

# 教育大数据视角下的内容语义 分析模型及应用研究

刘清堂<sup>1</sup>, 张 思<sup>2</sup>, 范桂林<sup>3</sup>, 王 洋<sup>1</sup>, 吴林静<sup>3</sup>

(1. 华中师范大学 教育信息技术学院, 湖北 武汉 430079;

2. 华中师范大学 教育大科学研究中心, 湖北 武汉 430079;

3. 信息化与基础教育均衡发展协同创新中心, 湖北 武汉 430079)

[摘要] 教育大数据为分析学习行为与过程、提升学习绩效、科学规划与决策提供了巨大可能。当面临社会化学习中大容量、增殖速度快、具有一定情境特征的海量教育数据时,传统的分析方法遇到了瓶颈。如何建立可靠的学习分析模型,对教育数据进行科学合理的分析和解释,成为当下研究的热点和难点。本文首先介绍了内容语义分析的概念、一般过程和分析方法,在此基础上提出了数字化课堂内容语义分析方法,以及社会化学习中的内容语义分析模型和应用框架,探讨了内容语义分析模型在数字化课堂教学、教育技术学领域本体库构建、网络研修社区教师反思水平分析中的应用,以期为探索教育大数据环境下的教育研究途径和方法提供借鉴。

[关键词] 教育大数据; 内容语义分析; 分析模型; 应用研究

[中图分类号] G434 [文献标志码] A

[作者简介] 刘清堂(1969—),男,湖北仙桃人。教授,主要从事数字化学习、学习分析、知识服务等研究。E-mail: liuqtang@mail.ccnu.edu.cn。

## 一、引言

大数据的概念最早出现于1997年,意指难以用传统方法和工具分析的大量复杂数据。大数据具有体量大(Volume)、种类多(Variety)、速度快(Velocity)、易变性(Variability)、真实性(Veracity)和价值密度低(Value)等六大特性<sup>[1]</sup>。教育大数据特指在教育领域中产生的大数据,如学习行为数据、文本数据、资源、教学管理数据等。教育大数据为精准分析学习行为与过程、提升学习绩效、科学规划和决策提供了巨大可能<sup>[2]</sup>。

学习过程是一种信息加工和传递的过程,包括信息的产生、信息的呈现、信息的传递、信息的分析与处理等<sup>[3]</sup>。在数字化课堂中,教育信息主要包括有关学习

目标和教材分析的教育信息、有关教学过程的信息、有关传递过程的教育信息以及有关测试的信息。典型的教育信息分析方法有S-T分析法、ISM分析法、项目反应理论、统计分析法等<sup>[4]</sup>。随着网络与信息技术的飞速发展,数字化学习、在线学习、网络研修逐渐成为重要的社会化学习方式,新的教育信息分析方法,如学习分析、教育数据挖掘应运而生<sup>[1]</sup>。特别是,当面临社会化学习中海量、增殖速度快、具有一定情境特征的教育信息时,传统分析方法遇到了瓶颈,如何建立可靠的学习分析模型,对教育数据进行科学分析和解释,成为当下研究的热点和难点<sup>[5]</sup>。

自2004年起,美国新媒体联盟每年发布《地平线报告》,预测未来五年内将会影响教育发展的新技术,

基金项目:教育部人文社会科学研究青年基金项目“教师网络研修社区中知识共享机制研究”(项目编号:15YJC880134);华中师范大学中央高校基本科研业务费项目(项目编号:CCNU16JCZX04、CCNU16JCZX05、XJT20150006)

表 1

内容语义分析的概念

视角	名称	概 念	
心理学	语义分析法	运用语义区分量表对事物的意义进行研究的一种方法,要求被试在若干个含等级的语义量表上对某一事物或概念进行评价,以了解该事物或概念在各维度上的意义和强度	
教育学	内容分析法	实用内容分析	对某些主题词或特定的词汇进行统计分析,并赋予不同的权重(如对不同的字体、版面位置、排名顺序等),推断其出现的原因和可能的后果
		语义内容分析	语义内容分析是从内容的语义层面进行分析,又可以进一步分为指定分析、归因分析、断言分析等
		符号媒介分析	只从字面上统计特定符号(如某些主题词)出现的频率,不考虑语义,也不引入权重等,从中进行内容分析
计算机	语义分析方法	语义分析是一个复杂的过程,涉及语言学领域的研究、文本相似度的计算、本体构建、本体推理等	

智能技术(如学习分析/教育数据挖掘)成为《地平线报告》近年来关注的焦点,代表着教育领域的技术应用发展趋势<sup>[6-7]</sup>。Web技术从Web1.0向Web3.0发展,其内涵表现为内容语义特征和社会交互特性的高度融合,内容语义分析方法受到越来越多的学者关注。面对教育信息带来的研究方法论问题,建立适合教育信息特征的内容语义分析模型,在教育领域进行尝试应用,是论文研究的主要目标,以期教育大数据视角下的教育研究提供可借鉴的参考。

## 二、相关研究基础

### (一)内容语义分析的概念

不同的学科研究视角,内容语义分析的概念和内涵具有不同的表征形式,见表1。内容分析法最早来源于传播学的研究<sup>[8]</sup>,主要原理是对文献内容所包含的信息及其变化进行分析,实现对文献内容再现和推断。心理学领域,美国著名心理学家奥斯古德(Osgood)提出了语义分析法,用于测量相同概念对于不同人群所产生的不同意义<sup>[9]</sup>。一个典型的语义差量表在“好”和“坏”之间建立七个等级,通常“0”代表“中立”,“1”代表“少许”,而“3”代表的是“极端”。语义分析法主要用于理论检验,个体理解力、态度和动机的测量,以及对个体的诊断。

教育领域,内容分析法是对教育内容进行客观、系统和定量的描述的研究方法<sup>[10]</sup>。内容分析法又可细分为实用内容分析、语义内容分析和符号媒介分析。如唐焯伟等人采用内容分析法,分析了2011—2014年微课的发展现状<sup>[11]</sup>。尹睿等从课程视角,采用内容分析方法研究了三大MOOC平台课程的课程内容、课程资源、课程实施、课程评价情况<sup>[12]</sup>。管珏琪等以教师访谈文本为内容分析对象,通过内容梳理、编码和分析后发现电子书包应用的区域推进现状及策略<sup>[13]</sup>。随着语义网的发展,在信息抽取、机器学习、社会问答

系统、自动文摘等领域的内容语义分析发挥着十分重要的作用<sup>[14]</sup>。

计算机领域,内容语义分析的研究主要包括:基于句法规则的语义分析、基于相似度计算的语义分析、本体库构建中的语义分析以及基于本体的推理中的语义分析<sup>[15]</sup>。针对群体情绪分析中话题内容挖掘和对应群体情绪分析的难题,研究者引入句法规则和情感本体库,实现话题内容的准确抽取和情绪单元的准确匹配,实现群体情绪的量化和可视化<sup>[16]</sup>。本体和推理等语义技术应用到e-Learning能大幅度地提升教育系统的个性化和智能化服务水平;设计学科领域本体库,可以为学科资源动态聚合和适应性学习路径推荐建立基础<sup>[17]</sup>。

### (二)内容语义分析的一般过程

内容语义分析的过程主要包括四步:(1)明确内容语义分析要达到的目标,进行数据收集与整理;(2)明确内容语义分析的要素,并对要素的特征进行说明,建立分析类目;(3)挑选部分内容,采用确定的分析类目,通过人工编码的方式或计算机编码的方式对内容进行分析;(4)对分析后的数据进行统计,检验编码者之间的一致性,最后对数据表达的含义进行解释与说明。在单标签分类中,一致性检验最常用的参数包括百分比、Cohen's Kappa, Krippendorff's Alpha值<sup>[10]</sup>。

在计算机领域,语义分析方法是层层处理和析教育信息获得用户意图和内容意义。内容语义分析过程如图1所示。内容语义分析过程包括以下四个主要步骤。



图1 基于语义网的内容分析过程

(1)语料库和词典。用户内容通过“语料库和词

典”获得所有词的信息,包括词形、词性和语义连接。

(2)语法解析。“语法规则”阶段,解析句子的结构,包括主语、谓语、直接和间接宾语和句子的语法图表规则。

(3)语义分析。在“语义规则和模板”中包含高端的语义分析规则和模式定义,根据语法规则、语义和语义相关性计算,系统理解内容并采用恰当的语义模板表达。

(4)可视化表征。在“概念类树”阶段,包含层次化、结构化的拓补网状语言定义,如本体,更抽象化的、更具体化的、领域相关的类定义等。

### 三、数字化课堂的内容语义分析

数字化课堂中的教育信息具有量度水平低、小样本数据、包含一些模糊信息的特征。教育大数据视角下,数字化课堂的教学分析是多个小样本、单个的课堂教学分析的整合应用研究,从而发现大规模教学场景和现象中的一般规律和特征。

从学习过程的角度,数字化课堂中学生的学习过程可以理解为是一个人机、师生、生生互动的过程,可以采用行为分析、语义分析、社会网络分析等研究方法。数字化课堂中的学习过程是一个信息传递过程,可利用信息理论分析其中的学习过程规律。从教学内容的角度,数字化课堂中的教学内容是一个复杂的模糊信息,可利用模糊数学、内容语义分析方法进行研究。在数字化课堂的资源主动服务分析中,通常采用本体、语义分析、图等方法进行分析处理。数字化课堂中的教育信息结构形式主要包括矩阵、时间序列、图等。数字化课堂的教学分析主要包括教学内容结构分析、教学过程分析、集团应答分析、资源语义分析等,见表2。

表2 数字化课堂教学分析

环境	目标	方法	测验品质
数字化课堂	教育信息的结构分析	S-P表	正答率、区分度、难度、学生曲线、问题曲线、差异系数、注意系数
		项目关联结构分析、(IRS)分析	测试结构顺序
	教学过程分析	S-T分析	教学模式
		ITIAS分析	交互行为率
学生集团应答分析	集团应答曲线	正答率、区分度、应答率	

#### (一)教育信息的结构分析

教育信息的结构分析是以可视化的方法对教育

信息的结构进行分析的方法,典型的分析方法包括S-P表分析、项目关联结构分析(IRS分析)等。

S-P表是对测试得分进行分析的一种方法,通过某种测试,得到由1(正答)、0(误答)组成的得分矩阵。S-P表法将1、0按正答率以及学生得分进行排序,由此对学生和问题的特性进行分析。

项目关联结构分析是基于学生对各个问题(项目)的理解程度排序,对问题间的关联结构进行分析的一种结构分析法。

#### (二)教学过程分析

教学过程是教师和学生的信息传递过程,是一种在时间轴上展开和变化的过程。通过记录教学过程中的各种数据,通过对这些数据进行分析、处理,获得课堂教学评价的信息。主要的分析方法包括S-T分析方法、ITIAS分析方法。

S-T分析法将课堂教学行为分为两大类,即教师行为(T)和学生行为(S),以一定的时间间隔对教学行为进行采样,记录T和S发生的次数,形成S—T图,用于教学模式分析、师生的行为转换分析。

基于信息技术的互动分析编码系统(ITIAS)主要是对课堂教学中的言语行为进行分析,将课堂教学行为分为教师言语、学生言语、沉寂和技术等四大类,共18种行为,也是以一定的时间间隔进行采样,记录四种行为发生的比例,用于师生互动分析。

#### (三)学生集团应答分析

数字化课堂中,对学生集团应答情况的分析,可用于教学的形成性评价,可用于学生特性的分析,还可用于教学过程的控制。学生集团应答分析是指作为一个学生集团的全班学生对教师给予问题的应答分析。

数字化课堂环境下的教学分析,还包括话语(逐语)分析。话语分析主要是分析师生、生生对话的词性,统计词语频度,计算信息熵,用于分析教学活动中学生的认知特性和情感特性。

上述研究方法中,大容量的、大规模的课堂教学分析将从时间、空间、主体等多维度汇聚教与学信息,分析主体行为、内容特征、交互特征等普适的规律和特性。

### 四、社会化学习环境下的内容语义分析

随着网络技术在教学中的应用不断深入,基于社会化学习平台的在线学习、远程培训逐渐成为主要的学习方式。社会化学习环境具有受众面广、平台规模巨大、海量学习资源等特点,同时也面临诸多问题,亟

需教育大数据支持下的知识智能发现、服务个性导航和师生情感交互等研究方法支持。社会化学习研究面临的主要问题包括:教育资源海量存储与超链接的资源组织方式增加了学习者的“认知负荷”,导致学习者“迷航”问题的出现<sup>[18]</sup>;教育资源总量迅猛增长,使得从海量资源中快速定位所需资源的难度不断增加;用户间存在严重的“情感缺失”和“重知轻情”的问题,导致学习者学习热情不高,影响了学习的效果;信息化环境下不能对学习者的学习风格和认知水平进行有效测量,难以为学习者提供个性化的学习支持<sup>[19]</sup>。社会化学习环境出现的问题,迫切需要为用户提供个性化的、适应性的学习支持。需要在认知科学理论指导下,结合最新的计算机和网络技术,探索研究符合人类认知特点的学习支持理论、方法和工具。

### (一) 社会化学习环境下的内容语义分析框架

在借鉴计算机信息处理、学习分析、数据挖掘等技术的基础上,内容语义分析应用框架如图2所示。内容语义分析应用框架主要包括三个部分:内容语义分析应用的环境、内容语义分析方法以及内容语义分析应用。

#### 1. 内容语义分析应用的环境——社会化教与学环境

内容语义分析的应用环境主要包括数字化学习、在线学习、网络培训等,如同步课堂、大规模在线开放课程(MOOC)、教师网络研修社区等。内容语义分析的对象主要包括:微博、博客、论文、网页、书籍、聊天记录、电子邮件、作业,本地文本类格式的文件、存储在数据库中的各类文本字段的语料等。

#### 2. 内容语义分析的方法

内容语义分析的基本过程和方法包括:内容采集、特征提取、内容编码、统计分析、阐释、内容理解、特征识别、内容组织、应用推理和可视呈现。

其中数据采集的方法包括问卷调查、访谈、观察、拍照和摄像、使用计算机自动采集软件进行采集等。

特征提取的方法包括TF-IDF方法、词频方法、文档频次方法、互信息、期望交叉熵方法、二次信息熵方法、信息增益方法、卡方统计量方法以及N-Gram方法等<sup>[20]</sup>。

内容编码的方法包括质性内容编码方法和量化内容编码方法等。

统计分析方法包括描述性统计分析、多元分析、相关分析、因子分析、聚类分析、判别分析、主成分分析方法。其他的分析方法包括社会网络分析法、时序

列分析法、信息熵函数、模糊数学的方法等<sup>[21]</sup>。

特征识别的方法包括统计模式识别、句法模式识别、模糊模式识别、基于规则的方法、神经网络法等<sup>[14]</sup>。

内容组织的方法包括表格、概念图、主题图、社群图、矩阵等。

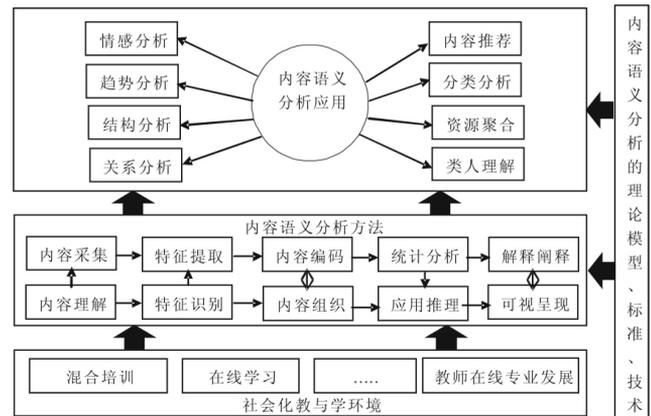


图2 社会化学习环境下的内容语义分析框架

### 3. 内容语义分析的应用

内容语义分析的应用主要包括情感分析、结构分析、趋势分析、关系分析、内容推荐、分类分析、资源聚合、类人理解等<sup>[22]</sup>。内容语义分析需借助句式识别、语法匹配、对话规则、基于语义概念库和语料库的自然语言理解方法和工具。

(1) 适应性学习内容服务。依据教学内容结构、学习者的行为特性、认知规律等,为学习者准确提供所需要的学习内容和学习资源。

(2) 个性化学习服务。依据学习者的学习状态、学习风格、认知能力等,为其提供适宜的学习路径、学习内容等。

(3) 教育知识服务(资源的组织)。对教育资源进行“理解和处理”的形式化表达,满足教育信息资源组织对信息、知识的收集、分析、描述、重组、存储、检索以及自动推理的需要。

(4) 交互信息分析。通过对用户的交互进行分析,研究用户的情感、态度、价值观、学习动机、参与度、认知度、网络关系等。

(5) 文献内容分析。对文献内容进行分析,了解所研究内容的发展趋势、内容的基本结构、内容的传播特征、对研究热点进行追踪等。

### 4. 典型的内容语义分析工具

近年来,研究者开发了多款内容语义分析工具,见表3,这些研究工具在很大程度上提升了内容语义分析的效果,促进了内容语义分析在各领域的应用。

表 3

内容语义分析基本工具

序号	软件名称	操作系统	处理语言	主要功能
1	CATPAC	Windows	ASCII	断词、词干提取、词频统计、聚类、多维标度
2	Word stat	Windows	英语、法语等	提供一个起始词和终止词的集合、列表、断词、词干提取、消除语义歧义、词频统计、共现索引、词语索引、聚类
3	DIMAP	Windows	ASCII	同上
4	Concordance	Windows	英语、汉语	提供一个起始词和终止词的集合、断词、词干提取、词频统计、共现索引、词语索引、聚类、多维标度
5	General Inquirer	Java 程序	ASCII	提供一个起始词和终止词的集合、提供工具词典列表、断词、词干提取、消除语义歧义、词频统计
6	Atlas-ti	Windows, MS-DOS	英语	数据管理、编码、检索、人工构建概念网络、多授权、输出分析数据进行统计处理和网上发布

### 五、教育大数据视角下内容语义分析的典型应用

华中师范大学一直从事教育大数据及学习分析的应用研究,典型的应用涉及数字化课堂分析、教育技术领域本体库的构建及应用、教师网络研修社区研究等方面,并取得了初步成果。

#### (一)数字化课堂教学分析应用

在多年教学和研究工作的基础上,华中师范大学教育信息技术研究团队开发了基于视频的课堂分析工具并应用于数字化课堂的分析和评价研究,如图 3 所示。

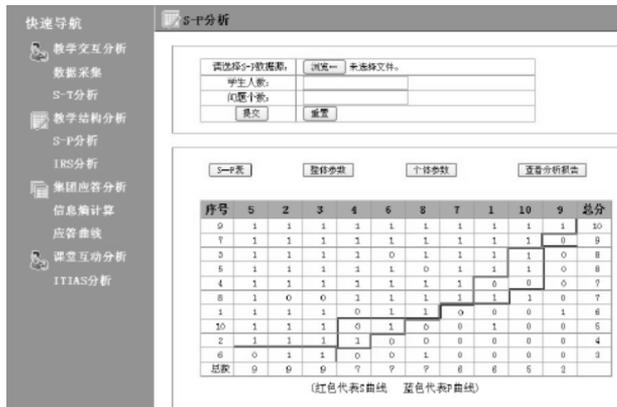


图 3 数字化课堂教学互动分析工具

课题组以中央电教馆 2013 年优质课视频《位似》为例,将该课堂的授课分为五个主要环节,分别为:(1)案例引入,概括定义环节,教师采用多媒体课件 PPT 和几何画板进行演示,时间为 5 分 18 秒;(2)探究新知,运用新知环节,主要活动是教师采用多媒体课件 PPT 和几何画板进行演示,时间为 4 分 38 秒;(3)回顾小结,梳理新知环节,主要活动是围绕多媒体演示的师生对话,时间为 1 分 56 秒;(4)达标检测,巩固新知环节,主要活动是围绕多媒体演示的师生对

话,时间为 2 分 26 秒;(5)总结知识,升华新知环节,课堂活动是围绕多媒体演示的师生对话。

运用数字化课堂分析工具中的 S-T 分析功能,对该课堂进行分析发现,在该课堂教学的不同环节,教学模式呈现不同的特征,其中第 1、3、5 环节为讲授型教学模式,特点是以教师讲授行为为主,在师生互动中,教师为主体,学生为受体;第 2 环节为对话型教学模式,特点是师生交互行为较多,互动密度高,学生互动探究比较多;第 4 环节为混合型教学模式,特点是教师引导,以练习的方式检查学生对知识的掌握程度,师生是互动的主体。最后,对整个课堂进行 S-T 分析表明,该堂数学课采用的是以教师讲授为主的讲授型教学模式。

对于数字化课堂的教学分析,我们可以借助教育大数据研究方法,对研究各变量进行大数据的统计分析。例如:对图 4 所示的类似课堂教学模式统计分析,发现不同学科的基本特征;新任教师与经验教师在师生互动、行为变化率、主题覆盖率等方面的显著特点。另外利用大数据分析,可以对课堂实时应答进行大规模的统计分析,发现未应答时间和应答时间之间的常模关系为 2~3 之间。

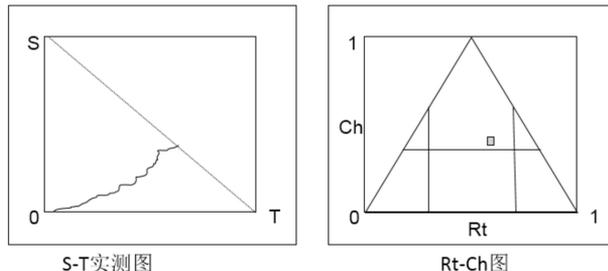


图 4 《位似》课堂教学第 2 环节的 S-T 分析图

#### (二)教育技术学领域本体库构建及应用

目前,人们处在“信息海洋”,却饱受“知识贫乏”困扰。海量的数字化教育资源无法为学习者提供个性

化学习的知识服务。资源语义聚类技术<sup>[23]</sup>通过语义聚合实现教育资源的二次组织,为学习者提供清晰的、歧义较小的、具有良好组织结构的资源集合,方便学习者快速找到自己所需要的资源,为学习者提供面向学科的知识服务,所研究的成果对于网络信息资源组织与检索、数字图书馆建设具有重要的意义。

### 1. 教育技术学领域本体库构建

面向教育技术学科需要,课题组构建了教育技术学领域学科本体库,主要包括数据源层、数据处理与语义聚类层、应用层,如图5所示。

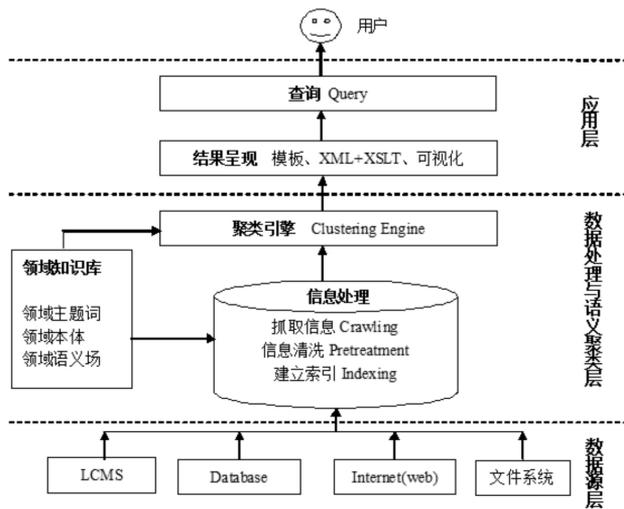


图5 教育技术学领域本体库构建系统

**数据源层:**收集教育技术学科相关的教育资源。教育技术学领域本体库的数据源主要包括教育技术领域8本主干学科教材、2009—2013年主要的期刊论文以及2009—2013年的学位论文。

**数据处理与语义聚类层:**主要作用是对收集到的教育资源进行处理和聚类。主要的方法包括数据预处理;对数据进行聚类,聚类的主要算法包括K均值算法、语义场聚类算法等;领域知识库在整个数据处理与聚类过程中提供支持。教育技术领域本体库共收录752个主题词,建立各主题词之间的相互关系,如同义关系、反义关系和上下位关系,并将领域本体按照知识元的方式进行组织<sup>[24]</sup>。

**应用层:**主要为用户提供资源检索与聚类服务,通过各种可视化方式展示聚类结果,方便用户浏览和查询。

### 2. 教育技术学领域本体库的应用

教育技术学领域本体库能提供多种应用,包括学科知识检索服务、学科资源聚合与可视化呈现、记忆知识融合与浓缩。

教育技术学领域本体库能提供学科知识检索、资

源聚合与可视化服务,如图6所示。当用户输入关键词后,可以检索相关的资源,并将资源以可视化形式提供给用户。



图6 学科资源检索、聚合与可视化呈现

教育技术学学科本体库的构建,为学科领域内的海量教育资源按需组织、深度聚合和个性化服务提供了可能。课题组对近年来教育技术学相关的期刊论文、著作和网络学习资源等进行汇聚和梳理,并通过可视化的方法进行展示。教育技术学领域本体库支持知识融合与浓缩,如图7所示。可以将学生讨论的信息、网上的信息等按照可视化的方式进行组织,呈现知识之间的关系,为用户的学习提供支持。

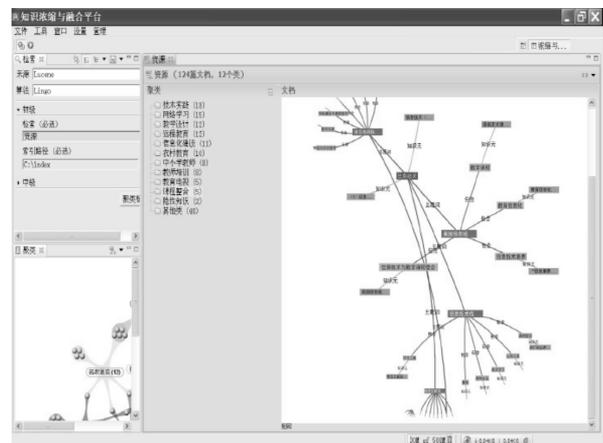


图7 领域本体库支持的知识融合与浓缩

### (三)网络研修社区中教师反思水平分析

自2013年起,我国开始实施新一轮的中小学教师信息技术应用能力提升工程项目<sup>[25]</sup>。在该项目中,提倡建立网络研修社区,促进区域间、校际间教师协同研修。教师研修的目标是促进教师积极反思,提升教师的教学实践能力。教师的教学反思是一个复杂的过程,国内外学者对教师的教学反思进行了积极的研究,国内外现有的教学反思过程模型和教学反思分析框架对我们分析网络研修社区中教师的研修情况提供了借鉴。在教师网络研修活动中,提出了教学反思的主题活动模型<sup>[26]</sup>,网络研修社区中的教学反思是一

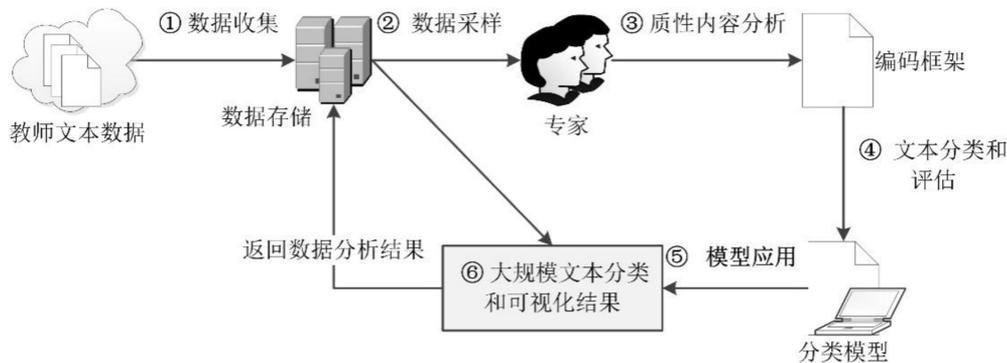


图8 网络研修社区中的教师反思分析过程

个螺旋式发展的过程,制定了教学反思的分析模型。

参与研修的教师人数多,产生了大量的学习过程数据、学习资料,如何对这些数据和资料进行分析带来了方法上的挑战。传统的内容分析方法费力、耗时,难以提出及时的干预策略。整合质性内容分析、教育数据挖掘和可视化学习分析技术的内容语义分析模型为大规模数据分析提供了可能,而完全无监督的算法得不出深度结论<sup>[27]</sup>。网络研修社区教师反思分析的过程如图8所示。

网络研修社区中的教师反思分析过程主要包括七个步骤。(1)收集教师网络研修数据。(2)从数据集中抽样部分数据,对数据进行预处理。(3)多个专家采取自下而上的方法对抽样数据进行分析,得出教师反思分析模型。在该反思模型中,教师的教学反思关注的维度包括“教学技术”(课堂教学方面)和教学理念(教学信念或对网络研修的认识方面)。教师反思的深度包括三级,即描述性反思、分析性反思和批判性反思。(4)依据质性内容分析的结果,对部分数据进行编码,然后采用分类算法对已标注的内容进行自动分类。(5)对分类模型进行训练和测试,对分类结果进行评估,比较分类器的性能。(6)对模型进行修正。(7)依据分类模型对大规模数据进行分析,将得到的结果进行可视化呈现,并将结果存储到数据库中。部分教师反思水平分析的结果如图9所示。

网络研修和校本研修是中小学教师信息技术能力提升工程的重要内容。教育大数据分析方法为1000万教师的信息技术能力提升水平的测量与评

估,以及绩效分析提供了可能性。教师反思水平的测量是其中的重要内容之一,尚需要从海量的混合培训行为数据中挖掘教育培养培训的基本特征和规律。

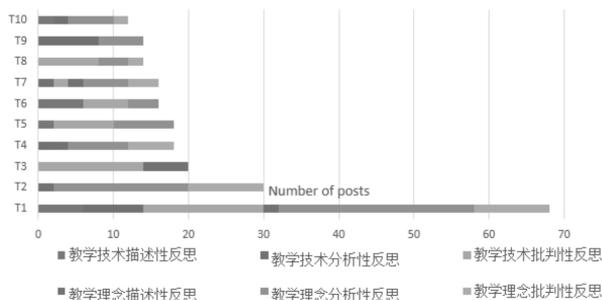


图9 教师反思分析的结果

## 六、结 论

大数据为教育带来巨大机遇的同时,也带来了巨大的挑战。如何建立合适的学习分析模型,为学习者提供学习过程支持和学习资源帮助,成为当前亟须解决的问题。本研究在对内容语义分析的概念、过程、方法和工具进行介绍的基础上,提出了内容语义分析基本框架和过程。在此基础上,论文提出了教育大数据视角下的内容语义分析模型,探讨了教育大数据视角下的内容语义分析模型在教育中的应用,分别介绍了数字化课堂的教学分析、教育技术学领域本体库的构建与应用、网络研修社区中的教师反思水平分析等。期望本文提出的模型和应用案例能对日益蓬勃发展的教育数据挖掘和学习分析相关研究提供借鉴。

### [参考文献]

- [1] 祝智庭,沈德梅. 基于大数据的教育技术研究新范式[J].电化教育研究,2013(10):5-13.
- [2] 刘雍潜,杨现民. 大数据时代区域教育均衡发展新思路[J].电化教育研究,2014(5):11-14.
- [3] 傅德荣,章慧敏,刘清堂. 教育信息处理[M].北京:北京师范大学出版社,2011.
- [4] 程云,刘清堂,王锋,王艳丽. 基于视频的改进型s-t分析法的应用研究[J].电化教育研究,2016(6):90-96.
- [5] CHEN X, VORVOREANU M, MADHAVAN K. Mining social media data for understanding students' learning experiences[C]// IEEE

- Transactions on Learning Technologies, 2014,7(3): 246-259.
- [6] JOHNSON L, ADAMS BECKER S, ESTRADA V, FREEMAN A. NMC Horizon Report: 2015 Higher Education Edition [M]. Austin, Texas: The New Media Consortium, 2015.
- [7] JOHNSON L, ADAMS BECKER S, ESTRADA V, FREEMAN A. NMC Horizon Report: 2014 K-12 Edition [M]. Austin, Texas: The New Media Consortium.
- [8] LOMBARD M, SNYDER-DUCH J, BRACKEN C C. Content analysis in mass communication: Assessment and reporting of inter-coder reliability. *Human Communication Research*, 2002,28(4): 587-604.
- [9] 廖春妍. 高校网络舆情的语义空间理论解读[D].南昌:江西师范大学,2012.
- [10] ROURKE L, ANDERSON T. Validity in quantitative content analysis [J]. *Educational technology research & development*, 2004,52(1): 5-18.
- [11] 唐焯伟,樊雅琴,庞敬文,钟绍春,王伟. 基于内容分析法的微课研究综述[J].*中国电化教育*,2015(4):74-81.
- [12] 尹睿,刘路莎,张梦叶,石娟. 国外百门大规模开放在线课程设计与开发特征的内容分析:课程视角[J].*电化教育研究*,2015(12): 30-37.
- [13] 管珏琪,苏小兵,钱冬明,余恩秀.电子书包应用区域推进现状及策略——基于教师访谈内容分析[J].*电化教育研究*,2014(10): 53-59.
- [14] 刘清堂,黄景修,吴林静,郭志强. 基于语义网的教育应用研究现状分析[J].*现代远程教育*,2015(1):60-65.
- [15] 王功辉,黄奇,秦超,杨呈中. 本体构建中的语义分析方法研究[J].*图书情报工作*,2014(4):106-111.
- [16] 林江豪,周咏梅,阳爱民,陈昱宏,陈晓帆. 基于概率潜在语义分析的群体情绪演进分析[J].*计算机应用*,2015,35(10):2747-2751.
- [17] 赵呈领,黄志芳,万力勇,陈智慧,张良,申苗苗. 基于初中物理课程的学科领域本体库构建研究[J].*电化教育研究*,2014(8):64-70.
- [18] 杨彦军,郭绍青. e-Learning 学习资源的交互设计研究[J].*现代远程教育研究*,2012(1):62-67.
- [19] 姜强,赵蔚,李松,王朋娇. 个性化自适应学习研究——大数据时代数字化学习的新常态[J].*中国电化教育*,2016(2):25-32.
- [20] LIU B. Web 数据挖掘[M]. 俞勇,薛贵荣,韩定一,译.北京:清华大学出版社,2009.
- [21] 郭志刚. 社会统计分析方法[M]. 北京:中国人民大学出版社,2015年.
- [22] 刘挺,秦兵,张宇. 信息检索系统导论[M]. 北京:机械工业出版社,2008.
- [23] 吴林静. 基于语义场模型的学科资源聚类及应用研究[D]. 武汉:华中师范大学,2013.
- [24] WU L J, LIU Q T, ZHAO G, HUANG H, HUANG T. Thesaurus dataset of educational technology in Chinese [J]. *British Journal of Educational Technology*, 2015,46(5): 1118-1122.
- [25] 郭绍青.《中小学教师信息技术应用能力培训课程标准(试行)》解读[J].*电化教育研究*,2015(9):11-15.
- [26] 刘清堂,张思. 教师混合式培训中主题研修活动设计模型研究[J].*中国电化教育*,2015(1):111-117.
- [27] ROST M, BARKHUUS L, CRAMER H, BROWN B. Representation and communication: challenges in interpreting large social media datasets[C]// *Conference on Computer Supported Cooperative Work*,2013:357-362.

## A Content Semantic Analysis Model and Its Application Research from The Perspective of Educational Big Data

LIU Qingtang<sup>1</sup>, ZHANG Si<sup>2</sup>, FAN Guilin<sup>3</sup>, WANG Yang<sup>1</sup>, WU Linjing<sup>3</sup>

(1.School of Educational Information Technology, Central China Normal University, Wuhan Hubei 430079;  
2.Research Center of Educational Science, Central China Normal University, Wuhan Hubei 430079;  
3.Collaborative & Innovative Center for Balanced Development of Information and Foundation Education,  
Wuhan Hubei 430079)

[Abstract] Owing to big data in education, it is possible to analyze learners' learning behaviors and

(下转第 93 页)

## Research on Knowledge Element-based Organization Model of Mobile Learning Resources

WANG Xiaogen, DENG Liejun, WANG Lulu, FAN Shuidi

(Research Center of Educational Informatization, Jiangnan University, Wuxi Jiangsu 214122)

**[Abstract]** In mobile learning, because of the resources distributed disorderly and aggregated insufficiently, once knowledge element is used, a complement knowledge system can be built to make knowledge acquisition more efficient. Based on the review of the present status of the organization of mobile learning resources and the concept of knowledge element, the particle size of knowledge element is determined as "terminology", "classification and category", "algorithm" and "strategy". Then a knowledge element-based organization model of mobile learning resource is put forward to represent the properties of knowledge elements, which is described as "extraction-index-link-representation". Finally, the model is applied to "Wu Cultural Heritage" WeChat public platform. The study results prove that the model can be applied to various fields such as knowledge retrieval, knowledge recommendation and knowledge visualization as well.

**[Keywords]** Knowledge Element; Mobile Learning; Organization Model of Resources; WeChat Platform

(上接第 61 页)

process, improve students' learning performances and make scientific planning and decision making. However, when it comes to the mass education data with large capacity, fast proliferation and contextual characteristics in social learning, the traditional analysis method encountered difficulties. How to establish a reliable model of learning analytics to analyze and explain the education data scientifically and reasonably becomes a heated research issue. In this paper, after the introduction of the concept, general process and analysis method of content semantic analysis, a semantic content analysis method in digital classroom, a semantic content analysis model and application framework of social learning are proposed. Then the application of that model in digital classroom teaching, construction of the ontology library in educational technology and the analysis of teachers' reflection in network research community are discussed in order to study educational research approaches and methods under the environment of big data.

**[Keywords]** Big Data in Education; Content Semantic Analysis; Analysis Model; Application Research