

基于网络学习行为的学习风格挖掘

吴青,罗儒国

(武汉大学 湖北 武汉 430072)

【摘要】学习风格是个体相对稳定的信息组织和信息加工方式,极大地影响着学习进程。掌握学习者的学习风格,有利于教师提供有针对性指导,从而提高网络学习有效性。本文基于数据挖掘技术,应用混合方法测量网络学习者的学习风格。在显式获取用户学习风格的基础上,提出运用J48算法,挖掘不同风格学习者的网络学习行为特征,并构建学习风格模型。其中重点探讨了应用最近邻居法进行异常点挖掘,以减少受其他因素使得网络学习行为无法客观体现学习风格的样本的干扰。以及利用属性选择方法,评估网络学习行为与学习风格的关联关系。并结合相关教学原则对构建的学习风格模型开展教学反思。

【关键词】学习风格;数据挖掘;网络学习行为

【中图分类号】G43

【文献标识码】B

【文章编号】1001-8700(2014)01-0054-09

一、引言

学习风格作为认知风格的应用,通常指学习者的喜好或信息处理偏好,而并非实际信息处理能力或技能^[1],学习风格差异决定着学习者的学习行为,体现了个体学习特征,是个体相对稳定的信息组织和信息加工方式,将极大地影响着学习过程。针对学习者的学习风格,采用适当的教学策略,将极大激发和维持学习者的学习动机,提高学习效率,使学习变得很容易。诚如所述“学习者分析的目的是为了了解学习者的学习准备情况及其学习风格,为……教学外因条件适合学习者的内因条件提供依据,从而使教学真正促进学习者智力和能力的发展。”对于接受远程教育的网络学习者而言,学习风格的确定变得尤其重要。网络学习的有效性不仅依赖于网络学习环境为学习者提供的外在支撑,更为重要的是,网络学习的环境和过程必须与学习者的个体特征相匹配^[2]。掌握学习者的学习风格,有利于教师合理组织在线课程。若教师更多更透彻地了解学生在网络学习中的行为特点和规律,就能够更好地组织和改进课程设计,引导和服务学生。同时,

有利于对学习资源及学习者做出有效的评价^[3]。

目前学习风格测量方法分为三大类,基于学习风格量表的显式获取法,基于网络学习行为的隐式获取法,以及显式获取为辅隐式获取为主的混合方法。其中,显式获取法需要学习者回答大量题目,加重了学业负担。并且由于该方法无法动态跟踪学习者的状态变化,只能静态表征学习者当前个性特征^[4]。而隐式获取法可以实时跟踪学习者学习行为,及时修正学习风格变化^[5]。然而,该方法的开展需要已知各种学习风格的网络学习行为特征,在方法应用初期容易出现“冷启动”问题。因此,混合测量法吸取上述方法的优势。在学习风格测量初期,使用学习风格量表显式获取学习风格,使用风格已知学习者的网络学习行为初始化学习风格模型。然后利用该模型评估学习者网络学习行为,预测其学习风格^{[6][7]}。

然而,学习风格的差别会造成网络学习行为差异,因此,根据学习者个性可以有效预测网络学习行为。这是由于学习者学习风格是影响网络学习行为的内部因素,而学习环境则是产生网络学习行为的外部因素。在特定的网络学习环境中,将学习风格

【基金项目】武汉大学自主科研项目(人文社会科学)研究成果;武汉大学“70后”学者学术发展计划支持;湖北省教育科学“十二五”规划2011年度重点课题(编号2011A001)“数字校园云平台设计与应用”阶段性研究成果。

【作者简介】吴青,武汉大学教育科学学院讲师,博士,计算机支持的协作学习(联系人);罗儒国,武汉大学教育科学学院副教授。

作为自变量,将网络学习行为作为因变量,研究学习风格与交互行为之间的关联关系是一个合理的推论。然而,影响网络学习行为的因素,除了学习风格外,还包含如前期准备、学习动机等多种干扰因素。如果不排除其他关键影响因素,仅仅根据网络学习行为来判断学习者个性特征,研究结论尚待推敲。此外,工作中学习风格模型的构建完全依赖手工完成,而网络学习者数量庞大,并且学习者在网络学习管理系统中遗留了海量学习行为数据,使得模型计算量过大并不适用于实际教学过程,影响该方法的应用和推广。

鉴于此,本文提出基于混合方法测量学习风格时,首先清理网络学习行为数据,剔除学习准备情况(例如在线学习时间、在线学习成绩等),以及学习参与积极性(例如在线讨论频次等)差异较大的学习者,在构建学习者风格模型时降低干扰数据的影响。并且,借鉴数据挖掘技术在电子商务领域的成熟运用,将该技术应用到学习风格测量过程中,挖掘出网络学习行为与学习风格之间的关联关系,从而获取不同学习风格的网络学习行为特征,由此作为有效判断参与者学习风格的依据。本文重点研究如下问题:

- (1) 哪些网络学习行为直接体现不同学习风格?
- (2) 学习风格模型体现了哪些网络学习行为特征?

二、学习风格理论模型

过去的20年,学习风格理论模型的构建引起研究者的极大兴趣,提出了71款学习风格理论模型,根据理论重要性、使用的广泛程度及其对其他模型构建的影响程度,其中13款作为主要模型^[8]。然而根据Curry的洋葱模型(Onion Model),学习风格模型可以被划分为四大类别:个性模型,注重学习者的个性特征及其对学习过程的影响。信息处理模型,重点在于获取、订购和参与信息的过程。社会交互模式,侧重学习过程中的协作方面。教学偏好模型,专注于学习者对环境、情绪和社会的偏好。虽然对学习风格的概念界定并不统一,但是对学习风格的特征的认识基本是一致的,即学习风格具有独特性、稳定性、兼有活动和个性两种功能。

其中,Kolb的学习风格分类模型按照学习者的知觉维度及其信息加工方式的偏好,用“具体经验/抽象概念”及“主动实验/反思观察”两个维度将学习者划分为四种风格类型,即分散者(Diverge)、聚

敛者(Converge)、同化者(Assimilator)及适应者(Accommodator)^[9]。分散者的学习方式倾向于具体经验和反思观察,具有较强的归纳推理及建立理论模式的能力。同化者的学习方式倾向于反思观察和抽象概括,有较强的想象力和理解能力,喜欢以想象与感觉来解决问题。聚敛者的学习方式倾向于抽象概括和主动实践,喜欢完成实际的计划或任务,善于挖掘理论观点的实际用途,通过不断探索的方法来解决问题、做出结论。适应者则倾向于主动实践和具体经验,喜欢由假设和演绎推理的方式解决问题与制定决策,以亲自实验的方式获得知识,并将理论及想法实际运用。

本文选择该理论模型,一方面由于“第三版Kolb学习风格量表”具有较高的信度和效度,说明该量表具有一定可信度^[10]。另一方面,该模型属于信息处理类型,指出学习风格是学习者在具体经验、观察后反思、形成抽象概念、行动后得到新的经验中的行为表现。有助于教师根据学习者的信息处理偏好,组织和改进课程设计,引导和服务学习者。

三、基于数据挖掘技术的学习风格分析方法

从学习者的网络学习行为数据中,基于数据挖掘技术分析出先前未知的、有效信息,并使用这些信息归纳为网络学习行为特征。主要包括数据获取、数据准备、数据挖掘和结果分析,如图31所示。学习风格挖掘的对象是,网络教学实施过程中学习者遗留在网络学习管理平台中的交互数据,该数据以记录集的形式存放在MySQL数据库中。数据获取阶段通过分析数据存储结构,筛选与主题目标相关的数据。在其他三大阶段中,使用通用的数据挖掘工具,依次处理获取的数据。

(1) 数据获取若缺乏人工指导,数据挖掘技术不会自动发现学习风格特征。因此需要根据学习风格分析需求,掌握网络学习行为数据的格式和特征。评估基于现有资源能否通过数据挖掘技术解决学习风格挖掘问题。

(2) 数据准备

该阶段用于准备数据挖掘需处理的数据。学习者遗留在网络学习平台中的交互数据全部存储在例如MySQL等关系数据库中。然而数据大多零散分布在不同的数据表中,无法直接用于数据挖掘。为保证学习风格挖掘的质量和时效性,充分利用有意义的网络学习交互数据,清除虚假干扰数据,需要开

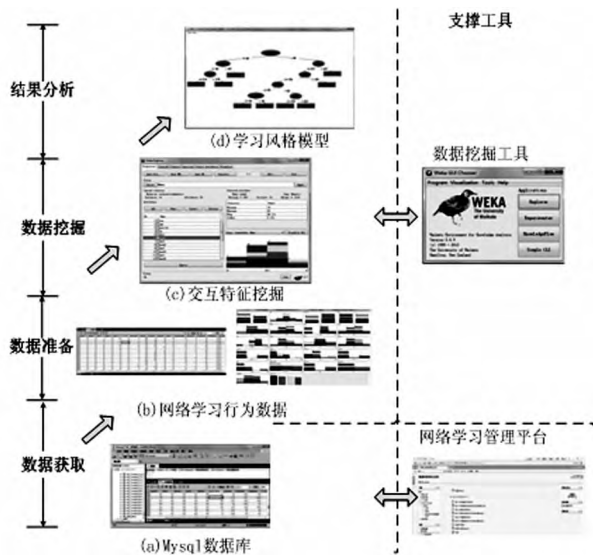


图 3-1 基于数据挖掘的学习风格分析步骤

展数据选择、数据预处理和数据转换等步骤。数据选择就是搜索所有与网络学习交互特征分析有关的数据信息,选择出适用于主题目标挖掘的数据集合。数据预处理就是研究网络学习交互数据的质量,为进一步的用户风格挖掘做准备,并确定将采用的具体挖掘技术。数据转换指将数据进行格式转换,使其适应挖掘软件的处理要求。

(3) 数据挖掘

本文欲解决两个问题,学习风格可以通过哪些网络学习行为描述,以及学习风格模型体现了哪些网络学习行为特征。对于第一个问题,指不是所有网络学习行为属性都可以代表学习者的学习风格。在构建学习风格模型时,为降低无关属性干扰,需要从中挑选出与学习风格关联度较大的一组属性。基于数据挖掘视角,其本质是一个属性选择问题。属性选择是指在初始的 N 个属性中选择出一个有 m ($m < N$) 个属性的子集,这 m 个属性可以像原来的 N 个属性一样用来正确区分数据集中的每个数据对象^[11]。而对于第二个问题,指如何构建有效的学习风格模型,以及该模型的具体涵义。基于数据挖掘视角,其本质是一个模型构建的问题,包括训练和测试两个步骤。在模型训练阶段,基于数据集的特点构造模型,由于本文采用 J48 算法,学习风格模型表现形式为决策树。在测试阶段,利用决策树学习风格模型预测新的数据,测试其分类准确率。

因此,本文的网络学习行为挖掘包括三步,如图 3-2 所示。首先,应用数据挖掘工具的属性选择功能,计算不同属性与学习风格类别之间的关联程度,如图 a 所示。接着,应用数据准备功能,移除相关性较低的属性,如图 b 所示。最后,基于筛选后的属性集合应用分类器功能,通过样本分析,自动构建学习风格模型,如图 c 所示。



图 3-2 学习风格挖掘过程

(4) 结果分析

通过数据挖掘得到的学习风格模型有可能是没有实际意义和价值,或者不能真实反映网络学习交互数据的实际涵义,甚至出现背离实际情况。因此需要对挖掘结果进行模式评估。以验证挖掘结果的有效性,并且最终能够转换成可被用户理解的知识。

四、基于网络学习行为的学习风格挖掘实证研究

(一) 实验样本

本研究以某学院 11 级和 12 级计算机技术专业教育技术方向的工程硕士研究生作为学习风格调查对象,欲从中选择分散型、聚敛型、同化型和适应型学习者各 10 名,共 40 名学习作为最终的研究对象。通过访谈得知,他们大多工作多年,具有一定的计算机操作基础。能够熟练操作计算机并上网进行交互,足以达到本次实验对学习者的计算机技术的要求。并且接受过一年的专业课教育,学习过《软件工程》、

《面向对象程序设计》等计算专业课,具有一定的专业基础,满足本次实验对学习者的专业技能的要求。

为筛选符合条件的学习者,向该专业的广州班和武汉班的学生发放第三版的 Kolb 学习风格测量表,作为学习风格测量问卷。总共发放 89 份,回收 76 份,其中有效问卷 65 份,无效问卷 11 份(2 份未署名,9 份答案雷同),问卷有效回收率是 73.03%。基于问卷反馈信息,将“主动实践/反思观察”作为 X 轴,将“抽象概括/具体经验”作为 Y 轴,运用分层抽样法剔除学习风格不明显的学习者,最终结果如表 4-1 所示。

表 4-1 学习风格分布

取值	样本数(分、同、聚、适)	占比比例
去除 $-12 < x < 12$, $-12 < y < 12$	18(7、4、2、5)	27.7%
去除 $-10 < x < 10$, $-10 < y < 10$	21(4、6、5、6)	32.3%
去除 $-8 < x < 8$, $-8 < y < 8$	44(11、12、10、11)	67.7%
去除 $-6 < x < 6$, $-6 < y < 6$	46(15、13、6、12)	70.8%
去除 $-5 < x < 5$, $-5 < y < 5$	53(22、10、6、15)	81.5%
去除 $-4 < x < 4$, $-4 < y < 4$	55(23、8、5、19)	84.6%

由上表可见,聚敛类型的学习者相对偏少。但考虑到需要平衡各种学习风格人数,选择取值标准为“去除 $-8 < x < 8$, $-8 < y < 8$ ”的一组。即分散学习风格 11 人,同化学习风格 12 人,聚敛学习风格 10 人,适应性学习者 11 人,总计 44 人作为实验样本,占问卷总人数的 67.7%。

(二) 实验环境

因此,本文将开源在线学习平台 Moodle 作为网络学习工作区,将数据挖掘工具 Weka 作为学习风格特征分析工作区,应用场景如图 4-1 所示。在线学习者在 Moodle 平台中产生的网络学习交互数据,例如人口统计学数据、自主学习日志、协作学习日志和测试成绩,作为 Weka 系统的输入。Weka 系统基于数据挖掘技术对样本数据进行特征抽样,构建学习风格模型。运用上述模型,可以基于网络学习交互数据,判断学习者的学习风格。

在网络学习工作区,学习者首先运用 Moodle 平台的用户注册功能,录入人口统计学数据。接着运用在线教程模块开展人-机自主学习,学习者在此

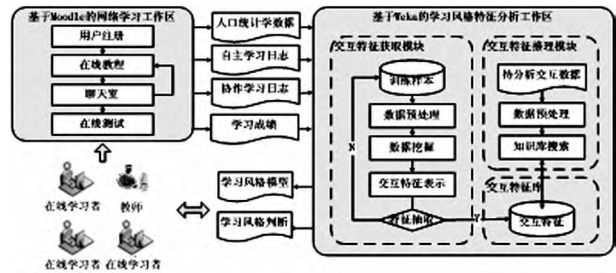


图 4-1 学习风格挖掘的应用场景

过程中在线浏览顺序和时长以及在线作业成绩作为自主学习日志保存在 MySQL 数据库中。与此同时,提供在线聊天室,学习者针对自主学习过程中不懂的问题开展学习互助,讨论过程作为协作学习数据,同样保存在 MySQL 数据库中。最后,通过在线测试模块,获取学习者在线学习效果和知识建构水平。

特别强调的是,在协作讨论过程中,本实验借鉴 EPSILON 系统提供句首词分类方法^[12],采用人工标注的方式,针对学习者讨论内容,标明讨论者的发言意图。该方法起源于 1996 年用于录像分析面对面协作学习,经过 5 年的发展,被应用到在线协作远程学习中。通过 EPSILON 系统的应用和推广,证明了其有效性和合理性。因此,本次实验借鉴其分类方式,将参与者的发言类型划分为情绪表达 (motivate)、信息共享 (inform)、讨论 (discuss)、请求 (request)、任务 (task)、确认 (acknowledge) 和讨论维持 (maintain) 七大类。

而 Weka 系统将存储在 MySQL 数据库中的输入数据划分为训练样本和测试样本。基于训练样本,经过数据预处理、数据挖掘和交互特征表示等步骤,抽取学习风格交互特征,构建学习风格模型。接着,运用该模型判定测试样本的交互特征,在获取学习者学习风格的同时,验证学习风格模型的有效性和正确性。

(三) 实验流程

1. 数据获取

Moodle 是一个开源的课程管理学习系统,可以替代商业在线学习解决方案,在开源许可下免费分发,以帮助教育者建立有效的在线学习社区。学习者在 Moodle 平台上的所有操作,均存储在 MySQL 关系型数据库中,共计数据表三百余张。为获取相关的网络学习行为数据,必须梳理三百多张数据表结构及其之间的关联关系。此项工作较为繁琐。一方面,由于 Moodle 系统的版本不断升级,相关数据

库设计文档并没有完全更新,导致数据库中存在大量无用的或者作用雷同的数据表缺乏说明。另一方面,该系统中表与表之间的联系并没有通过主、外键的方式建立引用关系,而是通过人工理解的形式建立其中的联系,但由于缺乏说明文档,增加了理解难度。

通过全面梳理学生个人信息以及在线学习过程,分析学生的学习风格并不需要分析所有表中存储的信息。存储学习风格相关信息的数据表结构如

图 42 所示。其中,学生个人信息,例如姓名、所在城市等存储在 user 表中。学生参与在线教程信息存储在以 lesson 作为前缀的一系列表中,主要包括学生浏览教程页面花费的时间,参与在线练习的时间以及分数。学生参与在线聊天时发送的消息等信息存储在以 chat 为前缀的表中,包括消息内容、消息类型以及消息接收者等。而学生在线测试的过程及其成绩存储在 quizz 为前缀的表中,包括在线测试花费的时间和最终成绩信息。

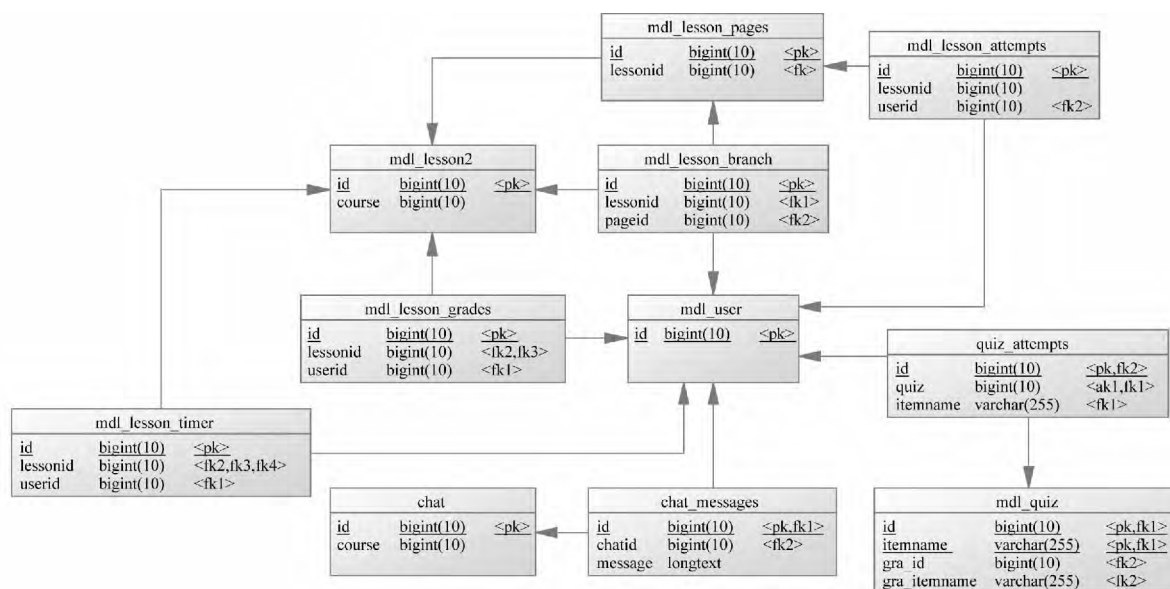


图 4-2 MYSQL 数据库中相关表结构

2. 数据准备

数据准备包括数据选择和数据预处理等操作,占整个挖掘过程的极大比重,其数据处理质量将极大影响最后挖掘结果,是数据挖掘的核心步骤。本阶段需要对图 42 所列表中的属性进行选择,对表中数据进行“去噪”操作,以保证数据能够真实反映学习风格。所以,本阶段的目的是为提高数据挖掘的质量,识别目标数据集,缩小挖掘范围。

本实验将网络学习行为数据主要划分为人口统计属性、人机自主学习属性和人人协作学习属性和最终测试成绩属性四大类型。

由于网络学习主要针对远程教育学习者,与全日制在职学生不同的是,这部分学生年纪跨度比较大,可支配学习的时间不同,专业基础存在差异。因此,需要收集这些学生的人口统计信息,能够有效帮助系统识别学生的学习能力和学习基础。其属性包括性别 sex、年龄 age、是否已婚 marital、专业背景

bg。系统通过记录学习者在线浏览教学资源所花费的时间以及在线答题成绩,作为学习者自主学习进程和知识掌握状态的主要标识。

本实验选择“关系数据库设计”章节作为自主学习课程,由“数据库设计概述”、“概念结构设计”、“逻辑结构设计”和“案例练习”四小节组成,每部分均提供相应的在线答题,以测试知识点的理解和掌握情况。目的在于培养学生的抽象思维能力、逻辑思维能力以及知识的灵活应用能力。因此,人机自主学习属性包括 dur1(第一节学习时长)、mark1(第一节在线作业成绩)、dur2(第二节学习时长)、mark2(第二节在线作业成绩)、dur3(第三节学习时长)、mark3(第三节在线作业成绩)、dur4(第四节学习时长)、mark4(第三节在线作业成绩)。

协作讨论属性反映了学习者在相互协作讨论,共同解决知识建构困难过程中,各种类型发言的总字数,体现学习者对协作学习讨论的贡献程度和专

注程度。其属性包括 motivate(情绪类型发言的总字数)、inform(信息共享发言的总字数)、discuss(讨论类型发言的总字数)、request(请求类型发言的总字数)、task(任务类型发言的总字数)、acknowledge(确认类型发言的总字数)、maintain(维持类型发言的总字数)、frequency(发言总次数)。

最终测试成绩属性是指学生在完成在线教程、在线讨论后,为测试学习效果,在制定时间区域内参加在线测试所获得的分数。属性表达为 finalMark(最终测试成绩)。

在数据选择阶段,创建新表,包含上述属性的以及学习风格属性 learnStyle。运用结构化查询语言(Structured Query Language)统计每一位学习者在上述属性的取值,将统计结果插入新表。并且利用学习风格调查的结论,为每位学习者的学习风格赋值,其中 ass 表示同化型学习者、div 表示分散型学习者、con 表示聚敛型学习者、acc 表示适应型学习者,从而获得完整数据集。结果如图 31 中(b)左侧所示。运用 Weka 工具访问该数据表,获取网络学习行为数据。

在数据预处理阶段,由于学习者网络学习行为受到学习者个性特征、知识基础、前期准备、学习积极性等多重因素的影响,因此需要排除其他因素影响过大的学习者,减少无关样本对学习风格测量的影响。然而,基于数据挖掘的视角,将数据集中少量与数据的一般行为不一致的数据对象称为异常点(Outlier),即孤立点^[13]。异常点的检测和分析称之为异常点挖掘。本文应用运用最近邻居法,对 44 名学习者样本进行异常点挖掘,以发现样本中反常实例或不满足规则的特例。该方法是一种实例学习法,计算学习者之间的向量空间距离来评估相似度,将相似度高的学习者聚集在一起形成簇,而游离在簇之外的学习者设定为异常实例。应用 Weka 提供的 IBk 算法,共筛选出 8 名“疑似异常点”,通过人工二次筛选,确定其中 5 名学习者的网络学习行为无法体现学习风格。理由如下:

两名学习者并非计算机相关专业本科毕业,由此可见其专业功底扎实的可能性较小。在自主学习阶段浏览网页的时间较短,但是作业成绩却比较好,在协作讨论过程中也没有任何发言,然而最终成绩却比较优秀。由此可见,该生可能出现在线作业和测试抄袭的情况。一名学生在自主学习阶段表现一般,作业成绩比较糟糕,协作讨论过程中发言极少,

然后最终成绩较之作业成绩反差过大。虽然不排除该生在课后可能选取其他教学资源或者将学习资料以图片的方式下载到本机开展学习,但考虑该生 31 岁,女性而且已婚,由于该阶段可能肩负抚养幼儿的重担,可支配的学习时间不会特别充裕。有理由认为该生或许在测试阶段并非独立完成,因此作为异常点排除。两名学生年龄偏大均超过 35 岁,在自主学习阶段花费的时间较少,协作学习阶段基本没有做出任何贡献,无论在线作业或在线测试都表现不佳。可能由于平时工作任务过于繁重,对于学业要求不高,并没有投入恰当的时间和精力参与学习,因此作为异常点排除。

因此,剔除上述 5 位学习者,剩余 39 名学生,其中分散型 10 人、同化型 9 人、聚敛型 9 人、适应型 11 人。利用 Weka 提供的 visualize all(可视化功能),直观看到所有数据的分类汇总结果,如图 31 中(b)右图所示。

3. 数据挖掘

(1) 哪些网络学习行为直接体现不同学习风格?

由于参与本次研究的学习者数目有限,缺乏大量数据参与训练和测试,因此在属性选择和模型构建时,均采用 10 折交叉验证的方法。即将数据大致分成十个分区,每个分区被轮流用于测试而剩余的则被用于训练。重复此过程十次,从而每个实例恰好有一次用于测试。

基于 Weka 的属性选择功能,输出如图 32(a)所示,属性选择输出为:0(0%) 1 sex、7(70%) 2 age、1(10%) 3 marital、1(10%) 4 bg、1(10%) 5 dur1、1(10%) 6 mark1、6(60%) 7 dur2、2(20%) 8 mark2、2(20%) dur3、0(0%) 10 mark3、6(60%) 11 dur4、2(20%) 12 mark4、8(80%) 13 motivate、10(100%) 14 inform、10(100%) 15 discuss、9(90%) 16 request、8(80%) 17 task、3(30%) 18 acknowledge、3(30%) 19 maintain、100(100%) 20 frequency、1(10%) 21 finalMark。其中 0(0%) 1 sex 代表在 10 折交叉验证中,第一个属性 sex 被选中 0 次,被选中概率为 0%。

因此,选择选中概率大于 60% 的属性,作为直接体现学习风格的属性,包括 age、dur2、dur4、motivate、inform、discuss、request、task、motivate、frequency。其中 age 属于人口统计属性,dur2 和 dur4 属性属于自主学习属性,其他属性均属于协作学习属性。由此可见协作学习属性体现了不同的学习风格。相比

协作学习属性,人口统计属性和自主学习属性并不能作为学习风格的显著性因素,无法体现不同学习者之间的学习风格差异。

(2) 学习风格模型体现了哪些网络学习行为特征?

在教育数据挖掘领域中,J48 分类算法已被广泛使用:例如发现潜在的学生群体对特定教学策略,具有类似的特性和反应^[14] 预测学生的表现和他们的最终等级;^[15] 检测学生的误用以及学生如何形成组;^[16] 预测学生的表现,以及评估相关的属性;^[17] 作为提示或反例对学生分组,以及发现学生常见的错误;^[18] 确定学生缺乏学习动力时,并找出补救行动,以降低辍学率;^[19] 预测学习过程成功^[20]。

因此本文利用 J48 算法,使用前文筛选后的属性集合,采用 10 折交叉验证,反复训练学习风格模型,发现当参数 confidentFactor = 0.3, minNumObj = 2 时,正确率最高,可以达到 76.92%。因此,构建的学习风格模型如图 4-3 所示。

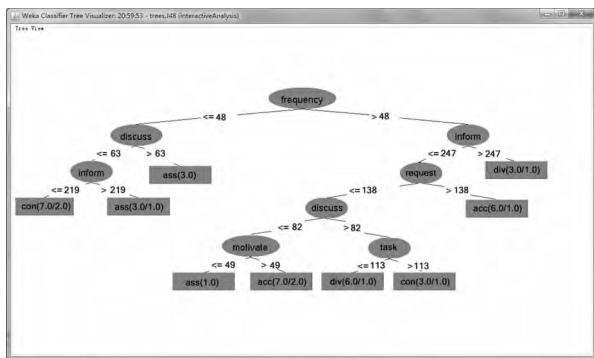


图 4-3 基于 J48 算法的学习风格模型

为了更加深度理解决策树,将其转为分类规则。从决策树上获取一组规则十分容易,每片叶子都可以产生出一条规则。规则的先决条件包含了从根到叶子路径上所有节点的条件,规则的结论是叶子上标注的类。规则的描述方式为“如果……那么……”,例如描述同化型学习者的规则一可以描述为“如果交互总次数小于等于 48 且讨论字数小于等于 63 时,那么为同化型学习者。其他规则描述类似,后文不再赘述。

同化型学习者相关规则集:

- ① if frequency < = 48 and discuss < = 63 then learningStyle = “ass”
- ② if frequency < = 48 and discuss > 63 and inform > 219 then learningStyle = “ass”
- ③ if frequency > 48 and inform < = 247 and re-

quest < = 138 and discuss < = 82 and motivate < = 49 then learningStyle = “ass”

聚敛型学习者相关规则集:

- ① if frequency < = 48 and inform < = 63 and inform < = 219 then learningStyle = “con”
- ② if frequency > 48 and inform < = 247 and request < = 138 and discuss > 82 and task > 113 then learningStyle = “con”

分散型学习者相关规则集:

- ① if frequency > 48 and discuss > 247 then learningStyle = “div”
- ② if frequency > 48 and inform < = 247 and request < = 138 and discuss > 82 and task < = 113 then learningStyle = “div”

适应型学习者相关规则集:

- ① if frequency > 48 and inform > 247 and request > 138 then learningStyle = “acc”
- ② if frequency > 48 and inform > 247 and request < = 138 and discuss < = 82 and motivate > 49 then learningStyle = “acc”

(四) 实验结果分析

1. 学习风格可以通过哪些网络学习行为描述?

由于 Kolb 学习风格理论模型建立在“学习是通过经验的转化创造知识的过程”的假设上。由于人口统计属性主要体现了学生作为人类个体的自然属性和社会属性,与知识学习关联不大,因此大部分属性不能作为决定学习风格的关键属性。最终学习成绩作为学习的最终结果,并没有过程性体现,并且与学习者的专业基础、学习积极性、学习投入程度等因素息息相关。因此,无法作为判断学习风格的关键指标。虽然自主学习属性体现了学生参与过程,但在线作业成绩作为阶段性学习成果,较多体现了知识建构发展程度,与学习风格的关联较弱。

发言次数 frequency 属性是区分主动实验或反思观察的关键指标,作为区分学习者风格的关键影响因素。发言次数较多说明学习者对于讨论交流具有较高的积极性,愿意参与协作讨论来共同构建知识。该类学习者具有一定信息加工偏好。发言次数较少则说明学习者更倾向倾听他人的意见,独立思考问题,不太愿意参与协作学习,此类学习者偏好通过独立思考,或者独自探寻问题的答案。与 Kolb 的理论一致, Kolb 认为发散型和适应型的学习者具有依靠具体经验的特点,该类学习者强调由情感体验

和实际经验而导致的学习,善于与学习伙伴间的沟通和交流。而聚合型和同化型的学习者具有抽象概括的特点,抽象概括的学习者注重对符号的理解,擅长在非人际环境中以及权威指导下的学习。

讨论类型发言总字数 discuss 和信息共享类型发言总字数 inform 则侧重体现学习者在“具体经验/抽象概念”维度的差异。探讨类型的发言主要记录学习者对知识的深度加工讨论过程,体现了知识建构中认知争辩过程,属于深层次交互。信息共享类型发言主要记录学习者相互分享教学资源过程。CSCL 交互的研究者大多将讨论发言字数作为交互贡献的重要指标,因此上述两种发言类型的总字数体现了不同学习风格的学习者对不同讨论深度的贡献层次。

请求类型发言总字数 request、任务类型发言总字数 task 和情绪类型发言总字数 motivate 同样对于区分学习风格具有重要作用。请求类型发言标志学习者是否愿意发布自己的困惑到小组,向他人求助。而情绪类型发言大多与学习目标和主题无直接关联。但体现了学习者比较容易表露自己的情感。而任务类型发言记录了学习者对任务的组织和总结能力,也属于深层次交互。

年龄 age 同样体现了学习者学习风格。这是由于学习者随着年纪增长、知识积淀和个人阅历的增加,学习风格可能出现动态变化,但具体风格转变趋势尚不清楚,有待进一步探讨。因此,年龄作为人口统计属性中唯一可能体现学习风格的属性。

学习者在第二节和第四节的自主学习时长 dur2 和 dur4,同样体现了学习者学习风格。这可能由于不同风格学习者的学习偏好导致的。部分风格学习者喜欢独自阅读、会花费更多的时间开展自主学习。而另一部分风格学习者更喜欢通过交流协作的方式,开展协同学习,因此花费在独立自主学习上的时间相对较少。然而,哪些风格学习者偏好自主学习尚不可知,有待进一步分析。

由此可见,体现学习风格的 CSCL 交互行为特征的属性包括年龄、第二节和第四节自主学习时长、发言次数、讨论类型发言总字数、信息共享类型发言总字数、请求类型发言总字数、任务类型发言总字数和情绪类型发言总字数。

2. 学习风格模型体现了哪些网络学习行为特征?

由学习风格模型可知,同化型学习者和聚敛型

学习者发言次数较少,相对于其他类型的学习者而言,参与协作学习的积极性不高。这一点符合 Kolb 的学习风格理论。Kolb 指出这两种类型的学习者更加注重抽象概括能力,习惯独立思考,擅长在非人际环境中以及权威指导下的开展自主学习。但由于自主学习属性受到其他因素影响,没有在构建的学习风格模型中体现出来。因此无法判断是否同化型学习者和聚敛型学习者在在线浏览和在线作业中花费更多的时间。

同化型学习者和聚敛型学习者相比较,后者更倾向于在倾听他人的讨论后,梳理思路提出自己的推论和总结。这点可以从判断分支中,聚敛型学习的讨论类型和任务类型发言字数总是相对偏大可以看出。这与 Kolb 的学习风格理论中提到的“有能力发现问题、解决问题并做出决策”相符合。然而由于同化者在交互讨论过程中发言次数较少,各种类型的发言总字数也是相对偏少,对其交互行为特征缺乏明显的判断依据。

分散型学习者和适应型学习者在协作学习讨论中处于积极主导地位。然而,适应型学习者在请求问题类型发言字数较多,该风格学习者倾向于在公共交流平台提出不懂的问题,需求他人帮助,并且积极参与讨论寻求问题的答案。这说明该类学习者可以充分利用协作交流平台的便利性,发挥协作学习的优势,通过积极探讨实现知识的协同构建,加深理解知识点。而分散型学习者则在教育资源共享方面贡献较大,体现了“喜欢收集信息,对人感兴趣,喜欢小组学习”的学习特性。但此类学习者很少参与深层次交互,缺乏信息的总结和归纳能力。

然而,令人意外的是,属性 age、dur2、dur4 并没有作为学习风格模型的状态节点出现,这是由于基于该属性判断学习风格的准确率不高,因此在构建决策树学习风格模型时被减枝。出现该状态的原因可能在于,协作学习属性相比人口统计属性和自主学习属性,更能体现学习风格。因此,学习风格模型只体现相关性较强的部分属性,而对于关联较弱属性并没有显式表示。

五、总结

本文基于数据挖掘技术,应用混合方法测量网络学习者的学习风格。首先,利用第三版的 Kolb 学习风格测量表,显式获取学习者学习风格。其次,运用最近邻居法进行异常点挖掘,减少不真实噪声样

本对学习风格测量的影响。接着,利用属性选择功能,筛选出与学习风格具有直接关联的属性集合。最后运用 J48 算法,构建决策树学习风格模型。该模型符合实际情况,遵从一般教学规律,并且预测正确率达到 76.92%。然而,研究过程中也发现一些问题,不少学习者需要指导教师不断鼓励和引导才愿意进入在线聊天室开展协作学习。而部分学习者依然习惯将网络协作学习变成网络独自学习,因此对学习风格模型的预测精度造成一定影响。通过本研究成果,可以从网络学习系统中遗留的海量交互数据中挖掘学习者学习风格,便于学生增强自我认知能力,便于指导教师快速掌握学习者的个性特征。该方法的实施并非局限与 Moodle 平台,而是适用于具有在线学习和在线聊天功能的学习平台,具有一定普适性。下一步研究重点在于,如何结合学习风格特征,实施协作学习异质分组自动化,促进在线学习交互行为,激发学习者的学习热情,实现成员之间的优势互补,从而获取更加高效的协作学习成果。

【参考文献】

- [1] D. H. Jonassen , B. L. Grabowski . Handbook of Individual Differences: Learning and Instruction [M] , Hove: LEA. 1993.
- [2] 吴鹏泽, 远程学习者内部学习动机的激发策略[J]. 现代远程教育, 2011. 2: 015.
- [3] 刘儒德, 江涛, 学习者特征对网络学习的影响[J]. 中国电化教育, 2004. 6(11): 67 - 72.
- [4] E. Triantafillou , A. Pomportsis , S. Demetriadis. The Design and the Formative Evaluation of an Adaptive Educational System Based on Cognitive Styles. Computers & Education , 41 , 2003: 87 - 103.
- [5] D. Wen , S. Graf , C. H. Lan , T. Anderson , Kinshuk , K. Dickson. Adaptive Assessment in Web - based learning [C]. Proc. IEEE International Conference on Multimedia and Expo (ICME 2007) 2007: 1846 - 1849.
- [6] 姜强, 赵蔚与王朋娇, 基于网络学习行为模式挖掘的用户学习风格模型建构研究[J]. 电化教育研究, 2012 (11): 55 - 61.
- [7] N. Stash Incorporating Cognitive/Learning Styles in a General - Purpose Adaptive Hypermedia System. [D] Eindhoven University of Technology , Netherlands. 2007.
- [8] F. Coffield , D. Moseley , E. Hall , K. Ecclestone Learning Styles and Pedagogy in Post - 16 Learning. A Systematic and Critical Review. Learning and Skills Research Centre , UK. 2004a: 231 - 245.
- [9] Kolb , A. Y. , The Kolb learning style inventory - version 3.1 2005 technical specifications. Boston , MA: Hay Resource Direct , 200
- [10] Felder , R. M. and J. Spurlin , Applications , reliability and validity of the index of learning styles [J]. International Journal of Engineering Education , 2005. 21(1) : 103 - 112.
- [11] Kohavi , R. and G. H. John , Wrappers for feature subset selection [J]. Artificial intelligence , 1997. 97(1) : 273 - 324.
- [12] Soller , A. , Computational modeling and analysis of knowledge sharing in collaborative distance learning. User Modeling and User - Adapted Interaction , 2004. 14(4) : 351 - 381.
- [13] BARNETT V , LEWIS T. Outliers in statistical data [M] . 2nd. New York : John Wiley & Sons , 1994.
- [14] Chen , G. , Liu , C. , Ou , K. , & Liu , B. Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology [J]. Journal of Educational Computing Research , 2000: 23(3) , 305 - 332.
- [15] Minaei - Bidgoli , B. , & Punch , W. Using genetic algorithms for data mining optimization in an educational web - based system [C]. In Genetic and evolutionary computation conference , Chicago , USA 2003 : 2252 - 2263.
- [16] Baker , R. , Corbett , A. , & Koedinger , K. Detecting student misuse of intelligent tutoring systems [C]. In Intelligent tutoring systems , Alagoas , Brazil 2004 : 531 - 540.
- [17] Kotsiantis , S. B. , Pierrakeas , C. J. , & Pintelas , P. E. Predicting students' performance in distance learning using machine learning techniques [J]. Applied Artificial Intelligence , 2004: 18(5) , 411 - 426.
- [18] Yudelson , M. V. , Medvedeva , O. , Legowski , E. , Castine , M. , Jukic , D. , & Rebecca C. () . Mining student learning data to develop high level pedagogic strategy in a medical ITS [C]. In Proceedings of AAAI workshop on educational data mining , Boston 2006 : 1 - 8.
- [19] Cocea , M. , & Weibelzahl , S. Can log files analysis estimate learners' level of motivation? [C] In Proceedings of the workshop week Lernen - Wissensentdeckung - Adaptivitat , Hildesheim 2006 : 32 - 35.
- [20] Hamalainen , W. , & Vinni , M. Comparison of machine learning methods for intelligent tutoring systems [C]. In Proceedings of the eighth international conference in intelligent tutoring systems , Taiwan. 2006 : 525 - 534.

(本文责任编辑: 闫 兵)