文章编号:1008-5564(2023)03-0040-04

基于 UGC 的网络视频舆情传播特征框架研究

音　坤
(滁州学院 文学与传媒学院,安徽 滁州 239000)

摘　要:新媒体的发展开启了交互、多向的信息模式,极大地改变了人们的生活方式.网络舆论生态变得更为复杂,给网络舆论引导工作带来一系列现实的和潜在的问题.对此,提出了网络视频舆情传播特征(NVPOSF)框架,用于分析视频舆情特征.NVPOSF 融合了模态内、模态间和双模态间的交互,构建了基于多头注意力的融合网络.NVPOSF 通过为声学—视觉、声学—文本和视觉—文本特征分配合理注意力以获得重要特征.实验评估的结果表明,与现有的框架相比,NVPOSF 具有更好的性能.为政府主动引领网络舆论,形成正面舆论强势,对社会重大事件的舆论引领起到重要作用.

关键词:用户生成内容;网络视频舆情;双模态框架
中图分类号:TP391　　文献标志码:A

Research on the Framework of Network Video Public Opinion Spread Feature Based on User Generated Content(UGC)

YIN Kun
(School of Literature and Media, Chuzhou University, Chuzhou 239000, China)

Abstract: The development of new media has opened up interactive and multi-directional information models, greatly changing people's lifestyles. The ecosystem of online public opinion has become more complex, bringing a series of practical and potential problems to the guidance of online public opinion. In response, the Network Video Public Opinion Spread Feature (NVPOSF) framework was proposed to analyze video public opinion features. NVPOSF integrates intra modal, inter modal, and bimodal interactions to construct a fusion network based on multi head attention. NVPOSF allocates reasonable attention to acoustic visual, acoustic text, and visual text features to obtain important features. The experimental evaluation results indicate that NVPOSF has better performance compared to existing frameworks. This actively leads online public opinion by the government, forming a strong positive public opinion and playing an important role in leading public opinion on major social events.

Key words: user generated content; online video public opinion; bimodal framework

伴随着网络的普及和新媒体的迅速发展,网络舆情呈现出高发多发态势,现有网络空间中各种形式的话语表达与激情讨论释放出来的倒逼力量十分强大,网络舆情所显现的政府信任危机成为当代网络

收稿日期:2022-10-29
基金项目:安徽高校人文社科研究重大项目(SK2019ZD36)
作者简介:音　坤(1983—),男,安徽滁州人,滁州学院文学与传媒学院副教授,硕士,主要从事新闻与传播研究.

政治学必须关注的重要议题.舆情分析是正确理解人们意图的最关键技术之一.从认知的角度来看,人类从现实世界的经验中学习通常是多感官的,学习不仅通过语义句法,还通过视觉和听觉[1].因此,用户生成内容(User Generated Content,UGC)中分析判断舆情不能仅依靠文字.面向网络视频的舆情分析涉及多模态数据,探索连接和挖掘互补信息是巨大的挑战.信息融合有利于模仿人类处理和分析文本的方式,从而克服标准方法的局限性来实现计算和情感分析.不同的模态反映了不同强度的情绪,为了有效地挖掘不同模态表达的情感倾向,注意力机制已被广泛用于多模态融合.受研究目标及视域的限制,以往研究在描写政务新媒体的发展以及在发展所面临的问题与不足方面投入了较多的精力,但对其传播路径与效果的研究等方面有所忽略.本文探讨了双模态对视频舆情分析的贡献,考虑了模态内以及模态间和双模态间的相互作用,基于双模态信息的多头注意力架构,设计了网络视频舆情传播特征(NVPOSF)框架.

1　NVPOSF 框架设计

在 NVPOSF 中,文本(T)、音频(A)和视频(V)是输入.此外,主要的两部分是模态间交互和双模间交互.对于模态间交互,先前关于多模态特征融合的工作表明,外积可以有效地学习不同特征之间的交互.因此,使用外积来表示视觉文本(VT)、声学文本(AT)和声学视觉(AV)特征.对于双模间交互,设计了一种扩展的多头注意力机制来计算双模态注意力.合并模态间特征和双模间特征,进行最终的情感预测.

该框架的体系结构如图 1 所示.从图 1 中,很容易注意到该模型由四部分组成,即单模表示学习、模态间交互、双模间交互和预测网络.单模表示学习用于模拟特定于模态的交互.模态间交互是获取每两个模态之间的交互信息.双模间交互是学习每两个双模态之间的双模交互.由于模间交互的输出特征具有不同的维度,这些特征被输入到两个线性层以适应双模间交互模块.第一个全连接层用于转换为统一维度 d,第二个全连接层是通过共享机制提取深度特征和减少参数.最后将得到的交互信息和原始信息作为预测层的输入得到最终的情感标签.

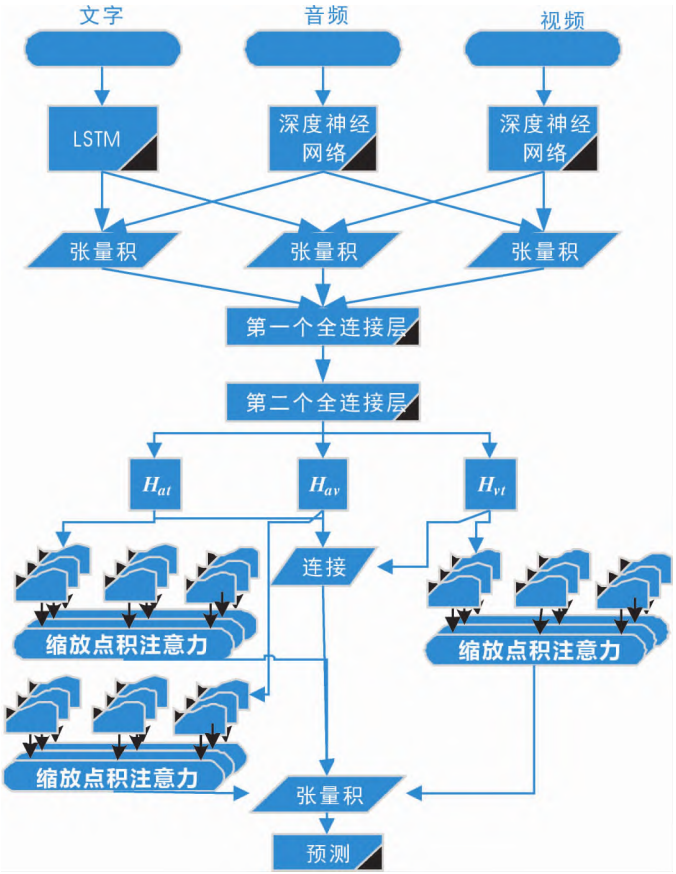

图 1　NVPOSF 框架的体系结构

1.1　单模表示学习模块

数据集中的视频被分成小的言语(Utt).每个言语包含三种单模态特征,即文本特征、声学特征和视觉特征.对于文本模态,使用预训练的中文 BERT[2] 来获得 d_t 维句子嵌入.在每一个句子中,词序的长度是不同的.对此,本文采用填充和截断使长度为 L.L 分两步计算:首先获得句子的平均长度并计算原始长度的标准偏差,然后将平均值乘以标准偏差的总和作为最终长度.引入填充以在末尾用特定字符填充短句.对于长句,取前 L 个向量构成句子嵌入.为了充分挖掘句子中单词之间的语义关系,使用 LSTM 网络为每个时间步生成融合特征.使用最终的隐藏状态输出作为具有 d_{t1} 维度的句子嵌入 t_i.对于声学和视觉模态,本文使用 LibROSA d_a 维声学特征,包括过零率、梅尔频率倒谱系数和恒定 Q 色谱图.使用 FFmpeg 以一定的速率对视频进行帧化,应用 MTCNN 来提取对齐的人脸,提取 d_v 维

人脸特征.用三层深度神经网络进一步提取具有 d_{a1} 维的声学特征 a_i 和具有 d_{v1} 维的视觉特征 v_i .

1.2 模态间交互模块

本文分别用 Z_t 、Z_a 和 Z_v 表示文本特征、声学特征和视觉特征的集合,用 N 表示数据集的样本数量.任意两种模态的张量融合是外积.AV 特征矩阵、AT 特征矩阵和 VT 特征矩阵是基于 Z_t 、Z_a 和 Z_v 学习得到,即:

$$\begin{aligned} Z_{av} &= Z_a \otimes Z_v \\ Z_{at} &= Z_a \otimes Z_t \\ Z_{vt} &= Z_v \otimes Z_t \end{aligned} \tag{1}$$

接下来,使用了两个包含 d 个单元的全连接层来对上述特征进行变换.第一个全连接层对 AV 特征矩阵、AT 特征矩阵和 VT 特征矩阵使用 ReLU 激活函,即:

$$\begin{aligned} \overline{Z}_{av} &= ReLU(W_{av} \times Z_{av} + b_{av}) \\ \overline{Z}_{at} &= ReLU(W_{at} \times Z_{at} + b_{at}) \\ \overline{Z}_{vt} &= ReLU(W_{vt} \times Z_{vt} + b_{vt}) \end{aligned} \tag{2}$$

为了进一步提取深度特征,$\overline{Z}_{av}$ 、$\overline{Z}_{at}$ 和 $\overline{Z}_{vt}$ 作为输入馈入到第二个全连接层,即:

$$H_s = FC(\overline{Z}_s, \theta) \tag{3}$$

其中,H_s 表示模态间交互特征,$s \in \{av, at, vt\}$.

1.3 双模态间交互模块

在多头注意力机制中,每个头的查询、关键词、值首先经过线性变换层处理,即:

$$\begin{aligned} \overline{X}^h &= X \times W_X^h \\ \overline{Q}^h &= Q \times W_Q^h \\ \overline{Y}^h &= Y \times W_Y^h \end{aligned} \tag{4}$$

其中,W_X^h 、W_Q^h 和 W_Y^h 为参数矩阵.缩放点积注意力的计算如下所示:

$$A^h = softmax\left(\frac{\overline{Q}^h \times (\overline{X}^h)^{\mathrm{T}}}{\sqrt{d}}\right) \times \overline{Y}^h \tag{5}$$

其中,d 是 X 的维度.将所有头的注意力分数连接起来作为线性变换的输入,以获得多头注意力的值,即:

$$\Omega(Q, X, Y) = [A^1; A^2; \ldots; A^n] \times W_O \tag{6}$$

其中,W_O 是参数矩阵.

为了进行双模态交互,计算双模态贡献并从不同的表示子空间捕获相关信息,将模态间特征连接起来,即:

$$D = Concat(H_{av}, H_{at}, H_{vt}) \tag{7}$$

其中,D 代表多模态特征.NVPOSF 旨在学习双模态注意力,鉴于成对特征相对于多模态特征具有不同的意义,NVPOSF 的任务是关注这些差异并找出它们的比例.在多头注意力的基础上,NVPOSF 包含三个输入形式略有不同的多头注意力.多模态特征被设置为源,而 AV、AT 和 VT 的特征分别被设置为目标. D 可被看作键和值,而一个双模态特征看作查询.

首先将多头线性投影应用于特征矩阵,并将它们映射到空间,即:

$$\begin{aligned} H_{D1}^i &= W_{D1}^i \times D \\ H_{D2}^i &= W_{D2}^i \times D \\ \overline{H}_s^i &= W_Q^i \times H_s \end{aligned} \tag{8}$$

其中 W_{D1}^i 、W_{D2}^i 和 W_Q^i 分别是 AV、AT 和 VT 的投影矩阵.对不同的双模态特征使用相同的参数矩阵 W_Q^i 以减少参数数量和内存消耗.在获得不同投影空间的特征后,利用注意力机制来探索成对模态之间

的互补关系.AV、AT、VT 注意力应用如下:

$$A_{av}^{i}=softmax\left(\frac{\overline{H}_{av}^{i}\times\left(H_{D1}^{i}\right)^{\mathrm{T}}}{\sqrt{d_m}}\right)\times H_{D2}^{i}$$
$$A_{at}^{i}=softmax\left(\frac{\overline{H}_{at}^{i}\times\left(H_{D1}^{i}\right)^{\mathrm{T}}}{\sqrt{d_m}}\right)\times H_{D2}^{i} \tag{9}$$
$$A_{vt}^{i}=softmax\left(\frac{\overline{H}_{vt}^{i}\times\left(H_{D1}^{i}\right)^{\mathrm{T}}}{\sqrt{d_m}}\right)\times H_{D2}^{i}$$

为了获得分配了注意力的双模态特征表示,将每个头部的 AV、AT 和 VT 注意力分别连接起来并进行线性层变换,即:

$$\Theta\left(H_{av},D,D\right)=\left[A_{av}^{1};\ldots;A_{av}^{h}\right]\times W_O$$
$$\Theta\left(H_{at},D,D\right)=\left[A_{at}^{1};\ldots;A_{at}^{h}\right]\times W_O \tag{10}$$
$$\Theta\left(H_{vt},D,D\right)=\left[A_{vt}^{1};\ldots;A_{vt}^{h}\right]\times W_O$$

其中 W_O 为权重参数.

1.4　预测网络

将获得的具有 A_{av}、A_{at} 和 A_{vt} 的双模态间交互特征连接起来并添加为原始特征 D 的残差函数,以避免梯度消失问题.最后,将它们输入到三层 DNN 中以生成输出.

2　实验评估

实验使用中文公共多模态情感分析数据集 CH-SIMS[3].实验所使用的基线框架的介绍如下.EF-LSTM 连接三种模态的原始特征并将它们输入 LSTM 以捕获模态序列之间的长期赖性[4].LF-DNN 使用 DNN 学习单模态特征,然后将它们连接起来作为预测层的输入[3].MFN 通过门控存储单元存储模态的内部信息和模态之间的交互信息,并添加动态融合图以反映有效的情感信息[5].Mult 使用其跨模态注意力模块提取每个模态内的关键信息,然后基于 Transformer 模型合并这些特征[6].MISA 结合了损失的组合,包括分布相似性、正交损失、重建损失和任务预测损失,以学习模态不变和模态特定表示[7].Self-MM 设计了一种基于自监督方法的单峰标签生成策略,然后引入单峰子任务以帮助学习模态特定表示.

模型在数据集上运行模型五次,统计测试集上的平均性能.在训练过程中,为不同数据集调整学习率、批量大小、dropout 和每个模态特定子网络的隐藏单元数等超参数.实验中都使用具有初始学习率的 Adam 优化器,并在 20 个 epoch 之前执行提前停止.

表 1 显示了 CH-SIMS 数据集的比较结果.由结果可知,提出的 BIMHA 在大多数指标上都优于其他框架.EF-LSTM 没有完全学习模态之间的交互信息,因此其综合性能是所有框架中最差的.基于后期融合的 LF-DNN 考虑了模态内交互并提取更多相关信息,因此其性能比 EF-LSTM 的稍有提高.虽然 Mult 使用了 Transformer 架构,但仍然无法达到令人满意的性能.Self-MM 是基于模型不断更新生成的单模态标签,而这些生成的标签可能会导致预测结果不准确.本文提出的框架考虑了成对模式之间交互的差异并引入了双模态注意力,因此性能在很大程度上优于其他模型.表 1 提供了每个框架的参数量来量化模型的复杂性.综上所述,结果表明结合模态内和模态间信息可以产生更好的性能,如果充分挖掘具有高贡献和互补信息的特征,双模态间交互可能有利于信息融合.

表 1　各个框架对比

框架	准确率/%	F1 度量	平均绝对误差	参数量/万
EF-LSTM	69.38	56.83	0.59	21.5
LF-DNN	77.12	77.38	0.45	63.5
MFN	77.98	77.99	0.44	6
Mult	78.65	79.64	0.45	180
MISA	69.56	57.12	0.59	12 300
SelfMM	80.15	80.32	0.43	10 200
NVPOSF	83.82	83.73	0.39	250

(下转第 87 页)

[26] 王刘煜，李冬，曾辉平，等.低温高铁锰氨氮地下水两级生物净化快速启动[J].中国环境科学，2019，39(6)：2361-2369.
[27] 张杰，杨宏，李冬，等.生物滤层中 Fe~(2+) 的作用及对除锰的影响[J].中国给水排水，2001(9)：14-16.
[28] 杨宏，熊晓丽，段晓东，等.贫营养条件下生物除铁除锰滤池生态稳定性研究[J].环境科学，2010，31(1)：99-103.
[29] CASALINI L C，PIAZZA A，MASOTTI F，et al.Manganese removal efficiencies and bacterial community profiles in non-bioaugmented and in bioaugmented sand filters exposed to different temperatures[J].Journal of Water Process Engineering，2020，36：101261.
[30] 武俊槟，黄廷林，程亚.同步去除水中铁、锰、氨氮滤池的快速启动与运行控制[J].中国给水排水，2016，32(15)：20-25.
[31] YANNA，LEGOUELLEC，MENACHEM ELIMELECH.Calcium sulfate(gypsum) scaling innano filtration of agricultural drainage water[J].JMembrSci，2002，205(1-2)：279-291.
[32] KIM HA，CHOI JH，TakizawaS.Comparison of initialnano filtration for drinking water treatment[J].SepPurif Technol，2007，56(3)：354-362.
[33] 徐满天，唐玉朝，胡伟，等.$KMnO_4$预氧化与混凝联合作用去除湖泊源水中 Mn~(2+) 的研究[J].水处理技术，2018，44(6)：46-51.
[34] ZOGO D，BAWA L M，SOCLO H H，et al.Influence of pre-oxidation with potassium permanganate on the efficiency of iron and manganese removal from surface water by coagulation-flocculation using aluminium sulphate：case of the Okpara dam in the Republic of Benin[J].Journal of Environmental Chemistry and Ecotoxicology，2011，3(1)：1-8.
[35] 雷晓玲，秦颖，文永林，等.预氧化强化混凝工艺处理含锰水实验研究[J].应用化工，2022，51(1)：110-113.
[36] 王文东，岳强，刘国旗，等.传统饮用水净化工艺对锰的去除特性[J].环境工程学报，2016，10(9)：4733-4736.

责任编辑　马云彤

(上接第 43 页)

3　结语

本文提出了网络视频舆情传播特征框架——一种基于双模态信息对舆情传播特征的分析方法.使用公共数据集对提出的框架进行评估.实验结果表明，所提出的框架是有效的，其性能优于现有的框架.未来的工作将使用更多相关的融合方法进行实验，为多模态舆情分析提供更多的基准结果.

参　考　文　献

[1] RAHATE A，WALAMBE R，RAMANNA S，et al.Multimodal co-learning：challenges，applications with datasets，recent advances and future directions[J].Information Fusion，2022，81：203-239.
[2] KENTON J D M W C，TOUTANOVA L K.BERT：Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of NAACL-HLT.2019：4171-4186.
[3] YU W，XU H，MENG F，et al.Ch-sims：A chinese multimodal sentiment analysis dataset with fine-grained annotation of modality[C]//Proceedings of the 58th annual meeting of the association for computational linguistics.2020：3718-3727.
[4] DING N，TIAN S，YU L.A multimodal fusion method for sarcasm detection based on late fusion[J].Multimedia Tools and Applications，2022，81(6)：8597-8616.
[5] BOEHM K M，KHOSRAVI P，VANGURI R，et al.Harnessing multimodal data integration to advance precision oncology[J].Nature Reviews Cancer，2022，22(2)：114-126.
[6] CHAN J Y L，BEA K T，LEOW S M H，et al.State of the art：a review of sentiment analysis based on sequential transfer learning[J].Artificial Intelligence Review，2023，56(1)：749-780.
[7] HAZARIKA D，ZIMMERMANN R，PORIA S.Misa：Modality-invariant and-specific representations for multimodal sentiment analysis[C]//Proceedings of the 28th ACM International Conference on Multimedia，2020：1122-1131.

责任编辑　王新奇