

基于情感分类和库普曼算子的网络舆情热度预测

王柯莹 高焯 李松
(河北大学管理学院, 保定 071002)

摘要: 为充分挖掘情感因素在舆情监测和管控中的作用, 精准预测网络舆情热度趋势, 本研究提出一种基于情感分类和库普曼算子的网络舆情热度预测方法。首先, 构建基于BERT-CGA多特征融合的情感分类模型, 充分提取情感特征并学习特征间关联, 对博文内容进行情感极性分类; 其次, 将情感特征和库普曼算子理论引入舆情热度预测, 构建融合情感因素的LSTM-EDMD热度预测模型, 以LSTM算法为观测函数实现低维舆情热度指标数据到高维空间的映射, 通过EDMD算法生成库普曼算子刻画系统动态特性实现热度预测, 增强模型可解释性; 最后, 以银川烧烤店爆炸事件为例进行实证研究, 实验结果验证情感因素在网络舆情热度预测中的关键作用, 同时表明所提方法具有较高的预测准确性。本研究有助于政府、企业等舆情应对主体及时捕捉情感动向, 准确预测舆情热度演变并制定针对性策略。

关键词: 网络舆情; 热度预测; 库普曼算子; 情感分类; 情感因素

中图分类号: G206.3 **DOI:** 10.3772/j.issn.1673-2286.2025.08.002

引文格式: 王柯莹, 高焯, 李松. 基于情感分类和库普曼算子的网络舆情热度预测[J]. 数字图书馆论坛, 2025, 21(8): 11-22.

随着社交媒体与移动互联网的深度融合, 网络空间已成为公众表达观点、情感和态度的重要场所。突发公共事件往往借助微博等平台迅速引发关注和讨论, 推动舆情热度急剧攀升, 形成强大的社会影响力。舆情热度虽在一定程度上有助于推动问题的高效解决, 但缺乏合理引导又会诱发谣言与极端情绪, 增加舆情治理难度。因此, 对舆情热度趋势进行精准预测是实现舆情科学管控的关键。大数据和人工智能等技术的迅猛发展, 为网络舆情热度预测研究开辟了新的空间: 一方面, 利用先进技术可以高效处理海量的网络数据并进行深度挖掘, 从而更精准地预测舆情走势; 另一方面, 新技术也有助于研究者深化对网络舆情的形成机制和传播规律的理解, 为舆情热度预测提供更加科学的理论支持。

在网络空间中公众针对各种事件发表的观点和态度常常带有情感色彩, 情感因素不仅影响公众对事件的认知、态度和行为, 更是推动舆情演变的关键驱动力^[1-2]。因此, 深入分析情感因素对洞察舆情热度发展趋势至关重要, 能够帮助政府、企业等有关部门及时采取应对措施, 从而有效防范舆情危机。同时, 基于库普曼算子理论的方法适合学习非平稳时间序列动态特性^[3], 能够提取数据随时间的动态演变特征, 通过高维线性近似揭示系统的内在演化机制, 将其与神经网络模型相结合可提升模型的可解释性。基于此, 本文提出了一种基于情感分类和库普曼算子的网络舆情热度预测模型, 从动力学视角完善情感因素驱动的舆情热度预测相关研究, 突破传统神经网络预测模型在可解释性上的固有限局, 为舆情热度研究提供跨学科的分析维度。

1 相关研究

1.1 基于深度学习的情感分类方法研究

基于深度学习的情感分类模型通常将词向量视为文本输入表示。Kim^[4]采用Word2Vec词嵌入与卷积神经网络(CNN)结合的情感分类方案,但Word2Vec的静态特性使其无法适应词语的多义性,为此Du等^[5]引入能根据上下文语境动态生成词向量的BERT模型以解决该问题。此外,为克服CNN因难以捕捉长距离上下文文依赖而导致未充分提取特征的局限,曹宇等^[6]构建基于双向门控循环单元(BiGRU)的中文文本情感分析模型,实验结果表明该模型的分类性能超越CNN和BiLSTM等基准模型。为了更好地捕获长期依赖关系和局部特征,Basiri等^[7]提出了一种基于注意力的双向CNN-RNN深度模型,该模型在长短文本情感极性分析中均取得较好的结果。

为综合发挥不同神经网络架构的优势,学者们普遍采用模型融合策略进行情感分析。曾子明等^[8]构建BERT-BiLSTM-Attention模型,通过集成BiLSTM长序列建模能力和注意力机制聚焦关键信息的能力,显著强化对文本中核心情感特征的识别,该模型在多项任务中展现出卓越的分类效果。韩坤等^[9]构建的文本分类模型BERT-MLFF-TextCNN具备多层次特征提取能力,不仅能挖掘文本的深层语义信息,还能有效捕捉表层语法结构和词汇特征。然而,目前多数模型在进行特征融合时直接采用特征向量拼接的方式,导致特征间关系学习不足^[10]。

1.2 舆情热度预测研究

为预测网络舆情热度,研究者普遍采用构建定量模型的方法,典型代表是时间序列聚类和时间序列预测。时间序列聚类方法将形态相近的热度序列归为同一类别,依据识别出的共性变化规律预测舆情热度的未来趋势,但此类预测具有片面性,忽视了舆情事件的相关影响因素。时间序列预测模型通过挖掘事件历史数据中的潜在规律并建模,实现对未来热度趋势的预测,主要分为两类:基于传统的时间序列预测模型和基于机器学习算法的预测模型。鉴于单一模型的预测性能有限,整合不同模型优势的混合策略日益成为主流研究方向。林育曼等^[11]结合ARIMA模型和BP神经网络

进行微信舆情的热度预测。景楠等^[12]基于ARIMA模型和LSTM网络进行预测和分析舆情数据。冯勇等^[13]引入SDZ(Suprisal-Driven Zoneout)方法改进LSTM,预测准确度和时效性均优于基于神经网络的预测模型。马永军等^[14]在LSTM中引入注意力机制,提高了预测的准确度。

情感因素作为网络舆情监测中的重要内容^[15],已被广泛引入热度预测研究。牟冬梅等^[16]设计了融合情感特征的网络舆情热度预测模型,实验结果表明情感特征的引入能有效提升预测准确率。杨柳等^[17]考虑情感因素构建KNN-SVR-RF组合预测模型,通过实例证明在引入情感特征且时间窗口滞后6天的设定下,其舆情热度预测性能最优。赵嵩正等^[18]则尝试基于负面评论率与热度突变频次来划分舆情预警等级。由于舆情发展受多种因素的影响,呈现出内在的非平稳特性,现有的基于情感分类的热度预测模型通常为传统时间序列分析方法,通过对舆情指标数据进行平稳化处理简化其复杂的演变过程,无法充分挖掘和利用时序信息。基于神经网络的热度预测模型通过多层非线性变换学习数据中的复杂模式,但其“黑箱”特性使最终决策结果难以与原始输入建立直观的因果关系,导致模型可解释性普遍较差。

库普曼算子理论^[19]的基本思想是将有限维非线性系统全局线性化,转换为无穷维线性动力系统。实践中计算无穷维库普曼算子的有限维近似是一大难点,近年来利用动态模态分解(DMD)、扩展动态模态分解(EDMD)及相关改进算法较好地解决了该问题。目前国外学者已将库普曼算子与深度学习方法结合进行时间序列预测,并应用于电力系统^[20]、控制论^[21]等领域。国内研究相对较少,如岳崇浩^[22]使用库普曼算子进行疫情预测。库普曼算子与深度学习模型融合进行非线性时间序列预测,能够为理解舆情热度的演化提供清晰的数学表征,显著提升舆情热度预测方法的科学性和结果的准确性。

综上所述,结合情感分析进行网络舆情热度预测已经成为研究的重点,学者们从多角度探索融合方案并取得了显著进展,但仍存在一些不足:①情感分类时特征提取不充分以及特征间关系学习不足;②传统的神经网络预测模型可解释性普遍较差。因此,本文从两个方面展开研究。首先,构建基于BERT-CGA(BERT-TextCNN-BiGRU-Attention)多特征融合的情感分类模型,通过集成BERT动态上下文编码、TextCNN局部

特征提取与BiGRU全局语义捕获的优势,并借助注意力机制实现特征间的关系学习与深度融合,有效提升情感特征表征效果与关联挖掘能力。其次,构建融合情感因素的LSTM-EDMD网络舆情热度预测模型,在情感分类结果基础上,利用LSTM将舆情指标数据映射到高维特征空间,并使用EDMD学习特征数据的动态演变特征,最终通过线性演化机制近似揭示舆情热度的内在动态规律,增强模型的可解释性,为舆情热度的精准预测提供更有效的技术路径。

2 研究设计

2.1 研究框架

研究框架如图1所示,包括数据获取与预处理层、特征提取层和热度预测层。

(1) 数据获取与预处理层:利用Python的Scrapy爬虫框架,经由微博端口获取指定舆情事件相关的原创微博及互动指标(转发量、评论量和点赞量);对博文内容进行数据清洗,剔除与话题无关的博文后按时间排序,并进行分词、去停用词操作。

(2) 特征提取层:从预处理后的数据中提取传播影响力特征、情感驱动特征和时序热度特征以支持舆情后续热度预测。传播影响力特征直接从原始数据中获取,包含博文数量和互动数据(点赞量、评论量和转发量);情感驱动特征通过构建的BERT-CGA多特征融合情感分类模型生成,获得积极情感值和消极情感值;时序热度特征来源于“知微事见”平台热度数据^[23]。

(3) 热度预测层:将传播影响力特征和情感驱动特征视为热度指标数据的相关因素,时序热度特征视为预测模型的舆情热度标签,辅助构建LSTM-EDMD模型实现舆情热度预测。

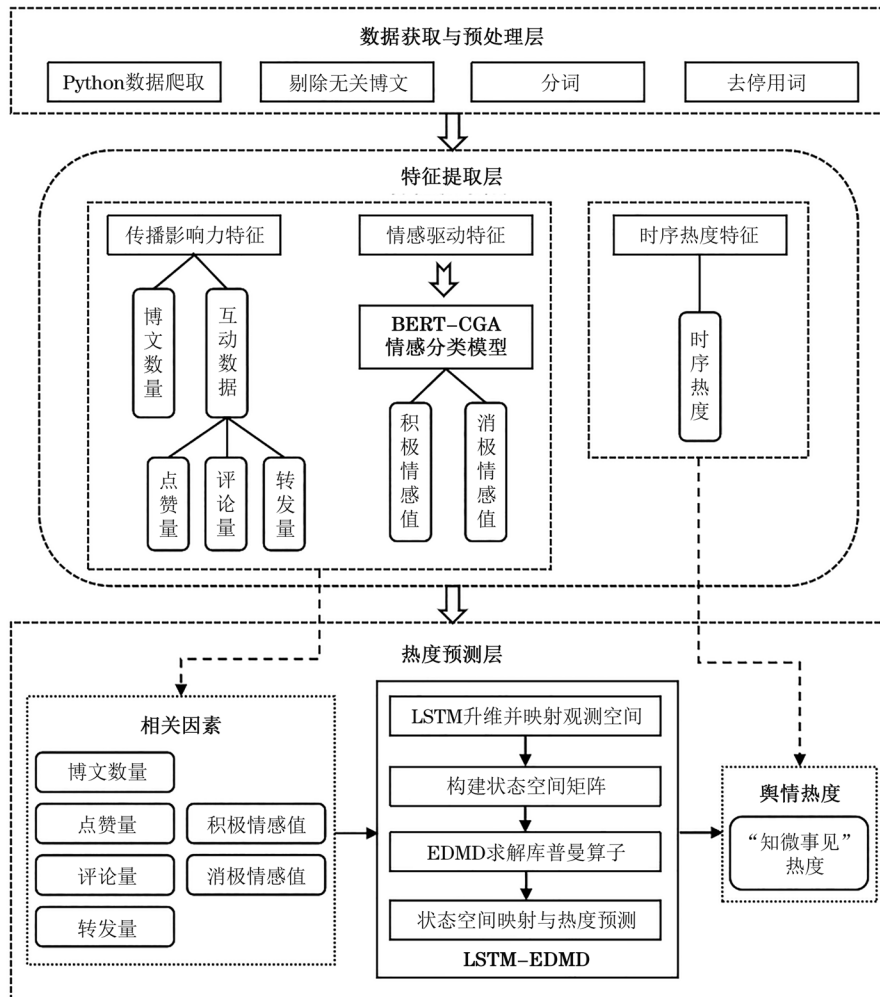


图1 研究框架

2.2 舆情热度预测的特征要素

网络舆情热度的演化是多重因素共同作用的结果，受事件传播范围和公众参与程度的直接影响，同时与群体情感倾向和热度演变规律密切相关。为构建严谨有效的网络舆情热度预测模型，须基于舆情传播规律与数据特性，系统分析驱动热度演化的关键特征要素，为后续模型构建提供清晰的变量基础。舆情热度预测的关键特征要素如表1所示。

(1) 传播影响力特征。现有研究普遍将博文数量和互动数据视为舆情热度的重要影响因素^[24-26]。原创博文是舆情讨论源头，单位时间内的数量体现公众参与意愿与话题实时活跃度。互动数据从多个维度刻画公众参与的深度：点赞量反映公众对感兴趣或欣赏的内容的认同或喜爱^[25]；评论内容承载公众对事件的具体观点，数量激增代表舆情进入聚集阶段；公众对事件的转发行为进一步扩大舆情覆盖范围。因此，博文数量和互动数据共同构成舆情传播影响力特征的核心要素，

直观反映舆情的扩散广度和公众参与深度。

(2) 情感驱动特征。情感驱动特征是反映公众情感倾向与情绪强度，推动热度非平稳变化的关键内在因素。公众的情感不仅影响个体舆情参与行为，也会通过群体情绪的聚集与扩散主导热度的突变趋势。牟冬梅等^[16]和杨柳等^[17]验证了情感值在突发事件网络舆情热度预测中的有效性。因此，通过构建高效的情感分类模型从原创博文中获取积极情感值和消极情感值两个核心要素，分别定义为单位时间内被判定为积极/消极情感的原创博文数量，以量化该时段公众对事件的正向/负向态度。

(3) 时序热度特征。时序热度特征能够反映舆情热度的历史演变轨迹，捕捉热度周期性与阶段性规律。舆情热度的波动存在显著的时间关联，呈现“日间高夜间低”的周期性波动和“潜伏期—爆发期—持续期—衰退期”的阶段性发展^[27]。知微事见平台热度数据综合博文量、互动量多维度信息，能真实反映舆情事件的整体热度，因此选取该平台小时级热度数据作为预测模型的真实标签。

表1 舆情热度预测的关键特征要素

特征类别	特征要素	要素含义	数据来源
传播影响力特征	博文数量	单位时间内原创博文量，体现公众参与讨论的意愿	微博平台原始数据
	互动数据	点赞量：反映公众对感兴趣或欣赏的内容的认同或喜爱 评论量：反映公众观点输出和深度参与意愿 转发量：反映舆情信息二次传播强度	
情感驱动特征	积极情感值	量化公众对舆情事件的支持、赞扬和乐观等正向态度	情感分类模型对原创博文的判定结果
	消极情感值	量化公众对舆情事件的批评、愤怒和担忧等负向态度	
时序热度特征	时序热度	反映舆情事件在时间维度上的热度变化趋势和规律	“知微事见”热度数据

2.3 基于BERT-CGA多特征融合的情感分类模型

基于BERT-CGA多特征融合的情感分类模型用于对原创博文数据进行情感分类，获取积极情感值和消极情感值，包含BERT编码层、多特征提取层、特征融合层和输出层等核心模块，模型结构如图2所示。在BERT编码层，利用BERT预训练语言模型生成上下文感知的词向量表示；在多特征提取层，使用由3个不同尺寸的卷积核组成的TextCNN模型进行文本局部特征的提取，使用多层BiGRU网络正向和反向捕获文本的全局特征；在特征融合层，将局部与全局特征向量共同输入多头注意力机制，捕捉特征之间的关系，得到关系特征向量，并引入软注意力机制动态拼接局部、全局

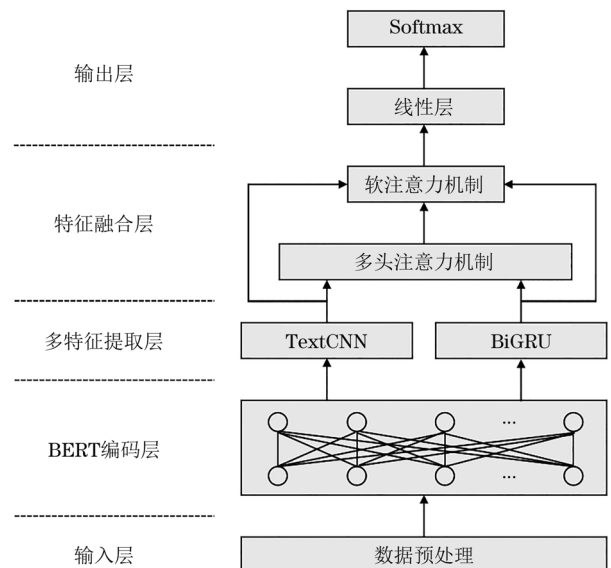


图2 BERT-CGA模型结构

和关系特征向量实现特征融合; 在输出层将融合后的特征向量输入线性层和Softmax层, 实现文本的情感分类。

2.3.1 BERT编码层

使用预训练bert-base-chinese模型对输入文本进行词向量编码, 该模型能够根据文本的上下文信息进行词向量的训练, 提取出丰富的语义表示。bert-base-chinese模型包含12层Transformer编码器, 每层编码器都会输出一组向量 E_i 。底层编码器输出的向量包含词性和语法结构等局部特征, 随着层数增加, 输出向量逐渐捕捉到更高级抽象的语义信息。每层编码器为[CLS]标识生成一个[CLS]向量, 综合了该层对整个输入文本的理解。为了融合不同Transformer层的语义信息, 将每层Transformer编码器的[CLS]向量 c_i 堆叠, 构造特征矩阵 C , 计算公式如式(1)所示。

$$C = c_1 \oplus c_2 \oplus \dots \oplus c_{12} \quad (1)$$

2.3.2 多特征提取层

(1) 局部特征提取。为了获取文本的局部语义特征, 在BERT词向量编码的基础上, 引入多尺度TextCNN模型。将特征矩阵 C 作为模型的输入, 选用3个不同尺寸的卷积核分别捕捉不同粒度的局部语义特征。池化层采用最大池化保留关键语义特征, 在防止模型过拟合的同时提升计算速度。进一步将不同尺度卷积池化的特征向量进行拼接, 获得最终的局部特征向量 A , 计算公式如式(2)~式(3)所示。

$$A_j = \text{Maxpool}(\text{ReLU}(\text{Conv}_j(C))) \quad (2)$$

$$A = A_1 \oplus A_2 \oplus A_3 \quad (3)$$

式中: A_j 表示第 j 个卷积核卷积池化后的输出特征。

(2) 全局特征提取。BiGRU模型的双向处理机制综合考虑文本从前到后及从后到前的信息依赖关系, 更全面地捕捉文本中的上下文信息, 实现对文本语义特征的深层次挖掘。BiGRU使用两个相反的门控循环单元(GRU)进行特征提取, 二者通过更新门和重置门更新隐藏状态, 具体结构如图3所示。

BiGRU层的输入为BERT模型最后一层编码器生成的输出向量 E_{12} , 经过BiGRU网络得到正向隐藏状态

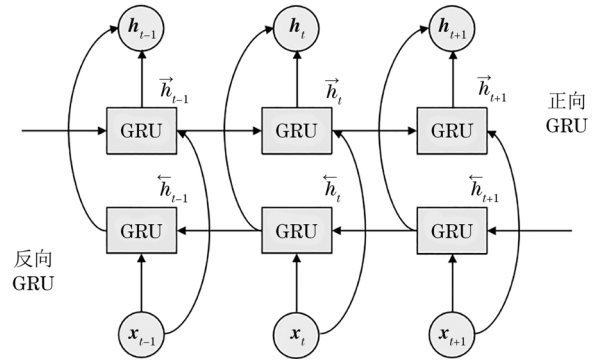


图3 BiGRU结构

输出 \vec{h}_t 和反向隐藏状态输出 \overleftarrow{h}_t , 通过拼接操作得到模型在 t 时刻的输出 h_t , 计算公式如式(4)~式(6)所示。

$$\vec{h}_t = \text{GRU}(\vec{h}_{t-1}, x_t) \quad (4)$$

$$\overleftarrow{h}_t = \text{GRU}(\overleftarrow{h}_{t+1}, x_t) \quad (5)$$

$$h_t = [\vec{h}_t \parallel \overleftarrow{h}_t] \quad (6)$$

式中: x_t 为模型在 t 时刻的输入向量。经过BiGRU的特征提取后得到最终的全局特征向量 B 。

2.3.3 特征融合层

为解决特征融合过程中特征间关系学习不充分的问题, 采用多头注意力机制学习局部特征向量 A 和全局特征向量 B 的关系, 得到关系特征向量 R 。多头注意力机制的计算公式如式(7)~式(8)所示。

$$H_i = \text{Attention}(Q_i^A W_i^Q, K_i^B W_i^K, V_i^B W_i^V) \quad (7)$$

$$R = \text{Concat}(H_1, \dots, H_h) W^0 \quad (8)$$

式中: Q_i^A 、 K_i^B 、 V_i^B 分别表示第 i 个头的查询、键和值; W_i^Q 、 W_i^K 、 W_i^V 、 W^0 表示对应的权重; h 表示“头”的数量, 通过多个“头”进行并行计算, 从不同的子空间学习局部与全局特征之间的关系, 捕获更丰富和全面的信息。

在进行多特征融合时, 由于不同特征对情感分类的重要性存在差异, 单一拼接操作忽略各类特征对情感分类的贡献程度。因此引入软注意力机制模块, 为局部、全局和关系特征动态分配权重, 最终得到加权融合后的特征向量 F , 计算公式如式(9)~式(10)所示。

$$\alpha_n^M = \text{Softmax}(\tanh(W_{\text{att}} M + b_{\text{att}})) \quad (9)$$

$$F = \sum_M \alpha_n^M M \quad (10)$$

式中： W_{att} 和 b_{att} 分别为软注意力机制的权重矩阵和偏置； α_n^M 为注意力分布， $M \in \{A, B, R\}$ 。

2.3.4 输出层

输出层由全连接层和Softmax层组成，全连接层通过非线性变换将特征映射至情感标签类别维数，经Softmax层输出分类概率矩阵，得到情感标签。

2.4 融合情感因素的LSTM-EDMD网络舆情热度预测模型

为有效分析网络舆情热度演化的复杂非线性动态过程，引入库普曼算子理论将网络舆情热度变化的状态空间映射到高维线性空间，从而揭示其内在规律并简化动态行为的描述与预测。在高维线性空间中，系统演化由库普曼算子驱动，特征模态具有明确的物理意义，能够清晰表征系统内在的演化规律，显著增强模型的可解释性。

此外，鉴于情感因素是驱动舆情热度演变的关键变量，情感值成为推算热度值的重要因素。因此，本文提出融合库普曼算子理论及其演化算法EDMD进行舆情热度预测，同时将情感因素有机融入热度指标，实现更精准的舆情态势洞察。

在EDMD模型中，观测函数的选取至关重要，需要实时监测数据的状态变化并反映在观测值中。学者们多采用多项式^[22]、径向基函数（RBF）^[28]等作为观测函数，但其无法灵活捕捉复杂非线性关系且对噪声数据敏感。LSTM具备卓越的时序建模能力，擅长捕捉时间序列中的长期依赖与跨周期关联，对于动态演变的舆情热度数据，能够深度挖掘其关键时序特征；同时LSTM可以学习数据中含非线性关系在内的复杂动态模式，为EDMD算法提供更丰富的观测信息。因此，选取LSTM作为观测函数学习每个时间片段中的动态关系，模型预测过程如图4所示。

时间序列预测通常使用之前 T 个时刻的值预测下一时刻的值，即使用 $t-T, t-(T-1), \dots, t-2, t-1$ 时间段内的热度指标数据预测 t 时刻的热度值 y_t ，则模型的输入数据为 $\mathbf{X} = [\mathbf{x}_{t-T}, \mathbf{x}_{t-(T-1)}, \dots, \mathbf{x}_{t-1}]$ 。根据库普曼算子理论， \mathbf{X} 为状态空间的热度指标数据，热度指标的相关因素包

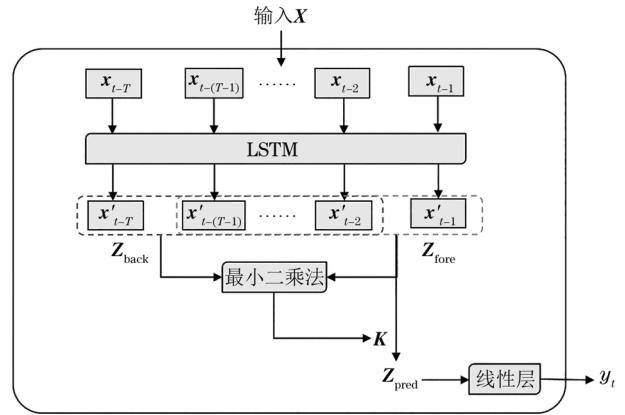


图4 融合情感因素的LSTM-EDMD网络舆情热度预测模型

含传播影响力特征（博文数量、转发量、评论量和点赞量）和情感驱动特征（积极情感值、消极情感值）两个部分，需要将其映射到观测空间。使用LSTM算法将状态数据 \mathbf{X} 升维并映射到观测空间中，得到高维观测数据 \mathbf{X}' ，计算公式如式（11）所示。

$$\mathbf{X}' = \text{LSTM}(\mathbf{X}) \quad (11)$$

为利用EDMD算法估计描述系统演化的库普曼算子，需要将高维观测序列划分为历史观测矩阵 \mathbf{Z}_{back} 和未来观测矩阵 \mathbf{Z}_{fore} ，矩阵的具体形式如式（12）~式（13）所示。

$$\mathbf{Z}_{back} = [\mathbf{x}'_{t-T}, \mathbf{x}'_{t-(T-1)}, \dots, \mathbf{x}'_{t-2}] \quad (12)$$

$$\mathbf{Z}_{fore} = [\mathbf{x}'_{t-(T-1)}, \mathbf{x}'_{t-(T-2)}, \dots, \mathbf{x}'_{t-1}] \quad (13)$$

式中： \mathbf{Z}_{back} 包含从 $t-T$ 到 $t-2$ 时刻的高维观测向量； \mathbf{Z}_{fore} 包含从 $t-(T-1)$ 到 $t-1$ 时刻的高维观测向量，可视为 \mathbf{Z}_{back} 在下一时刻的演进状态。综上，库普曼演化算子 \mathbf{K} 可以近似表示为式（14）。

$$\mathbf{K} \approx \mathbf{Z}_{fore} \mathbf{Z}_{back}^{\dagger} \quad (14)$$

式中： $\mathbf{Z}_{back}^{\dagger}$ 表示 \mathbf{Z}_{fore} 的Moore-Penrose伪逆。

利用最小二乘法最小化误差项来逼近库普曼算子 \mathbf{K} ，计算出的 \mathbf{K} 随时间窗口变化，有助于将局部时间变化视为线性系统进行分析。通过库普曼算子 \mathbf{K} ，可预测 $t-(T-2), t-(T-3), \dots, t$ 时间内各指标在观测空间的状态 \mathbf{Z}_{pred} ，计算公式如式（15）所示。通过线性层将其映射到状态空间中，得到舆情在第 t 时刻的热度值 y_t 。

$$\begin{aligned} \mathbf{Z}_{pred} &= [\mathbf{x}'_{t-(T-2)}, \mathbf{x}'_{t-(T-3)}, \dots, \mathbf{x}'_t] \\ &= \mathbf{K} [\mathbf{x}'_{t-(T-1)}, \mathbf{x}'_{t-(T-2)}, \dots, \mathbf{x}'_{t-1}] \end{aligned} \quad (15)$$

3 实证研究

3.1 数据获取及预处理

2023年6月21日的宁夏银川烧烤店爆炸事件经央视新闻新闻报道后在微博平台受到广泛关注与讨论。该事件属于典型的突发公共安全事件, 直接关联公众生命安全, 事后舆情热度高, 情感倾向明显, 具备完整的舆情生命周期, 能有效检验模型对非平稳舆情的预测能力。此外, 微博平台提供的原创博文、点赞/评论/转发等互动数据及“知微事见”热度数据能够满足实证分析的数据需求。因此, 选取银川烧烤店爆炸事件为研究案例。

由于该事件在2023年6月28日基本平息, 选取2023

年6月21—28日为研究时间段。通过Python网络爬虫, 以“银川烧烤店爆炸”为关键词, 爬取以上时间段内的所有原创微博数据, 具体字段包括发布时间、用户ID、用户昵称、微博正文、点赞量、评论量和转发量。通过剔除不相关博文, 得到11 355条有效微博数据。对微博正文依次进行数据清洗、中文分词和去停用词处理。

在“知微事见”舆情研究平台以小时为单位获取研究时间段内的热度数据并绘制舆情热度曲线, 如图5所示。从热度曲线中可以看出, 银川烧烤店爆炸事件舆情呈单峰传播趋势, 这是因为该事件属于突发公共安全事件, 具有强突发性和不可预测性, 通常在事件发生后迅速引发网民关注, 故将该舆情事件生命周期划分为潜伏期、爆发期、持续期和衰退期4个阶段。

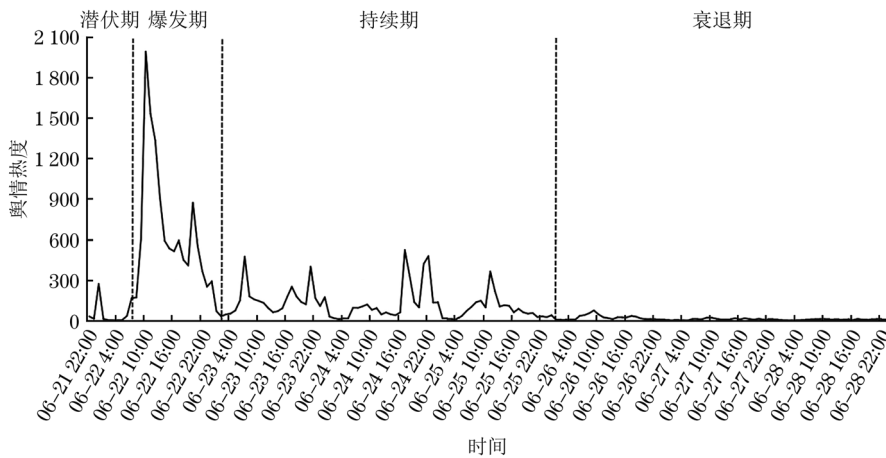


图5 银川烧烤店爆炸事件热度曲线

3.2 情感分类及分析

3.2.1 对比实验与结果分析

为了验证基于BERT-CGA多特征融合的情感分类模型的有效性, 设置Word2Vec-TextCNN、Word2Vec-BiGRU、BERT、BERT-TextCNN、BERT-BiLSTM、BERT-BiGRU、BERT-TextCNN-BiGRU这7个模型进行对比实验, 实验结果如表2所示。

通过对实验结果对比分析, 可得出以下结论: ①相较于传统的Word2Vec词向量模型, BERT预训练模型的准确率更高; ②BERT-TextCNN、BERT-BiLSTM、BERT-BiGRU这3个模型的各项结果较为接近, 在BERT模型的基础上加入TextCNN、BiLSTM、BiGRU特征提取层的准确率有所提升; ③与单特征提取层相比, BERT-TextCNN-BiGRU模型特征提取能力更强,

表2 对比实验结果

单位: %

模型	准确率	精确率	召回率	F1值
Word2Vec-TextCNN	80.24	77.60	72.47	74.51
Word2Vec-BiGRU	79.89	76.38	73.14	74.17
BERT	87.23	87.91	80.92	83.65
BERT-TextCNN	87.82	85.77	84.25	84.96
BERT-BiLSTM	87.80	86.01	83.82	84.84
BERT-BiGRU	87.80	85.78	84.82	84.93
BERT-TextCNN-BiGRU	88.18	86.32	84.64	85.41
BERT-CGA	88.93	87.22	85.36	86.28

指标值的上升证明同时提取局部特征和全局特征能够有效提高情感分类任务的准确率, 整体效果优于单一使用循环神经网络或CNN; ④BERT-CGA模型比BERT-TextCNN-BiGRU模型情感分类效果更好, 在多个评估指标上均有所提升。这一结果验证加入多头注意力机制能够学习文本不同特征间的内在关系, 能够

更全面准确地捕捉特征间的交互作用,进一步提高情感分类效果。

3.2.2 情感分类

统计银川烧烤店爆炸事件中所有博文的情感倾向,结果如表3所示。中性博文占比最高,主要源于大量媒体对该突发公共安全事件进行客观陈述与进展通报;消极博文占比次之,反映了事件引发的公众担忧与不满情绪较为显著;积极博文占比最低,表明该事件未产生积极影响,可能引发公众对公共安全与管理等方面的恐惧和深思,凸显出相关部门进行舆情管控和及时预测的重要性。

表3 银川烧烤店爆炸事件情感倾向占比

情感倾向	博文数量/条	占比/%
积极	966	8.50
消极	3 748	33.00
中性	6 641	58.50

3.3 网络舆情热度预测

在舆情爆发期和持续期,网民对事件关注度高,进行热度预测和管控至关重要。在舆情衰退期,热度明显下降且维持在较低水平,对社会和个人的实际影响减弱,预测价值较低。因此,将预测重点放在舆情热度较高且网民情绪较为明显的爆发期和持续期。

将2023年6月21—25日的原创微博按发布时间排序,以小时为单位提取博文数量、点赞量、评论量和转发量,与积极情感值和消极情感值进行合并求和,作为每小时舆情热度数据。将“知微事见”平台该事件热度数据作为预测模型的输出标签。

为了统一量纲并消除指标单位差异带来的影响,采用Min-Max归一化法对原始舆情数据进行标准化转换,计算公式如式(16)所示。

$$x'_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \quad (16)$$

式中: x_{ij} 表示第 i 个指标在第 j 个时间点的数据; $\min(x_{ij})$ 表示第 i 个指标的最小值; $\max(x_{ij})$ 表示第 i 个指标的最大值; x'_{ij} 表示标准化后的值,处于0~1。

使用BP神经网络、LSTM算法、SDZ-LSTM模型、BiGRU模型与所提出的基于情感分类和库普曼算子的

网络舆情热度预测模型进行对比实验。选取原始数据的75%作为训练集,剩余25%作为测试集。实验模型在Python 3.11平台构建,采用PyTorch框架搭建。实验采用均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)和判定系数 R^2 这4个指标对预测结果进行评价,计算公式如式(17)~式(20)所示。

$$I_{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (17)$$

$$I_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (18)$$

$$I_{MAE} = \frac{1}{n} \sum_{i=1}^n |(\hat{y}_i - y_i)| \quad (19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (20)$$

式中: \hat{y}_i 为预测值; y_i 为真实值, \bar{y}_i 为其均值; n 为预测值数量; R^2 取值范围为0~1,值越趋近1表明模型的预测结果与实际观测数据的一致性越高,反映出更优的拟合性能。

5个模型的预测误差对比如表4所示。相较于其他4个模型,提出的BERT-CGA+LSTM-EDMD模型在4个评价指标对比实验中表现最好。其中,相较于其他4个模型,本文提出的模型在MSE上分别降低46.50%、42.96%、34.09%、18.82%;在RMSE上分别降低26.85%、24.48%、18.81%、9.90%;在MAE上分别降低30.59%、14.23%、23.73%、10.95%; R^2 更趋近于1,模型拟合优度提升。

表4 模型预测误差对比

模型	MSE	RMSE	MAE	R^2
BP神经网络	16 267.77	127.54	91.58	0.84
LSTM	15 258.95	123.53	74.12	0.85
SDZ-LSTM	13 205.42	114.91	83.35	0.87
BiGRU	10 721.43	103.54	71.39	0.89
BERT-CGA+LSTM-EDMD	8 703.76	93.29	63.57	0.92

3.4 情感因素对舆情热度的影响验证

3.4.1 不同生命周期阶段的情感分析

舆情事件热度的波动呈现“潜伏期—爆发期—持

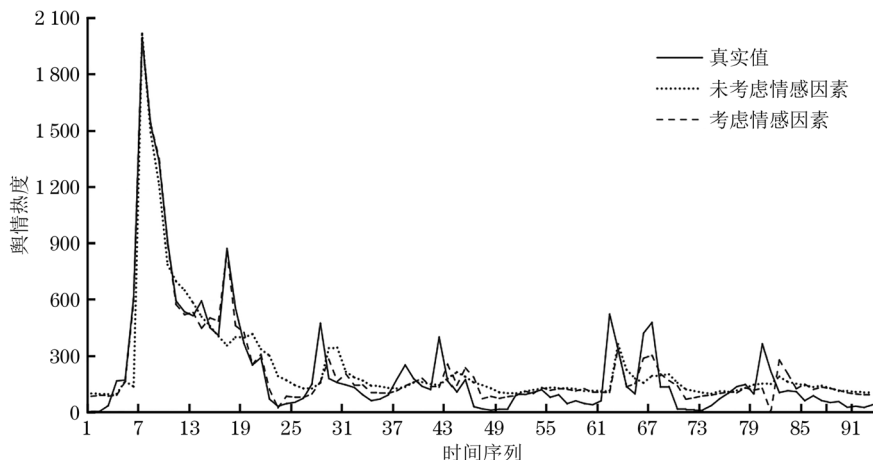


图8 有无情感因素预测结果对比

整理事件进展和公众关切回应材料备用。对社交平台而言,应建立突发公共安全事件标签识别机制,对单位时间内原创博文量突增的话题,自动向发布者推送信息真实性核验提示,并降低未标注来源博文的初始推荐权重,抑制虚假信息早期扩散。

(2) 爆发期: 消极情绪疏导与互动行为管控。爆发期是情感驱动热度攀升的关键阶段,应重点加强对消极情绪的疏导与网络互动管理,避免舆情热度失控式增长。对政府而言,应把握热度高峰期前1~2小时的关键窗口,通过官方微博、短视频平台发布可视化素材,重点解读事件处置进展、民生保障措施等公众关切内容,缓解消极情绪。对社交平台而言,应对单位时间内互动量激增且带有煽动性表述和极端情绪的博文实施临时限流,主动提升事件应对建议、风险防范思路等内容的曝光权重,引导公众焦点从情绪宣泄转向实际问题解决。

(3) 持续期: 情感波动跟踪与传播效果优化。持续期消极情感占比逐步下降,博文量和互动数据波动幅度减小,应密切跟踪情感动态,优化传播策略,推动热度平稳过渡。对于政府而言,应实时监测消极情感占比变化,当回升至10%时,须及时排查信息缺口并发布针对性回应。对社交平台而言,应实施差异化内容推送策略,结合“短视频+图文长帖”传播形式,根据互动反馈动态优化推送方式,提升传播有效性。面向普通群众,重点推送风险自查指南等实用信息,快速响应公众高频疑问;面向涉事领域企业,推送行业安全整改规范,内容须明确整改核心目标、时限要求和监督验收机制,要求企业将规范落实到位。

(4) 衰退期: 长尾风险防控与中性情感转化。衰退期呈现中性情感主导、消极话题存续的特征,应注重

长尾风险的防控,推动中性情感向正面价值转化,实现舆情价值延伸。对政府而言,应联合教育和应急管理部门,将事件中产生的中性信息转化为安全教育素材,通过学校与社区等渠道长效传播,将舆情关注转化为公众安全意识提升,为同类事件的舆情管控提供参考。对社交平台而言,应持续监测消极情感博文,将可能引发次生舆情风险的内容同步至政府舆情部门,由政府发布统一澄清声明并全平台转发,防范舆情反弹。

4 结语

网络舆情复杂多变,具有传播速度快和情感倾向明显的特点,而现有舆情监测和管控手段对情感因素的考虑相对不足。本文主要对融合情感因素的网络舆情热度预测展开研究,提出一种基于情感分类和库普曼算子的网络舆情热度预测方法。首先,构建基于BERT-CGA多特征融合的情感分类模型,有效解决特征提取不充分及特征间关系学习不足的问题;其次,将情感特征和库普曼算子理论引入舆情预测,提出融合情感因素的LSTM-EDMD网络舆情热度预测模型,实现多步长热度预测,显著增强模型的可解释性;最后,基于银川烧烤店爆炸事件的实证研究结果,面向政府和社交平台提出分阶段监控策略,以提升舆情治理效能,助力建设健康有序的网络舆论空间。本文仍存在一定的局限性:仅选取突发公共安全事件案例进行实证分析,尚未覆盖其他类型舆情事件;模型构建重点考虑情感因素,对事件关系、主题强度、人工智能深度伪造和平台传播等影响因素覆盖不足。未来研究须扩展实证案例类型,验证模型在多场景下的通用性,并完善特

征要素体系, 量化多维特征对舆情演变的贡献, 实现更科学全面的舆情管控目标。

参考文献

- [1] 史伟, 薛广聪, 何绍义. 情感视角下的网络舆情研究综述[J]. 图书情报知识, 2022, 39(1): 105-118.
- [2] 王虎, 吴浩伟, 江长斌. 聚焦热度变化、主题动态与情感趋势的微博舆情演化研究[J]. 情报杂志, 2024, 43(11): 144-151, 128.
- [3] LIU Y, LI C Y, WANG J M, et al. Koopa: learning non-stationary time series dynamics with Koopman predictors[EB/OL]. [2025-07-12]. <https://arxiv.org/abs/2305.18803>.
- [4] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: ACL, 2014: 1746-1751.
- [5] DU J F, GRAVE E, GUNEL B, et al. Self-training improves pre-training for natural language understanding[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2021: 5408-5418.
- [6] 曹宇, 李天瑞, 贾真, 等. BGRU: 中文文本情感分析的新方法[J]. 计算机科学与探索, 2019, 13(6): 973-981.
- [7] BASIRI M E, NEMATI S, ABDAR M, et al. ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis[J]. Future Generation Computer Systems, 2021, 115: 279-294.
- [8] 曾子明, 陈思语. 基于LDA与BERT-BiLSTM-Attention模型的突发公共卫生事件网络舆情演化分析[J]. 情报理论与实践, 2023, 46(9): 158-166.
- [9] 韩坤, 潘宏鹏, 刘忠轶. 融合BERT多层次特征的短视频网络舆情情感分析研究[J]. 计算机科学与探索, 2024, 18(4): 1010-1020.
- [10] 王金政, 杨颖, 余本功. 基于多头协同注意力机制的客户投诉文本分类模型[J]. 数据分析与知识发现, 2023, 7(1): 128-137.
- [11] 林育曼, 文海宁, 饶浩. 基于ARIMA-BP神经网络模型的微信舆情热度预测[J]. 统计与决策, 2019, 35(14): 71-74.
- [12] 景楠, 胡怡, 韩喜双. 基于ARIMA与LSTM的新冠肺炎网络关注度趋势研究[J]. 中国安全科学学报, 2020, 30(12): 37-42.
- [13] 冯勇, 吕红旭, 徐红艳, 等. 基于SDZ-LSTM的舆情事件网络趋势预测模型[J]. 情报理论与实践, 2021, 44(6): 158-163.
- [14] 马永军, 郎威. 基于融合注意力机制LSTM的网络舆情趋势预测[J]. 计算机仿真, 2023, 40(1): 493-498.
- [15] 刘璐, 谢耘耕. 当前网络社会心态的新态势与引导研究[J]. 新闻界, 2018(10): 75-81, 100.
- [16] 牟冬梅, 靳春妍, 邵琦. 基于情感分析的突发公共卫生事件网络舆情热度预测模型仿真[J]. 现代情报, 2021, 41(10): 59-66.
- [17] 杨柳, 曾爽, 黄宇. 融入情感分析的重大突发事件舆情热度组合预测模型[J]. 情报探索, 2022(11): 17-25.
- [18] 赵高正, 魏娜, 李美彦, 等. 基于情感分析和热度预测的网络舆情预测研究[J]. 西安石油大学学报(自然科学版), 2024, 39(1): 135-142.
- [19] KOOPMAN B O. Hamiltonian systems and transformations in Hilbert space[J]. Proceedings of the National Academy of Sciences of the United States of America, 1931, 17(5): 315-318.
- [20] PING Z W, YIN Z, LI X T, et al. Deep Koopman model predictive control for enhancing transient stability in power grids[J]. International Journal of Robust and Nonlinear Control, 2021, 31(6): 1964-1978.
- [21] FOLKESTAD C, PASTOR D, MEZIC I, et al. Extended dynamic mode decomposition with learned Koopman eigenfunctions for prediction and control[C]//2020 American Control Conference (ACC). New York: IEEE Press, 2020: 3906-3913.
- [22] 岳崇浩. 基于Koopman算子理论的新新冠肺炎预测与分析[D]. 武汉: 华中科技大学, 2022.
- [23] 知微事见. 宁夏银川烧烤店爆炸事故致31人死亡[EB/OL]. [2025-07-21]. <https://ef.zhiweidata.com/event/e80d35c38ca4664510075060/profileV2>.
- [24] 邓春林, 陈荃柳, 李建奇. 基于灰色-加权马尔科夫模型的微博负面舆情热度预测研究[J]. 科技情报研究, 2023, 5(2): 78-89.
- [25] 刘润东, 张鹏, 秦瑞青, 等. 基于深度学习的网络舆情信息热度早期预测与实证研究[J]. 情报科学, 2024, 42(2): 148-154.
- [26] 高焯, 李松, 郭海玲, 等. 基于LTTB-KSC算法的网络舆情事件热度高峰研究[J]. 情报杂志, 2023, 42(1): 198-207.
- [27] 马腾, 殷跃, 赵树宽, 等. 多维数据融合的突发公共卫生事件网络舆情演化特征研究[J]. 情报理论与实践, 2022, 45(12): 170-177.
- [28] WILLIAMS M O, KEVREKIDIS I G, ROWLEY C W. A data-driven approximation of the Koopman operator: extending dynamic mode decomposition[J]. Journal of Nonlinear Science, 2015, 25(6): 1307-1346.

作者简介

王柯莹, 女, 硕士研究生, 研究方向: 数据挖掘与舆情分析。

高焯, 女, 博士研究生, 研究方向: 数据挖掘与舆情分析。

李松, 男, 博士, 教授, 通信作者, 研究方向: 系统工程理论与方法、数据挖掘与舆情分析, E-mail: lees3432@163.com。

Heat Prediction of Online Public Opinion Based on Sentiment Classification and Koopman Operator

WANG KeXuan GAO Ye LI Song

(School of Management, Hebei University, Baoding 071002, P. R. China)

Abstract: To fully explore the role of sentiment factors in public opinion monitoring and control, and accurately predict the trend of online public opinion heat, a prediction method for online public opinion heat based on sentiment classification and the Koopman operator is proposed. Firstly, a sentiment classification model based on BERT-CGA multi-feature fusion is constructed to fully extract features and learn the correlation between features, and classify the sentiment polarity of the Weibo content. Secondly, the sentiment characteristics and Koopman operator theory are introduced into the prediction of public opinion heat. A heat prediction model of LSTM-EDMD integrating sentiment factors is constructed. The LSTM algorithm is used as the observation function to realize the mapping of low-dimensional public opinion heat index data to the high-dimensional space. The Koopman operator is generated through the EDMD algorithm to describe the dynamic characteristics of the system for heat prediction, and to enhance the interpretability of the model. Finally, an empirical study is conducted using the explosion incident at a barbecue restaurant in Yinchuan as an example. The experimental results verify the crucial role of sentiment features in predicting the heat of online public opinion and demonstrate that the prediction method proposed in this paper has high prediction accuracy. This research can assist public opinion response entities such as governments and enterprises in promptly capturing sentiment trends, accurately predicting the evolution of public opinion heat, and formulating targeted strategies.

Keywords: Online Public Opinion; Heat Prediction; Koopman Operator; Sentiment Classification; Sentiment Factor

(责任编辑: 王玮)