



近20年在线学习环境研究评述*

——基于LDA和DTM的动态分析

□ 牛晓杰 郑勤华

【摘要】

在线学习环境作为远程教育中教与学整合的媒体中介,是远程教育活动顺利开展的重要保障之一。对远程教育中在线学习环境发展动态的探索有助于进一步理解其发展路径,深刻和全面地把握其发展情况。本研究以经过人工筛选后的574篇CNKI数据库的中文文章和362篇Web of Science核心数据库的英文文章的摘要作为文本数据源,采用LDA主题分类模型和DTM动态主题模型作为分析方法,总结出近20年来国内外在线学习环境发展中“经验迁移与平台建设”“在线学习环境要素模型建构”“技术驱动的在线学习环境理念革新”“互联网背景下开放共享式在线学习环境构建”四个主要发展阶段以及每个阶段的发展特征,研究结论可为开展远程教育中的在线学习环境研究提供有益参考。

【关键词】 在线学习;远程教育;人工智能;平台建设;在线学习环境;学习环境;文献主题识别算法;文献分析

【中图分类号】 G434

【文献标识码】 B

【文章编号】 1009-458x(2021)7-0025-12

DOI:10.13541/j.cnki.chinade.2021.07.004

一、引言与问题提出

百年来,远程教育从那颗源于19世纪中叶出现在英国的函授教育萌芽开始,不断发展变化着,经历了函授教育、多媒体教学的远程教育和开放灵活的远程学习三个代表性的发展阶段(丁兴富,2000)。学习环境通常指促进学习者发展的各种支持性条件的综合体(钟志贤,2005)。远程教育中的学习环境作为远程教育中教与学再度整合的媒体中介,是保障远程学习教学交互顺利发生的基础(王志军,等,2016)。

随着计算机技术、网络技术和多媒体技术的发展,远程教育中学习环境的内涵也发生着翻天覆地的变化。最初学习环境的建立能够基本支持打破时空限制的远程教育活动的开展,学习交互形式单一,后随着技术和理念的发展增加了多种类型的学习交互(张伟远,2009),再到后来能够为学习者提供全面、贴身式的学习支持服务,技术驱动下远程学习环境的发

展愈加丰富和立体。

在线学习是一种学习过程基于技术媒介、完全通过因特网完成的师生之间时空分离的学习,是远程教育的一种形式(韩锡斌,等,2015)。换言之,在线学习环境特指以技术为媒介,基于因特网促进学习者在在线学习和发展的各种支持性条件的综合体。在线学习环境是以网络学习双向互动为主的第二代远程学习环境和基于一站式网络教学服务的第三代远程学习环境中主要的和重要的学习环境形式(卢方,等,2016)。最近互联网技术的迅猛发展与以“联通主义”为代表的新型学习理论的注入,赋予了在线学习环境新的机遇和新的可能。基于历史发展演变的视角对在线学习环境变化开展分析,有助于重新梳理和审视整个在线学习环境的发展脉络,明晰关键研究问题与研究进展,并通过不同时期在线学习环境的对比分析,挖掘在线学习环境中内在发展的规律与机制。

在远程教育领域,诸多学者在不同时期对在线学习环境的研究是理解其变迁的良好材料。因此,本文将使用远程教育语境中“在线学习环境”相关文献,使

* 本文系科技创新2030新一代人工智能重大项目“混合增强在线教育关键技术与系统研究”(项目编号:2020AAA0108800)的研究成果。

用文献主题识别算法 (Latent Dirichlet Allocation, LDA) 分年度探究文献主题变化情况。为进一步体现主题时序动态发展的状况, 研究选择在LDA基础上纳入时序发展要素考虑的改进算法 (Dynamic Topic Modeling, DTM) 进行文献分析, 最终根据结果分析远程教育中“在线学习环境”主题的发展特点、发展动态, 并就发现的结论与研究本身进行讨论。具体研究问题包括以下两个方面:

(1) 基于LDA和DTM分析的远程教育视角下的在线学习环境研究发展至今经历了几个主要阶段? 每个阶段的特征 (主题分布) 是什么? 如何进行描述?

(2) 基于DTM动态演化分析的远程教育视角下的在线学习环境研究发展变化的过程是怎样的? 如何解释这种变化?

二、定量文献分析相关研究

文献分析方法发展至今, 产生了适应不同需要的丰富的分析方法, 从定性分析到定量分析, 从文本描述到信息可视化, 从基于词频统计到基于模型的机器学习方法等 (王曰芬, 等, 2016)。LDA主题模型是数据挖掘尤其是文本挖掘和信息处理方面不可或缺的三层贝叶斯文本建模模型, 由Blei首先提出 (Blei, Ng, & Jordan, 2003), 随后掀起了主题识别与探测研究的热潮。LDA模型可以将文档的主题以概率分布的形式给出, 通过分析一些文档并抽取出它们的主题分布后, 便可以根据主题分布进行主题聚类或者文本分类。文献计量中LDA也是常用的进行主题提取的模型之一。然而, 不同主题之间存在相关性和继承性, 同一主题基于时序变化也会表现出动态性、发展性和差异性。此情景下LDA的功能是不足的, 因此各种纳入时序因素考虑的基于LDA的改进算法应运而生, 如将时间信息结合到LDA模型的TOT模型 (Wang & McCallum, 2006)、文本集合的后离散概率模型 (Griffiths & Steyvers, 2004)、文本集合的先离散概率模型DTM (Blei & Lafferty, 2006) 等。DTM模型是由LDA模型的开发者提出的, 它将文本数据按照时序进行离散化切片, 假定相邻时间片上主题分布以及主题内容都是随时间演化, 进而识别出时间连续的文本数据集合的主题链。

教育领域的文献分析研究呈现出多角度、多方面

的分析和呈现内容。表1展示了近几年国内教育技术领域基于文献分析开展的部分研究案例。从文献的数据来源来看, 国内文献数据多来源于CNKI数据库中CSSCI核心文章, 国外文献数据多来源于Web of Science平台中的核心论文或者领域内知名的SSCI索引期刊; 从使用的方法或工具来看, 近几年采用纯量化方法的研究占多数, 涉及DTM、LDA等主题分类模型, TF-IDF等分词算法, SPSS、Ucinet、Bibexcel等统计分析软件, 以及HistCites、CiteSpace等可视化软件。使用定性方案的案例在数量上不如使用量化方法的多, 但是定性分析方法是一种非常重要的分析方法, 能够达到量化方法所不能达到的深度水平, 如批判性分析法、内容分析法等, 也有研究结合定性和定量分析 (毕经美, 2016), 使分析结果更为立体。

表1 近几年国内教育技术领域基于文献分析的研究

作者	数据来源	方法/工具	分析内容与结果
徐晓雄, 等 (2016)	Web of Science 文献数据	HistCites 引文可视化软件	慕课领域的引文脉络、发展趋势
张蓉菲, 等 (2019)	Web of Science 数据库中文献关键词	词频统计: SATI 3.2 共词聚类: SPSS 19.0 知识图谱分析: Ucinet16	国外教育人工智能的四个主题研究领域和三个研究趋势
刘震, 等 (2019)	Web of Science 核心集	CitesPace 软件 批判性阅读方法	国外在线继续教育的两个阶段、四大主题
赵以霞, 等 (2019)	CNKI 数据库	CiteSpace 软件	国内学习分析技术研究、数据挖掘、个性化学习研究的发展演变路径
董伟, 等 (2019)	CNKI 数据库 Web of Science 数据库	高频词分析: TF-IDF 主题分析: DTM	智慧教育方面国内外的主题发展差异
徐杰, 等 (2018)	Web of Science 核心集	词频矩阵: Bicomb 系统 多元统计分析: SPSS	五个角度的国际游戏化学习的研究热点
兰国帅, 等 (2017)	Web of Science 数据库中的18本SSCI来源期刊	引文分析: Bibexcel 多元统计分析: SPSS	国外教育技术十大研究领域
毕经美, 等 (2016)	CSSCI 和北大核心期刊数据库	内容分析法	慕课在国内的研究现状、进展和发展趋势

三、研究思路与方法

(一) 研究思路

本研究的整体思路如图1所示。DTM模型是作为LDA模型在时序演变方面的补充, 主题分类结果



以LDA为主，在结论和讨论部分会参考DTM模型的在时间变化方面的结果。

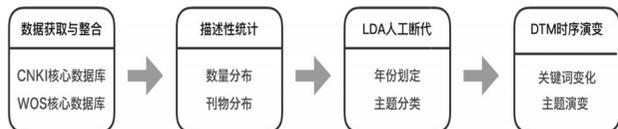


图1 研究思路示意图

具体研究计划分为“数据获取与整合”“描述性统计”“LDA人工断代”“DTM时序演变”四个部分。首先通过CNKI和WoS数据库收集与学习环境相关的研究文献，在简单的描述性统计的基础上使用LDA主题分类模型对文章进行主题的划分，考虑到“学习环境”的内涵发展是一个动态的过程，需要纳入时间的因素，因此进一步采用纳入时间维度的优化模型DTM进行计算，就最终得到的结果和文献情况进行讨论与解释。

(二) 研究方法

1. LDA主题分类模型

LDA是文本语义分析中应用广泛的一个模型，一个统计上的前提情景为：当某人写一篇文档时，会先根据一定的概率选定主题，然后根据与选定主题相关的概率生成文字。LDA遵循贝叶斯思想，文档-主题分布概率和主题-词分布概率是不确定的，但是它们服从一个已知参数的Dirichlet分布。详细的模型介绍可以参见Blei等(Blei, et al., 2003)的文章。图2是LDA的模型图，主要分为两个过程：

(1) 从Dirichlet($\bar{\alpha}$)的分布中生成针对一篇文档 m 的主题分布 $\bar{\theta}_m$ ，然后生成文档 m 的每个文字，对第 n 个字根据 $\bar{\theta}_m$ 给该词分配一个主题 $z_{m,n}$ ；

(2) 从Dirichlet($\bar{\beta}$)的分布中生成 K 个主题-词分布 $\bar{\varphi}_k$ ，选择编号为 $z_{m,n}$ 的 $\bar{\varphi}_{z_{m,n}}$ ，并根据这个分布生成文字 $\omega_{m,n}$ 。

2. DTM动态主题模型

DTM是在LDA模型的基础上融入时间因素的改进模型，能够挖掘文档主题，分析不同主题在时间序列中的演化趋势。图3是DTM模型的图(以三个时间片为例)，每个主题的自然参数 $\beta_{t,k}$ 以及主题比例的逻辑正态分布平均参数 α 随着时间而演变。从图3可以看出，当代表时间变化的水平箭头被移除时，模型打破时间动态，简化为一组独立的主题模型。利用时间动态，时间 t 处的第 k 个主题从切片 $t-1$ 处的第 k 个

主题平滑演化。详细模型说明可见Blei(2006)的文章。

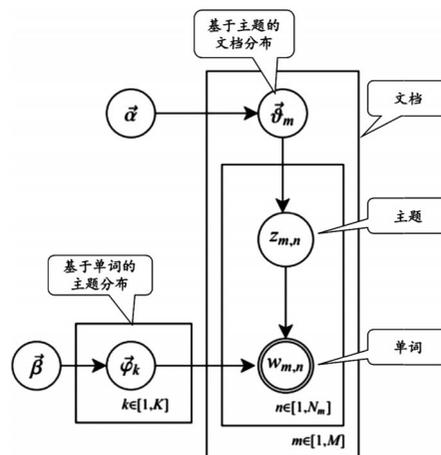


图2 LDA模型表示图(Blei, Ng, & Jordan, 2003)

(注：图中 K 为主题个数， M 为文档个数， N_m 为第 m 篇文档的主题个数。)

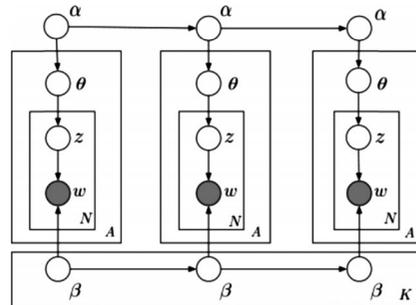


图3 DTM模型表示图(Blei & Lafferty, 2006)

四、研究结果

(一) 在线学习环境文献量化统计

为从整体上把握在线学习环境研究的初步情况，首先就在线学习环境文献发文进行主题搜索和描述性统计，此部分呈现简易量化文献综述的结果。

1. 数据获取与整理

中文文章基于中国知网CNKI平台获取，检索“篇名”包括“学习环境”并且发文时间在2020年1月19日之前的CSSCI期刊文献。检索方式没有选择“主题”包括“学习环境”的原因是搜索结果有相当数量并不是以“学习环境”为核心研究对象，研究者对比了“主题”“篇名”的检索情况，为了保证最后结果的准确性，因此将范围限制在了“篇名”。经过人工筛选，最终得到574篇中文期刊学术论文，选取其年份、期刊名、作者、题名、摘要、关键字字段作

为有效数据。

英文文章基于 Web of Sciences 数据库, 检索“topic”条目包括“learning environment”并且发文时间在2020年1月19日之前的特定期刊文章。由于本研究对象为远程教育语境下的在线学习环境, 因此首要考虑远程教育领域 SSCI 期刊, 最终结合数据检索情况确定 *Distance Education*、*American Journal of Distance Education*、*British Journal of Educational Technology*、*International Review of Research in Open and Distance Learning* 和 *The Turkish online journal of distance education* 五本 SSCI 为特定检索期刊。经过人工筛选, 最终得到 362 篇英文期刊学术论文, 选取其年份、期刊名、作者、题名、摘要字段作为有效数据。

值得说明的是, 在中英文文献中的“人工筛选”环节, 研究者依次浏览文章的摘要、架构和大致内容, 在此过程中首先筛选以“远程教育”为研究领域并且以“在线学习环境”为研究对象的文章。之所以在检索阶段不直接搜索“在线学习环境”和“Online learning environment”, 是因为研究者发现存在部分文章直接将“学习环境”默认成“在线学习环境”, 并且未在标题、摘要和关键字中进行具体说明, 因此需要由研究者人工判定其是否属于基于因特网技术的在线学习环境研究; 然后, 去除不符合研究主题的文章, 如研究面授学习环境的文章、访谈类文章、期刊卷首语、主编寄语、数据信息错误或者不全面的文章等, 确保最终入选的研究文章与“在线学习环境”主题研究相关并信息全面、完整。

2. 基于文献数据的描述性统计

对数据进行描述性统计可得, 英文文献数量在 2011 年和 2017 年的表现较为突出, 整体上呈现一个波动上升的趋势。英文文献中在线学习环境的发文数量正逐渐增加, 到研究统计截止的 2019 年其数量达到一个顶峰, 见图 4。从期刊数量分布来看, 英文文献中在线学习环境的文献较为均匀集中分布在 *Distance Education*、*British Journal of Educational Technology*、*International Review of Research in Open and Distance Learning* 和 *The Turkish online journal of distance education* 上。*American Journal of Distance Education* 中在线学习环境相关研究较少, 数量大大少于其他期刊; *British Journal of Educational Technology* 期刊的在线学习环境相关研究数量最多, 远超其他英文期刊

(如图 5)。

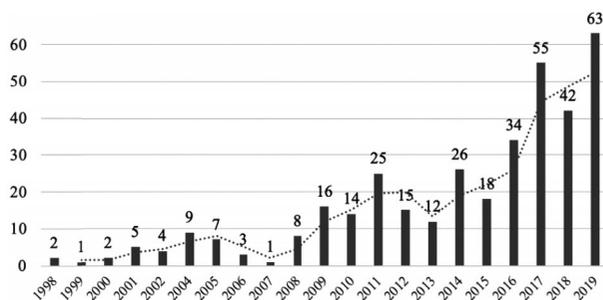


图4 英文文献年代数量分布图 (单位: 篇)

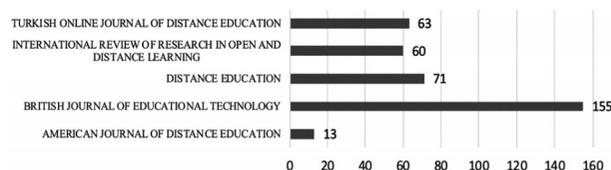


图5 英文文献期刊数量分布图 (单位: 篇)

如图 6 所示, 中文文献在 2011—2016 年间的数量有一个高峰, 2008—2019 年间文献数量相差不多, 因此中文在线学习环境相关文献在整体发文数量上呈倒 U 形的趋势。2011—2016 年产生发文数量上的相对高峰并持续了一段时间, 然而 2017 年至今数量逐年下滑。在发文期刊统计方面, 图 7 呈现了发文数量超过 5 篇的期刊名, 《中国电化教育》居首位, 总计有 105 篇, 《电化教育研究》《现代教育技术》《开放教育研究》《中国远程教育》也是在线学习环境发文较多的期刊。

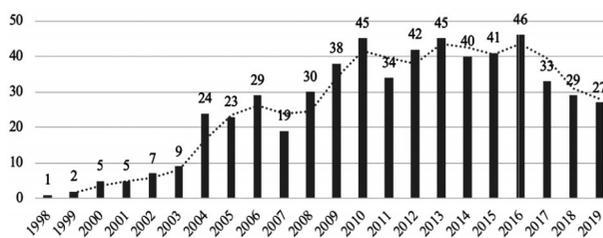


图6 中文文献年代数量分布图

上述基于文献的描述性统计能够看出 20 世纪末至今, 在线学习环境相关研究的发文数量、期刊分布。然而就分析结果而言, 目前这种统计性分析存在两个问题: ①无法更加确切地挖掘在线学习环境具体的内涵变化, 即信息呈现粒度不够, 需要能开展更深层次挖掘的方法和表征方式; ②阶段特征不明显, 无法从现有数据中提取在线学习环境发展的阶段及其主要特征和规律, 需要探索能够提炼其阶段特征的方式进行补充说明。

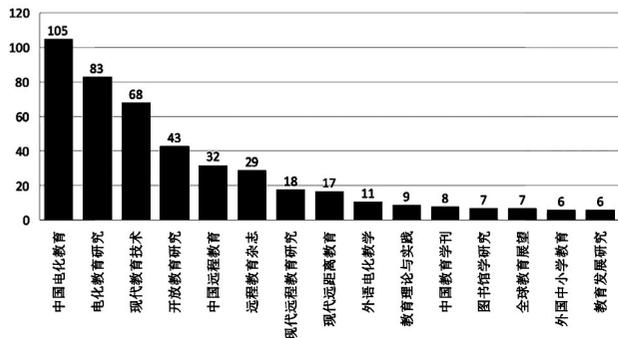


图7 中文文献期刊数量分布图 (单位: 篇)

(二) 关于在线学习环境研究的时代划分

1. LDA人工断代

本研究使用北京师范大学远程教育研究中心自主开发的DMTS可视化平台,平台将LDA的算法进行集成,使用Python语言研发可视化操作程序,目前平台正在内部测试之中。图8为DMTS平台的简易界面,其上为词频统计图,下为各主题分类及对应主题词。经过调试,研究发现将主题设定为4个时结果最好,因此LDA模型最终形成3个主题,在每个主题下面形成的概率排名前10的词语中选择合适的主题词用于分析。

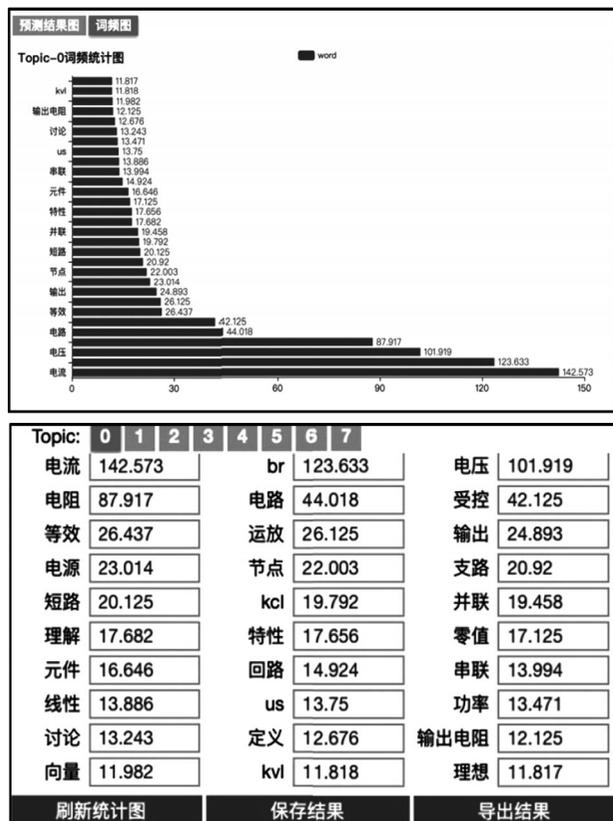


图8 DMTS系统中LDA的运行界面示例

依据文献时间分布和数量,人工将文献进行断代,分别进行LDA分析。人为设定断代的原则是五年为一代,若五年内包含文献数量较少则进行合并分析。最终中文文献分为三代,为1998—2009年、2010—2014年、2015—2019年三个时间段,英文文献分为四代,为1997—2004年、2005—2009年、2010—2014年、2015—2019年。人工断代较为均匀,说服力不强,因此后续的DTM作为年代发展的补充。

中文设置的停用词为“教学”“教师”“学生”“研究”“环境”“学习”;英文设置的停用词为“Learning”“Online”“Education”“Environments”“The”“We”“Based”“Study”“Learners”“Teaching”等一些参考价值低的实词和虚词。

2. DTM时序演变

DTM模型研究时需要首先对文本进行词的划分。对于中文文献,研究使用Python语言,通过jieba分词包(代码见<https://github.com/fxsjy/jieba>)进行中文文本分词,最终得到6,629个词语。对于英文文献,使用DTM中对于英文的分类方式text2ldac,Python代码以及使用指导(见<https://github.com/JoKnopp/text2ldac>),最终得到6,495个词语。

研究采用Blei发布的C语言编写的DTM源码进行建模(代码见<https://github.com/Blei-Lab/dtm/tree/master/dtm>)。DTM断代结果是基于时间演变的主题词结果和主题强度变化产生的,同一个时间段内的主题词和主题强度具有相对稳定性。最终,中文文献产生6个主题和4个时间段(1998—2000年、2001—2007年、2007—2012年、2013—2019年),英文文献产生4个主题和4个时间段(1998—2004年、2005—2012年、2013—2015年、2016—2019年)。

(三) 在线学习环境研究的主题分布情况

表2是中英文基于LDA分年代分词的结果。

1. 中文文章LDA情况

从1999年到2009年的四个主题可以分别用课程开发、研究方法、平台实现和教学设计来概括。这一时期在线学习环境的应用探索主要是从课程出发,从课程角度开展教学设计;出现了一批关于平台建设的实例或者学习平台建设新理念(钟志贤,2005),如基于建构主义的学习环境设计(杨开城,2000)。

从2010年到2014年,主题的聚焦点可以总结为资源开发、模型开发、以协作为代表的学习策略研究和评价体系的发展。在线学习环境的发展从其环境组成的要素入手,关注其中的资源、课程、策略、评价等各个环节的设计发展。在线学习环境的设计开始走向数据驱动,强调分析技术的必要性(王兄,2010)。在基于要素研究的基础上,从整体视角出发构建的模型也是这一时期的热点,适用于不同场景、不同目的的在线学习环境设计模型相继涌现,如基于语境感知的泛在学习环境模型(张洁,2010)、基于问题的情景式学习环境(谢涛,等,2010)、认知学徒制学习环境(陈家刚,2010)等。

表2 基于LDA的中英文主题跨年代分类结果表

时间段	中文主题词	时间段	英文主题词
1998—2009年	课程、探究、分析、科学、活动	1998—2004年	Distance, Adult, Constructive, Collaborative
	评价、质性、量化、方法、范式		Web, Experiment, Evaluation, Cognitive, Results
	设计、网络、构建、PLE、平台		Face, Design, Instructional, Development, Inquiry
2010—2014年	网络学习、学习者、思维、元认知、虚拟	2004—2009年	Retention, Quality, Sharing, Educational, Support
	设计、资源、文化、共同体、数据		Distance, Skills, Perceptions, Courses, Open
	模型、资源、分析、信息、服务		Worlds, Virtual, Active, Submission, Group
2015—2019年	网络学习、学习者、协作、目标、导向	2010—2014年	Social, Distance, APP, Web
	评价、技术、互动、实践、课程		Virtual, Flow, Social
	动机、情感、虚拟、学习者、调节		Mobile, Technology, Virtual, Research, Model
—	交互、学习者、具身、知识、支持	2015—2019年	Educational, Peer, Mobile, Open, Paper
	智慧、探究、科学、信息化、发展		Self, Analysis, Factors, Instructional, Presence
	智慧、知识、情景、要素、场所		Research, Social, Collaborative, Self
—	—	—	Design, Social, Collaborative, Communication, Analysis
—	—	—	Video, Social, Page, VL, Open
—	—	—	Data, Attention, Quality
—	—	—	APP, Teachers, Students, Design

从2015年到2019年,在线学习环境的焦点聚集在学习因素、环境交互、智能化学习环境方面。研究角度会从不同的在线学习环境对学习效果的实证检验方面开展,关注在线学习环境的设计对学习者的学习表

现、知识建构、能力生成的影响情况。另外,随着人工智能技术和互联网技术的迅猛发展,更加智能化、智慧化的学习环境设计被应用起来,如虚拟环境中虚拟教师的设计(夏志鹏,等,2016)、MOOCs学习环境中视频标注技术的研究(徐春华,等,2017)、基于云计算技术的智慧学习环境(杨澜,等,2018)、基于虚拟现实(VR)的学习环境(祝士明,等,2019)等。

2. 英文文章LDA情况

从1998年到2004年,在英文文章中在线学习环境的研究主题可以总结为基于建构主义的理论基础、网络环境效果、教学设计、质量保障。有研究者将建构主义与成人学习理论联系起来,主张使用建构主义的理论来设计在线学习环境(Huang,2002)。也有学者关注基于网络的学习环境的设计效果,关注从教学设计出发对环境的设计,关注对学习环境的评价和改进方式。

从2005年到2009年,在英文文章中在线学习环境的研究主题可总结为学习因素、虚拟环境、学习策略、社交。典型的虚拟环境有在线学习中的3D虚拟学习环境(Dickey,2005)和个性化移动的分布式学习环境(Ahmed & Sadeq,2006)等。社交作为一个关键词体现出了相关研究中对于学习环境之中开展社交的重视,如了解某学习环境中学生的社交情况(Kehrwald,2008)、设计SL学习环境中的社交支持等。也有学者对在线学习的问题进行反思,建议开发集成式的学习环境来切实提升学习效果(Hoskins & Van Hooff,2005)。

从2010年到2014年,在英文文章中在线学习环境研究的聚焦点可以总结为移动学习环境、模型构建、学习者表现和在线学习环境中的社交。似乎这一阶段与2005年到2009年间的差别不大,从各主题的主题词来看,这一阶段有着更为明显的社交学习的考虑,体现为对远程教育视角下计算机支持的协作学习环境的(CSCL)研究(O'Neill & Conboy,2011)。此阶段在理论层面出现了较为体系化且较为成熟的、适用于在线教育学习环境的学习理念和学习理论,如自我调节的学习(Barnard-Brak, Paton, & Lan,2010)、联通主义(Boitshwarelo,2011)等。另外,2012年作为慕课元年,一批关注构建慕课学习环境的研究也相应出现了。

从2015年到2019年,在英文文章中在线学习环



境研究的聚焦点为社交活动设计、视频研究、注意力、教学设计。社交活动设计延续了对社交的一贯关注,体现的是深刻了解之后的设计与干预。视频研究和注意力的研究则是从研究材料角度讲的,是对在线学习环境的构建倾向使用更加客观、更加真实的自然情景中的数据。教学设计则仿佛是一个轮回:从20世纪末教学设计角度出发构建学习环境到近20年后回归到教学设计之中来。而此时的教学设计是学习环境诸多丰富要素之中的重要因素,有着丰富的研究结论做支撑,因此更加有效、更加个性化。

3. 针对中英文文献LDA结果的总结

从上述LDA分类结果出发,回顾近20年来中英文文献在线学习环境研究整体发生的历程,可以发现中英文文献关于在线学习环境研究的共性与差异。从整体来看,中英文文献在线学习环境的研究视角经历了“大-小-大”的基本发展阶段,最初关注在线学习环境整体的平台建设和理论架构,然后基于已有的研究经验进行总结、概括、提炼,产生适用于在线学习环境的设计或者理论模型,最近十几年开始进一步的优化和重新审视,纳入多方位、细致的学习支持设计和学习要素研究,如关注在线学习环境的交互、学生注意力的研究和智能学习环境的设计。

LDA分析结果也显示出了中英文文献在在线学习环境方面的差异。中文研究文献始于在线学习环境的平台建设和教学设计,并进一步探究基于在线学习环境的各要素设计,如不同教学模式、教学策略的适用性,学习资源的设计,协作学习活动的设计,等等;近年研究则更关注在线学习环境的交互支持,如重视在线学习环境中的师生交互、生生交互以及学生和资源交互的设计与研究。

英文研究文献与中文稍有差异,始于学习理论和学习效果的呼唤,如20世纪末对建构主义在线学习环境的关注和对提升远程效果和学习质量的研究,后续进一步关注学生学习体会与学习效果,如研究在线学习环境中适合的学习策略、基于在线学习环境的学生社交表现,近几年更加关注影响在线学习环境的细致因素研究,如对在线学习环境中呈现资源方式的研究、对视频种类与学生注意力方面的研究等。

(四) 学习环境研究主题的动态演化情况

1. 关键词演化

基于数据的DTM分析得到的关键词演化过程如

图9和图10所示。由结果可以看出,在中文文章中与“在线学习环境”相关的关键词有智慧、认知、网络、设计、个人、具身、无障碍、远程、感知、分析、化学、建构主义、模型、技术等。在英文文章中与“在线学习环境”相关的关键词有Tool、Virtual、Internet、Web、Instructional、Self-regulated、3D、Learning、Achievement、Instructors、Format、Quality、Interaction、Engagement、Individuals、Ability等。从时间发展的角度看,中文文献发展根据关键词的变化被划分成四个阶段,分别为1998—2000年、2001—2007年、2007—2012年、2013—2019年;英文文献被划分成四个阶段,分别为1998—2004年、2005—2012年、2013—2015年、2016—2019年。

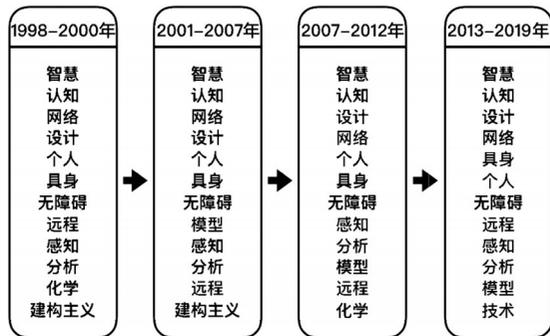


图9 中文DTM主题关键词演化情况

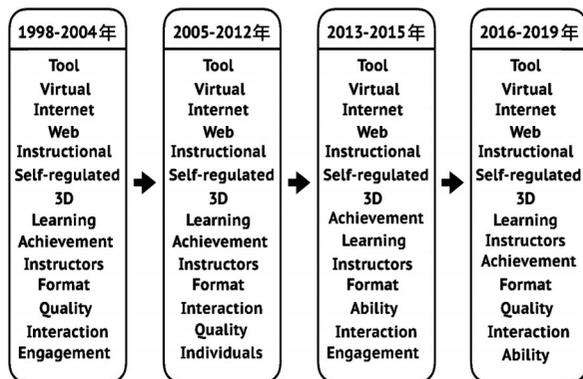


图10 英文DTM主题关键词演化情况

从关键词的变化来看,整体上国内外关键词的变化并不大。国内“模型”这个关键词在2001—2007年有较高的发生概率,对在线学习环境的模型构建关注度较高。“具身”近几年发生概率排位上升,是具身认知理论结合到在线学习环境设计中的一个特别体现。“技术”在2013—2019年作为热点进入关注视野,与新技术的迅速发展有着密切关系。在英文文章中2005—2012年首次出现“Individuals”,意味着对个

性化和个体学习的关注;“Instructor”“Achievement”“Ability”“Engagement”的出现和变化,说明从时间发展的角度看,近年国际上关于“在线学习环境”的关注着力基于环境提供的指导、学生的学习成就、能力生成以及学生在在线学习环境的沉浸度方面。

2. 主题演化

DTM对数据进行主题分类,并给出各主题下包含的主要关键词。中文文献划分出六个主题,研究赋予其中五个主题以具体含义,如表3所示;英文文献划分出四个主题,研究赋予其中三个主题以具体含义,如表4所示。中文主题四和英文主题一由于关键词宽泛而零散,难以赋予具体命名,因而暂用“主题四”和“主题一”来指代。图11为中英文主题随时间变化的趋势图,横轴为年份,纵轴为取幂运算的主题强度——主题强度是指每个主题在某时刻被生成的概率,一般被用于描述主题的受关注度或活跃度,将该强度置于时间序列中可以作为衡量主题演化的指标(李保利,等,2012)。

表3 中文DTM主题划分

主题一 功能属性	科学、社区、文化、真实、空间、工具、活动、共享、虚拟	主题二 理论特征	智慧、具身、感知、认知、支持、技能、创新、信息
主题三 技术方法	智能、数据、分析、混合、知识	主题四	认知、网络、无障碍、远程、分析
主题五 课程设计	设计、个人、化学、模式、课程	主题六 模型设计	网络、远程、建构、模型、设计

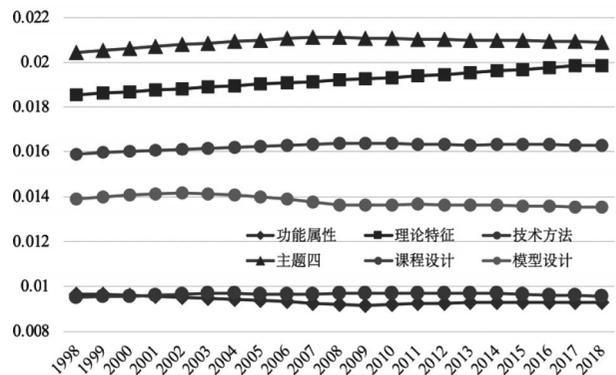
表4 英文DTM主题划分

主题一	Learning, Online, Distance, Students	主题二 工具支持	Tool, Achievement, Instruction, Internet, Web, Evolution, Space, Active, Inquiry, Psychology, Ability
主题三 环境交互	Instructors, Interaction, Engagement, Social, Individual, Questions, Improve, Create, Interest	主题四 学习策略	Virtual, Instructional, Self-regulated, 3D, Format, Quality, Cultural, Achieve

由图11可知,中文数据中“理论特征”主题在保持上升趋势,体现出“在线学习环境”的应用与实践对相关理论的丰富与发展,并且“理论特征”是所有主题中强度最高的主题。这一点是区别于LDA分析结果的新发现。“理论特征”下包含的词语主要有“具身”“认知”“感知”等,具身认知学习环境是一种

基于具身认知理论的身心融合、主客一体下建立起来的多种内嵌性的在线学习环境(李志河,等,2018)。具身认知的提出和研究对在线学习环境设计发展起了推动作用,是在线学习环境设计的重要理论基础之一。英文数据中三个主题基本保持平稳的趋势,“环境交互”主题稳中上升,“学习策略”主题是所有主题中强度最高的主题。这一点与英文文献的LDA相呼应,也进一步印证了“学习策略”在在线学习环境英文研究文献中是需要关注点。

中文文章主题随时间变化趋势图



英文文章主题随时间变化趋势图

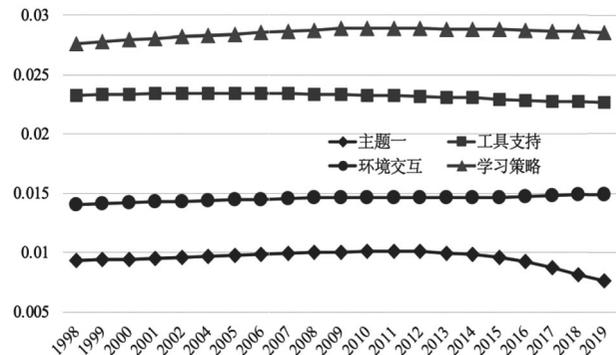


图11 中文文章(上)和英文文章(下)DTM主题随时间变化趋势图

五、研究结论

基于LDA和DTM模型的分析 and 计算,结合文献综述,最终本研究将近20年国内外远程教育中“在线学习环境”的发展分为如下四个阶段:

1. 20世纪末到21世纪初:基于课堂学习环境的经验迁移与平台建设

这个时期是第三代远程教育的起步阶段,远程教育进入开放灵活的在线教育时代。在线学习环境的建



设成为远程教育发展的要素,此时正在对其进行初步摸索。这一时期远程教育中的在线学习环境明显借鉴了课堂面授教学的经验,尤其是在教学模式和教学策略设计上缺乏远程教育特有的灵活性,而此时在线学习环境设计和建设的经验为后续进一步优化发展提供了良好的环境。

2. 21世纪初到21世纪10年代初:建构适用于在线学习环境的要素模型

经过一段时间的发展,在前一段时期的基础上有了较为成熟和深入的视角。技术运用方面变得更加体系化和全面化,如基于SNS网络技术设计的学习环境、人机交互技术的应用、SchoolLife平台的探索、桌面学习环境的应用尝试等。技术驱动下学习平台设计开始多样化、智能化、虚拟化,逐步体现出远程教育的灵活性,出现了众多类型的学习环境,如虚拟学习环境(VLE)、个人学习环境(PLE)、协同学习环境、混合学习环境等。此时的研究基于上个阶段的建设经验展开,注重从信息、社会和知识管理的角度探索适用于在线学习环境的整体设计模型,包括在线学习环境下各要素的设计研究模型,如在线学习支持模型、在线学习环境评价模型等。技术发展和研究视角的转变体现出此阶段远程学习关注的主体逐渐向学习者转移,更多强调学习者的感受与学习效果。

3. 21世纪10年代初到21世纪10年代中期:技术驱动下的在线学习环境理念革新

这一时期积累了丰富的学习环境设计实践与研究成果,出现了新思路、新理念、新方法。在技术运用方面,体现出跨学科的、复杂的、系统性的理论与工具运用形式,如工作流技术、大数据技术等。在线学习环境设计思路重视适合特定学习环境的学习策略设计,重视课程与学习方法的整合,关注相应的在线学习模式构建。丰富的理论发展与技术变革使得新型的学习环境设计大量涌现,如富交互型学习环境、参与型学习环境等各要素导向的学习环境,这种在线学习环境的创新并不仅仅是先进技术的应用,更体现在新的理念和新的应用方法上。2013年首次出现大数据支持和联通主义视角下的学习环境设计研究。

4. 21世纪10年代中期到21世纪10年代末:“互联网”浪潮下的开放共享式在线学习环境

这一时期人工智能技术、5G技术结合“互联网+”教育掀起的浪潮一波未平,一波又起,学习者

通过平板电脑或者手机参与到完全的网络学习环境中开展在线学习活动。在线学习环境的构建有了更加个性化、开放化、数据化的技术支持和设计理念,如具身认知的设计视角、分布式学习环境、生态学的设计视角以及基于大数据的设计视角等。在线学习环境已经不单单是知识共享的平台,更是知识创造的摇篮。技术的成熟应用为系统性的学习环境运行提供了良好的支持,这是经验积累与成果产出的新高潮,并且将持续性地促进新理论、新观念、新思路的进一步迭代。

综观近20年整个在线学习环境的发展变化,可以看出“技术到理念”“模仿到革新”的发展趋势。技术驱动是新理念、新模式产生的重要前提,此规律不仅适用于在线学习环境的发展历程,对于整个远程教育的发展也是如此。然而,从文献数据也可以看出,内涵日趋丰富的背后是鱼龙混杂的在线学习环境概念和模型,多种新型的在线学习环境理念甚至仅基于技术基础或者研究者个人经验便被提出,缺乏实践的检验和过滤。另外,关于在线学习环境建设的总结性、反思性研究不足,会导致经验积累追不上技术革新、理念指导跟不上产品迭代,使得技术实践应用效率低下,有造成资源浪费的风险,长此以往将不利于在线学习环境整体的建设与发展。

六、建议和反思

自20世纪末以来,在线学习环境的发展随着技术快速发展与理念的革新,正朝着开放化、信息化、大数据化和个性化的方向发展。从本研究结果来看,新技术、新应用促使学习环境一体化、共享化、创新化,优质高效的在线学习环境是当代远程教育所呼唤的。基于此,结合2020年新冠肺炎疫情期间我国在线教学实践反馈情况,本研究提出以下四个建议:

第一,强化基于数据驱动的“一站式”学习环境、多模态融合学习环境的研究与实践。大数据是一片包含着丰富信息的海洋,研究者和研发者可以采用数据挖掘技术和学习分析技术获取学习者个性化数据,如在线学习者自画像(肖君,等,2019)等。学习环境是开展数据收集和开展基于数据信息支持的重要媒介,如何基于特定的学习环境有效收集数据并利用学习分析结果为学习者提供优质的个性化服务,是

未来在线学习环境研究的追求。多模态融合学习环境是技术和理论整合的必要研究对象,有助于多功能、自适应、高效、精准学习环境的建立。

第二,开展远程在线学习环境建设标准、规范与评价的研究。我们正处于远程教育快速变革的时代,时代的特点就是积累了丰富的远程教育实践经验,未来研究可以聚焦如何利用现有的实践经验构建完整、具体和多类型的学习环境规范和评价模型,这对高效推动远程教育发展有着重要意义。疫情期间,复杂多样的在线学习环境在“停课不停学”实践中有着参差不齐的影响效果,规范、良好的在线学习环境能够有效、合理地利用学习数据进行学习过程监督和分析,提升教和学的效率;随意粗糙的在线学习环境甚至会影响正常的教和学,迫使教学各主体浪费时间和精力。2020年5月14日在教育部新闻发布会上,教育部高等教育司司长吴岩说:“我们再也不可能、也不应该退回到疫情发生之前的教与学状态。”融合“互联网+”与“智能+”技术的在线教学已经成为教育发展的重要方向。要达到这种状态,围绕在线教育行业各要素(学习环境、资源建设等)制定行业标准和规划是必要的环节,需要整合多类型在线学习环境的“散沙”,沉淀珍贵的经验和优质的学习环境,维持高效的使用和发展状态,更加彻底地革新在线教育。

第三,探索基于AI和5G技术的智能化在线学习环境。从研究结果可以看出,“技术驱动”是近20年来在线学习环境中内涵发展变化的重要推动因素,从技术创新到理念革新的发展规律是不可忽视的。当前人工智能技术和5G技术的兴起和迅速发展又为新型的在线学习环境建设提供了无限的可能,我们要把握时机,加快探索AI技术、5G技术与学习环境的融合,向智能化、开放化的全方位学习环境再迈进一步。

第四,开展动态可适配的环境计算研究。北京师范大学黄荣怀团队提出了数字学习环境的高端形态即智慧学习环境,并介绍了智慧学习环境对学习者的“3A”“4E”学习要求的支持(黄荣怀,等,2016)。动态可适配的学习环境计算指的是学习环境能够识别学习者特征、感知学习情境、利用计算和推理的方法,选择适合的教学交互策略,生成个性化学习路径,智能推荐匹配资源,从而建立一个智能、精准和全面的自适应学习支持服务(黄荣怀,等,

2019)。黄荣怀团队同时指出了学习环境计算研究的四种典型研究问题:保障数字隐私的可信学习环境安全体系、复杂场景下的基于多模态信息融合的学习支持服务机制、基于边云协同计算的学习环境计算模型、课堂教学支持系统的有效关联协同(黄荣怀,等,2019)。

七、结语

本文使用CNKI和WoS核心数据库为数据来源,采用LDA和DMT对近20年“在线学习环境”主题的中英文文章进行了动态分析,最后将“在线学习环境”近20年的发展划分成四个阶段,并总结了每个阶段的主要特征。

本研究有其局限性和不足之处。首先,从数据源角度来看,英文文献中的研究仅以远程教育领域几本期刊作为研究对象,未来研究可以考虑进一步扩大数据收集范围,构建体量与质量兼得的文献分析数据源。其次,研究中DTM主题词演变的变化并不是很明显,经分析认为有三点原因:①数据量不够。DTM属于概率模型,足够多的数据量能够将模型训练得更好,而本研究中中英文文献均各只有几百篇的数量级,后续研究可以酌情增加分析文本的数量。②数据的来源期刊需要更加广泛。研究本着收集高质量、有代表性和领域特色的文献数据,因此研究文章聚焦于远程教育领域的期刊,并且进行了人工筛选,这可能不利于探究“在线学习环境”在多学科领域中的多种发展状况。建议后续研究进一步扩大期刊检索范围,增加文献数量,以更加宏观的数据视角来归纳或验证“在线学习环境”的发展与变革历程。③主题过于聚焦。本研究的主题为远程教育中“在线学习环境”主题演变,采用概率模型进行主题分类不明显也可能是由词语之间语义相近、概率分布差别不大导致的。建议后续研究在采用量化方法进行文献分析时,将关注点适当扩大,从更为宏观的角度探究其内部的演变过程。最后,研究认为,基于量化的文献分析仍然是有其局限性的,需要进一步结合定性的理解与思考方可获得全面的认识,不要过于执着于量化分析结果而忽视了文献本身所体现的含义与价值。

[参考文献]

毕经美. 2016. 国内MOOCs文献研究的现状分析与趋势思考——



- 2009-2014年CSSCI和中文核心期刊文献的内容分析[J]. 黑龙江高教研究(3):134-139.
- 陈家刚. 2010. 技术支撑的认知学徒制学习环境设计[J]. 现代远程教育(4):68-72.
- 丁兴富. 2000. 三代信息技术和三代远程教育:远程教育中的信息技术和媒体教学(2)[J]. 中国远程教育(8):15-19.
- 董伟,陶金虎. 2019. 基于DTM的国内外智慧教育热点和主题演进比较[J]. 现代教育技术(7):4.
- 韩锡斌,王玉萍,张铁道,等. 远程、混合与在线学习驱动下的大学教育变革——国际在线教育研究报告《迎接数字大学》深度解读[J]. 现代远程教育研究,2015(5):3-11,18.
- 胡钦太,张晓梅. 2018. 教育信息化2.0的内涵解读、思维模式和系统性变革[J]. 现代远程教育研究(6):12-20.
- 黄荣怀,刘德建,方海光,等. 2017. 2016中国智慧学习环境白皮书[R]. 北京:北京师范大学智慧学习研究院.
- 黄荣怀,周伟,杜静,等. 2019. 面向智能教育的三个基本计算问题[J]. 开放教育研究,25(05):11-22.
- 兰国帅,汪基德,梁林梅. 2017. 国外教育技术十大领域与权威人物的知识图谱建构研究——基于18种SSCI期刊(1960—2016年)文献的可视化分析[J]. 远程教育杂志,35(2):74-86.
- 李保利,杨星. 2012. 基于LDA模型和话题过滤的研究主题演化分析[J]. 小型微型计算机系统,33(12):2738-2743.
- 李志河,李鹏媛,周娜娜,等. 2018. 具身认知学习环境设计:特征、要素、应用及发展趋势[J]. 远程教育杂志,36(5):81-90.
- 刘震,陈东. 2019. 近二十年国外在线继续教育研究综述——基于Citespace的可视化分析[J]. 清华大学教育研究(4):17.
- 卢方,尹学松,张吉先. 2016. 开放大学视域下的无缝学习环境设计[J]. 远程教育杂志,35(2):39-48.
- 王兄. 2010. 网络学习环境下的数据分析——评价的视角[J]. 现代教育技术,20(12):106-110.
- 王曰芬,傅柱,陈必坤. 2016. 基于LDA主题模型的科学文献主题识别:全局和学科两个视角的对比分析[J]. 情报理论与实践,39(7):121-126.
- 王志军,陈丽,韩世梅. 2016. 远程学习中学习环境的交互性分析框架研究[J]. 中国远程教育(12):37-42.
- 夏志鹏,刘革平. 2016. 三维虚拟学习环境中操作指导型虚拟教师设计与实现[J]. 中国电化教育(5):98-103.
- 肖君,乔惠,李雪娇. 2019. 大数据环境下在线学习者画像的构建[J]. 开放教育研究,25(4):111-120.
- 谢涛,刘革平. 2010. SecondLife中基于问题的情境式学习环境建构[J]. 现代教育技术,20(3):107-109.
- 徐春华,傅钢善. 2017. 视频标注工具支持的深度学习研究——以MOOC学习环境为例[J]. 现代教育技术,27(3):13-19.
- 徐杰,杨文正,李美林,等. 2018. 国际游戏化学习研究热点透视及对我国的启示与借鉴——基于Computers & Education(2013-2017)载文分析[J]. 远程教育杂志,36(6):73-83.
- 徐晓雄,郝建江,乔星峰. 2016. 基于Web of Science数据库的MOOCs文献分析[J]. 现代教育技术(4):61-67.
- 杨开城. 2000. 建构主义学习环境的设计原则[J]. 中国电化教育(4):14-18.
- 杨澜,曾海军,高步云. 2018. 基于云计算的智慧学习环境探究[J]. 现代教育技术,28(11):26-32.
- 赵以霞,王鑫,金昆,等. 2019. 国内大数据环境下学习分析技术研究路径及趋势分析. 现代教育技术,29(8),34-40.
- 张洁. 2010. 基于境脉感知的泛在学习环境模型构建[J]. 中国电化教育(2):16-20.
- 张蓉菲,赵磊磊,李明泓,等. 2019. 国外教育人工智能研究主题及趋势分析——基于Web of Science文献关键词的可视化分析[J]. 现代教育技术,29(12):5-12.
- 张伟远. 2009. 以互动为核心的网上教学原理及应用[J]. 现代远程教育研究(5):10-13,71.
- 钟志贤. 2005. 论学习环境设计[J]. 电化教育研究(7):35-41.
- 祝士明,陈静潇. 2019. 虚拟现实学习环境的作用、挑战以及应对策略[J]. 现代教育技术,29(2):39-45.
- Ahmed, I., & Sadeq, M. J. (2006). An autonomous mobile agent-based distributed learning architecture: a proposal and analytical analysis. *British journal of educational technology*, 37(4), 605-616.
- Barnard-Brak, L., Paton, V. O., & Lan, W. Y. (2010). Profiles in self-regulated learning in the online learning environment. *The International Review of Research in Open and Distributed Learning*, 11(1), 61-80.
- Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd international conference on machine learning* (pp. 113-120). ACM.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine learning research*, 3(1), 993-1022.
- Boitshwarelo, B. (2011). Proposing an integrated research framework for connectivism: Utilising theoretical synergies. *The International Review of Research in Open and Distributed Learning*, 12(3), 161-179.
- Dickey, M. D. (2005). Three-dimensional virtual worlds and distance learning: two case studies of Active Worlds as a medium for distance education. *British journal of educational technology*, 36(3), 439-451.
- Edirisingha, P., Nie, M., Pluciennik, M., & Young, R. (2009). Socialisation for learning at a distance in a 3D multiuser virtual environment. *British Journal of Educational Technology*, 40(3), 458-479.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5228-5235.
- Hoskins, S. L., & Van Hooff, J. C. (2005). Motivation and ability: which students use online learning and what influence does it have on their achievement?. *British journal of educational technology*, 36(2), 177-192.
- Huang, H. M. (2002). Toward constructivism for adult learners in online learning environments. *British journal of educational technology*, 33(1), 27-37.

(下转第44页)

- Frameworks. *International Working Group on Educational Data Mining*, (1): 210-219.
- Fu, S. W., Zhao, J., Cui, W. W., & Qu, H. M. (2017). Visual Analysis of MOOC Forums with iForum. *IEEE Transactions on Visualization and Computer Graphics*, 23(1): 201-210.
- Goh, C. H. (1997). *Representing and reasoning about semantic conflicts in heterogeneous information systems*. Cambridge USA: Massachusetts Institute of Technology.
- García-Saiz, Palazuelos, C., & Zorrilla, M. (2014). *Data Mining and Social Network Analysis in the Educational Field: An Application for Non-Expert Users*. Berlin Germany: Springer International Publishing.
- Lawrence, R. (2001). *Automatic Conflict Resolution to Integrate Schema*. Canada: University of Manitoba.
- Lin, F., Wang, L., Liu, S. L., & Liu, G. C. (2017). *Classification of Discussion Threads in MOOC Forums Based on Deep Learning*. In *Proceedings of 2017 2nd International Conference on Wireless Communication and Network Engineering: DEStech Transactions on Computer Science and Engineering* (pp. 506-511). USA: DEStech Publications.
- Min, S. D., & Zhu, B. J. (2014). Collection and Analysis of Emotional Data in Bulletin Board System Forum of University. *Applied Mechanics and Materials*, 513-517:2099-2102.
- Mokbel, M. F., Chow, C. Y., & Aref, W. G. (2015). The New Casper: A Privacy-Aware Location-Based Database Server. In *IEEE International Conference on Data Engineering* (pp. 1499-1500). Istanbul: Institute of Electrical and Electronics Engineers.
- Permana, F. C., Rosmansyah, Y., & Abdullah, A. S. (2017). Naive Bayes as opinion classifier to evaluate students satisfaction based on student sentiment in Twitter Social Media. *Journal of Physics Conference Series*, 893(1): 012-051.
- Rai, L., & Deng, C. R. (2016). Influencing factors of success and failure in MOOC and general analysis of learner behavior. *International Journal of Information and Education Technology*, 6(4): 262-268.
- Wei, X. C., Lin, H. F., Yang, L., & Yu, H. Y. (2017). A convolution-LSTM-based deep neural network for cross-domain MOOC forum post classification. *Information*, 8(3): 92.
- Wen, M. M., Yang, D. Y., & Rose, C. P. (2014). *Sentiment Analysis in MOOC Discussion Forums: What does it tell us*. In John, C. S., Zachary, A. P., Manolis, M., & Bruce, M. M. (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 130-137). UK: International Educational Data Mining Society.
- Wang, L., Hu, G. L., & Zhou, T. H. (2018). Semantic analysis of learners' emotional tendencies on online MOOC education. *Sustainability*, 10(6): 1-19.
- Yang, D., Piergallini, M., Howley, I., & Rosé, C. P. (2014). *Forum thread recommendation for massive open online courses*. In John, C. S., Zachary, A. P., Manolis, M., & Bruce, M. M. (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 257-260). UK: International Educational Data Mining Society.
- Yu, X., Yu, H., Tian, X. Y., Yu, G., Li, X. M., Zhang, X., et al. (2017). Recognition of college students from Weibo with deep neural networks. *International Journal of Machine Learning and Cybernetics*, 8(5): 1447-1455.

收稿日期: 2020-02-05

定稿日期: 2020-09-21

作者简介: 李彤彤, 博士, 副教授, 硕士生导师; 李坦, 硕士研究生; 郭翔宁, 硕士研究生。天津师范大学教育学部教育技术系(300387)。

责任编辑 韩世梅

(上接第35页)

- Kehrwald, B. (2008). Understanding social presence in text-based online learning environments. *Distance Education*, 29(1), 89-106.
- O'Neill, S., Scott, M., & Conboy, K. (2011). A Delphi study on collaborative learning in distance education: The faculty perspective. *British Journal of Educational Technology*, 42(6), 939-949.
- Wang, X., & McCallum, A. (2006). Topics over time: a non-Markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 424-433). ACM.
- 定稿日期: 2020-11-25
- 作者简介: 牛晓杰, 硕士研究生, 北京师范大学教育学部(100875)。
- 郑勤华, 博士, 教授, 博导, 通讯作者, 北京师范大学远程教育研究中心(100875)。

收稿日期: 2020-07-26

责任编辑 郝丹