

大数据思维下诚信建设网络关注度研究

——基于浙江网民 2014-2020 年

百度搜索指数的动态分析

赵蓓 毛通 王佳华¹

(浙江金融职业学院, 浙江 杭州 310018)

【摘要】: 公众对社会信用体系建设的关注程度是公众信用意识最直观有效的体现。本文依托百度指数和互联网大数据手段, 以 301 个社会信用热点关键词为统计依据, 以浙江省 11 个地市为实证研究对象, 通过对 2014-2020 年间浙江网民 2.34 亿次网络搜索行为特征的分析对比, 动态刻画公众对社会信用体系建设网络关注度的时空变化规律。

【关键词】: 诚信建设 网络关注度 百度指数

【中图分类号】:F27 **【文献标识码】:**A

1 引言与文献评述

2014 年, 国务院印发《社会信用体系建设规划纲要(2014—2020 年)》(以下简称“纲要”), 正式明确我国社会信用体系建设的顶层设计架构, 并全面启动长达 7 年的社会信用体系建设周期。2021 年, 我们再次站在了新一轮建设的起点, 面向国家提出的高质量社会信用体系建设新目标, 有必要对上一轮建设进行科学评估, 以更好的规划未来。

目前, 从整体社会信用环境评价, 到社会信用体系建设评价, 到政务、商务、社会和司法等重点领域诚信评价, 以及政府、企业、个人等重点主体诚信评价, 应该说, 学术界关于社会信用评价的研究成果已经十分丰富。但对文献梳理来看, 从社会公众的角度, 关于社会信用体系建设关注度的评价方法研究和实证分析并不多。公众对社会信用体系建设的关注程度是公众信用意识最直观有效的体现, 提高公众诚信意识和全社会信用水平是社会信用体系建设的重要内容和重要目标。如何客观评估公众对社会信用体系建设的关注度, 据此反映我国社会信用体系建设的成效和不足, 成为研究的当务之急。

网络行为数据是公众对事务关注度、认知程度和意见态度的一种体现。运用基于需求搜索数据对公共关注度进行测量的相关研究已取得不少成果。例如, Ginsberg 等人(2009)提出利用谷歌搜索数据开展流行病预测, Ripberge 等人(2011)运用与流病学相关的谷歌词汇搜索数据追踪传染病的扩散。国内也有不少学者运用百度指数用于经济问题和社会热点议题的趋势预测与

¹**作者简介:** 赵蓓 (2001-), 女, 浙江金融职业学保险专业学生;

毛通 (1980-), 浙江金融职业学院保险专业副教授, 研究方向: 信用大数据;

王佳华 (1999-), 浙江金融职业学保险专业学生。

基金项目: 浙江省大学生科技创新活动计划(新苗人才计划)项目“一种基于大数据的信用舆情监测系统”(2020R421002); 教育部人文社科研究青年基金项目(18YJC790117); 浙江金融职业学院 2019 年度基本科研业务费重点项目(2019ZD02)

时空分析。例如，孟天广、赵娟(2019)基于百度搜索数据采集了2011-2017年26个政策议题的2亿余条关键词指数，对公众的搜索行为与公共(政策)关注度进行研究。毛通、谢朝德(2019,2020)等人基于百度搜索指数，运用信用关联关键词来编制信用舆情指数、信用治理满意度指数，来动态跟踪评估网络公众的社会信用舆情变化态势。

文章正是在上述相关研究成果基础上，进一步提出运用互联网大数据，通过分析网络公众对社会信用体系建设相关的重大问题、热点话题的网络检索、评论、转发、点评、关注等网络行为，达到公众网络关注度评估的目的。文章的主要创新点在于依托百度指数和互联网大数据手段，以301个社会信用热点关键词为统计依据，以浙江省11个地市为实证研究对象，通过对2014-2020年间各领域各地区网民网络行为特征的纵向和横向比较，从而较为客观真实的还原公众对社会信用体系建设网络关注度的时空变化规律。

2 基于百度搜索指数的诚信建设网络关注度度量原理

百度搜索指数是一种以网民在百度的搜索量为数据基础，以关键词为统计对象，通过计算出各个关键词在百度网页搜索中搜索频次的加权，用以分析互联网用户对关键词搜索关注度及持续变化情况的工具。文章以《纲要》为诚信建设关键词的基础语料库，围绕诚信建设的重点领域和重点人群，从政务、司法、生产、流通、金融、互联网金融、电商、税务、价格、交通运输、医疗卫生、社会保障、劳动用工、教育科研、节能环保、知识产权、互联网、旅游18个领域，公职人员、社会公众和从业人员3类群体，初选诚信核心关键词，然后以其为基础，通过百度指数需求图谱关键词推荐功能，关联出与核心关键词存在高度相关性且搜索热度最高的关联关键词，剔除掉搜索热度低、关联度不达标的关键词。然后运用爬虫软件在百度指数中爬取每个关键词PC端和移动端的日搜索值，加总得到年搜索指数。通过将上述关键词搜索指数按照领域分类，得到不同领域、不同时间、不同地区的分类搜索指数，最后将全部关键词加总得到搜索总指数，并以此作为评估各地诚信建设网络关注度的依据。

3 浙江省2014-2020年诚信建设网络关注度测算结果

3.1 数据说明

文章最终甄选了301个诚信建设关键词，以浙江省为实证研究对象，通过浙江省11个地市2014-2020年间当地网民在百度搜索引擎中对全部诚信建设关键词的搜索量分析，对浙江及11个地市诚信建设网络关注度的动态变化过程进行分析，实证数据采集详细流程图如图1所示。

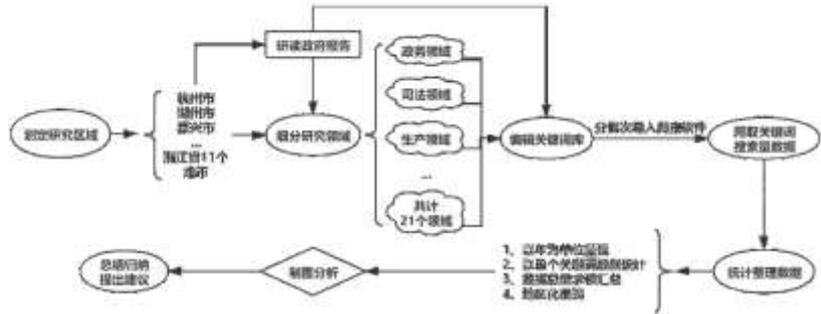


图1 实证数据采集流程

3.2 关键词统计

对关键词统计结果看，2014—2020年间，浙江省网民通过PC和移动端在百度搜索引擎中对301个关键词的总搜索量为2.34

亿频次，日均 9.14 万频次，七年间平均每个词搜索 77.24 万频次，其中最多一个关键词搜索量达 669.38 万频次，最小一个关键词搜索量为 0.96 万频次。

表 1 浙江网民 2014-2020 年间 301 个诚信建设关键词检索统计

总搜索量	日均搜索量	词均搜索量	最大搜索量	最小搜索量
2.34 亿	9.14 万	77.24 万	669.38 万	0.96 万

4 浙江省 2014-2020 年诚信建设网络关注度实证分析

4.1 搜索总指数趋势分析

从对浙江省 2014-2020 年间全部 301 个关键词百度搜索总指数统计结果分析来看(见图 2):2014-2018 年搜索总量呈逐年上升趋势,2018 年达到峰值,2019 年和 2020 年呈下降趋势。从中可以看出,网民对诚信建设网络关注度显著高于建设初期的 2014 年。至于 2019 年和 2020 年搜索总指数下降的原因,文章认为有两个可能性:第一,社会信用体系建设已经从快速发展期进入成熟期,公众早期基于认知需求的网络搜索行为态度和偏好开始发生转变;第二,受到短期干扰因素的影响,例如“问题疫苗”“集资诈骗”“债务违约”“学历造假”“阴阳合同”等事件型关键词的短期波动,此外,2020 年疫情对经济活跃度的干扰。上述两点中文章更倾向于前者是主因。

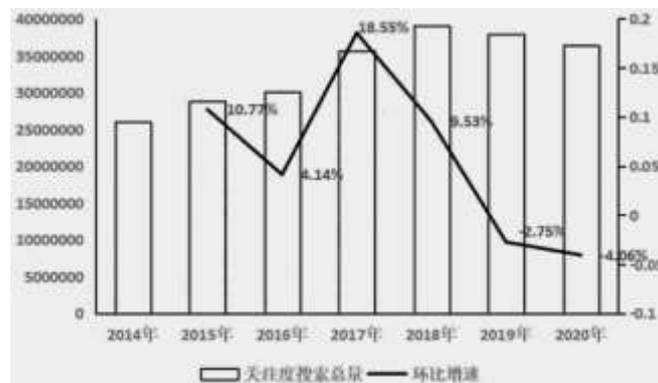


图 2 2014-2020 年浙江省 11 个地市搜索总指数

4.2 搜索总指数的领域对比

文章将 18 个细分领域合并为政务、商务、社会、司法四大领域,同时为了便于横向比较,将各领域搜索总指数取自然对数,得到各领域搜索指数趋势图。从四大诚信建设重点领域的搜索总指数趋势分析来看(见图 3):2014-2020 年间,政务、商务、社会和司法搜索指数年均增速依次为 21.96%、3.02%、1.52%和 4.87%,增速最快的为政务领域,增速远高于其他领域。这表明网民对政务诚信的关注度上升最快。从峰值分析看,商务和社会诚信关注度的峰值出现在 2018 年,政务诚信关注度峰值出现在 2019 年,司法诚信关注度的峰值出现在 2020 年。

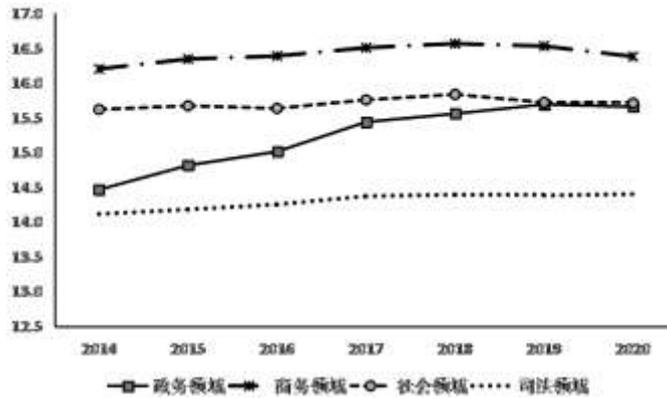


图 3 2014-2020 年间浙江省四大领域搜索总指数(对数值)

4.3 各领域搜索指数的年均增速比较

从对浙江省 2014-2020 年间 18 个细分领域、3 大类主体诚信建设关键词搜索指数年均增速测算结果分析来看(见图 4):浙江网民关注度增速上升最快的是政务领域诚信建设,搜索指数达 21.96%,排名第二的是互联网金融领域诚信建设(12.22%)和从业人员的诚信建设(11.64%);期间诚信建设关注度出现负增长的有价格领域、交通运输领域、社会保障领域、节能环保领域、旅游领域和社会公众等 7 个领域或主体,这表明网民对上述领域的诚信建设关注度有所下降。

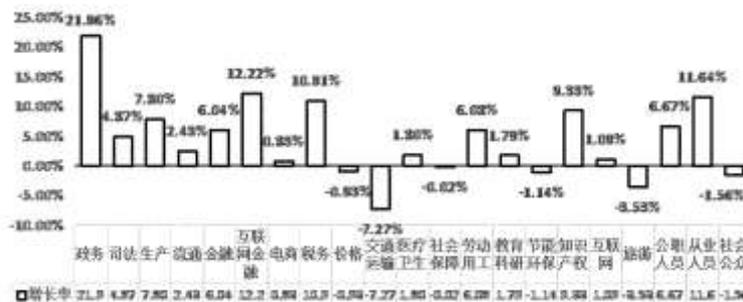


图 4 2014-2020 年间浙江省各细分领域关键词的平均增速

4.4 搜索总指数的地区比较

从浙江省 11 个地市 301 个关键词 2014-2020 年搜索总指数的横向对比来看(见图 4),搜索总指数排名前三依次为杭州、宁波、温州,排名后三依次为衢州、舟山和丽水。文章以 2019 年各地 GDP 为自变量,搜索总量为因变量,进行线性回归,决定系数为 0.911,两者关系成立。从回归结果来看,很显然,搜索总指数与各地 GDP 存在显著的正相关性。这表明,网民对信用的关注度与经济发展水平有关,经济发展水平越高的地区,网民对信用的关注度也就越高。

5 研究结论

第一,浙江网民对诚信建设的关注度呈现先上升后下降的变化特征。

这一变化符合浙江省社会信用体系从建设初期进入快速成长期再到成熟期的演变特征,即在建设初期,公众的整体关注度

不高，随着建设进程的加快，公众的关注度开始快速上升，当进入成熟期后，随着诚信建设中的热点和问题的转移，关注度开始下降。

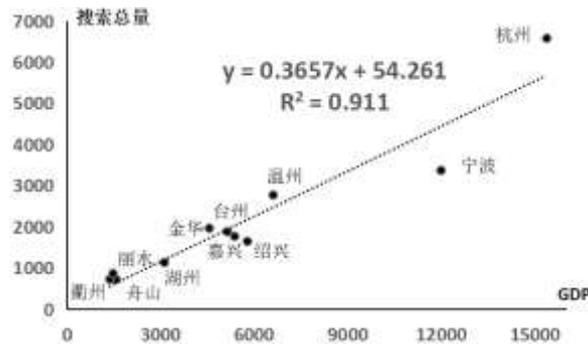


图 5 2014-2020 年浙江省 11 个地市搜索总指数与 2019 年各地 GDP 对比

第二，各领域间网民关注度的分化较为明显。

首先，从四大领域对比看，政务领域的诚信建设关注度上升速度远高于商务领域、社会领域和司法领域，这表明当地政务诚信建设仍然在持续发力，不断深入推进；其次，18 个细分领域和 3 大主体中，有 11 个是正增长，有 7 个是负增长，网民对各领域的诚信建设关注度有较大差异。

第三，搜索指数与各地经济发展水平正相关。

经济水平越是发达的地区，当地公众对诚信和信用的价值就越是重视，表现在日常的网络行为中，有关诚信和信用的话题或事件较为关注。

参考文献:

[1] 廖玲. 我国的社会信用体系建设现状分析[J]. 环渤海经济瞭望, 2020, (11): 125-127.

[2] 张俊杰. 大数据在网络舆情治理中的应用[J]. 中国国情国力, 2021, (1): 42-45.

[3] 毛通, 谢朝德. 基于舆情大数据的城市信用治理满意度评价——来自 17 个 GDP 超万亿元大城市的实证[J]. 征信, 2020, (09): 15-23.

[4] 毛通, 谢朝德. 基于百度大数据的信用舆情指数构建与实证研究[J]. 征信, 2020, (1): 11-20.

[5] 孟天广, 赵娟. 大数据时代网络搜索行为与公共关注度: 基于 2011-2017 年百度指数的动态分析[J]. 学海, 2019, (3): 41-48.