

运用社交媒体数据评估公共政策满意度： 方法构建与基于“双减”政策的应用

杨代福，刘 爽

重庆大学公共管理学院，重庆市公共经济与公共政策研究中心，重庆

收稿日期：2024年11月5日；录用日期：2025年1月19日；发布日期：2025年1月28日

摘 要

大数据时代构建一套运用社交媒体数据评估公共政策满意度的方法具有重要意义。文章从社交媒体数据与政策评估融合的角度，提出了运用社交媒体数据评估公共政策满意度的逻辑、维度与相应的方法和技术。其逻辑在于明确社交媒体数据和政策评估的共同性、互补性的基础上将社交媒体的优势、方法、技术与政策评估相整合，整合的维度包括政策后果话题识别、政策后果关注度、政策满意度、政策满意度影响因素四个方面，文本聚类、话题热度分析、情绪分析、基于情感极性的主题建模等有助于相关评估维度的实现。对C省份“双减”政策实施公众满意度评估的实践表明，该方法具有较强的优势和可行性。文章对大数据时代公共政策评估的创新进行了探索，并为“双减”政策实施的优化提供了一定启示。

关键词

社交媒体数据，政策满意度，政策评估，“双减”政策

Using Social Media Data to Evaluate Policy Satisfaction: Methodology Construction and Application of “Double Reduction” Policies

Daifu Yang, Shuang Liu

Public Economy and Public Policy Research Center, School of Public Policy and Administration, Chongqing University, Chongqing

Received: Nov. 5th, 2024; accepted: Jan. 19th, 2025; published: Jan. 28th, 2025

Abstract

In the era of big data, it is of great significance to construct a set of methods for evaluating policy

文章引用：杨代福，刘爽. 运用社交媒体数据评估公共政策满意度：方法构建与基于“双减”政策的应用[J]. 社会科学前沿, 2025, 14(1): 493-505. DOI: 10.12677/ass.2025.141062

satisfaction using social media data. From the perspective of the integration of social media data and policy evaluation, this paper puts forward the logic, dimension and corresponding methods of using social media data to evaluate public policy satisfaction. Its logic is to integrate the advantages, methods and technologies of social media with policy evaluation on the basis of clarifying the commonality and complementarity of social media data and policy evaluation. The integration dimension includes four aspects: topic identification of policy consequences, attention to policy consequences, policy satisfaction and influencing factors of policy satisfaction. Text clustering, topic heat analysis, sentiment analysis and topic modeling based on emotion polarity are helpful to the realization of relevant evaluation dimensions. The practice of public satisfaction evaluation on the implementation of “double reduction” policy in C province shows that this method has strong advantages and feasibility. This paper explores the innovation of policy evaluation in the era of big data, and provides some enlightenment for the optimization of the implementation of “double reduction” policy.

Keywords

Social Media Data, Policy Satisfaction, Policy Evaluation, “Double Reduction” Policy

Copyright © 2025 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

政策满意度评估是政策效果认知与检验、政策优化和改进的重要手段,然而,一些传统的评估方式正受到人们的批评和质疑。例如,大多数政策评估都是通过发送给目标群体的问卷调查或访谈来进行,但运用的问卷和访谈等评估方法存在着抽样困难、回答率低、回答真实程度不高的问题,制约了评估的信度与效度;又如,当前的一些评估往往依据“目标达成模式”进行,而对政策目标之外的效果缺乏考察,导致评估不够全面;再如,当前的评估往往只重视评估“政策有无效果或政策的效果是什么”,而缺乏对“政策为何有效或无效”进行回答,致使不能为政策的优化和改进提供充分、深入的证据。随着大数据时代的来临,社交媒体数据不断生成和累积。作为一种“社会传感器”,社交媒体数据与政策满意度评估拥有天然的结合优势。其一,广泛性。绝大多数公众都使用社交媒体,并可以发表对政策的评论和意见,因此数据体量巨大;而且,社交媒体数据的样本量通常远大于调查数据的样本量[1]。其二,实时性。社交媒体数据是实时生成的,用户可以全天候提供评论和意见,提高了“事件信息的创建、传播和评论的速度”[2],因此便于对政策进行实时性评估。其三,互动性。基于 web2.0 技术的社交媒体确保了互动性,允许参与其中的公众进行对话和交流,充分地发表对政策的意见和看法,这提升了评估的参与性和“建构性”。其四,真实性。社交媒体的匿名性、异步性和连接性促进用户可以更真实表达对政策的看法[3],增强了服务对象评价的真实性。其五,低成本。通过社交媒体平台公开 API 或网络爬虫等方法可以较为方便地采集社交媒体数据,全球定位系统的普及使电子设备中的(GPS)功能还可以收集带有地理标签的相应评论。由于这些特点,社交媒体数据可以与政策评估进行良好的融合,具有改善和优化传统政策评估方法的巨大潜力。因此,基于这些特点,构建一套运用社交媒体数据评估公共政策满意度的方法以改进传统方法具有重要的意义。

国内外学界目前已对运用社交媒体数据开展政策满意度评估进行了一些探索。在国外,意大利学者使用社交媒体进行消费者情绪分析来评估公共资助项目的执行程度及成效;联合国和印度尼西亚利用社

交媒体评估传染病控制效果; Fabra-Mata 等运用 Twitter 数据评估了挪威对哥伦比亚和平进程的贡献[4], Deborah 等运用 Twitter 数据评估了意大利某地区大学公共服务的绩效[2]。在国内, 孙玖陌等利用微博数据对我国政府工作满意度进行了评估[5]; 王建冬等综合运用微博、论坛、播客等数据对我国“双创”政策、供给侧改革、数字经济等政策的推进成效进行了评估[6]; Jiang 等运用微博数据评估了三峡工程的公众满意度[1]。这些研究为运用社交媒体数据开展政策评估进行有益的探索, 但目前仍然存在以下局限。一是相关研究绝大多数都是针对公共服务、公共项目、政府工作进行的评估, 没有专门聚焦公共政策的评估; 二是虽然相关研究都使用了一些社交媒体数据收集、分析方法, 但这些方法没有与政策评估的逻辑和程序进行很好的融合。三是已有研究对政策的评估基本停留在“政策是否有效”层面, 而忽视了对评估结果影响因素的挖掘, 导致评估缺乏全面性。总的来说, 目前国内外学界还没有形成一套专门的、全面的、系统的运用社交媒体数据评估公共政策满意度的方法。有鉴于此, 本文试图构建一套专门的、较为全面系统地运用社交媒体数据对公共政策满意度进行评估的方法, 并以 C 省份“双减”政策评估为案例检验该方法的有效性和可操作性, 最后反思了该方法的贡献与局限以及未来的研究方向。

2. 运用社交媒体数据评估公共政策满意度的方法构建

2.1. 运用社交媒体数据评估公共政策的逻辑与维度

本文运用大数据进行政策评估的核心逻辑是在明确大数据和政策评估的共同性、互补性的基础上将大数据的优势、方法、技术与政策评估相结合、整合或融合, 更好地实现政策评估的目的[7]。这一逻辑同样也适合运用社交媒体数据进行政策评估。从二者的共同性和互补性看, 二者都以数据为前提, 数据的处理过程都是一致的, 都经过数据的收集与分析, 这为二者的融合奠定了基础。从二者融合的目的来看, 可以实现以下重要评估目的: 其一, 可以从数据中识别出公众讨论了哪些方面的政策后果。从政策评估来看, 首先需要明确评估的指标, 即要明确评估哪些方面的后果。在传统的政策评估中, 往往通过政策目标分解、“政策变化理论”、逻辑模型来建构评估指标。而在社交媒体数据中, 在大型语料库中通过话题抽取, 可以识别出政策实施中或结束后公众关注哪些有关“后果”的话题, 这起到了类似于建构评估指标的作用。其二, 可以评估各个方面的后果的绩效。从社交媒体数据的特性、方法与技术上看, 特别适用于分析两类绩效: 各种后果的公众关注度和公众满意度。通过社交媒体数据中的话题讨论热度分析, 可以评估公众对各种后果的关注度, 判断哪些后果更受公众关注和讨论。通过对各种后果话题的情感强度分析, 可以评估公众对各种后果的满意度, 以及通过对所有评论文本的情感强度的加总评估公众对政策整体满意度。其三, 可以分析影响公众政策满意度的因素。尽管大数据“不追求因果关系而只关注相关性”, 但是仍然可以通过话题的深入挖掘、语义的关系分析、扎根理论的思维来发掘影响政策效果的某些因由。基于以上分析, 可以形成运用社交媒体数据评估公共政策的四个维度: 政策后果话题识别、政策后果关注度、政策满意度、政策满意度影响因素。

2.2. 运用社交媒体数据评估公共政策的方法

以上逻辑与维度的实现还需要具体的方法和技术支撑, 否则将只能是某种理论设想。以下提出针对以上各评估维度的、可能的数据收集和分析方法。

2.2.1. 评估数据收集方法

首先, 依据评估对象和目的、数据可获得性、平台活跃性、评估对象覆盖度, 确定社交平台进行数据收集。在此基础上, 确定内容形式、时间范围、地理位置, 选取收集方法, 以此保证数据的规模和质量。常用的数据收集方法包括 API 调用、Web 抓取、数据库查询和通过第三方数据获取平台进行采集。在数据采集完毕后, 需要对数据进行预处理。由于社交媒体数据的特性, 往往需要进行以下预处理: 一

是根据位置信息, 排除非本地的数据; 二是由于社交媒体数据中可能存在博流量、炒热点、打广告等情形, 可以采用关键词匹配算法或核查事实的工具对相关的词汇进行识别和过滤处理; 三是由于社交媒体数据具有一定的杂乱性, 需要运用去重、去短、去空白、去无关、中文分词、去停用词等方法对数据进行清洗。

2.2.2. 评估数据分析方法

(1) 政策后果的话题识别方法。政策后果的话题识别主要通过文本挖掘, 依赖于对文本内容的深入分析和处理。一是关键词提取, 从文本中自动识别并提取出最能够代表文本主题的词汇或短语与频率分析。二是频率分析, 对文本中词汇或短语的出现频率进行统计, 将高频词汇视为话题关键词。三是深度学习方法, 如循环神经网络(RNN)、卷积神经网络(CNN), 能够捕捉文本中的复杂语义关系, 实现更精确的主题识别。四是文本聚类, 适用于无监督学习场景, 能够自动发现文本数据中的主题分布, 被广泛应用于学术研究中。文本聚类等方法可以在大型语料库中识别出潜在话题[6]。其对围绕政策所表述的各种观点、诉求、建议等信息进行话题抽取, 识别政策实施后公众关注哪些话题。首先, 提取聚类关键词。TF-IDF是在信息检索与数据挖掘方面常用的一种加权方法, 一定程度上能够有效表示某个特征词在文本集或者整个语料库中的重要程度。其次, 训练政策语料库。Word2vec模型通过将所有词语向量化, 可以对词与词之间关系进行定量化度量, 从而挖掘词语之间的联系。最后, 进行文本聚类。聚类指将抽象对象或者物理对象分成相似的对象集合的过程。K-Means算法是文本聚类常用的算法之一, 是一种无监督的机器学习算法, 基本原理是根据样本之间的距离大小, 将样本集分为K个簇, 距离越小代表相似度越大。让簇内的点尽量紧密地连在一起, 而让簇间距离尽量大, 以此实现聚类目的。由此, 能够识别出政策后果的讨论话题。

(2) 政策关注度分析方法。政策关注度可以通过话题热度来进行分析, 而话题热度的衡量方法包括帖子强度和热度算法。帖子强度即在特定时间段和/或特定区域内发布的与特定话题相关的消息的数量。以往的研究表明, 社交媒体上的帖子强度直接反映了该话题的流行程度[1]。热度算法考虑多个因素, 如点击量、分享量、点赞量等, 通过综合计算得出一个热度值, 反映用户对政策及其相关后果的关注度。

(3) 公众政策满意度分析方法。讨论话题的情感强度往往运用情感分析方法、实体识别与关键词提取等方法进行评估。在识别话题满意度时, 实体识别和关键词提取可以帮助识别出具有特定意义的实体和分析文本数据中的关键信息, 通过分析这些实体和关键词的情感倾向, 可以了解公众对这些方面的满意度情况, 通常采用条件随机场(CRF)、神经网络等实现。情感分析的基本目的是从文本中识别用户观点, 分析其情感倾向(积极、消极或中立)。当前, 基于情感词典的方法、基于传统机器学习的方法、基于深度学习的方法是常用的三种文本情感分析方法。基于情感词典的情感分析是先制定一套情感词典与规则, 再对文本、段落、句子进行拆解、分析其内含的情感词所属, 然后判断出该文本的情感倾向。基于机器学习的情感分析方法是指通过大量有标注的或无标注的语料, 使用统计机器学习算法, 抽取特征, 最后再进行情感分析输出结果。基于深度学习的情感分析包括单一神经网络的情感分析方法、混合神经网络的情感分析方法、引入注意力机制的情感分析方法和使用预训练模型的情感分析方法。应根据以上方法的优劣和适用情形对以上方法进行恰当选择。依据以上情感分析方法, 首先可以计算出各个话题的情感强度; 其次通过计算整体的平均情感得分, 可以获得公众对政策的整体情感强度。

(4) 政策效果影响因素分析方法。社交媒体数据通常是非结构性数据, 对非结构性数据常用的影响因素挖掘方法有三种。一是特征选择与处理, 特征选择是指从原始文本数据中通过文本清洗、词干提取等步骤挑选出对分析结果最有影响力的特征。二是基于情感极性的主题建模方法, 其是根据得到的数据, 去寻找出这篇数据的主题, 以及这些主题所对应的词。在政策评估中, 可以将整个数据分为非负面语料

库(政策满意度高)、负面语料库(政策满意度不高)两类语料,然后在两类语料库中分别进行主题挖掘,由此识别政策满意度高或政策满意度不高的因由或因素。三是语义网络分析方法,其通常是输出一个带有箭头标示的有向图,这种表示法可以把各种事物有机地联系起来,反映事物之间的相关性。在实际评估中,可以对以上方法综合使用,例如以情感极性的主题建模方法为主,辅之以义网络分析,使政策满意度影响因素的挖掘更为可靠。基于以上,本文构建出了运用社交媒体数据评估公共政策满意度的框架和方法,如图1所示。

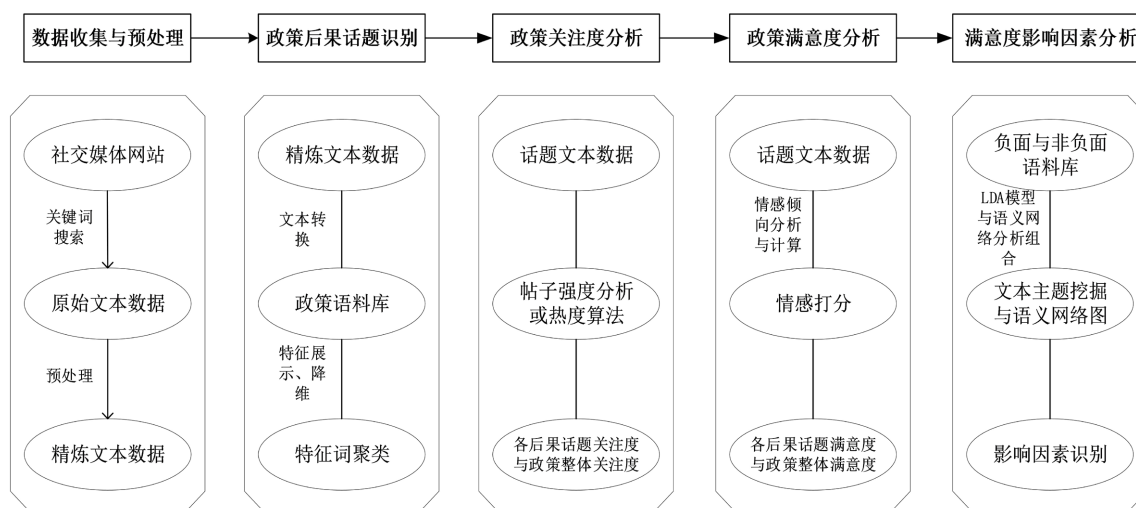


Figure 1. Methodology construction for evaluating public policy satisfaction using social media data

图 1. 运用社交媒体数据评估公共政策满意度的方法构建

3. 方法运用：C 省份“双减”政策满意度评估

为了检验上述方法的可操作性和有效性,本文选择 C 省份“双减”政策实施为案例,运用上述方法对其进行实证评估。C 省份位于我国西部地区,其积极推动“双减”政策的实施,多项实施举措先后入选教育部遴选的落实“双减”典型案例并获得推广,但也存在一些问题,在我国“双减”政策实施中具有较强的典型性¹。

3.1. 评估过程

首先,进行数据收集和预处理。选定抖音、微博、微信、知乎、C 省本地宝、C 省问政网作为主要数据来源;将时间设定为 2021 年 7 月 20 日至 2023 年 12 月 20 日,关键词设为“双减”、“双减政策”。运用 python 中的 requests 包和第三方采集器进行数据采集,除了采集每个消息的正文,还采集了时间戳和位置。然后对收集到的所有数据进行整理。一是依据每个信息的位置信息,对位置不在 C 省的信息进行删除;二是利用关键词匹配算法和核查事实的工具 Hoaxy,识别出与博流量、炒热点、打广告相关的词汇(如热门、必看、广告、转发抽奖),对其所在的文本段落进行过滤处理。三是通过数据合并、数据清洗(去重、去短、去空白、去无关)、中文分词、去停用词等对原始数据进行清洗。最终本文共采集符合搜索条件的信息 272,430 条。

其次,进行政策后果的话题识别。采用 TF-IDF 提取聚类特征词,训练 C 直辖市“双减”政策语料库,运用 K-Means 算法对降维后的特征词进行聚类。共提取了 149 个聚类高频关键词,然后进行文本向

¹C 省份位于我国西部地区,其积极推动“双减”政策的实施,多项实施举措先后入选教育部遴选的落实“双减”典型案例并获得推广,但也存在一些问题,在我国“双减”政策实施中具有较强的典型性。

量化, 再通过手肘法确定初始质心数 $k = 8$ 后进行聚类, 获得话题聚类结果。

再次, 进行政策关注度分析。通过计算各个话题的绝对帖子量和相对帖子量, 评估公众对各个话题的关注度。

从次, 进行政策满意度分析。运用百度 AI 开放平台的情感倾向分析功能, 输入清洗后的 C 省份“双减”政策相关评论文本, 并抽样完成有效性测试。以情感值 0.5 为分界点, 判定文本情感是偏向积极还是消极, 同时去除置信度低于 0.7 的文本数据, 以提升情感打分的准确度。然后分析各个话题的情感强度。按话题对已爬取的政策相关评论文本的情感得分进行整理和计算, 获得公众对各话题的情感强度。最后对所有的评论文本的情感得分进行综合计算, 获得公众对“双减”政策的整体情感强度。

最后, 进行政策效果影响因素分析。基于前述情感分类的基础上, 将评论文本分为负面和非负面语料库, 分别引入 LDA 模型进行文本主题挖掘, 同时辅助语义网络关系图来识别影响“双减”政策效果的主要因素。最终, 通过对多次运行、对比实验和人工编码相结合的方式, 进行结果稳定性和有效性验证。

3.2. 结果

3.2.1. “双减”政策后果的话题

表 1 显示了公众对“双减”政策后果的讨论话题。可以发现, 公众对于“双减”政策后果的话题讨论主要围绕政策实施后学生校内负担减轻情况、校外培训机构减少程度、校外培训机构规范性、家长经济负担、家长精力负担与焦虑、学校教学质量、教师教学压力、学生课后学习与发展情况 8 个话题展开。说明公众主要关注“双减”政策对以上八个方面产生的影响。这些话题也涵盖了对学生、家长、校外培训机构与学校及教师的影响。而从与政策目标的联系来看, 这些话题也与“双减”政策第一、二阶段的核心目标(有效或显著减轻学生过重作业负担和校外培训负担, 有效或显著减轻家庭教育支出和家长精力负担)紧密相关, 当然, 还涉及到政策的附带效果——对学生课后学习与发展的影响以及对教师工作压力的影响。

Table 1. Discussion topics on the consequences of the “Double Reduction” policy
表 1. “双减”政策后果的讨论话题

标签	话题	关键词
01	政策实施后学生校内学习负担	学生 学习 时间 延迟 晚自习 做作业 作业量 更累 减了 轻松 期末 难度 成绩 考试 跟不上 分数 公布 排名 等级 差距 学霸
02	政策实施后校外培训机构减少程度	学科 取缔 奥数 学而思 新东方 小区 家教 寒暑假 违规 补课 有钱 一对一 线上 网课 监管
03	政策实施后校外培训机构规范性	举报 艺术类 文化课 校外 培训班 线下 机构 退费 收费 费用 交钱 补课费 涨价 花钱 维权 托管
04	政策实施后家长经济负担	家长 教育 有钱人 两个 学费 太贵 工人 压力 家庭 现实 条件 上班 自愿 变相 强制 课外 免费 收钱
05	政策实施后家长精力负担与焦虑	读书 培养 优秀 孩子 焦虑 担心 职高 大学 分流 高考 中考 竞争 长大 群里 签字 家长会 App 手抄报 家庭作业 微信 打卡
06	政策实施后学校教学质量	学校 老师 布置 课堂 讲课 试卷 批改作业 专业 教学质量 教师 农村 教学内容 投诉 公立学校 私立学校 水平 疫情
07	政策实施后教师教学压力	太难 假期 待遇 绩效 佛系 检查 作业 辅导 负责 职业 责任心 班主任 班级 提高 收入 一线 辛苦 辞职 考核 工资 加班
08	政策实施后学生课后学习与发展	美术 书法 兴趣班 体育 音乐 手工 课后 延时 服务 服务费 开心 快乐 周末 游戏 看电视 放假 回家 休息 睡觉 手机

3.2.2. 相关后果话题的关注度

表 2 报告了相关后果话题的热度分析结果。政策实施后政策实施后校外培训机构减少程度是讨论最多的话题，占比达 22.3%。其次是对学生参加学校课后服务、课后兴趣培养、回家后的活动安排等学生课后学习与发展的影响的讨论，占比 18.7%。政策实施后家长精力负担与焦虑，帖子量占比 17%，受到较多讨论。政策实施后家长经济负担讨论热度同样较高，帖子量占比达 14.8%。这些话题居于公众讨论热度的前列，成为了公众关注的核心话题。对政策实施后学校教学质量、政策实施后学生校内学习负担、政策实施后教师教学压力、政策实施后校外培训机构规范性的讨论帖子量相对较少，分别占 9.3%、6.7%、5.9%、4.9%。公众对各个话题的关注程度的基本特点是更多的关注政策对校外或课后层面的影响，而对校内或课堂层面的影响的关注相对较少。

Table 2. Popularity of topics related to the “Double Reduction” policy
表 2. “双减”政策各话题讨论热度

标签	讨论话题	绝对帖子量(条)	相对帖子量
01	政策实施后校外培训机构减少程度	46,568	22.3%
02	政策实施后学生校内学习负担	14,017	6.7%
03	政策实施后校外培训机构规范性	10,176	4.9%
04	政策实施后学生课后学习与发展	39,048	18.7%
05	政策实施后家长经济负担	31,023	14.8%
06	政策实施后家长精力负担与焦虑	35,409	17%
07	政策实施后教师教学压力	12,303	5.9%
08	政策实施后学校教学质量	19,412	9.3%

3.2.3. 相关后果话题的关注度

Table 3. The overall emotion score of each topic and policy of “Double Reduction” policy
表 3. “双减”政策各话题与政策整体情感得分

维度	具体话题	积极得分	消极得分	置信度
讨论话题	政策实施后学生校内学习负担	0.60	0.40	91.0%
	政策实施后校外培训机构减少程度	0.38	0.62	91.0%
	政策实施后校外培训机构规范性	0.64	0.36	91.7%
	政策实施后家长经济负担	0.56	0.44	91.7%
	政策实施后家长精力负担与焦虑	0.40	0.60	91.7%
	政策实施后学校教学质量	0.52	0.48	90.8%
	政策实施后教师教学压力	0.31	0.69	91.7%
	政策实施后学生课后学习与发展	0.45	0.55	91.2%
政策整体	—	0.55	0.45	91.3%

表 3 显示各话题的情感强度与政策整体情感强度。该表表明，对 C 省份“双减”政策实施的情感打分均有较高置信度。从各讨论话题来看，政策实施后校外培训机构规范性(0.64)、学生校内学习负担(0.60)、家长经济负担(0.56)、学校教学质量(0.52)情感积极得分大于 0.5 分，情感偏向积极，说明公众对于这四个方面的政策后果较为认可。而政策实施后学生课后学习与发展(0.45)、家长精力负担与焦虑(0.40)、校外培训机构减少程度(0.38)、教师教学压力(0.31)情感积极得分均低于 0.5 分，情感偏向消极，说明公众对政策在此四个

方面的后果满意度不高。从政策整体上看, C 省份“双减”政策的情感得分为 0.55 分, 高于 0.5, 但得分并不高, 说明总体上公众对 C 省份“双减”政策的实施持积极态度, 但满意度不高, 尚有不少需要完善之处。

3.2.4. 相关后果话题的关注度

(1) 促进因素

运用 LDA 模型对非负面语料库进行主题挖掘的结果如表 4 所示, 基于非负面语料库绘制的语义网络图见图 2。结合两种结果, 可以发现公众对 C 省“双减”满意度较高的影响因素有三个:

其一, 主题 1 和语义网络关系图“全面→发展”、“人才→培养”反映政策方向的正确性促进了“双减”政策实施。如评论所说: “‘双减’的目的是培养德智体美劳全面发展的人才, 学而优则仕, 但是不是拖垮身体”, “双减”政策顺应当前教育发展规律, 有利于推进素质教育、培养高质量人才和促进教育公平。

其二, 主题 2 和语义网络关系图“市教委→学生”、“学校→学生”反映执行主体的贯彻促进了政策实施, 这包括各级教育行政部门、各个学校、教师的积极落实行动, 例如对校外培训机构的治理规范了培训结构的收费、运行等环节, 优化作业设计、减少考试次数、不公布成绩排名等减轻了学生的校内学习负担, 减少校外培训与相对较低的课后服务收费也一定程度上减轻了家长经济负担, 教师观念的一定转变、教学技能的一定提升以及一定的教学创新探索也提升了教学质量。

其三, 主题 3 体现了良好的政策监督促进了政策有效落地。一方面 C 省份建立了较为完善的政策监督机制, 如将“双减”政策落实成效纳入了政府履职督导评价, 将校外培训治理纳入区县的绩效考核指标体系以及畅通“双减”执行的外部监督, 如评论“督导组公布了举报电话可以打”。另一方面也设计了较为有效的监督方法, 如建立基础教育“双减”工作监测平台、对地方落实情况进行排名以及部门问责等。通过建立较为严密的监控机制和监督方法, 有利于发现政策执行的偏差并纠正偏差, 促进了“双减”政策的具体落实。

Table 4. Non-negative thematic emotion feature words

表 4. 非负面主题情感特征词

主题	特征词(部分)	主题归纳
P-1	素质教育 质量 人才 公平 评价	政策方向的正确性
P-2	教育部 案例 学校 市教委 执行	执行主体的贯彻
P-3	举报 督导 工作 监督 评价 绩效	良好的政策监督

(2) 阻碍因素

对负面评论语料库进行主题挖掘的结果如表 5 所示, 基于负面语料库的语义网络关系图见图 3。反映出 C 省份“双减”政策满意度不高的因素有以下八类:

其一, 主题 1 体现了政策本身的局限性影响政策的满意度。虽然政策的方向是正确的, 但是一些政策措施尚存在不足。如政策资源配备不够, 尤其是财政资金还不足; 一些家庭仍然抱怨课后服务收费较贵。文本中反映最多的问题在于“缺乏将来或未来的学生综合素质评价标准”, 现有的“双减”要求克服“唯分数、唯成绩”倾向, 改变以考试成绩为唯一标准评价学生的做法, 但是并未规定或提出合理的评价标准, 这给家长、教师或学生均带来了不确定性或茫然。

其二, 主题 2 和语义网络关系图“游戏→成绩”、“手机→成长”体现了家庭教育能力和学生自律性也影响政策的实施效果。由于中小学生自律性差以及家庭缺乏教育能力或对孩子的监管, 导致减负后的大量空闲时间被用于看电视、打游戏、玩手机等娱乐活动。这影响了学生的健康成长, 也影响了家长对政策的满意度。

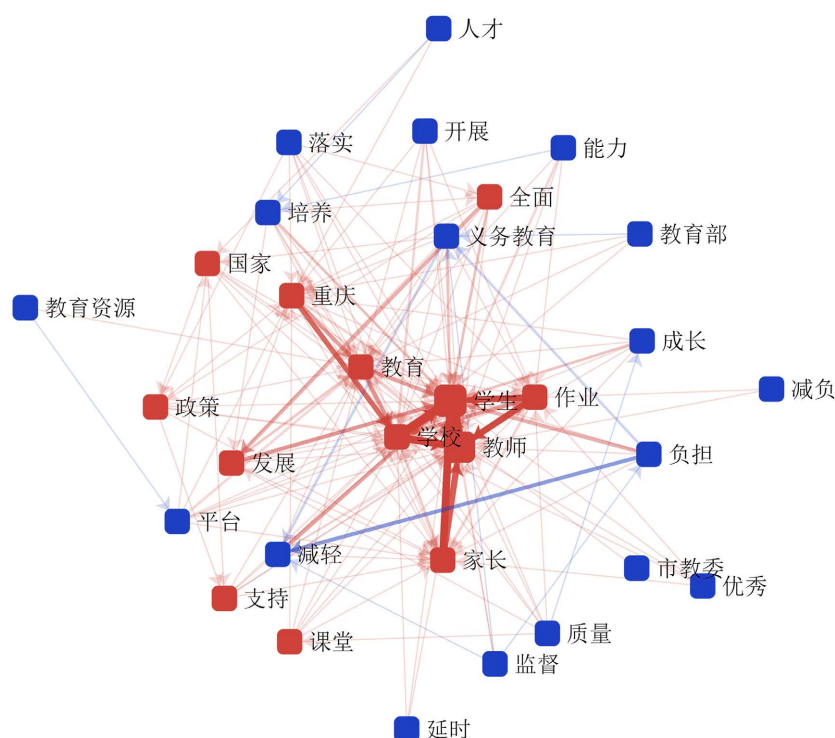


Figure 2. Semantic network diagram of the “Double Reduction” non-negative corpus
图 2. “双减”非负面语料库的语义网络图

其三，主题 3 和语义网络关系图“分数→辅导班”、“考试→培训”体现了考试和升学文化影响政策的实施效果。受到传统科举文化的影响以及长期的“唯分数”倾向，人们重成绩的观念根深蒂固，无法及时转变为促进学生全面发展的观念。这也导致家长和学生培训机构的依赖，甚至政策实施后家长仍然会不惜高价安排“一对一”家教为学生“增负”。

其四，主题 4 体现了校外培训机构的不当逐利影响政策的满意度。“双减”政策的实施对学科类培训机构进行了压减和规范治理，然而，由于以上家长和学生的“培训需求”仍然广泛存在，在逐利的动机下，“高端家政”、“众筹私教”、“住家教师”等隐形培训开始出现。由于相关部门对学科类校外培训的不间断查处和监管增加了培训供给的稀缺性，反而导致隐形培训的不间断增加甚至导致价格的高涨，以至于出现了“愈监管、愈增加、价格愈高”的怪圈，由此，校外培训治理成效大打折扣。

其五，主题 5 反映教师能力因素影响政策的有效实施。“双减”政策的实施要求教师转变教学理念、改革作业管理、改革课堂教学、改革考试管理、开展课后服务、善于家校沟通等，这对教师的能力也提出了全新的挑战，而且，社会对校内教育质量的追求和要求比以往更高。然而，习惯了传统教学模式和方式的教师目前尚不完全具备这些新的理念、素质和技能。比如有评论指出：“学校有些老师水平真的有限，还是在用老一套”。这制约了学校教学质量的提升、课后服务的开展，当然，也客观上增加教师的压力。

其六，主题 6 和语义网络图中的“中考→分数”、“分数→高考”反映当前的中考、高考等教育制度影响政策的有效实施。一方面，中考、高考仍然是通过考试成绩来选拔学生的，“成绩”这一指挥棒并没有发生改变；另一方面，中考分流制度、激烈的高考更是加剧了竞争。这种“上面不变”导致家长和学生难以认同和遵从仅仅“下面转变”的政策，因而补课和繁重的学习任务难以根本改变。

其七，主题 7 和语义网络关系图“水平→教师”、“农村→学校”体现教育不均衡问题影响政策实施效果。城市与农村、普通学校与重点学校的区分导致财政投入不均衡、办学规模不均衡、师资配备不

均衡、教师待遇及生活条件不均衡以及教学水平和质量不均衡。为取得一所更强学校的入学资格，各个家庭不得不再次投入和加压，繁重的学习也再一次挤进孩子们的世界。

其八，主题 8 体现当前激烈的社会竞争是阻碍政策有效落地的主要因素之一。除了中考、高考的竞争和进入更强学校的竞争，家庭和学生还面临着更宏观的社会激烈竞争。当前就业、收入、职称等无不面临着激烈竞争，这种竞争也传导至家庭和教育领域中。正如有评论指出的那样，“如果孩子学习不好，未来在社会竞争和阶层划分中就会处于不利位置”。这不仅会难以减轻家长的焦虑情绪，也难以弱化“成绩竞争”、“补课竞争”，从根本上制约了学生负担的减轻。

Table 5. Negative theme emotion feature words

表 5. 负面主题情感特征词

主题	特征词(部分)	主题归纳
N-1	教育资源 收费 太难 将来 未来	政策本身的局限性
N-2	学生 跟不上 游戏 看电视 休息	家庭教育能力和学生自律性
N-3	家长 焦虑 补课 升学 考试	考试升学文化
N-4	机构 退费 补课 小区	校外培训机构不当逐利
N-5	教师 教学水平 课后 服务 待遇	教师能力不足
N-6	高考 中考 职高 分数线	中考高考制度
N-7	有钱人 农村 普通 名校 城市	教育不均衡问题
N-8	社会 长大 技校 能力 竞争	社会激烈竞争

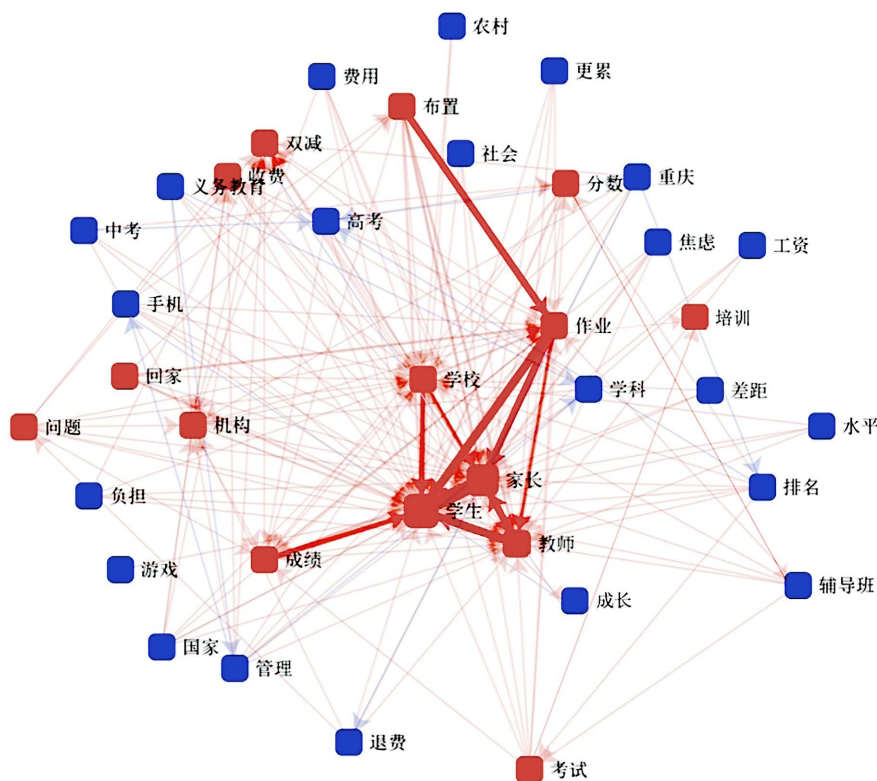


Figure 3. Semantic network diagram of the “Double Reduction” negative corpus

图 3. “双减”负面语料库的语义网络图

4. 讨论

通过 C 省份“双减”政策成效的实证评估,可以发现运用社交媒体数据进行政策评估的优势、可行性以及局限性。

4.1. 运用社交媒体数据进行政策评估的优势

一是运用社交媒体数据进行政策评估有利于更全面地评估政策的后果。以往的评估研究往往遵循的是一种“自上而下”的逻辑或者目标达成模式对政策进行评估,尽管这种评估逻辑具有重要价值,然而其对政策后果的考察仅仅局限于与政策目标相关的后果,而忽略了对政策产生的其它影响的考察。运用社交媒体数据的评估,则采用的是一种“自下而上”的逻辑或“数据驱动的知识生产模式”,即主要地依靠“数据来说话”,事先并不先入为主设定考察哪些后果,而是从数据中去识别公众讨论了哪些话题,最终获得了“最初不知道要问的问题的答案”[8]。从“目标达成”的角度看,本文表明学生校内学习负担、家长经济负担有所减轻,但校外培训负担、家长精力负担仍然不小。但除此之外,本文还发现了“双减”政策的其它的附带效果,如学校教学质量得到了改进,非学科类校外培训机构规范性得到加强,学生课后学习与发展出现新问题,教师的教学压力加大。运用社交媒体数据的评估不仅考察了与政策目标相关的后果,还考察了政策的附带后果,因而可以更全面地考察政策的后果。

二是有利于更准确的评估政策后果。在 C 省份“双减”政策实施的评估中,采集了有关政策后果的评论 272,430 条,这是采用问卷和访谈方法难以企及的样本量。而且,由于互联网的隐藏作用,发帖人员很少承受表达的群体压力和社交压力,这些评论数据很好地反映了相关公众对“双减”政策的“真实情感”。从 C 省份“双减”政策实施的评估结果看,相较于运用传统评估方法评估获得结论[9],此次评估的消极情感略高,这一定程度上说明此次评估更好地捕获了公众的真实看法,从而提升了评估结果的准确度。

三是增强了评估的深入性。本文运用 LDA 模型和语义网络图等方法对社交媒体数据进行挖掘,获得了影响“双减”满意度因由的认识。这使得政策评估不再停留在获知“效果如何”层面,还提供了更深入的、有关“双减”政策“为什么有效或无效”的洞见。对 C 省份的评估结果表明,政策方向的正确性、执行主体的落实行动和良好的政策监督促进了双减政策正面效果的出现,而政策环境因素、政策本身的局限性、执行主体因素和目标群体的特性则导致公众满意度不高。这说明,“双减”政策无论在政策指向上和政策导向上均是正确的、符合教育发展规律的,政策执行人员也进行了积极的贯彻。然而“双减”政策的实施目前还存在四大误区。一是忽略了政策环境的影响。现有政策尽管制定了各种具体细则,但是对当前我国社会竞争压力大、教育不均衡问题、考试和升学文化传统的约束考虑不够。二是改革缺乏系统性。在中考制度、高考制度甚至大学教育制度没有开展“基于学生的健康成长和全面发展”的改革的前提下,仅对基础教育这个子系统进行改革,会导致改革缺乏联动性。三是未能抓住改革的“牛鼻子”。在“双减”中,形成“德智体美劳全面发展”的观念与行为依赖于正确的、新的学生素质评价标准的建立。然而,现有政策并未这一“指挥棒”进行改革,导致家长、学生、教师等的观念和行为亦只能照旧,甚至在政策强制减负与评价标准依旧中,产生茫然与焦虑,减少对政策的认同甚至产生违背行为。四是政策实施理想化。政策有效实施有赖于充足的财政资源、教师的适应能力、家长开展家庭教育的能力、较强的学生自律性以及校外培训机构的配合。然而,现有这些条件都还不具备或不充分。总体来说,只有把“双减”政策实施视为一场社会改革,一场教育制度的系统改革,一种学生评价标准的改革,一种家长、学生、教师、校外培训机构观念、能力、行为的改革,政策才能得到真正有效地实施。这也说明,运用社交媒体数据进行政策评估可以获得政策运行的更深入的认识。

4.2. 运用社交媒体数据评估公共政策的可行性

通过实践的验证,可以发现该方法具有很强的可操作性。首先评估数据的收集、分析和挖掘,都具有技术上的可行性。抖音、微信、微博、知乎、地方问政网均是受不同群体关注和使用的平台,数据均对外开放、覆盖的人群广,而且近几年的数据都可以提取,这保证了数据的可获得性和丰富性。社交媒体数据可以实时进行采集和分析,能够帮助随时随地掌握一手资料,为政策评估提供及时有效的数据支撑;数据收集和分析依托互联网和互联网终端,借助爬虫工具、代码块和情感分析等工具进行,这些方法较易掌握。此外,整个评估过程对人力、物力和财力的要求不高,评估的成本也较低。

4.3. 运用社交媒体数据评估公共政策的局限性

实践的验证也揭示了目前的方法尚存一些局限。一是数据的非代表性问题。社交媒体数据可能无意中排除了部分人群^[10],在本文中由于中小学生对使用前述社交媒体并发表评论,因此评估缺少学生群体的声音,导致数据的覆盖性受影响。此外还可能存在数据噪声问题,社交媒体上存在某些营销号或个人会故意发布一些不实信息或带有误导性信息来达到某种目的,尽管本文采用了一定方法进行过滤,但这些数据难以全部清除。另外,对影响因素的分析还需要引入和开发更科学的大数据方法,以提升影响因素识别的可靠性。

5. 结论与展望

本文基于对社交媒体数据的特点和政策评估的逻辑和程序,构建了一套运用社交媒体数据对公共政策满意度进行评估的方法,然后以C省份“双减”政策为案例,进行了方法的实际运用。方法构建与实证运用的交互表明,社交媒体数据在政策评估中具有诸多的优势和可行性,但也存在一定的局限性。从其优势和可行性而言,本文构建的方法具有较强的应用价值和推广价值。从其局限性而言,未来的方法构建可以从以下方面进行进一步优化:一是纳入传统定性分析、定量分析进行补充,提升数据的覆盖面,解决部分利益群体的声音缺失问题。二是引入更科学的核查事实工具,进一步区分和排除数据噪声。三是引入更先进的影响因素分析方法,以使影响因素识别更为准确。四是从二者融合的角度,探索更多的评估维度,以使评估结果更为全面和深入。

基金项目

重庆市教委人文社科重点研究基地项目“基于混合方法的‘双减’政策绩效评估与作用机制研究”(22SKJD003);重庆市社科规划项目“运用大数据进行政策影响评估的理论与方法研究”(2023NDYB87);国家社科基金后期资助项目“西方政策评估理论与方法研究”(20FGLB043)。

参考文献

- [1] Jiang, H., Qiang, M. and Lin, P. (2016) Assessment of Online Public Opinions on Large Infrastructure Projects: A Case Study of the Three Gorges Project in China. *Environmental Impact Assessment Review*, **61**, 38-51. <https://doi.org/10.1016/j.eiar.2016.06.004>
- [2] Agostino, D. and Arnaboldi, M. (2016) Social Media Data Used in the Measurement of Public Services Effectiveness: Empirical Evidence from Twitter in Higher Education Institutions. *Public Policy and Administration*, **32**, 296-322. <https://doi.org/10.1177/0952076716682369>
- [3] Hu, C., Kumar, S., Huang, J. and Ratnavelu, K. (2019) The Expression of the True Self in the Online World: A Literature Review. *Behaviour & Information Technology*, **40**, 271-281. <https://doi.org/10.1080/0144929x.2019.1685596>
- [4] Fabra-Mata, J. and Mygind, J. (2018) Big Data in Evaluation: Experiences from Using Twitter Analysis to Evaluate Norway's Contribution to the Peace Process in Colombia. *Evaluation*, **25**, 6-22. <https://doi.org/10.1177/1356389018804259>

-
- [5] 孙玖陌, 张滨熠, 朱廷劭. 基于社交媒体大数据的政府满意度评估[J]. 中国党政干部论坛, 2017(2): 75-77.
- [6] 王建冬, 童楠楠, 易成岐. 大数据时代公共政策评估的变革: 理论、方法与实践[M]. 北京: 社会科学文献出版社, 2019.
- [7] 杨代福. 西方政策评估理论与方法研究[M]. 杭州: 浙江大学出版社, 2023.
- [8] Lemire, S. and Petersson, G.J. (2017) Big Bang or Big Bust? The Role and Implications of Big Data in Evaluation. In: *Cyber Society, Big Data, and Evaluation*, Routledge, 215-236. <https://doi.org/10.4324/9780203793909-12>
- [9] 北京师范大学教育国情调查中心全国‘双减’成效调查课题组. 全国“双减”成效调查报告[R]. 2022.
- [10] Rathinam, F., Khatua, S., Siddiqui, Z., Malik, M., Duggal, P., Watson, S., *et al.* (2021) Using Big Data for Evaluating Development Outcomes: A Systematic Map. *Campbell Systematic Reviews*, **17**, e1149. <https://doi.org/10.1002/cl2.1149>