

# 大型开放式在线课程(MOOC)数据可视分析综述

周志光<sup>1)</sup>, 王飒清<sup>1)</sup>, 刘玉华<sup>1)</sup>, 王勇<sup>2)</sup>, 王毅刚<sup>1)\*</sup>

<sup>1)</sup> (杭州电子科技大学人文艺术与数字媒体学院 杭州 310018)

<sup>2)</sup> (School of Computing and Information Systems, Singapore Management University Singapore 188065)  
(yigang.wang@hdu.edu.cn)

**摘要:** 大型开放式在线课程(massive open online courses, MOOC)是面向用户提供教学资源的线上学习平台. 从 MOOC 数据类型、MOOC 数据可视化方法和 MOOC 数据的可视分析应用 3 个角度进行综述. 首先介绍 MOOC 中的文本数据、视频数据、拓扑结构数据和时空数据; 然后对不同数据类型常用的可视化方法进行论述; 再介绍 MOOC 数据可视分析在在线学习行为分析、论坛交互活动分析和学生表现结果分析等方面的应用; 最后对 MOOC 数据可视分析的未来发展趋势进行展望.

**关键词:** 大型开放式在线课程; 可视分析; 可视化; 在线学习行为  
**中图法分类号:** TP391.41 **DOI:** 10.3724/SP.J.1089.2022.19536

## A Survey on Visual Analysis of Massive Open Online Courses (MOOC) Data

Zhou Zhiguang<sup>1)</sup>, Wang Saqing<sup>1)</sup>, Liu Yuhua<sup>1)</sup>, Wang Yong<sup>2)</sup>, and Wang Yigang<sup>1)\*</sup>

<sup>1)</sup> (School of Media and Design, Hangzhou Dianzi University, Hangzhou 310018)

<sup>2)</sup> (School of Computing and Information Systems, Singapore Management University, Singapore 188065)

**Abstract:** Massive open online courses (MOOC) are free online courses that provide educational resources to users. In this paper, related works are summarized from the following perspectives: data types of MOOC, MOOC data visualization, and visual analysis applications of MOOC data. First, MOOC data is categorized into four types, including text data, video data, topological data, spatio-temporal data, and their concepts and applications are introduced. Then, the common visualization methods of different data types are discussed. Furthermore, the major applications of visual analytics approaches in MOOC data analysis are summarized, including online course learning, online tests and exams, online forum interaction and so on. Finally, the challenges and future work on visual analytics of MOOC data are discussed and concluded.

**Key words:** massive open online courses; visual analysis; visualization; online learning behavior

随着互联网技术的快速发展, 以大型开放式在线课程(massive open online courses, MOOC)为核心的教育平台(如 Coursera, edX 等)吸引了数亿计的学习者在线上学习和交流. 在 MOOC 教学

过程中, 平台收集了丰富的学习者网络日志数据, 其中不仅包括文本、视频等媒体类型数据, 还包括拓扑结构数据和时空数据等. 这些数据全面记录了学生在课程学习过程中的行为表现, 为课程教

收稿日期: 2021-12-15; 修回日期: 2022-03-23. 基金项目: 国家自然科学基金(61802339, 61872314, 62177040); 浙江省科技计划(2021C03137); 浙江省科技厅公益项目(GF20G010005, GF20F020065); 浙江大学 CAD&CG 国家重点实验室开放课题(A2001, A2224). 周志光(1983—), 男, 博士, 教授, 硕士生导师, CCF 会员, 主要研究方向为数据可视化与可视分析; 王飒清(1998—), 女, 硕士研究生, CCF 学生会员, 主要研究方向为数据可视化; 刘玉华(1988—), 男, 博士, 讲师, 硕士生导师, CCF 会员, 主要研究方向为数据可视化与可视分析; 王勇(1989—), 男, 博士, 助理教授, 主要研究方向为信息可视化、视觉分析、人机交互和人工智能; 王毅刚(1971—), 男, 博士, 教授, 博士生导师, CCF 会员, 论文通信作者, 主要研究方向为虚拟现实与可视化.

师和教育研究人员评估教学方案并提升教学质量,以及学生提高学习效率等需求提供了素材和帮助。

可视化与可视分析技术以图形的方式直观地呈现隐含于数据中的特征模式,帮助用户从大量数据中推断出有用的信息。近年来,研究人员在 MOOC 数据可视化与可视分析方面展开了大量的工作,以支持用户通过各种视觉帮助和丰富的交互来分析数据。例如,在学习行为分析方面,它帮助普通用户更轻松地探索不同学生的行为模式,了解不同类型的学习群体<sup>[1]</sup>;在论坛交互方面,它能直观地展示学生的情感变化和交互频率,协助用户更深入地分析学生的情绪变化和社交动态<sup>[2]</sup>;在学生表现方面,它也为用户提供了许多关于学生表现趋势的有用见解<sup>[3]</sup>。

由此可知,可视分析技术在 MOOC 的各方面都发挥着重要作用。国内外已有一些相关的文献综述对其进行总结和讨论,如 Wu 等<sup>[4]</sup>从应用的角度出发,将 MOOC 论坛可视化分为预测表现、理解情绪和预测社会行为模式 3 个类别,并分别对每个类别中已有的可视化系统进行分析和总结;Qu 等<sup>[5]</sup>通过分析现有的几个典型的 MOOC 数据可视分析系统,进一步探讨和总结未来 MOOC 数据可视化面临的挑战以及可能的发展方向。

本文在前人的工作基础上,分别从数据类型、数据可视化方法、可视分析应用 3 个方面入手,对 MOOC 数据可视分析进行全面的调研和综述。首先,从数据类型出发,重点对文本数据、视频数据、拓扑结构数据、时空数据进行介绍;其次,对不同数据类型的可视化方法展开论述;接着,从应用的角度出发,对在线学习行为分析、论坛交互活动分析、学生表现结果分析的相关可视分析论文进行总结与归纳;最后,对 MOOC 教学数据可视分析领域的难点和发展趋势进行了总结和展望。

## 1 MOOC 数据类型

大量非结构化的多媒体数据如文本数据和视频数据,表达关系的拓扑结构数据如层次结构数据和网络关系数据,以及表达时间和空间信息的时空数据等,在 MOOC 教学过程中被记录下来,这些数据为教师和研究人员进行进一步分析和探索课堂学习、教学效果和学生反馈等提供了有力的支持。本节针对不同类型的 MOOC 数据展开介绍。

### 1.1 MOOC 文本数据

文本数据是 MOOC 中非结构化数据类型的典

型代表,也是人们日常工作与生活中接触最多的数据类型之一。课程的课件、论坛的互动和学生的测验中都包含着大量的文本信息。研究者通过分析网络日志文件或爬取 MOOC 网站等途径获取文本数据,采用文本挖掘技术发现其中有趣的模式,并帮助用户探索其背后的知识和情感等信息。

Huang 等<sup>[6]</sup>采用文本挖掘技术对课堂讲义和视频字幕中的文本数据进行关键词提取,并将其有序排列在词云中,帮助学习者更方便地复习课程知识。Hsu 等<sup>[7]</sup>基于自然语言处理技术,对 MOOC 论坛中与课程概念相关的讨论文本进行分析,帮助教师了解学生容易产生误解的一系列概念,从而有针对性地改进教学。此外,为了更好地理解文本数据中蕴含的情感信息,Hoque 等<sup>[8]</sup>采用语义取向计算器的方法,对论坛中的文本数据进行情感分析,帮助用户判断学生在学习过程中学习态度的积极和消极程度。

### 1.2 MOOC 视频数据

视频作为 MOOC 平台上传递知识的重要媒介,是学生学习体验的核心。MOOC 视频数据由一系列连续的帧和字幕组成,研究者通过下载或爬取视频的方式获取数据,并通过视频处理或字符识别等技术分析视频的内容。一项对 Coursera 上课程的研究<sup>[9]</sup>发现,许多学生跳过在线测验、在线讨论以及其他含有师生互动的课程内容,直接通过课程视频进行学习,由此可见视频在 MOOC 中的重要性,它在很大程度上决定着学生的学习状态和效果。

目前,大量关于 MOOC 视频数据的处理工作是针对视频内容的解读展开的。例如,TalkMiner 系统<sup>[10]</sup>基于光学字符识别和词法分析法提取并处理视频中的文本,以帮助用户快速地了解视频内容。MOOCex 系统<sup>[11]</sup>同时挖掘 MOOC 视频内容和课程大纲中的文本信息,提取每个视频的相关主题,帮助用户根据主题寻找感兴趣的视频。

### 1.3 MOOC 拓扑结构数据

MOOC 数据不仅拥有丰富的多维特征,同时蕴含着复杂的拓扑结构关系。拓扑结构数据分为层次结构数据和网络关系数据 2 种,节点和边是这类数据的重要组成部分。通过实体抽取和关系抽取等技术得到数据中隐含的拓扑结构,可以帮助用户了解和探索数据之间的关联关系。其中,层次结构数据一般存在于 MOOC 课程的流程与结构中。Pardos 等<sup>[12]</sup>提取课程的结构信息,通过构建课程树帮助用户了解课程知识中的章节脉络和

组织结构。

网络关系数据广泛存在于教师与学生、学生与学生之间的交互网络,以及课程与课程之间的先导关系网络中。其中,MOOC 论坛交互网络是 MOOC 数据分析的热点。Wong 等<sup>[13]</sup>提取论坛线程和讨论贴对学生论坛进行分析,了解学生与不同讨论组的互动情况。

#### 1.4 MOOC 时空数据

MOOC 中的时空数据有着明显的时间属性和空间属性,具有多源、异构和实时等特点。通过时空域分析技术,可以完成时空预测、时空模式挖掘和时空异常检测等诸多任务。其中,时间属性对应的时间序列数据在 MOOC 中分布广泛,如学生观看课程视频时产生的点击流数据、平台每天的注册人数以及论坛中每周的帖子数等。分析时序数据有利于用户了解事件的变化情况,并挖掘其中的发展规律。VisMOOC 系统<sup>[14]</sup>通过记录选定 MOOC 视频的观看情况,为用户分析视频的在线观看人数与重要时间节点(如考试测验周、期中期末考试)之间的关系提供可视工具。

一般通过调查学生登录时的 IP 地址以及注册时填写的位置信息,可获得空间属性对应的地理空间数据,如所在城市、国家或地区等。分析地理空间数据可以帮助用户了解不同国家和不同地区使用 MOOC 时的行为差异。如 PeakVizor 系统<sup>[15]</sup>通过分析点击流数据的空间属性与学习持久性之间的关系,帮助课程教师和教育专家了解不同地区学生的学习差异。

## 2 MOOC 数据可视化方法

可视化可以直观地呈现和表达数据,帮助用户快速掌握数据背后的含义。因此,研究人员针对不同类型的 MOOC 数据,提出了许多不同的可视化方法。本节对不同数据类型的可视化方法进行总结和回顾。

### 2.1 MOOC 文本数据可视化

MOOC 文本数据可视化为人们掌握文本中的知识和情感等信息提供了便利。与传统的文本可视化方法类似,MOOC 文本数据可视化工作可分为基于词频和基于语义 2 种。其中,基于词频的 MOOC 文本数据可视化方法有词云和柱状图等。词云和柱状图简单易懂,能直观地反映特定文本的内容,在 MOOC 文本数据分析中应用十分广泛。包昊罡等<sup>[16]</sup>通过词云将在线协作讨论中的高频关键词

按照不同的颜色和大小进行可视化。LDAvis<sup>[17]</sup>利用柱状图展示与论坛中某个特定话题最相关的一系列术语,浏览者可以通过查看每个主题中的相关术语(即关键词)探索这些主题的意义和流行程度。孙雯中<sup>[18]</sup>采用柱状图统计了论坛话题中的文本关键词的词频,从而帮助用户了解学习者讨论的热点话题。

研究人员也有从语义的角度出发对 MOOC 文本数据展开可视化工作。例如,Hoque 等<sup>[8]</sup>将常见的关键词提取技术与情感识别的文本分析结果进行组合,并与用户 ID 建立关联,直观地反映论坛对话的情绪状态,帮助使用者更好地实现论坛文本数据的多方面探索。

### 2.2 MOOC 视频数据可视化

在现代教育领域,已有许多与视频数据结合的可视化工作。EmotionCues 系统<sup>[19]</sup>通过情感识别模型自动检测课堂视频中学生的的情绪,帮助教师和家长快速地了解学生在课堂上的参与度。Li 等<sup>[20]</sup>提出的在线测验可视分析系统,通过多视图联动技术可视化学生考试视频中的姿势以及点击流行为,帮助监考人员快速地识别可能有作弊倾向的学生。如图 1 所示,该系统由问题列表视图、学生行为视图、视频回放视图等多个视图组成,监考人员可通过控制面板进行在线考试的选择和参数的调整。

MOOC 视频数据可视化方法注重以视频的片段或整个视频为基础,通过增加辅助图形增强用户对现有视频的理解。例如,Cooper 等<sup>[21]</sup>对用户选定的视频进行文本检测,并提取文本中包含的相关概念信息加以展示;Chatti 等<sup>[22]</sup>设计 Heatmap 组件增加视觉辅助,帮助学生快速地定位学习材料中的重要部分;Sharma 等<sup>[23]</sup>收集教师的眼动数据,通过可视化教师的凝视轨迹来突出幻灯片上的重要区域,进而帮助学生更好地掌握知识。

### 2.3 MOOC 层次结构数据可视化

MOOC 层次结构数据可视化可以帮助用户更好地理解数据之间的包含和从属关系,常见的可视化方法有树图和圆形嵌套图等。与传统的基于空间填充的树图<sup>[24]</sup>相比,MOOC 中的树图更注重表示节点和边之间的关系,H 状树、圆锥树和放射树均为 MOOC 树图常见的表现形式。Deng 等<sup>[25]</sup>采用可折叠的放射状树展示考试问题和作业问题之间的相关性,使教师可以根据该相关关系及时反馈和提高教学质量。当数据量增大时,树结构的枝干越发茂盛,会给用户的理解带来干扰。相比之

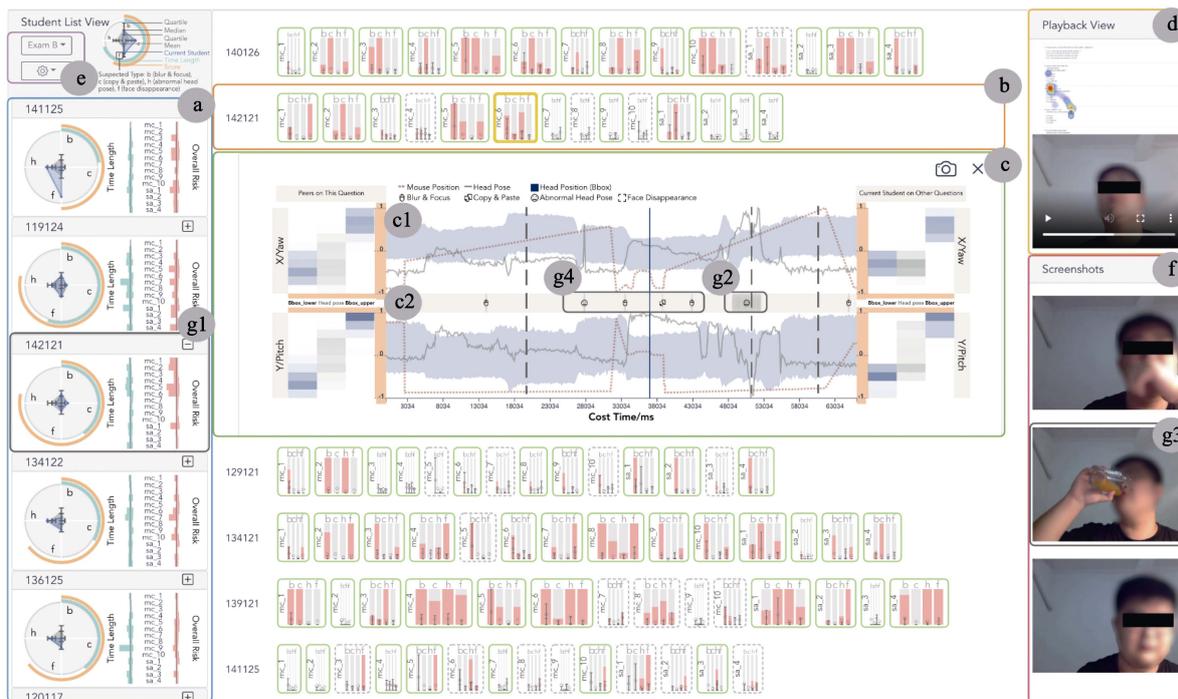


图 1 在线测验视频数据可视化<sup>[20]</sup>

下, 圆形嵌套图的方法虽然更占用空间, 但它固有的空白区域也使数据的层次感更容易理解. 例如, Silius 等<sup>[26]</sup>针对在线教学中学生小组的讨论内容, 采用圆形嵌套图直观地反映小组讨论过程中所涉及主题之间的关系, 并设置配置面板支持用户改变投票、视图和时间等变量的值, 以帮助用户根据自己的需求探索不同的可视化效果. 为了更好地理解 MOOC 背景下的在线论坛, Atapattu 等<sup>[27]</sup>也使

用圆形嵌套图对 MOOC 论坛帖子中的主题进行可视化, 以便用户对 MOOC 论坛中的讨论主题进行识别和分类.

此外, booc.io 系统<sup>[28]</sup>针对学习材料中的概念, 结合层次概念图和动态非线性学习路径, 将材料的线性结构表示为一个学习计划的形式, 如图 2 所示. 通过动态查询和添加学习材料等交互, 帮助学生探索个性化的学习路线.

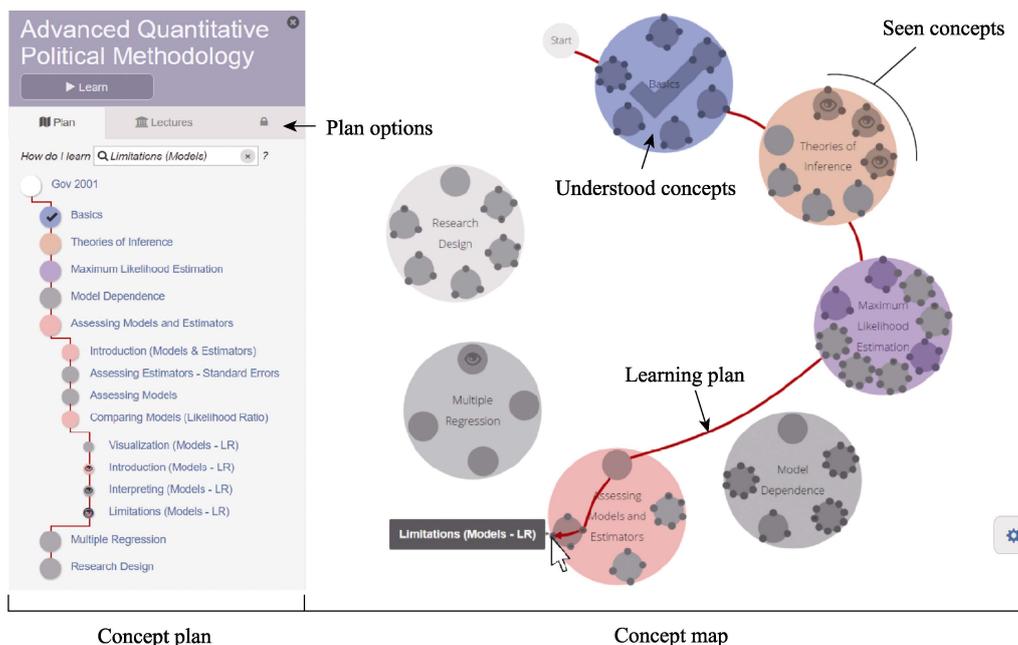


图 2 MOOC 层次概念结构图<sup>[28]</sup>

## 2.4 MOOC 网络关系数据可视化

可视化方法能够帮助用户全面而深入地探索网络中隐藏的复杂特征和关联模式,节点和边是网络数据可视化视图中最常见的元素.传统的节点链接图作为网络关系数据可视化最常用的布局方式之一,具有简洁和布局灵活等特点,在 MOOC 网络关系数据中也应用广泛. Saqr 等<sup>[29-30]</sup>采用带方向的节点链接图对老师和学生之间的交互进行分析,并用节点大小映射每个学生的活跃程度. Ho 等<sup>[31]</sup>采用节点链接图可视化哈佛大学课程和麻省理工学院课程的学生流入/流出流量情况,帮助教师了解不同课程的受欢迎程度,进而有针对性地改进课程安排.

力导向图有着良好的对称性和局部聚合性,常用于复杂网络可视化. VisForum<sup>[32]</sup>将一个线程中论坛用户之间的交互可视化,论坛成员以小圆的形式呈现,圆形半径和颜色分别映射该论坛成员的活跃度和影响力,链接宽度映射 2 个论坛成员之间的相互回复频率,帮助用户直观地了解节点之间的交互频率.

## 2.5 MOOC 时间序列数据可视化

时间序列可视化着重突出数据的时间属性,传统的时间序列数据可视化方法有日历图和三维图等,它们都有沿着水平时间轴排列事件的特点,在 MOOC 中使用广泛.例如, CourseVis 系统<sup>[33]</sup>采用三维图的方式按时间次序依次排列论坛上不同时期发布的讨论帖.该系统将论坛中的帖子用球体表示,结合坐标轴上的主题、参与者和发帖日期等信息确定球体的位置,帮助教师快速地掌握论坛帖子随时间的变化情况.

当数据集变大时,密集的可视化信息可能会增加用户的认知负担.因此,研究人员将大型数据集中的数据聚合分类,并采用数据流的方式模拟数据随时间的变化情况.例如, ViSeq 可视分析系统<sup>[34]</sup>针对学生的学习顺序,通过双向弦图表示学生学习序列的过渡情况.为了让用户一次看到更长的序列,系统将同一周内的事件(在视觉上以相同的颜色出现)压缩成一个颜色条,用户可以通过单击它展开任何段,从而清楚地感知学习者在不同周之间过渡的差异.

## 2.6 MOOC 地理空间数据可视化

地理空间可视化能够有效地融合数据挖掘技术和可视化设计,全面而细致地分析和探索地理空间数据中隐含的多维、时空、动态和关联等特征<sup>[35]</sup>.

地图作为地理空间数据常见的一种可视化方式,能很好地呈现数据的空间属性,与复杂的空间立体可视化方法<sup>[36]</sup>相比, MOOC 地理空间数据可视化方法注重点与区域的表达.其中,基于点状地图的可视化方法适用于表示数据在地理空间上的分布情况,而基于区域地图的可视化方法常用于区域之间数据的对比.

Emmons 等<sup>[37]</sup>借助点状地图的形式对 MOOC 学生的源地进行可视化的呈现.其采用圆圈的大小映射相关区域参与学生的数量,不同的圆圈颜色映射不同的年份,为用户了解不同年份使用 MOOC 的主流国家分布提供帮助. Dernoncourt 等<sup>[38]</sup>通过基于区域地图的方法,将每个国家的证书获得者与注册人数的比率呈现在地图上,颜色越深比率越高,为用户分析各个国家之间学生学习完成度的差异提供便利.

## 3 MOOC 数据的可视分析应用

对 MOOC 数据进行可视分析,不仅能直观地向用户展示学生的行为、互动和表现等信息,还能通过大量的数据识别其背后的规律,进而开发与设计一些实际性的应用.现有的 MOOC 数据可视分析工作根据其应用领域分为在线学习行为分析、论坛交互活动分析和学生表现结果分析 3 个方面,本节分别对它们进行总结和归纳.

### 3.1 在线学习行为分析

常见的在线学习行为可视分析工作是基于鼠标点击数据和学习序列数据展开的.例如, Wang 等<sup>[39]</sup>采用时间重映射的方式展示了点击流数据随时间出现、发展和消失的过程,帮助普通用户更轻松地了解学生的行为模式. Xia 等<sup>[40]</sup>提出了一个视觉分析系统来分析学生在 K-12 数学电子学习平台上的鼠标移动数据,帮助教师快速地发现学习材料设计中可能存在的缺陷,并对不同学生具体解决问题的方式有更深入的了解. Chen 等<sup>[34]</sup>针对学习序列数据提出了一个可视分析系统 ViSeq.如图 3 所示,系统使用序列模式挖掘方法处理不同粒度级别的数据,采用发现学习者群体的投影视图、挖掘学习者学习行为的模式视图和探索连续事件之间转换的序列视图等多个视图呈现每个学习者的学习顺序,并对相似个体进行比较,旨在帮助用户更好地理解学习序列和性能之间的相关性,进一步表征不同的学习群体.

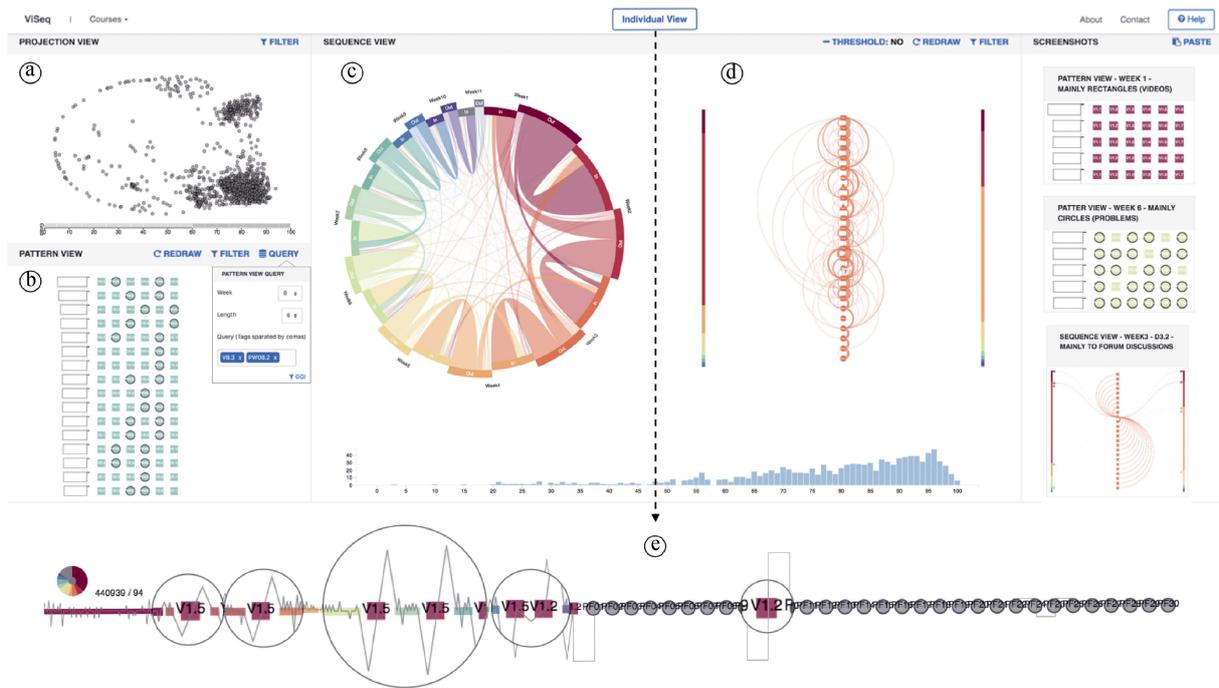


图 3 学习序列可视分析系统 ViSeq<sup>[34]</sup>

此外, 可视分析工具 CCVis<sup>[11]</sup>采用高阶网络和结构分类等先进技术, 对常见的学生学习行为进行分析、分类和总结, 并探索学生行为模式和等级分布的关系. Mu 等<sup>[41]</sup>开发了一个可视分析系统 MOOCad 以检测和解释异常学习活动. 该系统采用了一种集成

算法, 基于学习序列数据对异常群体进行识别和推理. 如图 4 所示, MOOCad 的界面由查询视图、阶段内学习模式的探索视图、针对个人的学习路径视图以及频繁模式视图组成, 通过多个角度和丰富的上下文信息对学生的非典型学习活动进行解释.

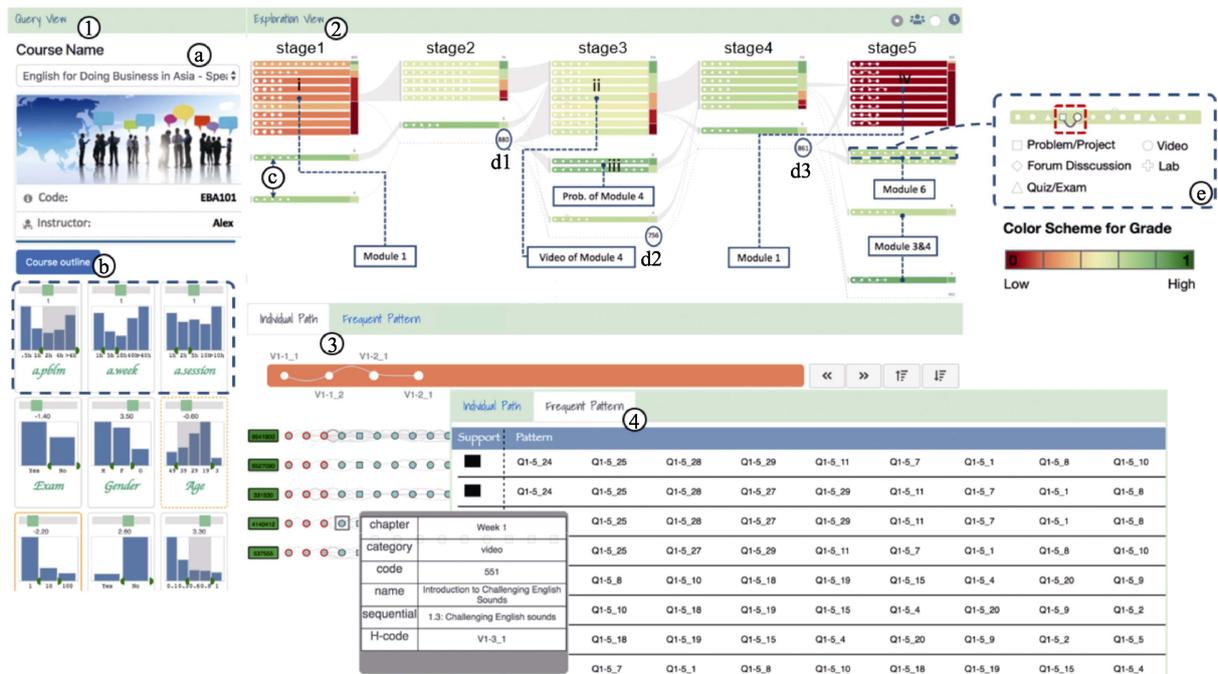


图 4 异常行为检测系统 MOOCad<sup>[41]</sup>

### 3.2 论坛交互活动分析

MOOC 论坛中的异步讨论为学生提供了互动

和交换信息的绝佳机会, 也为教师提供了关于课程内容和提高教学质量的宝贵意见. 目前与论坛

交互活动分析相关的工作可分为情绪理解和社会行为识别 2 类。

情绪理解的重点是通过分析讨论论坛表达数据, 从而洞察学生的情绪。例如, Schubert 等<sup>[42]</sup>提出了一个使用微软文本分析程序接口分析论坛讨论情绪的可视化系统。整个课程期间的班级整体情绪用箱线图中的分数表示, 使用户可以很容易地观察情绪随时间的变化情况。Wu 等<sup>[43]</sup>提出的可视分析系统 NetworkSeer, 通过平行坐标图和网络图可视化论坛中的交互信息, 将学生的时间导向行为和课程表现与社会互动相结合, 帮助用户了解不同类型学生对论坛的态度。

社会行为识别通过分析论坛中的讨论内容识别学生的社会行为。例如, iForum 系统<sup>[44]</sup>将异构论

坛数据(包括用户、帖子和线程)中的动态模式以 3 种不同的尺度进行呈现。如图 5 所示, 该系统可以方便教师识别不同成绩学生之间的行为差异, 并检查不同用户群体之间的社交互动。类似地, ForumDash<sup>[45]</sup>也对学生的社交行为进行了可视分析, 它不仅通过潜在语义分析和聚类方法构造论坛主题聚类散点图, 还利用学生回复网络图揭示了哪些学生有着最多的讨论数量, 从而帮助教师确定学生中的思想领袖。

可视分析系统 MessageLens<sup>[2]</sup>采用自动数据分析技术从不同的数据提取学习者活动信息, 构建有关论坛主题、交流文本和论坛用户交互的多个视图, 帮助 MOOC 教师从不同角度整合信息, 更全面地了解论坛。

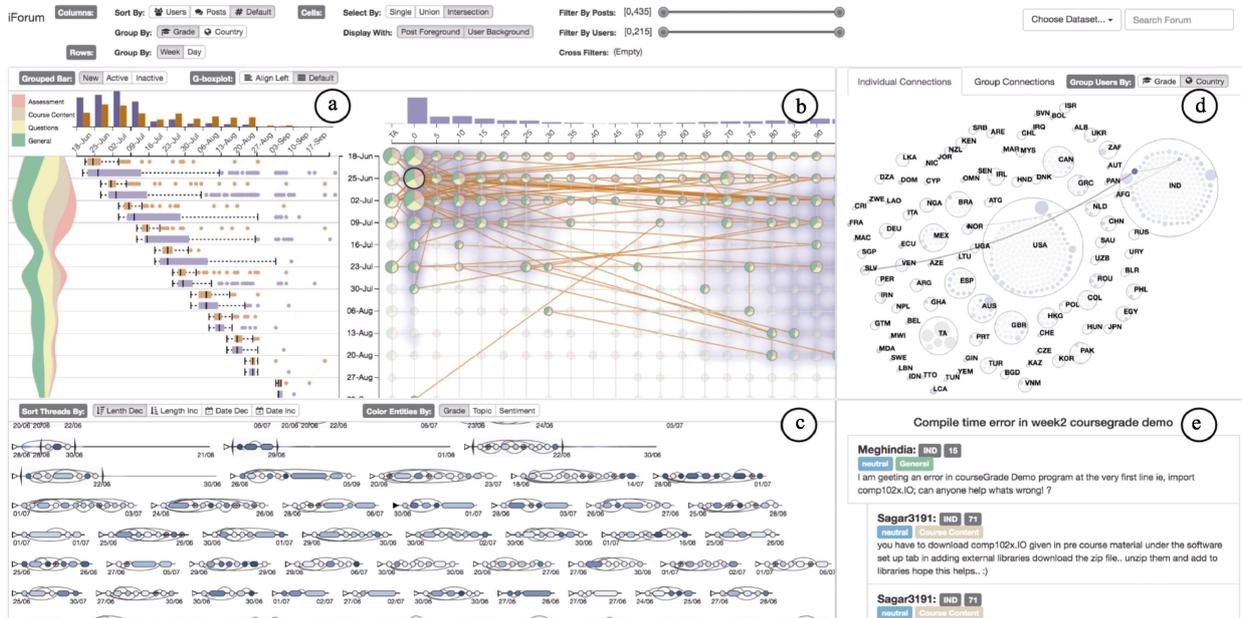


图 5 社会行为可视分析系统 iForum<sup>[44]</sup>

### 3.3 学生表现结果分析

学生的表现结果反映了学生对某一门课程的兴趣程度和收获水平。分析学生的表现情况, 既可以帮助教师了解学生的未来成绩走向, 还可以及时发现有辍学想法的学生。

例如, Trimm 等<sup>[46]</sup>根据学生的成绩进行分组, 并可视化每组的成绩变化情况。其中学生课程的历史数据用二维轨迹表示, 这些轨迹用学期成绩着色。教师根据轨迹的大致走向可以大致了解每组未来的成绩变化。不过轨迹的混乱降低了可视化的体验, 很难让人观察到单个学生的进步。相比之下, 使用节点链接图进行可视化表示的 KeyGraphs 系统<sup>[47]</sup>要简洁许多。该系统使用文本挖掘分析工具预测

学生的学习成绩, 帮助教师对教学效果和学习创新进行评估。为了推测学生是否会辍学, 大多数研究人员使用了累积平均绩点(cumulative grade point average, CGPA)<sup>[48-49]</sup>和平台日志数据等进行分析。Chen 等<sup>[3]</sup>提出可视分析系统 DropoutSeer 通过 3 种不同类型的学生活动日志(即点击流、论坛帖子和作业记录)分析学生的辍学倾向。如图 6 所示, 该系统集成了展示学生群体的聚类视图、分析学生表现的时间轴视图和了解论坛交流活动的流视图等不同类型的视觉设计, 不仅能使分析人员得知哪些学生有辍学倾向, 还能从多个尺度了解学生辍学的原因, 进而针对这部分学生采取一些措施。

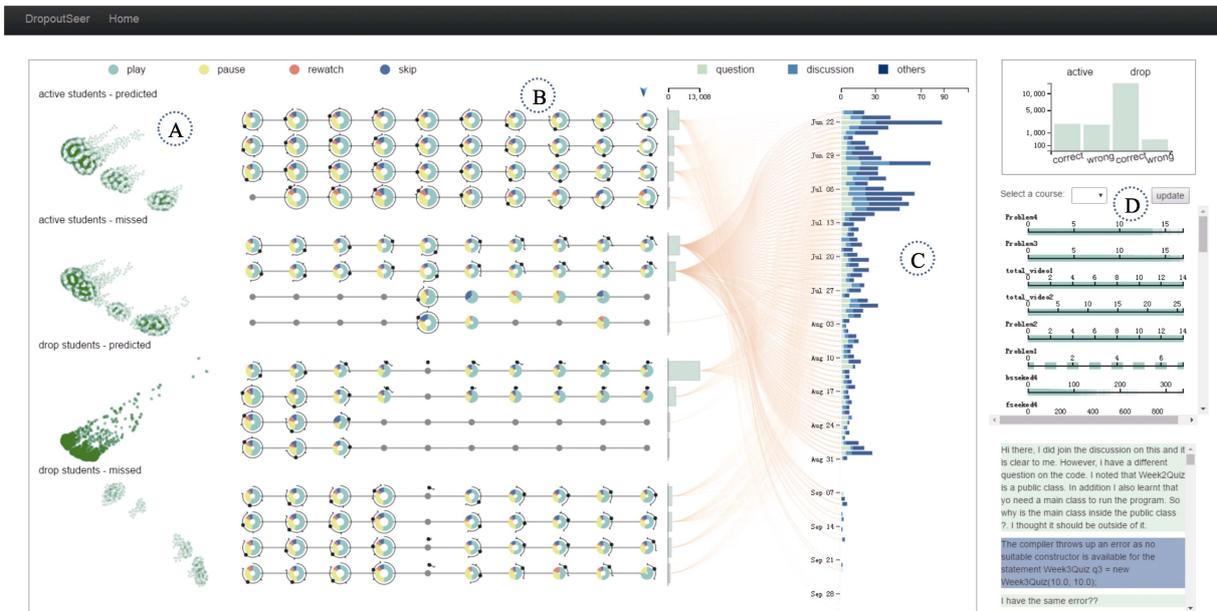


图 6 辍学倾向可视分析系统 DropoutSeer<sup>[3]</sup>

### 3.4 其他可视分析应用

也有研究人员致力于研究其他 MOOC 可视分析应用, 如在线编程和在线做题等. QLens 可视分析系统<sup>[50]</sup>通过分析学生解决问题时的行为日志, 帮助问题设计师检查详细的问题解决轨迹. 如图 7 所示, 系统由宏观层面的整体表现视图、中观层面的学生解题步骤视图以及微观层面的路径视图组

成, 从多个角度帮助用户分析学生解决问题时的行为表现. PeerLens 可视分析系统<sup>[51]</sup>针对现有的在线编程平台如 LeetCode, 通过识别相关的同伴学习群体, 比较他们的学习表现, 提出了一种通过提交类型建模学习路径的方法, 并通过一种整合的拉链式视觉隐喻方法促进用户对学习路径的理解与规划.

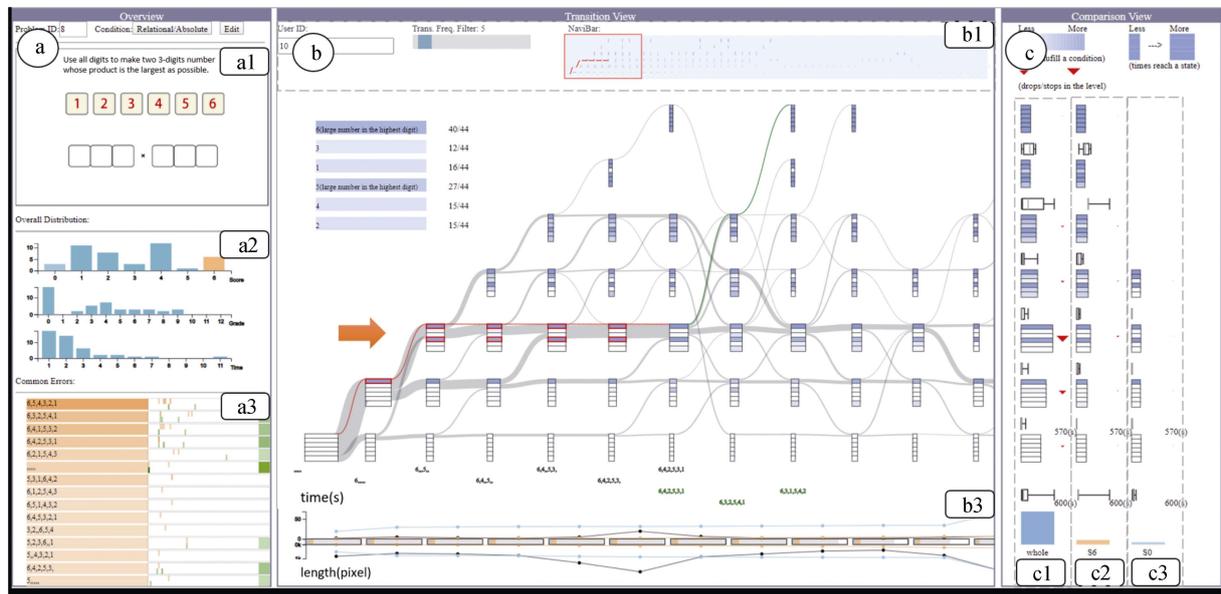


图 7 在线做题可视分析系统 QLens<sup>[50]</sup>

## 4 结 语

本文对 MOOC 数据现有的可视分析系统和方法进行了综述, 根据现有的成果, 讨论了 MOOC

中不同类型及其可视化方法, 以及 MOOC 数据可视分析的相关应用. 目前在线学习平台的注册学习人数不断上升, MOOC 教学数据也变得越来越复杂, 教学数据可视分析的需求只增不减. 然

而,该领域仍然存在很多问题和挑战,有待进一步研究和探索.

### (1) MOOC 教学多模态数据协同可视分析

MOOC 教学过程中产生的数据量庞大、种类繁多. 每种来源的数据都可以看作一种模态,如在线学习平台中的视频、音频和文本数据,以及智慧环境下收集的眼动和体态数据等. 多模态信息的整合与利用,对于教育数据分析与规律挖掘具有重要价值<sup>[52]</sup>. 目前,大多数 MOOC 数据可视分析工作只是对平台日志文件中的少数几种基本信息数据进行分析,并没有将各种多媒体数据以及物理或虚拟世界中的教学数据充分利用起来. 如何采用可视分析手段有效地结合各种模态数据,达到对在线学习平台中某一种现象、过程或环境的完整理解,将是未来值得探讨的一个方向.

### (2) MOOC 教学质量评估

对 MOOC 教学质量进行定性或定量评估,能帮助教师对教学的成果有清楚的了解. 然而,MOOC 本质上是一种“师生相对分离”的远程教育,教学过程中教师并不能像传统课堂教学那样实时监控教学过程、评估课程的教学质量. 目前大多是通过观察学生的成绩的高低这个唯一指标衡量教学的质量. 未来应从自主学习、协作学习、参与水平和信息处理等多个层面,全面地对 MOOC 教学质量进行评价.

### (3) 混合教学质量提升

由于近几年疫情的反复影响,许多学校推出线上线下混合的教学模式. 要想达到应有的教学质量,需要在线上提供优质的教学资源,帮助学生理解基础知识;线下组织活动,检测学生的学习成果或针对疑难点进行讲解. 然而,在实际教学过程中,学生的学习状态很难同步环境的变化,教师也很难时刻掌握学生的基本情况. 可视分析方法已在许多领域取得出色的成就,如何通过可视分析方法将在线教学与传统教学的优势充分结合,帮助教师有针对性地优化教学过程,改进教学目标,将是未来的一个重要方向.

## 参考文献(References):

- [1] Goulden M C, Gronda E, Yang Y R, *et al.* CCVis: visual analytics of student online learning behaviors using course clickstream data[J]. *Electronic Imaging*, 2019, 31(1): 681-1-681-12
- [2] Wong J S, Zhang X. MessageLens: a visual analytics system to support multifaceted exploration of MOOC forum discussions[J]. *Visual Informatics*, 2018, 2(1): 37-49
- [3] Chen Y Z, Chen Q, Zhao M Q, *et al.* DropoutSeer: visualizing learning patterns in Massive Open Online Courses for dropout reasoning and prediction[C] // *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Los Alamitos: IEEE Computer Society Press, 2016: 111-120
- [4] Wu M, Dewan M A A, Lin F H, *et al.* Visualization of course discussion forums: a short review from online learning perspective[C] // *Proceedings of the IEEE Canadian Conference of Electrical and Computer Engineering*. Los Alamitos: IEEE Computer Society Press, 2019: 1-4
- [5] Qu H M, Chen Q. Visual analytics for MOOC data[J]. *IEEE Computer Graphics and Applications*, 2015, 35(6): 69-75
- [6] Huang N F, Hsu H H, Chen S C, *et al.* VideoMark: a video-based learning analytic technique for MOOCs[C] // *Proceedings of the 2nd IEEE International Conference on Big Data Analysis*. Los Alamitos: IEEE Computer Society Press, 2017: 753-757
- [7] Hsu H H, Huang N F, Chen S C, *et al.* Misconceptions mining and visualizations for Chinese-based MOOCs forum based on NLP[C] // *Proceedings of the 2nd IEEE International Conference on Big Data Analysis*. Los Alamitos: IEEE Computer Society Press, 2017: 634-639
- [8] Hoque E, Carenini G. ConVis: a visual text analytic system for exploring blog conversations[J]. *Computer Graphics Forum*, 2014, 33(3): 221-230
- [9] Kizilcec R F, Piech C, Schneider E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses[C] // *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge*. New York: ACM Press, 2013: 170-179
- [10] Adcock J, Cooper M, Denoue L, *et al.* TalkMiner: a lecture webcast search engine[C] // *Proceedings of the 18th ACM International Conference on Multimedia*. New York: ACM Press, 2010: 241-250
- [11] Zhao J, Bhatt C, Cooper M, *et al.* Flexible learning with semantic visual exploration and sequence-based recommendation of MOOC videos[C] // *Proceedings of the CHI Conference on Human Factors in Computing Systems*. New York: ACM Press, 2018: 1-13
- [12] Pardos Z A, Kao K. moocRP: an open-source analytics platform[C] // *Proceedings of the 2nd ACM Conference on Learning @ Scale*. New York: ACM Press, 2015: 103-110
- [13] Wong G K W, Li S Y K, Wong E W Y. Analyzing academic discussion forum data with topic detection and data visualization[C] // *Proceedings of the IEEE International Conference on Teaching, Assessment, and Learning for Engineering*. Los Alamitos: IEEE Computer Society Press, 2016: 109-115
- [14] Shi C L, Fu S W, Chen Q, *et al.* VisMOOC: visualizing video clickstream data from massive open online courses[C] // *Proceedings of the IEEE Conference on Visual Analytics Science and Technology*. Los Alamitos: IEEE Computer Society Press, 2014: 277-278
- [15] Chen Q, Chen Y Z, Liu D Y, *et al.* PeakVizor: visual analytics of peaks in video clickstreams from massive open online courses[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(10): 2315-2330
- [16] Bao Haogang, Xing Shuang, Li Yanyan, *et al.* Teacher-oriented visual learning analytics tools in online collaborative learning

- [J]. Distance Education in China, 2019, 2019(6): 13-21(in Chinese)  
(包昊罡, 邢爽, 李艳燕, 等. 在线协作学习中面向教师的可视化学习分析工具设计与应用研究[J]. 中国远程教育, 2019, 2019(6): 13-21)
- [17] Sievert C, Shirley K. LDAvis: a method for visualizing and interpreting topics[C] //Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces. New York: ACM Press, 2014: 63-70
- [18] Sun Wenzhong. Research on topic mining in MOOC discussion forum[D]. Xi'an: Shaanxi Normal University, 2019(in Chinese)  
(孙雯中. MOOC 讨论区话题挖掘研究[D]. 西安: 陕西师范大学, 2019)
- [19] Zeng H P, Shu X H, Wang Y B, *et al.* EmotionCues: emotion-oriented visual summarization of classroom videos[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 27(7): 3168-3181
- [20] Li H T, Xu M, Wang Y, *et al.* A visual analytics approach to facilitate the proctoring of online exams[C] //Proceedings of the CHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2021: 1-17
- [21] Cooper M, Zhao J, Bhatt C, *et al.* MOOCex: exploring educational video via recommendation[C] //Proceedings of the ACM on International Conference on Multimedia Retrieval. New York: ACM Press, 2018: 521-524
- [22] Chatti M A, Marinov M, Sabov O, *et al.* Video annotation and analytics in CourseMapper[J]. Smart Learning Environments, 2016, 3(1): Article No.10
- [23] Sharma K, D'Angelo S, Gergle D, *et al.* Visual augmentation of deictic gestures in MOOC videos[C] //Proceedings of the 12th International Conference of the Learning Sciences. Singapore: International Society of the Learning Sciences, 2016: 202-209
- [24] Shneiderman B. Tree visualization with tree-map [J]. ACM Transactions on Graphics, 1992, 11(1): 92-99
- [25] Deng H, Wang X, Guo Z, *et al.* Performancevis: visual analytics of student performance data from an introductory chemistry course[J]. Visual Informatics, 2019, 3(4): 166-176
- [26] Silius K, Tervakari A M, Kailanto M. Visualizations of user data in a social media enhanced web-based environment in higher education[C] //Proceedings of the IEEE Global Engineering Education Conference. Los Alamitos: IEEE Computer Society Press, 2013: 893-899
- [27] Atapattu T, Falkner K, Tarmazdi H. Topic-wise classification of MOOC discussions: a visual analytics approach[C] //Proceedings of the 9th International Conference on Educational Data Mining. Frisco: International Educational Data Mining Society, 2016: 276-281
- [28] Schwab M, Strobel H, Tompkin J, *et al.* booc.io: an education system with hierarchical concept maps and dynamic non-linear learning plans[J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(1): 571-580
- [29] Saqr M, Alamro A. The role of social network analysis as a learning analytics tool in online problem based learning[J]. BMC Medical Education, 2019, 19(1): Article No.160
- [30] Saqr M, Fors U, Nouri J. Using social network analysis to understand online Problem-Based Learning and predict performance[J]. PLoS One, 2018, 13(9): e0203590
- [31] Ho A, Chuang I, Reich J, *et al.* HarvardX and MITx: two years of open online courses fall 2012-summer 2014[J]. SSRN Electronic Journal, 2015, 10: 1-37
- [32] Fu S W, Yong W, Yi Y, *et al.* VisForum: a visual analysis system for exploring user groups in online forums[J]. ACM Transactions on Interactive Intelligent Systems, 2018, 8(1): Article No.3
- [33] Mazza R, Dimitrova V. CourseVis: a graphical student monitoring tool for supporting instructors in Web-based distance courses[J]. International Journal of Human-Computer Studies, 2007, 65(2): 125-139
- [34] Chen Q, Yue X W, Plantaz X, *et al.* ViSeq: visual analytics of learning sequence in massive open online courses[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(3): 1622-1636
- [35] Zhou Zhiguang, Shi Chen, Shi Linsong, *et al.* A survey on the visual analytics of geospatial data[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(5): 747-763(in Chinese)  
(周志光, 石晨, 史林松, 等. 地理空间数据可视分析综述[J]. 计算机辅助设计与图形学学报, 2018, 30(5): 747-763)
- [36] Reddy M, Leclerc Y, Iverson L, *et al.* TerraVision II: visualizing massive terrain databases in VRML[J]. IEEE Computer Graphics and Applications, 1999, 19(2): 30-38
- [37] Emmons S R, Light R P, Börner K. MOOC visual analytics: empowering students, teachers, researchers, and platform developers of massively open online courses[J]. Journal of the Association for Information Science and Technology, 2017, 68(10): 2350-2363
- [38] Dernoncourt F, Taylor C, O'Reilly U M, *et al.* MoocViz: a large scale, open access, collaborative, data analytics platform for MOOCs[C] //Proceedings of NIPS 2013 Education Workshop. Amsterdam: Elsevier Science Publishers, 2013: 1-9
- [39] Wang Y, Chen Z T, Li Q, *et al.* Animated narrative visualization for video clickstream data[C] //Proceedings of SIGGRAPH Asia 2016 Symposium on Visualization. New York: ACM Press, 2016: 1-8
- [40] Xia M, Wei H, Xu M, *et al.* Visual analytics of student learning behaviors on K-12 mathematics E-learning platforms[OL]. [2021-12-15]. <https://arxiv.org/abs/1909.04749>
- [41] Mu X, Xu K, Chen Q, *et al.* MOOCad: visual analysis of anomalous learning activities in massive open online courses[C] //Proceedings of Eurographics Symposium on Parallel Graphics and Visualization. Aire-la-Ville: Eurographics Association Press, 2019: 91-95
- [42] Schubert M, Durruty D, Joyner D A. Measuring learner tone and sentiment at scale via text analysis of forum posts[C] //Proceedings of the 8th Edition of the International Workshop on Personalization Approaches in Learning Environments. London: CEUR-WS, 2018: 30-34
- [43] Wu T S, Yao Y, Duan Y Q, *et al.* NetworkSeer: visual analysis for social network in MOOCs[C] //Proceedings of the IEEE Pacific Visualization Symposium. Los Alamitos: IEEE Computer Society Press, 2016: 194-198
- [44] Fu S W, Zhao J, Cui W W, *et al.* Visual analysis of MOOC forums with iForum[J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(1): 201-210

- [45] Speck J, Gualtieri E, Naik G, *et al.* ForumDash: analyzing online discussion forums[C] //Proceedings of the 1st ACM Conference on Learning @ Scale Conference. New York: ACM Press, 2014: 139-140
- [46] Trimm D, Rheingans P, DesJardins M. Visualizing student histories using clustering and composition[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2809-2818
- [47] Wong G K W, Li S Y K. Academic performance prediction using chance discovery from online discussion forums[C] //Proceedings of the 40th IEEE Annual Computer Software and Applications Conference. Los Alamitos: IEEE Computer Society Press, 2016: 706-711
- [48] Shahiri A M, Husain W, Rashid N A. A review on predicting student's performance using data mining techniques[J]. *Procedia Computer Science*, 2015, 72: 414-422
- [49] Sweeney M, Rangwala H, Lester J, *et al.* Next-term student performance prediction: a recommender systems approach[J]. *Journal of Educational Data Mining*, 2016, 8(1): 22-50
- [50] Xia M, Velumani R P, Wang Y, *et al.* QLens: visual analytics of multi-step problem-solving behaviors for improving question design[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2021, 27(2): 870-880
- [51] Xia M, Sun M F, Wei H, *et al.* PeerLens: peer-inspired interactive learning path planning in online question pool[C] //Proceedings of the CHI Conference on Human Factors in Computing Systems. New York: ACM Press. 2019: 1-12
- [52] Lahat D, Adali T, Jutten C. Multimodal data fusion: an overview of methods, challenges, and prospects[J]. *Proceedings of the IEEE*, 2015, 103(9): 1449-1477