

融合语言特征的多模态中文反讽识别模型

胡文彬, 陈龙, 韩天乐, 仲兆满, 马常霞

(江苏海洋大学 计算机工程学院 江苏 连云港 222005)

摘要: 针对图文模态间不一致性问题和未充分关注文本内含信息的问题,提出了融合语言特征的多模态中文反讽识别模型。利用卡方统计方法提取具有反讽和非反讽含义的词语,构建语言特征体系。使用 TextCNN 提取语言特征,加强反讽特征与非反讽特征的区别。采用 TextCNN 和 ResNet 提取文本和图像特征,并引入交叉注意力机制,以残差连接融合图文特征,更好地保留了语言特征。使用了包含网民反讽评论的突发事件多模态数据集验证了所提模型的有效性,结果表明,该模型优于基线模型,关注文本语言特征有助于提高解决问题的效率。

关键词: 语言特征; 中文反讽识别; 突发事件; 多模态

中图分类号: TP391

文献标志码: A

文章编号: 1671-6841(2025)05-0016-08

DOI: 10.13705/j.issn.1671-6841.2024096

A Multimodal Chinese Sarcasm Detection Model Integrated with Linguistic Features

HU Wenbin, CHEN Long, HAN Tianle, ZHONG Zhaoman, MA Changxia

(School of Computer Engineering, Jiangsu Ocean University, Lianyungang 222005, China)

Abstract: In light of the disparity between graphic and text modes and insufficient attention to textual information, a multimodal Chinese sarcasm detection model integrated with linguistic features was proposed. The Chi-square statistical method was used to extract words with sarcastic and non-sarcastic meanings, forming the linguistic feature system. TextCNN was utilized to extract linguistic features, enhancing the distinction between sarcastic and non-sarcastic characteristics. TextCNN and ResNet were employed to extract text and image features, and a cross-attention mechanism was introduced. Residual connections were used to fuse text and image features, to help preserve language characteristics. The effectiveness of the proposed model was verified by using an emergency multimodal dataset containing sarcastic comments. The results showed that the model outperformed the baseline model, and focusing on textual linguistic features helped improve the efficiency of problem-solving.

Key words: linguistic feature; Chinese sarcasm detection; emergency event; multimodality

0 引言

当前,社交媒体中存在着大量的网民参与话题讨论的信息,通过情感分析可以了解网民的真实意图。然而,网民在参与一些突发事件话题讨论时,更倾向使用反讽来表达观点和想法^[1]。这不仅影响传统情感分析的识别精度,而且造成管理者误判舆

情趋势,甚至导致网络舆情发展不可控。因此,通过反讽识别研究有助于管理者了解公众对突发事件的真实想法^[2],及时控制舆情。

微博是国内最大的社交媒体平台之一,随着用户发布图文多模态消息占比的迅猛增长,开展图文信息多模态研究具有重大意义。然而,经过知网检索和谷歌学术搜索后发现,中文反讽识别研究文章数量占比很低,且这些研究很少涉及除

文本之外的其他模态的利用。《2020 微博用户发展报告》显示,微博评论充满随意和个性化的特点,主要表现在以下几个方面:使用大量网络语言,不注重语句逻辑,表达较为隐晦难懂以及内容简洁。因此,应充分利用这些特征开展研究,以提升反讽识别的准确性。

现有的多模态模型的整体设计大多集中在深度学习模型的训练上,以模态间信息融合与交互、捕捉“模态间矛盾”为重点来设计相关模型^[3-4],忽略了反讽语句本身显式的语言特征。针对以上问题,本文使用深度学习的方法挖掘语句深层语义特征,结合微博评论中的显式语言特征,提出了一种融合语言特征的多模态中文反讽识别模型。首次尝试在多模态反讽识别研究中引入语言特征模块,所提模型在准确率上表现突出,同时验证了语言特征的实质性作用。

1 相关工作

1.1 纯图文模态反讽识别

早期的反讽识别研究主要集中在纯文本领域,随着多模态消息的增长和深度学习技术的进步,近年来多模态反讽识别研究发展迅速。Schifanella等^[5]考虑了讽刺检测的文本和图像特征,并提出了两种方法来实现反讽识别任务,是模态间的简单融合。张继东等^[3]提出一种多模态深度学习模型,在模态内和模态间的特征融合层以简单加权的方式实现融合。但是,这些模型未充分考虑模态间的信息交互。

为了解决模态之间存在的非一致性问题,Wang等^[6]提出一个以BERT为基础的模型,设计了2D-Intra-Attention来探索文字和图像之间的差异。Pan等^[7]提出一个基于BERT架构的模型,该模型通过使用注意力机制来获取反讽识别模态内与模态间的非一致性。Gupta等^[8]提出了基于RoBERTa的协同关注模型,使用RoBERTa对文本编码,利用FiLMed ResNet模块对输入图像提取特征,通过构建联合注意力解决输入文本和图像的非一致性问题。Liang等^[9]对图像进行切片处理后,使用图卷积神经网络来建模图文模态间的非协调关系。Wen等^[10]提出了由两个分支组成的双重非一致感知网络,从事实和情感层面挖掘讽刺信息。Tian等^[11]提出了动态路由Transformer网络,利用动态路径来适应模态间非一致性分层。

1.2 辅助模态设计

一些研究者采用基于规则或统计的方法,将设计的文本或图片作为一种新模态添加到模型中,完成最终的模型设计。Cai等^[12]使用属性预测器对每个图像进行属性预测,将得到的图像属性作为连接文本和图像之间的额外模态,通过融合这些模态特征进行预测。Sangwan等^[13]使用OCR技术提取嵌入在图像中的文本,将其作为一种模态构建分类网络,以提高预测效果。Xu等^[14]通过从图像中提取不同词性的语义信息,设计了一个网络来表征文本和图像之间的多视图语义关联。Liang等^[15]采用目标检测的方法,在图像中识别与文本描述相符的物体,设计了一个基于图卷积神经网络的框架,捕捉模态间的非一致性。

2 微博评论文本语言特征体系构建

2.1 语言特征体系构建流程

反讽语句本身存在的一些特征,使得深度学习模型在学习反讽语句时面临较大的挑战。本文在文献[16]的基础上,结合微博评论的特点,通过文本评论的卡方检验,得到词语和词汇搭配。在一系列筛选和整理后,构建出反讽语言特征体系和非反讽语言特征体系。语言特征体系构建流程见图1。

2.2 反讽语言特征体系构建

2.2.1 反讽语言特征 本文旨在深入探讨微博评论中的反讽语言特征,并构建出相应的反讽语言特征体系,具体的反讽语言特征如下。

1) 固定词汇搭配。固定词汇的搭配在反讽语句中经常出现,因此这一特征被考虑最多^[16-17]。例如:“今天气温 35℃,还可以再热一点吗?”其中,“还可以再……一点……”是反讽语句中的固定搭配形式,表达了对燥热天气的不满。

2) 词义转化词。在反讽语句中,会使用某些褒义词来起到贬义的效果。在微博反讽评论中,这一现象尤为明显。因此,将这种现象作为微博评论反讽语言的一个重要特征加以研究。例如:“这次活动真是太有创意了,每天新花样令人惊叹,感谢主办方带来的精彩,首次体验了完美混乱,值得铭记。”其中,“惊叹”“精彩”“铭记”等褒义词与该语境产生强烈冲突,表达了对事件处理的不满。

3) 谐音词和网络词。谐音词和网络词皆是通过在语义上与原文产生反差来表达反讽效果。例如:“我劝你耗子尾汁!”其中“耗子尾汁”是“好自为之”的谐音词。

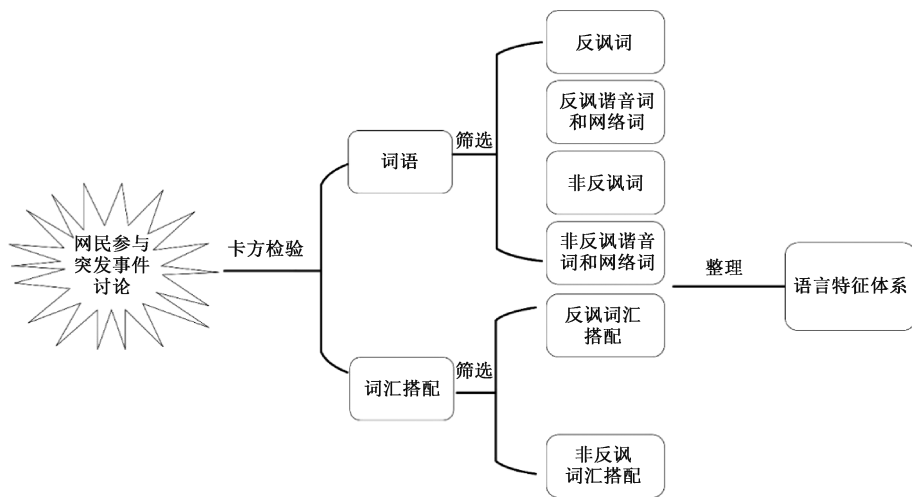


图1 语言特征体系构建流程

Figure 1 Process of constructing linguistic feature system

4) 特定的语气词。在反讽语句中,会将语气词用于表达戏谑、讽刺等意味。例如:“哈哈,对不起,真的要笑死了。”其中,“哈哈”在这句话中充满了反讽意味。

5) 特定的标点符号。标点符号可以通过特定的使用方式来表达反讽意味,不仅增加了语言的深度和趣味性,也促进了读者对信息更加细致的解析。例如:“在这种情况下竟然还能搞砸事情,真是太不可思议了!!!”其中,通过叹号表达了讽刺意味。

6) 特定的表情符号。某些表情符号在使用时含义已经开始变味,用于表达讽刺等意味。例如:“这次公司年会的组织真是典范标杆☺☺。”其中,通过使用“☺”这一表情符号,传达了一种反讽的语气。

7) 特定的数字。使用一些数字可以传达情感信息,用于表达嘲笑等意味。例如:“又被裁员了,看来在管理这一块儿玩得可真6啊!”结合语境,表达了作者的不满。

2.2.2 反讽语言特征体系选取方法与结果 使用卡方检验(χ^2)统计量的大小作为衡量指标,评估反讽语言特征在评论数据中的显著性和重要性。

1) 词汇搭配选取方法与结果

在词汇搭配研究中,采用一种自动的方法,使用nltk包对文本进行分词和预处理,并对词汇搭配进行统计和计算卡方值,最终选取了前1000条卡方统计值最大的词汇搭配。采用人工收集的方法,对前人工作中的词汇搭配结果和本文计算出的卡方值结果进行对比和整理,最终筛选出适用于本文数据集的22条词汇搭配规则。反讽词汇搭配规则的部分卡方统计值结果见表1。

表1 反讽词汇搭配规则的部分卡方统计值

Table 1 Partial Chi-square statistics results for sarcastic vocabulary collocation rules

词汇搭配规则	卡方值
前有……后有	1 530.000 6
听我说……谢谢你	1 528.002 6
不愧……是	405.152 6
好……呢	12.544 0
又……了	2.791 0

2) 词汇和符号特征选取方法与结果

通过对数据集进行分词和词汇的卡方值计算,输出了前6000条卡方值高的词汇,进一步研究、分析和筛选后,得到谐音词和网络词、词义转化词和语气词共136条,表情符号、标点符号和数字共57条。为了便于处理,将表情符号转换为相应的文本。反讽词汇的部分卡方统计值结果见表2。

表2 反讽词汇的部分卡方统计值

Table 2 Partial Chi-square statistics results for sarcastic vocabulary

特征类别	词汇和符号特征	卡方值
词义转化词	谢谢	55.952 2
	精准	41.973 1
谐音词和网络词	真实	14.996 5
	绝	9.998 0
语气词	哈哈	14.079 2
	呵呵	14.058 6
表情符号	😏	10.310 6
	☺	8.466 5
标点符号	?	526.927 8
	!	442.283 8
数字	6	30.985 3
	2 020	1.999 9

2.3 非反讽语言特征体系构建

2.3.1 非反讽语言特征 为了填补因缺乏语言特征而出现的特征空值,使语言特征更加对称,从而更全面地捕捉文本信息,构建了非反讽语言特征体系,其语言特征如下。

1) 明显的情绪表达词。明显的情绪表达词直接传达了作者的某种情感,没有任何隐藏的含义。

2) 非反讽性谐音词和网络词。这些词大多为与原词语发音相似的词语变形或网络中创造的新词汇和缩写,用于增添趣味性、幽默感或表达特定情绪。

2.3.2 非反讽语言特征体系选取方法与结果 通过构建非反讽语言特征体系,进一步提升模型对非反讽语句的分类准确性,并最终提升模型的整体准确性和鲁棒性。对数据集中前 6 000 条卡方值较高的词汇进行筛选和分析,得到明显的情绪表达词 667 条(积极词条 201 条,消极词条 176 条,中性词条 290 条),非反讽性谐音词和网络词 20 条。非反讽词汇的部分卡方统计值结果见表 3。

表 3 非反讽词汇的部分卡方统计值

Table 3 Partial Chi-square statistics results for non-sarcastic vocabulary

特征类别	语言特征	卡方值
情绪表达词	希望	173.875 5
	感谢	133.926 1
	垃圾	37.994 0
非反讽性谐音词和网络词	离谱	25.997 2
	躺平	5.999 8
	捡漏	3.999 9
	凡尔赛	3.999 9

1) 文本和语言特征模态表示。使用 TextCNN^[18] 获取文本和语言特征表示。

2) 图像模态表示。使用 ResNet^[19] 提取图像特征。

3) 模态融合。通过使用交叉注意力机制将图像和文本的特征表示融合在一起,捕捉反讽特征,利用语言特征增强对反讽句和非反讽句的区分能力。最后,将得到的特征输入至全连接层(FC)中进行二分类判别。

为了研究方便,把表情特征转化为文本形式,将其视为文本特征。融合语言特征的多模态中文反讽识别模型如图 2 所示。

3 多模态中文反讽识别模型构建

3.1 模型组成

所构建的模型主要包含以下三个部分。

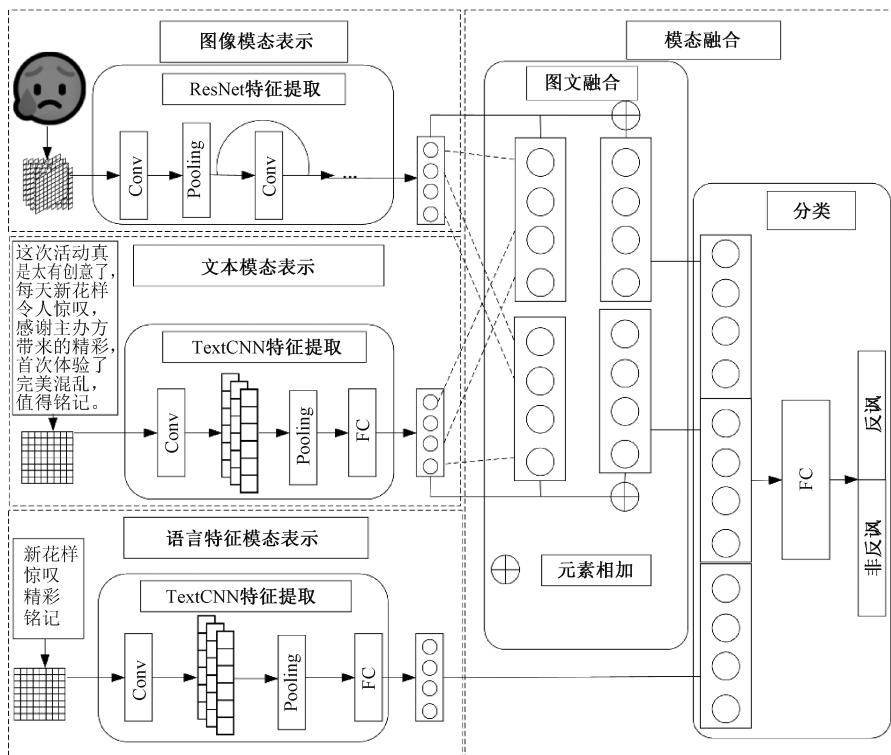


图 2 融合语言特征的多模态中文反讽识别模型

Figure 2 A multimodal Chinese sarcasm detection model integrated with linguistic features

3.2 文本和语言特征模态表示

使用 Word2Vec 模型从海量文本中学习语义信息,得到构建数据集中的每条文本向量化表示,用 $\mathbf{x}_{kn} = (x_{kn}^1, x_{kn}^2, \dots, x_{kn}^d)$ 表示长度为 n 的第 k 条文本的 d 维词向量表示。后续 TextCNN 模型的输入是一个 $n \times d$ 矩阵 \mathbf{X} , 其中矩阵的行表示一段文本中一个词的 d 维词向量表示。对矩阵 \mathbf{X} 进行卷积操作,在 \mathbf{X} 上进行滑动窗口计算,得到特征图 \mathbf{C} 。一次卷积计算表示为

$$\mathbf{C}_i = f(\mathbf{W} \times \mathbf{X}_{i:i+h-1} + \mathbf{b}), \quad (1)$$

其中: $\mathbf{X}_{i:i+h-1}$ 表示矩阵 \mathbf{X} 的第 i 行到第 $i+h-1$ 列; \mathbf{W} 是卷积核,其大小为 $h \times d$; \mathbf{b} 为偏置向量; f 为激活函数。对得到的每个特征图 \mathbf{C} , 通过最大池化操作得到池化向量 \mathbf{t} , 即

$$\mathbf{t} = \max(\mathbf{C}). \quad (2)$$

将所有的池化向量 \mathbf{t} 连接形成一个长向量 \mathbf{T} , 通过一个全连接层进行分类。以同样的方式获得语言特征 \mathbf{L} , 以便后续使用。

3.3 图像模态表示

使用 ResNet34 获取图像特征,将输入图像的像素大小设置为 $224 \times 224 \times 3$, 经过一系列卷积层、池化层和残差连接操作后,得到了输出像素大小为 $7 \times 7 \times 512$ 的特征图。为了后续任务使用,将模型的最后一个全连接层的输出维度调整为所需的维度,最终得到图像特征 \mathbf{I} 。

3.4 模态融合

将文本特征 \mathbf{T} 和图像特征 \mathbf{I} 通过交叉注意力机制学习到模态间的交互信息,以残差连接方式进行模态融合,建立它们之间的联系,最终设计出一个特征融合模块。将文本特征 \mathbf{T} 作为查询 \mathbf{Q} , 图像特征 \mathbf{I} 作为键 \mathbf{K} 和值 \mathbf{V} , 通过应用注意力机制得到文本层面的注意力特征 \mathbf{A}_T , 计算公式为

$$\mathbf{A}_T = \text{Softmax}\left(\frac{[\mathbf{W}^Q \mathbf{T}]^T [\mathbf{W}^K \mathbf{I}]}{\sqrt{d_k}}\right) [\mathbf{W}^V \mathbf{I}]^T, \quad (3)$$

其中: $\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V$ 是模型训练过程中可学习的参数矩阵。引入 d_k 的目的是平衡不同维度对相似度计算的影响。将多个头的输出连接起来,得到总的文本层面注意力特征。通过使用残差连接的方式,将原始的文本特征 \mathbf{T} 和经过交叉注意力机制得到的文本层面的注意力特征 \mathbf{A}_T 连接起来,这样既保留了原始文本特征的信息,又得到了图文交互信息。

至此,得到了文本层总输出 \mathbf{F} , 即

$$\mathbf{F} = \mathbf{A}_T + \mathbf{T}. \quad (4)$$

同样地,将图像特征作为查询 \mathbf{Q} , 文本特征作为键 \mathbf{K} 和值 \mathbf{V} , 其计算方法与计算文本层面的注意

力分数相同。通过图像引导文本信息的方式,计算图像层面的注意力分数,得到图像层面的注意力特征 \mathbf{A}_I 。按照同样的残差连接方式,将原始的图像特征 \mathbf{I} 和经过交叉注意力机制得到的图像层面的注意力特征 \mathbf{A}_I 连接起来,得到图像层总输出 \mathbf{M} 。

将文本层注意力特征 \mathbf{F} 和图像层注意力特征 \mathbf{M} 连接起来,形成图文模态交叉注意力融合模块的图文输出 \mathbf{R} , 即

$$\mathbf{R} = \text{cat}(\mathbf{F}, \mathbf{M}). \quad (5)$$

将获取到的语言特征 \mathbf{L} 和图文输出 \mathbf{R} 连接起来,得到模态融合的最终特征表示 \mathbf{S} , 即

$$\mathbf{S} = \mathbf{R} + \mathbf{L}. \quad (6)$$

将其输入至全连接层后,使用 Softmax 函数得到分类结果。

4 实验与结果分析

4.1 数据集介绍

目前,突发事件领域内的反讽识别研究尚未受到充分关注,特别是在中文方面的相关研究工作较少,公开的中文数据集更少。本文使用笔者在文献[20]中构建的数据集,该数据集包含了多模态反讽评论和非反讽评论各 1 179 条数据。

4.2 实验设置

模型基于 PyTorch 框架设计实现,文本数据的词向量维度设定为 300 维,文本数据特征提取模型 TextCNN 中使用了 2、3、4 和 5 四种不同的卷积核尺寸,每种尺寸的数量为 400。使用 Adam 优化器,将模型的学习率设置为 5×10^{-4} 。为进行高效训练,将批大小设置为 32,丢弃率为 0.2,并进行了 100 个轮次的训练。此外,还设置了一个早停机制,如果 50 轮后网络的准确率没有提升,则训练过程会提前终止。

由于绝大部分文本的长度都在较短范围内,且限制文本长度可以降低模型的计算复杂度,提高训练和推理的效率。因此,将文本和语言特征的最大长度皆限定为 64,这一设置兼顾了效率和实用性。

4.3 实验结果对比

将基线模型分为文本模态模型、图像模态模型和图文多模态模型三类,与本文模型进行对比实验,评价指标包括准确率、精确率、召回率和 F1 值,具体对比实验结果见表 4,从表 4 可以得出以下结论。

1) 引入交叉注意力机制能设计出更好的图文模态融合方式,增强了不同模态之间的关联性。同

表4 不同模型对比实验结果

Table 4 The experiment results of different models

单位:%

模型	方法	准确率	精确率	召回率	F1
文本模态	TextCNN ^[21]	86.76	87.06	86.04	86.39
	BiLSTM ^[22]	85.84	87.00	84.65	85.25
	BERT ^[23]	87.61	87.68	86.65	87.06
图像模态	ResNet ^[24]	59.82	58.84	57.14	56.19
图文多模态	TextCNN+ResNet ^[25]	89.04	92.35	83.25	87.56
	BiLSTM+CNN ^[3]	80.82	82.21	79.27	79.81
	BERT+ResNet ^[7]	82.65	81.59	80.79	81.19
	LFMIBi ^[23]	89.73	91.58	85.71	88.55
	LFMIT+Att	90.18	93.96	84.24	88.83
	本文	91.32	94.12	86.70	90.26

注:LFMIT+Att为本文模型的变体形式,仅在模型融合之前加入注意力机制。

时,通过文本数据手工设计、筛选和总结相关语言特征,使得模型能够学习更多的与反讽和非反讽相关的特征。

2) 仅使用图像特征的模型在反讽识别任务中表现较差,这表明图像特征在该任务中的有效性较低。相比之下,文本模型的表现更好。

3) 尽管预训练后的BERT模型在文本特征上取得了最优的性能,但与其他模型结合后,其准确率会下降。这可能是由于数据集的数量较少,特征表示不充分。

4) 综合来看,图文多模态模型的表现整体上优于单模态模型,进一步证实了融合多种模态信息对于提高模型性能的有效性。

5) 注意力机制并非在所有情况下都可以起到有效的作用。

4.4 消融实验

通过消融实验来探究交叉注意力机制模块和语言特征模块对整体模型的有效性,具体包括:① 去除图文交叉注意力机制融合模块,仅简单连接图文特征,并保持其他部分不变。② 去除语言特征模块,将经过交叉注意力机制模块后的图文特征输入全连接层进行分类。消融实验结果见表5。可以看

出,在模型设计中,具有图文交叉注意力机制同时融合语言特征模块,模型准确率为91.32%,F1值为90.26%,达到了最佳。通过分析表5可以得出以下结论。

1) 缺少图文交叉注意力机制融合模块会导致模型准确率下降,说明该模块促进了图文之间的信息交互和融合,且召回率达到了最佳。

2) 缺少语言特征模块时,从模型实验结果来看,四个指标较低,进一步证明语言特征的加入会增强模型对文本信息的理解。将语言特征看作一种模态,增强了模态融合效果。

表5 消融实验结果

Table 5 Ablation experiment results

单位:%

模型	准确率	精确率	召回率	F1
去交叉注意力	89.50	87.56	90.15	88.83
去语言特征	88.36	88.78	85.71	87.22
本文	91.32	94.12	86.70	90.26

4.5 参数分析

对模型进行了不同维度语言特征的评估实验,探索了最优的语言特征表示维度。不同维度语言特征的评价指标结果如图3所示。

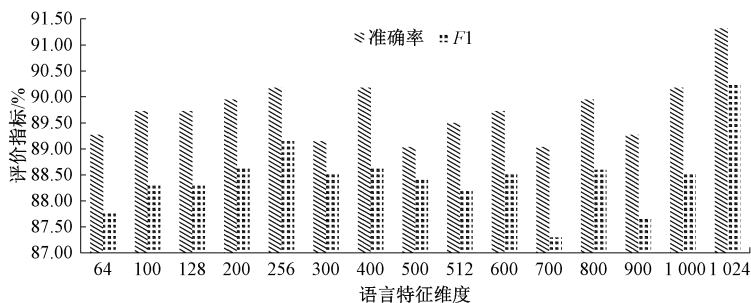


图3 不同维度语言特征的评价指标结果

Figure 3 Evaluation index results of linguistic features in different dimensions

通过准确率和 $F1$ 值来衡量模型性能,逐步增加语言特征维度,从 64 维升至 1 024 维,共进行了 15 组不同维度的实验。可以看出,当语言特征维度设置为 1 024 维时,模型的准确率和 $F1$ 值均达到最大值。这表明较高的语言特征维度对模型性能有积极影响。同时,较高的语言特征维度能够更好地表达反讽与非反讽信息。

5 结语

本文提出了一种融合语言特征的多模态中文反讽识别模型,通过图文交叉注意力机制的设计,使得图片和文本之间能够实现信息交互和有效融合。对文本手工提取了丰富的语言特征,并将其作为另一种模态引入模型中。通过与一系列基线模型的对比实验,验证了本文模型的优越性,并证明语言特征的加入对模型性能产生了积极影响。未来的研究工作主要包括:1) 积极探索注意力机制加入的位置和设计方式,以进一步优化模型性能;2) 继续扩展数据集的规模,获取更广泛的样本,以进一步提升模型的性能和泛化能力;3) 探索广义的上下文信息,以获得更丰富和有效的特征表示。

参考文献:

- [1] ZHANG M, ZHANG Y, FU G. Tweet sarcasm detection using deep neural network[C]//Proceedings of the 26th International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2016:2449-2460.
- [2] BILAL G, JIHEN K, FARAH B, et al. Overview of the track on irony detection in Arabic tweets[C]//Proceedings of the 11th Annual Meeting of the Forum for Information Retrieval Evaluation. New York: ACM Press, 2019: 10-13.
- [3] 张继东, 蒋丽萍. 基于多模态深度学习的旅游评论反讽识别研究[J]. 情报理论与实践, 2022, 45(7): 158-164.
ZHANG J D, JIANG L P. Research on irony recognition of travel reviews based on multi-modal deep learning[J]. Information studies: theory & application, 2022, 45(7): 158-164.
- [4] 张亚洲, 俞洋, 朱少林, 等. 一种量子概率启发的对话讽刺识别网络模型[J]. 电子学报, 2022, 50(8): 1885-1893.
ZHANG Y Z, YU Y, ZHU S L, et al. A quantum probability inspired network for dialogue sarcasm recognition [J]. Acta electronica sinica, 2022, 50(8): 1885 - 1893.
- [5] SCHIFANELLA R, JUAN P D, TETREAULT J, et al. Detecting sarcasm in multimodal social platforms [EB/OL]. (2016-08-08) [2024-03-15]. <https://doi.org/10.48550/arXiv.1608.02289>.
- [6] WANG X Y, SUN X W, YANG T, et al. Building a bridge: a method for image-text sarcasm detection without pretraining on image-text data[C]//Proceedings of the 1st International Workshop on Natural Language Processing Beyond Text. Stroudsburg: Association for Computational Linguistics, 2020: 19-29.
- [7] PAN H L, LIN Z, FU P, et al. Modeling intra and inter-modality incongruity for multi-modal sarcasm detection [C]//Findings of the Association for Computational Linguistics: EMNLP 2020. Stroudsburg: Association for Computational Linguistics, 2020: 1383-1392.
- [8] GUPTA S, SHAH A, SHAH M, et al. FiLMing multimodal sarcasm detection with attention[C]//Proceedings of the International Conference on Neural Information Processing. Cham: Springer International Publishing, 2021: 178-186.
- [9] LIANG B, LOU C W, LI X, et al. Multi-modal sarcasm detection with interactive in-modal and cross-modal graphs[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM Press, 2021: 4707-4715.
- [10] WEN C S, JIA G L, YANG J F. DIP: dual incongruity perceiving network for sarcasm detection[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 2540-2550.
- [11] TIAN Y, XU N, ZHANG R K, et al. Dynamic routing transformer network for multimodal sarcasm detection[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2023: 2468-2480.
- [12] CAI Y T, CAI H Y, WAN X J. Multi-modal sarcasm detection in Twitter with hierarchical fusion model[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 2506-2515.
- [13] SANGWAN S, AKHTAR M S, BEHERA P, et al. I didn't mean what I wrote! Exploring multimodality for sarcasm detection[C]//International Joint Conference on Neural Networks. Piscataway: IEEE Press, 2020: 1-8.
- [14] XU N, ZENG Z X, MAO W J. Reasoning with multimodal sarcastic tweets via modeling cross-modality contrast and semantic association[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Lin-

- guistics, 2020; 3777-3786.
- [15] LIANG B, LOU C W, LI 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. Stroudsburg: Association for Computational Linguistics, 2022; 1767-1777.
- [16] 卢欣, 李旸, 王素格. 融合语言特征的卷积神经网络的反讽识别方法[J]. 中文信息学报, 2019, 33(5): 31-38.
- LU X, LI Y, WANG S G. Linguistic features enhanced convolutional neural networks for irony recognition[J]. Journal of Chinese information processing, 2019, 33(5): 31-38.
- [17] TANG Y J, CHEN H. Chinese irony corpus construction and ironic corpus construction and ironic structure analysis[C]//Proceedings of the 25th International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2014; 1269-1278.
- [18] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2014; 1746-1751.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016; 770-778.
- [20] 胡文彬, 陈龙, 黄贤波, 等. 融合交叉注意力的突发事件多模态中文反讽识别模型[J]. 智能系统学报, 2024, 19(2): 392-400.
- HU W B, CHEN L, HUANG X B, et al. A multimodal Chinese sarcasm detection model for emergencies based on cross attention[J]. CAAI transactions on intelligent systems, 2024, 19(2): 392-400.
- [21] WU S Y, SU E T, LEI B Y, et al. TextCNN-based text classification for E-government[C]//Proceedings of the 6th International Conference on Information Science and Control Engineering. Piscataway: IEEE Press, 2019; 929-934.
- [22] SHARFUDDIN A A, TIHAMI M N, ISLAM M S. A deep recurrent neural network with BiLSTM model for sentiment classification[C]//International Conference on Bangla Speech and Language Processing. Piscataway: IEEE Press, 2018; 1-4.
- [23] JACOB D, CHANG M W, KENTON L, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2019; 4171-4186.
- [24] 陈俊松, 易积政, 陈爱斌. CNS-Net: 一种循环多注意力特征聚合架构[J]. 郑州大学学报(理学版), 2024, 56(2): 73-79.
- CHEN J S, YI J Z, CHEN A B. CNS-Net: a cyclic multi-attention feature aggregation architecture[J]. Journal of Zhengzhou university (natural science edition), 2024, 56(2): 73-79.
- [25] 计咏梅. 基于深度学习的卷积码盲识别方法研究[D]. 西安: 西安电子科技大学, 2023.
- JI Y M. Research on blind recognition method of convolutional code based on deep learning[D]. Xi'an: Xidian University, 2023.