Hans Journal of Data Mining 数据挖掘, 2025, 15(1), 82-93
Published Online January 2025 in Hans. https://www.hanspub.org/journal/hjdm
https://doi.org/10.12677/hjdm.2025.151007

Hans汉斯

短剧用户观感评价情感分析

丁文轩，李征宇

沈阳建筑大学计算机科学与工程学院，辽宁 沈阳

收稿日期：2024年12月24日；录用日期：2025年1月17日；发布日期：2025年1月23日

摘　要

短剧随着时代发展逐渐崛起，成为当今国内外新潮的娱乐载体。本文爬取腾讯短剧品牌十分剧场的短剧用户评价，对该不平衡样本数据进行情感分析，比较多种模型与模型组合的效率与效果。1) 使用Word2vec的连续词袋模型将预处理后的文本转为词向量，构建LSTM/BILSTM模型，两者无效果差别，LSTM所用时间最短；2) 构建TextCNN + LSTM/BILSTM模型，使用TextCNN获取向量特征，通过LSTM/BILSTM学习情感规律，稀少数据的F1-Score提升约10%；3) 构建TextCNN + LSTM + Muti_Head_Attention模型，添加多头注意力机制把握字与字之间的多重联系，耗时增加一倍，稀少数据的F1-Score上限再次提升1%；4) 使用随机删除增强数据会以降低20%的精准率的代价提高10%的召回率；5) 在第3点的基础上在卷积层中添加残差连接，稀少数据的F1-Score上限提高2%；6) 使用Bert/Roberta的分词器和模型取代Word2vec与传统RNN，得到的结果对比第5点，提升约为9%/12%，泛化性更强，时间和硬件成本大幅提升，但添加TextCNN、LSTM与多头注意力后，效果反而出现下降。

关键词

短剧，自然语言处理，情感分析，深度学习

Sentiment Analysis of the Users of Micro-Dramas

Wenxuan Ding, Zhengyu Li

School of Computer Science and Engineering, Shenyang Jianzhu University, Shenyang Liaoning

Received: Dec. 24th, 2024; accepted: Jan. 17th, 2025; published: Jan. 23rd, 2025

Abstract

As Micro-Dramas grow in popularity worldwide, this article evaluates user reviews from Tencent's "Shifen Theater", analyzing imbalanced data sentiment and comparing various models and combinations. 1) Word2Vec's bag-of-words model turns preprocessed text into vectors, building

文章引用: 丁文轩, 李征宇. 短剧用户观感评价情感分析[J]. 数据挖掘, 2025, 15(1): 82-93.
DOI: 10.12677/hjdm.2025.151007

LSTM/BiLSTM models—both perform poorly, with LSTM being the fastest; 2) The TextCNN + LSTM/BiLSTM model uses TextCNN for vector features and LSTM/BiLSTM for sentiment learning, boosting the F1-Score for rare data by about 10%; 3) Adding Multi-Head Attention to TextCNN + LSTM/BiLSTM captures intricate character relationships, doubling the runtime and increasing the F1-Score by 1%; 4) Random deletion enhances data but sacrifices 20% precision for 10% better recall; 5) Add residual connections to the convolution layers in model 3, improving the F1-Score by 2% on sparse data; 6) Replacing Word2Vec and traditional RNNs with Bert/Roberta improves results by 11%/14% over the third model, offers better generalizability, but increases time and cost significantly. However, incorporating TextCNN, LSTM, and Multi-Head Attention can decrease performance.

Keywords

Micro-Dramas, NLP, Sentiment Analysis, Deep Learning

1. 引言

根据国家广播电视总局的定义，微短剧指的是单集时长从几十秒到 15 分钟左右、有着相对明确的主题和主线、较为连续和完整的故事情节的网络剧集[1]。近两年来，随着影视行业投资缩水，作品数量下滑，大量行业相关工作者面临无戏可拍的困境，微短剧一跃成为新的行业风口。十分剧场是腾讯视频设立的首个微短剧品牌，标志着腾讯视频对短视频领域精品化探索的新阶段，旗下短剧《执笔》上线两周分账破千万，微短剧播放指数蝉联 17 日 TOP1，是腾讯视频短剧的代表作。

2. 相关研究

文本情感分析技术是自然语言处理的一部分，又称为意见挖掘，是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程[2]。文本情感发展至今，共有三种主要的研究方法，基于情感字典[3]，基于机器学习[4]，基于深度学习[5]。基于情感词典的方法通过人为编撰的对各种词汇情感赋分的词典，对文本进行情感分析，随着大数据发展，其滞后性与忽略上下文语义关系的缺点越发凸显。传统机器学习对于小规模样本数据集的效果突出，但需要建立复杂的特征工程，且随着数据量的增加，其准确率会逐步下降。深度学习在相关领域大行其道，省略了复杂的抽取特征的步骤，且对于大数据的效果比起传统机器学习方法要更加突出，例如顾昕健[6]使用 Mish 函数代替 TextCNN 中的 Relu 函数，一定程度上规避了负输入的梯度损失问题，陈可嘉[7]等人使用 Emoji2vec 获取表情向量，分别从局部级和全局级表情符号注意力机制层面加强表情符号与文本结合后的关键信息。臧洁[8]等人改进了 DPCNN 网络，添加多个多尺度卷积核，增加 Batch_Norm 与池化操作，可以提取更加丰富的文本长距离依赖关系并使用残差连接改善了反向传播中的梯度消失问题。鲁富宇[9]等人引入词语情感类别分布、情感倾向以及情感强度三个关键因子改进了词语的向量表示，使得模型可以更好地学习到向量中的信息。王海涛[10]等人提出了一种包括多层感知机、统计情感缩放、预构建词典、多通道处理和自注意力机制在内的模型，在融合颜文字符号的文本情感分类中有着良好的效果。

2.1. 情感分析

情感分析是 NLP 任务中的一项重要任务，通过挖掘目标文本中的信息与规律来判断出该文本的情绪

倾向。在经济、文化、政治等领域中有着广泛的应用，其方法不断演变更新，总结概括为三种。

2.1.1. 基于情感词典的情感分析

使用人工设计编撰的情感词典对文本中的情感词进行赋分，根据得分正负来判断文本的情感倾向。该方法非常依赖人工标注的词典的涵盖范围、时效性与网络性，导致对文本的情绪分类效果相对较差，用于特定领域的词典效果相对较好，结合深度学习效果更佳[11]-[13]。

2.1.2. 基于机器学习的情感分析

通过监督学习，将文本进行预处理并建立特征工程后，使用机器学习的方法，如线性回归，随机森林、支持向量机、朴素贝叶斯等进行分类[14] [15]。

2.1.3. 基于深度学习的情感分析

基于深度学习的方法有 Word2vec、FastText、Seq2Seq、LSTM、GRU、TextCNN、Transformer、GPT、Bert、XLNet、Roberta 等，自然语言处理的效果获得长足的进步。

2.2. Word2vec

自然语言处理(NLP)是一门研究计算机如何理解、处理、分析、生成、模拟的学科，自然语言处理中，最细粒度的单位为词语，词语组成句子，句子组成段落，段落组成文章，文章组成文档。计算机无法理解人类的语言，因此需要将文字变为计算机可以理解的语言，即词嵌入(word-embedding)：将文本映射到向量空间中，表示为实数向量。

Word2vec 是 Tomas.Mikolov 等人在《Efficient Estimation of Word Representation in Vector Space》[16]一文中提出，谷歌于 2013 年开放这款软件，根据给定的语料库，使用优化后的训练模型将文本高速有效地转化为向量形式。Word2vec 主要依赖两种模型，连续词袋(Continuous Bag-of-Words, CBOW)与跳字模型(Skip-Gram)。Word2vec 模型促进了自然语言处理各领域研究的发展，不足之处在于无法准确把握词语上下文之间的联系。连续词袋模型通过中心词文本序列周围的背景词来预测中心词，由于背景词存在多个，因此取各背景词的向量的均值，计算中心词条件概率，分别设 $v_i \in R^d$，$u_i \in R^d$ 为词典内词索引为 i 的背景词与中心词，设中心词 ω_c 在词典中的索引为 c，背景词，$\omega_{o1},\omega_{o2},\cdots\omega_{ok}$ 在词典中索引为 $o1,o2,\cdots,ok$，根据背景词预测中心词的公式(1)所示，连续词袋结构如图 1 所示。

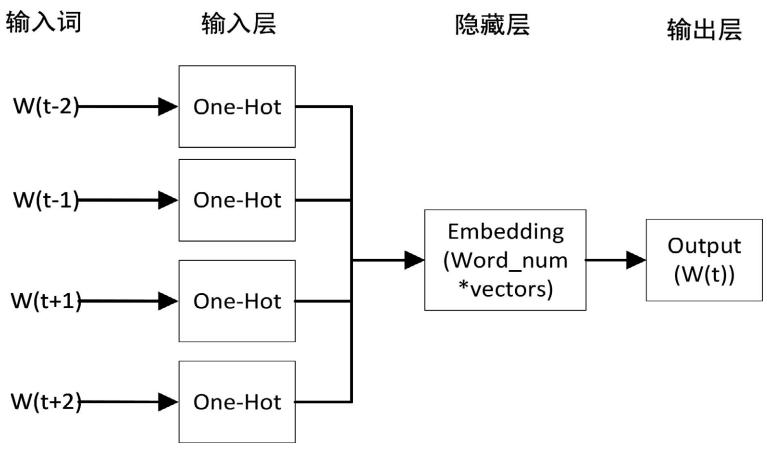

Figure 1. Word2vec's Continuous Bag-of-Words (CBOW) architecture
图 1. Word2vec 连续词袋结构

$$p\left(\omega_c \middle| \omega_{o1}+\omega_{o2}+\cdots+\omega_{o2k}\right)=\frac{\exp\left(\frac{1}{2k}u_c^T\left(v_{o1}+v_{o2}+\cdots+v_{o2k}\right)\right)}{\prod_{i\in v}\exp\left(\frac{1}{2k}u_i^T\left(v_{o1}+v_{o2}+\cdots+v_{o2k}\right)\right)} \tag{1}$$

2.3. TextCNN

卷积神经网络广泛应用于提取特征方面，尤其是计算机视觉领域的工作。2014 年，Yoon Kim [17]针对 CNN 模型进行了一定的改动，相对于一般的 CNN 模型，结构大体相同，但模型卷积核只有高度方向的特征捕获，更加适用于自然语言处理。其缺点在于，对于文本信息，往往将整句作为一个向量处理，不能很好地处理词序信息，卷积核较小，无法捕捉长距离特征，同时池化层可能会丢弃部分有用特征，单个模型效果不如 LSTM 模型。TextCNN 网络结构图如图 2 所示。

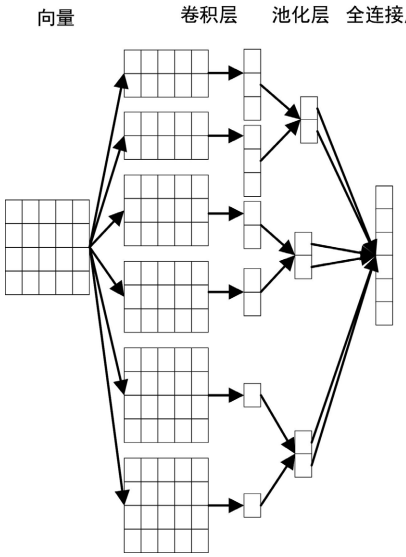

Figure 2. TextCNN network architecture
图 2. TextCNN 网络结构

2.4. LSTM

LSTM 的网络结构分为输入门，遗忘门，输出门，相对于传统 RNN 模型存在的长期依赖导致的梯度消失，与梯度中存在指数函数容易导致梯度爆炸的问题，LSTM 不同之处在于添加了细胞状态与遗忘门，细胞状态能让信息在序列中传递下去，遗忘门通过 Sigmoid 函数得到 0 和 1 的向量，从而控制遗忘或传递信息，梯度则是下一个留存的信息对前一个偏导的累乘，如果差别不大，则偏导无限接近于 1，使得在一定程度上抑制梯度消失或爆炸。BILSTM 是双向 LSTM，由两个 LSTM 组成，分别处理正反向的序列，从而更好地捕捉双向的语义依赖，再将双向结果拼接，得到结果。模型缺陷在于，其超参数过多，调优难度高，且解释性差，并非完全抑制梯度消失或爆炸，对序列特征的依赖使得并行处理能力被限制。LSTM 单元图与 BILSTM 网络结构图如图 3，图 4 所示。

2.5. Transformer

Transformer 的编码器与解码器分别由 6 个 Encoder Block 与 6 个 Decoder Block 组成，整个模型围绕注意力机制，为自然语言处理开创了新的领域。其中的多头注意力为多个自注意力模型组成，通过输入

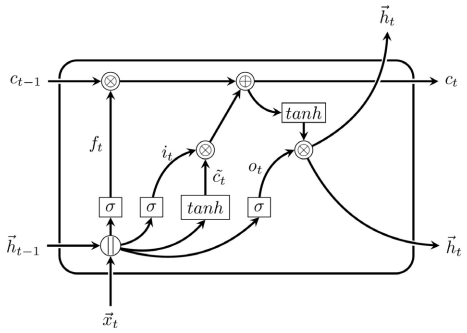

Figure 3. LSTM (Long Short-Term Memory) unit
图 3. LSTM 单元

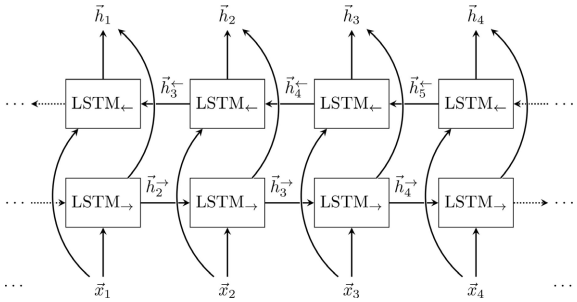

Figure 4. BILSTM Network Architectur
图 4. BILSTM 网络结构

矩阵 X 与权重矩阵 W^Q, W^K, W^V 相乘得到矩阵 Q(查询)，K(键值)，V(值)，从而计算出自注意力模型的输出。Transformer 相对于传统 RNN 模型，其对于长距离关系关系的捕捉表现更加优秀，并行计算能力大幅提高，但同时，计算成本，调参难度，数据量的需求，硬件成本都有大幅提高。结构图如图 5 所示。

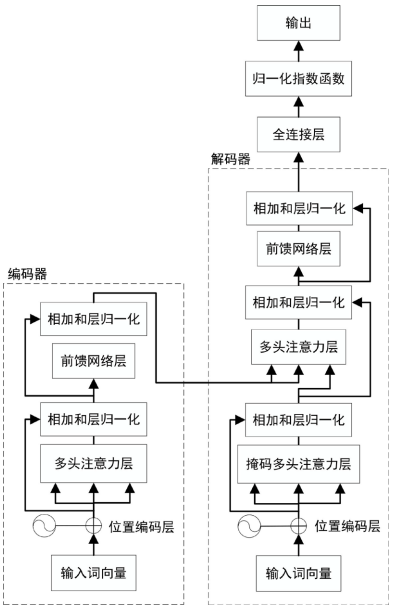

Figure 5. Transformer network architectur
图 5. Transformer 结构图

2.6. Bert

BERT (Bidirectional Encoder Representation from Transformers)是 2018 年 10 月由 Google AI 研究院提出的一种预训练模型，没有使用传统的 RNN 与 CNN 模型，而是使用了多层的 Transformer 结构，大致意思为双向 Transformer 的 Encoder 结构，相对于其他基于 Transformer 的模型，双向结构的 Bert 添加的 Masked Language Model 与 Next Sentence Prediction 进行预训练，效果更好，泛化能力与鲁棒性更强，在自然语言处理的各领域都获得了里程碑式的卓越成绩。Roberta 是 Bert 的增强版，更大的 batch_size，更庞大的训练数据，更长的语句，使用动态掩码并去除了 NSP。

2.7. 残差连接

深度学习相对于浅层学习，可以学习到更复杂的功能，与之相对的是，训练中，深度学习模型会随着架构加深而退化[18]。Srivastava R K [19]等人受到 LSTM 的启发，提出了残差连接(Residual Connection, ResC)的结构，即公式(2)，当 $T(x,WT)=0$ 时，$Y=x$，当 $T(x,WT)=1$ 时，$Y=H(x,WH)$。He K [20]等人简化了公式，即公式(3)，做了更多的实验验证，残差连接广泛应用于各种模型的架构中。残差连接保留了更多的原始文本中的信息，当输入权重矩阵完全退化后，残差连接使得模型恢复表达能力，有效缓解了模型退化的情况。结构图如图 6 所示。

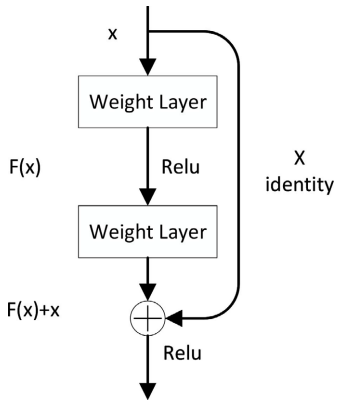

Figure 6. Residual connection
图 6. 残差连接

$$Y=H(x,WH)\cdot T(x,WT)+X\cdot(1-T(x,WT)) \tag{2}$$

$$Y=H(x,WH)+X \tag{3}$$

2.8. SwiGlu

SwiGLU [21]是 2020 年谷歌提出的一种基于 Swish 与 GLU 的激活函数。SwiGLU 广泛应用于各种大模型的 FFN 中，效果显著。GLU (Gated Linear Units，门控线性单元)通过引用两个不同的线性层，其中一个经过 Sigmoid 函数，两者逐元素相乘，从而实现门控并行处理时序数据。

$$Glu(x,W,V,b,c)=\sigma(xW+b)\odot(xV+C) \tag{4}$$

Swish 是一种自我门控的激活函数，与传统的 Relu 不同，它允许负值输入，并简化了门控机制，一定程度上缓解了梯度消失问题。其系数 $\beta=0$ 时，函数趋近于线性函数 $y=x$^2。$\beta=1$ 时，函数光滑且单调，等价于 Silu 函数。β 趋近于∞时，函数趋近于 Relu 函数。

$$\mathrm{Swish}_{\beta}(x)=x\times\mathrm{Sigmoid}(\beta x) \tag{5}$$

公式(5)带入公式(4)，取代 Sigmoid 函数，得到 SwiGLU 的表达式：

$$\mathrm{SwiGlu}(x,W,V)=\mathrm{SwiGlu}_{1}(xW)\odot(xV) \tag{6}$$

在前馈神经网络中的表达式如下：

$$\mathrm{FNNSwiGlu}(x,W_1,V,W_2)=\left(\mathrm{Swish}_{1}(xW_1)\odot(xV)\right)W_2 \tag{7}$$

Swish 激活函数克服了传统 Relu 函数负数区域神经元坏死的情况，而 GLU 的门控机制可以让神经网络学习到有用的表示，提高模型的泛化能力，对于长序列，长距离依赖的文本有良好效果，SwiGLU 结合了两者的优点，其中的参数可以根据不同的任务动态调整，使得模型有着良好的适应性。

2.9. 小结

本章介绍了后续章节所用的模型，通过构建 LSTM/BILSTM、TextCNN + LSTM/BILSTM、TextCNN + LSTM/BILSTM + Muti_head_attention、Bert/Roberta 模型，比较不同模型之间的效果与效率，上限与成本，添加数据增强方法与否，从而对模型选择有更清晰的认知。

3. 数据获取及数据预处理

3.1. 数据获取

本文以腾讯视频的十分剧场中，用户评价超过一千的短剧《执笔》《授她以柄》《唐朝异闻录》《江城诡事》《千年情劫》为研究对象，利用 python 设计爬虫，爬取评分页面数据，共有 18,862 份数据，信息囊括：评论 ID，打分，推荐态度，评论内容，观看时长，用户名。爬取数据无缺失值，筛选重复值并剔除，去除文本长度小于 2 的评价，评分推荐态度 1 至 5 分分别为：不推荐，一般推荐，推荐，力荐，强烈推荐，根据人工筛查，三分基本倾向于好评，因此设 3 分及以上为好评，以下为差评，人工去除差分好评的讽刺与无意义干扰评价，好评共 12,835 条，差评共 880 条，属于极度不平衡数据。

3.2. 数据预处理

对文本进行去除停用词处理，去除停用词，首先可以显著减少数据量，提高运行效率，其次是为了避免停用词影响更有意义的词语，从而提高特征提取的准确性。本文使用汇总的中文停用词表，综合了百度停用词、哈工大停用词及四川大学停用词。其次由于部分语句包含大量符号、表情，空格，因此通过正则表达式筛掉非中文、数字和英文的内容，并进行机械去重，去除重复文字，效果如表 1 所示。

Table 1. Preprocessing effect
表 1. 预处理效果

原评论	处理后
这剧真的好好看😍导演好会拍！太短了！都不够看！太喜欢了！	这剧 真的 好看 导演 好会 拍 太短 不够看 太喜欢
不落俗套，演技剧情颜值都过关，剧情转换流畅丝滑，拍出了电影的感觉。[一起冲]	不落俗套 演技 剧情 颜值 过关 剧情 转换 流畅 丝滑 拍出 电影 感觉
内容啰嗦又乱令人反感	内容 啰嗦 乱 令人 反感
好好好好好看	好看

3.3. 词向量化

本文使用 python 中 gensim 库的 Word2vec 模型的连续词袋模型，具体参数配置如表 2 所示。

Table 2. Word2vec hyperparameters
表 2. Word2vec 超参数

超参数	超参数值
sg	0
size	100
window	10
min_count	1
workers	4
epochs	20

训练并得到词典后，将文本数据转换为向量形式，不在词典中的文本转化为相同大小的零向量矩阵。

4. 多种模型构建

4.1. 数据集划分

短剧评论语料转换为向量后，以 8:1:1 的比例划分为训练集、验证集与测试集，将词向量转换为 torch 张量，文本向量为 float32，标签为长整型，并使用 pad_sequence 规定向量长度，超过则截断，少于则填充 0，转换类型后，数据按批次封装进 Dataloader 中。

4.2. 模型构建

4.2.1. 模型一

构建 LSTM/BILSTM 模型，学习率动态变化策略选择 warm-up + cos 退火，超参数如表 3 所示。

Table 3. LSTM/BILSTM hyperparameters
表 3. LSTM/BILSTM 超参数

超参数	超参数值
Input_size	vector_size(100)
Hidden_size	100
Num_layers	3
Num_epochs	15
Learning_rate	3e-4
Criterion	CrossEntropyLoss
Optimizer	AdamW
Num_Warmup_Steps	0.1*training_total_steps

4.2.2. 模型二

TextCNN 可以更好地抓住向量中有价值的特征，结合 TextCNN 与 LSTM/BILSTM 模型，TextCnn 的

网络结构分为四层，嵌入层，卷积层，池化层与输出层，卷积层对嵌入的词向量进行卷积操作，提取不同大小的 n-gram 特征，通过不同大小的卷积核获取更全面的局部特征。池化层对卷积结果进行池化，将结果剪切为固定长度的特征向量，减少特征维度，提取重要特征。TextCnn 提取重要特征后，传入 LSTM，探索特征中有价值的信息。相对于 LSTM/BILSTM，加上 TextCNN 后的运行效率差别不大，但效果得到一定提升，模型参数大致如图 7 所示。

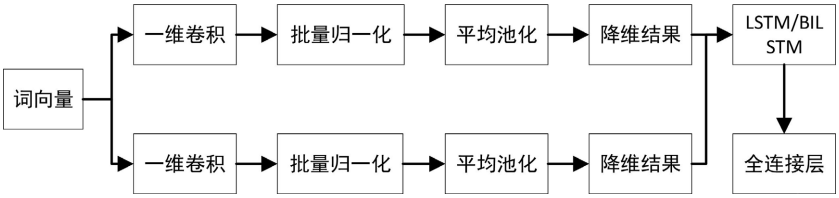

Figure 7. Model 2 structure
图 7. 模型二结构

本文构建三层 TextCNN，结构为 Conv1D + BatchNorm1d + AvgPool1D + Dropout + LSTM/BILSTM。导入 hyperopt，使用 fmin，tpe，hp，Trials 实现贝叶斯优化，对模型的各个参数进行择优，学习率动态变化策略选择 warm-up + cos 退火，在训练初期保持稳定，后期通过减少学习率增加模型收敛速度，使得模型效果更佳。参数如表 4 所示。

Table 4. TextCNN + LSTM/BILSTM hyperparameters
表 4. TextCNN + LSTM/BILSTM 超参数

超参数	超参数值
Input_size	vector_size (100)
Conv_First_Out_Channels	64
Kernel_Size	3
Activation Function	SwiGLU
Hidden_Size	100
Num_Layers	3
Num_Epochs	15
Learning_Rate	3.5e−4
Dropout_Rate	0.1
Criterion	CrossEntropyLoss
Optimizer	AdamW
Num_Warmup_Steps	0.1*training_total_steps

4.2.3. 模型三

多头注意力机制(Muti_head_attention, MHA)联合多个自注意力模型，可以最大程度注意到每个词之间的多种联系与差别。构建 TextCNN + (LSTM/BILSTM) + MHA，参数如表 5 所示。

Table 5. TextCNN + LSTM/BILSTM + MHA hyperparameters
表 5. TextCNN + LSTM/BILSTM + MHA 超参数

超参数	超参数值
Num_Heads	16

4.2.4. 模型训练、验证及测试

每轮训练后预测验证集，选择验证集效果最好的轮次模型预测测试集，Precision，Recall，F1-Score展示正/负面评价的情况，结果如表 6 所示。

Table 6. Model Test Results
表 6. 模型测试结果

模型	秒时	Accuracy	Precision	Recall	F1-Score
LSTM	20.64	0.94	0.97/0.65	0.97/0.59	0.97/0.62
BILSTM	49.57	0.94	0.97/0.61	0.97/0.62	0.97/0.61
TextCNN + LSTM	204.69	0.95	0.96/0.83	0.99/0.51	0.98/0.63
TextCNN + BILSTM	352.99	0.95	0.96/0.80	0.99/0.51	0.98/0.62
TextCNN + BILSTM + Random_Delete	471.24	0.94	0.97/0.58	0.96/0.63	0.97/0.61
TCNN + MHA	195.36	0.95	0.96/0.75	0.99/0.52	0.97/0.61
TCNN + LSTM + MHA	433.65	0.95	0.96/0.76	0.99/0.53	0.97/0.64
TCNN + BILSTM + MHA	579.58	0.95	0.96/0.75	0.99/0.52	0.97/0.62
TCNN + ResC + BILSTM + MHA	189.14	0.95	0.97/0.74	0.98/0.59	0.97/0.66

由于数据不平衡，因此以 F1-Score 为模型效果的主要衡量标准。由表 4 可以看出，BILSTM 与 LSTM 在该数据集内效果几乎相同，BILSTM 训练所需的时间比 LSTM 模型高了 84.65%，好评的各项数据基本一致，差评数据的 F1-Score 最高可达到 0.62。添加 TextCNN 后，模型的准确率提高了 1%，差评预测的精准率有 18%以上的提升，同时召回率下降约 10%，召回率最高可达 0.63，提升了 1%，差距不大，所耗时间为无 TextCNN 的 2~3 倍，通过随机删除的方式增强负面评价，只能以精准率降低 20%左右的代价提高 10%左右的召回率。TextCNN 添加多头注意力模型后，效果略差于 LSTM/BILSTM，添加 LSTM/BISLTM 后，F1-Score 最好成绩突破到 0.64，提高了 1%，所耗时间提高了一倍。在模型中加入残差连接，训练验证的时间大幅缩短，且差评数据的 F1-Score 最高可达 0.66，效果显著。

4.2.5. 模型四

构建 Bert/Robert 模型，本文使用的是谷歌团队针对中文进行预训练的 bert-base-chinese 模型，和哈大讯飞与科大讯飞联合实验室的中文 Roberta 预训练模型，使用分词器对文本进行向量化，超参数如表 7，表 8 所示。

Table 7. BertTokenizer hyperparameters
表 7. BertTokenizer 超参数

超参数	超参数值
add_special_tokens	vector_size (100)

续表

max_length	64
padding	3
truncation	SwiGLU
return_tensors	100

Table 8. Model hyperparameters
表 8. Model 超参数

超参数	超参数值
Num_Epochs	15
Learning_Rate	3e−4
criterion	CrossEntropyLoss
optimizer	AdamW

每轮训练后预测验证集，选择验证集损失最小轮次模型预测，表现结果如表 9 所示。

Table 9. Bert/Roberta Test Results
表 9. Bert/Roberta 测试结果

模型	Accuracy	Precision	Recall	F1-Score
Bert	0.97	0.98/0.84	0.99/0.66	0.98/0.74
Roberta	0.97	0.98/0.92	1.00/0.66	0.99/0.77

可以看出，即使没有经过超参数微调或更复杂的结构，效果也超过之前的所有模型，但需要更高的硬件要求与时间，从经济的角度看，TextCNN + LSTM + Muti_head_Attention 已经足够，从结果来看，Bert/Roberta 效果提升明显。而添加 TextCNN、LSTM、多头注意力后，模型效果反而出现一定下降，如表 10 所示。

Table 10. Bert/Roberta test results
表 10. Bert/Roberta 测试结果

模型	Accuracy	Precision	Recall	F1-Score
Bert + LSTM	0.96	0.97/0.81	1.00/0.36	0.98/0.50
Roberta + LSTM + MHA	0.97	0.97/0.84	0.99/0.52	0.98/0.64

5. 结论

经过多种模型的对比，可以看出，相对于 Word2vec，Bert 在中文文本情感分类中展现出了更良好的效果，同时所耗成本也远高于 Word2vec，使用 Word2vec + TextCNN + LSTM + Muti_head_attention + 模型效果最优，时间合理，经过贝叶斯优化调整超参数后，以 F1-Score 为衡量标准，模型效果相对于未调参提升约为 10%。单纯使用 Bert/Roberta，非调参情况下，效果提升约为 11%与 14%，各项指标都十分突出，泛化性更强，添加 TextCNN、LSTM、多头注意力后，效果反而下降。

参考文献

[1] 国家广播电视总局办公厅. 国家广播电视总局办公厅关于进一步加强网络微短剧管理实施创作提升计划有关工作的通知[EB/OL]. https://www.nrta.gov.cn/art/2022/12/27/art_113_63062.html, 2024-08-31.

[2] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.

[3] 赵妍妍, 秦兵, 石秋慧, 等. 大规模情感词典的构建及其在情感分类中的应用[J]. 中文信息学报, 2017, 31(2): 187-193.

[4] 严军超, 赵志豪, 赵瑞. 基于机器学习的社交媒体文本情感分析研究[J]. 信息与电脑(理论版), 2019, 31(20): 44-47.

[5] 梁宁. 基于注意力机制及深度学习的文本情感分析研究[D]: [硕士学位论文]. 北京: 华北电力大学, 2019.

[6] 顾昕健, 陈涛. 基于深度学习的评论文本情感分析[J]. 信息技术与信息化, 2024(7): 38-42.

[7] 陈可嘉, 夏瑞东, 林鸿熙. 融合双重表情符号注意力机制的文本情感分类[J]. 北京航空航天大学学报, 2024: 1-15.

[8] 臧洁, 鲁锦涛, 王妍, 等. 融合双通道特征的中文短文本情感分类模型[J]. 计算机工程与应用, 2024, 60(21): 116-126.

[9] 鲁富宇, 冷泳林, 崔洪霞. 融合 TextCNN-BiGRU 的多因子权重文本情感分类算法研究[J]. 电子设计工程, 2024, 32(10): 44-48+53.

[10] 王海涛, 何鑫, 施屹然, 等. 基于结构性符号文本的多通道情感分类模型构建[J]. 智能计算机与应用, 2024, 14(10): 164-169.

[11] 李琳, 韩虎, 范雅婷. 基于旅游领域的细粒度情感词典构建方法[J]. 北京航空航天大学学报, 2024: 1-12.

[12] 卫青蓝, 何雨, 宋金宝. 基于语义规则的自适应情感词典自动构建算法[J]. 北京航空航天大学学报, 2024: 1-11.

[13] 王浩畅, 王宇坤, Marius Gabriel Petrescu. 融合情感词典与深度学习的文本情感分析研究[J]. 计算机与数字工程, 2024, 52(2): 451-455.

[14] 胡梦雅, 樊重俊, 朱玥. 基于机器学习的微博评论情感分析[J]. 信息与电脑(理论版), 2020, 32(12): 71-73.

[15] 张财, 马自强, 闫博. 基于机器学习的政务微博情感分析模型设计[J]. 计算机工程, 2024, 50(12): 386-395.

[16] Mikolov, T., Chen, K., Corrado, G., *et al*. (2013) Efficient Estimation of Word Representations in Vector Space.

[17] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the* 2014 *Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), Doha, October 2014, 1746-1751. https://doi.org/10.3115/v1/d14-1181

[18] Orhan, A.E. and Pitkow, X. (2017) Skip Connections Eliminate Singularities.

[19] Srivastava, R.K., Greff, K. and Schmidhuber, J. (2015) Highway Networks.

[20] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[21] Shazeer, N.M. (2020) GLU Variants Improve Transformer.