

# 基于数据挖掘的网络学习行为与学习效果研究

傅钢善<sup>1</sup>, 王改花<sup>2</sup>

(1.陕西师范大学 现代教学技术教育部重点实验室, 陕西 西安 710062;

2.西安职业技术学院 资产处, 陕西 西安 710077)

[摘要] 随着计算机与网络的快速发展和广泛应用,网络学习已经是学校教育教学的重要组成部分。已有研究表明,网络学习行为与学习效果密切相关,因此探究网络学习行为与学习效果对有效开展网络学习具有重要意义,也是目前研究的热点与趋势。本研究以陕西师范大学参加“现代教育技术”网络学习系统学习的2801名学习者作为研究对象,以数据库数据作为网络学习行为特征数据来源,采用数据挖掘方法与统计学方法对网络学习者的行为特征进行定量分析,并探讨了网络学习行为特征与学习效果的关系。

[关键词] 网络学习; 行为; 学习效果; 数据挖掘

[中图分类号] G434 [文献标志码] A

[作者简介] 傅钢善(1955—),男,陕西富平人。教授,主要从事网络与远程教育、信息技术教育应用、教育信息化理论与实践研究。E-mail: fughsh@snnu.edu.cn。

## 一、问题的提出

根据中国互联网络信息中心第33次全国互联网发展状况统计报告显示,截至2013年12月31日,中国网民数量达6.18亿。<sup>[1]</sup>网络已经渗透到经济发展和社会生活的各个方面,人们的生产、生活方式以及学习方式正在发生深刻的变化。<sup>[2]</sup>网络学习已经是学校教育教学的重要组成部分,并且正在成为未来教育信息化发展的重要任务之一。因此网络学习也已经引起越来越多研究者的关注,成为研究的热点。研究者发现网络学习行为与学习效果密切相关,因此研究网络学习行为特征及其与学习效果的关系,有助于教师与课程开发者更好地了解网络学习者特点,制定有针对性的教学策略,进行科学的教学评价,提供有效的学习支持服务,并为适应性学习系统的设计与开发提供依据,从而使学习者达到有效学习。

## 二、文献综述

目前关于网络学习行为的研究主要有以下特点。

第一,研究内容主要包括网络学习行为内涵的界

定、网络学习行为影响因素、网络学习行为现状调查、网络学习行为评价等。<sup>[3]</sup>很少有研究者从群体差异角度来研究不同性别、不同学历、不同学科背景的学习者的学习行为差异。但是差异是普遍存在的,了解不同学习群体的行为特征有助于教师进行决策,有助于为学习者提供更加有效的支持服务。而且,关于学习行为与学习效果的关系研究也甚少,结论的可信度也有待进一步验证,例如台湾研究者刘惠如探讨了网络学习行为与学习效果的关系,研究表明登录次数、浏览总时间、浏览页数、在线测验次数、参考资源浏览次数、讨论区发帖次数等学习行为均与学习效果有显著的正相关。<sup>[4]</sup>但台湾研究者杨奕晨和柴蕙质的研究表明登录次数、参与讨论次数等因素对学习效果没有显著影响。<sup>[5]</sup>

第二,国内研究方法单一,主要是文献分析与调查研究。但问卷调查存在一定的局限性,学习者可能有意地输入不准确的数据或完成问卷需要花费时间,增加了学习者的时间负担,可能存在胡乱填答的情况。<sup>[6]</sup>近年来国外越来越多的研究利用学习者的Web日志文件、网上留言信息、数据库数据来进行定量分

析,但国内利用这些方法开展研究还相对薄弱。

因此,本研究尝试以数据库数据作为网络学习行为特征数据来源,采用数据挖掘方法与统计学方法对网络学习者的行为特征进行定量分析,并探讨网络学习行为特征与学习效果的关系。

### 三、研究方法

#### (一)研究框架

本研究中所提到的学习行为特征指网络学习系统数据库中所记录的学习者参与网络学习时的行为。本研究对学习行为特征进行以下几方面探讨:

1. 网络学习行为特征总体情况;
2. 不同性别、不同学科背景、不同学历学习者在网络学习行为特征方面有何差异;
3. 网络学习行为与学习效果的关系。

#### (二)研究对象

本研究的研究对象为陕西师范大学 2012~2013 学年第二学期参加“现代教育技术”网络学习系统学习的 2687 名 2009 级大三本科生和 210 名成人教育硕士,共 2936 人(16 个学院、20 个专业)。通过数据挖掘前期数据预处理,研究对象共剩余 2801 人,其中性别分布为男生 943 人,女生 1858 人;学历分布为本科生 2657 人,硕士生 144 人;学科背景分布为理科生 1399 人,文科生 995 人,艺术生 407 人。

#### (三)网络学习系统介绍

陕西师范大学“现代教育技术”网络学习系统主要由课程学习、辅导资源、媒体学习、讨论交流、软件下载和其他辅助功能(优秀作业、信息订阅、学习日记、作业管理、给教师留言)等模块组成。课程采用网上学习与专题讲座相结合的混合学习方法。考核采用过程性评价和总结性评价相结合的多元评价方式。

#### (四)研究工具

本研究采用 SQL Sever 2008 Data Mining 数据挖掘工具对数据库中的信息进行处理,采用 SPSS17.0、Excel2007 进行统计分析。

“数据挖掘又称数据库中的知识发现,就是从大量数据中获取有效的、新颖的、潜在的、有用的、最终可理解的模式的非平凡过程,简单地说,数据挖掘就是从大量数据中提取或挖掘知识。”<sup>[7]</sup>数据挖掘的基本步骤一般包括问题定义、数据准备、数据预处理、数据挖掘、模型评价。<sup>[8]</sup>

“现代教育技术”网络学习系统数据库共有 40 余张数据表,本研究首先选取了学习者基本特征、学习者行为特征、学习者学习效果等三类数据表,其次采用

SQL 语言对数据进行预处理,包括数据清洗、数据集成、数据变换、数据规约。最终确定了九种行为特征:

- (1)学习时间跨度,即学习者从第一次登录系统到最后一次退出系统的时间周期;
- (2)学习总时长,即学习者在网络学习系统上累计的学习时间长短;
- (3)学习次数,即学习者登录网络学习系统进行学习的次数;
- (4)平均每次在线学习停留时长,即学习者平均每次学习持续的时间长度;
- (5)重复学习率,即学习者平均每个知识点重复学习的次数;
- (6)讨论交流,即学习者参与 BBS 讨论交流的情况,由发主题帖(数量、点击量、被回帖数量)、回帖(数量、点击量)共同构成;
- (7)学习笔记,即学习者参与在线学习笔记的情况,由在线学习笔记的数量、更新量、点击量共同构成;
- (8)接收短信数量,即学习者在学习过程中接收督促学习短信的数量;
- (9)学习效果,即学习者总成绩,由 10%网上作业成绩、20%实践技能成绩、20%网上考试成绩、50%期末笔试成绩共同构成。

### 四、结果分析

#### (一)总体情况

学习者行为特征总体情况见表 1。

##### 1. 学习时间跨度分析

学习时间跨度的范围在 14~128 天,均值为 109.54 天,主要分布在 100~128 天。系统规定的学习时间是 2013 年 2 月 22 日至 2013 年 6 月 6 日,最小时间跨度是 105 天,78%的学习者的学习时间跨度达到了系统规定时间。

##### 2. 学习总时长分析

学习总时长的范围在 10.2~118.5 小时,均值为 28.5 小时,主要分布在 20~40 小时之间。系统规定的最低学习总时长是 20 小时,99.7%的学习者达到了系统规定的学习时长。

##### 3. 学习次数分析

学习次数的范围为 12~190 次,均值为 51 次,主要分布在 30~70 次。系统规定的最少学习次数是 18 次,99.8%的学习者达到了系统规定的学习次数,且学习次数明显高于系统规定次数。

##### 4. 平均每次在线学习停留时长分析

平均每次在线学习停留时长范围为 6~109 分,均

值为 35.3 分,主要分布在 20~50 分钟。

#### 5. 重复学习率分析

重复学习率的范围为 0.41~8.84,均值为 1.75,主要分布在 1~3 次之间。

#### 6. 学习笔记(Blog)分析

发表学习笔记的人数为 2202 人,未发表人数为 599 人,所有学习者共发表 21908 篇 Blog,平均每篇 Blog 被浏览 37 次。Blog 范围为 0~0.95,均值为 0.57,发表学习笔记的学习者分布符合正态分布。其中每人发表学习笔记数量的范围为 0~192 篇,均值为 10 篇,主要分布在 0~15 篇;每人学习笔记更新量范围为 0~79 次,均值为 3 次,主要分布在 0~5 次;每人学习笔记点击量的范围为 0~786 次,均值为 38 次,主要分布在 0~40 次。

#### 7. 讨论交流(BBS)分析

BBS 范围为 0~0.98,均值为 0.50,主要分布在 0.3~0.4 之间,此类人是未参与 BBS 讨论交流的学习者,占总数的 42.4%。

从发主帖情况看,每人发主帖数量范围为 0~32 条,均值为 0.71 条,813 人发表了 BBS 主帖,未发表 BBS 主帖人数为 1988 人(占总人数的 71%)。所有主帖被点击 75348 次,平均每条主帖被点击 38 次,最多被点击 540 次,最少 0 次;每人发表的主帖最多被回帖 435 条,最少 0 条,平均每条主帖被回帖 9.5 条,其中有 151 人的主帖没有得到回复。

从回帖情况看,每人回帖数量范围为 0~230 条,均值为 6.75 条,1519 人参与了 BBS 回帖,未参与回帖的人数为 1282 人(占总数的 45.8%),共有 18901 条回帖。回帖共被点击 1405741 次,平均每条回帖被点击 74 次,点击量范围为 0~538 次。

#### 8. 短信接收数量分析

短信接收数量范围为 2~14 条,均值为 3.7 条,大部分学习者接收了 2 条短信。

表 1 学习者行为一般特征(N=2801)

行为特征	均值	标准差	极小值	极大值
学习时间跨度(天)	109.54	12.654	14	128
学习总时长(时)	28.49	7.15	10.21	118.48
学习次数(次)	51.32	17.64	12	190
平均每次在线学习停留时长(分)	35.30	10.84	6	109
重复学习率(次)	1.75	0.669	0.41	8.84
学习笔记(分数秩)	0.57	0.195	0	0.98
讨论交流(分数秩)	0.50	0.205	0	0.95
接收短信数量(条)	3.70	1.828	2	14

## (二)分类分析

### 1. 性别差异

为了了解不同性别学习者在行为特征上的差异,本研究对不同性别学习者的行为特征使用独立样本  $T$  检验进行比较,结果表明:

(1)男女生在平均每次在线学习停留时长、学习总时长、重复学习率、学习笔记、讨论交流、短信接收数量等六方面均有显著性差异( $P<0.01$ ),且在平均每次在线学习停留时长、学习总时长、重复学习率、学习笔记、讨论交流等五方面女生均值明显高于男生,接收短信数量男生明显高于女生;

(2)男女生在学习时间跨度、学习次数方面没有显著性差异( $P>0.05$ )。

总体看女生学习行为参与度明显高于男生。

### 2. 学历差异

为了了解不同学历学习者在行为特征上的差异,本研究对不同学历学习者的行为特征使用独立样本  $T$  检验进行比较,结果表明:

(1)本科生与硕士生在学习时间跨度、学习总时长、学习笔记、讨论交流等四方面均有显著性差异( $P<0.01$ ),在学习时间跨度、讨论交流方面本科生明显高于硕士生,学习总时长、学习笔记方面硕士生明显高于本科生;

(2)本科生与硕士生在学习次数、平均每次在线学习停留时长、重复学习率、短信接收数量等四方面没有显著性差异( $P>0.05$ )。

### 3. 学科背景差异

为了了解不同学科背景学习者在行为特征上的差异,本研究对不同学科背景学习者的行为特征使用单因素方差方法进行比较,结果如下:

(1)不同学科背景学习者在学习时间跨度、学习总时长、平均每次在线学习停留时长、重复学习率、学习笔记、讨论交流、短信接收数量等七方面均有显著性差异( $P<0.01$ );

(2)不同学科背景学习者在行为特征上没有显著性差异( $P>0.05$ )。

## (三)行为特征与学习效果关系

为了了解网络学习行为特征与学习效果之间的关系,本研究对学习者的各行为特征与学习效果进行相关分析,结果见表 2。结果表明:学习效果与讨论交流、学习次数、学习时间跨度、学习笔记、重复学习率、学习总时长均呈显著正相关,与短信接收数量、平均每次在线学习停留时长呈显著负相关。存在显著相关的变量相关系数介于-0.209 至 0.263 之间,即均是低度相关。



表 2

行为特征与学习效果相关分析

	学习时间跨度	学习总时长	学习次数	平均每次在线学习停留时长	重复学习率	学习笔记	讨论交流	接收短信数量
学习效果	.137**	.068**	.085**	-.054*	.184**	.263**	.252**	-.209**

(\* $p < 0.05$ , \*\* $p < 0.01$ )

## 五、小结与讨论

### (一) 小结

#### 1. 总体情况小结

早在 1989 年,著名远程教育专家 Moore(1989)提出远程教育的交互包括师生交互、生生交互和学生与学习内容之间的交互,且这一理论一直被国内外研究者广泛认可。通过分析发现,本研究中讨论交流与留言数量是人与人之间交互的外在表现,能够一定程度反映师生互动与生生互动;学习时间跨度、学习总时长、学习次数、重复学习率、平均每次在线学习停留时长、学习笔记是学生与学习内容之间交互的外在表现,能够一定程度反映学生与学习内容之间的交互。

研究表明,学习者在学生与学习内容之间的交互方面表现较好。因为本网络课程学习采用多元评价方式,对学习者的学习时间跨度、学习总时长、学习次数均有最低要求,大部分学习者能够按照网络学习要求和计划完成学习任务。通过问卷调查与访谈发现,大多数学习者表示在网上学习时容易受到网上无关信息的干扰或受电脑辐射影响容易造成视觉疲劳,因此注意力集中时间有限,平均每次在线学习停留时长主要分布在 20~50 分钟,且需要增加学习次数与重复学习次数或记录学习笔记来达到学习效果。

研究表明,学习者在师生交互、生生交互方面表现相对较差。笔者分析主要有以下原因。(1)用户因素,包括学习者的学习风格、动机、态度、学习策略、情感等因素。(2)网络学习系统因素。网络学习缺乏面对面的情感交流。(3)教师因素。本次参与网络学习的人数达 2963 人,而主讲教师只有 3 位,辅导教师 25 位,因此教师没有太多的时间主动引导学习者参与讨论交流。(4)其他因素,包括评价、网络学习环境等。该网络课程学习对学习者的交流互动并没有强制要求,也没有将其作为评价的指标,因此学习者的外部动机不足;部分学习者没有电脑,网络学习条件较差。

短信接收数量反映了学习者的自我调控能力,大部分学习者在网络环境下学习的自我调控能力较弱,需要外部的督促、鼓励等调控。

#### 2. 分类分析小结

##### (1) 性别差异原因分析

调查发现,在学习过程中,女生往往要比男生自觉、认真、努力,大多数女生会按照学习要求或教师建议积极参与网络学习,因此女生比男生的行为参与度高,且对外部调控的需求低于男生。

##### (2) 学历差异原因分析

参与网络学习的硕士是成人联考硕士,他们肩负着工作和家庭的双重责任,其参与网络学习的时间和精力非常有限,因此希望能够充分利用有限的学习时间,从而忽视与其他同学和教师进行非智力因素的交流;硕士生本身就是教师,他们在工作中已经形成了良好的教学反思能力,因此其在记录学习笔记方面的能力要明显高于本科生。

##### (3) 学科背景原因分析

统计分析发现,大多数艺术生成绩低于文理科生,造成其成绩较低的原因并非是他们认知能力差,而是他们的学习习惯和自我调控能力较差,因此在网络学习过程中互动表现较弱。

#### 3. 行为特征与学习效果关系小结

研究表明,学习者网络学习交互越强,学习效果越好;学习者外部调控能力越强,学习效果越好。

### (二) 讨论

#### 1. 开发适应性学习系统

20 世纪 90 年代以来,研究者开发了大量的适应性教学系统来适应和支持网络学习,以期实现差异教学,但是这些系统的有效性所需的实证证据贫乏,理论基础薄弱(Shapiro 和 Niederhauser, 2004)。[10]因此根据前人的研究成果与本研究的实证结论,建议开发者根据学习者的行为特征,创建一个灵活、友好的学习环境来支持不同行为特征学习者的学习。

#### 2. 创建微知识模块

研究发现,学习者在网络学习时注意力集中时间有限,因此根据 Mayer 多媒体设计的原则与认知负荷理论,建议课程开发者建立微学习模块,将课程知识点碎片化,例如以 5~10 分钟甚至更短的时长为单元的视频或 50~200 字为单元的文本。这些都可以通过手机短信、学习弹送框、E-mail、微博等形式简明扼要地推送给学习者,为学习者提供碎片化、移动化的网络学习新体验。

#### 3. 组织多样化的交互活动

研究表明,交互与学习效果有显著的正相关,交互是保证网络有效学习的关键因素之一,因此教师要有意识地主动引导学习者积极参与交互。第一,教师应对参与者的提问予以及时回复;第二,采取鼓励及奖励的措施,例如依据学习者的参与情况设置参与度排行榜,定期发布公告,对参与度高的学习者给予鼓励,对参与度低的给予提示;第三,设计更多的协作学习活动,采用基于问题的学习、基于项目的学习、基于探究的学习等方法增加交互行为。

#### 4. 提升自我调控,加强外部调控

学习者应有意识地提升自我调控能力,积极参与网络交流互动。该网络学习系统已经具备很强的监控能力,在学习者有效完成网络学习中起到了很好的作用,例如学习者登录学习页面后清晰地显示学习的进度,并提醒学习是否达标;若学习者学习次数偏低或间隔时间较长,会及时地发送 E-mail 或手机短信进行提醒;对页面学习时间进行监控等。这些外部调控措施具有很好的借鉴价值。

#### [参考文献]

- [1] 第33次中国互联网络发展状况统计报告[EB/OL].[http://www.cnnic.cn/gywm/xwzx/rdxw/2014/201401/t20140116\\_43823.htm](http://www.cnnic.cn/gywm/xwzx/rdxw/2014/201401/t20140116_43823.htm).
- [2] 教育部关于印发《教育信息化十年发展规划(2011—2020年)》的通知.[EB/OL].[2012-3-30][http://www.gov.cn/gzdt/2012-03/30/content\\_2103672.htm](http://www.gov.cn/gzdt/2012-03/30/content_2103672.htm).
- [3] 万海林.近十年我国网络学习行为研究述评[J].河北大学成人教育学院学报.2012,(9):57~60.
- [4] 刘惠如.整合式网路教学之教学设计与评量[D].台湾:国立中山大学资讯管理研究所,2000.
- [5] 杨奕晨,柴蕙质.非同步网路学习成效及影响因素之计量分析[C].台湾:非同步网路教学课程中心建置研讨会,2000.
- [6] J.Michael Spector, M.David Merrill, Jeroen van Merriënboer, Marcy P.Driscoll.教育传播与技术研究手册[M].任友群,焦建利,刘美凤,等主译.上海:华东师范大学出版社.2012.
- [7] 元昌安.数据挖掘原理与SPSS Clementine应用宝典[M].北京:电子工业出版社,2009.
- [8] 熊平.数据挖掘算法与Clementine[M].北京:清华大学出版社,2011.
- [9] Moore, M. . Three Types of Interaction[J]. The American Journal of Distance Education, 1989, (2):1~6.
- [10] Shapiro, A. M. and Niederhauser, D.. Learning from Hypertext:Research Issues and Findings [A]. In Handbook of Research on Educational Communications and Technology[C]. Mahwah,NJ: Lawrence Erlbaum Associates,2004.

(上接第52页)

- [20] 黄建军,张璐.移动学习应用环境构建研究[J].电化教育研究,2013,(7):59~63.
- [21] Chen,G. M. & Starosta,W. J.. The Development and Validation of the Intercultural Communication Sensitivity Scale [J]. Human Communication,2000,(3):47~56.
- [22] Kim, Young Yun. Becoming Intercultural:An Integrative Theory of Communication and Cross cultural Adaptation[M].CA:Sage Publications, 2001.
- [24] [25] 王伟,钟绍春,吕森林.大学生移动学习实证研究[J].开放教育研究,2009,(4):81~86.
- [26] 王萍.移动增强现实型学习资源研究[J].电化教育研究,2013,(12):60~67.