

危机情景下群体情感表达的动力学机制研究

安 宁¹, 安 璐^{1,2}

(1. 武汉大学 信息管理学院, 湖北 武汉 430072; 2. 武汉大学 信息资源研究中心, 湖北 武汉 430072)

摘 要:【目的/意义】探索危机情景下的群体情感表达的动力学机制, 对于危机情境下的网络舆情管理具有重要的理论价值与现实意义。【方法/过程】研究获取与“群体免疫”相关的微博数据, 利用 SKEP 模型计算每日情感倾向和情感值以构建情感时间序列, 对情感序列的平稳性、纯随机性以及混沌性进行分析。【结果/结论】研究结果表明, 舆情是一个过程系统, 群体情感表达是该过程系统的一种连续过程。在不同阶段群体情感表达呈现迥异的动力学特性, 起始期的情感序列是由一个随机系统产生的; 爆发期主导群体情感表达机制的是一个混沌系统; 平稳期群体情感表达过程是一个二阶马尔可夫过程, 是由一个稳定的非混沌系统主导情感的表达。【创新/局限】本研究挖掘了危机情境下群体情感表达的动力学机制, 丰富了理论视角。但本研究仅针对危机情境展开, 在未来研究中将进一步研究比较危机情境下与其他情境下群体情感表达动力学机制的差异。

关键词: 群体情感表达; 时间序列分析; 动力学机制; 过程系统; 危机情景

中图分类号: G250.2 **DOI:** 10.13833/j.issn.1007-7634.2022.01.020

1 引 言

近年来, 世界各地的自然灾害、事故灾害等多种危机事件逐渐呈现出日益高发、规模扩大、危害程度加深等态势^[1-2]。这些具体的危机事件与社会情境的综合作用产生了特定的危机情境^[3]。正如危机管理专家 Mitroff^[4]所指出: “危机已经成为一种自然现象, 它不再是社会异常的、罕见的、任意的或者外围的特征, 它根植于当今社会的经纬之中”。不同于其他情境, 危机情境是人类生活与社会秩序受到严重威胁的状态, 通常伴随不确定性、破坏性、聚焦性等特征。在危机情境下, 社交媒体在信息传播方面发挥至关重要的作用^[5]。例如, 在 COVID-19 肆虐全球之际, 社交媒体成为公众信息交流、情感表达和维持社会化关系的重要渠道, 被列为访问紧急信息的第四大最受欢迎来源^[6]。习近平总书记强调“要把网上舆论工作作为宣传思想工作的重中之重来抓, 坚持正能量是总要求、管得住是硬道理、用得好是真本事”^[7]。危机情境下的网络舆情是公众在危机期间信息表达行为的结果。因此, 为高效开展网上舆论工作, 明晰危机情境下群体信息表达的行为规律和生成机制等动力学信息尤为重要。

情感是人类认知的重要组成部分, 情感的传播将不断促进公众在危机情境下的情感共振与话语协同, 影响着公众对

危机事件本身的思考与解读^[8]。从某种层面来说, 情感是危机情境中网络舆情演化的重要动力^[9], 也是舆情系统的一种表征量。基于系统论的思想, 决定系统长期演化的任一变量的时间演化均包含了系统所有变量长期演化的信息^[10]。也就是说, 研究可以通过对单一变量时间序列来研究系统的动力学机制。有鉴于此, 本研究以新浪微博作为基础信息源, 以情感为切入口, 对情感时间序列进行构建与分析, 探究危机情境下群体表达蕴含的动力学信息。本研究深入挖掘舆情系统中群体情感表达的动力学机制, 以期进一步丰富对群体信息表达机制的研究与理论, 同时为管理部门在对危机事件网络舆情进行引导与决策时提供参考。

2 文献回顾

2.1 危机情境下网络舆情的情感研究

社交媒体的发展为危机情景下公众的互动提供了渠道, 网络舆情的发生机制与演化过程往往存在一定的规律性, 可将危机事件网络信息传播的生命周期划分不同阶段^[11], 舆情传播用户和舆情主题等方面也将会影响突发事件网络舆情的演变过程^[12]。其中, 情感不仅是网络舆情演化的重要动力, 也是人类认知的重要组成部分。情感的表达会产生一定的价值, 其本身也是一种价值的体现^[13]。自 Picard^[14]首次

收稿日期: 2021-10-19

基金项目: 国家自然科学基金面上项目“危机情境下网络信息传播失序识别与干预方法研究”(72174153); 国家自然科学基金重大课题“国家安全大数据综合信息集成与分析方法”(71790612); 国家自然科学基金创新研究群体项目“信息资源管理”(71921002)。

作者简介: 安宁(1993-), 女, 吉林长春人, 博士研究生, 主要从事危机信息管理、信息行为研究; 安璐(1979-), 女, 湖北武汉人, 教授, 博士生导师, 主要从事网络数据分析、应急情报研究, 通讯作者。

提出“情感计算”的概念之后,学者开始关注有关情感、情感源和情感传播等方面的计算研究。社交媒体利用带有快速和短消息的认知原理使其内容更多带有情感色彩^[15],而危机事件通常易于造成公众情绪波动,在此情境下公众情感不仅影响突发事件网络舆情信息的传播,还会造成情感极化现象^[16]。这一现象的产生主要是一方面由于公众在突发事件同一主题下所表达的情感具有协同性^[17],同类事件往往会引起相似的情绪表达^[18];另一方面性别、年龄、学历等特性相似的用户对突发事件情感倾向接近^[19]。目前,研究者多基于舆情评论文本信息挖掘公众情感倾向,如安璐^[20]等人指出结合评论回复的情感支持,将有助于识别微博评论中有影响力的意见领袖。

2.2 人类动力学研究

人类动力学是一个跨学科的研究主题^[21-23],旨在通过借鉴物理的方法和系统的思想对人类行为进行定量研究,从而发现人类行为统计规律及其生成机制^[24]。2005年,Barabási^[25]研究表明人类行为发生时间服从幂律分布,并引发学术界对人类动力学研究兴趣的新浪潮。人类动力学成为研究人类行为的新范式^[26],逐渐被引入到舆论的形成^[27]、传播^[28]以及调控^[29]等方面的研究中。例如,刘海鸥^[30]等人基于人类动力学研究视角构建了在线社交用户的舆情画像模型,对在线社交用户信息传播行为特征进行了实证分析。事实上,舆论的形成与传播是公众在社会中互动与交流的结果^[31],公众过去的意见、其邻域意见以及主题的上下文信息都对舆论观点的形成和传播都具有重要影响^[32]。目前,学者们通常基于 Ising 模型^[33]、Sznajd 模型^[34]、Deffuant 模型^[35]、Hegselmalm-Krause 模型^[36]等舆论动力学模型来探究舆论的动力学机制。动力学机制是系统运行的规律与机理,是使系统保持平衡和动态演化的能力^[37]。危机事件爆发后,尤其是在社交媒体回声室中极易造成群体极化等舆情衍生效应^[38]。针对网络空间舆论场,当情景与危机事件的话题关联度越大时,情景则会具有较强的主题提示作用,使用户的群体性思维易于集中于同一方向,从而在特定主体上形成共同认识^[39]。通过情境推演有助于对重大突发事件整体安全态势的理解、解释和预测^[40]。近年来,学者采用知识元、本体、系统动力学模型等不同方法对危机情境进行了描述^[41],逐渐发现掌握危机情境下的网络舆情发展的动力学机制尤为重要。例如,赵玲^[42]等人探究了微博对群体性事件助燃的动力学机制;虞铭明^[43]等人分析了环境群体性事件的网络舆情扩散动力学机制。

通过对相关文献进行梳理发现,关于危机情景下的网络舆情情感研究方面已取得了较为丰富的成果,多集中于探索舆情情感的极性、传播以及影响因素等方面。随着人类动力学理论与方法的发展,学者们逐渐将其引入至舆情研究中,但鲜有挖掘危机情境下群体情感表达背后所隐藏的动力学机制。因此,本文基于特定危机事件下的网络舆情信息构建情感时间序列,分析情感时间序列的平稳性、纯随机性以及

混沌特征,挖掘危机情景下群体情感表达的动力学机制。

3 研究方法

3.1 研究框架

本研究的总体框架如图1所示,主要包括对危机情境下舆情信息收集、情感倾向计算以及情感表达动力学机制挖掘三个部分。首先在信息收集过程中,研究主要以新浪微博平台为数据源,获取与“群体免疫”相关的博文及评论内容,通过对信息进行预处理构建本研究的语料集。其次在情感倾向计算方面,研究基于情感知识增强的情感预训练算法^[44](Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis, SKEP)分析文本的情感倾向,计算每日情感值以构建情感时间序列值。同时,研究根据微博数量和危机传播四阶段理论^[45]对危机情境下的网络舆情演化进行阶段划分。最后,研究检验在不同阶段中情感时间序列的平稳性和纯随机性,对情感序列进行混沌性分析与建模,进而挖掘危机情境下的群体情感表达的动力学机制。

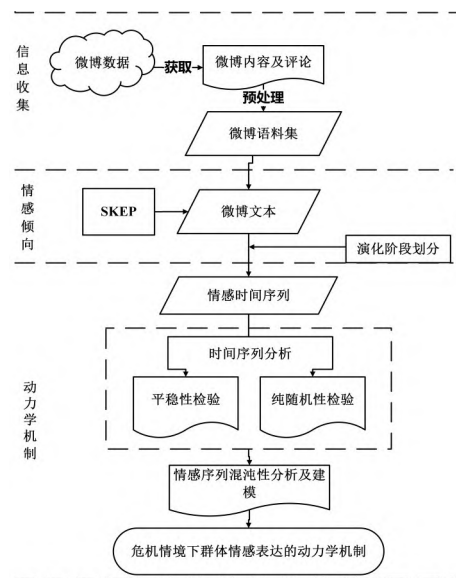


图1 研究框架

Figure 1 Framework of research

3.2 情感时间序列构建

危机事件的爆发通常会引起群体情绪的波动,呈现一定的情感倾向。因此,研究需要对舆情文本信息进行情感倾向分析与计算,获得每日情感倾向值以构建情感时间序列,进而把握公众对危机事件的情感倾向变化。研究利用百度研究团队在2020年提出的基于情感知识增强的情感预训练算法 SKEP(Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis),对网络舆情信息进行情感倾向计算。SKEP为各类情感分析任务提供统一且强大的情感语义表示,可用于句子级情感分析任务预测。实验表明^[44],以通用预训练

模型 ERNIE 作为初始化, SKEP 在 14 项中英情感分析典型任务上具有较好表现。

研究在基于 SKEP 算法进行情感倾向计算时, 首先以“日”为采样频率, 将舆情文本内容切割成句, 并基于统计方法从大量无标记的微博舆情数据中自动挖掘情感词、情感词极性以及观点搭配等情感知识。然后, 基于所挖掘的情感知识, 对原始输入句子中的部分词语进行屏蔽, 即替换为特殊字符【MASK】。最后, 设计情感优化目标, 复原被屏蔽的情感信息, 进而计算每条微博内容的正、负情感倾向得分。

研究获得每日每条微博内容的正、负情感倾向得分后, 取全部微博内容平均得分作为当日的正、负向情感值; 并以正向情感减去负向情感作为当日情感倾向, 进而构建情感时间序列 $\{X_t, t \in T\}$, 其中 X_t 为每日情感值, 取值范围为 $[-1, 1]$ 。情感值越高, 则表明情感正向倾向越大; 反之, 则负向倾向越大。

3.3 时间序列预处理

研究基于上述方法构建危机情境下群体情感时间序列, 通过对情感时间序列进行观察、研究, 有助于发现危机情境下公众情感变化的统计规律, 同时可以从情感时间序列数据中提取和恢复舆情系统高维空间下的轨迹规律^[10]。因此, 在获得情感时间序列的观测值后, 研究首先对时间序列进行预处理, 进而初步检验其系统内部的平稳性和纯随机性特征。

3.3.1 平稳性检验

平稳性作为时间序列的一种特征, 强调的是时间平移的不变性。若时间序列观测值具有平稳性, 则说明生成该序列的随机过程是一个平稳随机过程。因此, 研究获得危机事件舆情情感时间序列后, 需根据变量的收敛情况而判断其是否为平稳时间序列。通常, 时间序列平稳性有严平稳与宽平稳两种定义, 其中严平稳序列要求其联合概率分布族不随时间平移而变化。由于在实际研究中无法通过微博语料测算情感时间序列的联合概率分布, 因此研究提到的平稳序列均为宽平稳时间序列。

对于情感时间序列 $\{X_t, t \in T\}$, 若满足以下三个条件:

$$\left\{ \begin{array}{l} \text{对于任意 } t \in T, EX_t^2 < \infty; \\ \text{对于任意 } t \in T, EX_t = \mu; \\ \text{对于任意 } t_i, t_j, t_m \in T, \text{ 且 } t_i + t_j + t_m \in T, \\ \gamma(t_i, t_j) = \gamma(t_m, t_m + t_j - t_i) \end{array} \right. \quad (1)$$

则情感时间序列为宽平稳序列。其中, EX_t 为序列 $\{X_t, t \in T\}$ 在 t 时刻的均值函数, μ 为常数, $\gamma(t_i, t_j)$ 为序列 $\{X_t, t \in T\}$ 的自协方差函数。

研究结合图检验和单位根检验两种方法对情感时间序列的平稳性进行判断。其中, 图检验是基于描述性统计方法, 通过观察时序图与自相关图, 分析时间序列的趋势、周期波动以及短期相关性等性质, 对时间序列的平稳性做出初步

判断。图检验方法能够有效识别非平稳性较为显著的时间序列, 而对于平稳特征不明显的时间序列则需进一步利用单位根检验进行判断。本研究中单位根检验使用 ADF 作为检验统计量, 假设时间序列 x_t 的确定性部分可以由过去 p 时期的历史数据描述, 即:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \xi_t \quad (2)$$

其中 ξ_t 为序列的随机部分, $\xi_t \sim N(0, \sigma^2)$, $\phi_1, \phi_2, \dots, \phi_p$ 为序列的回归系数。构建参数 ρ 为:

$$\rho = \phi_1 + \phi_2 + \dots + \phi_p - 1 \quad (3)$$

若时间序列非平稳, 则存在特征根, 序列的回归系数之和为 1, 则 ρ 等于 0; 若时间序列平稳, 则不存在特征根, 序列的自回归系数之和小于 1, 则 ρ 小于 0。

由此, 构造 ADF 检验统计量:

$$\tau = \frac{\hat{\rho}}{S(\hat{\rho})} \quad (4)$$

其中 $S(\hat{\rho})$ 为 $\hat{\rho}$ 的样本标准差。检验的原假设为时间序列非平稳, 备择假设为时间序列平稳。若统计量的 P 值小于显著性水平, 则拒绝原假设, 判定序列为平稳序列。

3.3.2 纯随机性检验

宽平稳时间序列具有二阶矩平稳的性质, 但是并不能由此判断情感时间序列中的情感值之间是否具有密切相关性。纯随机时间序列(如白噪声序列)虽然表现出显著平稳, 但事实上其序列值之间不具有相关性, 无法根据历史数据对其进行建模。因此, 研究需要对情感时间序列进行纯随机性检验, 以判断是否有进一步分析的价值。根据 Bartlett 定理^[46], 若情感时间序列具有纯随机性, 则序列的延迟非零期样本自相关系数将近似服从均值为 0、方差为观测期倒数的正态分布, 即

$$\hat{\rho}_k \sim N\left(0, \frac{1}{n}\right), \forall k \neq 0 \quad (5)$$

其中, $\hat{\rho}_k$ 为自相关系数, k 为延迟期数, n 为序列观察期数。由此, 可以构造 LB 统计量^[47]:

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right) \quad (6)$$

其中, n 为序列观察期数, m 为指定延迟期数。同时提出检验假设, 原假设: 延迟期数小于或等于 m 期的序列值之间相互独立; 备择假设: 延迟期数小于或等于 m 期的序列值之间有相关性。若 LB 统计量的 P 值小于显著性水平, 则拒绝原假设, 判定该序列为非纯随机序列。

3.4 混沌时间序列分析

混沌作为非线性动力系统的重要特性, 其内在随机性会使得确定性系统表现出类似随机的行为, 即具有混沌性质的系统的部分变量经常会以离散的时间序列形式被观察到^[48]。因此, 为了进一步探究舆情系统随机性产生的本质, 研究将进一步利用最大李雅普诺夫指数对情感时间序列的混沌性进行判断, 以期深入理解产生该序列观察值的随机过程的演化机理。主要步骤如下:

3.4.1 计算序列平均周期

对时间序列 $\{X_t, t \in T\}$ 做傅里叶变换:

$$y_{k+1} = \sum_{t=0}^{n-1} \omega^{tk} x_{t+1} \quad (7)$$

其中, $\omega = e^{-2\pi i/n}$, i 为虚数单位。对于 x, y , 索引 t 和 k 的取值范围为 0 至 $n-1$ 。由此得到时间序列频率, 进而计算得到序列平均周期。

3.4.2 相空间重构

相空间是一个用以表示系统全部可能出现的状态的空间, 其中的每个点都对应着系统的某个状态, 相空间重构则是根据某一变量的观察序列近似恢复出原动力系统的技术。对于混沌时间序列而言, 相空间重构能够将代表原动力系统演化规律的混沌吸引子复现出来。给定时间序列 $\{X_t, t \in T\}$, 重构相空间 $\{X_t^\tau\}$ 为:

$$\{X_t^\tau\} = (x_t, x_{t+\tau}, \dots, x_{t+(m-1)\tau}) \quad (8)$$

其中 τ 为延迟时间, m 为嵌入维数, $i = 1, 2, \dots, N - (m-1)\tau$, N 为序列长度。显然, 相空间重构需要预先设定延迟时间 τ 与嵌入维数 m 这两个重要参数。本研究利用平均交互信息算法^[49] (Average Mutual Information) 估计相空间重构的延迟时间 τ , 计算方法如下:

$$AMI(\tau) = \sum_{i=1}^N p(x_i, x_{i+\tau}) \log_2 \frac{p(x_i, x_{i+\tau})}{p(x_i)p(x_{i+\tau})} \quad (9)$$

其中, τ 为向量 $(1, 2, \dots, [\text{最大延迟}])$, 以 $AMI(\tau)$ 的第一个局部极小值所对应的延迟时间 τ 作为重构相空间的延迟时间。

研究利用伪最近邻算法^[50] (False Nearest Neighbor) 估计嵌入维数 m 。在维数 m 下, 如果:

$$\sqrt{\frac{R_i^2(m+1) - R_i^2(m)}{R_i^2(m)}} > O \quad (10)$$

则重构相空间 $\{X_t^\tau\}$ 中的点 x_i 与其最近邻点 x_{i^*} 是伪最近邻点。其中, $R_i^2(m) = \|x_i - x_{i^*}\|_2^2$, O 为阈值。对于维数 m , 计算伪最近邻点数与重构相空间中全部点数的比值, 以使得该比值达到最小值的 m 作为重构相空间的嵌入维数。

3.4.3 混沌性判定

李雅普诺夫指数能够度量因微小初值误差而在相空间中产生的两条相邻轨迹随时间迭代发生分离的程度。李雅普诺夫指数均为负值, 则相邻轨迹趋向于平衡; 李雅普诺夫指数存在正值, 相邻轨迹将远离平衡, 系统出现混沌现象。因此, 本研究利用小数据量算法^[51], 通过计算最大李雅普诺夫指数判断时间序列的混沌性。具体步骤如下:

(1) 寻找重构相空间 $\{X_t^\tau\}$ 中的每个点 x_i 与其最近邻点 x_{i^*} 。由于小数据量算法限制短暂分离, 最近邻点在时间上至少相差一个平均周期, 即 $|i - i^*| > \text{序列平均周期}$ 。

(2) 计算重构相空间 $\{X_t^\tau\}$ 基本轨道上每个点的平均发散率:

$$y(j) = \left(\frac{1}{q\Delta t} \sum_{i=1}^q \ln d_i(j) \right) \quad (11)$$

其中, $d_i(j)$ 为每个点 x_i 与点 x_{i^*} 的第 j 个时间步长后的距离, q 为非零 $d_i(j)$ 的数量, Δt 为采样周期。

(3) 利用最小二乘法对 $y(j)$ 作回归直线, 该直线的斜率即为最大李雅普诺夫指数。

3.5 ARIMA 建模

差分运算能够有效提取时间序列中蕴含的确定性信息。非平稳的时间序列经差分运算处理后有时能够表现出显著的平稳性质。因此, 为了判别情感时间序列是否为此情况下的非平稳序列, 研究利用求和自回归移动平均模型 $ARIMA(p, d, q)$ 对序列进行拟合, 表达形式如下:

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) (1 - L)^d x_t = \left(1 + \sum_{i=1}^q \theta_i L^i\right) \varepsilon_t \quad (12)$$

其中 x_t 为序列值, ε_t 为随机干扰, ϕ_i 为自回归项系数, θ_i 为移动平均项系数, p 为自回归阶数, q 为滑动平均阶数, d 为差分运算阶数, L 为滞后算子。对于滞后算子 L , 有 $x_{t-i} = L^i x_t$ 。

以 $ARIMA(1, 1, 1)$ 为例, 模型表达为:

$$(1 - \phi L)(1 - L)x_t = (1 - \theta L)\varepsilon_t \quad (13)$$

其等价于:

$$x_t = (1 + \phi)x_{t-1} - \phi x_{t-2} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (14)$$

最后, 利用 LB 统计量对模型拟合残差项进行检验。检验通过, 则说明模型拟合残差项为纯随机序列, 不残留相关信息, 证明模型拟合显著。

4 实证分析

4.1 实验数据

在新冠疫情的社会背景下, 研究以新浪微博为数据源, 于 2021 年 7 月 1 日以“群体免疫”为关键词, 检索新冠疫情事件的全部相关微博。经过去重、剔除无意义内容等数据清洗过程, 最终数据集包含 98,437 条原创微博、775,567 条转发微博以及 685,779 条微博评论, 共计 1,559,783 条, 涉及用户 458,701 人。研究数据的时间跨度为 2020 年 1 月 20 日至 2021 年 6 月 30 日, 共计 528 天。其中, 2020 年 1 月 20 日是在新冠疫情危机爆发后, 公众第一次在社交媒体上谈及群体免疫相关内容的起始时间。

4.2 情感时间序列构建

研究根据上述情感时间序列构建方法, 计算微博情感倾向, 获得 2020 年 1 月 20 日至 2021 年 6 月 30 日的情感时间序列, 具体如表 1 所示。与此同时, 研究以时间为横坐标, 分别绘制每日微博数量及情感倾向值, 如图 2 所示。其中图 2(a) 纵坐标为微博数量, 以对数坐标展示; 图 2(b) 纵坐标为当日情感倾向值。

根据危机传播四阶段理论, 网络舆情一般分为起始期、爆发期、平稳期和衰退期四个生命周期阶段^[45]。虽然本研

究关于“群体免疫”的微博内容时间跨度长,涉及用户多,有别于一般的突发事件网络舆情,但图2(a)中微博数量仍然呈现出明显的不同发展阶段。根据图2(a)走势,可以将研究数据分为以下三个阶段:起始期(第1-52天)、爆发期(第53-164天)和平稳期(第165-528天)。其中,微博数量走势中未能发现对应“衰退期”的演化阶段,可能是随着新冠疫情的演变,公众关于“群体免疫”概念的认知正在持续发展,则当前数据尚不能观察到“衰退期”阶段。因此,研究将实验数据划分为三个阶段并不违背四阶段理论。同时,通过观察各阶段情感时间序列可以初步判断,不同演化阶段下情感倾向演化的趋势、周期等性质具有明显差异,说明随着危机情境下舆情的发展演化,推动群体情感表达的动力学机制可能不是一成不变的。为了证明这一研究假设,需要对各阶段情感时间序列做进一步分析以探究群体情感表达的深层机理。

表1 微博数量与情感序列

Table 1 The number of microblog posts and the emotional series

时间序列	微博数量	正向情感	负向情感	情感倾向
2020/1/20	3	0.342971867	0.657028122	-0.314056255
2020/1/21	1	0.223727416	0.776272625	-0.552545209
2020/1/22	2	0.261028963	0.738971058	-0.477942095
...
2020/3/12	431	0.287485113	0.712514893	-0.42502978
2020/3/13	14647	0.529689331	0.470310669	0.059378662
2020/3/14	48552	0.361141481	0.638858519	-0.277717038
...
2021/6/28	982	0.322752326	0.677247692	-0.354495366
2021/6/29	512	0.453615872	0.546384128	-0.092768256
2021/6/30	250	0.525224085	0.474775908	0.050448177

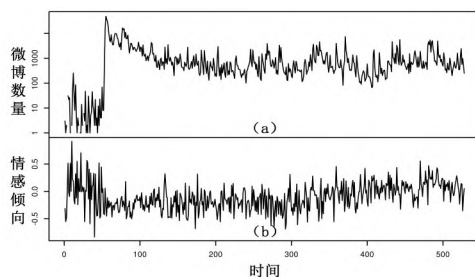


图2 各时间窗口下微博数量及情感值

Figure 2 The number of microblog post and emotional tendency value over time

4.3 情感时间序列预处理

研究根据微博量的走势将舆情演化分为三个阶段,分别对应起始期、爆发期和平稳期。各阶段时序图和自相关图分别如图3、图4所示。其中,图3的横坐标为时间,纵坐标为情感值;图4的横坐标为延迟时期数,纵坐标为自相关系数,虚线为2倍标准差;序号a,b,c分别对应起始期、爆发期和平稳期。

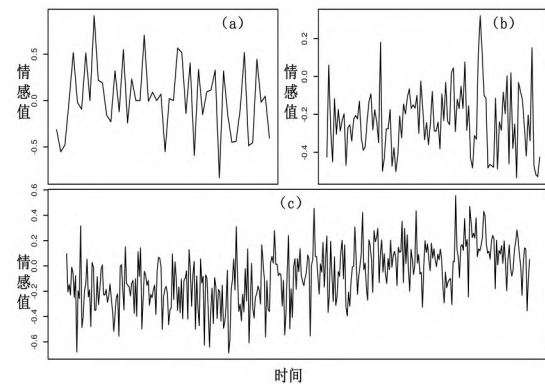


图3 各阶段情感时间序列的时序图

Figure 3 The emotional tendency value over time at different periods

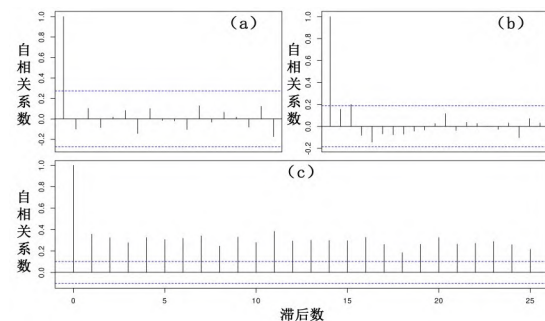


图4 各阶段情感时间序列的自相关图

Figure 4 The autocorrelation function of the emotional time series at different periods

由图4可以发现,平稳期的时间序列的自相关系数延迟25阶后仍然显著非零,其自相关系数长期位于零轴上方,表明该时期的情感时间序列显著非平稳。在起始期和爆发期阶段,情感时间序列的自相关系数在2倍标准差之内,自相关系数快速衰减至零,具备平稳时间序列特征。因此需要进一步对这两个阶段的情感时间序列进行单位根检验,结果如表2所示。

表2 起始期和爆发期的情感时间序列的单位根检验

Table 2 The unit root test of the emotional time series at the initial and outbreak periods

滞后数	起始期		爆发期	
	ADF统计量	P值	ADF统计量	P值
0	-7.750	0.010	-4.370	0.010
1	-4.840	0.010	-2.150	0.033
2	-4.460	0.010	-1.920	0.055
3	-3.670	0.010	-1.860	0.064

根据表2的结果可以发现,起始期 ADF 统计量的 P 值小于 0.01 水平,爆发期 ADF 统计量的 P 值大于 0.01 水平,说明起始期显著平稳,而爆发期不显著平稳。为了判断情感时间序列中的情感值之间是否具有密切相关性,研究进一步利用 LB 统计量对三个时期的情感时间序列进行纯随机性检验,结果如表3所示。

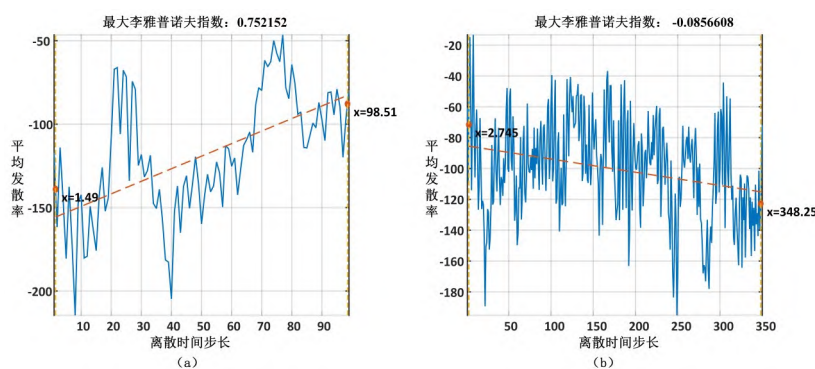


图5 爆发期和平稳期情感时间序列的最大李雅普诺夫指数

Figure 5 The largest Lyapunov exponent of the emotional time series at the outbreak and chronic periods

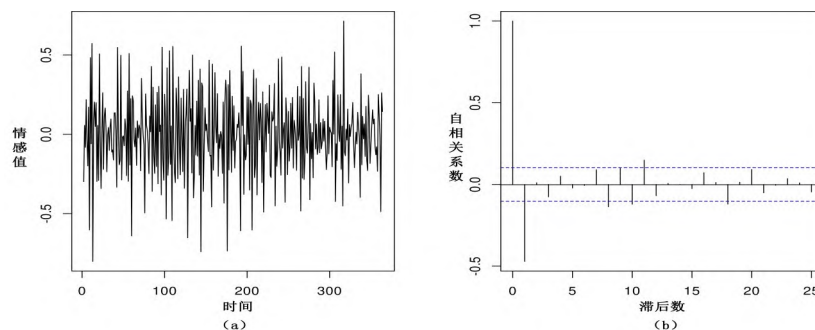


图6 一阶差分后的平稳期情感时间序列的时序图与自相关图

Figure 6 The value and autocorrelation function of the emotional time series over time after first-order difference

表3 各阶段情感时间序列的纯随机性检验

Table 3 The randomness of the emotional time series at different periods

滞后数	起始期		爆发期		平稳期	
	LB统计量	P值	LB统计量	P值	LB统计量	P值
5	1.973	0.853	11.229	0.047	188.965	0.000
10	4.577	0.918	12.992	0.224	361.544	0.000
15	6.602	0.968	15.149	0.441	549.460	0.000

表3中,起始期和爆发期的P值都大于0.01水平,说明两个阶段具有纯随机性;平稳期的P值小于0.01水平,说明第三阶段平稳期不具有纯随机性。因此,综合上述分析可以初步说明,起始期的情感时间序列是平稳随机序列,情感序列值彼此之间没有任何相关性,不具备进一步分析和建模的价值。事实上,这一时期通常作为舆情演化的初始阶段,群体参与少,舆情热度低,群体情感表达往往具有较高的随机性,说明该阶段情感序列是由一个主要受随机因素影响的动态系统产生的。同时,爆发期的情感时间序列不显著平稳且具有较高随机性,而平稳期情感时间序列非平稳且不具有纯随机性。在爆发期阶段,舆情热度通常会在较短时间达到峰值,群体参与热情迅速攀升;在进入平稳期后,舆情热度降低,舆情演化保持相对稳定,但群体参与的累计值较高。爆发期与平稳期作为群体观点和情感的主要形成阶段,蕴含着极为丰富的动力学信息。从目前的分析结果来看,有两种系统能够输出类似爆发期和平稳期的情感时间序列观测值:一

种是由确定性与随机性叠加而成的系统,其随机性主要来自于外部的随机作用;另一种是混沌系统。因此,为了深入探究在爆发期和平稳期阶段群体情感表达的动力学机制,研究需要进一步对这两个时期的情感时间序列的混沌性进行分析。

4.4 情感时间序列混沌性分析及建模

混沌系统是一种确定性系统,但由于具有内在随机性往往会表现出貌似无序的伪随机运动状态。因此,为更加明晰爆发期和平稳期群体情感表达的动力学机制,研究需进一步对这两个时期的情感时间序列进行相空间重构,利用最大李雅普诺夫指数判断情感序列是否具有混沌特征,结果如图5所示。

表4 一阶差分后的平稳期时间序列平稳性和纯随机性检验

Table 4 The unit root test and randomness of the emotional time series after first-order difference

单位根检验			纯随机性检验		
滞后数	ADF统计量	P值	滞后数	LB统计量	P值
0	-31.8	0.010	5	84.584	2.2e-16
1	-21.6	0.010	10	104.02	2.2e-16
2	-18.9	0.010	15	114.34	2.2e-16
3	-14.5	0.010	20	125.23	2.2e-16

根据图5(a)可知,爆发期情感时间序列的最大李雅普诺夫指数为正数,表明爆发期情感时间序列具有混沌特征。

也就是说,在爆发期阶段的舆情系统本质上是一个混沌系统,群体情感表达的随机性在一定程度上是该系统混沌性的表现。图5(b)中平稳期情感时间序列的最大李雅普诺夫指数为负数,说明在这一阶段的情感时间序列不具有混沌特征,其背后的动力学系统稳定收敛。因此,可以利用常规方法对平稳期序列进行模型拟合,即首先对序列做一阶差分,提取确定性信息,然后对差分得到的序列进行平稳性和纯随机性检验,其结果如图6和表4所示。

根据图6和表4的实验结果可以发现,一阶差分后平稳期阶段的情感时间序列具有平稳序列的特征,能够通过平稳性与纯随机性检验。因此,研究可进一步基于ARIMA模型对该时期的情感序列进行建模,利用ARIMA(1,1,1)模型拟合该平稳期阶段的情感时间序列,模型表达为:

$$x_t = 1.0688x_{t-1} - 0.0688x_{t-2} + \varepsilon_t - 0.9491\varepsilon_{t-1}, \text{var}(\varepsilon_t) = 0.03705 \quad (13)$$

其中, x_t 为序列值, ε_t 为随机干扰。研究对拟合结果进行模型显著性检验,结果如图7所示。

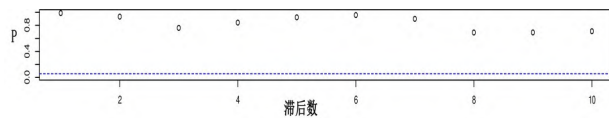


图7 平稳期情感时间序列拟合模型显著性检验

Figure 7 The model significance test at the chronic period

根据图7可以发现,残差序列的LB统计量的P值都高于显著水平参考线(图7中的虚线),这说明残差序列不残留信息,即该拟合模型显著有效。这说明在舆情进入平稳期后,群体的情感表达过程是一个二阶马尔可夫过程,即该时期情感时间序列的序列值主要受到前两天序列值影响。此时,群体情感表达是由一个稳定的非混沌系统所主导。

5 实证结果讨论

通过上述实证分析,可以观察到在此次危机情境下,随着舆情热度的增加,群体情感表达形成三个不同阶段。值得注意的是各阶段表现出迥异的动力学特性,如表5所示。起始期的情感序列是一个平稳随机序列,说明该期间的情感时间序列是由一个随机系统产生的;爆发期的情感序列具有混沌特征,这说明随着舆情热度升高,进入爆发期后主导群体情感表达机制的是一个混沌系统;平稳期的情感序列能够完美拟合ARIMA模型,说明平稳期群体情感表达过程是一个二阶马尔可夫过程,是一个稳定的非混沌系统。对于危机情境下群体情感表达的动力学机制的挖掘,将有助于为管理部门进行舆情引导与决策时提供理论支持。根据本研究实证结果可推断,在爆发期对舆情进行引导将最为有效。群体在爆发期虽会产生大量信息及情感表达等行为,但该阶段群体表达的动力机制主要是由混沌系统主导。混沌对于初始值极为敏感,即通过微小的改变可以改变情感时间序列的演变

模式,因此管理部门应更注重在爆发期阶段对群体情感与思想进行引导。

表5 各阶段情感时间序列的动力学机制

Table 5 The dynamic mechanism of group emotional expression at each period

阶段	动力学机制
起始期	情感时间序列是一个平稳随机序列,由随机系统主导
爆发期	群体情感表达具有混沌性,是一个混沌系统
平稳期	群体情感表达过程是一个二阶马尔可夫过程,是稳定的非混沌系统

同时从实证分析来看,舆情是一个过程系统,群体情感表达是该过程系统的一种连续过程,如图8所示。以往关于舆情演化的研究受信息生态理论影响,普遍认为舆情中的信息、信息环境与人在一定的模式下会形成某种信息生态系统,并且该系统能够在主体、客体、本体等各类要素的协同交互下发挥整体功能。研究人员往往重视舆情信息生态系统的框架体系、自组织机制、主要社群和关键传播节点、路径等问题,认为此类问题的解决是理解舆情演化周期、传播机制以及观点与情感形成过程的关键所在。然而从系统角度看,以往研究本质上将舆情视作一个由各要素组成的对象系统,舆情中信息传播、情感表达等现象是该对象系统的一种功能表达。这与实证结果存在矛盾。如果群体的情感表达是某个对象系统的功能输出,那么该对象系统,也就是舆情信息生态系统在实证数据的三个阶段(528天)内不会经历数次变动,即不会从随机系统转变为混沌系统,再变为非混沌系统。因为在缺乏重大社会变革与信息技术革新的推力下,舆情信息生态系统的组织架构、要素分布以及信息、信息环境与人的相互关系是相对平稳的。因此针对某个舆情问题,舆情信息生态系统应该稳定发挥情感表达功能。某个话题的热度高低只会对情感表达的强烈程度产生影响,并不会改变舆情信息生态系统的内部结构与功能运作。也就是说,舆情热度不会轻易破坏舆情信息生态系统的动力学机制。可见,虽然信息生态理论有助于理解群体信息表达机制的形成机理,但是基于信息生态理论的舆情研究在解释群体情感表达过程方面存在局限。

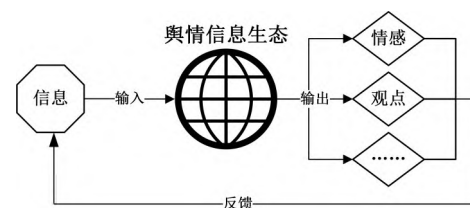


图8 舆情过程系统

Figure 8 The public opinion as a process system

本研究并不否定舆情信息生态系统的研究假设。从实证分析来看,一个能够影响舆情演化的信息生态系统是可以客观存在的,但是群体情感表达并不是该系统的内在功能。同时,具体实践工作中,热度、流量等过程控制量往往更受舆情管理部门的关注。种种迹象表明,舆情应该是一种过程系统,是以舆情信息作为输入量,并在固有的信息生态作用下,

将信息转换为观点与情感等输出量。在舆情过程系统架构下,舆情热度、传播程度、情感强度等观测量可以被视为系统的各类参量,这些观测量之间有着复杂的关联关系。正是这些观测量的复杂关系,才使得舆情过程系统在舆情信息生态保持相对稳定的前提下,表现出丰富多变的动力学特性,即在不同的机制作用下输出各个阶段的情感序列值。换句话说,舆情信息生态系统是舆情过程系统关键组成部分,在舆情演化中发挥重要作用,但其并不是产生舆情动力学现象的唯一机理。

6 结 语

本研究以情感为切入口,通过对情感时间序列进行构建与分析,探究危机情景下群体表达蕴含的动力学信息,对于舆论工作具有重要的理论意义与现实意义。研究表明,在危机情境的不同阶段下群体情感时间序列具有不同特性,其背后所蕴含的动力学机制迥异。换句话说,在危机情境下群体情感表达在起始期由随机系统主导,在爆发期的舆情系统本质上是混沌系统,而平稳期群体情感表达过程是由一个稳定的非混沌系统主导。同时,本文为舆情研究提供了一个新的研究思路,即可视舆情是一种过程系统,群体情感表达是该过程系统的一种连续过程。这一研究发现破解了舆情研究内容中关于对象与过程的概念混淆,指出关于舆情的研究应该分别从对象系统和过程系统两方面展开。本文仅探讨了在危机情境下群体情感表达的动力学机制,在未来将进一步探究其他情境下群体情感表达的动力学机制。

参考文献

- 1 刘冰,肖高飞,晁世育.重大突发公共卫生事件风险研判与决策模型构建研究[EB/OL].[2021-11-09].信息资源管理学报:1-11. <http://kns.cnki.net/kcms/detail/42.1812.G2.20210723.1440.002.html>.
- 2 张海涛,李佳玮,刘伟利,等.重大突发事件事理图谱构建研究[EB/OL].[2021-11-17].图书情报工作:1-8. <https://doi.org/10.13266/j.issn.0252-3116.2021.18.014>.
- 3 文宏.危机情境中的人群“圈层阻隔”现象及形成逻辑——基于重大传染病事件的考察[J].政治学研究,2021(4):134-142,159.
- 4 Mitroff I. Managing Crises before Happen[M]. New York:American Management Association,2001:1-9.
- 5 Kim J, Bae J, Hastak M. Emergency information diffusion on online social media during storm Cindy in US [J]. International Journal of Information Management, 2018,40:153-165.
- 6 Adekoya C O, Fasae J K. Social media and the spread of COVID-19 infodemic[J]. Global Knowledge, Memory and Communication, 2021,10.1108/GKMC-11-2020-0165.
- 7 中华人民共和国国家互联网信息办公室.网络强国建设的思想武器和行动指南——学习《习近平关于网络强国论述摘编》[EB/OL].[2021-09-19]. http://www.cac.gov.cn/2021-02/01/c_1613753152020388.htm.
- 8 阳长征.网络空间中情感扩散,信息级联与舆论偏差的内生影响效应研究——基于2015-2020年突发事件面板数据动态分析[J].情报学报,2021,40(5):448-461.
- 9 Fu P, Jing B, Chen T, et al. Modeling network public opinion propagation with the consideration of individual emotions[J]. International Journal of Environmental Research and Public Health, 2020,17(18):6681.
- 10 吕金虎,陆君安,陈士华.混沌时间序列分析及其应用[M].武汉:武汉大学出版社,2005:1-20,40-60.
- 11 陈思菁,李纲,毛进,等.突发事件信息传播网络中的关键节点动态识别研究[J].情报学报,2019,38(2):178-190.
- 12 成全,刘国威,李艺全.考虑用户主题兴趣度的网络舆情演化建模与仿真研究[J].管理评论,2020,32(11):128.
- 13 Mark B. 信息论:本质·多样性·统一[M].王恒君,嵇立安,等,译.北京:知识产权出版社,2015:127-146.
- 14 Picard R W. Affective computing[M]. Cambridge:MIT Press, 1997.
- 15 Kermer J E, Nijmeijer R A. Identity and European Public Spheres in the Context of Social Media and Information Disorder[J]. Media and Communication, 2020,8(4):28-39.
- 16 安璐,欧孟花.突发公共卫生事件利益相关者的社会网络情感图谱研究[J].图书情报工作,2017,61(20):120-130.
- 17 曾子明,万品玉.融合演化特征的公共安全事件微博情感分析[J].情报科学,2018,36(12):3-8,51.
- 18 Harb J G D, Ebeling R, Becker K. A framework to analyze the emotional reactions to mass violent events on Twitter and influential factors[J]. Information Processing & Management, 2020,57(6):102372.
- 19 任中杰,张鹏,兰月新,等.面向突发事件的网络用户画像情感分析——以天津“8·12”事故为例[J].情报杂志,2019,38(11):126-133.
- 20 安璐,胡俊阳,李纲.基于主题一致性和情感支持的评论意见领袖识别方法研究[J].管理科学,2019,32(1):3-13.
- 21 Thow-Yick L. General information theory: Some macroscopic dynamics of the human thinking systems[J]. Information processing & management, 1998,34(2-3):275-290.
- 22 Mryglod O, Holovatch Y, Mryglod I. Editorial process in scientific journals: analysis and modeling[J]. Scientometrics, 2012,91(1):101-112.
- 23 Vermeer W, Head B, Wilensky U. The effects of local network structure on disease spread in coupled networks[C]// International Workshop on Complex Networks and their Applications. Springer, Cham, 2016:487-498.
- 24 Shih-Lung S, Tsou M H, Xinyue Y. Editorial: Human dy-

- namics in the mobile and big data era[J].International Journal of Geographical Information Science,2016,30(9):1.
- 25 Barabasi A L.The origin of bursts and heavy tails in human dynamics[J].Nature,2005,435(7039):207-211.
- 26 Yi M,Lu Y,Deng W,et al.Understanding verified users' posting behavior from the perspective of human dynamics:a case study of Sina micro-blog[J/OL].Aslib Journal of Information Management,2020,10.1108/AJIM-03-2020-0069.
- 27 Amelkin V,Bullo F,Singh A K.Polar opinion dynamics in social networks[J].IEEE Transactions on Automatic Control, 2017,62(11):5650-5665.
- 28 徐涵,张庆.复杂网络上传播动力学模型研究综述[J].情报科学,2020,38(10):159-167.
- 29 王家坤,王新华.双层社交网络上的企业舆情传播模型及控制策略研究[J].管理科学,2019,32(1):28-41.
- 30 刘海鸥,刘旭,姚苏梅,等.基于舆情画像的在线社交用户信息传播特征统计分析[J].现代情报,2019,39(9):64-73.
- 31 Zhang L,Fung A Y H.Opinion Dynamics Research on Social Media:Breakthroughs and Challenges[J].Telematics and Informatics,2020,46:101314.
- 32 Zhu L,He Y,Zhou D.Neural opinion dynamics model for the prediction of user-level stance dynamics[J].Information Processing & Management,2020,57(2):102031.
- 33 Jiang L L,Hua D Y,Zhu J F,et al.Opinion dynamics on directed small-world networks[J].European Physical Journal B,2008,65(2):251-255.
- 34 Sznajdweron K.Sznajd model and its applications [J].Acta Physica Polonica,2005,36(8):2537-2547.
- 35 Deffuant G,Neau D,Amblard F,et al.Mixing beliefs among interacting agents[J].Advances in Complex Systems,2001(3): 87-98.
- 36 Hegselmann R,Krause U.Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation[J].Journal of Artificial Societies & Social Simulation,2002,5(3):2.
- 37 蔺楠,覃正,汪应洛.基于Agent的知识生态系统动力学机制研究[J].科学学研究,2005(3):406-409.
- 38 张天荣,周文扬,张颢君.我国信息异化的研究热点与发展趋势解析——基于2008—2017年CNKI刊载相关文献的知识图谱分析[J].情报理论与实践,2018,41(11):127-132.
- 39 阳长征.危机事件中网络空间场域情境对舆情偏差的影响研究[J].现代情报,2019,39(10):46-57.
- 40 张海涛,李佳玮,周红磊,等.重大突发事件演变机制:认知框架与理论方法[J].情报学报,2021,40(9):914-923.
- 41 张海涛,刘伟利,栾宇,等.重大突发事件的情景图谱构建[J].情报学报,2021,40(9):924-933.
- 42 赵玲,张静,王欢.微博对群体性事件助燃的动力学机制分析——以“昆明PX事件”为例[J].情报杂志,2013,32(8): 50-56.
- 43 虞铭明,朱德米.环境群体性事件的网络舆情扩散动力学机制分析——以“昆明PX事件”为例[J].情报杂志,2015, 34(8):115-121.
- 44 Tian H,Gao C,Xiao X,et al.SKEP:Sentiment knowledge enhanced pre-training for sentiment analysis[C]//In:Proceedings of 58th Annual Meeting of the Association-for-Computational-Linguistics (ACL),2020:4067-4076.
- 45 Fink S.Crisis management:Planning for the Inevitable[M]. New York:American Management Association,1986:20.
- 46 Bartlett M S.On the theoretical specification and sampling properties of autocorrelated time-series[J].Supplement to the Journal of the Royal Statistical Society,1946,8(1):27-41.
- 47 Ljung G M,Box G E P.On a measure of lack of fit in time series models[J].Biometrika,1978,65(2):297-303.
- 48 张海涛,周红磊,李佳玮,等.信息不完全状态下重大突发事件态势感知研究[J].情报学报,2021,40(9):903-913.
- 49 Fraser A M,Swinney H L.Independent coordinates for strange attractors from mutual information[J].Physical review A,1986, 33(2):1134-1140.
- 50 Rhodes C,Morari M.False-nearest-neighbors algorithm and noise-corrupted time series[J].Physical Review E,1997,55(5): 6162.
- 51 Rosenstein M T,Collins J J,De Luca C J.A practical method for calculating largest Lyapunov exponents from small data sets[J]. Physica D:Nonlinear Phenomena,1993,65(1-2):117-134.

(责任编辑:孙晓明)

The Dynamic Mechanism of Group Emotional Expression in the Crisis

AN Ning¹, AN Lu^{1,2}

(1. School of Information Management, Wuhan University, Wuhan 430072, China;

2. Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China)

Abstract: [Purpose/significance] It is of great significance to explore the dynamic mechanism of group emotion expression in the crisis, which can provide theoretical support for the guidance and decision-making of public opinion. [Method/process] This study obtains microblog posts and comments related to "herd immunity", and calculates daily emotional inclination and values based on the SKEP model. We construct sentiment series, and analyze the stationarity, randomness, and chaos of emotional series. [Result/conclusion] The research results show that the formation of public opinion is a process system. The group emotion expression is a continuous process of this system. At different periods, group emotion expression presents different dynamic characteristics. In the initial period, the sentiment series is generated by a random system. During the outbreak period, the chaotic system dominates the group emotion expression mechanism. The expression process of group emotion in the chronic period is a second-order Markov process, which is dominated by a stable non-chaotic system. [Innovation/limitation] This study explores the dynamic mechanism of group emotional expression in the crisis, and enriches the theoretical perspective. However, this study only focuses on the crisis. In the future, we will study and compare the differences of group emotional expression dynamic mechanism between crisis situations and other situations.

Keywords: group emotional expression; time-series analysis; dynamic mechanism; process system; crisis

(上接第 147 页)

Combine Deep Learning with the Rules of Requirement to Construct the "Target Data" Extraction Model for the Academic Literature

——Taking the Resources of the South China Sea as an Example

PENG Yu-fang^{1,2}, CHEN Jiang-hao³

(1. School of Economics & Management, Nanjing Institute of technology, Nanjing 211167, China;

2. Department of Information Management, Nanjing University, Nanjing 210046, China;

3. School of mathematical sciences, University of Science and Technology of China, Hefei 230026, China)

Abstract: [Purpose/significance] Extracting the target data needed by researchers from the massive academic literature content is conducive to improve the research efficiency of researchers on the one hand, and improve the retrieval services of current literature databases on the other hand. [Method/process] According to the academic needs of scientific researchers, first extract target data from a large number of academic documents through deep learning methods. Secondly, NER and TF-IDF are used to extract the "5W" rule of the target data, and then the target data is filtered by the second-level requirement rule. Any data that meets the "5W" rule is identified as the target data. Finally, the third layer of manual verification is performed on the target data, and the academic literature "target data" is finally generated. [Result/conclusion] The accuracy rate of the "target data" extraction model for academic literature constructed in this paper can reach 86.6%, and the integration of "5W" rule filtering and final manual verification will not only improve the academic literature search of scientific researchers accuracy rate, and to a certain extent assist the retrieval work of literature database institutions. [Innovation/limitation] Combine the deep learning with rules of requirement to realize the search results of academic literature from the bibliographic information level of academic literature to the data level of the content of academic literature.

Keywords: deep learning; named entity recognition; bag-of-words model; TF-IDF; "5W" rule