

# 基于文本挖掘的高校网络舆情用户情感分析研究

柳向阳 胡浩 陈则通 王智勇 李玉香<sup>通讯作者</sup>

(河北科技师范学院, 河北 秦皇岛 066004)

**[摘要]**近年来,大数据等新兴技术蓬勃发展,高校学生获取信息的途径日益多样化,互联网平台逐渐成为学生表达情感、互相沟通的主要阵地。大数据等新兴技术的发展为生活带来了诸多便利,但是在一定程度上致使高校网络舆情的管理工作面临更大的挑战。本文结合高校网络舆情的热点话题展开了深入探究,了解高校用户个人情感,结合实际情况分析了用户网络关系和情感走势,得出一系列结论,供大家参考。

**[关键词]**文本挖掘;高校;网络舆情;用户情感;分析研究

**[DOI]** 10.12252/j.issn.2096-627X.2021.09.1091

互联网技术的发展促使高校将网络舆情列入重点工作进行管控。随着新时代网民数量大度增加,网络已经成为高校学生宣泄情感、表达观点的主要场所,网络舆情不时发生,严重影响了高校日常管理。所以,本文就高校网络舆情产生的背景和用户情感进行了深入分析,为后续大家研究高校网络舆情提供经验借鉴。

## 一、网络舆情用户情感分析研究背景

大数据等新兴技术的出现为人类生活带来诸多便利,人们对于网络信息和舆情的了解更加便捷。互联网技术的完善创新了网络舆论环境,就目前的实际发展情况来看,网络管理工作的重点逐渐转向对网络舆情的管控。关于网络舆情,其实指的是因为社会中某一类或者某一个体事件的刺激,在互联网进行传播,用户对于此类事件的个人情感和态度的汇集。目前,对互联网用户组成进行分析,可以发现,高校学生占据了网络用户中一大部分。依托互联网平台和手机等终端设备,高校学生可以快速获取网络信息,针对部分热点事件表达自身想法。随着科技的进步,高校学生已经逐渐成为网络舆情的发表者和参与者。

一项调查研究结果表明,高校网络舆情已经逐渐成长为网络舆情的重点。高校网络舆情不断扩大,已经被行业内广大学者和相关部门所重视,互联网时代网络治理能力的提升需要从治理高校网络舆论入手。

随着新时代的发展,各种新兴媒体平台应运而生,建立在用户之间关系交换等基础上,社交媒体平台蓬勃发展,成为新时代网络用户的主要阵地,其中最突出的集中社交媒体有网络论坛、通讯平台及社交网站等。目前,高校学生在互联网浏览信息,发表建议的渠道越来越多,其中最为突出的几个平台就是微博、微信等媒体平台。在手机等终端设备的支持下,高校学生浏览网络信息更加便捷,在一定程度上有利于网络舆情传播。由于高校学生浏览网络信息时,个人评论和信息传播渠道相对私密,极易被网络舆论所影响。再者,高校学生尚未步入社会,对社会的认知依旧停留在学校层面,缺乏必要的安全意识和情感分析能力,对于网络舆情信息容易片面化,导致个人情感不受约束,发表一些不切实际的言论。所以,在这样的背景下,有必要对高校学生在网络舆论中的部分评论进行情感分析,了解学生内心时间情感,促进学生全面发展。

## 二、情感分析基本流程

针对互联网时代高校网络舆情,应当结合用户情感进行分析,了解高校网络用户的自身想法和内心情感,有利于高校网

络舆情工作的管控。一般来说,高校网络舆情用户情感分析工作主要从以下几个方面入手:第一,结合相关技术收集网络热点话题。以疫情期间“考研人数扩招”和“高校宿舍征用”等风靡一时的话题为例,应当利用互联网技术对用户个人评论进行收集,利用相关手段,对收集到的互联网数据开展预处理,之后结合词频统计制作热点话题云图,以便开展热点词汇分析工作。第二,结合收集到的两个网络热点话题进行情感分析,结合情感分析的方式对热点话题的影响做出判断,了解网络热点话题的负面、正面影响。建立在情感分析的基础上,针对某一个具体话题,像是“考研人数扩招”的用户情感开展深入探究,了解用户内心情感。第三,结合上述工作的完成,针对收集到的网络热点话题开展社会分析,结合互联网用户的评论数据,进行更深层次的研究。第四,结合网络舆论的分析结果,针对互联网时代高校网络舆情管控工作,有针对性的提出整改建议,促进高校网络舆情管理。

## 三、用户情感分析过程及数据处理

### (一)前期数据收集

随着疫情的影响,社会各种热点事件频频出现。其中,疫情初始“考研人数扩招”以及“高校宿舍征用”等热点话题在网络上迅速铺开,一时间在高校引发了广泛讨论,因此,以上两个话题极具代表性,对于高校舆论研究有着重要意义。目前,高校用户了解社会热点话题、发表个人言论的主要平台就是以微博、微信等为代表的社交媒体。所以,对于高校网络舆论工作的研究数据主要从以上几个社交媒体中获取。结合当时对于两个热点话题研究的数据收集工作来看,初步收集了大约超过3万条用户评论,对于高校网络舆论研究工作有着重要的意义,提供了大量数据以供研究,有利于提升研究结论与用户实际情感的契合性。

### (二)互联网数据预处理

#### 1.剔除不规范数据

初始阶段的数据收集工作,在开展过程中并不顺利,遇到了诸多问题。其中,数据信息格式不符合要求以及数据重复等现象频频出现,组织工作人员对不符合要求的数据进行预处理,几乎删除了将近一半的数据,可供研究的数据经过处理后仅剩16000条左右。

#### 2.文本预处理

这项工作简单来说就是对关键词的提取以及文本符号处理等工作的统称,文本数据完成预处理工作之后,可以有效提升文本质量,有利于研究工作的开展。一般来说,高校网络舆

情研究工作的文本处理通常结合Python语言进行展开。具体来说,主要涵盖了以下几个步骤:

首先是对收集的数据进行清洗。一般来说,数据清洗工作通常会载入附件,将数据的形式进行转换,然后存储信息。用户评论数据涵盖了英文字符和数字字符,所以可以结合表达式删除其他与研究无关的符号。其次是分词,这是文本分析的重点。做好分词工作,有利于开展算法分析、建构数据模型。再就是剔除停用词,不管什么类型符号,都有部分缺乏实际应用意义的词语,与研究并没有直接的联系,仅仅是在语句中起到连接上下文的作用。因此,对于此类词语,可以选择词语停用表辅助工作,结合实际情况删除停用词语。

### 3. 网络舆情文本挖掘

网络舆情分析工作中经常的一项工作叫作词频统计,简单来说就是对分词之后的词语进行统计,按照词频的方式进行表达。结合词频统计方式可以发现,对于“高校宿舍征用”这一热点话题,高校学生更加注重的是疫情结束之后能否对征用宿舍进行彻底消毒、个人物品能否妥善处理、宿舍设施是否会被破坏等问题。对于疫情期间高校宿舍被征用,学生大多表示认同和支持,但是更多的希望学校可以利用图书馆或者体育馆等大型场所。对于考研学生扩招这一事件,大多数学生期望个人能够被研究生学校录取,促进个人发展,提升个人学习能力。就考研专业的报考统计情况来看,计算机专业和人工智能专业成为学生报考的热门。

### 4. 高校用户情感分析

针对疫情期间“高校宿舍征用”这一热点话题进行分析,可以结合互联网工具,对用户评论中的个人情绪进行分类,一般将用户评论情绪分为负面情绪、中性情绪和正面情绪三种。通过对用户个人情绪进行分析,可以制作成柱状图的形式,便于大家了解。结合统计的用户情感数据可以发现,其中情绪类型最多的一种就是负面情绪,其比例占据了将近全部用户情绪的一半,另外两种情绪比例相当,其中正面情绪的评论比例略高于中性情绪。结合收集的用户情绪评论数据,对负面情绪用户的个人社交账号和媒体平台进行分析可以发现,其中负面情绪的用户集中于高校在校学生,在学生中,宿舍被征用的学生发表负面情绪的比例几乎占据了负面情绪用户的绝大部分。对于此类学生的负面情绪,在评论中大多集中于疫情期间学生个人物品的保护情况,学生回校之后宿舍是否经过彻底消毒、是否会出现学生被感染等类型的评论。不过,针对“高校宿舍征用”这一话题,可以发现仍然存在不少正面情绪,特别是在发布宿舍征用公告之后,部分学生坚决支国家决策,愿意贡献自己宿舍,大力支持一线工作人员。随着时间的推移,正面评论逐渐占据了该话题的主体地位,在一定程度上体现了新时代高校学生的担当。

对于疫情期间另一热点话题“考研人数扩招”来说,几乎是在知乎上收集的用户评论数据。结合互联网工具对用户评论进行情感分析,可以看到,对于考研人数扩招这一话题,绝大多数评论表现为正面情绪。对各类用户情绪数据进行统计,正面情绪占据了该话题将近一半的比例,中性情绪次之,比例最小的一部分是负面情绪。用户评论中不乏有“考研希望增

加”、“有利于国民素质提升”等类型的正面评论。此类评论的主体用户大多是即将参加研究生招考或者是未来希望参加研究生考试的在校学生,对于国家颁布的此项政策表示大力支持。不过,值得注意的是,对于此项热点话题,仍然存在诸多负面情绪,像是“工作更难找”、“学历未来贬值”等类型的评论,此类用户评论大多是由于研究生扩招所造成的一系列就业困难和学历价值降低等情况。对于部分已经读过研究生或者没有研究生考试想法的同学来说,更多的保持中立态度。

### 四、结语

随着互联网技术的不断发展,社交媒体逐渐成长为网络媒体主流,高校网络舆情逐渐成为社会舆情的主体。与社会网络舆情相比,高校网络舆情网络传播速度更快,相对更为敏感。并且,高校学生思想尚不成熟,对待问题容易感情用事,高校网络舆情会引起多个区域间高校共鸣,因此相较于社会舆情的影响范围更为广泛。所以,高校网络舆情要想实现有效管控,应当结合社会热点情况进行分析,引导舆论走向,把控舆论尺度,促进学校网络环境和谐。

### 参考文献

- [1]林智林.新媒体环境下高校网络舆情应对的困境与路径选择[J].国际公关,2021(08):124-126.
- [2]王沁,郑福民,戴溢雯.高校网络舆情导控策略研究[J].学园,2021,14(19):42-44.
- [3]董积有.高校网络舆情信息监测系统的优化与实现[D].广西大学,2021.
- [4]李静怡.中国高校网络舆情的传播与应对策略研究[D].山东大学,2021.
- [5]羊悦.融媒体时代高校网络舆情引导策略分析[J].黑龙江高教研究,2021,39(04):128-131.
- [6]隗玮.新媒体背景下高校网络舆情特征与引导策略[J].新闻研究导刊,2021,12(06):43-44.
- [7]赵晓霞,王超.我国高校网络舆情研究的可视化分析[J].改革与开放,2020(23):50-56.
- [8]杨单,程键,姚怡琦,李晓雨.基于文本挖掘的高校网络舆情用户情感分析研究[J].武汉纺织大学学报,2020,33(05):74-80.

作者简介:柳向阳(1999年10月-),男,河北省张家口市人,汉族,本科在读,研究方向:数据分析与处理。胡浩(2000年9月-),男,河北省张家口市人,汉族,本科在读,研究方向:数据分析与处理。陈则通(2002年4月-),男,河北省文安县人,汉族,本科在读,研究方向:数据分析与处理。王智勇(2000年2月-),男,河北省怀安县人,汉族,本科在读,研究方向:数据分析与处理。

通讯作者:李玉香(1964年7月-),女,河北省怀来县人,汉族,教授、硕士生导师,硕士,研究方向:数据分析与处理,农业工程与信息技术。

基金项目:2021年河北科技师范学院大学生创新训练项目“基于聚类与LDA文本挖掘的新冠疫情舆情分析”(2021-100)、2018年教育部产学合作协同育人项目“基于Python的数据分析与智能开发人才培养实践”(201801037002)。