2—项集  $A = \{\text{面向对象程序设计 (及格)}, \text{软件工程 (及格)}\}$  同时又满足 Apriori 算法的最小置信度阈值  $conf_{min}$ , 则  $A$  可产生一条由蕴含式表示的规则:  $\{\text{面向对象程序设计 (及格)} \Rightarrow \text{软件工程 (及格)}\}$ , 其中“面向对象程序设计 (及格)”为规则的前件, “软件工程 (及格)”为规则的后件。规则的支持度计算公式如下:

$$support(X \Rightarrow Y) = count(X \cup Y) / count(D) = p(X \cup Y) \quad (1)$$

其中,  $count(X \cup Y)$  表示  $X$  与  $Y$  的并集在事务集  $D$  中的计数,  $count(D)$  表示事务集  $D$  中事务数。规则的置信度计算公式如下:

$$confidence(X \Rightarrow Y) = p(X \cup Y) / p(X) = p(Y|X) \quad (2)$$

支持度与置信度是对规则筛选最重要的两种度量, 规则的支持度反映所发现规则的价值性及在事务集中出现的频率, 置信度则反映了规则的确定性<sup>[19]</sup>。

### 1.2 传统的 Apriori 算法课程预警规则挖掘算法

传统的 Apriori 算法课程预警关联规则挖掘利用逐层迭代的搜索方法, 即由生成的频繁  $k$ —项集确定频繁  $(k+1)$ 项集的思想. 传统的 Apriori 算法课程预警规则挖掘的算法流程如图 1, 其主要步骤为两个阶段.

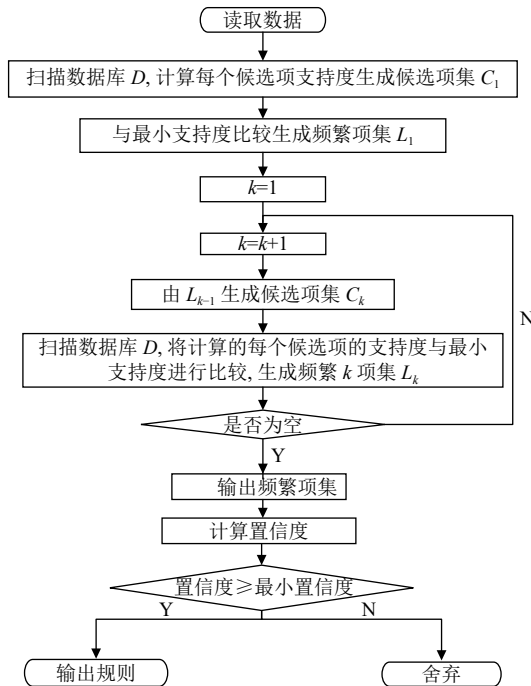


图 1 课程预警关联规则挖掘算法

第 1 阶段, 算法执行后扫描事务集  $D$ , 记录每个项的频数, 并根据最小支持度阈值筛选出符合要求的候选频繁项集, 记为  $L_1$ ; 然后利用  $L_1$  找出满足最小支持度阈值频繁 2-项集, 记为  $L_2$ , 如此反复迭代, 直到没有满足最小支持度阈值的频繁  $k$ -项集.

第 2 阶段, 再次扫描所有的频繁  $k$ -项集,  $k = 1, 2, \dots$ ; 利用第 1 阶段中得到的支持度数据计算每一个频繁  $k$ -项集的置信度, 将满足最小置信度阈值的频繁  $k$ -项集挖掘筛选出符合课程预警的规则, 进行构建规则库.

### 1.3 改进的 Apriori 课程预警关联规则算法

#### 1.3.1 增强冗余规则筛选机制

传统的关联规则挖掘算法使用支持度-置信度框架作为筛选依据, 但支持度-置信度框架存在一定的缺陷. 其中支持度的缺点是显而易见的, 在事务集中会有许多潜在的有意义的模式由于其频繁项集  $L_k$  支持度小而被忽略或者产生一些冗余的规则; 而置信度的缺陷却比较特殊, 它主要在于忽视了规则后件的项集在

事务集中的支持度. 因此, 通过传统的 Apriori 算法生成的预警规则库会出现很多无用、冗余的关联规则, 对成绩预警分析产生干扰, 降低预测的准确性, 因而需要对预警关联规则作进一步地度量分析, 从而筛选出合适的预警规则. 对此, 本研究在支持度-置信度框架基础上引入提升度 Lift<sup>[20]</sup>、兴趣度 Interest<sup>[21]</sup> 对挖掘关联规则算法进行改进.

提升度方面, 某一频繁  $k$  项集生成的某规则  $\{X \Rightarrow Y\}$  的提升度被定义为:

$$Lift(X \Rightarrow Y) = \frac{p(X \cup Y)}{p(X)p(Y)} = \frac{confidence(X \Rightarrow Y)}{support(Y)} \quad (3)$$

当该条规则的提升度值小于 1 时表示  $X$  和  $Y$  是负相关的, 一个出现可能导致另一个不出现; 当值  $Lift$  近似等于 1 时表示  $X$  和  $Y$  是独立的, 彼此之间无相关性; 当  $Lift$  值大于 1 时表示  $X$  和  $Y$  是正相关的, 其中一个的出现会蕴涵另一个的出现. 提升度越高则表示由规则前件主导与规则后件共同出现的程度越高, 反之提升度越小时则二者互斥的程度越高.

兴趣度方面, 本研究引入基于差异思想的兴趣度<sup>[22,23]</sup>. 某一频繁  $k$  项集生成的某规则  $\{X \Rightarrow Y\}$  的兴趣度的定义如下: 由该条规则  $\{X \Rightarrow Y\}$  的置信度与规则后件的支持度之间的差值作为分子; 分母则取该条规则  $\{X \Rightarrow Y\}$  的置信度与规则后件支持度的最大值, 也即引入了一个标准化因子, 整体上使得该条规则的兴趣度绝对值小于 1.

区别于提升度筛选前后件关联程度较高的规则, 兴趣度可对某频繁  $k$ -项集产生的正反面规则筛选得到用户更感兴趣的规则. 如对于某一频繁 2-项集  $A = \{\text{面向对象程序设计 (及格)、软件工程 (及格)}\}$ , 由于前后件的不同可产生两条规则:  $\{\text{面向对象程序设计 (及格)} \Rightarrow \text{软件工程 (及格)}\}$ 、 $\{\text{软件工程 (及格)} \Rightarrow \text{面向对象程序设计 (及格)}\}$ , 根据前文兴趣度的定义, 二者的兴趣度值显然不一; 那么, 通过兴趣度阈值就可筛选出以“面向对象程序设计 (及格)”为前件的预警规则. 引入兴趣度的目的在于可选择某些不及格的重点课程作为前件, 挖掘出相关预警规则, 从而摒弃某些兴趣度不高的规则.

#### 1.3.2 计算候选集频数优化策略

在 Apriori 算法中, 由频繁项集  $L_{k-1}$  自连接得到候选集  $C_k$  的过程中, 存在两大缺陷: 一是由于  $L_{k-1}$  中存在很多不必作自连接的项集, 导致产生冗余  $C_k$ ; 二是需要

多次扫描数据库确定候选项集支持度<sup>[24]</sup>. 针对上述问题, 提出一种计算候选集频数优化策略.

首先通过矩阵运算的方法获得频繁一项集 $L_1$ , 并对 $L_1$ 进行预剪枝得到 $L_1$ , 然后利用 Apriori 算法的自连接性质对 $L_1$ , 进行操作生成候选二项集 $C_2$ . 在计算候选二项集 $C_2$ 的支持度计数之前, 首先要根据 $L_1$ 明确 $I_x$ 和 $I_y$ 中的最小项, 接下来在搜索频繁项集时候只需要扫描特定分支即含有最小项事务结点链而不是整个数据库, 并结合该节点处标记 $flag_i = (x, y)$ 中 $y$ 的数值, 即可得到 $C_2$ 的计数, 最后对 $C_2$ 进行剪枝得到频繁二项集 $L_2$ , 对获得的频繁二项集 $L_2$ 作同样的操作就可以快速计算出 $C_3$ 支持度计数, 节省了空间和时间资源.

### 1.3.3 基于矩阵的存储方式改进

将事务数据库 $D$ 转换为矩阵, 每一事务以矩阵的行来表示行纬度, 事务中的项用列表示列纬度, 当要对数据进行挖掘时就可以直接调用转换为矩阵中的信息, 在行列指定的位置处若存在数据则为1, 否则记为0, 删除事务长度小于最小置信度的行与列, 免去了对 $D$ 再次扫描的过程, 节省了I/O的开销与时间.

改进 Apriori 的课程预警关联规则算法伪代码如算法1.

算法1. 改进的课程预警关联规则算法

输入: 事务集 $D$ , 最小支持度,  $sup_{min}$  最小置信度 $conf_{min}$ , 最小提升度 $Lift_{min}$ , 最小兴趣度 $Interest_{min}$   
输出: 预警规则列表 $R$

- 1) 扫描事务集 $D$ , 生成候选项集矩阵 $C_1$ 并计算其  $support$ ;
- 2) 将候选项集 $C_1$ 的  $support$  与 $sup_{min}$ 进行比较, 生成频繁项集 $L_1$ ;
- 3) for( $k=2; L_{k-1} \neq \emptyset; k++$ ):
- 4) 剪枝并生成新的候选项集 $C_k$ ;
- 5) 再次扫描事务集 $D$ , 计算 $C_k$ 的  $support$ ;
- 6) 将 $C_k$ 的  $support$  与 $sup_{min}$ 比较, 生成频繁项 $L_k$ ;
- 7) 输出频繁项集列表 $L$ ;
- 8) for( $k=2; k \leq \text{len}(L); k++$ ): //遍历频繁项集列表 $L$
- 9) 提取每一个 $L_k$ 并计算置信度; 并与 $conf_{min}$ 比较后筛选;
- 10) 计算每一个 $L_k$ 的提升度  $Lift$ , 与 $Lift_{min}$ 比较后筛选;
- 11) 计算每一个 $L_k$ 的兴趣度  $Interest$ , 与 $Interest_{min}$ 比较后筛选;
- 12) 由筛选后的 $L_k$ 生成预警规则 $R_k$ ;
- 13) 输出预警规则列表 $R$ .

## 2 实验结果与分析

### 2.1 实验环境与数据来源

本研究的实验硬件环境为 Intel Core i7 处理器、16 GB 内存, 1 TB 硬盘; 操作系统选用 Mac OS 系统; 选用 Python 3.7 并在 Jupyter notebook 开发环境上对实

验进行编程实现.

实验数据来源于新疆师范大学计算机科学技术学院毕业生的19门基础课与核心专业课的成绩, 共3000条记录, 部分原始数据如表1所示.

表1 部分原始数据

姓名	数据结构	计算机组成原理	Web技术与应用	计算机网络
贾**	75.4	60.5	86.5	62.1
蔡**	69.7	76.5	83.9	80.4
白**	76.8	81.0	81.2	81.1
岳**	75.1	81.0	74.0	84.6
纪**	76.0	89.5	87.6	88.8
郭*	46.4	78.5	60.2	35.6
刘**	67.4	77.8	86.8	81.1
伏**	71.2	74.5	67.2	66.4
张**	89.7	71.5	95.8	81.1
王**	80.0	84.8	82.7	79.7
高*	75.2	71.5	91.8	63.6
戴*	75.2	63.5	80.9	75.4
马*	68.4	76.5	81.2	78.3

### 2.2 数据的清洗

由于部分原始数据中出现了不同专业间相关开设的课程名称前后不一致的现象, 必须对其进行一致化处理: 如高等数学(1)与高等数学(I)统称为高等数学一、面向对象程序设计(Java)与面向对象程序设计统称为面向对象程序设计, 计算机网络与计算机网络技术统称为计算机网络.

本研究主要挖掘是专业课程的相关关系, 特别是不同学期开始的专业基础课与核心专业课程之间的联系, 可将原始数据中与专业无关的公共类、博雅类课程进行剔除, 仅选取如高等数学、数据结构、Web技术与应用等19门学科课程作为研究对象.

对于某些学生成绩大面积缺失的, 则采取舍弃该样本, 若学生出现补考成绩的需进行成绩的一致性处理. 对于某些学生出现某科科目成绩出现空值的情形, 则采用该学科的均值成绩填充.

### 2.3 数据的离散化

课程成绩本身也就是一种离散化数据, 若将课程成绩的分值直接作为 Apriori 算法的事务集进行输入, 其挖掘出来的规则适用性较低且模型效率低. 由于考试方式、考试难度、评卷的标准存在差异性, 课程成绩的分值有可能相对集中在某一个小区间内, 而非在各个分值区间中呈现出相似的密度. 本研究针对事务集中550条记录, 决定采用基于单值的划分方法, 设定

课程成绩区间,划分课程成绩等级,以便于更好的挖掘课程之间的联系.所得的成绩等级划分表如表2.

表2 成绩等级划分表

等级	不及格	及格	良好	优秀
分数区间	[0,60)	[60,75)	[75,85)	[85,100]

### 2.4 预警规则挖掘对比实验及分析

为验证改进算法的效率,分别采用传统 Apriori 算法, MIFP-Apriori<sup>[24]</sup>, RCM-Apriori<sup>[25]</sup>, 与本文提出的算法进行预警规则的挖掘,对效率进行对比.在事务集为 3000 的数据集上,选用最小值支持度阈值 0.1,最小置信度阈值 0.4 的提升度阈值 1.4、兴趣度阈值 0.1 执行算法实验,4 种算法所挖掘出的无用冗余规则条数和运行时间见表3,由表3可以看出本文提出的改进的 Apriori 算法可以大大降低冗余规则条数,算法运行时间也得到缩短,提高了预警规则的挖掘效率.

表3 预警规则挖掘对比实验

算法	冗余规则条数	时间(ms)
传统Apriori算法	2510	4500
MIFP-Apriori算法	2034	4200
RCM-Apriori算法	1872	4000
本文算法	1324	3200

为进一步测试在不同事务集大小下的各算法执行效率,将事务集按照 500~3000 进行递增,考察不同数据规模下算法的运行时间,实验结果如图2所示.

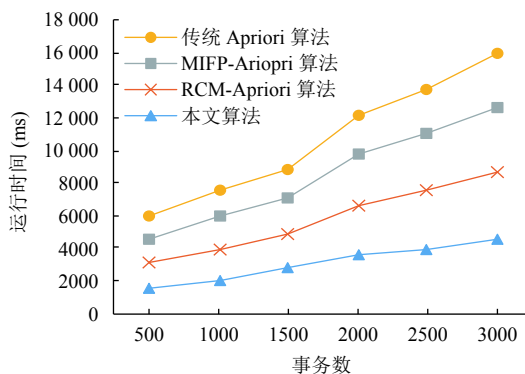


图2 不同数据规模下4种算法性能比较

由图2可以看出,当数据规模较小时,4种算法的执行时间相差不大.当数据规模逐渐增加时,传统的 Apriori 算法斜率增加比较明显,执行时间最长,本文提出的算法在候选项集生成,冗余规则筛选,存储方式等方面进行了优化,曲线斜率比其他3种算法小,运行时

间最短,并且性能稳定.

本文提出的改进的 Apriori 算法在剔除掉无关的 1324 条冗余规则后,共挖掘到有效规则 430 条.部分规则如表4所示.

表4 部分课程预警规则表

强关联规则	置信度	提升度	兴趣度
{'高等数学二(不及格)}⇒{'概率与数理统计(不及格)}	0.4320	1.9414	0.4849
{'网页设计基础(不及格)}⇒{'Web技术与应用(不及格)}	0.4347	1.8249	0.4520
{'数据结构(不及格)}⇒{'Android程序开发(不及格)}	0.4368	1.6791	0.4044
{'离散数学(不及格)}⇒{'数据结构(不及格)}	0.5352	1.6575	0.3967
{'程序设计基础(不及格)}⇒{'Android程序开发(不及格)}	0.4266	1.6398	0.3901
{'程序设计基础(不及格)}⇒{'数据结构(不及格)}	0.5733	1.7756	0.4368
{'程序设计基础(不及格)}⇒{'网页设计基础(不及格)}	0.4266	1.4794	0.3240
{'网页设计基础(不及格)}⇒{'Web技术与应用(不及格)}	0.4347	1.8249	0.4520

表4中第2条规则{网页设计基础(不及格)}⇒{'Web技术与应用(不及格)}揭示了Web技术与应用的与先导课网页设计基础存在紧密联系,若学生出现《网页设计基础》课程不及格时,那么很有可能在学习《Web技术与应用》课程时成绩同样会不及格.再如第4条规则{离散数学(不及格)}⇒{'数据结构(不及格)},离散数学作为数据结构的重要基础之一,一旦学生出现离散数学不及格的情况,则有大概率发生数据结构同样不及格的情况.

### 3 结束语

高校课程预警规则库可以为学生成绩预警提供决策支持,本文对学生成绩进行预处理后,采用 Apriori 算法,重点挖掘不及格课程之间的关联关系,构建基础预警规则库,挖掘强关联关系的课程,扩充基础预警规则库,采用多种方式对冗余规则进行过滤,确保预警规则库的有效性.挖掘出的课程预警规则,可为课程成绩的预测,高校教学方案的设计提供数据支撑,有助于提升高校的教学质量.

#### 参考文献

1 陆冷飞. 高校成绩分析模型及应用研究. 中国教育信息

- 化·高教职教, 2017, 9(5): 62-65.
- 2 梁啸. 基于数据挖掘的高校学生成绩预警技术的研究 [博士学位论文]. 武汉: 武汉理工大学, 2014.
  - 3 吴文玲. 基于数据挖掘技术的课程相关性分析及其应用研究 [硕士学位论文]. 成都: 四川师范大学, 2018.
  - 4 Wu XD, Zeng YZ. Using Apriori algorithm on students' performance data for association rules mining. Proceedings of 2nd International Seminar on Education Research and Social Science. Kuala Lumpur, Malaysia. 2019. 403-406.
  - 5 吴蓓. 基于决策树算法的成绩预测模型研究及应用 [硕士学位论文]. 西安: 西安理工大学, 2019.
  - 6 张甜, 尹长川, 潘林, 等. 基于改进的聚类和关联规则挖掘的学生成绩分析. 北京邮电大学学报(社会科学版), 2018, 20(2): 91-96.
  - 7 都娟, 翟社平. 基于改进 Apriori 算法的高校学生成绩预警系统的研究. 现代信息科技, 2018, 2(2): 108-109, 112. [doi: 10.3969/j.issn.2096-4706.2018.02.037]
  - 8 许琦, 祖燕飞. 我国高校大学生学习预警援助机制的建构及其作用. 科技信息, 2009, (31): 145, 189.
  - 9 刘娜. 基于数据分析的高校教学管理决策支持方法研究 [硕士学位论文]. 杭州: 浙江大学, 2019.
  - 10 谢雨婷. 基于学生成绩的课程关联性和学生综合素质评价研究 [硕士学位论文]. 武汉: 华中师范大学, 2019.
  - 11 崔妍, 包志强. 关联规则挖掘综述. 计算机应用研究, 2016, 33(2): 330-334. [doi: 10.3969/j.issn.1001-3695.2016.02.002]
  - 12 陆丽娜, 陈亚萍, 魏恒义, 等. 挖掘关联规则中 Apriori 算法的研究. 小型微型计算机系统, 2000, 21(9): 940-943. [doi: 10.3969/j.issn.1000-1220.2000.09.013]
  - 13 何楚, 宋健, 卓桐. 基于频繁模式谱聚类的课程关联分类模型和学生成绩预测算法研究. 计算机应用研究, 2015, 32(10): 2930-2933. [doi: 10.3969/j.issn.1001-3695.2015.10.011]
  - 14 郭鹏, 蔡骋. 基于聚类和关联算法的学生成绩挖掘与分析. 计算机工程与应用, 2019, 55(17): 169-179. [doi: 10.3778/j.issn.1002-8331.1902-0223]
  - 15 王华, 刘萍. 改进的关联规则算法在学生成绩预警中的应用. 计算机工程与设计, 2015, 36(3): 679-682, 752.
  - 16 刘华婷, 郭仁祥, 姜浩. 关联规则挖掘 Apriori 算法的研究与改进. 计算机应用与软件, 2009, 26(1): 146-149. [doi: 10.3969/j.issn.1000-386X.2009.01.054]
  - 17 袁路妍, 李峰. 改进的关联规则 Apriori 算法在课程成绩分析中的应用. 中国教育信息化·高教职教, 2017, (9): 62-65. [doi: 10.3969/j.issn.1673-8454.2017.09.017]
  - 18 赵峰, 刘博妍. 基于改进 Apriori 算法的大学生成绩关联分析. 齐齐哈尔大学学报(自然科学版), 2018, 34(1): 11-15.
  - 19 Han JW, Kamber M. 数据挖掘: 概念与技术. 范明, 孟小峰, 译. 北京: 机械工业出版社, 2012. 157-176.
  - 20 毕建欣, 张岐山. 关联规则挖掘算法综述. 中国工程科学, 2005, 7(4): 88-94. [doi: 10.3969/j.issn.1009-1742.2005.04.016]
  - 21 周欣, 沙朝锋, 朱扬勇, 等. 兴趣度—关联规则的又一个阈值. 计算机研究与发展, 2000, 37(5): 627-633.
  - 22 张玉芳, 熊忠阳, 彭燕, 等. 基于兴趣度含正负项目的关联规则挖掘方法. 电子科技大学学报, 2010, 39(3): 407-411. [doi: 10.3969/j.issn.1001-0548.2010.03.018]
  - 23 周皓峰, 朱扬勇, 施伯乐. 一个基于兴趣度的关联规则采掘算法. 计算机研究与发展, 2002, 39(4): 450-457.
  - 24 曾子贤, 巩青歌, 张俊. 改进的关联规则挖掘算法——MIFP-Apriori 算法. 科学技术与工程, 2019, 19(16): 216-220. [doi: 10.3969/j.issn.1671-1815.2019.16.033]
  - 25 廖纪勇, 吴晟, 刘爱莲. 基于布尔矩阵约简的 Apriori 算法改进研究. 计算机工程与科学, 2019, 41(12): 2231-2238. [doi: 10.3969/j.issn.1007-130X.2019.12.019]