

基于学习行为和内容的网络学习论坛分析模型

胡庆春, 蔡建华

(1.华东理工大学信息学院, 上海 200236;
2.华东师范大学计算中心, 上海 200062)

摘要: 在线学习论坛中学习者之间的互动和文本内容可以作为评价学习的重要因素。本文提出了一个基于学习行为和内容的二维分析模型, 分析在线学习论坛。该模型通过分析同伴之间学习行为的互动, 可视化呈现问题讨论的热点, 探究学习行为和内容是如何促进知识的获取。同时该模型也有助于分析并挖掘学习的困难并通过可视化的词云反馈给学习者。此研究可以为设计自动应答的实时反馈系统和自适应学习系统的体系结构提供实践的建议。

关键词: 网络学习; 在线论坛; 学习分析

中图分类号: TP311 **文献标识码:** A

An Analysis Model of Online Learning Forum Based on Learning Behavior and Content

HU Qingchun, CAI Jianhua

(1. School of Information Science & Engineering, East China University of Science and Technology, Shanghai 200236, China;
2. Computer Center, East China Normal University, Shanghai 200062, China)

Abstract: The interaction between learners and text contents in online learning forums can be used as an important factor in evaluating learning. This paper proposes a two-dimensional analysis model based on learning behavior and content to analyze online learning forums. By analyzing the interaction of the learning behaviors between peers, the model visualizes the focus of discussion and explores how learning behaviors and content promote knowledge acquisition. At the same time, the model also helps analyze and mine learning difficulties and offer feedback to learners through word clouds. This research provides practical suggestions for designing the real-time feedback system of automatic response and the architecture of the adaptive learning system.

Keywords: online learning; online forums; learning analysis

1 引言(Introduction)

如今, 数字化学习时代已经到来并深深地影响了当前的高等教育^[1-3]。随着技术对教与学的影响, 教育技术的关注点已经从技术怎样在学习中被运用, 进一步转变为深入研究技术是如何助力学习, 如何更好地挖掘学习行为数据, 对学习的过程进行评价并给学习者提供自适应的指导^[4]。在网络学习平台中, 在线论坛常常用于学生之间和师生之间的交流, 这是一个必不可少的功能模块, 即通过论坛去支持学生随时随地发布问题, 及时地反映他们的学习。随着时间的推移, 网络学习平台中会积累大量的数据, 例如, 在我校的程序设计课程, 其网络平台的在线论坛有自2010年至今的海量数据, 在线论坛中的这些海量数据可以被看作是一个巨大的学习资源库, 不仅是对当前学生的学习具有价值, 也是对未来学生

的一个很好的学习材料。然而, 这些海量数据与学习过程和结果之间的关系还需要进一步深入探究。

在线论坛的一个主要功能和设计关注点是及时促进用户之间的通信。论坛中最基本的模块是搜索模块, 也即用户可以通过键入关键字信息来搜索论坛。虽然是可以搜索到与关键字匹配的信息, 但是, 不能清晰并深入地了解论坛中的活动。本研究旨在挖掘学习行为和讨论区的内容, 可视化展示其特点, 以便通过可视化图能够像脚手架一样支持(scaffold)学习者的学习。因此, 本文通过对论坛主题的归纳分析, 提出了论坛的概念图分析模型。在该模型中, 分析学习行为的两个方面: 同伴之间的交互, 发帖的主题和内容。首先, 对于同伴之间的交互, 本文通过分析帖子的回复和响应时间进行聚类, 以找出最热门的帖子内容; 其次, 对于发帖的主题和内

容,通过可视化的词云显示帖子内容,构成一个可视化的学习困难概念图,目的是不仅促使教师在教学中更加关注此类问题,同时也是促使学习者反思其所学的内容。

本研究分析的数据源自C程序设计课程,这是我校大多数理工科专业一年级本科生的必修课程。选取了一个学期的论坛数据进行分析。可视化学生同伴之间的互动和论坛内容,探究论坛发帖的主题和学习的知识点之间的联系。由于程序设计课程学习中很多困难的本质就是提高调试能力,本研究使用文本挖掘工具的可视化分析,揭示了在编程学习中的主要困难。同时,分析了同伴之间学习行为的不同模式。

行为和内容的二维分析模型加深了教师和学生对于在线论坛学习过程的理解;吸引了讨论参与的积极性;提供了证明在线内容讨论与知识点学习之间的重要性。

下面的内容首先回顾了相关工作;接着讨论了在线学习中的学习分析,在此基础上提出了模型框架;然后对数据进行采集和分析;最后是总结,建议进一步完善此方法有助于自动应答系统和自适应学习系统的设计。

2 相关的研究(Related work)

2.1 程序设计课程的网络教学环境

本研究的背景是C程序设计课程,这是我校大部分理工科本科生的必修课。在学习程序设计的过程中,学生遇到的主要问题集中在代码调试上^[5,6];而且,在课堂教学中师生之间的互动不够理想。因此,迫切需要一种有效的方法来提高教与学的效果。

随着数字时代对高等教育的影响,传统的课堂正在延伸。教师和学生可以通过网络环境进行交流和互动。自七年前,Moodle作为支持教和学的网络平台被应用在我们的教学中。Moodle平台有很多实用的功能^[7-9],例如,随时随地、方便的内容更新、灵活的师生互动、及时的答疑和反馈机制,这些都非常适合我们课程的要求。可见,一方面,在线学习环境为学生和教师提供了非常灵活的交互方式;另一方面,也提供了实时记录保存学生的学习信息。显然,这些海量的信息包含了学习者的隐性行为特征。本研究希望剖析学生的学习活动,深入了解在线学习的认知过程。但是,目前大多数在线学习系统分析学生行为大多来自日志活动,鲜有整合在线论坛的内容^[10-12]。而学习过程是行为和内容的一个复杂认知过程,需要综合考虑多个因素的相关性和影响。当学生在论坛上发表问题时,表明其对于相关的知识点是有疑问的,通过对某一个阶段的数据分析,可以挖掘学习困难的薄弱环节。为此,通过对网络学习平台的学习行为和论坛内容进行洞察分析,并将其融入到学习评价中是必要的,这与学习分析研究领域有关。

2.2 学习分析

学习分析是数据挖掘技术在教育中的一种实践,它来源于大数据的研究背景^[13,14]。在学习分析和知识的国际会议中,学习分析被描述为对学习者和学习情景数据集的测量、收集、分析和报告,以理解和优化学习和情景^[15]。学习分析需要对海量数据中的学习过程进行进一步的挖掘,以评估学生的学业进步,预测其未来的结果,找出影响学习的因素。它的目的不仅是评估学生,而且还可以发现潜在的问题,并最终优化学习。通过在一段时间内收集和分析大量学习者的相关数据,探索学习过程和变化,以及学习者的当前学习行为与学习结果之间的联系。另一个目标是用既定的相关性,根据当前的学习行为,预测他们未来的学习趋势和结果。这种分析和预测不仅对教育管理者、教师,而且对学生都有积极的意义。在线学习论坛中累积的数据其主要特征是规模大和多样性。这些数据隐含着学习者的实时性特征。学习分析的目的和方法与学习目标紧密结合。没有学习目标,学习分析会过于宏观和抽象。本研究与当前教育技术领域的国际关注是一致的^[16],未来的教育技术需要从学习、教学、管理、评价、基础设施等五个方面进行阐述。其中,未来的教学系统应具备以下特点:自适应地评价学习者的知识和能力、嵌入式的学习过程、实时反馈等。要达到这一目标,就需要学习分析技术。

当前已有的分析学习论坛的方法主要有:使用日志文件分析、页面链接的点击率分析、可视化文本分析和评估。这些方法在一定程度上反映了学生的学习过程和对知识的掌握;但是其主要关注的是单个因素在学习中的作用,数据来源较单一,鲜有研究综合了多个因素。网络平台的讨论区其数据来源源于多种学习活动,有师生之间,学生之间的互动数据,也有涉及学习内容的文本信息,需要考虑对数据进行分类和提取。本文从论坛发帖主题的互动,时间节点和内容方面,对论坛活动进行了剖析。

本研究提出的模型是在一个真实环境中的实践模型,用此模型可以让教师知道该何时和如何关注学习者贴出的话题,旨在探索Moodle论坛中讨论内容及互动关系,希望在学习论坛中找出需要紧急响应的问题帖子,并且可以帮助教师有效地浏览学习论坛和及时干预以辅助学生的学习。

3 基于行为-内容的模型框架(A model framework based on behavior-content)

本研究构建了一个基于同伴互动和内容分析的模型(图1),其主要构成部分是活动模块和内容模块。数据采集、存储和分析是在该模型框架下进行的。本文从帖子回复期的时间跨度;同伴之间的帖子回复;帖子内容分类这些方面展开分

析。正如上面提到的，在C程序设计课程中，学生遇到的主要困难是调试问题。但是在一个班级或者本学期的学习中，哪些概念导致调试困难还不清楚。为此，本研究把学习内容分为两类：概念和调试。前者是指发布与概念相关的帖子；后者是发布编码调试与请求帮助。这两种帖子均与学习内容有关，即和所学习的知识点有关。分析模型中的两个方面：互动和内容。这两个方面均可以看作是脚手架(scaffold)机制。一是来自认知支持(cognitive scaffold);另一种是同伴互动支持(peer scaffold)。基于此，可以挖掘发现论坛中的热点问题。该框架具有跨平台的潜在应用，帮助教师有效地浏览在线论坛，并及时干预学习者的学习^[17]。另一方面，也有助于学生对所学知识的反思。

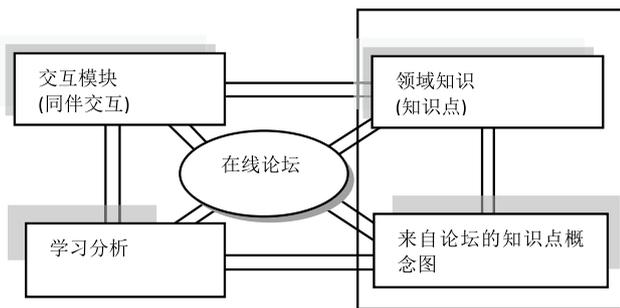


图1 模型框架

Fig.1 The framework of Model

4 数据分析(Data analysis)

本研究采用了开源的工具软件Kibana进行可视化分析，根据图1的模型框架，在活动的维度上，分析了用户之间发帖的交互、回帖交互及其时间节点序列；另一个内容维度用论坛内容中的词云来可视化。在Moodle的后台数据库系统选取和问题讨论区有关的内容，论坛中的数据如图2所示。



图2 可视化的学习分析

Fig.2 The Visual learning analysis

4.1 学习者之间的交互

学习者之间的交互主要体现在每个发帖、回帖和浏览帖子。本研究可视化了阅读数按照用户的分布(图3(a))，阅读数按主题分布(图3(b))。前者展示了学习者的活跃程度，可以清晰的看到最活跃的学习者；后者，可以看到各个主题的访问量，例如，在图示的这段时间内，关于“return 0”的这个主题引起大家的关注。同时，本研究还可视化了帖子数按照用户分布(图4(a))，帖子数按照主题分布(图4(b))。

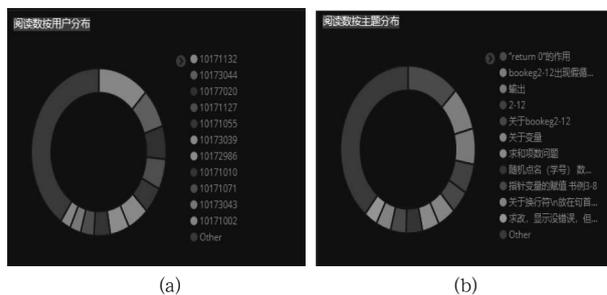


图3 阅读数按用户分布和主题分布

Fig.3 The number of readings based on users distribution and topics distribution

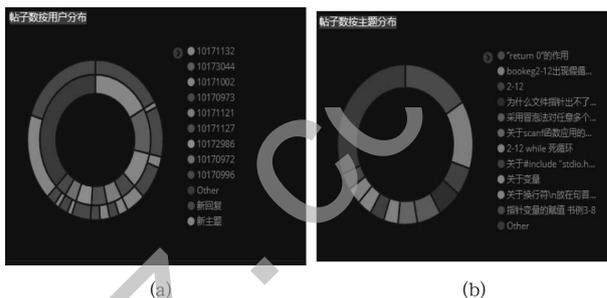


图4 帖子数按用户分布和主题分布

Fig.4 The number of posts based on users distribution and topics distribution

在图2的基础上，针对一个学期的数据分析并聚类了关于帖子的时间节点特征。在一个学期中，某个班级有141名学生参加了C程序设计课程。发布229个主题。其中，173个帖子得到回复。回复这些帖子而衍生出来的帖子回复数量是360。在这些回复中，图5(a)显示了每个帖子得到第一次回复的时间。可以看到大多数帖子在五天内回复。如果这个帖子在五天内无法得到回复，那么它可能最终不会得到回复。因此，教师应该在学习过程中进行干预。图5(b)显示了每个帖子的回复时间跨度。例如，如果一个帖子得到了第一个回复和最后一个回复，在此期间称之为每个帖子的回复时间跨度。通过可视化发现时间序列模式是聚集的。很明显地，时间序列分为四个时期：两天、五天、十天，超过十天。为此，可以把帖子分为两类。一是得到积极的回复，另一个则是惰性的回复。

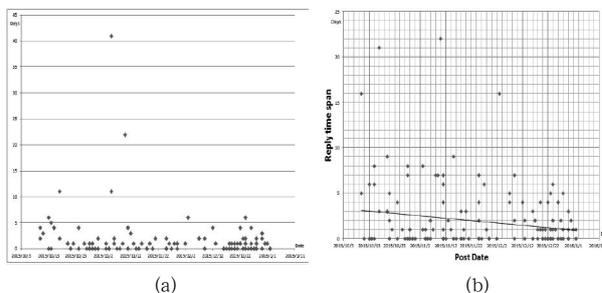


图5 帖子的回复时间节点特征分布

Fig.5 The feature distribution of posts' reply time

