

基于关联规则的学习行为关联度分析方法研究

李 飒

(辽宁石油化工大学, 辽宁 抚顺 113000)

摘 要: 目前主要以在线访谈和发放调查问卷方式对学习者的学习行为进行调查, 没有充分利用网络在线远程教学平台中遗留的大量数据, 忽略了学习者的个性特征, 导致分析的结果可推广性较差. 提出基于管理规则的学习行为关联度分析方法, 选取某大型网络在线远程教育平台中记录的学习者学习行为数据进行采集和处理; 采用关联规则挖掘算法对学习者的学习行为与学习效果之间存在的规律性进行了关联分析. 实验结果表明, 在课堂实验上的大部分学习行为上学生个体之间存在显著差异性, 且学生的 Diagram, FSM, Study, Properties 和 TextEditor 这五种学生学习行为参与频率与学习效果之间存在显著的正相关关系.

关键词: 关联规则; 学习行为; 关联度

中图分类号: TP182

文献标识码: A

文章编号: 1000-7180(2018)06-0065-04

Study on Correlation Analysis Method of Learning Behavior Based on Association Rules

LI Sa

(Liaoning Shihua University, Fushun 113001, China)

Abstract: The current study mainly in the online interview and questionnaire method on learners' learning behavior investigation, did not make full use of large amounts of data from the network distance online learning platform, ignoring the personality characteristics of learners, leading to the results of the analysis extension is poor. This learning behavior association degree analysis method based on Rules select the record, a large network of online learners in distance education learning behavior data acquisition and processing; the regularity of mining association rules based on learning behavior and learning algorithm of learners among the association analysis. The experimental results show that the experiment in the classroom There are significant differences between students in part learning behavior, and there is a significant positive correlation between the participation frequency of five kinds of student learning behaviors, Diagram, FSM, Study, Properties and TextEditor.

Key words: association rules; learning behavior; association degree

1 引言

教育与信息技术二者的融合发展越来越相互适应. 网络在线教育便是信息化教育蓬勃发展下衍生出的一种新型的具有典型代表性的教育形式^[1]. 根据《中国互联网教育用户行为分析报告》可知^[2], 当前我国利用互联网教育满足对知识、技能和兴趣爱

好需求的群体呈现逐年递增趋势. 因此, 对网络学习行为关联度进行研究分析势在必行, 且未来前景不可估量^[3-4].

本文研究在已有的学生学习行为数据基础上, 利用关联规则挖掘中最具影响力的 Apriori 算法对虚拟实验平台下记录的学生学习行为进行分析, 具体研究了学生学习行为与学习效果之间的关联度.

收稿日期: 2017-10-10; **修回日期:** 2017-11-01

基金项目: 辽宁省教育厅“十二五”规划项目(JG14DB226); 辽宁省教育厅教改项目(辽教发[2016]23-21)

2 基于关联规则的学习行为关联度分析方法研究

2.1 学习行为分析数据的采集与处理

学习行为数据主要由四部分构成^[5],即:人口统计属性数据、人机交互学习属性数据、人人协作学习属性数据、期末测评成绩数据^[6].其中人口统计属性数据记录了学习者的一些基本资料,如学习者的性别、年龄、受教育情况以及学习者本身的学习风格;人机交互学习属性数据记录了学习者在网络环境下浏览课件资料(包括文本和视频)的总耗时和学习者课后作业的平均成绩;人人协作学习属性数据记录了一些能够表现学习者对网络在线学习不同会话层次的贡献程度,从学习者在网络协作学习过程中,发表的不同类型言论的总字数与所有发言的总次数都能体现出来^[7];对学习者期末总测评成绩设置以下四个等级:

$$\text{Grade} = ("A", "B", "C", "D") \quad (1)$$

式中,学习者期末总测评成绩小于 60 的属于 D 级;学习者期末总测评成绩大于等于 60 并且小于 70 的属于 C 级;学习者期末总测评成绩大于等于 70 并且小于 80 的属于 B 级;学习者期末总测评成绩大于等于 80 的属于 A 级.

2.2 学习者学习行为数据的处理

假设 $I = \{i_1, i_2, \dots, i_n\}$ 表示学习者学习行为项目集; D' 表示学习行为全体事务的集合;学习行为事务 T 是项目集 I 上的一个子集,满足 $T \subseteq I$,每个学习行为事务采用唯一的标识 TID' 来表示^[8].学习行为的关联规则计算公式如下:

$$X \Rightarrow Y \quad (2)$$

式中, X 表示学习行为关联规则的条件; Y 表示学习行为关联规则的结果;且满足 $X \subseteq I, Y \subseteq I, X \cap Y = \emptyset$.

2.3 基于关联规则的学习行为关联度挖掘

对学习行为数据的挖掘分析得到的 50 条规则,筛选学习行为数据中前件和后件中包含学习风格属性的数据或者具有一定指导意义的规则,共计 9 条^[9].

(1) 与学习者学习风格不相关的普适性规则分类和注释

规则 ①:假设学习者自主学习时间较长,且课后作业平均成绩和期末总测评成绩成正相关关系.则计算公式如下:

$$\begin{aligned} \text{duration} &= "(15.1 - \text{inf})" \text{homework} = \\ "&(85 - \text{inf})"10 \Rightarrow \text{test} = "(84 - \text{inf})"9 \text{conf}(0.9) \end{aligned} \quad (3)$$

规则 ②:假设学习者之间协作频率较低,且课后作业平均成绩和期末总测评成绩都较差,则计算公式如下:

$$\begin{aligned} \text{frequency} &= "(-\text{inf}, 41]" \text{homework} = \\ "&(-\text{inf}, 81]"11 \Rightarrow \text{test} = \\ "&(-\text{inf}, 78]"10 \text{conf}(0.91) \end{aligned} \quad (4)$$

规则 ③:假设是女性学习者前期学历可能为计算机相关专业,并且期末总测评成绩较好,则计算公式如下:

$$\begin{aligned} \text{test} &= "(84 - \text{inf})" \text{sex} = \text{female}24 \Rightarrow \text{bg} = \\ \text{yes}22 \text{conf}(0.92) \end{aligned} \quad (5)$$

规则 ④:假设学习者之间的进行较高频率的协作学习,并且其花费在网络在线自主学习的时间一般,则计算公式为:

$$\begin{aligned} \text{frequency} &= "(79 - \text{inf})"51 \Rightarrow \text{duration} \\ &= "(13.6 - 15.1]"45 \text{conf}(0.88) \end{aligned} \quad (6)$$

规则 ⑤:假设学习者参加的讨论与信息共享比较频繁,学习者之间进行较高频率的协作学习,则计算公式如下:

$$\begin{aligned} \text{discuss} &= "(106 - \text{inf})" \text{inform} = "(337 - \text{inf})" \\ 32 \Rightarrow \text{frequency} &= "(79 - \text{inf})"26 \text{conf}(0.81) \end{aligned} \quad (7)$$

规则 ⑥:假设学习者在线自主学习时间较长,参与的协作学习和讨论频率较低,但在协作学习和讨论时也会发表自己的言论,其期末总测评成绩较好,则计算公式为:

$$\begin{aligned} \text{frequency} &= "(-\text{inf}, 41]" \text{duration} = \\ "&(15.1 - \text{inf})" \text{motivate} \\ &= "(32, 67]" \text{learningStyle} \Rightarrow \text{test} = \\ "&(84 - \text{inf})"4 \text{conf}(0.8) \end{aligned} \quad (8)$$

(4) 适应性学习者学习行为相关规则分类和注释

规则 ⑦:假设学习者参与的讨论类很少发言,并且期末总测评成绩较低,则计算公式为:

$$\begin{aligned} \text{discuss} &= "(-\text{inf}, 74]" \text{learningStyle} = \\ \text{acc}14 \Rightarrow \text{test} &= "(-\text{inf}, 78]"11 \text{conf}(0.79) \end{aligned} \quad (9)$$

规则 ⑧:如果学习者为男性同化类型,其参与情感类讨论和信息共享类协作较少,表达式为:

$$\begin{aligned} \text{sex} &= \text{male} \text{learningStyle} = \text{ass}17 \Rightarrow \text{motivate} \\ "&(-\text{inf}, 32]" = "(-\text{inf}, 213]"12 \text{conf}(0.71) \end{aligned} \quad (10)$$

规则 ⑨:如果属于同化类型的学习者花费的自

主学习时间较多,其更容易获得较好的期末总测评成绩.表达形式为:

$$\begin{aligned} learningStyle &= assduration = "(15.1 - inf)" \\ 13 \Rightarrow test &= "(84 - inf)" 10conf(0.77) \end{aligned} \tag{11}$$

根据上述分析的学习者学习行为之间的关联度,建立基于关联规则的学习行为关联度分析模型,一方面能够帮助学习者更加了解自己当前的学习情况;另外一方面,能够帮助教师更好的选择有针对性的教育教学资源和教学模式,提供科学恰当的教学指导,为实施合理的教育教学干预提供更加准确的理论基础^[10].

3 实验结果与分析

本文研究采用的实验数据来源于 Wahdat 等(2014)收集的某大学计算机工程专业大二学生在 DEEDS 平台上在线学习的操作数据集.实验采用基于问题解决的教学方式,让学生根据预设问题在网络实验环境下进行自主探究以完成实验课程学习.

(1)学生学习行为的总体情况分析
首先对学生的学习行为进行提取,表 1 记录了从学生的 15 种学习活动中提炼出的 9 种学习行为名称以及具体描述;

表 1 学生学习行为及其具体描述

学生学习行为	学习行为具体描述
Aulaweb	代表学生在 DEEDS 平台上学习时上传和下载相关学习资料
deeds	在 DEEDS 平台上进行的一系列实验模拟操作行为
Diagram	代表学生绘制相应逻辑电路的时序电路图
FSM	代表学生在“有限状态机”上进行的电路连接模拟操作
Properties	代表学生使用属性窗口调整任务所需参数
Study	代表学生学习或者复习与实验课程相关的学习材料
TextEditor	代表与文本编辑相关的学习行为,例如填写实验报告数据在 excel 上
Blank	代表学生所访问的窗口标题没有被记录
Other	代表除上述学习行为之外,实验课程操作过程中可能发生的相关学习行为,例如用办公软件交流

在完成上述学生学习行为数据提取和统计之后发现,有一部分学生有学习行为记录,但缺少期末总测评成绩,还有一些学生有成绩但是没有学习行为记录.为了实现学生学习行为与学习效果之间的关联度,实验舍弃了这两种情况的所有样本数据,最终

得到的学生学习行为研究样本数目为 92.

(2)学习行为参与频率分析结果

研究结果表明,对于每一位学生个体来说,平均有 2869.95 次学习行为被统计记录了下来,但是标准差却显示学生的所有学习行为数据分布非常不均匀,即不同学生个体在学习行为参与频率总和上存在显著差异.

课堂实验上的 9 种学生学习行为中,TextEditor,Deeds 的发生频率分别占总体频率的 18.2%和 20.61%,是在课堂实验中学生参与频率较高的两种学习行为;TextEditor 代表学生在从事与课堂实验报告相关的行为活动;Deeds 代表学生在操作课堂实验电路模拟器上的相关活动.实验结果表明,学生的这两种学习行为是参与度差异性最大的两种学习行为.

Aulaweb 的学习行为参与频率最小,它代表学生进行与课堂实验操作密切相关的行为活动.Diagram,FSM,Properties 和 Study 四种学习行为是学生在课堂实验中总体参与频率比较接近,分别占比为 10.00%,10.53%,9.60%和 10.64%.根据表 2 中的标准差可以看出,这四种学习行为发生频率比较不稳定的是 FSM 和 Study,分别为 111.48 次和 10.3.15 次.

(3)学生学习行为与学习效果之间的关联度分析

通过对学生学习行为的相关分析结果发现,不同学生个体在大部分学习行为发生频率上均存在较大的波动性.研究两个变量之间相关关系习惯用的统计学方法是 Bi-variate 双变量相关分析法,常用的相关系数有三种:Pearson 相关系数、Spearman 相关系数以及 Kendall 相关系数.由于本文研究分析的是两个连续数据之间的相关关系,且各个变量之间基本符合正态分布,采用 Pearson 相关系数对学生学习行为指标与学习效果进行关联度分析,分析测试结果如表 2 所示.

表 2 学生学习行为与学习效果之间的关联度分析

	Time	Aulaweb	Deeds	Diagram	FSM
成绩	0.152	0.037	0.109	0.254 *	0.253 *
	Properties	Study	TextEditor	Blank	Other
成绩	0.331 **	0.278 *	0.220 *	0.081	0.112 *

注:表 2 中的 * p<0.05, ** p<0.01.

从表 2 的研究分析结果可以看出,Time,Aulaweb,Deeds,Blank 和 Other 这四种学生学习行为参与频率和学习成绩(即学习效果)并无显著相

关性.

其中,Diagram、FSM 和 TextEditor 这三种学生学习行为和学习效果在 0.05 水平上呈现明显相关,其余的 Properties 与 Study 这两种学生学习行为与学习效果在 0.01 水平上呈现明显相关.

综上分析,影响学生学习效果的主要学习行为包括 Diagram, FSM, TextEditor, Properties 和 Study 五种学生学习行为. 这五种学生学习行为不仅充分说明了学生在学习过程中可能存在的问题,同时也为教师后续改进教育教学方法提供了相应依据.

4 结束语

本文研究只是对学习行为关联度分析的初步探索. 未来将开展的工作包括:利用多种数据挖掘技术对学习行为数据进行进一步挖掘,发现更多隐藏在学习行为背后的认知过程,为教师后续教育教学提高更加准确的数据基础;综合运用多种技术手段增强学习行为关联度分析结论的完整性与客观性.

参考文献:

[1] 姜永超. 基于数据挖掘的学生选课及学习行为分析算法研究[J]. 现代电子技术, 2016, 39(13):145-148.

[4] 张鹏, 赵齐乐, 李由, 等. 基于 PDR、WiFi 指纹识别、磁场匹配组合的室内行人导航定位[J]. 测绘地理信息, 2016, 41(3): 29-32.

[5] 薛洋, 刘龙坡, 金连文. 一种基于加速度传感器的实时步长估计方法[P]. 中国, CN103983273A. 2014.

[6] 谢宏伟. 基于智能手机平台的地磁室内定位系统[D]. 南京:南京大学, 2015.

[7] 陈斌涛, 刘任任, 陈益强, 等. 动态环境中的 WiFi 指纹自适应室内定位方法[J]. 传感技术学报, 2015(5): 729-738.

[8] Zhu J, Luo X, Chen D. Maximum like lihood scheme for finger printing positioning in LTE system[C]// Communication Technology (ICCT), 2012 IEEE 14th International Conferenceon Communication Technology. Piscataway, NJ: IEEE, 2013:428-432.

[9] Xiong K, Zhang H Y, Chan C W. Performance evalu-

[2] 王斌, 王金, 宁小倩. 基于黄金比例的关联规则挖掘算法[J]. 计算机仿真, 2015, 32(8):302-305.

[3] 陈妍, 张新慧, 朱海萍, 等. 面向网络教育学院的学习行为分析[J]. 计算机应用, 2016, 36(s1):224-227.

[4] 张继荣, 王向阳. 基于 XML 数据挖掘的 Apriori 算法的研究与改进[J]. 计算机测量与控制, 2016, 24(6): 178-180.

[5] 陈康, 王丹丹. 关联规则数据结构分布重排的数据库索引算法[J]. 科技通报, 2015, 31(10):178-180.

[6] 蒋卓轩, 张岩, 李晓明. 基于 MOOC 数据的学习行为分析与预测[J]. 计算机研究与发展, 2015, 52(3): 614-628.

[7] 陈妍, 张新慧, 朱海萍, 等. 面向网络教育学院的学习行为分析[J]. 计算机应用, 2016, 36(s1):224-227.

[8] 樊超, 宗利永. MOOC 在线学习行为的人类动力学分析[J]. 开放教育研究, 2016, 22(2):53-58.

[9] 师黎, 陶梦妍, 李志辉. 鸽子强化学习过程中内部学习状态的动态建模研究[J]. 科学技术与工程, 2017, 17(13):120-125.

[10] 陈宏. 基于关联规则挖掘算法的用电负荷能效研究[J]. 电子设计工程, 2017, 25(4):79-82.

作者简介:

李 飒 女, (1970-), 博士, 副教授. 研究方向为数据挖掘、信息管理、互联网+. E-mail: lisa-ty@sohu. com.

(上接第 64 页)

ation of UKF-based nonlinear filtering[J]. Automatica, 2006, 42(2): 261-270.

[10] 汪秋婷, 胡修林. 基于 UKF 的新型北斗/SINS 组合系统直接法卡尔曼滤波[J]. 系统工程与电子技术, 2010, 32(2): 376-379.

作者简介:

宋 宇 男, (1969-), 硕士, 教授, 硕士生导师. 研究方向为嵌入式系统设计与研究、数学信号处理.

喻文举(通讯作者) 男, (1988-), 硕士研究生. 研究方向为室内定位技术. E-mail: yuwenjuju@163. com.

程 超 男, (1984-), 博士, 讲师. 研究方向为无线通信理论与技术.

王 磊 男, (1988-), 硕士研究生. 研究方向为图像、数字信号处理.