

基于情感分析的网络舆情主题演化分析

——以双减政策为例

金百川, 曹旭

大连外国语大学, 辽宁 大连

收稿日期: 2022年5月3日; 录用日期: 2022年6月7日; 发布日期: 2022年6月14日

摘要

本文通过python编写爬虫程序, 收集有关双减政策的微博评论并通过snowNLP对其情感分析, 在积极与消极情感分别通过LDA模型做主题分析, 得出了“政策实施效果”、“时间分配”、“教培老师转行”、“疫情开学”、“教育公平”、“教师待遇”、“家长陪伴”、“学业减负”、“兴趣培养”、“人才培养”、“教育改革”等13个主题关键词。通过分析得知, 双减政策对于处于义务教育阶段学生的学习模式有较大的改变, 也对课外教育机构进行了有效的打压。可以看出双减政策的强势性。从长远的角度看, 该项政策有效的推动了人才强国战略的实施, 有助于教育的良好发展。

关键词

双减政策, 微博评论, 情感分析, LDA

Internet Public Opinion Theme Evolution Analysis Based on Emotion Analysis

—Taking the Double Reduction Policy as an Example

Baichuan Jin, Xu Cao

Dalian University of Foreign Languages, Dalian Liaoning

Received: May 3rd, 2022; accepted: Jun. 7th, 2022; published: Jun. 14th, 2022

Abstract

In this paper, a crawler program was written by Python to collect Weibo comments about the double-reduction policy and analyze their emotions through snowNLP. In terms of positive and

negative emotions, subject analysis was conducted by LDA model. Thirteen key words were obtained, including “policy implementation effect”, “time distribution”, “teachers changing careers”, “epidemic reopening”, “education equity”, “teachers’ treatment”, “parents’ companionship”, “academic burden reduction”, “interest cultivation”, “talent cultivation” and “education reform”. Through the analysis, the double-reduction policy has greatly changed the learning mode of students in the compulsory education stage, and implemented a strong pressure on extracurricular education institutions. We can see the strength of the double-reduction policy. However, in the long run, this policy effectively promotes the implementation of the strategy of reinvigorating the country with talents and contributes to the sound development of education.

Keywords

Double-Reduction Policy, Weibo Comments, Sentiment Analysis, LDA

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

国务院办公厅印发的《关于进一步减轻义务教育阶段学生作业负担和校外培训负担的意见》于2021年7月24日正式落地实施,要求减少义务教育阶段学生的作业负担和校外培训负担(以下简称双减)。

双减政策出台之后得到了民众的热切关注,公众快速通过各种社交网络平台(微博、微信、论坛、短视频平台等)展开了讨论,并形成了大量的舆情。然而网络舆情的影响力很大,当大部分民众偏言舆论,有可能导致不稳定的因素产生。因此如何有效挖掘出网民对双减政策隐含的情感态度和了解双减政策的实施效果,是实施双减政策过程中必须探究的问题。

过去的研究显示,许多家长不能满足学校的教学质量,对课外补课的需求越来越严重,家长希望通过额外的教育提高孩子的学习成绩,然而在这全民补课的环境下,滋生出了许多课外培训机构,同时教育的成本大大提高,学生的学习压力越来越大。这影响了教育和培训行业的正常发展。

所谓双减政策就是要减轻校内的学业负担,禁止学校占用学生寒暑假,剥夺挪用学生体育音乐美术课等,同时也要减轻校外学生负担,治理校外学科类的培训机构,严格禁止校外培训占用学生寒暑假。国家希望通过双减政策,减轻学生的学业压力,解决严重的内卷现象。

综上,本文对人们在微博关于双减政策讨论的评论进行爬取、情感分析、分类,以了解双减政策的实施效果以及人们对双减政策的态度。

2. 实验流程图

数据收集和处理流程如图1所示。



Figure 1. Flow chart of Weibo data collection and processing

图1. 微博数据收集和处理流程图

3. 数据收集

互联网的迅猛发展滋生出了各种各样的社交媒体, 然而微博数据具有普遍性、动态性、民主性的特征。作为公开的社交平台, 微博能够集合媒体、当事人、政府官微以及各领域大 V, 多角度解读事件, 挖掘真相[1]。据《2020 年微博用户发展报告》, 微博 2020 年 9 月月活用户数为 5.11 亿, 9 月日活用户 2.24 亿[2]。所以说当一件热点事件发生后微博的数据更能反映出公众对此的普遍看法。

微博数据在以往的突发公共卫生事件的研究中已多有应用。安璐[3]等在研究突发公共卫生事件的预测时选择的是与埃博拉爆发相关的新浪微博数据作为微博影响力建模的样本。赵丹[4]等以新浪微博“埃博拉”热点话题数据为样本, 进行新媒体环境下的微博舆情传播态势模型构建的研究。周清清[5]等在研究微博短文本上的基于迁移学习的情绪分类任务时, 采用了 H7N9 的微博语料。王睿哲[6]等以 H7N9 禽流感疫情首周的微博内容进行情感倾向分析、共词网络分析。

本文的数据是通过使用 Python 内置的 urllib.reques 库指定 url, 获取网页数据, 使用 bs4 库中 BeautifulSoup 进行网页解析, 编写新浪网微博评论的爬虫程序, 爬取了从 2021 年 7 月 24 日双减政策正式颁布到 2022 年 4 月 24 日在微博“双减”话题下的微博评论数据, 共计 12,195 条。

4. 实验流程

4.1. 数据清洗

为了更加清晰的抓取评论, 本文通过 python 中的 re 库编写正则表达式, 将评论数据中的网页源代码过滤。并通过 Excel 筛选的方式去除重复数据和空数据。

4.2. 分词

分词选用的工具是 python 库中的 jieba。由于许多词语是随着一些热点时间的发展而形成, 例如“双减”, “教培”, “新冠”等, jieba 不会准确将它们分隔, 所以设置了自定义分词表将这些词添加进去。例如: “双减下的多彩课堂”分词结果“双减/下的/多彩/课堂”。

4.3. 情感分析

目前情感倾向分析的主要方法分为两类: 一种是基于情感词典的方法; 另一种是基于机器学习的方法, 如基于大规模语料库的机器学习。前者需要标注好的情感词典; 后者则需要大量的人工标注的语料作为训练集, 通过提取文本特征, 构建分类器来实现情感分类。本文情感分析选用的是 python 的 snowNLP 库中基于情感词典的文本情感分析, 通过打分, 将情感倾向划分为积极和消极。

4.4. 主题分析

网络舆情是随着网络和信息技术快速发展和普及产生的, 聚类分析是网络舆情分析中获取热点话题的主要手段。通过聚类析可以发现潜在的隐性知识[7], 客观地揭示对象间的相似程度, 从而发现隐含的客观规律[8]。

本文选用的模型是隐含狄利克雷分布(Latent Dirichlet Allocation, LDA)模型。LDA 主题模型是能够提取数据中的隐结构的统计模型[9]。LDA 利用一组组显变量(特征词), 构成一个个隐变量(主题), 通过主题上的概率分布来表达每篇文档[10]。

本文将积极情感文本和消极情感文本分别做隐含狄利克雷分布的主题分析, 这样就可已得到双减持积极态度的主题和消极态度的主题。然而主题的数量通常需要人为指定, 而模型的困惑度(perplexity)会随着主题数量的不同而发生变化, 在自然语言处理中困惑度是用来衡量语言概率模型的优劣。本文通过

设置不同的主题数量, 绘制困惑度曲线, 为了避免过拟合现象, 最后积极情感的主题数量确定为 9 个, 消极情感的主题数量确定为 4 个。

5. 数据分析

5.1. 主题分类

将消极和积极的评论分别通过 LDA 模型分析得出对应的主题, 如表 1、表 2 所示。

Table 1. Negative theme

表 1. 消极主题

主题编号	主题名称	主题名称
1	政策问题	家长 问题 政策 工资 学生 工作 家教 国家 韭菜 专升本 作业 时间 公务员 办法 教师 服务 教育 老师 英语 评论 政府 有点 机构 地方 教培
2	时间分配	孩子 老师 时间 教育 教育局 辅导 办法 疫情 教师 政府 机会 作业 小时 家长 领导 教培 工资 学校 事情 学科 国家 机构 家教 问题 政策
3	教培老师转行	机构 培训 学科 教培 评论 学校 维权 公务员 服务 事情 作业 家教 政府 影响 教育 家长 学生 老师 教育局 英语 辅导 问题 工资 国家 疫情
4	疫情开学	开学 影响 英语 公司 地方 有点 领导 小时 疫情 学校 作业 政策 学生 孩子 教育局 机构 事情 辅导 公务员 办法 培训 学科 家长 工资 老师

Table 2. Positive theme

表 2. 积极主题

主题编号	主题名称	主题名称
1	教育公平	公平 教师 专业 作业 教育资源 小孩 英语 教培 优质 创业 课堂 资源 专升本 视频 收费 一节课 工作 意义 重点 教学 竞争 自律 时间 差距 待遇
2	教师待遇	学校 老师 行业 工资 能力 社会 服务 责任 小时 待遇 基本 公立学校 地方人民 活动 工人 资源 农村 班主任 教培 毕业 情况 竞争 辅导 原因
3	家长陪伴	家长 事情 陪伴 教育部 数学 作业 小孩子 职业 指导 压力 精英 历史 新东方 家教 关键 世界 布置 检查 情况 结果 个人 家庭 领导 原因 时间
4	学生减负	老师 学生 时间 作业 放学 小学 活动 上学 回家 辅导 人家 课外 辅导班 教培 教育局 布置 班主任 意义 教学 检查 领导 结果 新东方 毕业 网课
5	兴趣培养	兴趣 政策 交流 感兴趣 情况 新东方 领导 辅导班 个人 生活 教育 小学 毕业 目的 教培 结果 活动 时间 国家 政府 全面 影响 放学 素质教育 读书
6	人才培养	国家 问题 人才 合理 文化 企业 影响 地区 方式 文科 经济 教学 农村 技术 原因 社会 能力 差距 辅导 主管 收费 工人 重点 世界 领导
7	教育改革	教育 发展 全面 改革 裁员 水平 主管 目的 职高 私信 精彩 职业 资本 素质教育 工作 工人 新东方 素质 家庭 情况 关键 回家 方式 原因 大学
8	打击教培	机构 培训 考试 大学 政府 学科 个人 观点 家庭 培训班 转型 自律 网课 意见 基础 疫情 成绩 素质 毕业 家教 小学 教育资源 地方 教育局 差距
9	丰富生活	孩子 工作 开学 父母 生活 成绩 时间 家庭 读书 机会 竞争 天赋 博物馆 结果 领导 辅导 毕业 家教 工人 能力 差距 作业 素质教育 教育局 地方

通过 LDA 主题模型分析, 得到了不同的主题分布, 图 2 和图 3 分别是积极和消极主题分布饼状图, 以显示出不同主题的情绪分布。

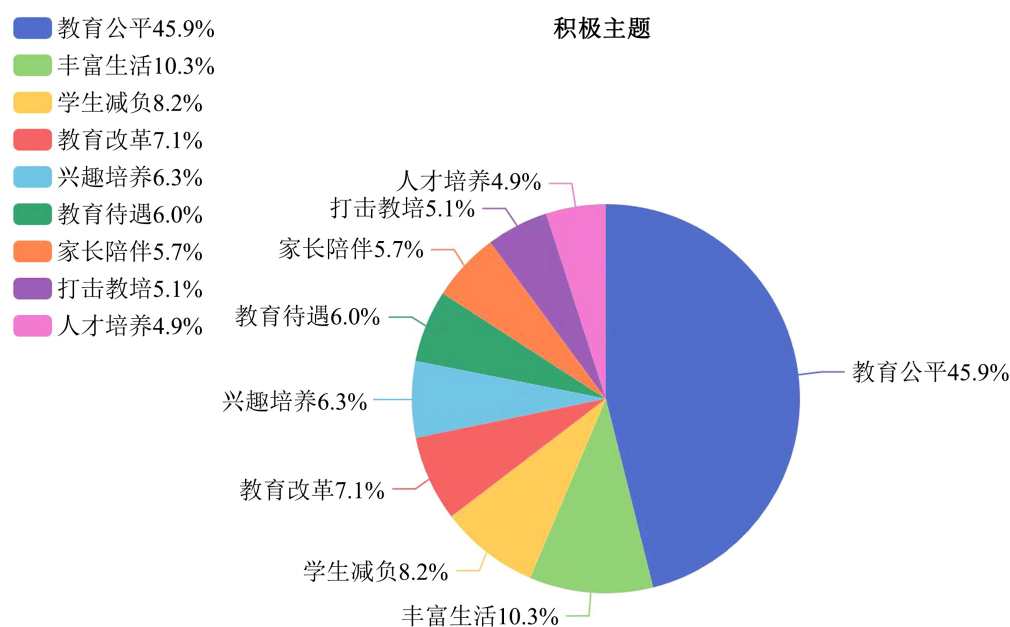


Figure 2. Positive topic pie chart

图 2. 积极主题饼状图

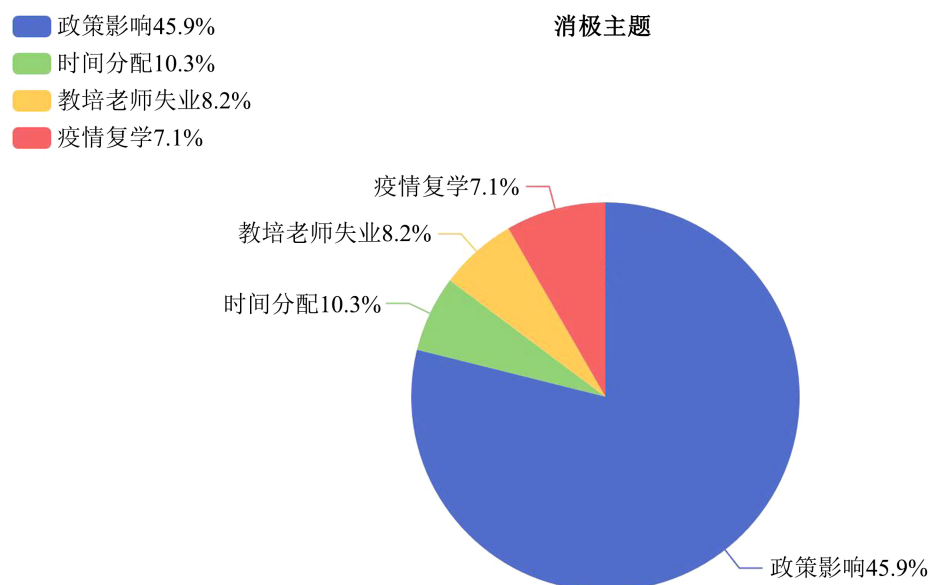


Figure 3. Negative topic pie chart

图 3. 消极主题饼状图

5.2. 主题分析

图 2、图 3 反映出人们对不同话题的关注程度不同, 其中积极方面是以“教育公平”为主, 可以看出, 公众主要关心的是一直以来未能解决的教育不公平的问题, 然而通过双减政策的强制性, 有效的避

免了这一点。其在所有主题中“丰富生活”，“学生减负”，“教育改革”，“时间分配”，“打击教培”，“教培老师失业”，“家长陪伴”等主题反应了双减政策的作用对象不仅仅是学生，国家的教育体制、教培老师、家长也受到了一定的影响。以下将针对每个主题找出代表性的关键句子做出分析。

(一) 消极主题

1) 政策实施效果关键句

相关关键句有“说实话，这个政策一刀切，差距更大了，影响的就是孩子，学不会的也没地方补，一大批人失业，到时候又得重新恢复何必呢，出政策也得根据实际情况出发吧。”、“6.17号才买的8月550的鲸鱼线上英语的课程，还没开课，双减政策来了，退课怎么办，销售顾问告诉我退课的很多要排队，如果支撑不下去申请破产坑的不是老百姓的钱吗？”从这些关键句中，可以看出，许多家长对于政策实施效果产生了抱怨，主要表现对于培训机构的一刀切现象，导致许多家长交的费用不能退还；还有的人担心因为双减政策许多条件优厚的家庭的孩子仍可以请一对一家教，但是大多数普通人却不能，导致原本想解决的教育平等问题反而加重了。

2) 时间分配关键句

相关关键句有“我不太明白作业时间为什么要控制在一个半小时内，且不说每个人速度不同，初中这么多学科，平均每科作业时间十几分钟？”王治峰^[11]对中小学落实双减政策的实践路线研究中表明要严格控制作业总量，作业基本在课堂或课后服务时间完成，未完成的作业带回家完成时间最多不超过30分钟。可以看出许多学校实施双减政策时限制了学生的作业时间，在升学的压力下，许多同学认为他们没有足够的时间分配学习任务。

3) 教培老师转行关键句

相关关键句有“听朋友说班主任和老师的工资都欠着不发了，他们是昨天晚上凌晨被离职了，企业微信全被删除了。”、“深圳市美博美智教育有限公司卷钱，跑路，将法人股权100%转给70岁老人至告不到，对方因为了解到司法不会管，政府各部门不管，可以不退钱。”王蓓蓓^[12]等双减政策下教培机构的转型发展研究策略表明，双减政策下教培机构面临上市融资渠道关闭、主营业务收入下降、获得市场更加困难、运用监管更加严格、内部人力资源供过于求的困境。可以看出在双减政策下教育培训行业十分不景气，教培老师失业的问题不只在如何复工，在双减政策下许多校外补课机构不能正常营业，甚至出现了部分培训机构直接“卷钱跑路”的社会乱象。

4) 疫情开学关键句

相关关键句有“高三学生到现在都没法返校怎么办？”、“高三走读这个学期到底放不放，不放也就说一声，外面租的房子白白空几个月还要交房租，不放也说一声好让我们提前把房子退了。”在双减与疫情的大环境下，学生上课的教学时间变少，导致学生的学业压力和升学压力加大，所以有很多学生和家针对疫情学校迟迟不能开学的问题产生了抱怨。

(二) 积极主题

1) 教育公平关键句

相关关键句有“公立学校也有尽力在做啦，公办民办本质上都是非营利性的，是公益性的……而且现在上学摇号(珠三角有些城市民办也开始摇号了)。”、“教师轮岗制度，优秀教师不会一直在一所学校，这都是国家为了教育公平推出的政策，虽然说要普及这些政策还是需要时间，但是大趋势还是尽力追求公平的。”由此可以看出双减政策的出台呈现出了致力于缩小学校教育的差异化使得教育更加公平，这可以极大地促进教育的公平化和避免教育内卷化的产生，使得教育良性发展。魏继宗^[13]等对双减政策促进教育公平的价值意蕴研究中也表明，双减政策通过对校外培训的规范治理、重构学校教育的主体地位、提供优质的课后服务资源等措施，消除校外教育无序竞争、家庭经济差异造成的教育不公平。

2) 教师待遇关键词

“大大提高学校在职老师待遇, 壮大教师队伍, 提高教师业务能力, 同时建立完善的透明公平的举报投诉系统, 让教育真正回归学校。”可以看出, 许多家长认为发展教育业的关键在于提高教师的薪资和社会地位, 并且督察系统的完善对于整治各种补课等教育乱象起着至关重要的作用。

3) 家长陪伴关键词

“这样也好, 每天有陪伴孩子的时间, 宝宝也高兴。”、“家长的陪伴真的是特别的重要啊, 给孩子一个好的成长规划。”可以看出, 双减政策下孩子有更多的时间与家人在一起, 同时家长们也更加意识到陪伴对孩子人生轨迹的重要影响。

4) 学业减负关键词

“双减, 减轻了家长的精神负担和家庭的经济负担, 让学科教育重新回归学校主阵地。”可以看出双减政策的落实能有效减轻学生的学业负担, 也能有效缓解家庭的经济负担。

5) 兴趣培养关键词

“双减之后可以给孩子提供更多的兴趣培养。”可以看出, 学生们的课余时间的增加, 使家长更加积极地支持孩子用更多时间去培养和支持他们的兴趣。

6) 人才培养关键词

“让教育回归校园才能通过中高考筛选出真正的人才, 而非在内卷浪潮里挣扎下来的普通人。”“双减”的系统性改革为创新型人才的培养提供了前所未有的变革契机, 它引发对于基础教育人才培养命题的深层思考, 也促进了教育生态优化的正向改变^[14]。教育行业的问题不在于教学质量的不足, 而是高等教育资源的不足和社会对高等岗位需求的不足。提高职高、中专和大专的教学质量、社会地位与职位需求、提高社会生产力、提高人民对工业制品的消费能力, 才是缓解学生内卷现象的关键。

7) 教育改革关键词

“真正的让孩子全面发展, 这个政策我太喜欢了。”在双减政策下孩子有更多的时间和兴趣学习艺术科目, 才能进行体能、艺术、思想的全面发展。我国从建国到现在经历了经济、科技、生产力的复兴, 而国民文化的全面发展也尤为重要, 国民文化的发展不仅能提高国民的民族认同感, 还能增加人民生活的幸福感。文化发展, 是从古至今每一个文明在发展中都不可或缺的必要工具, 这同样是一项重要且长远的发展道路。

8) 打击培训机构关键词

“家长辛苦了一年, 工资都用来交补课费了。为了让家长的腰包鼓起来, 国家采取打击校外培训班的方式, 这更有利于教育公平!”可以看出, 双减政策实行之前, 培训机构机构大行其道, 掏空了普通家长的腰包, 双减政策实施后, 有效打击了机构补课的乱象, 给学生减负的同时给家长减负。

9) 丰富生活关键词

“这个政策太好了, 可以看到学生们丰富多彩的课后生活。”、“不能及就不能及, 这有啥, 结果是好的就行了, 中国那么多好大学, 非要上清华北大吗, 多少孩子专科、二本三本, 不照样在好好生活!”可以看出, 相比在学业上的成就, 许多家长也十分关注孩子的课余生活是否丰富, 在双减政策实施后, 课余时间的增加, 使丰富学生的课余生活这项任务有了更多可能。

5.3. 词云图

经过分词筛选去停用词并统计词频后, 绘制的词云图如图 4 所示。不难看出, 在围绕以“双减政策”为主题的评论中, 词频最高的词语以“孩子”为首, 同时还出现了如“学生”、“学校”、“家长”、“老师”和“机构”等高词频的词汇, 这反映了在双减政策的实施影响下, 广大人民群众密切关注双减

-
- [5] 周清清, 章成志. 基于迁移学习微博情绪分类研究——以 H7N9 微博为例[J]. 情报学报, 2016, 35(4): 339-348.
- [6] 王睿哲, 洪建中, 刘静君. H7N9 禽流感疫情首周微博研究——基于内容分析法[C]//第二届湖北省青年心理学人论坛论文集. 武汉: 第二届湖北省青年心理学人论坛组委会, 2013: 19.
- [7] 张小平, 周雪忠, 黄厚宽, 等. 基于词相似性与 CRP 的主题模型[J]. 模式识别与人工智能, 2010(1): 72-76.
- [8] 李金广. 数据挖掘中聚类算法研究综述[J]. 中国科技信息, 2010(17): 48-49.
- [9] 张连文, 袁世宏. 隐结构模型与中医辨证研究(I)——隐结构法的基本思想以及隐结构分析工具[J]. 北京中医药大学学报, 2006, 29(6): 365-369.
- [10] 张小平. 主题模型及其在中医临床诊疗中的应用研究[D]: [博士学位论文]. 北京: 北京交通大学, 2011.
- [11] 王治峰. 中小学落实“双减”政策的实践路径研究——以甘肃省兰州市 X 学校为例[J]. 西北成人教育学院学报, 2022(2): 96-99.
- [12] 王蓓蓓, 王静怡. “双减”政策下教培机构的转型发展策略探析[J]. 中外企业文化, 2021(11): 194-195.
- [13] 魏继宗, 吴璇, 韩威. “双减”政策促进义务教育公平的价值意蕴与路径探析[J]. 成都师范学院学报, 2022, 38(4): 9-14.
- [14] 李晓雅. “双减”: 基础教育阶段创新型人才培养的新契机[J]. 广西师范大学学报(哲学社会科学版), 2021, 57(6): 107-115.