

基于文本挖掘的学习分析应用研究

刘三女牙, 彭 晔, 刘 智, 孙建文, 刘 林, 郑年亨

(华中师范大学 国家数字化学习工程技术研究中心, 湖北 武汉 430079)

[摘 要] 当前,关于学习分析的研究,国内外研究者主要关注学习者在网络教学环境下产生的结构化数据,而伴随学习交互模式的多元化,以文本为主的非结构化数据正在不断生成。近年来,通过对文本数据的挖掘来测评学习者的知识能力以及甄别其心理与行为已成为一种新的学习分析方法。本文首先描述了学习分析、文本挖掘的概念,通过文献分析法,对国内外采用文本挖掘技术的学习分析研究现状进行了综述;其次,介绍了学习分析中文本挖掘的数据来源、方法和工具;最后主要从课程评价支持、学习者知识能力测评、学习共同体分组、学习行为危机预警、学习效果预测和学习状态可视化等6个方面阐述了文本挖掘应用于学习分析中的具体实例,并对未来作了进一步的展望和探讨。

[关键词] 教育大数据;学习分析;文本挖掘;文本分析

[中图分类号] G434

[文献标志码] A

[作者简介] 刘三女牙(1973—),男,汉族,安徽桐城人。教授,主要研究方向为计算机应用、人工智能、教育信息技术。

E-mail:lsy5918@mail.ccnu.edu.cn。

近年来,“大数据”成为各行各业关注的热点话题,教育作为大数据的一个重要应用领域,相关学者已预测大数据将给教育带来革命性的变化。开放式网络课程的井喷式涌现和在线学习平台的普及,使教育领域中的“大数据”技术获得了更为广阔的发展空间,为利用这些数据中隐含的、有价值的信息,进行智能化的教学决策与优化的学习服务,学习分析技术应运而生。当前国外一些学习分析研究小组和机构已开展了一些前瞻性项目和应用,例如:2009年美国普渡大学推出一个名为信号(Signals)的项目,^[1]旨在利用课程管理系统中的数据来预测学业成功与失败;2011年比尔和梅琳达·盖茨基金会资助开发的大数据分析系统——“学位罗盘”(Degree Compass),^[2]它能在对学生信息管理系统中数据分析的基础上向学习者推荐合适的课程;2012年密歇根大学开发了用于鼓励和指引学习者有效完成交互活动的 Gradecraft 平台等。^[3]我们可以发现,目前的学习分析应用大多是利用学习

管理系统、教学软件来收集学习者个体及群体在网络教学活动中产生的结构化数据,然后从中提取知识并存入关系型数据库。

然而最近的趋势是交互科技(Interactive Technologies)在教学中得到了广泛应用,通过在线学习系统、社交网络平台、移动终端生成了越来越多复杂的非结构化数据。Gartner 和 IDC 称:近 80%的网络数据都是非结构化,^{[4][5]}这些数据蕴藏了大量有价值的信息,但是目前人们还很难管理和分析。毫无疑问,文本是非结构化数据中最典型、显性的实例来源。^[6]通过捕获和分析在商业、生物医学、政府等中的文本数据^{[7][8][9]}已改变了其决策的制定和资源的分配。然而,在教育领域中,致力于优化学习服务的分析技术,通常注重于数字的分析,^[10]比如在线学习系统中学习者的登录/退出次数,上传/下载文件次数,提交/未提交作业次数,浏览课件时长等外显式行为数据,但较少考虑学习者的态度与感受等内隐式行为信息,例如学习

基金项目:2014 年度国家社会科学基金项目“面向高校青年网络行为的情感识别关键技术研究”(课题编号:14BGL131);2015 年度中央高校基本科研业务费项目“大数据环境下的学习行为分析与预测研究”(课题编号:CCNU15A050100)

者群体的在线对话交流与讨论,学习者个体的课程评价、笔记等数据信息。因此**传统的学习分析技术难以准确发现学习者的个体意愿和内在心理状态,并不能深入解释影响学习者学业成功与否的关键因素**。文本作为教育大数据中一种特质的类型,最真实、直接地反映了学习者的学习动机、认知发展、情感态度、学习体验。

通过文本挖掘,可提取出学习者文本表达中隐藏的关键信息,实时监控学习者的意见和想法,将有助于学习者在文本交互中的自动干预和推理。并且从2011年起,连续四届“学习分析技术与知识国际会议”中,文本作为协作学习和知识传播的重要途径已成为了一个研究重点。由此可见,文本挖掘在学习分析中的地位已日益凸显,那么如何将文本挖掘的优势与学习分析应用相结合就给我们带来了新的研究视角和挑战。

一、核心概念

(一)大数据环境下的学习分析

随着教育信息化的不断深入和“大数据”时代的来临,学习分析作为大数据在教育中的重要应用之一,为达到支持教学决策、优化学习服务的目的提供了一系列方法和策略。加拿大阿塞巴斯卡大学的 G. Siemens 教授认为,学习分析是“关于学习者以及他们的学习环境的数据测量、收集、分析和汇总呈现,目的是理解和优化学习以及学习情境”。^[11] 新媒体联盟(NMC)将学习分析定义为:利用松散耦合的数据收集工具和分析技术,研究分析学习者学习参与、学习过程和学习表现的相关数据,进而对课程、教学和评价进行实时修正。^[12] 并且新媒体联盟《地平线报告》(2014 高等教育版)预测学习分析将在 1 年内成为主流。^[13]

(二)面向学习者行为的文本挖掘

文本挖掘作为辅助学习分析的核心技术之一,可以帮助辨识和解释学习者的心理及行为过程。文本挖掘作为一种主流的方法论,在许多领域都有广泛的应用。**文本挖掘一般指从非结构化的文本数据(例如:文档、聊天信息和邮件)中发现和提取有用的模式、模型、方向、趋势或者规则,**^{[14][15]}**最终形成用户可理解的信息与知识的过程。**它是语言学、统计学、机器学习、数据挖掘及自动文本处理等理论和技术相结合的产物,更多地被用在生命科学研究、政府智能、商业等应用领域。目前,已有不少研究将文本挖掘引入教育领域中,辅助文本内容研究,实现面向学习者行为的文本数据挖掘,支持学习设计和评估反馈。

二、国内外研究现状

在查阅国内外著名文献资料数据库基础上,本文尝试运用文献分析法对基于文本挖掘的学习分析应用研究的现状进行综述和梳理,具体情况如下。

对于国内研究,文献分析的主要来源是“中国知网”学术文献库中的相关文献。笔者在“中国知网”学术文献库中以最近十年(2005 年 1 月至 2014 年 12 月)为时间跨度,对“文本挖掘”和“在线学习”、“文本分析”和“在线学习”、“学习分析”等关键词进行高级搜索,统计时剔除与主题关联性不大的文章,最后获得相关文献 11 篇。由此可见,国内运用文本挖掘技术来调整和改善教学的研究较少,缺乏具体的案例实践经验,有待进一步思考和研究。

对于国外研究,笔者以 Google 学术搜索、ACM 美国计算机学会数据库(国际站)、Elsevier Science Direct Online 数据库为国外文献搜索来源,按照如上相同方法进行高级搜索和筛选,最后获得与本文主题相关学术期刊或会议论文共 103 篇(Google 学术搜索 43 篇,ACM 数据库 31 篇,Elsevier 数据库 29 篇)。国内外各年份研究论文的数量分布如图 1 所示。

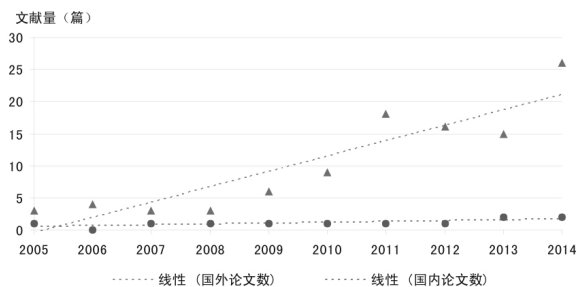


图 1 国内外研究论文分布趋势图

图 1 显示,国内对文本挖掘学习分析应用的研究论文数量整体上呈现小幅度上升。而国外相关研究则表现出强劲的发展趋势,尤其自 2010 年以后,研究论文增长的幅度更为明显,可能主要因为在国际上有专门针对学习分析研究和应用而召开的国际会议“学习分析技术与知识国际会议”和美国新媒体联盟与美国高等教育信息化协会(EDUCAUSE)对学习分析研究的高度重视。根据文本挖掘在学习分析应用中的研究与实践,我们可以将获取的 114 篇文章划分到七个应用类别:“课程评价支持”、“学习者知识能力测评”、“学习共同体分组”、“学习行为危机预警”、“学习效果预测”、“学习状态可视化”、“其他”(包括知识构建、文本注释、同伴互评等方面研究)。文本挖掘在学习分析中的具体应用研究分布情况如图 2 所示(因笔者查阅到的国内文献较少,特将国内外相关研究进行统筹处

理分析)。

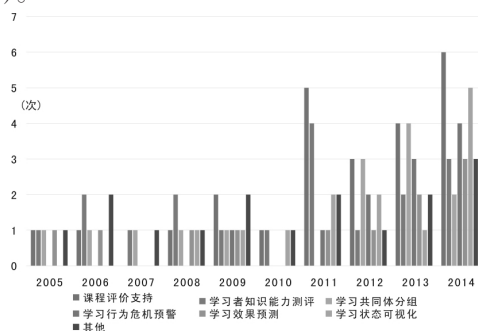


图2 文本挖掘在学习分析中的具体应用研究分布

从图2可看出,往年关于课程评价支持的研究相对较多,一定程度上反映出文本挖掘在支持课程教学方面具有较高的实用价值。最近几年,各类具体应用研究的比重都有所提高,这也表明将文本挖掘作为一种技术手段来调整和优化教学过程,已日渐得到本领域研究者的认可。

三、基于文本挖掘的学习分析方法

为了更全面准确地揭示学习系统的概览,学习分析中不仅应包括结构化数据,也应包括以文本为主的非结构化数据。我们需充分利用文本挖掘技术的优势来剖析学习者文本表述背后的深层次含义,理解其认知过程,辅助分析、优化学习及其发生情境。图3是文本挖掘应用于学习分析的系统框架图,左侧列举了制约文本挖掘在教学中继续深入推进的关键因素,中间描述了文本挖掘作为支撑学习服务的一种辅助技术

的过程,右侧总结了文本挖掘应用于学习分析中的一些具体实践项目。

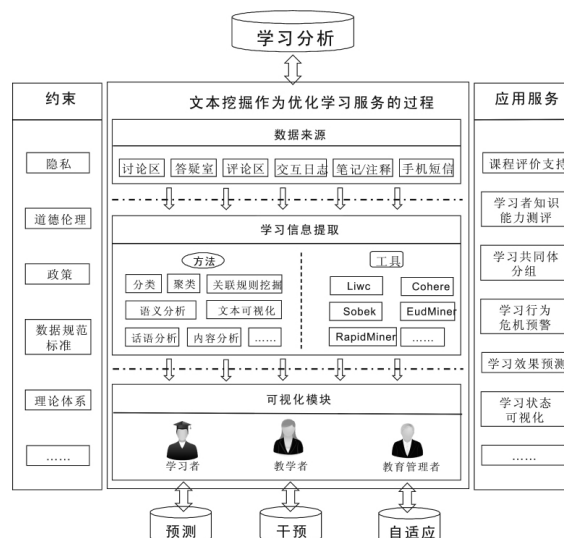


图3 文本挖掘应用于学习分析的系统框架

(一)数据来源

目前,在线学习系统、教学网站和网络公开平台的讨论区、实时答疑室、评论模块、学习文档、学习笔记和日志已成为学习者产生文本数据的主要来源,并且随着广播式社交网络应用的盛行,校园 SNS(Social Networking Services)、微博、博客等 Web 应用以及移动手机、平板上所产生的文本数据都为密切跟踪学生学习的感受和需求提供了充足且真实的数据源。与现实面对面交流不同,虚拟网络为学习者提供了更为自由和开放的言论空间,这使得学习者在在线互动过

表1 2005—2014年文本挖掘在学习分析应用中的工具概述

工具	次数	特点点	目的
1 LIWC	5	语词分析、词频统计、心理评估	通过统计学习者文本表述中情感词的频率来识别其持有的态度和观点;可分析学习者不同形式写作内容中(论文、自我介绍短文、日记等)的语词构造和心理意义,来揭示学习者自我管理学习和反思过程
2 Cohere	3	语义连接、社交网络、可视化	使用该工具学习者可以将注释文本作为一种信息资源进行索引和检索,实现语义连接;可发现持有相似观点的学习者,构建学习者在线交互的协作网络概念模型,旨在集体智慧最大化
3 Sobek	2	可视化	从学习者短文写作中实现相关概念提取,观点、关系、事实和事件的发现,并以交互式图形可视化呈现,进而支持学习者和教学者任务的顺利开展
4 RapidMiner	1	观点挖掘	收集学习管理系统中学习者关于课程评论的文本数据来识别其对平台功能、教学设计、学习持有的态度和观点
5 Dissertation Browse	1	相似性检测、可视化	可视化呈现不同学科间学术论文的相似性检测结果
6 EduMiner	1	实时反馈、可视化	实现对学习过程的自动形成性评估;在动态的追踪学习者交互过程中,使用该工具不仅可减轻教学者的压力,而且能够及时将交互绩效的可视化结果反馈给学习者,学习者进行自我调整,有助于促进在线学习者对话交流过程中完成更深层次的表达
7 GCS	1	分类	对课程管理系统讨论区中产生的帖子进行分类,帮助教学者监控和调整其在线交互活动
8 Toreador	1	难度评定、标注	根据学习者年龄、性别、先验知识水平的差异,评估在线阅读资料的难度系数,预测并标注对于学习者的难度词汇;主动推送与其阅读水平更加吻合的阅读内容,激发其学习兴趣

程中“无意”生成的数据里蕴含大量潜在的、有价值的个体或群体信息。

(二) 分析方法

通过对收集文献的总结归纳,基于自动挖掘的方法可以涵盖以下几种:(1)分类;(2)聚类;(3)关联规则分析;(4)语义分析;(5)可视化;(6)话语分析;(7)内容分析;(8)其他。

针对不同教学情境中文本数据的复杂性、内隐性、交互性等特点,一些特色各异的文本挖掘工具已经被设计和开发,极大地加快了人们从中提取信息的能力,为快速、精准、深入解读学习者表述中的隐含内容和丰富意义提供了便利。表1是基于文献分析基础之上,对学习分析具体实例中文本挖掘工具使用次数进行的排序和概述。

四、文本挖掘的学习分析应用综述

当前文本挖掘与学习分析正在不断地深度融合,这一交叉领域已产生了大量研究成果,具有广泛的应用前景。根据差异化的教学目的,针对不同的学习情境,研究者开发出一些新工具和提出新模型、方法,并在新研发平台或已有平台上开展了大量实例研究。笔者对其进行详细的总结和归纳,主要从以下六个方面来探讨。

(一) 课程评价支持

作为一种有效的教学方法,课程评价在教学过程中的作用至关重要,能为教师和学生提供及时的反馈和建议。如一项探讨挖掘学习者反馈的教学文本情感的研究,^[16]提到在每门课程结束后,学习者通过移动设备发送短消息来评价课程教学。该研究旨在利用SMS(Short Message Service)文本内容和表情符号来呈现课程评价的积极和消极内容。与此类似的一项研究中,Kontogiannis等人提出了一种新的课程教学评价框架来自动挖掘学习者情感观点,^[17]他们通过收集学习者在社交网络微博客中产生的有关学习课程活动的文本数据,并利用观点挖掘技术来判断学习者对每门课程所持有的积极或消极态度。此外,一项由美国国家科学基金会资助的学习系统GIS研究,^[18]通过收集GIS系统中大学生的文本评论内容,使用文本挖掘技术来发现和抽取潜在的知识,旨在将学习者评论划分为不同类别,从而辅助学习平台的教学者可以从大量文本数据中快速识别出关键主题,为进一步指导课程教学提供依据。

(二) 学习者知识能力测评

相比传统研究中利用学习者的客观题解答结果来评估知识掌握程度,我们可以利用学习者在教学活

动中产生的主观文本数据来测评其知识结构、高级思维技能等。以下我们将主要从学习者写作能力和阅读理解能力测评两个方面来阐述。在威特沃特斯兰德大学计算机科学学院,^[19]研究者专注于开发一个自动短文评分系统,它通过对学习者提交的短文本内容采用潜在语义分析方法进行自动评分。实验结果指出,该系统在理想的配置下可以准确测评出学习者的写作水平,并且自动评分的准确度可超过80%。此外,中国香港大学计算机科学与工程系研究者开发了一个名为VeriGuide的平台,^[20]旨在提供专业的写作工具集,测评学习者能力并促进写作的教学和学习。在智利大学,一项基于情境化模型和潜在语义特征的自动化文本理解分类器研究中,^[21]以来自工程学和语言学学院的大学一年级学生为实验对象,通过分析学习者关于阅读资料问题的文本作答来检测发现其阅读理解能力的不足。此外,就学习者的阅读理解任务而言,^[22]该项研究描述了如何结合多维K-means聚类方法和布鲁姆教育分类理论来确定他们采取的积极和消极的认知策略。

(三) 学习共同体分组

大规模在线学习环境中,小组学习已经成为一种主流的学习方式,我们应结合实际的教学需要,根据学习者特征来确定不同类型的学习者,促进学习共同体学习效果的最优化。例如,在加拿大英属哥伦比亚大学,一项辨识学习者在线交互行为模式的研究中,^[23]研究者基于学习者与在线学习平台的交互日志进行分析,构建用户模型框架,对具有相同学习偏好、兴趣、主题等的学习者进行聚类和分组,提供合适的学习支持和交互体验。在此基础上,西班牙科尔多瓦大学计算机科学学院研究者提出了一种改进的聚类模型,^[24]相比传统收集学习者所有在线交互的日志数据,该模型在划分学习共同体组别方面具备更好的适应性。在美国哥伦比亚大学计算学习系统中心,研究者介绍了一种非监督式方法来自动化检测讨论区主题内持有相似观点的学习者。^[25]实验结果指出,学习者含蓄态度的表达(倾向于使用相似的文本内容)对于直接态度的表达(讨论话题内隐含的情感)是一个很重要的补充,并且两者的结合可以有效提高构建学习共同体的准确度。

(四) 学习行为危机预警

根据言语行为理论,^[26]人们在书写文字的同时也在实施某种行为,从文字中传递出来的言语信息可反映个体的意志和心理行为现象,能为教师及时掌握学习者的思想状况及行为危机预警提供可靠的决策支

持。在澳大利亚昆士兰大学,研究者使用 SNAPP (Social Networks Adapting Pedagogical Practice) 工具来分析学习者与讨论区的交互活动,^[27]旨在识别处于高风险状态的学习者,引导教学者及时实施干预。由美国国家科学基金会(National Science Foundation)资助的卡耐基梅隆大学计算机科学学院的一项关于情感分析预测学习者退课行为的研究中,^[28]以 Coursera 平台中 3 门课程的讨论区为数据来源,通过分析学习者集体的情感来监测学习者在课程中的情感趋势。实验结果指出,学习者在课程讨论区内的集体情感比率与其退课率有显著的关联。此外,在美国德州大学达拉斯分校行为与脑科学研究院,Underwood 教授开展的关于中学生反社会行为挖掘的开创性研究,^{[29][30]}该研究通过采集学习者在手机通讯中产生的短信内容,利用发布时间、关键不良情绪词、发送对象等信息推测群体性不良行为,并得出了触发该群体不良行为的一系列量化因素,为学习者行为的分析和危机干预起到了重要的推动作用。

(五)学习效果预测

研究者通过采集和分析学习者学习经历相关文本数据,探索学习者在不同阶段中学习效果的变化,发现学习者学习行为与学习效果的相关关系。例如,在美国德克萨斯大学,^[31]研究者以 363 个大学生的自我介绍简文为数据来源,使用 LIWC 工具对其进行分析来预测学习者的课程效果。实验结果指出,某些特殊语词类别的词语构造使用概况确实可作为预测学习者课程效果的重要指标。西班牙科尔多瓦大学计算机科学与数据分析学院提出了从定量、定性和社交网络等三个测量角度来评判学习者在讨论区内的参与行为,^[32]并结合不同的文本挖掘方法来提高学习者最终学习效果预测的准确度。实验结果表明,基于学习者在讨论区内产生的文本数据,使用聚类及关联规则分析方法有助于预测学习者课程通过情况。美国马里兰大学和加利福尼亚大学研究者在 LDA 主题模型基础之上提出了种子主题模型。^[33]实验结果表明,该模型能够通过分析 MOOC 平台中学习者交互讨论的文本数据来预测学习者学业完成效果。

(六)学习状态可视化

使用文本数据的交互式图形来评价学习者的知识结构、认知能力、情感态度等多维度的状态特征,有助于教学者的教学决策和学习者的自我学习监控。例如,英国开放大学研究者提出了一个自然语言处理工具 XIP,^[34]旨在分析学习者的学术写作文档,并通过仪表盘来可视化其中的关键特性,如语句的修辞类型、关

键概念等。台湾一项分析学习者知识构建过程的研究中,^[35]以 56 个来自信息管理专业的研究生为实验对象,通过分析处理学习者在在线学习社区交互过程中产生的文本内容,按照布鲁姆的教育目标分类法,从知识层次、领会层次、应用层次、分析层次、综合层次和评价层次等 6 个方面为教学者和学习者呈现实时可视化的认知能力评估图形。研究结果指出,该评估方法可通过有效激发学习者的学习动机来发布高阶认知层次的内容,并促进学习者之间更深层次的对话交流。在一项旨在追踪学习者情绪变化过程的研究中,^[36]教学者通过自主开发平台来分析学习者的学习体验日记,依据普拉特切克的情绪三维理论,自动抽取学习者的 8 类基本情绪,实现学习者情绪流的动态可视化呈现。

除了以上描述的六个主要应用研究方面外,也有部分考察和促进学习者高阶思维能力发展的研究。比如,Gasevic 等人通过使用特有的文本挖掘框架 LIWC,^[37]收集学习者产生的文本数据来分析他们的心理和语言加工过程,识别和抽取其中的关键变量,从而培养自我反思和元认知建构的技能;此外,Rosen 等人采用语义模型来实现文本挖掘,^[38]研究结果表明,学习者的分析推理和探索分析能力都有所提高,学习者之间的协作明显加强。最近研究发现,研究者相当重视一定社会文化情境下学习者与在线对话交流实践中的文本话语,相比学习者外显行为数据,这些数据更能表征出学习者在特定情境中的内隐思维过程。并且运用话语分析技术,可以帮助我们理解不同类型文本消息或消息序列在知识构建过程中存在的时间和因果关系,而且有助于我们探究意义学习的发生模式。^{[39][40][41]}

五、总结和展望

众所周知,一场由网络教育 2.0 为标志的新革命正强烈冲击着传统教育的生态环境,学习的交互方式和组织形式不断被重塑和变革,个性化学习需求受到各界重视,以学习者为中心的新的教学模式催生了大量文本数据的产生,如何利用文本挖掘技术对其进行深度挖掘,提取潜在的、有价值的规律信息及群体或个体学习行为模式已经刻不容缓。将文本挖掘应用于学习分析中,不仅可以帮助学习者快速定位和检索关键信息,也能够引导学习者进行自我反思、自我管理、自主学习,促进学习者的知识构建和有意义学习。

然而,由于技术与现代化教育理念的融合尚浅、文本挖掘技术的发展瓶颈、研究者单一学科背景的局限等多方面的不足,目前采用文本挖掘的学习分析研

究仍处于初级阶段。未来相关研究可以从以下几个方面深入展开。

1. 多样性文本数据的标准研究

随着泛在学习环境的不断丰富,面对学习过程中产生的大量不同来源的文本数据,如何规范化管理和共享,实现数据的兼容性和统一性是当前急需解决的问题。

2. 面向学习服务的文本挖掘模型和方法研究

文本挖掘作为一种面向学习服务的社会情境性辅助技术,所涉及的多种模型和方法大多局限于某一学科或情境,为避免资源的浪费,应设计一些通用的文本挖掘模型和方法来实现跨平台、跨学科的应用。

3. 学习者在线交互文本数据的深度学习技术研究

深度学习技术有利于对学习本质的认识,应加大研究力度和资助经费来挖掘交互式文本数据中的各种学习行为和心理信息,通过深入分析和掌握学习

者文本表达的内在含义,理解学习者在协作学习中知识信息的建构过程,实现知识驱动的深度学习研究,探讨意义学习的演化规律以及揭示其发生机制。

4. 跨领域的多学科交叉理论研究

目前此项研究主要是从自然科学领域的角度出发,如机器学习、自然语言处理、统计学等,难以挖掘文本数据中隐藏的深层次内涵和解释学习者的心理状态变化,应融合人文科学领域视角,如学习科学、学习心理学、教育学等,进一步为其提供理论依据。

笔者相信,为了更好地理解和优化学习以及学习情境,文本挖掘将成为学习分析研究中一种有力的论证方法。我国广大教育工作者有必要充分意识到文本挖掘在学习分析应用中所扮演的重要角色,借鉴国外先进的理念和研究经验,将此项新型的研究运用于在线学习中,支撑数据驱动的学习和评估,优化教学方法和教学活动,从而进一步推动教育教学的创新研究。

[参考文献]

- [1] 祝智庭,沈德梅.学习分析学:智慧教育的科学力量[J].电化教育研究,2013,(5):5~12.
- [2] West, Darrell, M.. Big Data for Education: Data Mining, Data Analytics, and Web Dashboards. Governance Studies at Brookings [R]. Washington: Brookings Institution, 2012: 1~10.
- [3] [13] The New Media Consortium. NMC Horizon Report 2014 Higher Ed Edition [EB/OL]. [2014-01-10]. <http://www.nmc.org/publications/horizon-report-2014-higher-ed-edition>.
- [4] Steve Kearns. Six Reasons Businesses Need to Pay Attention to Unstructured Data [EB/OL]. [2013-12-01]. <http://www.itbusinessedge.com/slideshows/six-reasons-businesses-need-to-pay-attention-to-unstructured-data-05.html>.
- [5] Xu, Y., Reynolds, N.. Using Text Mining Techniques to Analyze Students' Written Responses to A Teacher Leadership Dilemma[A]. Proceedings of the 2011 4th IEEE International Conference on Computer Science and Information Technology [C]. New York: ACM Press, 2011: 93~97.
- [6] Seth Grimes. Unstructured Data and the 80 Percent Rule [EB/OL]. [2008-08-01]. <http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>.
- [7] [15] Byrd, R.J., Steinhubl, S.R., Sun, J., et al. Automatic Identification of Heart Failure Diagnostic Criteria, Using Text Analysis of Clinical Notes from Electronic Health Records[J]. International Journal of Medical Informatics, 2014, 83(12): 983~992.
- [8] Melville, P., Chenthamarakshan, V., Lawrence, R.D., et al. Amplifying the Voice of Youth in Africa via Text Analytics[A]. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. New York: ACM Press, 2013: 1204~1212.
- [9] Enck, W., Xie, T.. Tutorial: Text Analytics for Security [A]. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security [C]. New York: ACM Press, 2014: 1540~1541.
- [10] 吴永和,曹盼,那万里,马晓玲.学习分析技术的发展和挑战——第四届学习分析与知识国际会议评析[J].开放教育研究,2014,20(6):72~80.
- [11] 魏雪峰,宋灵青.学习分析:更好地理解学生个性化学习过程[J].中国电化教育,2013,(9):1~4.
- [12] The New Media Consortium. NMC Horizon Report 2011 Higher Ed Edition [EB/OL]. [2011-01-10]. <http://www.nmc.org/publications/horizon-report-2011-higher-ed-edition>.
- [14] Reategui, E., Epstein, D., Lorenzatti, A., et al. Sobek: A Text Mining Tool for Educational Applications[A]. Proceedings International

- Conference on Data Mining (DMIN)[C]. Las Vegas: ACM Press, 2011:59~64.
- [16] Leong, C-K., Lee, Y-H., Mak, W-K.. Mining Sentiments in SMS Texts for Teaching Evaluation [J]. Expert Systems with Applications, 2012, 39(3): 2584~2589.
- [17] Kontogiannis, S., Valsamidis, S., Kazanidis, I., et al. Course Opinion Mining Methodology for Knowledge Discovery, Based on Web Social Media[A]. Proceedings of the 18th Panhellenic Conference on Informatics[C]. New York: ACM Press, 2014: 1~6.
- [18] Yu, W.B., Luna, R.. Exploring User Feedback of A E-learning System: A Text Mining Approach [M]. Las Vegas: Springer Berlin Heidelberg, 2013.
- [19] Klein, R., Kyrilov, A., Tokman, M.. Automated Assessment of Short Free-Text Responses in Computer Science Using Latent Semantic Analysis [A]. Proceedings of the 16th Annual Joint Conference on Innovation and Technology in Computer Science Education[C]. New York: ACM Press, 2011:158~162.
- [20] Lau, T.P., Wang, S., Man, Y., et al. Language Technologies for Enhancement of Teaching and Learning in Writing [A]. Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee[C]. New York: ACM Press, 2014: 1097~1102.
- [21] Bravo-Marquez, F., L'Huillier, G., Moya, P., et al. An Automatic Text Comprehension Classifier Based on Mental Models and Latent Semantic Features [A]. Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies[C]. New York: ACM Press, 2011: 158~162.
- [22] Peckham, T., McCalla, G.. Mining Student Behavior Patterns in Reading Comprehension Tasks [J]. International Educational Data Mining Society, 2012, (34): 87~94.
- [23] Kardan, S., Conati, C.. A Framework for Capturing Distinguishing User Interaction Behaviors in Novel Interfaces [A]. EDM [C]. New York: ACM Press, 2011: 159~168.
- [24] Bogarín, A., Romero, C., Cerezo, R., Sánchez-Santillán, M.. Clustering for Improving Educational Process Mining [A]. Proceedings of the Fourth International Conference on Learning Analytics And Knowledge[C]. New York: ACM Press, 2014: 11~15.
- [25] Dasigi, P., Guo, W., Diab, M.. Genre Independent Subgroup Detection in Online Discussion Threads: A Pilot Study of Implicit Attitude Using Latent Textual Semantics[A]. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2[C]. New York: ACM Press, 2012:65~69.
- [26] L.J. Austin. How to Do Things with Words [M]. Oxford: Oxford University Press, 1962.
- [27] Bakharia, A., Heathcote, E., Dawson, S.. Social Networks Adapting Pedagogical Practice: SNAPP[J]. Same Places, Different Spaces, 2009: 49~51.
- [28] Wen, M., Yang, D., Rosé, C.P.. Sentiment Analysis in MOOC Discussion Forums: What Does It Tell Us[A]. Proceedings of Educational Data Mining[C]. 2014: 1~8.
- [29] M.K., Underwood, L.H., Rosen, D.. More, et al. The BlackBerry Project: Capturing the Content of Adolescents' Text Messaging [J]. Developmental Psychology, 2012, 48(2): 295~302.
- [30] Ehrenreich, S.E., Underwood, M.K., Ackerman, R.A.. Adolescents' Text Message Communication and Growth in Antisocial Behavior Across the First Year of High School[J]. Journal of Abnormal Child Psychology, 2014, 42(2): 251~264.
- [31] Robinson, R.L., Navea, R., Ickes, W.. Predicting Final Course Performance From Students' Written Self-Introductions: A LIWC Analysis[J]. Journal of Language and Social Psychology, 2013: 1~11.
- [32] Romero, C., López, M.I., Luna, J.M., et al. Predicting Students' Final Performance from Participation in On-line Discussion Forums [J]. Computers & Education, 2013, 68: 458~472.
- [33] A. Ramesh, D. Goldwasser, et al. Understanding MOOC Discussion Forums Using Seeded LDA [A]. Proceedings of the 9th ACL Workshop on Innovative Use of NLP for Building Educational Applications[C]. New York: ACM Press, 2014:28~33.
- [34] Simsek, D., Shum, S.B., De Liddo, A., et al. Visual Analytics of Academic Writing [A]. Proceedings of the Fourth International Conference on Learning Analytics And Knowledge[C]. New York: ACM Press, 2014: 265~266.
- [35] C Hsu, J.L., Chou, H.W., Chang, H.H.. EduMiner: Using Text Mining for Automatic Formative Assessment [J]. Expert Systems with Applications, 2011, 38(4): 3431~3439.

- [36] Munezero, M., Montero, C.S., Mozgovoy, M., et al. Exploiting Sentiment Analysis to Track Emotions in Students' Learning Diaries [A]. Proceedings of the 13th Koli Calling International Conference on Computing Education Research [C]. New York: ACM Press, 2013: 145~152.
- [37] Gašević, D., Mirriahi, N., Dawson, S. Analytics of the Effects of Video Use and Instruction to Support Reflective Learning [A]. Proceedings of the Fourth International Conference on Learning Analytics And Knowledge [C]. New York: ACM Press, 2014: 123~132.
- [38] Rosen, D., Miagkikh, V., Suthers, D.. Social and Semantic Network Analysis of Chat Logs [A]. Proceedings of the 1st International Conference on Learning Analytics and Knowledge [C]. New York: ACM Press, 2011: 134~139.
- [39] De Liddo, A., Shum, S.B., Quinto, I., et al. Discourse -Centric Learning Analytics [A]. Proceedings of the 1st International Conference on Learning Analytics and Knowledge [C]. New York: ACM Press, 2011: 23~33.
- [40] Chen, B., Resendes. Uncovering What Matters: Analyzing Sequential Relations among Contribution Types in Knowledge - Building Discourse [A]. Proceedings of the Fourth International Conference on Learning Analytics and Knowledge [C]. New York: ACM Press, 2014: 226~230.
- [41] Chiu, M. M., Fujita, N.. Statistical Discourse Analysis of Online Discussions: Informal Cognition, Social Metacognition and Knowledge Creation [A]. Proceedings of the Fourth International Conference on Learning Analytics And Knowledge [C]. New York: ACM Press, 2014: 217~225.

Research on the Application of Learning Analytics Based on Text Mining

LIU San-ya, PENG Xian, LIU Zhi, SUN Jian-wen, LIU Lin, ZHENG Nian-heng

[Abstract] Current research on learning analytics in China and abroad mainly focuses on structured data generated by learners in networked learning environment. However, because of the diversity of learning interaction patterns, more and more unstructured data primarily consisted of text are generated. In recent years, evaluating learners' knowledge and ability as well as identifying their psychology and behavior by mining text data have become a new trend in learning analytics. This paper first discusses the concept of learning analytics and text mining. It then reviews studies of learning analytics that utilize text-mining technology. Next, this paper introduces the data sources, methods and tools of text mining in learning analytics. This paper then elaborates specific examples of the application of text mining in learning analytics from six aspects: course evaluation and support, learners' knowledge and ability evaluation, grouping in learning community, learning behavior crisis warning, learning effect prediction, and visualization of learning statues. At conclusion, this paper discusses the future research.

[Keywords] Educational Big Data; Learning Analytics; Text Mining; Text Analytics