教学数据的关联挖掘算法[[1]](#footnote-1)\*

阮 帅

(北京航空航天大学 计算学院,北京 100191)

通讯作者: 阮帅, E-mail: @buaa.edu.cn

摘 要: 随着大数据时代的到来,数据挖掘的需求越来越广泛,其应用领域也越来越广泛.近年来,随着人们对教育愈发的重视,教学数据不断积累,如何从教学信息中挖掘出有用的信息成为一个研究热点.本文首先简单介绍了教育信息的数据挖掘工作,并指出传统关联挖掘算法在挖掘教学信息上的缺陷.接着,提出关联模式数据挖掘算法,并通过真实的教学数据进行模拟验证.实验验证表明:新提出的教学数据挖掘算法能够较为准确的挖掘出两门课程之间的关联关系。

关键词: 数据挖掘;教学数据;关联模式;数据验证

中图法分类号: TP311

随着大数据时代的到来,每时每刻均会产生数量巨大的数据.在这些数据中,教学数据占据了很大比例.如何在如此多且纷杂的教学数据中挖掘出有价值的信息,用以优化教育服务成为了数据挖掘方面的一个热点.不少研究者将数据研究算法与教学数据结合,进行了教学数据的挖掘探索.

在传统教学中,Snajeev[1]首次将数据挖掘应用到大学数据库中挖掘满足特定模式的人群,最后将实验的结果作为学校制定政策的参考.Dillon[2]利用多种聚类方法将1278名学生聚类,根据聚类的结果分析学生内在的驱动力以及以后从事研究的潜力.Becker[3]利用数据挖掘技术探究课程的改革是否对学生的学习或者其他方面产生了影响,他们通过关联规则和分类等挖掘方法对课程改革的结果进行评估.Ma[4]在关联规则的基础上通过打分函数对所有的学生进行打分,然后选择潜在落后的学生,并向他们推荐课程.M Ramaswami[5]利用1000条数据对730名学生的表现行为进行建模,并且将模型应用到学生的行为以及表现的预测上,以便对表现不好的学生给予建议.Luan[6]利用聚类以及决策树的算法帮助学校管理者更好的分配学习资源.HaiyunBian[7]分析了针对教学数据中不同的聚类算法的优缺点.Xin[8]通过给予规则权重对关联规则挖掘方法进行改进,并将改进后的算法应用到教学数据中,实验表明改进的算法发现了更多有价值的规则.Yadav[9]将决策树用于48名学生以往的数据上建立模型,然后将模型用于预测其他学生的变现.SinghC[10]通过将聚类和关联规则挖掘算法用于学生对教学的反馈信息中,发现教师表现以及行为的潜在倾向性.González-Brenes[11]通过将隐马尔科模型用于128名学生的数据对学生进行建模,以便发现学生对于某项技能的熟练程度.Toscher A[12]将改进协同过滤算法用于对学生问答问题可能性的预测上.

针对在线学习系统的数据挖掘也已经存在了很多的研究成果,当前对于在线学习中教学数据的挖掘主要是针对日志进行挖掘.如梁燕红[13]利用Web日志挖掘技术对在线教学数据进行分析,通过对知识点网页访问的频度进行统计,以此确定学生兴趣点以及对应的知识难点,通过这些分析可以有效的调整和改进教学策略.卢朝晖等[14]提出了基于Web的教学反馈系统,通过挖掘学生用户日志中的联系,挖掘出更多的信息和规则,以便加强对学生的了解.Bedi[15]分析了在线学习平台的发展.[Chellatamilan](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=p_Authors:.QT.Chellatamilan,%20T..QT.&newsearch=true)[16]讨论了如何通过对于在线学习平台数据的分析,更好的向用户推荐课程以及分配资源.Zaı¨ane[17]针对推荐系统的建立和优化进行了研究.Klašnja-Milićević[18]构建了一个能够根据用户兴趣自动调整的推荐系统.AntonenkoPD[19]将聚类分析应用到在线学习系统中,通过对学生行为的特征进行抽取,利用k-means和层次聚类算法对学生在解决问题时的行为进行分析.Le Blanc[20]利用聚类算法做数据的前期处理,并且将聚类的结果用于学生群组分类.Hamalainen[21]通过将学生进行聚类,根据聚类发现的不同学生团体,学校将给予不同的指导.JovanovicM[22]利用在线系统的数据对学生进行聚类以及成绩预测.Lu[23]等通过关联规则挖掘算法挖掘在线学习系统的日志,根据挖掘出的关联规则主动引导学生的学习活动,并且推荐相关的资料.Hwang[24]通过关联关系挖掘算法挖掘学生学习中的问题,并且给予相应的建议.García E[25]实现了一个关联关系挖掘算法为基础的工具,并且利用该工具分析在线学习系统的数据,进而引导教学改进以及课程.C Romero[26]利用以往大约600名学生的数据预测特定的一门课程的成绩.Kim[27]通过文本挖掘的方法对在线学习系统论坛中讨论的主题进行分类.

尽管已经存在了如此多对于教学数据挖掘的研究,但是并没有人针对教学数据中的课程间关系进行研究,本文将主要致力于通过分析两门学科之间的学生成绩挖掘这两门课程间的关联关系.本文提出上升与下降模式挖掘算法与正常规则与异常规则挖掘算法.上升与下降模式挖掘算法主要针对有等级的事务,挖掘出它们等级上升或者下降的关系.在得到其上升或者下降关系之后,可以通过正常与异常规则可以挖掘出两门课程之间的异常规则集合,挖掘算法挖掘出的异常规则集合之中包含的异常规则越多,则说明二者关联度越小.最后通过数学公式进行规范,可以清晰地的表示出两课程之间关联关系的强弱.

本文的第1节主要对传统的关联关系（association rules）挖掘的主要缺陷进行分析.第2节针对关联关系挖掘的主要缺陷提出一种新的挖掘关联模式挖掘算法.第3节利用真实的教学数据进行检验.第4节总结全文.

# 传统关联关系（association rules）挖掘缺陷

传统的关联规则挖掘分为两个步骤:第一步是频繁项集挖掘;第二步是强关联规则挖掘.频繁项集挖掘的主要算法包括Aprori[28]算法和FP－Tree[29]算法:Apriori算法通过将上层频繁项集进行连接来获取下层频繁项集.算法需要产生大量的候选项集合,然后通过支持度阈值将不符合阈值的项删除掉,由于每次迭代都需要计算规则对应的支持度,所以需要对数据进行多次扫描,当频繁项集比较多的情况下,需要的时间复杂度比较高,为了解决这个问题提出了FP-Tree算法,FPTree算法是在不生成候选项的情况下,完成频繁项集的挖掘.强关联规则挖掘是根据挖掘出的所有的频繁项集产生出对应的强关联规则.

关联关系挖掘算法能够挖掘出符合支持度阈值以及置信度阈值的所有规则的集合,但是挖掘出的规则集合是”杂乱”的,以至于很难直接从挖掘出的关联规则中发现有价值的信息,例如对表1 中的带有等级性的数据（A,B,C代表的是等级）进行关联规则挖掘.设置支持度阈值为10%,置信度阈值为10%,挖掘出来的关联规则如表2.

表1 事务数据

|  |  |  |
| --- | --- | --- |
| 事务1 | 事务2 | 数目 |
| A | B | 5 |
| B | C | 20 |
| C | B | 10 |
| A | A | 35 |
| D | C | 20 |
| B | D | 10 |

表2 关联规则

|  |  |  |  |
| --- | --- | --- | --- |
| 规则标号 | 规则 | 支持度 | 置信度 |
| 1 | A🡪A | 35% | 87.5% |
| 2 | D🡪C | 10% | 100% |
| 3 | B🡪C | 20% | 66% |
| 4 | B🡪D | 10% | 33% |
| 5 | C🡪B | 20% | 100% |

由表2中的关联规则,支持度以及置信度,可以得到如下信息:事务1为A,事务2为A对应的规则出现的次数占整个规则集合的35%,并且如果事务1为A可以认为有87.5%可能性认为事务2也是A等级,这些关联规则可以传递一些信息.例如可以知道事务位于某个等级的情况下与另外一个事务处于的等级的最大可能性.如果关联规则中存在上万条这样的规则,用户将很难直观的得到想要的数据.

本文针对上面问题提出关联模式挖掘概念以及挖掘算法,并且进一步提出关联模式挖掘的应用.

# 关联模式挖掘算法

## 基础概念说明

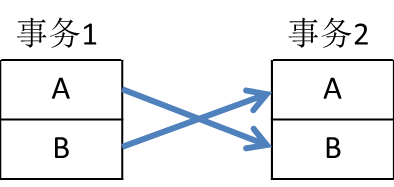


图1规则交叉图

事务等级:如表1中数据,事务中的数据为AB,C,D.这个数据可以认为是事务的等级,并且定义等级大小为A>B>C>D.

关联规则:形如”“的式子,左边代表的是事务1的等级,右边代表的是事务2的等级.例如表2中每个标号对应的记录可以认为是一个规则.

关联规则趋势:如果规则的右侧对应的等级高于规则左侧对应的等级,那么认为规则是上升趋势,例如对应的就是上升趋势,反之为下降趋势.

关联规则支持度:设规则X属于规则集合,规则X的支持度=规则集合中包含X的条数/数据规则的总条数;支持度代表的规则X出现的频率,例如规则集合中存在1000规则,其中含有X的规则有100条,那么规则X的支持度是100/1000=0.1.

支持度阈值:用户定义的支持度的最小值.

规则交叉:两条规则对应的级别是交叉的,两条规则就是交叉的.如图1所示事务1与事物2存在两条关联规则分别为和这两个规则是交叉的.

正常规则集合:规则集合中不存在交叉点,整体的规则存在一定的趋势性.

异常规则集合:按照一定的约束,不断将规则集合中引起交叉的规则剔除并添加到新的规则集合中,直到剩余的规则集合变成正常规则集合,这个新的规则集合我们称为异常规则集合.

规则等级差:按照(1)中给出的等级的大小,按照A对应5,B对应4,C对应3,D对应2,E对应1的映射方式将规则左右等级映射为数字,等级差代表的是左侧等级对应的数值与右侧等级对应的数值的差.

规则集合趋势:如果规则集合下降趋势规则的支持度和大于上升趋势规则的支持度和,那么该规则集合趋势为下降,相反为上升.

## 关联模式概念说明

基于关联分析算法挖掘出关联规则存在的缺陷,即产生大量的并且离散型细节规则,对规则进行整合再挖掘,将规则整合到特定的模式下,这些模式能够向用户提供更清晰更有价值的信息,这些模式我们称为关联模式.关联模式是一类约束,通过约束将关联规则集合划分为满足特定约束的规则集合以及不满足特定约束的规则集合.

本文定义的关联模式包括如下两种模式:

上升与下降模式:该模式反映规则集合中规则的趋势性,上升模式对应的规则趋势都是上升,下降模式对应的规则趋势都是下降的.

正常与异常模式:该模式按照基础概念定义中对正常规则集合和异常规则集合的定义将原始的规则集合划分为正常规则集合和异常规则集合.

## 关联模式挖掘算法

### 关联规则挖掘

针对带有等级的数据利用Aprioir算法获得满足最小支持度的阈值的频繁项集合,由频繁项集产生所有满足最小置信度的关联规则集合.

### 上升与下降模式挖掘算法

#### 算法思路

通过计算关联规则集合中所有规则对应的趋势,按照规则不同的趋势将其加入到上升规则集合中或者下降规则集合中.

#### 算法示意图

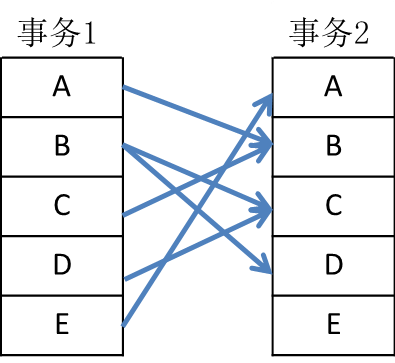


图2 规则图

图2为规则图,图中每一个连线代表的是一条规则,图中左侧和右侧的表格中的字符代表的是数据等级,左侧和右侧分别代表两个事务的数据.图3（左）为算法挖掘出的下降规则集合,图3（右）为算法挖掘出的上升规则集合.

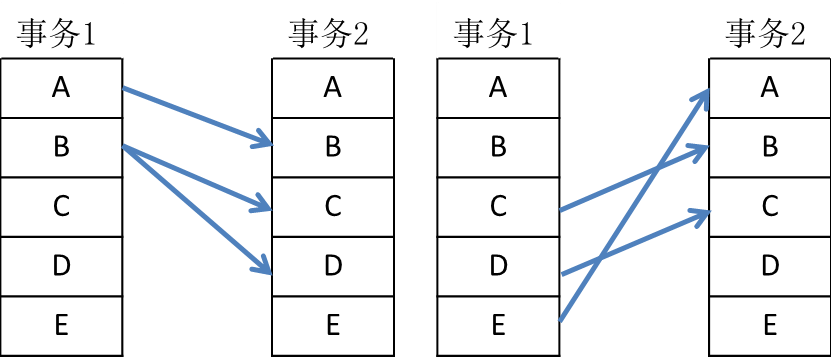


图3 上升模式与下降模式

#### 算法流程



图4 算法流程图

#### 算法伪代码

|  |
| --- |
| 算法1:上升与下降模式挖掘算法 |
| 输入:关联规则集合  输出:上升规则集合,下降规则集合   |  |  | | --- | --- | | 1 | ascendRuleSet={},descendRuleSet={} | | 2 | while IsEmpty(RuleSet): #判断规则集合是否为空 | | 3 | rule=GetOneRule(RuleSet) | | 4 | leftRuleLevel=GetLeftRuleLevel(rule) #计算规则左边数据对应的等级 | | 5 | rightRuleLevel=GetRightRuleLevel(rule) #计算规则右边数据对应的等级 | | 6 | if leftRuleLevel-rightRuleLevel>0: | | 7 | AddRule(rule,ascendRuleSet) #将规则加入到特定规则集合中 | | 8 | end if | | 9 | if leftRuleLevel-rightRuleLevel<0: | | 10 | AddRule(rule,descendRuleSet) | | 11 | end if | | 12 | end while | | 13 | rutern ascendRuleSet,descendRuleSet | |

图5 算法伪代码

### 正常与异常挖掘模式算法

#### 算法思路

算法通过启发式的思想将规则集合划分为正常规则集合和异常的规则集合.启发式的方法主要分为以下几个筛选异常规则的方法:选择交叉点最多的规则,选择等级差最大的规则,选择与大趋势相反的规则,选择支持度小的规则.算法具体步骤如下:

首先根据规则交叉的定义,计算每个规则对应的交叉点的数量.根据等级差定义,计算每个规则对应的等级差.根据规则结合趋势定义,计算规则集合的趋势.

然后选取交叉点最多的点对应的规则加入到异常规则集合中,如果存在交叉点相同的情况,选择等级差绝对值最大的规则加入到异常规则集合中,如果上面两个条件都相同的话,选择与大趋势相反的规则加入到异常规则集合中,如果上面的三个条件都相同的话,选择支持度最低的规则加入到异常集合中.

最后每当有规则加入到异常集合中,需要重复前两个步骤,并且在完成第一个步骤时,需要对剩余规则集合进行判断,如果剩余的规则集合中不存在交叉点说明规则划分完成,反之继续迭代.

#### 算法示意图

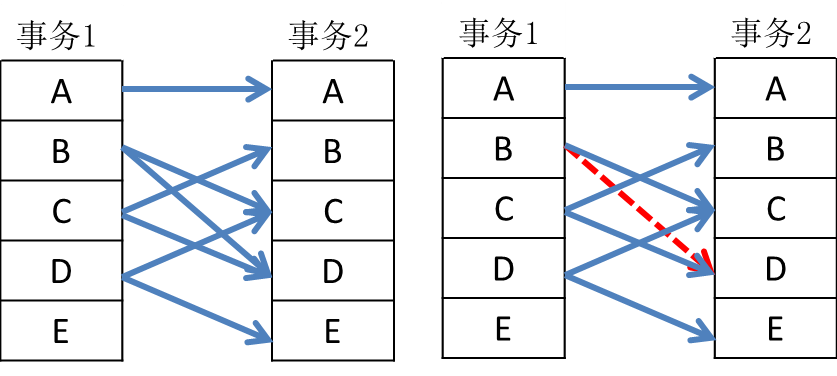


图6 正常与异常模式挖掘算法示意图-1

图6到图7给出了挖掘算法的解释,图6（左）图7（右）选择规则B->D的原因在于同样拥有相同交点的C->B和D->C,D->C对应的等级差小于B->D对应的等级差,所以选择B->D并将其加入到异常规则集合中,图7（左）选择D->C的原因在于虽然规则B->C,C->B,C->D,D->C有同样的交叉点以及相同的等级差,但是规则D->C与规则C->B与规则集合的大趋势相反,同时规则D->C对应的支持度小于规则C->B对应的支持度,所以应该选择规则D->C,并将其加入到异常规则集合中.同样图7（右）基于相同的方法选择将C->B加入到异常规则集合中.

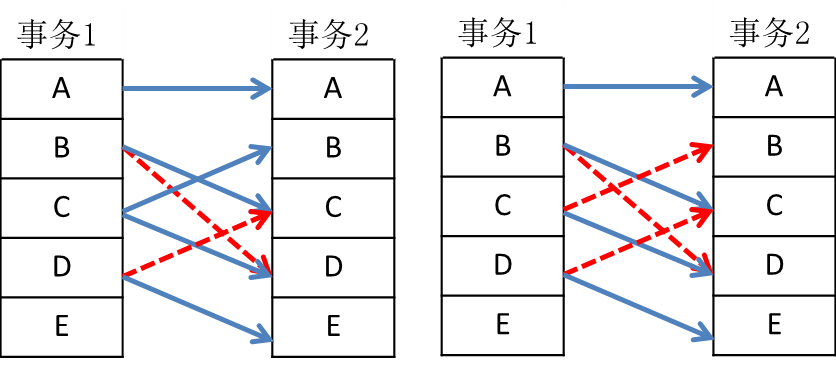


图7 正常与异常模式挖掘算法示意图-2

#### 算法流程

图8 正常与异常模式挖掘算法

#### 算法伪代码

根据上面介绍的算法思路算法伪代码主要包含以下几个规则:

选择交叉点最多点的规则:交叉点代表的是与该规则有交叉的规则的数量,交叉点多的规则意味着该规则在整个规则集合中有交叉或者冲突的规则数量的最大或者反超该规则的规则数量大,该规则造成全局规则混乱的可能性最大,如果剔除了该规则,规则集合变为正常规则集合的可能性也最大.

选择等级差最大的规则:等级差大的规则相对于等级差小的规则来说在两个事务间等级变化的更大,等级变化更大意味着更有可能是异常规则.

选择与全局大趋势相反的规则:全局大趋势代表的是规则整体的趋势,与规则整体趋势相反的规则更容易成为异常规则.

选择支持度小的规则:异常的规则支持度较低的可能性比较大,所以选取支撑度小的规则.

|  |
| --- |
| 算法2:正常与异常模式挖掘算法 |
| 输入:关联规则集合  输出:正常规则集合,异常规则集合   |  |  | | --- | --- | | 1 | normalRuleSet={} abnormalRuleSet={} tmpRulset={} | | 2 | while RuleSetHaveCross(RuleSet): #判断当前规则集合是否存在交叉点 | | 3 | for rule in RuleSet: | | 4 | CaculateCross(rule)#计算每个规则的交叉点数量 | | 5 | end for | | 6 | ruleList=FindRulesWithMostCrossNum(RuleSet) #寻找规则集合中交叉点数据最多的规则子集 | | 7 | if len(ruleList)==1: | | 8 | AddRule(ruleList,abnormalRuleSet) #添加规则到特定规则集合 | | 9 | RemoveRule(ruleList,RuleSet) #在特定规则集合中删除规则 | | 10 | Continue | | 11 | endif | | 12 | ruleList=FindRulesWithHighLevelDif(ruleList) #寻找规则集合中等级差最大的规则子集 | | 13 | if len(ruleList)==1: | | 14 | AddRule(ruleList,abnormalRuleSet) | | 15 | RemoveRule(ruleList,RuleSet) | | 16 | continue | | 17 | endif | | 18 | ruleList=FindRulesWithDifThrend(ruleList) #寻找规则集合中与大趋势相反的规则子集 | | 19 | if len(ruleList)==1: | | 20 | AddRule(ruleList,abnormalRuleSet) | | 21 | RemoveRule(ruleList,RuleSet) | | 22 | endif | | 23 | ruleList=FindRuleWithLeasetSupport(ruleList) #寻找规则集合中支撑度最小的规则 | | 24 | AddRule(ruleList,abnormalRuleSet) | | 25 | RemoveRule(ruleList,RuleSet) | | 26 | end while | | 27 | return abnormalRuleSet,normalRuleSet | |

图9 正常与异常模式挖掘算法伪代码

# 教学数据验证

## 实验数据来源说明

实验数据来自于北京航空航天大学计算机学院本科学生在校成绩,其中包含590名学生的成绩数据,对应的是40门课程（不包括实验课程）.

课程分类:分为必修课（公共必修课,专业必修课）,选修课.必修课每个学生都会有成绩,选修课只有部分学生有成绩.

成绩数据形式:一种该学生在该门课程的评级（A,B,C,D）,另外一种是学生在该门课程的综合成绩（考勤成绩,作业成绩,试卷成绩的加权成绩）.

## 实验数据预处理

扫描所有的成绩数据按照表4中的成绩区间对应的成绩等级将成绩数据离散到相应的等级上.

表3 成绩数据离散表

|  |  |
| --- | --- |
| 成绩区间 | 成绩等级 |
| 100-85 | A |
| 84-70 | B |
| 60-69 | C |
| 59-40 | D |
| 0-39 | E |

## 实验一:关联模式在两门课程上面的应用

### 实验方法

表4成绩数据表

|  |  |
| --- | --- |
| 数学分析1 | 数学分析2 |
| C | C |
| B | B |
| A | A |
| B | C |
| A | B |
| ……. | …… |
| D | D |
| B | A |
| B | D |

本实验将关联模式挖掘算法应用到数学分析1以及数学分析2两门课程对应的成绩数据上.表4中的数据是学生在数学分析1与数学分析2上的离散化后的成绩数据图.图10和图11为关联模式挖掘算法应用到这两门课程后得到的规则集合图.

### 实验结果与分析

图10反映的是上升与下降模式挖掘算法针对表4的数据挖掘出的两个规则集合.图10（左）可以看出有248人数学分析1到数学分析2上出现了上升,图10（右）可以看出有45人出现了下降的趋势,整体能够给出的信息是数学分析1到数学分析2上学生成绩出现上升的比较多.

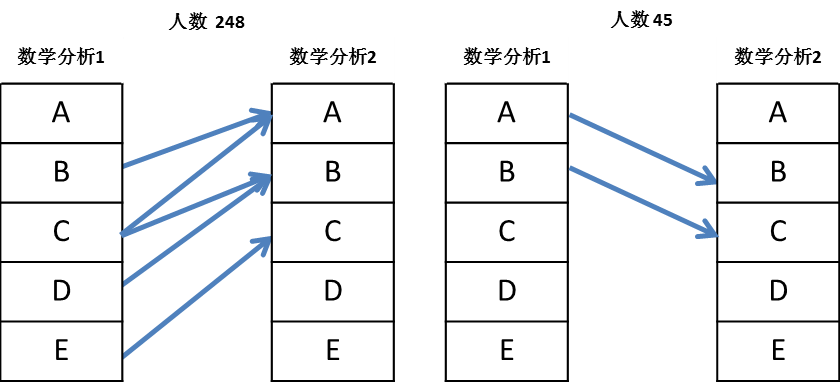


图10 上升与下降模式实验数据图

图11反映的是正常与异常模式挖掘算法针对表4的数据挖掘出的两个规则集合.图11（左）代表正常规则集合,图11（右）代表的是异常规则集合.通过分析正常模式对应的规则集合可以看出处于不同成绩等级的大部分学生成绩都出现了上升,并且针对B,C两个等级来说上升的普遍范围在0到1个等级,而分析异常模式规则集合可以看出相对于正常模式中的规则,异常规则中的规则或者是下降的,或者是上升幅度过大的,而这些规则为以后分析相应成绩等级下的学生的行为提供了依据,并且从通过挖掘的两种模式对应的学生的数量可以看出两门课程存在较强的关联性.

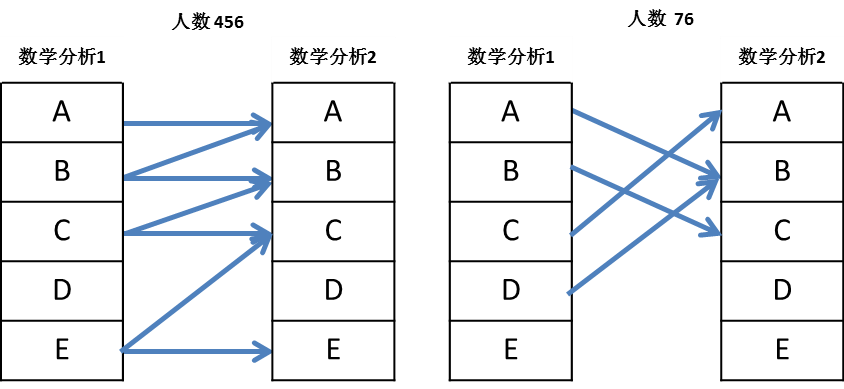


图11 正常与异常模式实验数据图

## 实验二:关联模式挖掘在课程相关度计算上的应用

### 实验方法

通过关联模式中正常与异常模式的挖掘,可以将规则集合划分为异常规则集合以及正常规则集合,如果异常规则集合与正常规则集合对应的规则的支持度和比较接近,两事务的关联度低,相反关联度高.

本章利用以下方法计算事务间关联度,其中Sim为相关度,N对应的是正常规则集合,D对应是异常规则集合,Support对应规则集合中支持度和.

1．当时,.

2．当时,利用公式1计算关联度,Sim值越小对应的关联度越小,Sim值越大对应的关联度越大.

 (1)

本实验中将关联模式挖掘算法以及关联度计算方法应用教学数据中,计算40门课程两两课程之间的关联度.实验结果如表5,表中选取了部分课程进行演示,针对某一门课程本文将所有课程与该课程的关联度排序后选择关联度最高的四门课程显示出来.并且为了更清楚的展示将关联模式挖掘应用到计算课程关联度上的结果,本实验利用挖掘出的课程间的关联度对课程数据进行层次聚类,层次聚类结果如图12.

表5 相关课程实验结果表

|  |  |  |
| --- | --- | --- |
| 课程编号 | 课程名称 | 关联紧密的课程列表 |
| 1 | 数学分析1 | 数学分析2,高等数学,离散数学2,物理1 |
| 2 | 编译原理 | 模拟电子,计算机组成原理,操作系统,计算机网络 |
| 3 | 高等数学 | 物理2,数学分析1,数学分析2,电路分析 |
| 4 | 多媒体技术 | 虚拟现实技术,信号与系统,模式识别,计算机图形学 |
| 5 | 模拟电子 | 汇编语言,信号与系统,计算机接口与技术,EDA设计 |
| 6 | 操作系统 | 计算机网络,编译原理,数据库原理,模拟电子 |
| 7 | 汇编语言 | 模拟电子,数字逻辑,电路分析,数据库原理 |

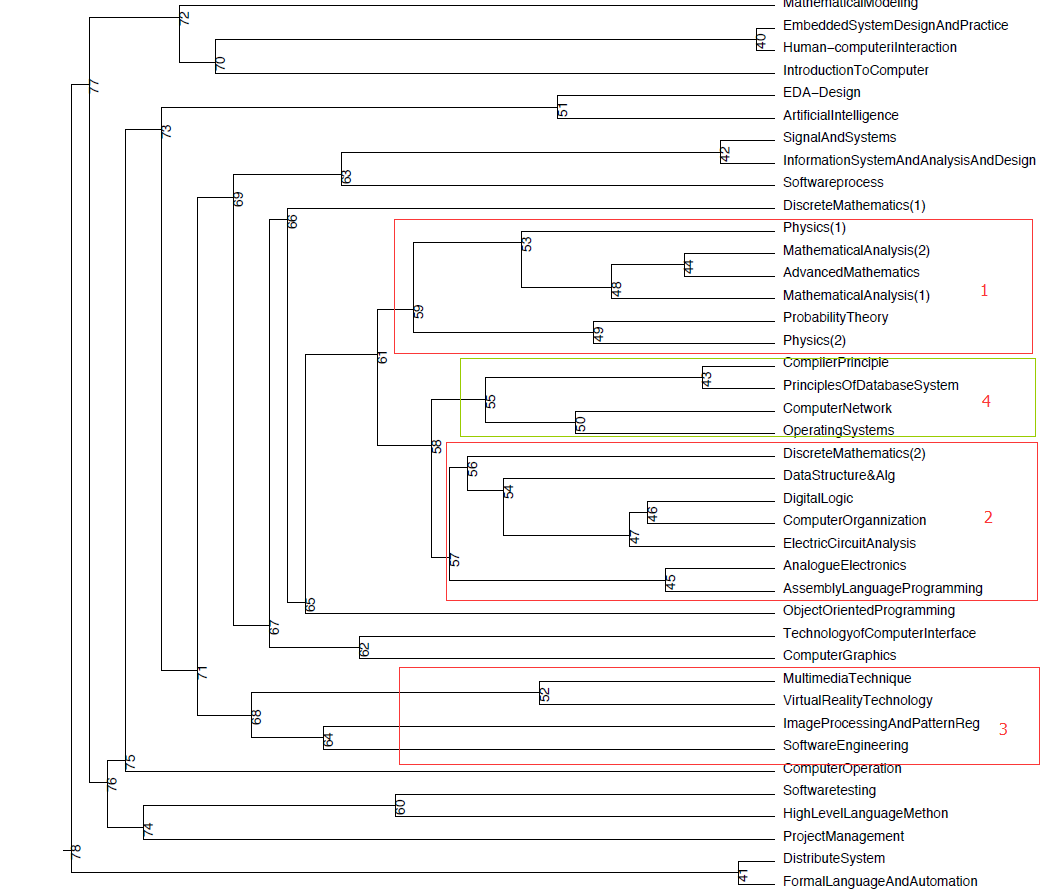


图12 课程层次聚类图

### 实验结果分析与解释

表5中第1行对应的是与数学分析1相关的课程列表,按照学习经验数学分析2,高等数学2,物理与数学分析1都有着较强的关联,而离散数学与数学分析1的关系无法按照经验知识判断.

表5中第2行对应的是与编译原理相关的课程列表,按照学习经验,我们认为挖掘出的与编译原理相关的课程列表是不准确的.

表5中第3行对应的是与高等数学相关的课程列表,按照学习经验,这些课程都有着将强的关系.

表5中第4行对应的是与多媒体技术相关的课程列表,按照学习经验,该行对应的课程认为有一定的关联关系.

表5中第5行对应的是与模拟电子相关的课程列表,按照学习经验,表中的课程大部分对应的是硬件方向的课程所有具有较强的关联关系.

表5中第6行对应的是与操作系统相关的课程列表,按照学习经验,这些课程关联关系较弱.

表5中第7行对应的是汇编语言相关的课程列表,按照学习经验,模拟电子,数字逻辑以及电路分析与汇编语言一定的关联关系,但是数据库系统关联的强弱无法确定.

图12为根据课程相关度进行课程层次聚类后的结果图,图中用红色方框括起来的课程是相关性较强的课程.第一个框内的课程.包括数学方面的课程以及物理方面的课程;第二个框内课程主要包括有硬件相关的课程;第三个框内课程主要包括图形图像以及模式识别方面的内容,红色框内的课程被较早的聚集在一起说明这些课程之间的相关性比较强,这些课程能够符合我们的经验知识.绿色框内的课程包括了编译原理,操作系统,计算机网络和数据库原理,这四门课程为计算机最主要的课程,虽然在知识上可能联系不大,但是也被较早的聚集在了一起,这些课程可能在学生学习方法或者学习的关注度上存在一定的关联性,所以反映到成绩上后具有较强关联.

实验结果中有些课程的关联关系符合我们的经验知识,有些不符合我们的经验知识,本文给出两种产生这种误差的可能原因:

1．启发式的算法存在一定的误差性,算法基于的假设是如果两门课程存在强关联性,那么学生在这两门课程上的成绩应该是相当的,算法基于上述的假设通过成绩数据来推导出课程相关度,但是由于学生的学习过程是一个很复杂过程,学习过程会受到很多因素的影响,这些影响的因素会使得课程间的关联变的不够明显,由此导致了关联度计算的偏差.

2．两个课程间存在隐含的关联性,并且反映到了学生的成绩中,但是这种关联性还没有被认识到,这需要教育学家对课程以及学生的学习习惯进一步的进行分析.

# 总结

本文针对关联关系挖掘的主要缺陷提出了关联模型的概念以及关联模式挖掘算法,算法通过对关联规则进行整合与挖掘,生成一系列富含有价值信息的关联模式.并且本文将关联模式挖掘应用到教学课程数据,挖掘出课程间关联模式,并在此基础上提出了一个计算课程关联度的应用实例.通过真实教学数据实验验证,关联模式能有效为用户提供清晰的且有价值的信息,并且,关联度计算方法能反映课程间的关联关系.

算法首次提出了上升模式与下降模式,正常模式与异常模式的挖掘算法.主要针对带有等级的数据,通过挖掘出两个数据集合之间上升与下降的关系,从而运用正常与异常规则的挖掘算法,挖掘出两个数据集之间的正常规则集合和异常规则集合.如果异常规则小于正常规则,则说明两个数据集间具有关联关系;反之,二者就不具有关联关系.在最后将挖据出的正常与异常数据集的个数代入量化公式,就能够清晰地反映出两个对象关联关系的强弱.

数据挖掘在教学数据的研究上仍具有很大的研究空间,如何设计准确的算法将课程进行科学分类,将是我们未来要研究的方向之一.

References:

1. Sanjeev A P, Zytkow J M. Discovering Enrollment Knowledge in University Databases[C].KDD. 1995: 246-251
2. Dillon A, Stolk J. The students are unstable! Cluster analysis of motivation and early implications for educational research and practice[C].Frontiers in Education Conference (FIE), 2012. IEEE, 2012: 1-6.
3. Becker K, Ghedini C G, Terra E L. Using KDD to analyze the impact of curriculum revisions in a Brazilian university[C].AeroSense 2000. International Society for Optics and Photonics, 2000: 412-419.
4. Ma Y, Liu B, Wong C K, et al. Targeting the right students using data mining[C].Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2000: 457-464.
5. Ramaswami M, Bhaskaran R. A CHAID based performance prediction model in educational data mining[J]. arXiv preprint arXiv:1002.1144, 2010.
6. Luan J. Data Mining and Knowledge Management in Higher Education-Potential Applications[J]. 2002.
7. Bian H. A Preliminary Study on Clustering Student Learning Data[C].MAICS. 2011: 128-132.
8. Xin-huaZhu, Zhang J, Deng Y. Application of Mixed-Weighted-Association-Rules-Based Data Mining Technology in College Examination grades Analysis[J]. JDCTA: International Journal of Digital Content Technology and its Applications, 2012, 6(10): 336-344.
9. Yadav S K, Bharadwaj B, Pal S. Data mining applications: A comparative study for predicting student's performance[J]. arXiv preprint arXiv:1202.4815, 2012.
10. Singh C, Gopal A, Mishra S. Extraction and analysis of faculty performance of management discipline from student feedback using clustering and association rule mining techniques[C].Electronics Computer Technology (ICECT), 2011 3rd International Conference on. IEEE, 2011, 4: 94-96.
11. González-Brenes J P, Mostow J. What and when do students learn? Fully data-driven joint estimation of cognitive and student models[C].Proceedings of the 6th International Conference on Educational Data Mining. 2013: 236-240.
12. Toscher A, Jahrer M. Collaborative filtering applied to educational data mining[J]. KDD Cup, 2010
13. 梁燕红. Web 数据挖掘技术在网络教学中的应用[J]. 科技信息, 2011 (22): I0244-I0245.
14. 卢朝晖, 刘家宁. Web 日志挖掘技术在网络教学平台中的应用研究[J]. 制造业自动化, 2009 (11): 184-185.
15. Bedi K, Milic M, Stedul I. Information society and e-learning[C].MIPRO, 2012 Proceedings of the 35th International Convention. IEEE, 2012: 1249-1253.
16. Chellatamilan T, Ravichandran M, Suresh R M, et al. Effect of mining educational data to improve adaptation of learning in e-learning system[C].Sustainable Energy and Intelligent Systems (SEISCON 2011), International Conference on. IET, 2011: 922-927.
17. Zaíane O R. Building a recommender agent for e-learning systems[C].Computers in Education, 2002. Proceedings. International Conference on. IEEE, 2002: 55-59.
18. Klašnja-Milićević A, Vesin B, Ivanović M, et al. E-Learning personalization based on hybrid recommendation strategy and learning style identification[J]. Computers & Education, 2011, 56(3): 885-899.
19. Antonenko P D, Toy S, Niederhauser D S. Using cluster analysis for data mining in educational technology research[J]. Educational Technology Research and Development, 2012, 60:383-398. DOI:10.1007/s11423-012-9235-8.
20. Le Blanc L A, Rucks C T. Data mining of university philanthropic giving: Cluster-discriminant analysis and Pareto effects[J]. International Journal of Educational Advancement, 2009, 9(2): 64-82.
21. Hämäläinen W, Suhonen J, Sutinen E, et al. Data mining in personalizing distance education courses[C].World conference on open learning and distance education. 2004.
22. Jovanovic M, Vukicevic M, Milovanovic M, et al. Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study[J]. International Journal of Computational Intelligence Systems, 2012, 5(3): 597-610.
23. Lu J. Personalized e-learning material recommender system[C].International conference on information technology for application. 2004: 374-379.
24. Hwang G J, Hsiao C L, Tseng J C R. A computer-assisted approach to diagnosing student learning problems in science courses[J]. J. Inf. Sci. Eng., 2003, 19(2): 229-248.
25. García E, Romero C, Ventura S, et al. A collaborative educational association rule mining tool[J]. The Internet and Higher Education, 2011, 14(2): 77-88.
26. Romero C, Espejo P G, Zafra A, et al. Web usage mining for predicting final marks of students that use Moodle courses[J]. Computer Applications in Engineering Education, 2013, 21(1): 135-146.
27. Kim J, Chern G, Feng D, et al. Mining and assessing discussions on the web through speech act analysis[C].Proceedings of the Workshop on Web Content Mining with Human Language Technologies at the 5th International Semantic Web Conference. 2006..
28. Agrawal R, Mannila H, Srikant R, et al. Fast Discovery of Association Rules[J]. Advances in knowledge discovery and data mining, 1996, 12(1): 307-328.
29. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation[C].ACM SIGMOD Record. ACM, 2000, 29(2): 1-12.

1. [↑](#footnote-ref-1)