Advances in Social Sciences 社会科学前沿, 2022, 11(4), 1489-1494
Published Online April 2022 in Hans. http://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2022.114203

Hans汉斯

基于对社交网络词频分析的大学生思想动态的研究

——以内蒙古医科大学为例

何　佳

内蒙古医科大学计算机信息学院，内蒙古　呼和浩特

收稿日期：2022年3月24日；录用日期：2022年4月20日；发布日期：2022年4月27日

摘　要

现在对大学生思想动态的调查方法更多的以传统调查方式为主，具有直观性，但也存在一定不足，在大数据分析技术迅速发展的时代，高校更加注重互联网与思想政治动态调研的结合。当代大学生更倾向于在社交网络上表达意见与情绪，本文就以内蒙古医科大学新浪微博超话社区近一年的发帖作为研究案例，通过词频分析的方法，主要体现为以下的特点：排在前三名的高频词汇为“考研”、“调剂”、“复试”，出现次数分别为136、108、97。排名前50高频词汇中涉及到的考研相关专业为“护理”、“口腔”、“中医”、“药学”，出现次数分别为85、39、29、18。通过以上分析结果，学生在超话社区最主要的话题是集中在考研升学方面，这说明医学生群体对考研的关注度以及需求极大，主要把超话平台作为一个信息与资源共享的平台，发表的情绪言论较少。

关键词

社交网络，词频分析，思想动态，考研

Research on College Students' Ideological Dynamics Based on Social Network Word Frequency Analysis

—Taking Inner Mongolia Medical University as an Example

Jia He

College of Computer and Information, Inner Mongolia Medical University, Hohhot Inner Mongolia

Received: Mar. 24th, 2022; accepted: Apr. 20th, 2022; published: Apr. 27th, 2022

文章引用: 何佳. 基于对社交网络词频分析的大学生思想动态的研究[J]. 社会科学前沿, 2022, 11(4): 1489-1494.
DOI: 10.12677/ass.2022.114203

Abstract

At present, the investigation methods of College Students' ideological dynamics are more traditional investigation methods, which are intuitive, but there are also some deficiencies. In the era of rapid development of big data analysis technology, colleges and universities pay more attention to the combination of Internet and ideological and political dynamic investigation. Contemporary college students are more inclined to express their opinions and emotions on social networks. This paper takes the post of Sina Weibo Chaohua community of Inner Mongolia Medical University in recent one year as a research case, through the method of word frequency analysis, it is mainly reflected in the following characteristics: The top three high-frequency words are "postgraduate entrance examination", "adjustment" and "retest", and the occurrence times are 136, 108 and 97 respectively. The top 50 high-frequency words related to postgraduate entrance examination are "nursing", "stomatology", "traditional Chinese medicine" and "pharmacy", with the occurrence times of 85, 39, 29 and 18 respectively. Through the above analysis results, the most important topic of students in Chaohua community is to take the postgraduate entrance examination, which shows that the medical student group lays stress on the postgraduate entrance examination. It mainly takes Chaohua platform as a platform for information and resource sharing, and less emotional speech is published.

Keywords

Social Networks, Word Frequency Analysis, Ideological Dynamics, Postgraduate Entrance Examination

1. 引言

党的十八大以来，以习近平同志为核心的党中央高度重视网络文化建设和青年思想政治教育工作，习总书记强调，要高度重视对青年一代的思想政治工作，完善思想政治工作体系，不断创新思想政治工作内容和形式，教育引导广大青年形成正确的世界观、人生观、价值观[1]。现阶段高校较为广泛使用的对学生思想动态的调查主要是抽样调查、访谈、统计分析等方式，这些方法具有直观性，但也存在一定局限：一是参与的学生是少数而不是全部，二是学生对此类问卷填写的认真程度无法保证，三是这样的调查都是主观的预设而不能挖掘出学生的深度思想状态。在今天的互联网大数据时代，如何改进传统的调研方式，有效的挖掘当代大学生思想动态数据，是当今大学教师和思政教育工作者的一个思考题。高校思政教育工作者已经意识到在互联网时代，可以通过对大学生呈现的数据分析来研究他们的思想动态与行为特点，来提高思想政治教育工作的成效[2] [3]。现阶段，相关学者正在将数据挖掘与数据分析融合进入思政教育工作，提出对改进传统高校思政教育工作的针对性建议，但研究多停留在宏观的政策改进方面[4] [5] [6] [7]。此类研究中，对大学生思想动态的数据分析限于校内数据的挖掘和汇总分析，更多是客观数据的分析，而对于语言表达等非数字数据的挖掘研究较少。目前，互联网+教育的理念不断发展，提高使用新媒体[8]等信息能力成为思想政治教育者保持自身先进性的重要途径。大学生思想政治教育过程中越来越重视与新媒体的结合，尤其是社交网络，它已经逐渐深入学生们的学习和生活中，并且对学

生们的世界观、人生观及价值观产生了很大的影响力。社交网络是大学生表达言论抒发情感的平台，学生几乎每天都会使用智能手机刷微博、聊微信等，通过了解学生在公共社交网络上的发言情况就可以侧面了解学生对学校教学管理、生活环境、爱情交友等方面的情况，从而跟踪大学生的思想动态。那么，就可以通过计算机爬虫理论抓取学生在校园或者其他公共社区平台上的发言、关注点等相关信息，运用大数据思维分析学生的思想动态，挖掘学生个人思想与现阶段的关注动态，从而及时调整思政教育的方向，改进思政教育措施，提升思政教育工作方法。因此，本文针对大学生在社交网络发表的言论数据，运用词频分析法，对大学生的思想动态数据进行挖掘分析，以新浪微博内蒙古医科大学超话社区近一年的微博文本作为案例，基于 Scrapy 框架编写 python 程序对文本数据进行采集和挖掘，分析在微博超话社交平台学生的主要关注点和动态，以便高校管理者以及思政教育工作者更加了解学生思想动态，为他们的工作提供新思路、新方法，开展有针对性的思想政治教育工作，助推思政教育工作不断完善，同时也对大学生起到更大的引领作用。

2. 研究思路与方法

微博是在新媒体时代年轻网民进行情感交流、意见表达的重要平台，新浪微博是国内微博社区目前的主流媒体，而超话社区是按照一定具有聚合性话题的集中微博内容，用户具有一定群体性。所以本文选择以新浪微博内蒙古医科大学超话近一年的发帖内容作为案例，基于 Scrapy 框架爬取浏览器数据，通过指定的 url，直接返回给用户需要的数据，提取结果。

Scrapy 爬虫框架主要由 5 个部分组成，分别是：Scrapy Engine (Scrapy 引擎)，Scheduler (调度器)，Downloader (下载器)，Spiders (蜘蛛)，Item Pipeline (项目管道)。爬取过程是 Scrapy 引擎发送请求，之后调度器把初始 URL 交给下载器，然后下载器向服务器发送服务请求，得到响应后将下载的网页内容交与蜘蛛来处理，尔后蜘蛛会对网页进行详细的解析。本文爬取目标为 https://weibo.com/p/100808b71e19c228c8eee615841b85978b8a62/super_index，在 Scrapy 框架中导入 requests 库，设置有效的 cookie、header 和目标界面 url 进行爬取，设置循环翻页，获取近一年超话文本信息。数据清洗后使用 jieba 分词对关键词频进行统计分析，获取词云图以及词频关系图。分析框架见图 1。

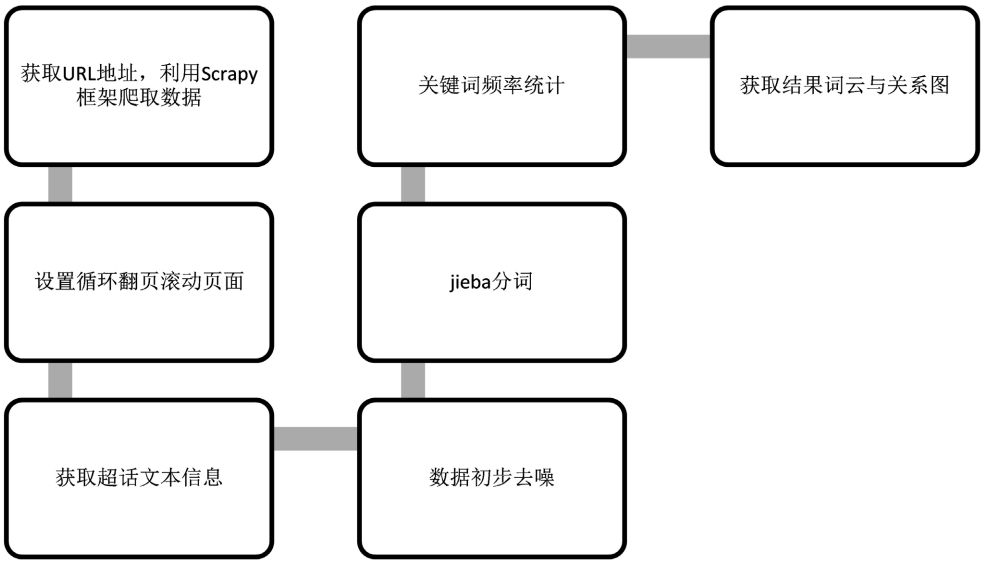

Figure 1. Content analysis framework of microblog hyperphone post
图 1. 微博超话发帖内容分析框架

3. 案例与数据分析

3.1. 案例选择

本文以新浪微博内蒙古医科大学超话社区近一年的微博文本作为案例，基于 Scrapy 框架编写 python 程序对文本数据进行采集和挖掘，本地文件共收集内蒙古医科大学微博超话社区近一年的发帖量 763 条，数据包含发帖时间、用户、微博内容等关键信息。

3.2. 文本库收集与预处理

本项目利用爬虫原理在指定的集合中读取 URL，访问相应的 web 内容，抓取有效信息的脚本，从而完成数据的收集、分类和整理。实现流程如下，首先模拟访问内蒙古医科大学超话社区，将网页 url 循环存入列表中，访问并滚动页面，遍历页面内容同时写入本地文件中。为了更好的对该语料库做分析，剔除一些无用信息以及重复度高又无具体分析意义的词汇，例如“学长学姐”、“我们”、“学校”等词汇，对数据进行降噪处理。采用“搜索引擎模式”的结巴分词，将句子精确切开，基于树结构实现高效的词图扫描。基于 TF-IDF 算法进行关键词频率统计，选取前 50 的关键词生成词图以及词图关系。

3.3. 数据分析与讨论

通过对收集语料库的词频统计，表 1 为排名前 20 的高频词汇。排在前三名的高频词汇为“考研”、“调剂”、“复试”，出现次数分别为 136、108、97。根据排名前 50 的高频词汇生成的词云图(图 2)以及词图关系(图 3)得到，学生在超话社区最主要的话题是集中在考研升学方面。这个结果符合如今的“考研热”现象，尤其对于医学生来说，国家卫计委规定住院医师要经过规范化培训，那么考研无疑是更好的选择，对于今后的发展也更加具有竞争性。排名前 50 高频词汇中涉及到的考研相关专业为“护理”、“口腔”、“中医”、“药学”，出现次数分别为 85、39、29、18。从另一个角度也可以看出，高频词汇主要围绕着考研的不同阶段进行，在备考阶段的关键词主要为“有偿”、“资料”“咨询”等，分别为 61、52、49，此阶段的发帖主要集中在对于考研事宜进行咨询以及对考研资料以及试题的收集。在复试以及调剂阶段，关键词更多为“贵校”、“谢谢”、“成功”等词汇，出现频率为 33、17、15，说明到了复试调剂阶段，除了在研招网等渠道获得信息以外，在学校的公共社交平台，可以通过师兄师姐来获得更多更为直观的有效信息。

在学业就业方面，“大赛”“方向”“导师”“毕业”“机会”等词汇也较高频次出现，说明在此社交平台上高年级的学生更为活跃。高年级的大学生面临着工作、考研等诸多选择，同时面临着在校期间是否要多参加比赛增加实践经验、就业的方向、考研如何选择导师、学校所在城市与家乡的工作的机会等等问题。而这些问题可以通过学校的心理健康中心与招生就业处进行合作来为学生排忧解难。在生活方面，“早安”“晚安”“青春”“快乐”等词汇出现较多，但关于个人情感情绪以及其他方面的词汇较少，也无其他敏感词汇。以此来分析大学生的政治观、人生观、价值观等思想动态是不足够的，缺少更多的样本数据。

4. 结论与展望

通过对内蒙古医科大学超话社区语料库的分析，发现该校学生在此平台的发帖主要集中在考研升学的信息交流和资料获取方面，这说明医学生群体对考研的关注度以及资料需求很大，学校可以围绕考研方向给学生提供更多支持，也可通过校内平台收集考研资料或组建学习小组，形成资源共享。结果显示，由于现在朋友圈、抖音、快手等更多情绪表达途径的存在，大学生的情绪宣泄越来越多样化也更私人化，

Table 1. Microblog ultra high frequency words (top 20)
表 1. 微博超话高频词汇(前 20)

序号	关键词	词频	序号	关键词	词频
1	考研	136	11	咨询	49
2	调剂	108	12	口腔	39
3	复试	97	13	贵校	33
4	医科大学	89	14	中医	29
5	护理	85	15	医学	24
6	专业	69	16	希望	22
7	研究生	64	17	同学	20
8	晚安	64	18	专业课	19
9	有偿	61	19	想要	19
10	资料	52	20	历年	18

Figure 2. Word cloud generated from the top 50 high-frequency words
图 2. 根据排名前 50 的高频词汇生成的词云图

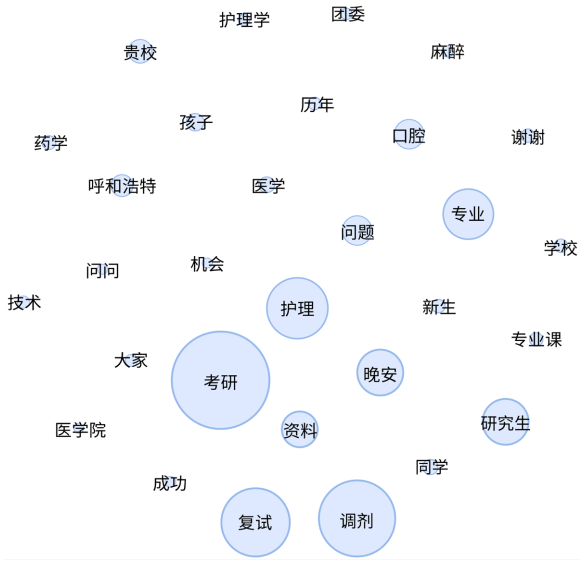

Figure 3. High frequency word graph relationship
图 3. 高频词图关系

通过公共社交网络平台收集学生的情感诉求和情绪表达不具备更多优势，可以通过学校的心理健康中心以及“互联网”+平台开创更多创新形式对学生的心理健康问题进行关注并收集更多素材进行数据挖掘与分析。

基金项目

内蒙古医科大学思想政治科研项目 YKD2021SXZZ015。

参考文献

[1] 郭鲁江. 网络文化与青年思想政治教育[EB/OL]. 中国共产党新闻网. http://dangjian.people.com.cn/n1/2017/0821/c117092-29483343.html, 2017-08-21.

[2] 武昭阳, 陈阳, 贾红达. 大数据时代思想政治教育管理模式创新研究[J]. 教育教学论坛, 2021(52): 13-16.

[3] 苏国辉, 李萌. 高校思想政治教育工作中的大数据思维[J]. 湖南工业大学学报(社会科学版), 2018, 23(3): 60-64.

[4] 任琳. 基于网络媒体和数据挖掘的大学生思想动态评估[J]. 微型电脑应用, 2020, 36(9): 136-138.

[5] 王萌萌. 大数据时代大学生思想政治教育创新研究[D]: [硕士学位论文]. 沈阳: 沈阳师范大学, 2020.

[6] 熊校良. 大学生精准引领目标下的多学科协同育人平台构建[J]. 学校党建与思想教育, 2021(5): 81-83.

[7] 王串, 章怡芳, 伍诗萌, 李芳芳, 毛星亮. 基于线上线下数据分析的大学生思想动态研究[J]. 教育观察, 2021, 10(47): 9-13.

[8] 孙丽艳. 新媒体时代高校思想政治教育的创新路径研究[J]. 文化创新比较研究, 2020, 4(25): 52-54.