

社会情绪的结构性分布特征及其逻辑^{*}

——基于互联网大数据 GDELT 的分析

龚为纲 朱 萌

内容提要:本研究试图整合情感社会学相关理论资源,借助互联网大数据及其相关算法,对当前中国社会各阶层情绪体验的突出特征进行归纳,对阶层互动各场景中的情感氛围进行比较。研究发现,从各阶层的总体性情绪体验指标来看,社会结构骨干阶层与中间阶层情绪体验的积极程度明显高于社会结构普通阶层。情感体验的阶层分布逻辑在于,社会结构普通阶层较之其他更多地暴露于多种类型的压力与风险中,而且这些压力更容易唤起其负面情绪。从阶层互动场域的“情感氛围”来看,不同阶层互动场域中的情感氛围明显不同,社会结构的骨干阶层内部互动的情感氛围相对和谐,而这一阶层与普通阶层、中间阶层与普通阶层互动场景中情感氛围相对紧张。上述有关情绪体验的阶层分布特征,有助于透视社会转型时期的阶层关系与社会结构,加强预防和化解社会矛盾机制建设,正确处理人民内部矛盾。

关键词:社会情绪 大数据 阶层 阶层互动

一、研究问题

随着 1978 年以来经济体制转轨的深入,中国社会结构剧烈分化,强调获致性因素的职业机制逐渐成为新的社会分化机制,社会经济资源的获取途径日益多样化,从而使阶层结构在改革之后进入快速分化的进程^①。社会分化的后果表现在一系列政治、社会领域。已有研究对阶层分化过程中各阶层的消费倾向、居住空间、文化资本与文化品位、社会资本、阶层意识等议题进行了富有成效的探索。

值得注意的问题是,社会的阶层分化在现实层面既表现为阶层之间在生活水平和生活方式上的差异日益明显,更使得人们的认知方式、价值观念和情感心理开始呈现出根据所属阶层而分化的特征^②。故而,对不同阶层的社会心态进行研究具有重要理论意义和现实意义。党的十九大报告指出,要提高保障和改善民生水平,加强和创新社会治理。深入研究当前社会情绪的阶层

* 本文为国家社会科学基金重大项目“大数据时代计算社会科学的产生、现状与发展趋势研究”(16ZDA086)、武汉大学青年学术发展计划团队项目“新媒体与新社会化”(Whu20160008)的研究成果。武汉大学罗教讲教授对本文写作有实质性贡献。特此致谢。

① 陆学艺:《当代中国社会阶层的分化与流动》,《江苏社会科学》,2003 年第 4 期。

② 马广海:《存在、认同与冲突:转型期我国社会的阶层意识概览》,《山东社会科学》,2011 年第 5 期。

分布问题,有助于促进社会公平正义,形成有效的社会治理、良好的社会秩序。已有社会心态相关研究侧重于社会分化进程中各阶层的阶层意识^①、幸福感、不公平感^②等方面,而社会情绪作为社会心态中的核心要素,情感社会学领域仅有少量关于中国不同阶层之社会情绪的相关讨论,但主要聚焦于理论探索层面^③,在实证层面则语焉不详。另外也有研究针对一些具体的负面情绪做过分析^④。总的来说,社会情绪相关研究还存在以下问题需进一步推进。

一是将社会情绪纳入社会结构中进行思考。社会情绪是大众在社会心态层次上的主观性体验,这种主观性体验既是当前社会分化的产物,也是透视社会结构及其问题的社会表征。通过分析社会各个阶层中的突出性情绪问题,寻找其根源,进而为塑造“理性平和、积极乐观”的社会心态提供治理之策。

二是将社会情绪嵌入到社会情境中进行分析。社会情绪的产生与社会互动有关系,不同阶层的互动关系是完全不同的。如果研究能够呈现不同阶层互动时的“情感氛围”,不但有助于了解社会情绪生成的结构性背景,还有助于理解社会转型时期的阶层关系。

本文试图回答,在社会转型过程中,中国社会不同阶层除了在主导性生产方式、消费倾向等领域存在阶层分化之外,在社会情绪上是否也存在系统性差异?这种差异的形成机制是怎样的?社会情绪的产生往往有阶层互动背景,那么不同阶层互动场景中的“情感氛围”是否有所不同?阶层互动的情感氛围折射出当前中国阶层关系的何种特征?随着互联网与大数据分析技术日新月异的发展,对网络大规模人群情绪的实时监测成为可能,为以往难以想象的研究策略的实现提供了可能。本文试图借助于谷歌云计算所提供的超强算力,对全球网络空间中门户网站、网络新闻、网络论坛等 190531 个网站所包含的 7.3 亿个新闻类网页的信息进行收集、分析和编码,借助于自然语言处理等算法,对海量的文本信息进行“情感计算”,进而实现对社会情绪的动态监测,并据此透视社会系统的结构与运行逻辑。

二、文献回顾与分析框架

从社会心理学角度看,社会情绪是一种集体情绪,即一定社会环境下某一群体或者某些群体或整个社会中多数人所共享的情绪体验,社会情绪并非个体情绪的叠加或混合,而是个体之间、个体与群体之间、群体与群体之间互动的结果^⑤。

故而,社会因素作为影响社会情绪的重要源头长期以来被学界所重视。已有社会情绪研究常把个体的情绪、情感放在社会结构中理解。这方面,肯珀(Theodore Kemper)的权力-地位理

① 刘欣:《转型期中国大陆城市居民的阶层意识》,《社会学研究》,2001 年第 3 期;陈云松:《阶层自我定位、收入不平等和主观流动感知(2003~2013)》,《中国社会科学》,2016 年第 12 期;马广海:《存在、认同与冲突:转型期我国社会的阶层意识概览》,《山东社会科学》,2011 年第 5 期;张文宏、刘永根:《社会分化、生活体验与阶层冲突的主观建构》,《社会科学战线》,2017 年第 1 期。

② 翁定军:《阶级或阶层意识中的心理因素:公平感和态度倾向》,《社会学研究》,2010 年第 1 期。

③ 成伯清:《心态秩序危机与结构正义:一种社会学的探索》,《福建论坛》,2016 年第 11 期;王鹏:《情感社会学的社会分层模式》,《山东社会科学》,2013 年第 3 期。

④ 成伯清:《从嫉妒到怨恨》,《探索与争鸣》,2009 年第 10 期;王俊秀:《警惕群体性怨恨成为“社会心病”》,《中国教育报》,2012 年 1 月 11 日;马广海:《从群体性事件看转型期社会心态》,《中国海洋大学学报》,2012 年第 6 期;王佳鹏:《羞耻、自我与现代社会》,《社会学研究》,2017 年第 4 期;杨宜音:《浮躁怎样成为一种社会心态》,《江苏行政学院学报》,2014 年第 6 期;周晓虹:《中国社会心态危机蔓延》,《人民论坛》,2014 年第 9 期;李春玲:《中国中产阶级不安全感和焦虑心态》,2016 年第 4 期。

⑤ 王俊秀:《社会心态理论》,第 104 页,社会科学文献出版社,2014 年版。

论、柯林斯(Randall Couins)的仪式互动理论堪称经典,这两种理论的核心逻辑认为,客观的社会结构性位置会影响公众的社会心态和情绪体验,不同阶层的社会位置塑造了不同的生活模式,而生活模式塑造了人的性格结构,故而不同的结构位置会对人的性格结构与情绪体验产生影响。^①肯珀、柯林斯等人有关情感的社会分层模式极有启发意义,其所作的机制解释也极富洞见。但问题在于,与心理学中有关情绪个体和微观层面的实证研究相比,很难采取实证的方法去验证已有情感社会学有关情感分层的理论,这是情感社会学理论的劣势。就情绪体验的社会分层而言,能否在实证层面得到检验呢?

客观性的结构性位置是否会转化成为情绪体验上的差异,这个问题在实证层面有待检验。更为重要的是结构性位置作用于情绪体验的具体机制。

本研究认为,不同的社会结构条件会诱发不同的情绪类型。这里所谓的结构性条件主要是指阶层结构。阶层结构既包括行动者所在阶层位置,也包括不同阶层行动者互动所形成的关系结构。这种考虑,主要与社会分层领域两种研究社会阶层的视角有关。在已有的社会分层的理论脉络中,认为既定社会的阶层结构既可以被视为阶梯性的,也可以被视为关系性的。阶梯性模型依据社会成员拥有资源的多寡来确定其社会阶梯中的地位,用该模型描述阶层结构,能发现主要资源、生活机遇的不平等状况。关系性模型把社会成员间的权力-利益关系视为阶层分析的核心议题,强调以社会关系确定阶层地位。^②

如下这两种社会分层视角,有助于揭示社会情绪生发的结构性条件。

第一,从阶梯性视角看,不同社会成员处于不同的结构性位置,有不同的社会资源,正如情感社会学领域的权力-地位模型所描述的那样,不同的结构性位置产生不同的情绪体验。结构性位置与情绪体验之间的关联机制很多。考虑到本研究主要分析社会负面情绪的阶层分布及其逻辑,我们试图借助精神健康社会学领域的压力-过程模型理解结构性位置与情绪体验之间的关联机制。在社会分层与心理健康研究领域,社会结构被看作影响个体健康的根本原因。阶层等结构性因素主要通过影响社会压力的阶层分布塑造社会情绪,其核心机制主要有两层:一是压力暴露存在阶层差异,二是在应对压力时,不同阶层的脆弱性不同,即压力易感性^③。

压力暴露的阶层差异。压力过程模型的第一个研究维度多涉及对于具体压力来源的探索,例如,急性生活事件、慢性生活压力,以及日常生活烦恼,等等。本研究试图将这作为社会情绪阶层分化的一层重要机制。其中的逻辑在于,不同社会地位的群体处于不同生活环境,因而受到不同类型和不同程度的压力影响。即压力暴露的阶层差异构成社会情绪阶层分化的第一层机制。

压力易感性的阶层差异。所谓压力的易感性,是指容易因为生活压力而产生情绪问题或罹患心理疾病的程度。由于各阶层的群体性差异,不同社会地位造成不同的生活环境和社会经历,同样面对风险与压力,不同阶层可能会产生不一样的情绪反应,社会地位较高的人群对压力的感受更为积极。故而,易感性的阶层差异构成社会情绪阶层分化的第二层机制。

第二,从关系性视角看,社会情绪的生成机制,与社会情绪本身的属性有关。正如有研究者指出的,社会情绪是一种集体情绪,即一定社会环境下某一群体或者某些群体或整个社会中多数人所共享的情绪体验,社会情绪并非个体情绪的叠加或混合,而是个体之间、个体与群体之间、群

① 参见兰德·柯林斯:《仪式互动链》,第366页,商务印书馆,2016年版;特纳、斯黛兹:《情感社会学》,第178~179页,上海人民出版社,2007年版。

② 刘欣:《协调机制、支配结构与收入分配:中国转型社会的阶层结构》,《社会学研究》,2018年第1期。

③ 梁樱:《心理健康的社会学视角》,《社会学研究》,2013年第2期。

体与群体之间互动的结果。^① 那么不同阶层互动所构成的场域中的“情感氛围”是否有所不同呢? 在不同阶层相遇的场域中,如果情感氛围存在显著性差异,那么可以进一步认识情绪体验的阶层特征,进而理解阶层互动对于情绪体验的影响,这时的情绪体验,不是单个阶层所独享的情感体验,而是两个或多个阶层互动场景中的“情感氛围”。

基于上述分析,笔者提出如下研究假设:

研究假设 1:不同的结构性位置上会产生不一样的情绪体验,就总体性情绪指标而言,情绪体验的积极程度与阶层地位成正向相关,负面情绪体验程度与阶层地位呈现负向相关。

研究假设 2:从压力过程模型的视角看,普通阶层的压力暴露程度和压力易感性高于骨干阶层,二者的阶层分布构成社会情绪阶层分化的两层重要机制。

研究假设 3:社会情绪的形成与阶层互动有关,不同阶层互动下的场域,有可能存在完全不同的情感氛围,缓和的阶层关系更有助于产生积极情感氛围,而对立紧张的阶层关系,更倾向于产生消极的情感氛围。

三、数据与方法

本研究的数据来源为互联网大数据 GDELT(全球事件、语调与语言数据库, Global Database of Events, Language and Tone),该数据库由谷歌开发者卡列夫·利塔鲁(Kalev Leetaru)所创建。

2011 年,《科学》杂志上发表了由哈佛大学、麻省理工学院与谷歌研究团队利用谷歌图书(Google Books)的海量数字化资料所做的研究。他们分析了公元 1500 年到 2000 年间 500 多万本书籍、高达 5000 多亿单词的语料库,并提出了“文化组学”(Culturomics)的概念,所谓“文化组学”是指用计算语言学的方法对大量电子文本信息进行定量分析进而研究人类行为和文化趋势的方法^②。利塔鲁(Kalev Leetaru)等人在此基础上提出了“文化组学 2.0”的概念,试图通过汇聚互联网空间中的信息理解整个世界,探索人类社会如何运行,特别是人类的社会互动和社会行为。GDELT 数据库就是在这一思想的基础上建立的,其运作方法是,实时监测全球网络空间、门户网站、印刷媒体、电视广播、网络媒体、网络论坛中的新闻事件,对其进行分析提取,提炼出新闻事件相关的人物、地点、组织和事件类型等关键信息。自 2015 年 2 月以来,GDELT 已经汇聚全球网络空间中的 738,405,547 条信息,数据体量高达 7.7TB^③,这是典型的互联网大数据。

本研究以 GDELT 中有关中国社会生活的 2900 万个事件构成的案例库为依据,借助社会学领域有关社会分层的操作化概念,将网络信息文本中常见的 1600 多种社会角色与社会活动主体概括为 9 大社会阶层,进而在大数据的语境下将社会阶层的概念进行操作化处理;借助于机器学习、WordNet^④ 情感词库等自然语言处理技术,对中国社会的总体性情绪特征以及 150 多种负面情绪进行测量,进而尝试将社会情绪进行量化,呈现社会各阶层的多重情感面向。

① 王俊秀:《社会心态的理论》,第 107 页,社会科学文献出版社,2014 年版。

② Jean-Baptiste Michel, et al, “Quantitative Analysis of Culture Using Millions of Digitized Books”, *Science*, Vol. 331, No. 6014, 2011.

③ 本研究的计算所涉及的时间跨度为 2015 年 2 月~2018 年 3 月。

④ “语言获得与词汇计数程序”(Linguistic Inquiry and Word Count, LIWC)能够在线测量网络文本信息中的数十种情绪和行为维度,包括积极、消极情绪,以及焦虑、愤怒和行为抑制等方面。

(一) 阶层与压力指标的操作化

在 GDELT 数据中与本文密切相关的信息包含 Themes(主题)、V2Tone(情绪)、GCAM(全球内容分析量表)等维度。Themes 这一列信息包含 GDELT 团队利用各种自然语言处理算法,从文本信息中提取 53,165 个反映社会生活的主题^①。本文以下信息都从 Themes 中提取。

一是职业群体的信息,GDELT 数据库可以识别的职业群体类型包含 2243 个,如果某个个案中包含某种职业角色,比如说警察,则赋值为 1,没有则赋值为 0。这样,针对一个文本个案,就形成 2243 个 0/1 变量。本文从中提取了其中的 1600 多个能够反映职业特征的类型信息,然后结合李路路等人的社会分层框架^②,根据其所占有的生产资料、组织权威和专业技能,将这 1600 多个职业群体划分为 9 大社会阶层,这 9 大阶层分别是:处于社会结构骨干阶层的官员阶层、企业家阶层、知识分子阶层,处于社会结构中间阶层的技术人员阶层、小业主阶层、办事人员阶层、技术工人阶层,处于社会结构普通阶层的非技术工人阶层、农民阶层。

二是压力过程模型中的相关信息,本文主要从 6 种社会生活中常见的压力与风险类型对压力来源进行测量,包括危机情境、冲突情境、自然灾害、人为灾害、贫困、失业风险等。其中危机情境涉及 20 多种类型^③,冲突情境涉及 51 种类型,人为灾害涉及 95 种类型,自然灾害涉及 163 种类型。由于每一种压力类型下的子类繁多,本文通过因子分析提取主成分的方法,对这些信息进行综合,分别形成了 6 个压力指数,即冲突情境指数、危机暴露指数、自然灾害暴露指数、贫困暴露指数、失业暴露指数、人为灾害暴露指数。

(二) 社会情绪指标的操作化

社会情绪是指人们对社会生活情境的知觉,是通过群体成员之间相互影响、相互作用而形成的较为复杂而又相对稳定的态度体验,这种知觉和体验对个体或群体产生指导性和动力性的影响^④。按照社会情绪的性质与投射方向,可以将它划分为积极社会情绪与消极社会情绪。

本文对社会情绪的测量分为两个层次,首先是总体性层次,包含积极情感指数、消极情感指数以及综合情感指数。总体性情感指数用于比较各阶层情绪状态的总体状况。GDELT 针对每一个个案,通过高性能自然语言处理技术,给出其综合情感指数(tone)、积极情感指数(positive score)、消极情感指数(negative score)。对各阶层的总体性情绪特征,我们主要从 V2Tone 中提取信息。

其次是具体情绪类型层次,基于 GCAM 中 WordNet Affect 1.1 所监测的 150 多种负面情绪类型^⑤,本文侧重分析典型负面情绪的阶层分布。

① 通过机器学习与自然语言处理提取文本等非结构化数据中的主题和情感信息,是目前计算社会科学领域的主流做法,具体可以参见 Scott A. Golder and Michael W. Macy, "Diurnal and Seasonal Mood Vary With Work, Sleep, and Daylength Across Diverse Cultures", *Science*, Vol. 331, No. 6051, 2011.

② 李路路、朱斌:《当代中国的代际流动模式及其变迁》,《中国社会科学》,2015 年第 5 期。

③ 关于这 20 种危机的类型,具体内容请参见 GDELT 官网, <https://blog.gdeltproject.org/crisislex-taxonomies-now-available-in-gkg/>。

④ 沙莲香主编:《社会心理学》(第二版),第 171 页,中国人民大学出版社,2006 年版。

⑤ WordNet Affect 1.1 根据文本的内容,对文本中含有情绪色彩的语义信息进行提取,故而这种提取不仅是情绪状态,还包含那些蕴含情感色彩的社会行动与肢体语言等。WordNet Affect 1.1 涉及到对 150 种左右的消极情绪和 130 种左右的积极情绪的识别,限于篇幅,本文主要展示各阶层的消极情绪。

四、社会情绪的阶层分布特征及其逻辑

GDELT 数据库对每个案例中的积极情绪与消极情绪做了精确计算。通过这两个指标比较各个阶层的情绪状态,即积极情感指数越高,说明该阶层在生活世界中体验着更多的积极情绪,具有更高的精神健康程度;反之,如果消极情感指数越高,说明该阶层在生活世界中体验着更多的消极情绪,精神健康程度相对较低。

(一) 积极情感指数、消极情感指数与综合情感指数的阶层分布

根据 GDELT 数据库中的信息,计算各阶层消极情感指数和积极情感指数,为了便于阶层比较,笔者对积极情感指数和消极情感指数进行标准化处理(见图 1)。

从阶层之间相对比较角度看,农民、非技术工人、技术工人、办事人员等,其消极情感指数明显高于官员、企业家、知识分子、技术人员、小业主等社会阶层。消极情感指数的阶层分布符合研究假设 1(图 1)。

为了比较更为简便,笔者通过综合情感指数进行比较。综合情感指数越高,说明情感体验的积极程度越高,反之则说明消极情感体验越强。透过这个指数,可以对各阶层情感体验的总体状况进行比较。综合 2015 ~ 2017 年的数据发现,综合情感指数存在鲜明的阶层分化特征(图 2)。

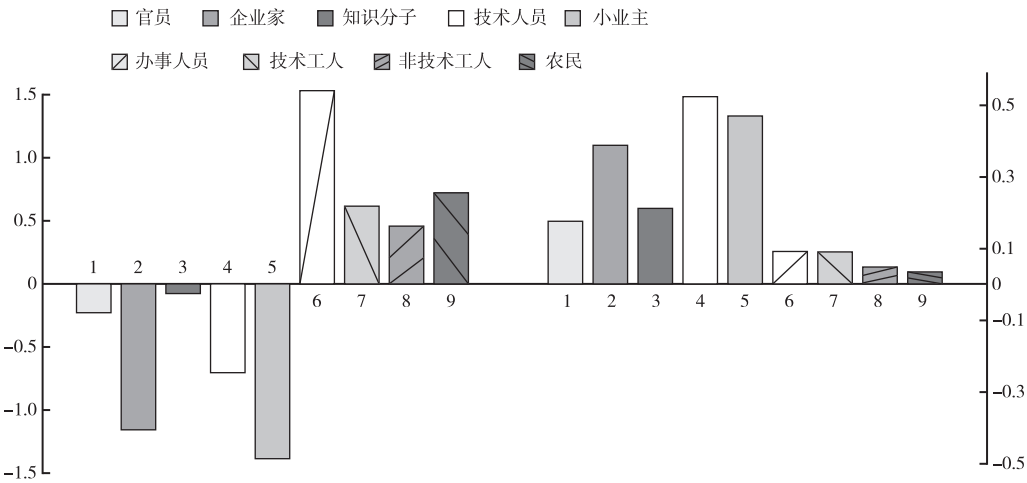


图 1 消极情感指数

图 2 综合情感指数

图 2 显示:官员、企业家、知识分子和技术人员、小业主的综合情感指数远远高于办事人员、技术工人和非技术工人、农民;中间阶层内部分化明显,技术人员、小业主远远高于办事人员和技术工人;办事人员、技术工人又明显高于非技术工人、农民。综合情感指数的阶层分布,大致符合阶层地位与情绪体验的积极程度成正向相关的研究假设。

图 2 还显示:情感体验积极程度最高的是处于社会结构中间位置的技术人员和小业主阶层,消极情感体验最强的是非技术工人与农民阶层。技术人员、小业主的积极情感体验明显高于其他阶层,这能否反映出中间阶层的一些社会结构属性呢? 比如已有研究认为,这一阶层属于社会结构中相对理性、乐观的阶层,是社会结构中的稳定器,本文从大数据与情感计算的角度,对这些判断提供了初步证明。

官员、企业家、知识分子阶层,尽管在权力、财富、声望与知识、技术等稀缺资源的占有上处于绝对优势地位,但是这三个阶层并不是积极情绪体验最高的。这三个阶层的情绪体验远远优于普通阶层,但是不及技术人员、小业主。很显然,这很难从物质财富的角度去解释,而可能与这三个阶层面临更大的工作压力有关系。

(二)代表性负面情绪的阶层分布特征

通过上文的分析,我们已经发现总体性情绪指标具有阶层差异性。下面结合数据挖掘过程中的一些发现,对各阶层存在的典型负面情绪作简要分析。

1. 负面情绪主要集中在弱势群体底层,负面情绪体验强度与阶层地位成负向相关。当前中国社会心态领域的突出问题是一定程度的负面情绪普遍蔓延。那么这一情绪具有怎样的阶层分布特征呢? 数据分析显示,该情绪的阶层分布,基本上呈现出根据阶层地位从低到高,而呈现出由强到弱的态势:阶层地位越低,程度越强;阶层地位越高,程度越弱。所谓社会弱势群体,比如说非技术工人、农民等阶层中的负面情绪明显超过其他阶层(图 3)。这主要与这一情绪本身的属性有关系,不满主要与“无能感”有关:受到伤害,马上给予回击或者自卫,则不会心怀怨恨,但假如由于无能、恐惧或者害怕,必须压抑伴随的愤怒情绪,则这种隐忍就容易酿成不满,这是此情绪突出的主要原因。

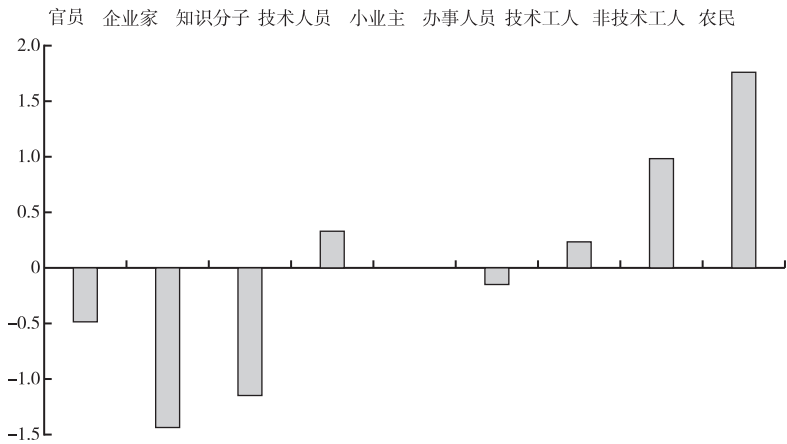


图 3 各阶层中突出性负面情绪的阶层分布

2. 反腐败高压形成“不敢腐”的氛围,使部分“问题官员”中的忧虑、恐慌情绪相对突出。数据分析结果显示,与其他阶层相比,官员阶层的恐慌、忧虑等情绪体验更为明显(图 4),一些“问题官员”被查、落马,在总体上增强了官员阶层中的“忧虑”、“恐慌”情绪,官员阶层中的“忧虑”、“恐慌”情绪相对较高主要与这些反腐败事件有关系。

3. 商业气候、大力反腐与企业家阶层所谓的“不安全感”有关。与其他阶层相比,不安全感在企业家阶层中最为明显(图 4)。自 2008 年金融危机爆发以来,全球经济复苏长期乏力,而近年来中国经济也进入去过剩产能、调结构等新常态,一些企业债务危机严重,这无疑会让企业家阶层压力加重。更为重要的是,企业家阶层的不安全感或许与反腐败的高压态势有关。企业家阶层,作为社会转型的得利群体,在过去积累了巨额财富。一些企业运营过程中存在着一定的官商勾结与腐败行为,使得该群体忧虑财富存在风险,以至于不安全感比较突出。近年来,大量腐败官员落马的同时也有大量企业家被查,为这种不安全感提供了注脚。

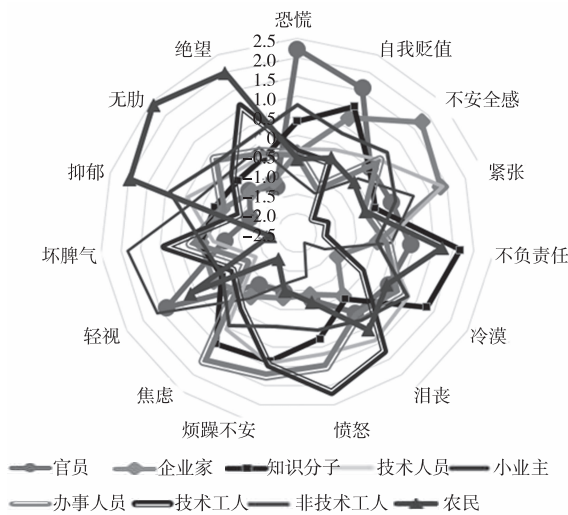


图 4 各阶层中突出的负面情绪体验

4. 人民日益增长的美好生活需要和不平衡不充分的发展之间存在结构性矛盾,相对弱势阶层对美好生活的需求尚未得到有效满足,并表现为相应的负面情绪。

比如,办事员、技术工人的生存压力与焦虑情绪。与其他各阶层相比,在中间阶层中盛行焦虑情绪,突出地体现在办事人员阶层(图4)。办事人员阶层是典型的城市白领阶层,根据已有的大量研究,“白领”阶层最为突出的情绪特征就是“焦虑”,这一判断在本研究中得到来自互联网大数据的支持(图4)。“焦虑”、“焦躁”情绪来源于该群体的结构性位置与生存压力。

又如,技术工人阶层的不满,甚至是愤怒情绪。技术工人阶层是工人阶层的主体部分,该阶层预期在城市立足生存,但是近年来房价上涨,对这个阶层诸多人来说,在北上广一线城市买房立足已经是不可能,在二线城市同样需要支付高房租,背负高额的房贷,以至于很多人自嘲为“房奴”。而对于没有买房子的群体而言,这样一种不满与愤怒表现更为强烈。

最后,非技术工人团体“挫败感”强烈。他们是城市社会的底层,是靠体力获取收入维持生存的阶层,主要由农民工以及城市弱势群体(比如小摊贩、各种苦脏累的活、建筑工人等)构成。这个阶层的市民化历程十分艰难,其中很多人都因为无法在城市真正立足,而不得不返回乡村或者继续维持现状。挫败感作为其情绪体验的突出特征,与该阶层缺乏改变命运的机会、向上流动缺乏相应的能力和空间有关。

(三) 社会情绪的阶层分化机制

精神健康社会学认为,人类的负面情绪主要与生活压力有关,这种压力可以是来自各方面的,比如说生存压力、工作压力等。下面我们结合前文介绍的压力-过程模型,对综合情感指数的阶层分布以及在中国社会中普遍流行的负面情绪的阶层分布进行回归分析。

1. 风险与压力暴露的阶层差异。社会生活中存在大量的危机性情境,这是一个客观事实。风险的发生会中断人们的正常生活,或者造成人员伤亡、财产损失,或造成生活不便、情感受伤,故而,作为一种反常的生活状态,风险性情境会唤醒不安、悲伤、恐惧等负面情绪。然而,风险的暴露并不是在社会各个阶层均匀地分布。与财富向上层聚集相反,社会风险是向社会下层

集中的。

本文从大数据的角度证明了这一判断:即各种生活压力的阶层分布都呈现出明显向下层集中的特征,社会结构中的农民阶层、非技术工人阶层,在各种生活压力上的暴露指数要远远高于其他阶层(见表 1)。

首先,我们将社会生活中的压力归纳为 6 个方面:天灾(自然灾害)、人祸(人为灾害)、贫困、失业、社会冲突、以及各种危机性情景。显然,对于公众而言,只要暴露于上述任何一种压力,都会中断正常生活,产生精神压力。上述 6 个方面的压力涵盖了社会生活中最主要的压力类型。

其次,将上述 6 个方面的压力信息从 GDELT 数据中提取出来。具体做法是,GDELT 数据库根据自然语言处理和概率主题模型算法,从超高维的层面对文本中的信息进行提取,GDELT 数据库目前可以提取的主题信息包含 53165 种类别,几乎涵盖了人类社会生活中绝大部分社会现象。也就是说,GDELT 可以从 53165 个维度对一个文本个案中包含的信息进行提取,对于任意一个文本案例,通过自然语言处理可以识别该案例是否包含了某一个主题的相关信息,这样就可以形成 53165 个 0/1 变量,透视这一个文本案例。

本研究所需要的上述 6 个方面的压力信息,全部包含在这 53165 个主题中。以人为灾害为例,GDELT 数据库中目前已经识别出包括交通事故、列车出轨、房屋着火、建筑倒塌、桥梁坍塌、瓦斯爆炸、管道爆炸等 120 种类型,这样就相应地形成 120 个有关人为灾害的 0/1 变量。显然,我们不可能对 120 种人为灾害的阶层分布一一列举,而只能根据主成分分析方法,对这 120 种人为灾害的信息提取主成份,以最主要的因子作为人为灾害暴露指数,然后进行阶层比较。其他压力暴露指数的计算方式类似,分析结果见表 1^①。

表 1 显示:在 6 项压力指标上,农民、非技术工人、技术工人阶层,其暴露指数远高于其他阶层。故而,研究假设 2 关于压力暴露的阶层分布假说得到检验。

2. 风险与压力易感性的阶层差异。压力-过程模型不但强调分析压力暴露因素,而且还分析面临压力时,易感性机制对于负面情绪唤起的阶层差异。那么不同类型的压力与风险,对不同阶层的负面情绪唤起程度是否存在显著性差异呢?

数据分析结果大致证明了压力易感性的阶层差异。面对同样的风险与压力,压力与风险在唤起不同阶层的负面情绪时,效应系数不一样。表 2 显示:所列举的 6 种压力因素中,不同阶层对积极情绪、消极情绪的作用机制与力度是不一样的。在所列举的 6 项压力指标中,冲突暴露指数、危机暴露指数、贫困暴露指数、失业风险指数等 4 项指数对普通阶层的负面情绪的激化效应和对正面情绪的消减效应,都明显超过其他阶层。而剩下的人为灾害暴露指数和自然灾害暴露指数,即天灾人祸在骨干阶层的易感性更强,这个分析结果与研究假设 2 略有出入。

综上所述,压力的阶层暴露差异以及易感性差异已经被大致证明。首先,6 项压力指数在普通阶层的暴露水平明显高于骨干阶层,压力暴露的阶层分布假说成立;其次,在 6 项压力指数中,有 4 项指数符合压力易感性的阶层分布假说,而剩下的自然灾害暴露指数和人为灾害暴露指数则主要以压力暴露机制对情绪状态的阶层分化发挥作用,易感性机制的作用方向与研究假设不符合。

^① 对具体编码规则感兴趣的读者,可以参考 GDELT 数据库中概率主题模型的提取文件说明。也可向作者索取。

表 1 各阶层的压力、风险与危机暴露指数比较^①

阶层	自然灾害 暴露指数	人为灾害 暴露指数	危机暴露 指数	冲突暴露 指数	失业风险 指数	贫困暴露 指数
官员	-0.10	-0.05	-0.09	-0.21	0.169	-0.033
企业家	-0.19	-0.09	-0.32	-0.23	0.190	-0.003
知识分子	0.01	-0.04	-0.07	-0.01	0.144	-0.129
技术人员	-0.11	-0.01	0.02	-0.12	0.162	-0.076
小业主	-0.13	-0.08	-0.16	-0.26	0.215	-0.049
办事人员	-0.06	-0.03	0.03	0.14	0.194	-0.022
技术工人	-0.02	0.07	0.09	-0.04	0.278	0.041
非技术工人	0.01	-0.02	0.19	-0.03	0.306	0.041
农民	0.54	0.70	0.31	0.14	0.233	0.574

表 2 不同阶层的压力易感性比较 (OLS 回归)

压力因素	压力因素对各阶层消极情感指数的影响系数 (Beta)			压力因素对各阶层积极情感指数的影响系数 (Beta)		
	普通阶层	中间阶层	骨干阶层	普通阶层	中间阶层	骨干阶层
冲突暴露指数	0.482 **	0.437 **	0.357 **	-0.136 **	-0.130 **	0.002
危机暴露指数	0.465 **	0.258 **	0.239 **	-0.167 **	-0.095 **	-0.010 **
贫困暴露指数	0.160 **	-0.013	-0.022	-0.108 *	-0.023	-0.060
人为灾害暴露指数	0.062	0.029	0.022	-0.080 *	-0.101 *	-0.132 *
自然灾害暴露指数	0.079	0.076	0.014	0.011	-0.058	-0.139 **
失业风险指数	0.122 **	0.059 *	0.008	-0.090 *	-0.051	-0.064
模型拟合度:R square	0.684	0.579	0.597	0.293	0.270	0.211

说明:分别以消极情感指数、积极情感指数为因变量,以 6 种压力与危机暴露指数为自变量进行回归分析,上表列举的是回归方程在不同阶层中运行时得出的不一样的回归系数,通过比较不同阶层中回归系数的大小,测量 6 种压力对不同阶层的总体情绪指标的影响,进而比较这些压力在不同阶层中的影响力度。**, * 分别表示回归系数的显著性为 P<0.1 和 P<0.05。

五、阶层互动、场域与情感氛围

一般而言,新闻案例中的事件,很少只涉及到一个阶层,而往往有多个阶层同时出现,这时,这个事件实际上就是两个或者多个阶层之间的互动,或者说两个阶层遭遇时的场景。如果事件中涉及到两个或多个阶层,那么它反映的是与这些阶层有关的互动场景。这种互动,可以是冲突、合作,也可以是群体性事件,具体的互动类型涵盖了各个阶层之间所有的互动类型,互动过程中的情感性成份,则构成两个阶层或者多个阶层互动情境中的情感氛围。

在微观个案层面,这种阶层互动的情感氛围随机性非常强,阶层互动场景多种多样。但是将这些海量的情境汇聚起来,对场景进行结构化,按照阶层之间互动类型对这些海量个案进行分类,然后看不同阶层互动时的情绪氛围的总体状况,并据此折射当前的阶层关系状况:如果阶层

^① 表中 6 项压力指数的计算都达到因子分析对提取主成分的基本要求,最大公因子的载荷都具有理想的表现。限于篇幅,在此不展开。

互动更多地是合作、友好、融合,那么这种阶层互动类型下积极情感指数就会相对更高;如果阶层互动更多地是冲突、对立与排斥,那么这种阶层互动类型下的消极情感指数就会更高。

我们对中国社会各阶层之间互动的“情感氛围”以矩阵的形式反映出来,借此透视中国社会各阶层之间的关系(见表3)。

表 3 中国社会阶层互动“情感氛围”矩阵

	官员	企业家	知识分子	技术人员	小业主	办事人员	技术工人	非技术工人	农民
官员		0.47	-0.04	0.22	0.21	-0.22	-0.05	-0.09	-0.20
企业家	0.47		0.05	0.34	0.34	-0.01	0.17	-0.02	-0.18
知识分子	-0.04	0.05		0.54	0.29	-0.10	0.29	-0.15	-0.34
技术人员	0.22	0.34	0.54		0.66	0.33	0.58	0.33	0.70
小业主	0.21	0.34	0.29	0.66		2.60	0.46	-0.14	-0.74
办事人员	-0.22	-0.01	-0.10	0.33	2.60		-0.83	-0.57	-0.36
技术工人	-0.05	0.17	0.29	0.58	0.46	-0.83		-0.19	-0.93
非技术工人	-0.09	-0.02	-0.15	0.33	-0.14	-0.57	-0.19		-0.24
农民	-0.20	-0.18	-0.34	0.70	-0.74	-0.36	-0.93	-0.24	

说明:表格中的数据为综合情感指数。

(一)官员阶层与各阶层互动的情感氛围

这里最有趣的现象在于,官员阶层与企业家之间的互动,其综合情感指数远高于官员阶层与其他阶层。这说明,中国社会最主要的精英阶层之间的友好合作要大于对立冲突,这有助于社会结构的稳定;另外,这也有可能是“精英联盟”正在形成的征兆,是社会结构演变过程中需要注意的趋势。

表3还显示,官员阶层与农民阶层关系较为紧张。官员与农民的冲突,某种程度上是政府与农民关系的一种表现形式。政府与农民关系最为紧张的时期可能是农业税费改革前夕,当时,乡镇官员与村干部在收取农业税费和实施计划生育政策时与农民的矛盾非常激烈,形成了所谓的“治理性危机”^①。而农业税费改革之后,基层政权“悬浮”,国家与农民关系发生历史性变迁,那么国家与农民关系是否得到改善呢?从阶层比较的角度来看,我们发现官员与农民之间的紧张关系,依然是阶层关系紧张的主要表现形式。近年来,政府与农民的冲突,不再是源于计划生育与收取农业税费,而是城市化进程中由地方政府推动的征地拆迁,以及为了获取建设用地指标的各种所谓“圈地运动”,比如“整村推进”“农民上楼”等。已有大量分析显示,暴力拆迁引发大量官民冲突,引发的群体性、抗争性事件不计其数。这些冲突性事件是当前引发官-民关系紧张的主要原因,官-民冲突主要发生在上述场域。

(二)企业家与各阶层互动的情感氛围

企业家阶层与官员阶层互动的情感氛围最好,企业家与技术人员、小业主的关系都还不错;而企业家与非技术工人阶层、农民阶层之间的互动关系总体紧张。企业家与不同阶层互动场域的这样一种情感氛围,总的来说,反映了这样一个规律:企业家与同样掌握稀缺资源的阶层之间更多的是友好合作关系,而与其他阶层之间则更多的是一种紧张冲突关系,这种紧张关系在现实生活中表现为企业家与工人阶层之间的“劳资冲突”,以及资本下乡、征地拆迁过程中围绕土地

^① 李昌平:《我向总理说实话》,光明日报出版社,2000年版。

问题而展开的诸多博弈。对于转型社会而言,在社会结构重组与利益结构分化的过程中,这是很正常的,问题的关键在于如何减少社会结构转型过程中的“阵痛”,以实现社会的良性运行。

(三)中国社会阶层互动情感氛围矩阵分析

从阶层之间互动的情感氛围矩阵来看,阶层互动的情感氛围呈现出如下特征:骨干阶层互动场景中的情感氛围总体上呈现积极态势,而其与办事人员、技术工人、非技术工人、农民互动时则更偏向消极情绪氛围;技术人员、小业主、技术工人与骨干阶层互动时偏向积极情绪,与非技术工人、农民互动时则明显偏向消极情绪;非技术工人与农民阶层和其他各阶层的互动时明显偏向消极负面。

上述阶层之间互动场域中的逻辑大致在于:

第一,当前中国社会阶层关系紧张的压力来源于普通阶层。对社会财富分配不均、群体生存状况改善乏力,以及社会利益结构重构过程中利益相对受损等的不满,带来其与其他阶层的紧张性关系。

第二,技术人员阶层与其他各阶层之间的关系普遍较好,小业主阶层与非技术工人、农民冲突较多,与骨干阶层关系积极良好;办事人员阶层则与其他阶层的关系都比较紧张,而与中间阶层的关系相对较好。

第三,中国社会的骨干阶层之间关系相对“融洽”。

六、总结与讨论

(一)社会情绪阶层分化的主要内涵

本研究发现,社会情绪的阶层分化,既表现在总体性情绪指标上,也表现在具体性情绪类型的差异上。伴随着财富集中,社会风险向下集中,抑郁、不满、绝望等负面情绪具有向下聚集的趋势。社会情绪的阶层分化是社会转型的必然结果。不同阶层中突出性负面情绪问题差异明显,在情感治理进程中,应该结合不同阶层所存在的突出性情绪问题,不同情绪的具体形成机制进行分类治理。

负面情绪的阶层分化,一方面与压力暴露的阶层分布有关;另一方面也与压力易感性的阶层差异有关系,社会经济地位低的阶层与社会经济地位高的阶层相比,更容易被不良事件影响情绪,前者面对大部分压力事件时表现出更高的易感性。这样,压力暴露机制和易感性机制的阶层分布叠加在一起,构成社会情绪的阶层分化现象。

(二)不同的阶层互动场景中的情感氛围差异显著

中国社会正在发生快速的社会结构转型,“社会结构转型”被认为是“另一只看不见的手”。理解阶层关系是分析社会结构的关键,而如何理解阶层关系,尤其是社会转型时期的阶层关系,在已有的研究方法中只能借助问卷调查或者基于个案的归纳等方法,这些方法对于透视具有宏观特征的阶层关系力有不逮。本文借助互联网大数据,将媒介信息中记录的有关阶层互动的场景进行结构化,通过分析不同阶层互动场景中的“情感氛围”,借以透视各阶层之间的关系。

研究发现,官员、企业家、知识分子之间更多的是资源共享与互利互补的关系,而官民、劳资等阶层的互动关系则显得相对紧张,阶层互动中更容易产生冲突性事件。这些分析结果与社会学研究中有关“精英结盟”的判断相契合。

(三)本文在研究方法上的创新

本文借助于云计算的强大信息处理能力,对海量的互联网大数据进行分析,以透视中国社会

运行中情感的阶层分布特征。这是“计算社会科学”的一种初步实践。

从数据的属性来看,GDELT 数据库主要由主流媒体、网络新闻、网络论坛等不同来源信息构成。那么该数据是否适用于分析中国社会的阶层以及情绪状态呢?

传播学研究认为,媒介数据,无论是社交媒体数据,还是门户网站与网络新闻数据,都是一种“拟态真实”。在媒体高度发达的社会条件下,媒介所建构的拟态环境越来越接近真实环境。在 web2.0 时代,媒体已覆盖和渗透到现代社会生活的方方面面,成为人们的必备工具。快速、高效、便捷的媒介手段不仅可以为人们实时提供真实环境的动态信息,而且这种媒介所建构的拟态环境越来越接近真实环境,可以使受众具有身临其境的现场感。除此之外,拟态环境与现实环境之间的关系也在不断深化,表现为拟态环境的现实化。“拟态现实”在客观、真实、能动地反映现实环境的基础上,对现实环境发展变化又起着引导、转化和塑造作用^①。正是基于上述考虑,本研究认为,一个具体的新闻案例充满了建构性、选择性以及媒介框架的运用,但是对于海量的个案数据而言,因为信息来源多种多样,表述者的类型多种多样,以及媒介高度发达,这使得在较长跨度的历史过程中积累的数据与现实生活越来越接近,故而通过 GDELT 数据库反映现实生活的一些方面,具有较强的可行性。大数据与小数据的结果比对,以及不同来源的数据结果比对具有一致性,为这一判断提供了注脚。

作者:龚为纲,武汉大学社会学系(武汉市,430074)

朱萌(通信作者),湖北省农业科学院农业经济技术研究所(武汉市,430064)

(责任编辑:林立公 王炳权)

^① 曹劲松:《论拟态环境的主体建构》,《南京社会科学》,2009 年第 2 期。