

基于学习者评价文本挖掘的在线课程质量提升研究

■ 刘各巧^{1,2} 陆海峰¹ 金莉苹¹

(1.四川广播电视大学工程技术学院,四川 成都 610021;

2.国家开放大学教育信息管理与信息系统研究中心,四川 成都 610021)

[摘要]随着大规模开放在线课程的涌现,在线教育正从注重课程资源积累向强调课程质量转型。作为表征隐性学习行为的课程评价数据,其蕴含着学习者对课程质量的感知,对其进行分析有益于为在线课程的质量提升提供参考。本文利用从中国大学MOOC和Coursera平台获取的2000余门课程的21万余条评价数据,通过数据预处理、分词及词性识别、依存句法分析、评价和期望观点抽取、评价情感识别等步骤对其进行文本数据挖掘。分析结果表明:一方面,学习者的评价多关注课程的内容、设计及质量,教师思路清晰、课程内容讲解详细、授课方式生动有趣、具有实践教学内容的课程更受欢迎;另一方面,学习者对于课程的改进期望主要聚焦于课程学习支持、课程内容及设计、授课方式及交流互动。上述发现可以帮助教育机构和教师从课程建设、教学模式改革以及平台功能优化等方面提升在线课程质量。

[关键词]在线课程质量;学习者评价;文本挖掘;依存句法;情感识别

[中图分类号]G72

[文献标识码]A

[文章编号]1673-0046(2020)8-0106-05

在知识开放共享的时代背景下,大量在线学习平台不断涌现。丰富的学习资源、开放的平台以及不受时空限制的学习模式吸引了大量学习者参与在线学习,与此同时,平台中积累了大量的学习行为数据。学习分析专家George Siemens教授^[1]指出,对教学过程中的行为进行分析加速了教育数据挖掘和学习分析两个重要分支的发展,其中学习分析旨在理解学习者的学习行为和心理特征。学习行为是学习分析的主要研究对象。在线学习分析不仅应当关注显性学习行为,如浏览、做作业和考试等,还需关注一些隐形学习行为,如对课程的评论和论坛中的讨论等。学习平台中存储的隐性学习行为极具价值,但在大多数研究中这些隐形学习行为数据往往被忽略,没有得到更好的利用。随着大规模开放在线课程的蓬勃发展,在线教育已从课程资源积累期逐步进入课程质量提升期,这种变化使得针对在线课程质量的评价数据受到重视。此外,这些评价数据中蕴含了学习者的情感痕迹,可以为教学决策、平台建设、资源开发提供重要参考,因而受到研究者的重视。

为了能够更好地获取在线学习平台中课程评价数据所蕴含的价值信息,本文通过依存句法分析和模板匹配方式抽取课程评价数据中的评价观点和期望观点,随后采用基于情感词典方式识别评价观点的情感倾向,进而从课程、授课教师、授课方式、学习感悟以及在线教学平台等5个评价维度对学习者的评价和期望进行分析,以期对在线课程的质量提升提供参考。

一、在线教育背景下的文本挖掘

近年来,学习者在线文本交互中的情感和主题检测日益受到关注。目前,已有部分针对在线学习平台中文本内容进行分析研究,例如,开放大学Ferguson等人^[2]构建一套基于标注特征的训练模型来自动化探究话语的类型,完成对论坛中话语的二元分类;马里兰大学Ramesh等人^[3]提出了一种基于种子词的话题模型方法来挖掘MOOC平台中学习者的话语内容,旨在帮助预测其课程通过率;卡耐基梅隆大学的Wen等人^[4]通过分析Coursera平台中的讨论帖,采用情感分析技术来监测学习集体在课程中的情感演化趋势,并发现其情感比率与退课率有显著的关联;Wyner等人^[5]在在线异步讨论论坛上开展了问答情绪分析,提出了在讨论一定教学主题引起对话中产生的五种情绪类别,如紧张、挫折、高确定性、低确定性和礼貌等对话中产生的五种情绪类别,以量化学习效果;Munzero等人^[6]采用八种情绪分类对在线学习日记进行情绪识别,以跟踪学生的情绪状态;Tian等人^[7]提出了电子学习环境下交互式中文文本的情感识别框架。该框架可以提取互动对话中的主题,识别学习者的情感状态。

总体来看,上述文献大多数主要是针对某一门课程、某一专业或某一类话题等进行观点分类和情感识别,这会导致算法应用及数据的分析具有一定局限性。本文将从当前国内外最流行的且具有完善学习者评价体系的MOOC平台上采集2000余门课程的21万余条评价数据,并对其进行分析(正反面)抽取和情感分析,

基金项目:四川省2018-2020年高等教育人才培养质量和教学改革项目“基于学习者反馈分析的远程高等教育质量提升研究与实践”(项目编号:JG2018-1242);四川广播电视大学2018-2019年青年教学改革项目“基于课程评价文本挖掘的在线教学质量提升研究与实践”(项目编号:XMZXJY2018002Q)

同时开创性地对评价数据中学习者的期望观点进行挖掘,旨在充分地了解学习者的实际需求和期望,从学习者视角来对课程建设、教学模式改革以及平台功能优化等方面提出意见和建议。

二、研究方法过程

本文所述的研究框架如图 1 所示,研究过程包括评价数据采集及预处理、评价数据分词及词性识别、评价数据依存句法分析、评价观点和期望观点抽取、评价情感识别等 5 个步骤。

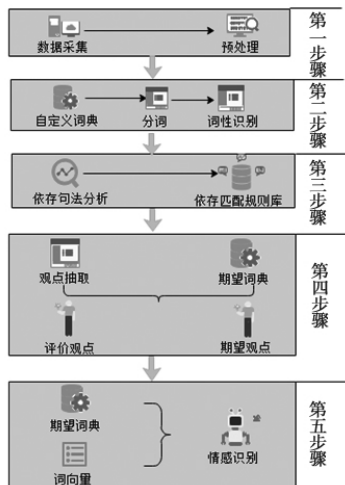


图 1 课程评价数据的观点抽取和情感识别框架

(一)评价数据采集及预处理

本文的课程评价数据来源于中国大学 MOOC(慕课)(<https://www.icourse163.org/>)和 Coursera (www.coursera.org)两个平台上 2474 门课程的共计 21 万条课程评价数据,具体数据概况如表 1 所示。使用自行编写的爬虫程序从中国大学 MOOC 平台和 Coursera 平台上采集课程评价数据,数据项包括平台名称、课程 ID、课程名称、评论文本、用户名等。在采集过程中去除一些不相关的信息和重复的样本,对评价数据进行预处理操作。

表 1 评价数据概况

平台名称	课程数	评论条数	平均每门课程评论条数
Coursera	1064	161021	151
中国大学 MOOC	1210	50052	41

(二)评价数据分词及词性识别

本文使用的分词及词性工具是 HanLP^[8],HanLP 是一系列模型与算法组成的 NLP 工具包,提供词法分析(中文分词、词性标注、命名实体识别)、句法分析等功能。但在进行分词及词性识别过程中,发现通用的分词算法对于本领域内的一些专业词汇不能做很好的区分,如表 2 中展示的例句“内容很好,但希望授课方式可以再生动流畅一些。PPT 有些欠条理,如果能总结知识点,并且提供下载就更好了”。其中“欠条理”“如果能”“就更好了”可以看作一个词语,“欠条理”可以看作是一个用来修饰 PPT 的形容词,“如果能”“就更好了”可以视为期望词(将在后续作详细介绍)。为了解决该问题,我们整理出一份行业词典,并将其作为自定义词典加载至分词算法中,对分词算法进行优化。经验证,采用该方法的分

词结果更为准确。

表 2 评价语句的分词及词性分析示例

分词方法	分词结果
分词及词性识别(原始)	[“内容/n 很好/a, /w 但/c 希望/v 授课/v 方式/n 可以/v 再/d 生动/a 流畅/a 一些/m” “PPT/nx 有些/r 欠条/n 理/n, /w 如果/c 能/v 总结/v 知识/n 点/q, /w 并且/c 提供/v 下载/v 就/d 更/d 好/a 了/u1”]
分词及词性识别(优化)	[“内容/n 很好/a, /w 但/c 希望/v 授课/v 方式/n 可以/v 再/d 生动/a 流畅/a 一些/m。/w” “PPT/nx 有些/r 欠条理/a, /w 如果能/c 总结/v 知识/n 点/q, /w 并且/c 提供/v 下载/v 就更好了/c”]

(三)评价数据依存句法分析

依存句法是由法国语言学家 L.Tesniere 最先提出^[9],其主要是将句子分析成一棵依存句法树,描述出各个词语之间的依存关系。基于前序步骤,本文将使用依存句法分析算法确定评价数据的语法体系,然后推导出句法结构,分析评价语句中所包含的句法单位及之间的关系。例如,“内容很好,但希望授课方式可以再生动流畅一些。PPT 有些欠条理,如果能总结知识点,并且提供下载就更好了。”的依存句法分析结构如表 3 所示,在表 3 中依存关系表示当前词语和关系 id 对应的词语之间的句法关联关系,其中核心关系表示该词是整个句子的核心,它的关系词语 id 是 0,是整个依存句法树的根。句子中其他词语直接或者间接与核心词语存在关联关系。比如 id=1 的“内容”这个词语和 id=2 词语“很好”之间存在一种主谓关系的依存关系,其中名词性词语“内容”作为主语是评价的对象,形容词性词语“很好”作为谓语是评价的情感特征。

表 3 评价语句的依存句法分析示例

例句	词语 id	词语	词性	依存 id	依存关系
句子一 内容很好,但希望授课方式可以再生动流畅一些	1	内容	n	2	主谓关系
	2	很好	a	0	核心关系
	3	,	w	2	标点符号
	4	但	c	5	状中结构
	5	希望	v	2	并列关系
	6	授课	vn	7	定中关系
	7	方式	n	10	主谓关系
	8	可以	v	10	状中结构
	9	再	d	10	状中结构
	10	生动	a	5	动宾关系
	11	流畅	a	10	并列关系
	12	一些	m	10	并列关系
句子二 PPT 有些欠条理,如果能总结知识点,并且提供下载就更好了	1	PPT	nx	3	主谓关系
	2	有些	r	3	状中结构
	3	欠条理	a	0	核心关系
	4	,	w	3	标点符号
	5	如果能	c	6	状中结构
	6	总结	v	3	并列关系
	7	知识点	n	6	动宾关系
	8	,	w	6	标点符号
	9	并且	c	10	状中结构
	10	提供	v	6	并列关系
	11	下载	v	10	动宾关系
	12	就更好了	y	11	动宾关系

从语言学的角度来看,评价对象与其对应的评价情感特征之间满足一定的依存关系。基于此,我们在采集的评价数据集上通过统计词语的词性和依存关系出现的频率,选取出现频率较高的规则来建立评论观点抽取规则,形成依存匹配规则库,如表4为部分评论观点抽取规则。为了提高规则的适用度,在选取依存匹配规则时,需考虑选择名词性词语或代词作为评价对象,形容词、副词、动词以及情感词典作为评价情感词。

(四)评价观点和期望观点的抽取

评价数据不仅仅包括对课程、授课教师及授课方式等进行客观评价,同时学习者还会根据自身需求提出主观期望。为能够更好地抽取评价数据中的评价观点和期望观点并识别情感,基于第三步基础上,通过将依存匹配规则抽取评价观点的直接关系词,与期望词典中的词语进行比对判定当前的观点属于评价观点还是期望观点,为此根据课程评价的数据特征手工整理了一份期望词典,如表5所示。基于此,本文中的评论观点将使用<评价对象,情感特征>二元组表示,其中评价对象为评论观点的评价维度,如课程、授课教师、授课方式以及在线教学平台等,情感特征为对评价对象发表的观点或情感表达。但仅仅将情感特征抽取出来往往还不够完善,缺乏情感倾向分析,我们将在第五步评价情感识别中补充情感的程度修饰词和否定修饰词,完善观点的情感特征,并计算情感极性。

(五)评价情感识别

本步骤主要工作是基于第四步抽取出的评价对象和情感特征基础上识别出评价数据中情感特征的情感倾向。为了准确描述评价的情感特征,用定义<情感特征,否定修饰词>程度修饰词三元组来表示一个情感特征。

首先,提取情感三元组。在实际语言表述中,情感表述往往会搭配一些副词来修饰。比如:“不”“很”“非常”“有一点”等,这种修饰词语可以分为两类:一类是程度修饰,如表6中展示的“非常”“有些”等,这类修饰词会增强或者减弱被修饰词所表达的情感特征,而不会完全改变被修饰词的情感极性。另一类是否定修饰,如表6中展示的“不”“非”等,这类修饰词会将被修饰词所表达的情感特征置反。我们将使用依存句法分析的方式来提取评价数据中表达情感倾向特征词的否定修饰词和程度修饰词,即在依存句法树上提取出与情感特征词有直接关系的否定修饰词和程度修饰词。为了解决程度修饰词和否定修饰词互相修饰的情况,再递归查找提取出的

表4 依存匹配规则库部分规则表

编号	依存关系	词性约束	例子
1	主谓关系	名词+形容词	内容好
2	主谓+动宾关系	名词+动词+形容词	内容结束有点突兀
3	主谓+动补结构	名词+动词+形容词	老师讲得通俗易懂
4	主谓+定中关系	名词+副词+形容词	课讲得漂亮
5	动宾关系	动词+名词	强化知识
6	定中关系	形容词+名词	很好的课程
7	定中关系	形容词+动词	很好的讲解
8	定中关系	动词+形容词	组织很好
9	状中结构	形容词+动词	不够详细
10	动补结构	动词+形容词	讲得很好

与程度修饰词和否定修饰词还有直接关系的否定修饰词和程度修饰词。

表5 期望词典

编号	期望词	编号	期望词
1	希望	7	渴望
2	如果能	8	盼望
3	建议	9	需要
4	如果可以	10	就好了
5	要是	11	但愿
6	期待	12	就更好了

表6 部分常用情感特征修饰词表

修饰词类型	修饰词	
否定修饰词	忌、否、没有、不要、不是、不用、不必、不曾、无须、并非、毫无、决不、休想、永不、未尝、从不、未必、未曾、未能、难以、尚未、从没、绝非、远非、切莫、切勿、禁止、尚无、绝不、毫不、不……	
程度修饰词	强化修饰词	万、倍加、非常、极、极其、十分、不得了、要命、甚为、万分、越来越、无以、过于、尤为、更加、愈加、颇为、最为、百分、何等、万般、远远、愈来愈、尤甚、太……
	弱化修饰词	略为、有些、稍稍、一点、一些、不大、弱、略加、不过、些小、些微、有点、相对、半点、不甚、稍许、或多或少、多多少少、不丁点儿……

其次,确定情感倾向的极性(正评价、负评价和中性评价)。我们采用基于情感词语词典的方法计算评价情感特征词语的情感极性,参考知网 HowNet^[9]的情感词语极性表对评论数据集中抽取出的评价情感词语进行极性的设置,如果评价情感词语不存在情感词语词典中,则通过计算该词语与词典中词语的相似度判断它的极性。具体的做法是引入 Word2vec 词向量模型^[10],Word2vec 是 2013 年 Google 开发的一款用于词向量计算的工具有,可以很好地度量词与词之间的相似性。我们在通用文本数据集和本文所使用的评价数据集上训练了 Word2vec。对于不存在情感词语词典中的情感词极性的计算,采用已训练的 Word2vec 模型获取情感特征词语的稠密向量表示,并计算该向量与情感词典中词语向量表示最相似的 5 个情感特征词语,统计排名前五个的词语极性,采取投票机制确定该词的最终极性。

最后,结合上述步骤获取的程度修饰词、否定修饰词及情感词极性共同确定情感特征的情感倾向。

三、实验结果和分析

本文采用“数据采集—分词及词性识别—依存句法分析—评价观点及期望抽取—评价情感识别”的技术路线对 21 万条课程评价数据进行观点抽取及情感识别,将抽取出的观点分为评价观点和期望观点,其中评价观点 172 种 81277 条,期望观点 56 种 1289 条。下面将对这两种观点分别进行分析:

(一)学习者评价观点分析

学习者的课程评价是反映在线课程质量的一种重要途径,与通过问卷调查

获悉教学质量不同,学习者评价和反馈是以非结构化文本数据的形式嵌入到课程评论区中,这无疑增加了数据处理的难度。本研究采取基于依存句法分析和模板匹配方式,将蕴含在课程评论中学习者对课程内容、授课教师、授课方式以及在线学习平台等方面真实的评价观点抽取出来,并识别出观点的情感极性。根据评价维度的主题方向,将评价维度分为对授课教师、课程、教学方式、在线教学平台、学习感悟等五个方面。如图2所示,学习者的评价观点中对课程内容的评价占比最高,达到

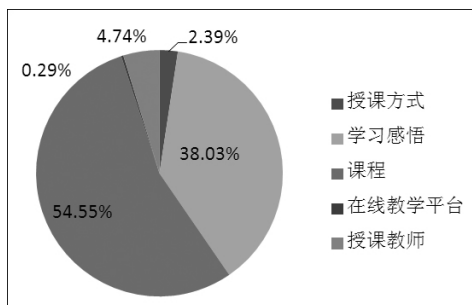


图2 评价维度类别占比

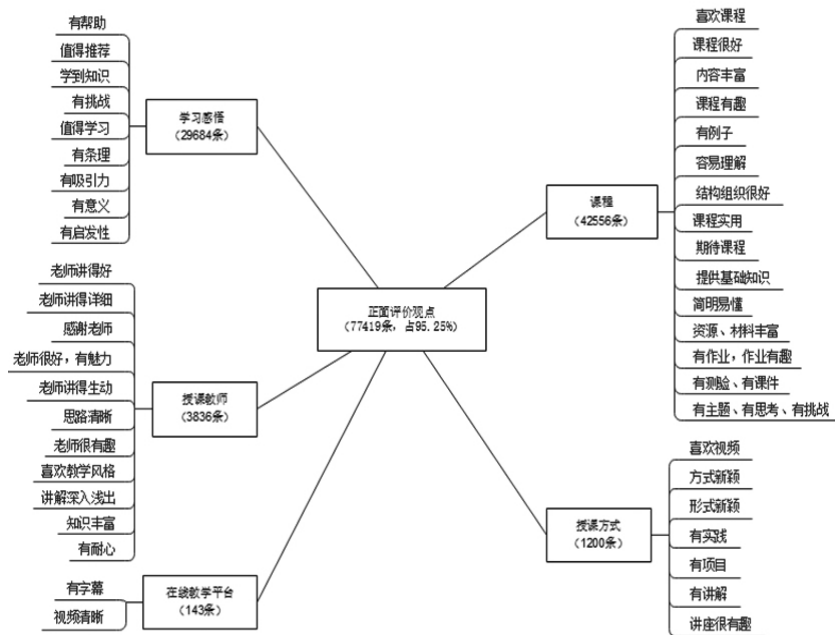


图3 正面评价观点

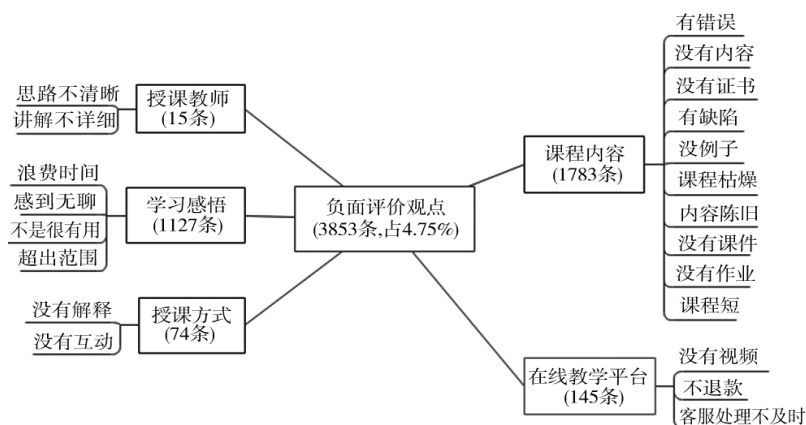


图4 负面评价观点

54.55%,排名第二的学习感悟类占比为38.03%,对授课教师、授课方式、教学平台的评价占比较少,分别为4.74%、2.39%、0.29%。学习者在评论课程时更聚焦于课程的内容、设计及质量,同时也有大量的学习者在课程评论中表达自己的学习感悟,这些感悟绝大多数体现在课程的实用性以及对工作、生活能有所帮助。对感悟类频度较高的评价观点进行课程评论追溯后,发现这些课程大多都是集中于生活、健康、工作,比如“心理急救”“运动科学”“斯坦福食品与健康导论”等,这从侧面说明了新时代快节奏的工作压力下,学习者更加注重生活和健康。从对教师及授课方式的整体评价来看,学习者更偏好教师思路清晰、课程内容讲解详细、授课方式生动有趣的教学模式,同时课程若能够提供视频、教学实践或项目会更好,这有利于学习者课后查漏补缺,完善知识框架。

在评价观点的情感极性方面,正面评价的占比达到95.25%,如图3所示,其中课程类评价占比最高,这说明大部分学习者对课程内容、设计以及授课方式感到满意,同时也说明了中国大学MOOC和Coursera平台上的课程绝大多数都是优质的,这离不开知名教育机构的参与,更离不开课程教师对课程结构的精心设计、对课程

资源的不断丰富完善以及生动有趣的授课方式等。但仍有4.75%的负面评价观点,如图4所示,这些负面观点主要是因为课程有错误、没有解释及课程无趣等引发的。为了更好地了解产生错误的原因,追溯课程评论后发现课程存在的错误主要分布在案例、课件(PPT)、测验、练习及习题等,课程没有解释评论主要集中在计算机、数学类课程中,比如“掌握Excel中的数据分析师”“大数据建模与管理信息系统”等,具体体现为代码没有注释、关键概念没解释、关键操作步骤没解释、推导过程没解释等,课程无聊及浪费时间主要体现在课程内容设计不符合学习者现实需求、课程内容粗糙、平台在处理用户撤销购买课程不及时导致收费、反馈的问题没能得到解决以及课程的任务结构设置不合理等问题。相对于正面的评价,负面评价更能帮助授课教师、平台管理者及时掌握学习者的学习状况、对课程的满意度以及对平台的诉求等,与此同时做出适当调整。

(二)学习者期望观点分析

通过对图5中展示的期望观点及频度进行分析,期望观点主要聚焦于课程学习支持、课程内容及设计、授课方式及互动交流等方面。关于课程学习支持的总体期望方面,较多学习者提出了期望课程有帮助和可以获得知识,追溯到原始评价中发现,这类期望大部分是在课程新开设不久或学习者刚学课程不久发表的占居多。一方面可能是学习者对课程整体内容设计不够了解,另一方面学习者从自身现实需

求出发 对后续课程内容表达学习期望。同时,也提醒了授课教师应注重课程目录的建设和提前公布,让学习者在选择课程时能更清晰地了解到课程的内容和所包含的知识,减少选择课程时的迷茫。

1.关于课程内容及设计方面,学习者期望课程涵盖基础知识,授课教师能对必要的概念进行解释,同时能够丰富课程内容,对于一些常规课程或实践操作类课程能够设计得更具有挑战性。这无疑对授课教师提出了更高的要求,教师在进行课程内容设计时需多注意突出课程的新颖、实用、有趣等,让学习者保持放松状态就能高效地获取知识。部分学习者期望课程能够提供资源下载的功能,包含课程视频、课件(PPT)、作业及习题等下载。平台管理者在满足学习者下载功能需求时,还应该注意保护课程制作者的知识产权,促进在线教学平台健康持续的发展。

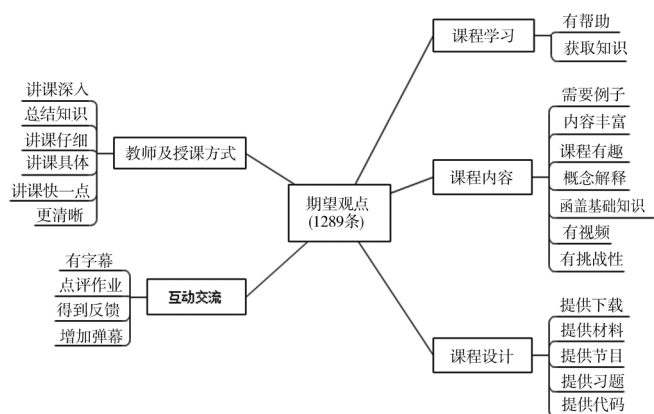


图5 期望观点

2.关于授课方式方面,少部分学生提出期望授课教师能够总结知识点、课程内容讲得仔细和具体些、讲课速度快一些,这些需求的提出主要是因为每位学习者在知识储备、理解能力等方面存在个体差异而导致的,这也是“因材施教”在线上教学的一个具体体现。

3.关于互动交流方面,学习者期望平台能提供作业点评、字幕以及弹幕等功能,这些功能的核心在于互动交流,包括学习者与学习者、学习者与教师之间的交流。课程字幕功能可以解决学习者在学习过程中因教师口音差异、语言差异而导致学习中断的问题,辅助学习者更清楚了解课程内容;同时还有助于课程的跨语种推广,扩大课程的覆盖面。弹幕是指直接显现在视频上的评论性字幕,以滚动、停留甚至更多动作特效方式在视频上呈现,是视频观看者发送的简短评论,是近年来一种新兴的视频观看体验。部分学习者期望课程能增加弹幕的功能,一方面将有助于学习者之间的沟通交流,另一方面还可以增加学习的乐趣。但过多的字幕可能会干扰学习者的学习进度和效果,同时弹幕是由学习者自主发表的,这会导致弹幕质量参差不齐,因此还需要平台管理者能积极地引导学习者发表正面向上的弹幕,为在线教学建立和谐的学习环境。

参考文献:

- [1] G Siemens, R S J D Baker. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration [C]//Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. New York: The ACM Guide to Computing Literature Press, 2012: 252-254.
- [2] Ferguson R, Wei Z, He Y, et al. An evaluation of learning analytics to identify exploratory dialogue in online discussions [C]// Proceedings of the Third International Conference on Learning Analytics and Knowledge. New York: ACM Press, 2013: 85-93.
- [3] Ramesh A, Goldwasser D, Huang B, et al. Understanding MOOC discussion forums using seeded LDA [C]// Proceedings of the ninth workshop on innovative use of NLP for building educational applications. New York: ACM Press, 2014: 28-33.
- [4] Wen M, Yang D, Rose C. Sentiment Analysis in MOOC Discussion Forums: What does it tell us [J]. Proceedings of educational data mining, 2014, 45(4):1-8.
- [5] Wyner S, Shaw E, Kim T, et al. Sentiment analysis of a student Q&A board for computer science [C]. Proceedings of the 9th KOCSEA Technical Symposium, 2008.
- [6] Munezero M, Montero C S, Mozgovoy M, et al. Exploiting sentiment analysis to track emotions in students' learning diaries [C]// Proceedings of the 13th Koli Calling International Conference on Computing Education Research. New York: The ACM Guide to Computing Literature Press, 2013: 145-152.
- [7] Tian F, Gao P, Li L, et al. Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems [J]. Knowledge-Based Systems, 2014 (55):148-164.
- [8] 何晗.自然语言处理入门 [M].北京:人民邮电出版社, 2014: 32-124.
- [9] Zeng X, Yang C, Tu C, et al. Chinese liwc lexicon expansion via hierarchical classification of word embeddings with sememe attention [C]. Thirty-Second AAAI Conference on Artificial Intelligence 2018.
- [10] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [EB/OL]. (2013-08-18)[2020-05-07]. <http://arxiv.org/abs/1301.3781>.