

基于系统聚类的网络舆情失焦分类及预测研究

摘要：随着我国网民规模不断扩大，互联网普及率飙升，网络舆情失焦现象日益显著。本研究基于网络舆情基本理论构建指标体系，通过聚类分析、判别分析模型搭建以及案例预测检验，将网络舆情失焦划分为情感宣泄和群体心态异常、价值观念和认知失调、存在潜在立场对立三类。研究揭示了这一现象的多维度特征，为深入了解舆论动态及其社会影响提供了准确视角。本研究创新性地拓展了研究路径，整合多个案例进行综合分析，为网络舆情政策的制定提供参考。通过构建指标体系，也为科学评估舆情传播水平提供了实质性支持。

关键词：舆情失焦 Q型聚类 类平均法 全模型贝叶斯判别分析

一、引言

随着互联网的快速发展，公众可以通过社交媒体表达他们对热点事件的观点和意见。第 52 次《中国互联网络发展状况统计报告》显示，截至 2023 年 6 月，我国网民规模达 10.79 亿人，较 2022 年 12 月增长 1109 万人，互联网普及率达 76.4%。这些数据的变化在一定程度上促使我国舆论场中出现了一种新的舆情现象，即网络舆情失焦现象。网络舆情失焦现象是指在网络平台上，大量的舆情信息涌现，但热点事件或话题的焦点往往无法集中在最重要的问题上。例如雷洋事件、女童被咬事件。此类事件借助互联网快速传播对正常网络舆情的发展具有破坏性，并且容易引发衍生舆情等一系列问题，所以关注网络舆情失焦现象，并对此做好分类预警工作十分必要。

为此，本研究尝试构建基于网络舆情基本理论的指标体系，以对网络舆情失焦现象进行聚类分析，同时利用 2016 年至 2023 年的案例对该理论进行实证检验。研究发现，网络舆情失焦现象可以大致分为三类：情感宣泄和群体心态异常类；价值观念和认知失调类；存在潜在立场对立类。通过对这些类别的分析，我们可以更准确地理解网络舆情失焦现象，从而有助于深入研究舆论动态及其对社会和个体的影响。

本研究的贡献主要表现在三个方面：**一是视角新颖。**已有文献往往从舆论反转（凌双和刘颖,2023）、舆情失控（余璐,2020）等角度开展传播研究，鲜有

文献从舆情失焦这一视角开展研究，本文创新性地研究舆情失焦现象。为后续网络舆论政策的制定提供更好的参考依据。**二是路径创新。**现有研究网络舆情现象的文献，往往只选择单一案例展开分析，本文通过整合多个案例，将网络舆情失焦问题进行整体分析和解决，为后续全面科学的探讨网络舆情失焦现象提供了一定的参考。**三是指标创新。**本文借助网络舆情基本理论构造了网络舆情事件的指标体系，并基于 28 个案例进行了统计测度，丰富了已有舆情失焦现象指标的相关制定方法，也为后续科学评估舆情传播水平提供了重要参考。

二、网络舆情失焦的风险因素选取

网络舆情失焦指网络上出现的舆情事件，情绪和信息传播逐渐失控、难以管理和引导。其表现为信息迅速传播、舆情情感极端化、虚假信息扩散等，对个人、组织或社会可能造成负面影响。研究关注网络舆情失焦的爆发原因、信息传播机制和用户行为等，这些研究通过分析大量的社交媒体数据，揭示了网络舆情失焦的一些共性和规律。

本文从网络舆情基本理论出发，识别网络舆情失焦影响因素。识别网络舆情失焦影响因素是辨识其风险，本质是分析评价其发展过程。网络舆情的基本组成要素包括主体、客体和本体（凌双和刘颖,2023）。主体是信息传播发起者或参与者，起到推动或引导舆情发展的作用。客体是信息传播的对象，即被讨论、被评论、被关注的事物。本体指网络舆情中信息的客观真实性，准确性和客观性对公

众认知重要。虚假信息、谣言或夸大事实可能导致舆情失焦。主体、客体、本体构成舆情基本框架，理解关系有助于分析舆情起因、发展和影响。

本文聚焦网络舆情爆发初始阶段，通过案例分析失焦特点，借助网络舆情基本理论，识别网络舆情失焦风险要素如图1所示，并作为聚类分析指标体系。

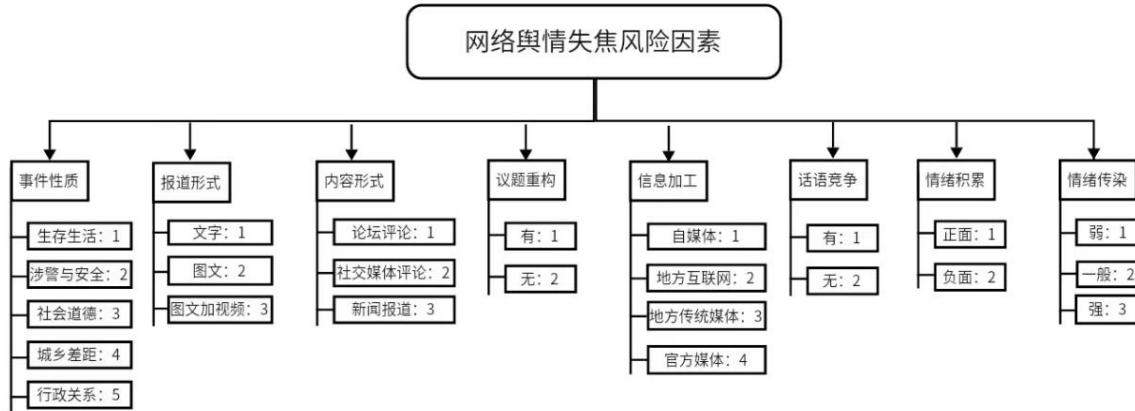


图1 网络舆情失焦风险因素识别的框架

1. 事件性质。

事件性质涉及事件是否与公民的日常生活直接相关。例如，教育问题、房产问题等与生存生活密切相关，而警民关系、生命财产安全、暴力执法等则涉及公共安全。行政纠纷涉及政府官员和政府部门，而社会道德类舆情可能涉及性别歧视、不公正等，城乡差距主题可能触及经济和文化差距。事件性质的差异影响事件的复杂性，是影响事件发展和焦点偏移的潜在因素。

2. 报道形式。

报道形式描述首发网络舆情信息的传播方式，包括纯文字、图文并茂和图文加视频。随着互联网和移动终端的发展，图文并茂和图文加视频的报道形式更受公众信任，但有些事件可能利用这一趋势通过断章取义、制造虚假视图迷惑公众，导致网络舆情失焦。

3. 内容形式

对网络舆论焦点的影响在于它们各自的特点。新闻报道可能因为选择性报道和标题党而使信息失焦，社交媒体评论则因碎片化和情感化言论而导致舆论过度关注情感而非事实，而论坛评论可能因为意见集中和匿名性而使得某些观点难以被充分讨论。这些特点共同影响着信息的传播和解读，使得网络舆论容易受到偏颇、情感化以及信息片面性的影响，从而失焦于核心议题。

4. 议题重构

议题重构指新闻发布或自媒体讲述事件时，是否设置了标签。如#受害者有罪#等，标签化的题目容易吸引人们的注意力，并且快速的形成情感认同，如果一开始就以标签化的题目进行舆论报道，容易造成后续舆论失焦的现象，转变为对标签的讨论。

5. 信息加工

信息加工指是舆情信息首次出现的平台，包括自媒体、盈利性互联网平台、地方传统媒体和央视官方媒体。这些平台的权威性不同，影响网络舆情是否失焦。

6. 话语竞争

在网络舆论中存在话语权竞争，这竞争往往受到性别、阶级、贫富等因素的影响，从而影响舆论的焦点。这种竞争可能使得某些议题受到过度强调，而其他重要议题被忽视，使网络舆论失焦于核心问题，从而影响了网络舆论的焦点和深度讨论内容呈现形式。

7. 情绪积累

情绪积累指媒体对事件的整体偏向。网络舆情失焦事件的媒体报道往往带有明显的情感偏向，影响民众关注点的变化。报道可划分为正向和负面，对网络舆情是否失焦起到关键作用。

8. 情绪传染

情绪传染用网民相关度来评估舆情事件与网民利益的关联程度。事件的关联度越高，越容易导致失焦。普通社会现象属于弱关联，而涉及公共服务、

基础设施建设的事件属于中关联度，涉及生命、财产安全的事件则属于强关联度。一般情况下，与网民关联度越高的事件越容易出现失焦现象。

三、网络舆情失焦的分类及预测模型搭建

网络舆情失焦的分类与预测是紧密相关的两个问题，前者为后者提供分类依据，后者是前者的验证。在识别出网络舆情失焦的风险因素后，采用聚类分析和判别分析法对样本进行分类和预测。

1. 分类模型构建

系统聚类法可将网络舆情失焦根据其不同的性质进行类别划分，其中 Q 型聚类是可以基于网络舆情失焦观测记录进行分类的一种方法，根据末级指标的属性把性质相似的网络舆情失焦案例分在同一个类，性质差异较大的则分在不同的类。由于本文的数据样本均为实际观测案例，因此选用 Q 型聚类分析方法契合本文的研究。

聚类统计量是表示样品之间的相似或相关程度的度量，是对样品进行分类的基础。设有 n 个样品，每个样品测得 p 项指标（或变量），

x_{ij} ($i = 1, \dots, n, j = 1, \dots, p$) 表示第 i 个样品的第 j 个

指标的观测数据。在系统聚类分析的计算过程中，本文拟采用类平均法对网络舆情失焦案例进行聚类。类平均法定义两类之间的距离平方为这两类元素两两之间距离平方的平均。

$$D_{pq}^2 = \frac{1}{n_p n_q} \sum_{X_{(i)} \in G_p} \sum_{X_{(j)} \in G_q} d_{ij}^2$$

其中 d_{ij} 表示样品 $X_{(i)}$ 和 $X_{(j)}$ 之间的距离； D_{pq} 表示类 G_p 与 G_q 之间的距离。设聚类到某一步将与 G_p 与 G_q 合并成 G_r ，则任一类 G_k 与 G_r 之间的距离的平方为 $D_{kr}^2 = \frac{n_p}{n_r} D_{kp}^2 + \frac{n_q}{n_r} D_{kq}^2$ 。

2. 预测模型搭建

判别分析是一种用于分类问题的统计方法，可根据一组已知分类的数据建立分类模型，然后利用该模型对新样本进行分类预测。

全模型贝叶斯判别法是判别分析的一种方法，它基于贝叶斯定理，将条件概率转化为后验概率进行分类。全模型指的是对每个类别建立独立的概率密度函数模型。通过计算样本在每个类别下的后验概率，即给定观测数据时，样本属于每个类别的概率，然后将样本分类为具有最大后验概率的类别。

设有 k 个总体 G_1, G_2, \dots, G_k ，它们的先验概率

分别为 q_1, q_2, \dots, q_k ，各总体的密度函数分别为：

$f_1(x), f_2(x), \dots, f_k(x)$ ，可求得判别函数和判别准则为：

$$y(g/x) = \ln q_g - \frac{1}{2} \mu^{(g)} \sum^{-1} \mu^{(g)} + x' \sum^{-1} \mu^{(g)}$$

$$y(g/x) \xrightarrow{g} \max$$

$$\text{后验概率为: } P(g/x) = \frac{\exp\{y(g/x)\}}{\sum_{i=1}^k \exp\{y(i/x)\}}$$

全模型贝叶斯判别法在新闻案例中能够更好地应对复杂的数据分布、考虑先验概率、处理特征相关性，并具备鲁棒性和可扩展性等优点，有助于提高判别预测的准确性和可靠性。

四、网络舆情失焦案例聚类及预测分析

1. 样本选取及赋值

本文从人民网选取 2016 年至 2023 年间 24 件具有代表性的网络舆情失焦案例，利用 Q 型聚类方法进行聚类分析，划分网络舆情失焦案例的类别，再根据分类结果，对 24 个案例进行判别分析，得出判别函数，最后选取另外 4 件案例进行实证分析。聚类案例选取如表 1 所示。

表 1 网络舆情失焦案例表

序号	名称	序号	名称
G1	刘鑫江歌案	G13	德阳女医生自杀事件
G2	雷洋事件	G14	学术姐已事件
G3	重庆万州公交坠江事件	G15	汤兰兰案

G4	王凤雅事件	G16	重庆东方之星轮船沉没事件
G5	林生斌事件	G17	张环宇故意杀人案
G6	埃塞俄比亚航班坠毁女大学生事件	G18	导演王全安嫖娼事件
G7	余秀华被家暴事件	G19	国企高管牵手异性逛街
G8	河南女童遇害	G20	上海杀妻藏尸案
G9	5·6郑州空姐打车遇害案	G21	漫展中女子拍照姿势事件
G10	女子在公厕被人殴打	G22	张桂梅反对当全职太太
G11	长春生物公司疫苗舆情	G23	“货拉拉”女孩坠亡事件
G12	教育部开学第一课	G24	武汉妈妈坠楼

本文使用个案聚类分析，列出的指标均为名义标度变量，其观测值仅代表不同状况、类别，无大小次序关系，采用编制序号的方式对每个案例的相关要素进行赋值，赋值依据如表 3 所示。

根据以上依据，对 24 个案例（G1-G24）进行赋值。如表 2 所示：

表 2 网络舆情失焦案例赋值表

序号	X1	X2	X3	X4	X5	X6	X7	X8	序号	X1	X2	X3	X4	X5	X6	X7	X8
G1	2	3	2	2	1	1	2	1	G13	3	3	2	2	1	2	2	1
G2	2	3	2	2	1	1	2	1	G14	3	3	2	2	1	1	2	1
G3	2	3	3	1	3	1	2	1	G15	2	1	3	1	2	1	2	1
G4	3	2	2	2	1	1	2	1	G16	2	1	3	2	2	1	1	3
G5	3	3	2	2	1	2	2	1	G17	2	2	3	2	2	1	1	1
G6	1	2	3	2	4	2	2	1	G18	3	1	3	1	3	1	1	3
G7	3	3	2	2	1	2	2	1	G19	3	2	2	2	2	1	2	1
G8	2	3	3	2	4	2	2	3	G20	2	3	3	1	3	1	1	3
G9	2	2	3	1	3	2	2	1	G21	3	3	2	2	1	2	2	1
G10	2	2	2	1	1	1	2	1	G22	3	3	3	1	3	1	1	1
G11	2	1	3	1	3	2	2	3	G23	2	2	2	2	1	2	2	1
G12	2	2	3	1	4	2	2	1	G24	2	3	3	1	1	1	1	1

2. 聚类分析过程及结果

使用 SPSS26 软件对以上数据进行聚类分析。

根据表 5 的处理结果得出，24 个案例的赋值均有效。

表 3 分类情况表

类别	序号
情感宣泄和群体心态异常类	1、2、4、5、7、9、10、12、13、14、15、19、21、23
价值观念和认知失调类	3、16、17、18、20、22、24
存在潜在立场对立类	6、8、11

3. 命名规则及应对策略

依据本文的聚类分析结果，将网络舆情失焦分为三类，通过分析这三类舆情事件，给出各类网络舆情失焦应对策略，为政府部门快速高效采取应对措施提供支持。

1) 情感宣泄和群体心态异常类

这类网络舆情失焦主要是因为人们对于奇闻轶事或悲惨事故表现出猎奇心理，甚至产生幸灾乐祸和隔岸观火的态度。比如余秀华被家暴事件中，起初人们鼓励她拿起法律手段维护自身合法权益。后

通过 Q 型聚类，采用类平均距离计算，得出树状图如图 2 所示。

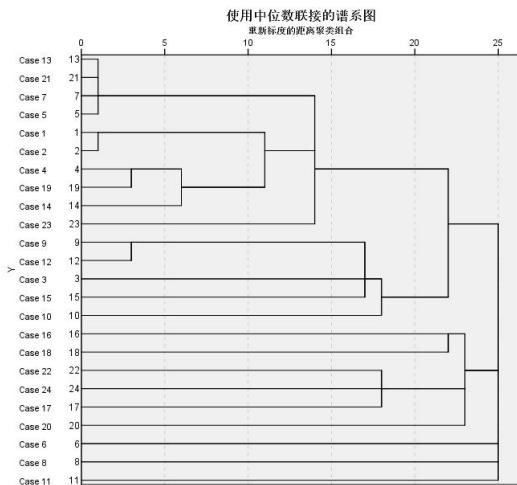


图 2 聚类分析树状图

根据树状图，将 24 个聚类案例分为 3 类，分类情况如表 3 所示。

来，人们开始热议 46 岁的余秀华，一个行动不便的脑瘫患者还想义务反顾的追逐爱情。由本身的家暴慢慢舆论失焦成为了爱情观。余秀华遭家暴最主要的舆论发酵平台就是微信和微博，在微信成熟的闭环交流通路和微博的“用脚投票”等机制下，人们很容易会基于共同的价值观念抱团取暖，多个公众号热门文章或微博高赞评论下，都是同质化的声音，人们在回音壁效应的作用下进一步导致舆论彼此的激烈冲突。

在情感宣泄和群体心态偏差方面，媒体应承担社会责任，不仅报道事实真相，也传递积极向上、

正能量的信息，引导公众参与社会改善和帮助。政府需要培养意见领袖（谢耘耕和荣婷,2011），包括官员、专家学者和媒体人积极参与社交媒体，发布积极正面、权威的言论，引导健康发展的网络舆论环境。

2) 价值观念和认知失调

这类网络舆情失焦主要涉及到价值观念和认知方面的失衡，例如过于追求外貌和颜值，对受害者有罪的偏见观念等。比如上海杀妻藏尸案中犯罪嫌疑人朱晓东婚后出轨，与多名女性发生关系，在一次与妻子的争吵中，残忍杀害妻子并藏之冰柜，然而在庭审视频曝光时，许多网友在下面称赞杀人犯的帅气，替他惋惜，完全忽略了残忍的案件本身，这样的案例数不胜数如北大高材生吴谢宇弑母案。

在价值观念和认知失调方面，政府应加强公民教育，从小培养正确的价值观念，包括多元化的美丽观念和性别平等观念。同时，规范自媒体平台的使用，健全法律法规。媒体应秉持正面的价值导向，避免渲染片面的颜值观念，并通过全面真实的报道消除对受害者的偏见。

3) 存在潜在立场对立类

这类网络舆情失焦主要体现在立场对立方面，例如性别之间的冲突和仇视富人的态度以及政府和普通群众之间的冲突。例如在埃塞俄比亚航班坠毁女大学生事件，人们不再关注逝世的女生，而是对其穿着、长相、家境评头论足，以及在河南女童遇害案中，人们热议为什么不公开嫌疑人的照片反而公开受害者照片，政府公信力受到质疑；同时近年来在很多热点事件的评论下，人们的关注焦点逐渐偏离事件本身，不去客观、理性地分析事件本身，而是打着“女权”“男权”的旗号将责任归单纯归咎于性别差异；性别对立逐渐发展成一个“热梗”，向着娱乐化的方向发展。

在立场对立方面，媒体不能只顾自己的发展，要肩负起媒体的社会责任，起到积极、正向的引导作用。媒体应避免使用容易引起争论的表达方式，

不强调性别、财富差异等对立性的字眼。网友需理性对待他人评论，学会甄别真伪，尤其对于信源模糊的评论要谨慎对待。

4. 网络舆情失焦的预测

1) 样本判别

根据上文聚类分析的结果，本文使用全模型贝叶斯判别分析法，通过 SPSS26 对 24 个案例进行判别分析。所选 24 个案例均通过验证，属于有效数据，如表 4 所示。

表 4 分析个案处理摘要表

分析个案处理摘要		
	数 量	百分 比
排 除	未加权个案数	
	有效	24 85.7
	缺失或超出范围组代码	4 14.3
	至少一个缺失判别变量	0 0
	总计	4 14.3
		总计 28 100

通过 SPSS26 得出 Fisher 的线性判别式函数如表 5 所示，本文主要通过 Fisher 线性判别函数来构建判别方程，理论上说：如果知道某个案例在 8 个指标上的得分，就可以估计出所属网络舆情失焦类型。

表 5 分类函数系数

聚类 指标	cluster		
事件类型	27.656	28.385	24.503
报道形式	11.642	15.799	13.26
内容形式	127.567	142.363	135.804
议题重构	38.83	39.81	43.866
信息加工	-22.083	-22.662	-20.697
话语竞争	-7.036	-13.194	-5.819
情绪积累	89.611	67.594	92.417
情绪传染	13.429	14.92	17.879
(常量)	-298.808	-309.095	-344.549

列出判别系数如下：

$$F_1 = 27.656X_1 + 11.642X_2 + 127.567X_3 + 38.830X_4 - 22.083X_5 - 7.036X_6 + 89.611X_7 + 13.429X_8 - 298.08$$

$$F_2 = 28.385X_1 + 15.799X_2 + 142.363X_3 + 39.810X_4 - 22.662X_5 - 13.194X_6 + 67.594X_7 + 14.290X_8 - 309.095$$

$$F_3 = 24.503X_1 + 13.260X_2 + 135.804X_3 + 43.866X_4 - 20.697X_5 - 5.819X_6 + 92.417X_7 + 17.879X_8 - 344.549$$

2) 合理性验证

据表 6 可知，该判别函数正确地对 95.8% 个原始已分组个案进行了分类。同时，根据组质心处函数得出分类图如图 2 所示。

表 6 预测分类结果

		分类结果				
		预测组成员信息			总计	
		cluster	1	2	3	
计数	1		14	0	0	14
	2		1	6	0	7
	3		0	0	3	3
未分组个案			2	1	1	4
预测成功率	1		100	0	0	100
	2		14.3	85.7	0	100
	3		0	0	100	100
未分组个案			50	25	25	100

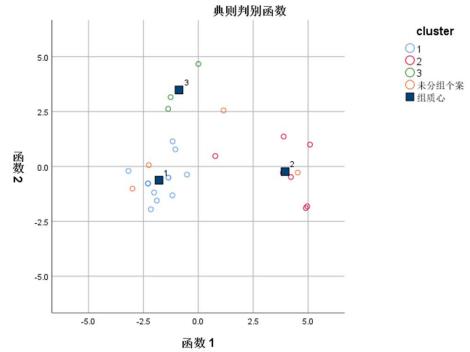


图 2 组质心处函数散点图

由图 2 可知，四类样本案例均围绕四个组质心分布，说明该分类方案较为合理，可以依据此以上判别式对新发生的舆情失焦事件进行分类。

3) 实例预测

通过 4 组其他网络舆情失焦案例对判别结果进行验证。对 4 组验证案例 (G25-G28) 的打分赋值如表 7 所示。

表 7 验证案例打分表

序号	事件名称	X1	X2	X3	X4	X5	X6	X7	X8
G25	《1818 黄金眼》小张维权	1	3	3	1	3	1	1	1
G26	两岁女童遭恶犬咬伤	2	3	3	1	3	1	2	3
G27	因染粉色头发被网暴的女生去世	1	2	2	2	1	1	2	1
G28	女研究生回家休学待产	3	2	2	2	1	2	2	1

同时判别函数计算如表 8 所示。

表 8 验证结果

案例	F1	F2	F3	聚类
25	215.0565	234.9196	213.395	第二类
26	359.1809	360.7376	366.072	第三类
27	248.4564	229.4847	242.0087	第一类
28	296.7311	273.0594	285.1948	第一类

根据判别函数我们得知 G25 中《1818 黄金眼》小张维权事件为价值观念和认知失调类，该事件由于小张帅气的外貌引起了网友的关注。人们纷纷要求物业赔偿其损失，显然这符合第二类中的颜值至上的类型。

在 G26 中一位母亲和 2 岁女儿遭到巨大黑狗袭击，导致女童重伤。母亲发起了 200 万众筹，引发争议。不少网友的言论，是对弱势群体遇难的姿态要求。网友们既要求女孩母亲承受女孩遇难的痛苦，又对“200w”眼红。很快又迎来第二次舆论失焦：是否抓捕流浪狗？舆论的关注点，应该是除了对流浪猫狗展开行动外，我们还必须对弃养动物的主人

进行惩治。而不是浅层的“捕猎流浪狗”，更不是单方面“替流浪狗声援”。通过分析可以发现在两次失焦中，第一次是人们潜在的仇富心理在作怪，第二次变成了执法者和普通人之间的话语权竞争，都属于第三类，与判别分析得出的结果一致。

G27 事件中，郑灵华因染粉色头发遭受网络暴力，一些网友以她照片中的粉色头发为由，展开攻击和谩骂。在 G28 事件中，南依考入社科院后选择和家乡一位理发师结婚，休学生子。这条本来普通的微博最终引起了更多人的关注，甚至上升到职场和学界歧视的议题，相关话题登上热搜。G27 和 G28 两个案件都表现出舆论失焦的主要原因在于群体心态失常，随意造谣的心理。与判别函数的结果较为一致。综上所述，判别函数的结果较为准确。

五、结语

综上所述，本研究在现有的网络舆情失焦案例基础上进行了初步探索，但对未来可能发生的特殊案例未进行深入考虑，因此还需要进一步跟踪和研

究网络舆情失焦的发展情况及规律。在方法使用上，由于可查证的案例有限，聚类分析和判别分析的准确性仍有待提高。因此，未来我们需要持续追踪和分析最新的舆情失焦事件，改善模型的合理性，提高预测的精确度。面对舆情失焦现象日益严重的情况，本研究将聚类分析和判别分析方法相结合，在SPSS26软件的支持下，对网络舆情失焦进行了多维度分析，识别了与失焦相关的风险因素，并提出了分类方案和相应的应对策略，这为应对网络舆情失焦现象提供了有效的依据，帮助相关主体能够高效、准确地应对网络舆情失焦问题。未来的研究应继续深入挖掘和探索，扩大样本量，以更好地应对不断变化的网络舆情失焦挑战。

参考文献

- [1] 凌双,刘颖.新媒体情境下重特大安全事故舆情风险何以被放大[J].中国应急管理科学,2023,(08):61-72.
- [2] 余璐.5G背景下突发事件网络舆情失控现象及其规避策略[J].新媒体研究,2020,6(08):58-60.DOI:10.16604/j.cnki.issn2096-0360.2020.08.019
- [3] 谢耘耕,荣婷.微博舆论生成演变机制和舆论引导策略[J].现代传播(中国传媒大学学报),2011(05):70-74.DOI:10.19997/j.cnki.xdcb.2011.05.013.