

面向在线教育领域的大数据研究及应用

陈 池 王宇鹏 李 超 张 勇 邢春晓

(清华大学计算机系 北京 100084)

(清华大学信息技术研究院 北京 100084)

(清华信息科学与技术国家实验室 北京 100084)

(chenchi14@mails.tsinghua.edu.cn)

The Research and Application of Big Data in the Field of Online Education

Chen Chi, Wang Yupeng, Li Chao, Zhang Yong, and Xing Chunxiao

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

(Research Institute of Information Technology, Tsinghua University, Beijing 100084)

(Tsinghua National Laboratory for Information Science and Technology, Beijing 100084)

Abstract Online education played an important role in the field of education and is becoming more and more attracted people's attention with the development of the Information technology and video technology. Because of the growth of the amount of data on online education platform, it is very important to research the big data of online education. In this paper, we give the classification and characteristics of the big data in the field of online education. Then talk about Education Data Mining, Learning Analytics and Knowledge Graph which are used in our design. Next we describe the overall framework of big data in the field of online education, and emphasize the applications which can be realized according to the different service objects. Our goal with this paper is to outline the view of the application and to contribute to the development of the online education system with big data.

Key words online education; big data; application; model; Intelligent

摘 要 随着互联网和视频通信技术的发展,在线教育如火如荼地展开,在线教育平台产生的数据日益增多.面向在线教育领域的大数据研究及应用对于在线教育的高效性至关重要.首先明确了在线教育平台上大数据的基本概念、分类及特点,其次介绍教育数据挖掘、学习分析和知识图谱等有关技术,然后构建面向在线教育领域的大数据模型,最后按照服务对象的分类重点描述有关大数据的功能,勾勒出未来在线教育平台上大数据应用的基本轮廓,为之后在线教育领域大数据的研发起到指导性作用.

关键词 在线教育;大数据;应用;模型;智能化

中图法分类号 TP319

在线教育(e-Learning)是一种通过信息通信技术(ICTs)来增长知识和技能的方式^[1].随着互联网和 Web 技术的迅猛发展,在线教育受到越来越多人的关注.在线教育的方式方法随着计算机技术和人们教育观念的改变而产生了重大的变革^[2].从 20 世

纪末的远程教育模式到资源平台模式,再到目前流行的商业在线教育平台模式^[3].从开始静态的共享资源逐步走向互动性的开放课堂,在线教育的理念和模式正在发生着重大的改变.

随着在线教育浪潮的掀起,在线学习平台纷纷

收稿日期:2014-09-23

基金项目:国家“九七三”重点基础研究发展计划基金项目(2011CB302302);新闻出版重大科技工程项目数字版权保护技术研发工程多硬件环境版权保护技术研发(GXTC-CZ-1015004/02);国家科技支撑计划基金项目(2012AA09A408);清华大学自主科研计划(20131089235)

闯入人们的视线,其中具有代表性的平台之一就是大规模在线开放课程(massive open online course, MOOC).它是一种面向大众的开放式大规模网络课程.该平台借用发达的网络和视频技术缓解了教育资源缺乏、教育资源分配不均衡的问题;相比于MOOC, SPOC(small private online course)平台主要面向几十人到几百人的课堂,对于注册课程的学生也有严格的限制,只有满足对应课程要求的申请者才有资格学习该课程. SPOC这种小型课堂教学模式让教师成为真正的课程掌控者^[4];除了MOOC, SPOC之外,还有在线学习社区、移动端数字图书馆、移动阅读等等.这些在线教育平台为学生们营造了一个资源齐全、积极互动的氛围,强化了学生课前自主预习与课后巩固练习的环节.

国外知名MOOC平台——Coursera,截止2013年8月注册人数已经超过450万,推出近4个月全球范围内就有196个国家的学生参与其中.截止2012年,面向K12教育的可汗学院参与人数达到1000万,视频每天的访问量在10万到20万次,视频累计观看次数为2亿1千万次.由清华大学打造的中国的MOOC平台——学堂在线,截止2013年10月选课总人数超过了13000人次,单一门《电路原理》课程就吸引了来自150多个国家的11000余名学生.

面对在线教育平台上如此庞大繁杂的数据,面向在线教育领域的大数据研究及应用显得尤为重要.“大数据”是需要经过一定的处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产^[5].本文提到的“大数据”并不仅仅指海量数据,更多的是一种“数据为大”的思维方式,也就是说,使用客观的用户数据来指导决断,进而减少个人喜好等因素的干扰.它的关键之处就在于从数据到价值的转换过程.本文研究的目的在于利用在线教育领域中的数据来辅助和指导学生的学习、教师的教授以及人才的选拔.

本研究旨在使学生拥有个性化、多元化的学习路径,实现系统、文化、资源的全球整合^[6].利用在线教育平台上的大数据可以:1)跟踪学生的行为档案,实时记录学生对知识点的掌握情况以及在其上花费的时间.根据这些信息可以更加真实的了解学生的学习状况从而进行个性化辅导;2)辅助教师优化教学方法,改进教程设计从而制定更加合理的教学大纲;3)探究学习规律,并预测学生未来的发展,为选拔发掘不同领域的人才提供更加便捷和可靠的途径.

1 在线教育领域中的数据

在线教育领域中的数据复杂多样,本节从数据的分类和特点两大方面来介绍在线教育系统中的大数据.

1.1 数据的分类

在在线教育平台上学生可以随时随地观看视频、学习课程内容,教师可以在任何时间、任何地点创建课程、上传通告等.这里按照数据的获取方式、数据的来源和数据的存储形式这3种方式将在线教育平台产生的数据进行分类.

1.1.1 按获取方式分类

按数据的获取方式可将其分为显式数据和隐式数据.

显式数据包括用户注册信息、兴趣标签、已选课程、课程练习成绩、课程总成绩、对课程的评分、对讨论内容的评价等.

隐式数据包括搜索的关键词、浏览内容及次数、课程学习实践、练习时间、翻看讨论内容时间、下载文件、拖动滚动条次数、视频暂停次数、浏览器收藏记录等.

1.1.2 按来源分类

按数据的来源分为用户信息、课程信息、习题信息、视频信息、论坛信息、用户行为信息以及知识图谱.

用户信息是指用户的个人信息,包括用户的基本信息、学习课程的成绩信息和对知识点的理解程度等;课程信息指课程的内容和组织结构等;习题信息指习题内容和相关的习题提示信息等;视频信息包括视频、音频和字幕等;论坛信息包括论坛的组织结构和帖子内容等;用户行为信息指用户浏览网页、播放视频等的行为记录;知识图谱^[7]是知识点的集合,初期由教师绘制,后期采用自动结合人工的方式对其进行维护和更新.

1.1.3 按存储形式分类

按数据的存储形式分为结构化数据和非结构化数据.

结构化数据是指可以用二维表结构来逻辑表达实现的数据,在在线教育平台上包括用户基本信息、课程信息等;非结构化数据就是指那些不方便用数据库二维逻辑表来表现的数据,其中包括视频、音频和知识图谱等.

1.2 数据的特点

处在数据爆炸性增长的时代,“大数据”无处不

在,渗透了各行各业。“大数据”普遍具有4V的特征,包括大量(Volume)、高速(Velocity)、多变(Variety)和真实(Veracity)^[8]。在线教育领域的数据也具有不同于其他领域的特点。

第一,数据间联系紧密。在线教育系统中的数据来源各不相同,但其间却有千丝万缕的联系,如图4展示了底层数据库间的关联关系,比如学生将课程、论坛、成绩等结合在一起,知识点把知识图谱、习题库结合到一起,每个学生和每个知识点又有联系。可见在线教育领域的数据联系紧密,关系复杂。

第二,数据存储异构。视频数据采用非结构化形式存储,论坛、课程的结构采用树形模式存储,用户的基本信息采用简单的二维表结构化形式存储。故而在在线教育系统中存储形式具有多样性,从而形成了异构数据库,对数据的管理提出了更高的要求。

2 相关工作

大数据处理技术的难点主要集中于数据的庞杂以至于没有现成的软件工具能够对其直接开发,而这海量数据中又蕴含着能够帮助决策非常有意义的价值信息。目前,大数据的研究已经渗透到各个行业,其中包括互联网、政治^[9]、医疗^[10]等。这里主要研究在线教育领域的大数据。

在线教育上大数据的研究主要分为教育数据挖掘和学习分析技术两大方向^[11-12]。

2.1 教育数据挖掘

数据挖掘(DM)是指从大量数据中提取或“挖掘”知识的方法和应用。数据挖掘和知识发现(KDD)同义,其一般需要经过以下处理过程:数据清理、数据

集成、数据选择、数据变换、数据挖掘、模式评估和知识表示。教育数据挖掘^[13-14](EDM)就是数据挖掘在教育领域的应用,指从人类教育及学习过程所产生的数据中自动提取出有价值信息的技巧、工具和研究,这些信息可以为教育者、学习者、管理者、教育软件开发者和教育研究者等所利用。在线教育中数据具有存储异构的特点,故而数据清理和集成的过程较为复杂。同时,由于数据来源多样而且联系紧密,在挖掘的过程中需要考虑多个特征和其间的关联关系。

2.2 学习分析技术

学习分析技术^[15](LA)在早期被定义为“使用智能化的数据,学生产生的数据以及分析模型以发现信息和社交关系,从而达到预测和提高人们学习的目的”^[16]。2012年,Greller和Drachsler提出了一个更加系统全面的定义^[17],从六大关键维度来定义学习分析学,如图1所示。这六大关键维度分别是内部限制(internal limitations)、外部限制(external constraints)、工具(instruments)、数据(data)、目标(objectives)和利益者(stakeholders)。利益者包括机构(institution)、教师(teachers)、学生(learners)等,其中学习分析技术所分析的数据来自于学生所产生的数据,主要面向教师、机构;主要目标是反思教育模式(reflection)和预测学生发展前景(prediction);学习分析技术所使用的工具有技术(technology)、算法(algorithm)、理论(theories)等;其所包含的数据分为公开(open)数据和隐私(protected)数据;内部限制包括学生的能力(competences)和对相关知识的接受度(acceptance);外部限制包括平台的规范(conventions)及与学生教师的约定(norms)。应用学习分析技术过程中信息在机构(institution)、

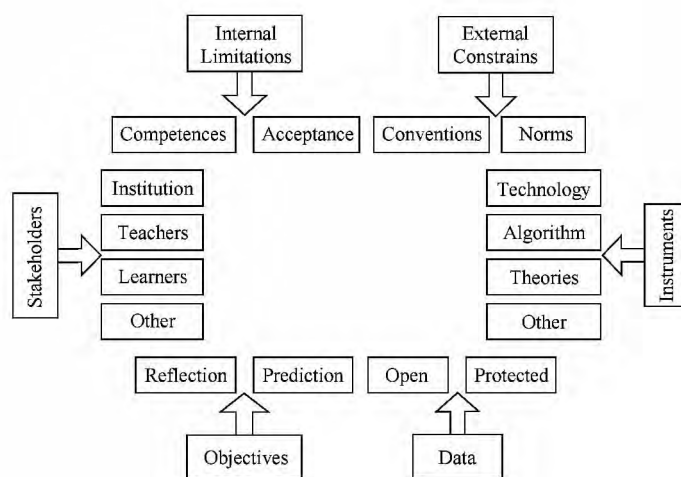


图1 学习分析技术的六大维度

教师(teachers)和学生(students)之间流动,如图2所示.运用该技术可以针对学生、教师和机构制定个性化的方案来实现自我反馈(self-reflection),同时,也可以对其上的数据进行统计、分析和研究(research),从而有利于政府(government/policy)作出较为明智的决策.

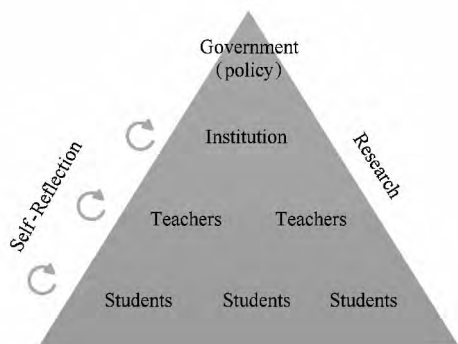


图2 LA 关益者信息流图

教育数据挖掘和学习分析技术的共同目标都是构建一个智能的学习平台,有利于提高教师的教学水平,激发学生的学习兴趣,促进学习的均衡发展.除了共性,两者之间也有着以下3点区别^[12]:

1) 教育数据挖掘主要强调具体使用的工具、技巧和算法;学习分析主要着眼于整个系统以支持不同对象的决策.

2) 教育数据挖掘侧重于细节和算法;学习分析技术侧重于模型的建立.

3) 教育数据挖掘的目的在于搭建一个智能即能够自动应答的系统;学习分析技术更多的是立足于教育领域以满足各类对象的需求.

2.3 知识图谱

1965年,普莱斯发表的《科学论文的网络》一文研究了科学论文之间的引证和被引证关系并形成了引证网络.该引文分析的技术就是知识图谱的雏形^[18].2012年,Google公司为了改进语义搜索而系统地提出了一种知识库——知识图谱^[19].其目标是减轻用户收集、总结和归纳的负担,通过知识图谱快速、便捷地掌握查询主体的脉络.

知识图谱是把应用数学、图形学、信息可视化技术、信息科学等学科的理论与方法与计量学引文分析、共现分析等方法结合,用可视化的图谱形象地展示学科的核心结构、发展历史、前沿领域以及整体知识架构的多学科融合的一种研究方法^[18].在知识图谱中查询某一个关键词的结果是一系列与之相关的内容,并不仅仅是该关键词的内容.知识图谱不仅能够帮助学生建立较为完整的知识体系,明确学习目标,提高学习效率;而且助于教师迅速发现学习弱点,进行有针对性的答疑,从而提升教学质量.

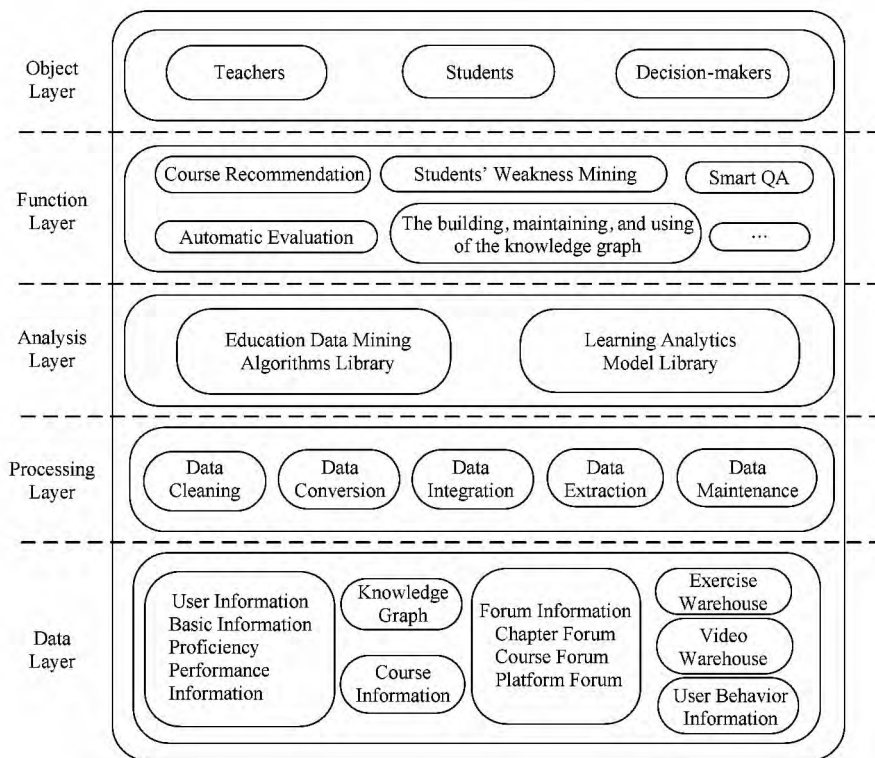


图3 面向在线教育领域的大数据模型

3.5 对象层

对象层主要包括学生(students)、教师(teachers)和决策者(decision-makers)、学生学习、教师授课。这里的决策者分为系统管理者和机构决策者,系统管理者主要负责系统的运营管理;机构决策者针对学生的学习情况和教师的授课情况对学习过程和人才选拔提出战略性的建议。在线教育系统一般提供LMS^[23](学习管理系统)和CMS(课程内容管理系统)两大子系统。LMS是面向学生的学习系统,为学生提供观看视频、学习教材、讨论、提问等功能;CMS面向教师,是教师制作课程、发布通告、制定教学大纲的平台。

4 大数据支撑的在线教育系统应用

图3展示了基于在线教育平台的大数据研究模型,主要强调了数据库层的内容和结构,本节站在学生、教师和决策者的角度来详细描述具体已经实现和可能实现的功能。

4.1 面向学生的智能化功能

在线教育的主要服务对象就是学生,这里从知识图谱、课程、视频、习题、论坛等角度来具体介绍面向学生的大数据功能。

4.1.1 基于知识图谱的学习

学生学习和教师教学都是从知识点出发考虑的,知识图谱就是一个可以将所有知识点总结汇总到一起的知识架构图。在知识图谱中可以找到每门课程包含的所有知识点以及各个知识点之间的联系。知识图谱将课程知识模块或学科内各课程以架构图形式呈现,点击跳转至相应课程或知识模块,方便学生了解知识脉络。

4.1.2 个性化课程服务

在线教育为广大学生提供了一个自主学习平台,在该平台上,学生可以注册自己感兴趣的课程,有选择性地观看视频。随着平台上的课程逐渐增多,查找合适课程所花费的时间越来越长,有时还会面临不知如何选择的困扰,因此提出面向学生的课程推荐系统这一构想。该推荐系统根据学生自己和其他学生的学习情况以及课程涉及到的知识点等情况向学生推荐适合的课程。相关课程推荐(course recommendation)所考虑的问题主要包括以下几个方面:

1) 课程间关联性。由领域专家教师对课程进行相似度划分,建立知识架构体系,推荐相应基础课程

或进阶课程。

2) 学生学习弱点/优势。根据该学生学习过程中习题、测验和实验等的完成情况,总结学生学习弱点,推荐相关课程。

3) 相似学习者。推荐所学课程相似、水平相当的学习者选修的其他课程,既弥补专家推荐存在的学科领域局限性,又可顺应学习研究的热点趋势。

4.1.3 结合知识点的字幕定位

学生在在线教育平台上学习的主要形式之一就是观看视频。为使能够精准定位到感兴趣的片段,需要提供基于字幕的视频跳转定位功能;此外,还可将字幕上的专业词汇链接到知识图谱上的知识点,点击相应词汇显示该词汇注释或注释链接,有利于学生更全面的学习。

4.1.4 智能化习题训练

习题训练是强化巩固知识的必要阶段,借助知识图谱和习题库可以提供以下功能:

1) 智能生成习题。系统能根据用户答题正误的情况,从与知识图谱关联的习题库中随机提取习题,供反复练习直至掌握。

2) 习题巩固。遵循记忆规律,间隔一定周期出题巩固(如按答题时间分为Practice→Level One→Level Two→Mastered四个阶段^[24])。对于之前错误率较高的知识点,会增强练习强度,使得练习更有针对性,提高复习效率。

3) 习题配套提示。学生可以选择查看提示信息,并且提示信息采用逐条递进方式,即可以一直点击下一步,直到学生有解题思路为止。同时,还配有视频形式的讲解供有需要的学生自主参考。

4) 习题纠错。学生可以质疑习题的答案,提出自己的观点。该纠错信息将发送给备课老师,提醒教师详细说明该问题,供之后进行聚类分析。

4.1.5 基于论坛数据的智能化服务

除了习题练习,学生和学生之间与学生和老师之间的讨论也是能够大幅度提升学生成绩的板块之一。论坛以讨论主题、帖子、帖子评论形成树形结构。论坛中可挖掘的信息有很多,这里从以下几个方面进行讨论。

1) 论坛挖掘。可以对论坛进行结构挖掘、帖子自动打分、主题挖掘、帖子推荐。论坛结构挖掘主要指分析论坛讨论模式,包括围绕一个话题讨论、进阶式讨论等;帖子自动打分可以根据帖子长度、内容以及该帖所获顶或踩的数目等对帖子进行自动打分;主题挖掘是对主要讨论主题进行抽取,并挖掘出较

好的回帖.论坛推荐可以根据对用户点击和发表帖子的挖掘为用户推荐感兴趣的板块或讨论主题.论坛挖掘可以自动从海量帖子中发掘出有意义、有针对性的精华部分,减少学生查找答案的时间,提高了学习效率,同时,为教师实时掌握学生动态提供新的途径.

2) 学生分级^[25].根据学生在论坛和课堂的表现对学生进行分类和分级,以激励学生发表有价值的帖子和有趣的观点,进而促进交流,提升学习效率.

3) 用户建模及信息推荐.包括个人标签生成和基于讨论内容的信息推荐.个人标签指根据讨论区语义分析,发现学生兴趣爱好和知识掌握的薄弱点.基于讨论内容的信息推荐是根据学习者反复观看的视频、答错的题目、搜索词条等推送相关讨论信息,实现实时答疑和学习引导的效果.

除了以上模块化的功能,还有一些简洁、智能、方便的功能,如资源推荐、学习分组推荐等.资源推荐指根据课程论坛的讨论内容,并结合网络上的相关搜索结果推荐课程相关的外部资源.学习分组是根据学生的学习模式、学习内容和学习行为进行学习小组成员的推荐.

4.2 面向教师的智能化功能

相比于传统课堂中,教师根据对课程的理解和授课经验备课;在线教育平台上教师还可以根据学生学习情况的反馈进行有针对性的备课.通过统计、分析和挖掘学生观看视频、完成习题、课程讨论等行为数据,可从中发现学生学习难点、教师教授弱点等.此外,基于庞大数目学生提出的大量问题,不仅可以对众多问题进行聚类分析,划分为几类具有代表性的问题以方便教师回答,而且还可以借助论坛中的内容实现自动推送优秀回答等功能.

为了提高教师的教学水平,也可以根据教师的备课情况,参与论坛情况和学生评教的结果等,建立教学效果/教师评估模型,对老师进行综合自动评教.同时,为了使教师之间形成更加良好的学术氛围,可以根据教师的备课行为和所授课程内容为教师推荐好友.

4.3 面向决策者的智能化功能

在线教育平台的大数据可以为决策者提供以下功能:一是面向整个平台的数据可视化,二是面向系统优化的统计分析.数据的可视化使用统计图表的形式展现学生的地域分布、性别分布、成绩分布、作为完成情况分布等等,为机构决策者提供决策数据;系统优化指可以使用学生的浏览数据对网站结构进

行优化^[26-27],一个页面化为一个点,访问 A 页之后访问 B 表示 AB 相连,边权表示连续访问的次数,点权表示页面访问的次数.系统管理者可以借用这种直观的表现方法改善网站的内容结构,从而吸引更多的用户.

5 结 论

本文从数据的分类及特点的角度阐述了在线教育平台上大数据的概念,并简单介绍教育数据挖掘、学习分析和知识图谱等有关在线教育上的大数据技术,重点设计面向在线教育领域的大数据模型,同时,按照服务对象的分类详细描述大数据的有关功能,为今后教育领域大数据研究打下基础.目前的在线教育平台实现了观看课程、录制课程等基本功能,对于大数据方面的探究正处于起步阶段,本文为在线教育领域大数据的研究提供了思路.对于功能、模型库和算法库的详细实现还有待进一步的探究.

面向在线教育领域大数据是一个既有学术意义又有应用价值的研究方向,随着大数据技术的发展,人们对在线教育理解的加深,会有越来越多的组织及个人参与到在线教育相关的研究中.相信通过大家的努力,一定可以实现更方便、更高效、更优质获取掌握知识,提高自身能力的梦想.

参 考 文 献

- [1] McGreal R, Anderson T. E-learning in Canada. *Inter Journal of Distance Education Technologies (IJDET)*, 2007, 5(1): 1-6
- [2] Aparicio M, Bacao F. E-learning concept trends // *Proc of the 2013 Inter Conf on Information Systems and Design of Communication*. New York: ACM, 2013: 81-86
- [3] 王嘉华,王靖.大规模在线课程(MOOC)全球趋势及对中国的意义.埃森哲,2013
- [4] 康叶钦.在线教育的“后 MOOC 时代”——SPOC 解析. *清华大学教育研究*, 2014, 35(1): 85-93
- [5] Kaisler S, Armour F, Espinosa A. Introduction to big data: Scalable representation and analytics for data science minitrack // *Proc of the 46th Hawaii Inter Conf on System Sciences (HICSS)*. Piscataway, NJ: IEEE, 2013: 984-984
- [6] 祝智庭,贺斌.智慧教育:教育信息化的新境界. *电化教育研究*, 2012 (12): 5-13
- [7] Pechsiri C, Piriyakul R. Explanation knowledge graph construction through causality extraction from texts. *Journal of Computer Science and Technology*, 2010, 25(5): 1055-1070

- [8] Shi Y, Xie W, Xu G. Smart remote classroom: Creating a revolutionary real-time interactive distance learning system // *Advances in Web-Based Learning*. Berlin: Springer, 2002: 130-141
- [9] 刘琼. 大数据时代的美国经验与启示. 人民论坛, 2013 (10): 30-31
- [10] 许德泉, 杨慧清. 大数据在医疗个性化服务中的应用. 中国卫生信息管理杂志, 2013, 10(4): 301-304
- [11] Bienkowski M, Feng M, Means B. Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. US Department of Education, Office of Educational Technology, 2012: 1-57
- [12] 陆璟. 大数据及其在教育中的应用. 上海教育科研, 2013, 2013(9): 5-8
- [13] Baker R S J D, Yacef K. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 2009, 1(1): 3-17
- [14] 李婷, 傅钢善. 国内外教育数据挖掘研究现状及趋势分析. 现代教育技术, 2010, 20(10): 21-25
- [15] 李青, 王涛. 学习分析技术研究与应用现状述评. 中国电化教育, 2012, 2012(8): 129-133
- [16] Siemens G. What are learning analytics. Retrieved March, 2010, 10: 2011
- [17] Greller W, Drachsler H. Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 2012, 15(3): 42-57
- [18] 杨思洛, 韩瑞珍. 知识图谱研究现状及趋势的可视化分析. 情报资料工作, 2012 (4): 22-28
- [19] Katrine Juel Vang. Ethics of Google's knowledge graph: Some considerations. *JICES*, 2013, 11(4): 245-260
- [20] 张存刚, 李明, 陆德梅. 社会网络分析——一种重要的社会学研究方法. 甘肃社会科学, 2004 (2): 109-111
- [21] 吴畏. 基于话语分析方法的案例研究——以实习教师网络反思为例. 现代教育技术, 2010, 20(2): 33-36
- [22] 王子平, 张延飞. 内容分析法及其在现代远程教育中的应用. 继续教育研究, 2009 (5): 55-58
- [23] Ellis R K. Field guide to learning management systems. Alexandria, VA: ASTD Learning Circuits, 2009
- [24] Bloom B S. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 1984: 4-16
- [25] Anderson A, Huttenlocher D, Kleinberg J, et al. Engaging with massive online courses // *Proc of the 23rd Inter Conf on World Wide Web*. New York: ACM, 2014: 687-698
- [26] 程舒通, 徐从富. 网站结构优化技术研究进展. 计算机应用研究, 2009, 26(6): 2013-2015
- [27] 杜华, 王锁柱. 网站结构优化模型及算法分析. 计算机工程与设计, 2008, 29(21): 5591-5594

陈池女, 1992年生, 博士研究生, CCF会员, 主要研究方向为数据库与知识工程.

王宇鹏男, 1985年生, 硕士研究生, 主要研究方向为数据库与知识工程(babulba@163.com).

李超女, 1978年生, 博士, 副教授, CCF高级会员, 主要研究为存储系统、数据管理及其在数字图书馆/档案馆和在线教育领域的应用(li-chao@mail. tsinghua. edu. cn).

张勇男, 1973年生, 博士, 副研究员, CCF高级会员, 主要研究方向为数据库技术(zhangyong05@tsinghua. edu. cn).

邢春晓男, 1967年生, 博士, 教授, 博士生导师, CCF高级会员, 主要研究方向为数字图书馆和数据库技术(xingcx@tsinghua. edu. cn).