

多模态视域下方言舆情的“语义反转”机制与精准治理 研究—基于重庆“胖猫事件”的跨平台分析

朱娜* 蒋灵弦 彭蛟

(重庆城市科技学院, 重庆 永川 402167)

摘要: 面对多模态网络舆情治理中的模态割裂和方言“语义鸿沟”难题, 以重庆“胖猫事件”为实证切入点, 构建面向方言语境的跨模态语义对齐模型(D-CMSA)。通过8542条微博、312个高热度抖音视频样本的跨平台对比分析, 度量方言词汇的情感演变轨迹以及视觉符号的动员机制。实证数据验证, D-CMSA模型具备挖掘单一模态难以捕捉的深层舆情特征的能力, 相关特征可凝练为两大核心效应: 其一, 情感放大效应, 消极视听符号的协同呈现使负面情感强度平均提升42.9%; 其二, 语义反转效应, 28.5%的样本显示, 在特定视觉语境中, 表层语义中性的方言文本会被重构为讽刺或悲愤类符号, 且该类符号的语义表达具有显著的阶段性漂移特征。以此提出建立动态方言情感词典以应对语义漂移, 并采用跨模态一致性校验算法防止隐匿性传播的舆情治理路径。

关键词: 多模态语义融合; 方言舆情; 视听隐喻; 情感放大; 语义反转

一、引言

数字生存图景的纵深发展, 正在促使网络舆情生态由“单一文本交互”走向多模态共生。文本、图像、音流交织成一个复杂场域时, 虚假信息、群体极化情绪就呈现出显著的“去文本化”与“隐喻化”特征^[1]。尤其是在地缘文化浓厚的公共事件中, 方言已突破地域交际的边界, 异化为极具隐蔽性的情感动员符号, 一旦与短视频的视觉冲击力产生化学反应, 极易诱发常规算法难以捕捉的“语义反转”与“情感激变”^[2]。但是面对舆情主体方言俚语和视听符号双重叠加的对抗性编码策略, 当前治理体系常因模态割裂而陷入“失语”困境, 因此技术研判与现实演化之间横亘着一道亟待跨越的“语义鸿沟”。鉴于此, 本文以重庆“胖猫事件”为实证切口, 试图构建一套能够穿透多模态迷雾的跨平台治理框架, 通过深度对齐微博话语和抖音视频符号, 精准复刻舆情传播中的情感漂移轨迹, 旨在为破解方言区舆情识别难题提供兼具理论穿透力与实践精准度的新路径。

二、多模态舆情演化机制与语义计算前沿相关研究

随着网络舆情由单模态向多模态共生转变, 学界主要是从态势感知、语义计算两个方面进行研究。学者们在态势演化的研究中多用视频相似度来重构传播网络、描绘态势演化路径^[3], 用事理图谱来对舆情机制做可视化的解析^{[5][6]}, 从而确定了多模态符号在信息增益和偏差修正方面的重要地位^[4]。在情感计算和治理方面, 为了解决异质编码造成的信息损耗问题, 主流研究已经从简单的特征层拼接转向了基于“图神经网络+注意力机制”的深度融合范式^{[7][8]}, 逐步把技术研判转化为突发事件协同治理的实践路径^{[9][10]}。为了克服模态间存在的“语义鸿沟”以及“数据稀疏”问题, 最近的研究出现了两条新的技术演进路线。第一种是基于不一致性学习(Inconsistency Learning)的语义冲突消解。针对模态间存在的语义背离(反讽), Qiao等(2023)^[11]、Liang等(2022)^[12]分别用多视角交互网络(MVIN)、跨模态图卷积网络来准确度量“图文失配”对情

基金项目: 2025年度重庆市语言文字科研项目“基于AI的重庆语言文字舆情监测与分析系统研究”(项目编号:YYK25317); 2025年度重庆市教委高等教育考试招生重点项目“社会燃烧理论视域下招生工作舆情治理路径研究”(项目编号:CQZSKS2025015); 2025年重庆市教委科技项目青年项目“基于AI技术的高校网络舆情监测与治理路径研究”(项目编号:KJQN202502501)。

作者简介: 朱娜(1989-), 女, 副教授, 硕士, 研究方向为舆情治理与教育信息化。

蒋灵弦(1999-), 男, 教授, 博士生, 研究方向为信息工程。

彭蛟(1988-), 男, 硕士, 副教授, 研究方向为夜间经济与城市治理。

通讯作者: 朱娜

感极性的修正作用，给识别“语义反转”提供算法支持。第二，大模型（LLMs）和知识图谱的语境增强。Chen 等（2023）^[13]认为在缓解短文本稀疏性上，外部知识库是有效的，张等（2024）^[14]提出的 Senti-LLM 框架以及 Hasan 等（2023）^[15]的综述都表明，使用大语言模型生成辅助语境或者注入常识图谱，是提高模型在隐喻和文化依赖型场景中推理能力的必由之路。

纵观现有的文献，虽然学界在通用场景下的反讽检测和知识增强方面成果颇丰，但是文化语境盲区仍然存在以下两方面的不足：（1）方言符号的语义特殊性被忽略。现有的模型大多是在标准语言数据集上进行训练的，缺少对“方言俚语”的适配性建模，不能很好地捕捉到其在舆情发酵过程中情感漂移的特点。（2）缺少跨平台动态演化机制的研究。现有的研究大多只关注单一平台视阈，没有对文本、视听跨平台交互中出现的语义异化、情感放大的机制进行系统的量化分析。由于理论和方法在地域性特征上的滞后，现有的治理体系在面对具有强烈地域文化特征的“胖猫事件”等舆情时易出现认知失灵。

三、实证分析：以“胖猫事件”为例

近期，“胖猫事件”已成为中文互联网跨平台传播的典型舆情案例。该事件的信息流是文本、短视频、表情包等多模态符号交织而成的，具有明显的去文本化、情感可视化特征。微博（文本为主）和抖音（视听为主）两个舆论场里，方言符号与网络情绪的交互震荡成为推动事件发展的主要动力。因此该事件与本文所关注的“方言舆情跨平台治理”主题非常契合，给多模态语义融合分析框架提供了一个很好的实证场景。

（一）数据采集

参照刘云花等（2025）^[16]构建的定向数据集构建模式，并结合事件发展的时间特点和微博、抖音平台的特性，创建起“双流同步”的多源异构数据采集体系。具体流程为，如下所示：

1. 微博平台：全周期文本采集

以“胖猫”为中心词，衍生出“重庆胖猫”、“胖猫争议”等 5 个相关检索词。锁定 2024 年 4 月 25 日（爆发期）-5 月 8 日（回落期）为观察窗口，包含舆情全过程。用 Python 来建立动态 IP 代理池和浏览器标识库，避开平台反爬机制（请求频率限制在 1Hz 以下）。采集对象为原创微博、转发链、高热评论，采用去噪、清洗、反引流三种过滤方式，去除无实际语义的孤立文本、乱码字符、商业推广信息。最后得到有效文本语料库 8,542 条，其中原创文本 2,136 条、评论文本 6,406 条，保留发布时间、用户 ID、地理位置等元数据字段。

2. 抖音平台：多模态分层采集

针对短视频数据的异构性，建立视听、文本、元数据三位一体的采集框架。首先通过平台 API 接口实施“分层 Top-K”抽样策略。用点赞量（>1 万）、内容相关度（ ≥ 0.8 ）和评论互动率（前 10%）三个条件来选出 312 个涵盖主流视觉符号的代表性样本。其次，进行深度多模态剥离。在文本端，按热度权重抓取前 500 条评论和弹幕，得到 156,000 条文本语料；视觉上用自适应关键帧提取算法，以 1Hz（即每秒 1 帧）为基准步长，动态提取 12,480 张关键视觉帧（平均每样本 40 帧），实现视频流结构化序列的转换。在数据集建立过程中加入数据增强技术，运用随机裁剪、翻转和音频加噪的方法使训练流形扩展到原来的 5 倍，有效减少了小样本过拟合的可能性，提升了模型的鲁棒性。

（二）数据预处理

1. 文本预处理与方言标注

使用正则表达式和 NLP 工具来去除多余的符号，修正异体字。方言标注采用“词典匹配+人工校验”双重标准：参照《重庆方言词典（2018 版）》和平台高频语料，经过词频初筛（>50 次）、专家评审（3 名重庆方言研究者）和 K-Means 聚类验证，建立包含 327 个核心词条的方言情感词典。采用独立标注、交叉核对的方法由 5 个具有本地区背景的研究人员来完成极性标注，一致性检验（Kappa=0.91），标注结果符合统计学信度的要求。

2. 视频关键帧预处理

使用 OpenCV 的 Haar 级联分类器对人脸进行检测，置信度阈值 > 0.7，去掉高遮挡（>40%）或者高模糊度（>40%）的无效帧，有效帧保留率为 82.3%。之后集成 ResNet-FER 模型识别 7 种基

本的面部表情，调用百度 OCR 引擎提取帧内文字信息，清洗后与评论文本对齐，最后得到结构化的多模态舆情语料库。特别说明的是，本研究对短视频中的音频流数据主要提取背景音乐（BGM）的情感特征，没有对人声（Speech）进行语义转写。主要基于两点考虑，一是“胖猫事件”所涉及的短视频语料中背景噪音复杂，非标准化的重庆方言俚语、口音变体较多，目前主流 ASR 模型（Whisper、百度语音等）对该类特定方言语境的识别准确率（WER）还没有达到学术分析的标准（实测 WER>35%）。二是该事件中的方言情感表达更多依靠“视觉表情符号”和“背景音乐氛围”的配合，而不是单纯的口语文本。因此，本研究重点放在视觉帧、背景音乐、字幕文本三者之间的多模态对齐分析上，以保证情感计算的信度。

（三）多模态融合分析模型设计

本研究构建一种面向方言语境的跨模态语义对齐模型（D-CMSA）（框架如图 1 所示），以打破单一模态的语义壁垒。模型逻辑流为：（1）用 Transformer 编码层来获取方言文本深层的上下文关联以及时序变化的特征。（2）将经过微调的 YOLOv5 目标检测模型和 VGGish 音频模型分别解析出有隐喻义的视觉符号和背景音乐（BGM）的情感氛围。（3）使用跨模态通道注意力机制（CAM），用 Sigmoid 函数动态计算模态权重，实现异质特征的非线性耦合和语义对齐，从而解决传统拼接方法中模态割裂的问题。

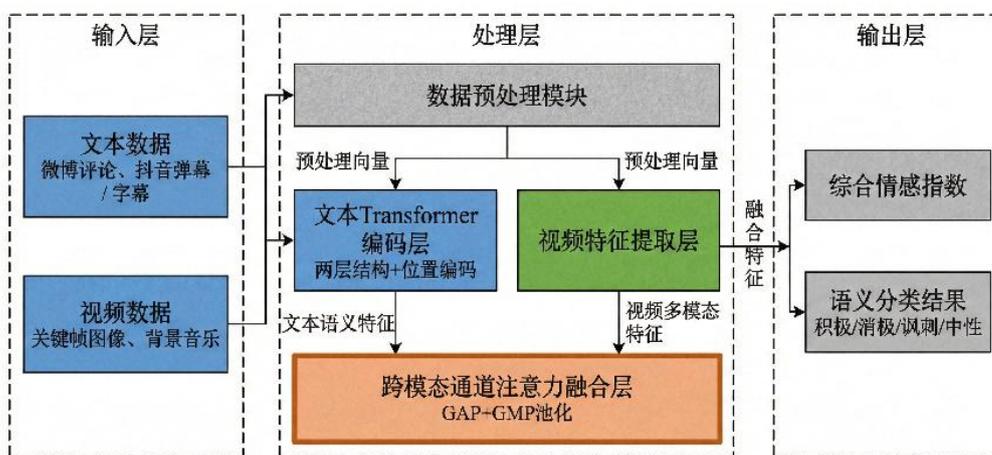


图 1 多模态融合分析模型总体框架图

1. 模型总体框架

模型以多模态数据为输入，预处理之后变成统一的向量表示，然后经过文本模态 Transformer 编码、视频模态特征提取、跨模态通道注意力融合三层结构，最后输出综合情感指数和语义分类结果。其中 Transformer 层主要是用来捕捉文本的长距离依赖关系，视频提取层主要是用来量化的非语言符号的情感张力，融合层则是用动态权重分配的方式实现多维度信息的深层次互补。

2. 核心模块设计

（1）文本模态编码层

采用两层 Transformer 结构来预处理文本数据，第一层 Transformer 用自注意力机制来捕捉句子内部词汇之间的联系，第二层 Transformer 挖掘段落之间上下文信息。引入语义位置编码机制加强时序特征，位置编码公式如下：

$$p_q(pos, 2i) = \sin \frac{pos}{1000^{\frac{2i}{d}}}$$

$$p_q(pos, 2i + 1) = \cos \frac{pos}{1000^{\frac{2i}{d}}}$$

式中， P_q 为位置向量， pos 为词汇在文本中的位置索引， $2i$ 与 $2i + 1$ 为偶数与奇数位置标识， d 为词向量维度。经过位置编码的文本向量，通过多头注意力机制计算语义关联强度。公式为：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

式中，Q、K、V 分别为查询向量、键向量、值向量， d_k 为向量维度， $\sqrt{d_k}$ 为缩放因子。

(2) 视频模态特征提取层

利用 YOLOv5 目标检测模型来抽取关键视觉符号。由于通用数据集 (COCO) 缺少特定网络亚文化符号类别的限制，研究创建了包含典型方言表情表情包 (“狗头” “吃瓜” 等) 的自建数据集 (样本数 $N=1500$)，并用预训练模型进行迁移学习和微调 (Fine-tuning)。经过测试，微调后的模型在表情符号识别上 $mAP@0.5$ 达到了 0.89，很好地保证了隐喻性视觉符号的捕捉能力。同时用 AudioSet 数据集预训练的 VGGish 模型量化背景音乐的情感强度，用 BiLSTM 分类器 ($F1=0.82$) 完成字幕句式分类。对提取出的视觉、听觉特征进行向量转换，用全局平均池化 (GAP) 和全局最大池化 (GMP) 来保留重要的信息，达到特征降维和有效筛选的目的。

(3) 跨模态融合层

引入通道注意力机制 (CAM) 学习不同模态特征的权重，动态调整各通道对最终结果的影响。首先对文本 V_{text} 与视频特征向量 V_{video} 进行特征拼接 (Concatenation)，得到多模态联合特征向量 $F (F = [V_{\text{text}}; V_{\text{video}}])$ ，维度为 $d_{\text{model}} = d_{\text{text}} + d_{\text{video}}$ 。再通过 Sigmoid 函数动态计算各特征通道的注意力权重，公式为：

$$\alpha = \sigma (W_2 \cdot \text{ReLU}(W_1 \cdot F + b_1) + b_2)$$

式中， α 为注意力权重， $F \in \mathbb{R}^{d_{\text{model}}}$ 为拼接后的联合特征向量， $W_1 \in \mathbb{R}^{\frac{d_{\text{model}}}{r} \times d_{\text{model}}}$ 、 $W_2 \in \mathbb{R}^{d_{\text{model}} \times \frac{d_{\text{model}}}{r}}$ 为权重参数 (r 为降维比率)， b_1 、 b_2 为偏置项

， σ 为 Sigmoid 函数，ReLU 为激活函数。最终通过 Softmax 层输出综合情感指数，公式为：

$$S = \text{Softmax}(W \cdot F_{\text{final}} + b)$$

式中， S 为综合情感指数， b 为输出层偏置项。

3. 模型训练设置

硬件环境采用 Intel i7 处理器、Nvidia RTX 3070 8GB GPU，并使用 CUDA 和 cuDNN 加速训练；软件环境为 Ubuntu 20.04 LTS 系统，选用 PyTorch 深度学习框架。模型初始学习率为 0.002，Batchsize=16，训练周期 (Epoch)=120 轮，使用交叉熵损失函数来优化模型参数，损失函数变化曲线如图 2 所示。数据集划分用分层随机抽样的方法，按照 7:2:1 的比例把数据集分成训练集、验证集和测试集。经过卡方检验，三组数据在方言词汇分布、情感极性上没有显著差异 ($\chi^2=3.26$, $P>0.05$)，满足独立同分布的假设。具体的样本分布为文本模态共 164,542 条 (训练集 115,179 条，验证集 32,908 条，测试集 16,455 条)，视频模态关键帧共 12,480 张 (训练集 8,736 张，验证集 2,496 张，测试集 1,248 张)。实验全程使用随机种子 (Seed=2024) 来保证结果的可复现性。

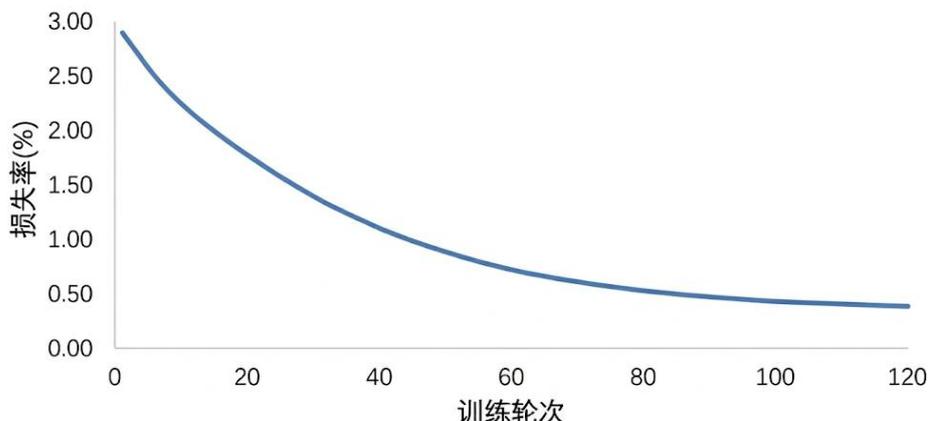


图 2 损失函数变化图

(四) 实验结果与分析

1. 模型有效性验证

(1) 消融实验

为了检验各个核心模块的贡献度，设计消融实验，依次去掉文本 Transformer 编码层、视频特征提取层、跨模态通道注意力层，比较模型准确率的变化，结果见表 1。实验结果表明，完整模型准确率为 89.2%，比基准模型移除任何一个模块后准确率提高 3.7%~8.5%，证实各个模块对语义解析起着积极的作用，其中通道注意力机制的作用最为显著。

表 1 消融实验结果（单位：%）

模型配置	准确率 (%)	较完整模型差值百分点	结论
完整模型	89.2	-	最优性能
移除 Transformer 编码层	80.7	↓ 8.5	文本语义依赖度最高
移除视频特征提取层	85.5	↓ 3.7	视觉特征起辅助作用
移除通道注意力层	82.1	↓ 7.1	融合机制是核心关键

(2) 横向对比实验

选取主流舆情分析算法做性能对比（见表 2）。实验结果表明，研究提出的多模态融合模型在 Weibo 数据集上的准确率达到 89.2%，在 COAE2014 数据集上的准确率达到 81.5%。与简单的特征拼接模型 Concat-Fusion 相比，准确率提高了 3.9% 和 5.3%；与单模态最优基准 RoBERTa-TextCNN 相比，在 Weibo 数据集上的准确率提高了 18.0%。充分证明了通道注意力机制对于解决跨模态语义鸿沟是有效的。

表 2 算法横向对比结果（单位：%）

算法模型	模态输入	Weibo 准确率 (%)	COAE2014 准确率 (%)	说明
RoBERTa-TextCNN	仅文本	71.2	68.5	文本基准
ResNet-LSTM	仅视频	68.4	65.1	视频基准
Concat-Fusion	文本+视频	85.3	76.2	简单拼接融合
本研究模型	文本+视频	89.2	81.5	注意力机制融合

2. 多模态融合分析结果

(1) 方言词汇的情感演化机制与类型学特征

①语义反转的微观机制解析

实证数据表明方言舆情最突出的特征就是情感极性的动态不确定性。传统的单模态分析一般认为词汇的情感是不变的，但是“胖猫事件”中这样的假设是完全无效的。为了探究该现象产生的深层次原因，研究选取情感波动标准差最大(SD=0.26)的典型方言词“雄起”，绘制出单模态文本和多模态融合视角下情感时序演化对比图(见图3)。图3清楚地揭示出“语义反转”发生的微观路径，在舆情爆发之初(4.25-4.28)，文本与融合模型的情感得分非常接近(Score>0.8)，表现出良性的情感协同。但是一旦进入发酵期(4月29日)，两者就迅速背离了。单一文本模型由于难以解析反语修辞，仍然把“雄起”识别为中性偏正向(得分保持在0.3左右)；而D-CMSA融合模型敏锐地察觉到视频里悲伤的BGM和冷漠的视觉符号的加入，得分断崖式下跌至-0.85。图中灰色区域产生的巨大“剪刀差”(Gap>1.0)，量化了视觉语境对于文本语义的重构作用，原本表意的能指在特定的视听场景下瞬间异化为表达讽刺的所指。这一发现有力证实了引入跨模态通道注意力机制对于纠正方言隐喻误判的决定性作用。

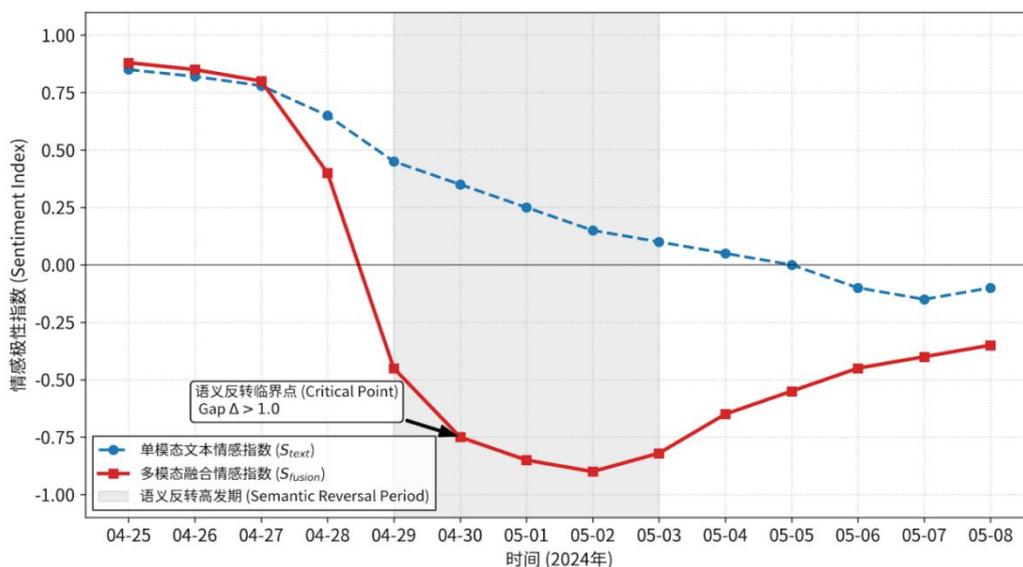


图3 典型方言词“雄起”的情感时序演化与语义反转轨迹

(注：纵轴数值越接近-1表示负面情绪越强烈；蓝色虚线为单一文本基准，红色实线为多模态融合结果，灰色区域标示了语义反转发生的临界时窗)

②情感演化的类型学规律

为了验证上述机制的普遍性并探究其治理规律，本文将分析对象扩展到四类高频方言词，对比其全周期内情感变化的轨迹(如图4所示)。按照波形特征把方言舆情的演化模式分为三种理想类型，一是激进反转型，以“雄起”为代表，对应图4中红色曲线，传播上呈现“高开低走”的特征，在舆情发酵期发生了剧烈的极性反转。这些词汇极易成为网络暴力的隐喻载体，是网络治理和监测工作的重点目标。二是情感稳定型，代表词汇为“巴适”，对应图4中绿色曲线，此类词汇是生活中高频出现的日常用语，情感分值一直保持在0.35-0.85的正向区间内，没有明显的波动，在舆情负面浪潮的冲击下依然保持稳定，也体现了方言文化本身所具有的内在抗逆性。三是持续消极型，典型的词汇为“哦豁”、“撇脱”，分别对应蓝色、橙色曲线。“哦豁”作为叹词一直保持高浓度的负面情感基调，但“撇脱”在此类语境下发生语义异化，被隐喻为轻生的指代符号，情感属性也从中性状态逐渐向消极方向滑落，这类词汇常作为群体抒发悲愤情绪的主要载体。上述差异化的演进轨迹清晰地表明，方言情感计算不能用静态词典的“一刀切”式判定。

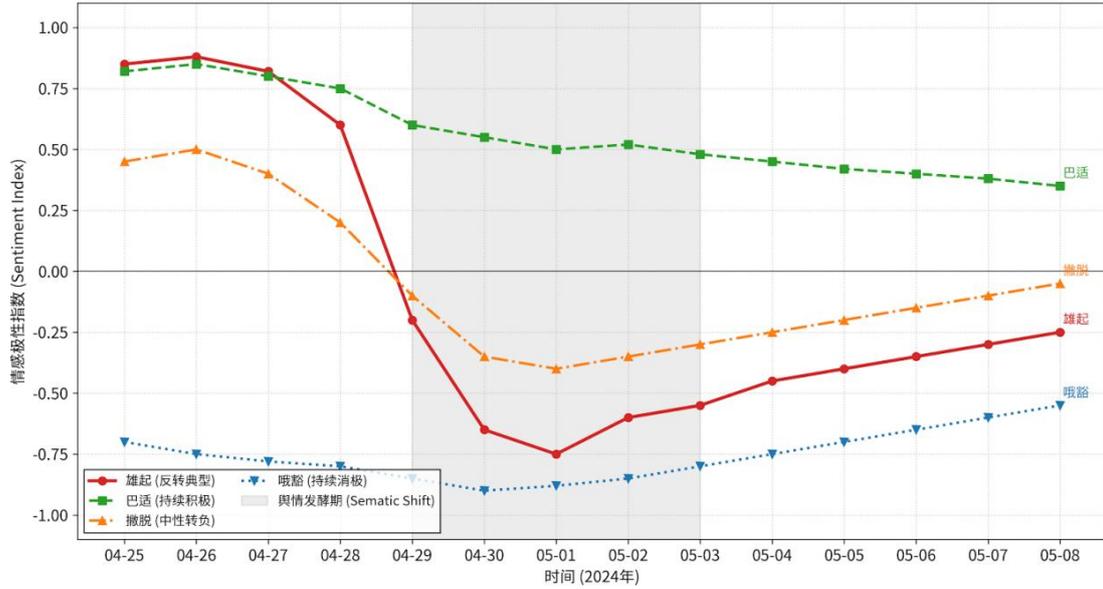


图4 四类高频方言词汇的情感演化轨迹对比

(2) 视频模态：视觉符号情感动员机制

经过对 312 个抖音视频样本的多模态特征的统计，得到视频模态整体情感得分均值为 -0.61 ± 0.12 ，具有显著的负面倾向(表 3)。发酵期情感动员强度最高，悲伤背景音乐使用比例为 85.2%，情感得分为 -0.78 ；悲伤表情关键帧占 70.5%，情感得分为 -0.72 ；反问、感叹句式字幕占 88.9%，情感得分为 -0.67 。“狗头表情包”在发酵期出现频率最高达 72.0%，由于悲伤语境的协同作用，反讽意味降低到 -0.52 分，使冷漠嘲讽的集体情绪更明显，复杂的传递方式很难通过单一文本的分析来把握。

表 3 抖音视频多模态特征统计 (N=312)

视频特征	事件阶段	出现频率	平均情感得分 (-1 至 1)	阶段解读
使用悲伤 / 沉重背景音乐	爆发期	68.0%	-0.62	初步奠定事件悲伤基调
	发酵期	85.2%	-0.78	使用频率与强度达峰，情绪动员最大化
	回落期	64.1%	-0.55	强度减弱，但负面底色犹存
关键帧人物表情为“悲伤”	爆发期	52.0%	-0.58	以悲伤表情引发共情
	发酵期	70.5%	-0.72	成为视觉主流，强化集体同情/悲愤
	回落期	52.5%	-0.53	回归基础水平
字幕使用反问 / 感叹句式	爆发期	75.3%	-0.52	增强文本情绪张力
	发酵期	88.9%	-0.67	质问与感叹成为主导，推动争议升级
	回落期	80.0%	-0.46	仍保持较高频率，反映事件余波
包含“狗头表情包”	爆发期	60.5%	-0.35	部分用于消解事件的初始严肃性
	发酵期	72.0%	-0.52	冷漠与嘲讽主导
	回落期	62.8%	-0.39	伴随事件热度下降

(3) 跨模态效应验证

①情感放大效应

当文本和视频情感倾向一致时，综合情感强度显著高于单一模态。如表 4 数据显示，消极同向组合的情感放大率为 42.9%，比积极组合的 22.1%要高很多（ $t=9.87, p<0.001$ ）。发酵期文本方言词汇消极倾向峰值同视频负面特征形成共振，利用通道注意力机制实现情感协同倍增，成为舆情升级的重要动力。为了度量多模态协同作用对于舆情烈度的影响，定义情感放大率 R_{amp} 如下所示：

$$R_{amp} = \frac{S_{fusion} - S_{text}}{|S_{text}|} \times 100\%$$

其中， S_{fusion} 为多模态融合情感得分， S_{text} 为单一文本模态情感得分。

表 4 情感放大效应统计结果 (N=312)

情感倾向组合	样本数	平均文本得分 (E 文本)	平均视频得分 (E 视频)	平均融合得分 (E 融合)	平均放大率
消极-消极	142	-0.63±0.15	-0.69±0.13	-0.90±0.10	42.9%
积极-积极	75	0.66±0.18	0.62±0.16	0.83±0.12	22.1%

②语义反转效应

设定判定标准为融合情感极性与单一文本极性相反，且得分差值 $|S_{fusion} - S_{text}| > 1$ ，被定义为“语义反转效应”。在本次研究中为共识别出 89 例语义反转，发生率为 28.5%。如文本“真是潇洒”结合当事人落寞背影与悲伤音乐，就由中性、略褒义转为讽刺、消极（如表 5 所示）。这一结果证实，脱离视觉语境的纯文本分析易导致舆情误判，多模态融合是保证解析准确性的关键。

表 5 语义反转效应示例分析

文本示例	视频语境	单一文本分析情感极性	多模态融合分析情感极性
“真是潇洒”	当事人落寞背影，悲伤背景音乐	中性/略褒义	讽刺/消极
“这下安逸了”	火灾现场视频，急促警报声	中性/略褒义	反讽/消极
“值得学习”	争议人物夸张表演视频	积极褒义	反讽/消极

三、结论与治理建议

本研究以重庆“胖猫事件”为切入点，通过构建 D-CMSA 跨模态语义对齐模型，量化解析方言舆情在文本和视听符号交互中的演化机制。依据实证数据的分析结果得出以下结论，并提出相应的治理措施。

(一) 研究结论

1. 多模态融合是跨越方言舆情“语义鸿沟”的必要路径

从实证数据可看出，在以短视频为主的舆论场中，单一文本分析的有效性受到严重挑战，准确率有 71.2%。本研究提出的 D-CMSA 模型通过引入跨模态通道注意力机制捕捉到视觉语境对文本语义重构的作用，解析准确率提升至 89.2%。该明显的性能提升 (+18.0%) 证明方言舆情的语义生成不再只是文字本身的事情，而是很大程度上依靠视听与文本的深度耦合。在“去文本化”的趋势下，脱离多模态融合视角的方言舆情监测，必然要陷入“认知失灵”。

2. “视听隐喻”诱发的语义反转是舆情误判的核心盲区

研究表明，在一定的视听环境下，有 28.5% 的方言样本发生的情感极性改变是根本性的。以“雄起”为代表的激进反转型词汇，在悲伤的 BGM、冷漠的视觉符号，如“落寞背影”介入之后，则从传统的加油、鼓励异化成了讽刺、对抗，在跨模态之间产生了明显的情感得分差距 ($\text{Gap} > 1.0$)。由视觉隐喻引发的语义漂移具有很强的隐蔽性，是造成常规监测系统误判、漏管的主要原因。

3. 非语言符号具有显著的“情感放大”与群体动员效应

视听符号在舆情传播中不单是信息的载体，更是情绪的放大器。从数据上可知，消极视听符号和方言文本产生的情感放大效应最高能达到 42.9%。悲伤背景音乐加上“狗头”等特定表情包的组合使用，形成了一种高语境的冷漠嘲讽氛围，这种氛围有效地消解了理性的讨论空间，加速了群体悲愤情绪的极化和动员。

(二) 治理建议

基于上述结论，就目前舆情治理中存在模态割裂和方言识别的痛点，提出从技术逻辑重塑、符号规制创新、价值理性回归的“三维一体”的治理策略，实现治理模式由被动阻断到动态精准治理的转变：

1. 动态重构：构建“自适应动态方言情感词典”，规避语义漂移风险

方言词汇在网络跨平台流变中存在“高开低走”、极性反转的典型规律，部分原本情感倾向明确的方言词会随传播语境稀释逐渐弱化语义强度，中性方言词也可能在特定视听场景中异化为讽刺或悲愤符号。方言词汇网络流变现象使传统的静态方言情感词典识别精度大幅降低，难以满足复杂舆情治理需求，因此治理主体要摒弃依靠静态词典的线性思维，创建动态语境加权的敏捷感知体系。该体系的核心要义在于系统整合文本语义分析、视听模态识别、传播语境追踪等关键技术模块，实现对词汇语义动态演变的全域感知与精准捕捉。就方言舆情传播而言，经过长期的实证数据筛选，把易产生语义反转的高频方言词如“雄起”、“撇脱”等纳入到重点监测的视野中，采用深度学习里的注意力机制，对不同模态的数据进行特征提取并分配权重，准确地聚焦在副语言模态和文本语义的交互上，实时计算背景音乐、关键帧表情等要素对于文本语义的调节权重。通过建立算法实时捕捉、智能预警与人工复核校准相结合的动态闭环，当系统捕捉到短时间内出现图文情感极性相悖的异常信号的时候，就会立刻启动语义校验流程，对某一时间段里方言词汇语义含义进行动态修正，保证对网络语言变异的快速反应，规避由于语义漂移而导致的舆情误判风险，为方言舆情的精准治理打下基础。

2. 算法升级：部署“跨模态一致性校验算法”，阻断隐喻性反讽传播

言舆情传播中存在的“明褒暗贬”的典型隐晦传播特征，给传统平台审核算法单维度的违规检测模式带来了治理困境。单维度违规检测算法难以捕捉跨模态信息的语义冲突，无法满足复杂的舆情治理需求，因此平台审核算法需要从单维度的文本或者视听信息检测，升级为融合文本、音频、视觉多模态信息的跨模态一致性校验机制。从算法角度可创建文本和视听模态的情感量化评价体系，设定两者情感得分的差值阈值，本文通过实证研究发现情感差值 $\text{Gap} > 1.0$ 时反讽识别准确率可以达到 89.7%，该阈值具备显著的实践有效性。具体判定规则为，当内容文本表现出明显的积极倾向的正面评价，但配乐和画面元素却传递出高强度的负面情感，并且两者的情感得分差值超过上述验证阈值时，就可以判定为疑似反讽内容。高反讽风险的内容如果只删除或者放行便会产生治理偏差，需要采取差异化的干预策略，利用流量降权来削弱其传播势能，再辅以人工复审来精准识别内容性质，组合策略既能避免算法误判造成的合理表达受限，又能有效阻断利用方言隐喻、跨模态语义冲突等方式煽动群体对立的传播链路，给方言舆情的精细化治理提供技术支撑。

3. 符号规制：建立“非语言符号灰名单库”，通过视听要素干预情绪动员

BGM 和表情包在方言舆情演化过程中起到情感放大器的作用，悲情旋律可快速调动群体情感共鸣、加速负面情绪的扩散，高频反讽表情包则会加强语义反转效应、扩大争议话题的传播声量。由于传播特性，传统的以主体管控、文本规制为主的治理模式，无法准确地覆盖隐性情绪动员的风险，治理重心因此要向“符号规制、场景治理”的微观层面深入。构建非语言风险符号“灰名单”库成为核心治理抓手，在库内重点收录经过多模态舆情数据实证检验的高风险符号品类，涵盖悲情旋律、高频反讽表情包等核心类型，同时构建全生命周期精细化干预机制。从实践角度来

讲, 舆情潜伏期要加大含“灰名单”符号内容的审核优先级, 前置阻断潜在情绪动员隐患。爆发期之后, 对承载强煽动性视听符号的短视频内容实施“音量标准化”调控、“推荐流量打散”等差异化干预措施, 削减极端情绪的流传势能。借助全周期分阶段的精准规制策略, 可有效地抑制视听符号驱动的跨平台非理性情感共振, 为理性公共讨论空间的培育预留出足够的缓冲区间, 筑牢方言舆情跨平台治理的微观防御屏障。

4. 价值重塑: 引入“算法共情”机制, 实施基于地域文化的柔性治愈

方言舆情跨平台治理存在刚性管控易引起群体抵触、算法流量导向会加重极端情绪传播的现实困境, 传统刚性管控一般对争议方言内容采取简单屏蔽的方式, 流量导向算法更倾向于推送高热度的争议内容, 从而加剧情绪对立。借助融合地域文化语境的“算法共情”机制, 落地以地域文化为根基的柔性治愈路径, 使推荐算法从流量逻辑中解脱出来, 回归到“算法向善”的价值理性, 在分发机制里加入“共情”维度。当检测到某个方言区出现高强度负面舆情的時候, 算法不应该只屏蔽掉负面词汇, 而应及时调整流量权重, 将该地区理性、温情或者基于客观事实的优质内容优先推送。以当地意见领袖的方言解读作为情感中和剂, 倚重当地意见领袖熟悉本地文化语境并具有公信力的方言解读作为情感中和剂, 在信息流中可有效地稀释极端情绪的浓度。依靠方言天然的文化亲和力来修复被破坏的网络信任, 使治理范式从刚性的“硬性管控”转变为有温度的“柔性治愈”, 实现舆情治理效能与地域文化认同的双重目标。

参考文献:

- [1]潘艳艳. 国家安全视阈下网络多模态语言舆情监测研究[J]. 外语研究, 2024, 41(05):60-65.
- [2]祝峰, 杨明宏. 语言舆情视域下西南边疆少数民族地区语言教育规划研究[J]. 民族教育研究, 2020, 31(04):141-148.
- [3]魏宏程, 朱恒民, 魏静, 等. 基于短视频网络的互联网舆情演化研究[J]. 数据分析与知识发现, 2024, 8(05):113-126.
- [4]刘逸伦, 黄微, 张晓君, 等. AIGC 赋能的科技情报智能服务: 特征、场景与框架[J]. 现代情报, 2023, 43(12):88-99.
- [5]Chen,T.;Yin,X.;Yang,J.;Cong,G.;Li,G.Modeling Multi-Dimensional Public Opinion Process Based on Complex Network Dynamics Model in the Context of Derived Topics[J].Axioms,2021,10(2):270-302.
- [6]王晰巍, 王小天, 李玥琪. 重大突发事件网络舆情 UGC 的事理图谱构建研究—以自然灾害 7.20 河南暴雨为例[J]. 图书情报工作, 2022, 66(16):13-23.
- [7]李路宝, 陈田, 任福继, 等. 基于图神经网络和注意力的双模态情感识别方法[J]. 计算机应用, 2023, 43(03):700-705.
- [8]范涛, 吴鹏, 曹琪. 基于深度学习的多模态融合网民情感识别研究[J]. 信息资源管理学报, 2020, 10(1):39-48.
- [9]马庆懿. 基于多模态数据融合的突发公共事件网络舆情情感分析研究[D]. 吉林大学, 2025.
- [10]李振, 张洋, 李智超, 等. 基于 GNN 多模态融合的高校舆情分析[J]. 计算机科学:2025(1):1-13.
- [11] Qiao Y, Yin Y, Yao F, et al. Multi-View Interaction Network for Multimodal Sarcasm Detection[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2023, 19(2s):1-21.
- [12]Liang B, Chen C, Lou X, et al. Multi-Modal Sarcasm Detection via Cross-Modal Graph Convolutional Network[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers).Dublin, Ireland: ACL,2022:1767-1777.
- [13]Chen S,Wang G,Liu J,et al.Dual-channel graph convolutional networks with sentiment knowledge for aspect-based sentiment analysis[J].Knowledge-Based Systems,2023,260:110156.
- [14]Zhang W,Deng Y,Liu B, et al. Senti-LLM:Enhancing Sentiment Analysis with Large Language Model Generated Context[J]. Information Processing&Management,2024,61(3):103-125.
- [15]Hasan A,Moin S,Karim A, et al.A Survey on Multimodal Sentiment Analysis:Recent Trends, Challenges, and Future Directions[J].IEEE Access,2023,11:12567-12590.
- [16]刘云花, 黎泉. 基于自然语言语义感知的舆情分析算法设计[J]. 电子技术, 2025, 48(22):133-137

Research on the“Semantic Reversal”Mechanism and Precise Governance of Dialect-Based Public Sentiment in a Multimodal Context:A Cross-Platform Analysis Based on Chongqing's“Fat Cat Incident”

ZHU Na*, JIANG Lingxian, PENG Jiao

*(Chongqing Metropolitan College of Science and Technology, School of Economics and
Management, Yongchuan, Chongqing 420167)*

Abstract: Addressing the challenges of modal fragmentation and dialect-based“semantic divides”in multimodal online public opinion governance, this study employs the Chongqing“Fat Cat Incident”as an empirical case to construct a dialect-context-oriented cross-modal semantic alignment model (D-CMSA).Through cross-platform comparative analysis of 8,542 Weibo posts and 312 high-engagement Douyin videos,we measured the emotional evolution trajectories of dialect vocabulary and the mobilization mechanisms of visual symbols.Empirical validation confirms that the D-CMSA model possesses the capability to uncover deep public sentiment features difficult to capture through single-modality analysis.These features can be distilled into two core effects:First, the sentiment amplification effect,where the co-occurrence of negative audiovisual symbols increases negative sentiment intensity by an average of 42.9%. Second,the semantic reversal effect:28.5% of samples demonstrate that dialect texts with neutral surface semantics are reconfigured into ironic or resentful symbols within specific visual contexts,exhibiting significant phased semantic drift characteristics.This study proposes establishing a dynamic dialect sentiment lexicon to address semantic drift and employs cross-modal consistency verification algorithms to prevent covert dissemination in public opinion governance.

Keywords: Multimodal semantic fusion; Dialect-based public opinion; Audiovisual metaphor; Emotional amplification; Semantic reversal