

基于 ERNIE-BiGRU 和注意力机制的 民航旅客评价情感分析

许雅奎, 鲁健平

(中国民用航空飞行学院经济与管理学院, 四川 广汉 618307)

摘要: 针对传统文本情感分析获取词向量信息不充分以及情感资源未得到充分利用, 导致在判断评论情感极性所依赖的文本信息不足, 提出 ERNIE-BiGRU-Attention 民航旅客评价情感分类模型。首先, 应用简单数据增强技术 (easy data augmentation, EDA) 对数据集进行处理。然后基于预训练语言模型 (enhanced representation through knowledge integration, ERNIE) 对文本进行情感知识提取。在特征提取方面, 引入双向门控循环单元 (bi-directional gate recurrent unit, BiGRU) 与注意力机制。结果表明, 该模型在分类上表现优异, 综合 F_1 为 0.975 9, 准确率较对比模型提升 0.73%。

关键词: 情感分析; BiGRU 模型; 数据增强; 注意力机制

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1671-1807(2024)16-0103-06

为贯彻数字中国建设整体布局, 推动智慧民航建设更好发展, 加快推进数字技术创新应用, 要充分发挥民航大规模市场和海量数据规模以及丰富应用场景的优势, 以智慧民航建设推动行业高质量发展^[1]。智慧民航建设其中一部分是智慧出行建设, 对民航旅客评价做情感分析, 充分了解旅客民航出行体验, 有利于航空公司对市场做出调整, 进一步提升民航服务质量与效益。

随着全球民航业的快速发展和竞争的加剧, 航空公司越来越关注客户体验和满意度, 民航旅客评价是航空公司了解旅客对服务的态度和感受的重要途径。随着社交媒体和在线评价平台的兴起, 大量的旅客评价数据涌现出来, 手动分析这些数据变得非常困难和耗时。因此, 基于自然语言处理技术的民航旅客评价情感分析研究应运而生, 旨在利用自然语言处理技术和情感分析算法来自动化地分析和理解旅客的评价, 通过分析旅客评价的反馈, 挖掘出当前服务所存在的缺点及旅客偏好, 规划提升服务质量方案, 为旅客提供更好的服务。

文本情感分析是自然语言处理领域的一个研

究方向, 旨在自动识别文本中所包含的情感倾向, 通常包括正面情感、负面情感和中性情感。情感分析可以帮助公司了解用户对产品、服务、事件等的情感态度, 从而更好地理解用户需求和市场反应^[2]。情感分析可采用多种技术, 常见的有基于情感词典的方法、基于机器学习的方法、基于深度学习的方法, 还有基于预训练语言模型的方法^[3]。

本文使用航空公司旅客评价数据集, 但是该数据集样本量偏小, 因此对其进行数据增强, 使用基于 ERNIE3.0 的预训练语言模型获取文本向量特征表示, 引入双向门控循环单元加注意力机制在情感提取层进行特征提取, 最后用 Softmax 函数分类, 以此来提高民航旅客评价文本特征的利用率, 进而提升旅客评价情感分类的准确率。

1 情感分析研究现状

目前有多种情感分析算法, 包括基于情感词典的方法、基于机器学习的方法、基于深度学习的方法和基于预训练语言模型的方法。

情感词典是传统情感分析方法中的重要工具^[4], 是一种基于预定义的情感词典进行情感分析的技术, 它的基本思想是通过匹配文本中出现的情

收稿日期: 2024-04-13

基金项目: 中央高校基本科研业务费专项资金 (ZJ2022-008); 面向民航高质量发展目标的地面服务复合型人才校企协同培养基地建设 (MHJY2024039)

作者简介: 许雅奎 (1976—), 女, 四川成都人, 硕士, 副教授, 研究方向为系统决策与优化、数据挖掘; 鲁健平 (2000—), 男, 河南汝州人, 硕士研究生, 研究方向为预测与决策分析、数据挖掘。

感词和对应的情感极性,来判断文本的情感倾向。该方法需要构建一个情感词典,其中包含各种情感词汇和对应的情感极性,然后将待分析的文本中的情感词与情感词典进行匹配,根据匹配到的情感词和极性计算文本的情感得分,最后根据情感得分将文本划分为积极、消极或中性情感类别。情感词典分析方法的优点是简单易实现,不需要大量标注数据,对于一些简单的情感分析任务效果较好。然而,它也存在一些缺点,如对于歧义性较强的文本,可能会导致情感分析结果不准确。Ahmed等^[5]通过学习句子的整体表示来捕捉情感,提出了一种结合手动和自动方法的弱监督神经网络模型,通过重新构建输入句子的表示,成功构建了多语言情感词典。研究结果显示,该情感词典显著提升了情感极性判别的有效性。

基于机器学习的情感分析算法通常使用支持向量机(support vector machines, SVM)、随机森林(random forest)、朴素贝叶斯(naive Bayes)等^[6]特征,并使用这些特征训练分类器。尽管这些方法在某些场景下表现良好,但其效果通常受限于特征的质量和数量。Rathor等^[7]提取了亚马逊评论,结合字母加权对比分析了三种机器学习算法,结果表明机器学习算法在加权一元组上运行良好,并且SVM实现了最佳精度。

基于深度学习的情感分析方法通常采用卷积神经网络(convolutional neural networks, CNN)、循环神经网络(recurrent neural network, RNN)、长短期记忆网络(long short-term memory networks, LSTM)等^[8]。神经网络在文本生成、向量表示、词表示估计、句子分类、句子建模和特征表示方面表现不错。使用深度学习模型,可以自动学习文本中的语义和上下文关系,周湘贞等^[9]采用RNN做情感分类,并结合注意力机制对词特征进行加权,结果表明该模型有很好的分类性能。李海峰和周壁刚^[10]采用CNN-BiLSTM,并融合多头自注意力机制的电商评价情感模型,结果表明该模型优于其他模型。

在基于预训练语言模型的情感分析方法研究领域,Peters等^[11]提出ELMo模型采用两层双向LSTM语言模型学习上下文本信息,结果表明该模型有效学习了上下文本信息。Devlin等^[12]提出BERT模型旨在从未标记的文本中学习深度双向表示,结果表明该模型对上下文特征提取效果更好。相较于BERT,ERNIE改进了掩码策略并且

融合自回归网络和自编码网络以此来增强对情感信息的学习,因此可以更好地来处理自然语言理解任务^[13]。沈彬等^[14]针对情感资源未充分利用等问题,提出一种基于ERNIE和双重注意力机制的微博情感分析模型,结果表明模型效果良好。本文采用百度提出的ERNIE3.0预训练语言模型来对民航旅客评价进行情感分析,并结合注意力机制提取特征,借此来提高模型的情感分析效果。

2 模型构建

在对旅客评价数据进行处理后,先使用预训练语言模型ERNIE3.0获取文本向量特征表示,引入双向门控循环单元结合注意力机制在情感提取层进行特征提取,最后用Softmax函数分类,模型总体结构如图1所示。

2.1 输入层

预训练语言模型已被证明可以有效地提高各种自然语言处理任务的性能,在本文中ERNIE3.0模型扮演着民航旅客评价语义特征提取器的角色。ERNIE^[15]的编码层采用了Transformer结构,该结构由多个编码器层(encoder layer)组成,每个编码器层又由多头注意力机制和前馈神经网络构成^[15]。使用ERNIE3.0对文本进行特征编码,首先将处理

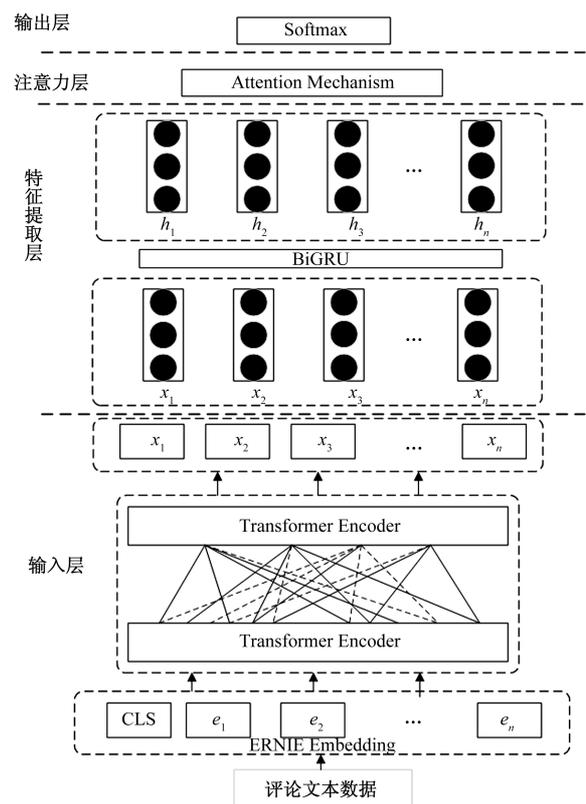


图1 ERNIE-BiGRU-Attention模型结构

后的文本输入到 ERNIE 的 Embedding 层得到词嵌入 $E = \{e_1, e_2, \dots, e_n\}$, 然后将词嵌入输入到 ERNIE 的编码层中, 经过计算后得到评论文本的特征向量矩阵 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ 。

2.2 特征提取层

2.2.1 BiGRU 双向门控单元上下文特征提取

语义特征抽取层主要是对 ERNIE 编码层提取的词嵌入进行语义的细化。RNN 在自然语言处理、语音识别、时间序列预测等任务中表现出色。然而, 传统的 RNN 在处理长序列时常面临梯度消失和梯度爆炸等问题, 导致难以捕捉长期依赖关系^[16]。为了克服这一问题, 出现了一些 RNN 的变体, 如 LSTM 和 GRU。这些变体通过引入门控机制有效地解决了梯度消失和梯度爆炸问题, 从而使得 RNN 能够更好地处理长序列数据。胡荣磊等^[17]提出一种基于长短期记忆网络和注意力模型的文本情感分析方法, 研究结果表明该方法具有显著的有效性和可行性。另一方面, Cho 等^[18]在 2014 年提出的 GRU, 根据 Chung 的实验, 通过在某些数据集上对所有模型使用固定数量的参数, GRU 在 CPU 时间的收敛、参数更新和泛化方面都优于 LSTM 单元。因此, 本文采用 BiGRU 作为主要模型以提取特征。GRU 的特点是引入了重置门和更新门两个门控。在每个时间步, GRU 接收一个输入, 然后根据当前输入和上一个时间步的隐藏状态计算新的隐藏状态。BiGRU 是由两个单向且方向相反的 GRU 组成, 输出由两个 GRU 共同决定, 计算公式如式(1)~式(3)所示:

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

$$\overleftarrow{h}_t = \text{GRU}(x_t, \overleftarrow{h}_{t-1}) \quad (2)$$

$$h_t = w \vec{h}_t + v \overleftarrow{h}_{t-1} + b_t \quad (3)$$

式中: \vec{h}_t 为 t 时刻正向 GRU 的输出状态; \overleftarrow{h}_t 为 t 时刻反向 GRU 的输出状态; BiGRU 在 t 时刻的 h_t 隐层状态由 t 时刻的输入 x_t 、 $t-1$ 时刻正向的隐层状态输出 \vec{h}_{t-1} 和反向的隐层状态输出 \overleftarrow{h}_{t-1} 三个部分共同决定; h_t 为两个方向 GRU 隐层状态的加权和。对 ERNIE 预训练模型编码后的特征向量矩阵 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$ 进行深度提取评价文本局部特征。

2.2.2 注意力层

为了提取出评论文本中更关键的信息, 将 BiGRU 模型的输出 $\mathbf{H}_t = [h_t^1, h_t^2, \dots, h_t^n]$ 输入到注意力机制中, 通过注意力机制对每一时刻的输入状态进行加权, 对 BiGRU 输出 h_t 应用激活函

数, 然后使用权重矩阵 \mathbf{W} 与经过激活的 h_t 进行相乘 $h'_t \mathbf{W}_a$ 得到注意力权重分数矩阵 \mathbf{s} , 计算公式如式(4)~式(6)所示:

$$h'_t = \tanh h_t \quad (4)$$

$$\mathbf{s} = h'_t \mathbf{W}_a + b_t \quad (5)$$

$$\alpha_t = \text{softmax}(\mathbf{s}, \text{axis} = 1) = \frac{\exp s_t}{\sum_t s_t} \quad (6)$$

$$h'_a = \alpha_t \odot h_t \quad (7)$$

式中: \tanh 为激活函数, 目的是将输出值缩放到 $[-1, 1]$, 以便后续的注意力权重计算; \mathbf{W}_a 、 b_t 分别为注意力权重参数矩阵和偏置向量; α_t 为注意力权重; h'_a 是一个加权的注意力表示; \odot 为逐元素相乘。然后通过序列求和池化操作和激活函数 ReLU 得到全连接层输入文本特征表示。

序列求和池化和 ReLU 计算公式为

$$\mathbf{H}_s = h_a^1 + h_a^2 + h_a^3 + \dots + h_a^n \quad (8)$$

$$\mathbf{H}'_s = \text{ReLU}(\mathbf{H}_s) \quad (9)$$

将激活后的特征向量 \mathbf{H}'_s 送入两个全连接层得到最后的输出。用这样方法可以更好地捕捉文本特征关系, 从而提高模型的性能和泛化能力。

2.3 输出层

在模型输出层中, 包含有两个全连接层, 第一个全连接层公式为

$$\mathbf{H}'_{s1} = \mathbf{W}_1 \mathbf{H}'_s + b_1 \quad (10)$$

式中: \mathbf{H}'_{s1} 为经过序列求和池化和 ReLU 激活后的向量; \mathbf{W}_1 为第一个全连接层的权重矩阵; b_1 为第一个全连接层的偏置向量。

第二个全连接层公式为

$$\mathbf{H}'_{s2} = \mathbf{W}_2 \mathbf{H}'_{s1} + b_2 \quad (11)$$

式中: \mathbf{H}'_{s1} 为第一个全连接层的输出; \mathbf{W}_2 为第二个全连接层的权重矩阵; b_2 为第二个全连接层的偏置向量。

接下来, 可以对预测向量 \mathbf{H}'_{s2} 应用 Softmax 函数, 将其转换为概率分布公式为

$$\text{probs} = \text{Softmax}(\mathbf{H}'_{s2}) \quad (12)$$

通过训练损失函数对预测概率分布与真实标签之间的差异进行优化, 从而学习模型的参数, 使得模型能够更好地拟合训练数据并泛化到新数据。训练过程中使用优化算法来更新权重矩阵。

2.4 优化器与损失函数

对于模型训练时的优化器的选择, 选用 AdamW 优化器^[19], 它是在 Adam 优化器加 L2 正则化的基础上进行改进的算法, 解决了 Adam 优化器函数中 L2 正则化失效的问题, 相比较于传统的随机

梯度下降算法有更好的泛化能力^[20]。模型选用的损失函数为分类问题常用的交叉熵损失函数。

3 实证分析

3.1 数据集及实验环境介绍

采用来自 Kaggle 的“Twitter US Airline Sentiment”数据集,包含用户在推特上发表的与美国 6 家航空公司有关的个人评价,根据评价内容情感极性每一条评价都有对应的一个情绪标签,情绪标签共分为负向、正向或中性三种,分别对应类别 0、类别 1、类别 2。该数据集共包含 14 640 条数据记录,其中包括 9 178 条负向评价、2 363 条正向评价以及 3 099 条中性评价。取所有数据进行实验,实验环境如表 1 所示。

3.2 数据分析及预处理

数据处理部分采用孟天宇等^[21]对评价文本与处理的方法,在预处理过程中消除标点符号。然后把所有的单词都小写,去掉停止词。停止词是那些对判断评论是正面还是负面没有帮助的词。停顿词包括 are,has,the 和 others。为了降低模型的复杂度,做初步简单处理,以增强原始文本的情感特征。考虑原始数据量少且样本不均衡,可用 EDA^[22]提高文本分类任务的性能,EDA 包含同义词替换、随机插入、随机交换和随机删除 4 个简单但功能强大的操作。在五个文本分类任务上,表明 EDA 提高了卷积和循环神经网络的性能。因为数据集中句子的长度不同,而神经网络希望每个输入具有相同数量的时间步长,同一批次下样本长度需一致,为了满足这个标准,本文将所有的句子标准化为相同的长度。要做到这一点,本文确定句子的最大长度为阈值。对于不足该长度的样本在右侧进行填充,在这个阶段之后,所有的短语都具有相同的长度。表 2 展示了训练集数据进行数据处理前后的样本信息。

表 1 实验环境信息

硬件信息		环境配置		框架信息	
CPU	2 核	Python	3.10.10	框架版本	Paddle 2.5.0
GPU	Tesla V100				
总显存	16 GB				
总内存	16 GB				

表 2 数据样本信息

数据	最大长度	最小长度	平均长度	数据量
原始	67	5	32	14 640
处理后	57	3	24	29 280

3.3 实验指标

对于本文提出的模型,实验以准确率(accuracy, acc)、精确率(precision)、召回率(recall)和 F_1 值为指标,用于衡量模型性能。首先计算二分类混淆矩阵,定义如表 3 所示。

各指标计算公式为

$$\text{accuracy} = \frac{TP + TN}{TP + EP + FN + TN} \quad (17)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (18)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (19)$$

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (20)$$

3.4 ERNIE-BiGRU-Attention 模型训练与评估

本模型实验过程分析结果对于本文提出的 ERNIE-BiGRU-Attention 航空公司旅客情感多分类模型,选用分类模型常用的准确率、损失值评价指标。本次实验分为四组实验,第一组分析不同分类模型在二分类任务上的效果,第二组分析预训练模型的加入在二分类任务上的效果,第三组实验是在原始数据与增强后的数据在不同模型上的效果,第四组实验是有注意力与无注意力在不同模型上的效果。通过四组实验来验证本文所提模型的构建过程和有效性。实验参数设置如表 4 所示。

不同分类模型对比结果如表 5 所示。Text CNN 模型的评估准确率相对较低, BiGRU 模型的评估准确率较高,表现不错。ERNIE 模型的评估准确率高,表现良好,添加了 CNN 层的 ERNIE 模型,准确率略有提升,但模型结果较差。结合了 BiGRU

表 3 二分类混淆矩阵

样本情况	预测正	预测负
样本正	TP	FN
样本负	FP	TN

表 4 参数设置

名称	初始设置
hidden_size	128
dropout	0.1
TransformerEncoderLayers	12
Gru_num_layers	3
AM_size	32
Batch_size	32
epoch	5
Num_classes	2
learning_rate	2×10^{-5}
num_aug	1

的 ERNIE 模型,准确率显著提升,因为预训练模型可以有效提取文本情感信息,结果也表明预训练模型的加入确实可以提高模型评价能力。在 ERNIE-BiGRU 模型上添加 Attention 层,准确率提高, F_1 稍有下降,但 EDA-ERNIE-BiGRU-Attention 模型显示 Attention 对于提高准确率和 F_1 仍然是有效的,在模型上添加 Attention 层,准确率提高,是因为通过赋予不同的注意力权重,模型能够更有效地利用文本中的语义信息。可见在大多数情况下,具有 Attention 的模型在评估准确率上表现更好,并且在使用数据增强技术后的 ERNIE-BiGRU 模型上添加 Attention 层, F_1 提升明显,表现出色。

预训练模型在使用不同分类模型上的效果如表 6 所示。预训练模型 ERNIE 的 F_1 值和准确率较高,TextCNN 模型结果差,并且预训练模型与 CNN 结合运行结果较差,而结合了预训练模型的 BiGRU 效果最佳,因为预训练模型可以有效提取文本情感信息,结果也表明预训练模型的加入确实可以提高模型评价能力,证明了预训练模型和 BiGRU 结合的可行性。

不同原始数据与增强后的数据在不同模型上的效果如表 7 所示。对于民航旅客评价数据匮乏,情感信息不足,数据增强技术可以充分利用现有数据并且增强文本情感表示。所有具有 EDA 数据增强的模型在评估准确率上表现更好。它们的 F_1 普遍高于 0.93,而没有 EDA 的模型 F_1 则在 0.9~0.93 波动。EDA-ERNIE-BiGRU-Attention 模型表现最佳,具有最高的评估准确率(0.9721)。结合 EDA 数据增强和多个模型架构(ERNIE、BiGRU、Attention)有更高的准确率和 F_1 值。这说明数据

增强可以有效提升模型的性能,尤其是在结合多种模型架构的情况下。

有注意力与无注意力在不同模型上的效果如表 8 所示。尽管 ERNIE-BiGRU 模型加入 Attention 机制后准确率提高, F_1 稍有下降,但 EDA-ERNIE-BiGRU-Attention 模型显示 Attention 对于提高准确率和 F_1 仍然是有效的,在模型上添加 Attention 层,通过赋予不同的注意力权重,模型能够更有效地利用文本中的语义信息。可见在大多数情况下,具有 Attention 机制的模型在评估准确率上表现更好。Attention 机制有助于模型更好地关注关键信息,提高了模型性能。

4 结论

从所提供的模型结果来看,可以明显观察到一个普遍的趋势:在添加数据增强、注意力机制以及结合不同的模型架构时,模型的性能通常会有所提升。这突显了模型的组合和创新在提高性能方面的重要性。在添加了 EDA 数据增强的情况下,大多数模型的 F_1 值都有所提升。这表明数据增强在增强模型的泛化能力和性能方面起着积极的作用,特别是在数据有限的情况下。之后加入注意力机制的模型在准确率方面表现更为出色,注意力在捕捉输入序列中不同部分的重要性方面具有优势,能够帮助模型更好地理解输入。从结果中可以看出,应用数据增强技术后 ERNIE-BiGRU-Attention 模型在评估准确率上表现最佳。这个模型结合了数据增强、ERNIE 模型、BiGRU 架构以及注意力机制,达到了较高的性能水平。在实际应用中,选择适合特定任务的模型需要权衡 F_1 值、准确率、模型复杂性等不同因素,还需要根据任务的目标和数据的特点来进行最终决策。

表 5 不同模型评估结果

模型	准确率	精确率	召回率	F_1
TextCNN	0.832 9	0.801 1	0.874 0	0.817 8
BiGRU	0.908 7	0.896 1	0.910 3	0.891 6
ERNIE	0.928 9	0.939 4	0.942 7	0.933 6
ERNIE-CNN	0.838 4	0.825 8	0.887 8	0.837 3
ERNIE-BiGRU	0.931 6	0.945 6	0.944 7	0.939 2
ERNIE-BiGRU-Attention	0.932 3	0.932 0	0.955 0	0.935 4
EDA-ERNIE-BiGRU-Attention	0.972 1	0.974 8	0.981 7	0.975 9

表 6 有无预训练模型结果对比

模型	F_1	模型	F_1
TextCNN	0.817 8	ERNIE-CNN	0.837 3
BiGRU	0.891 6	ERNIE-BiGRU	0.939 2
ERNIE	0.933 6		

表 7 有无数据增强模型结果对比

模型	F_1	模型	F_1
BiGRU	0.891 6	EDA-BiGRU	0.938 9
ERNIE	0.933 6	EDA-ERNIE	0.956 9
ERNIE-BiGRU	0.939 2	EDA-ERNIE-BiGRU	0.956 1
ERNIE-BiGRU-Attention	0.935 4	EDA-ERNIE-BiGRU-Attention	0.975 9

表 8 有无注意力机制模型结果对比

模型	F_1	模型	F_1
EDA-BiGRU	0.938 9	EDA-BiGRU-Attention	0.949 0
ERNIE-BiGRU	0.939 2	ERNIE-BiGRU-Attention	0.935 4
EDA-ERNIE-BiGRU	0.956 1	EDA-ERNIE-BiGRU-Attention	0.975 9

参考文献

- [1] 景云麒, 杨璐源. 我国智慧民航建设探讨[J]. 交通企业管理, 2022, 37(3): 4-7.
- [2] Wankhade M, Rao A C S, Kulkarni C. A survey on sentiment analysis methods, applications and challenges[J]. *Artif Intell Rev*, 2022, 55: 5731-5780.
- [3] 钟佳娃, 刘巍, 王思丽, 等. 文本情感分析方法及应用综述[J]. 数据分析与知识发现, 2021, 5(6): 1-13.
- [4] 李阳, 王石, 朱俊武, 等. 方面级情感分析综述[J]. 计算机科学, 2023, 50(S1): 34-40.
- [5] AHMED M, CHEN Q, LI Z. Constructing domain-dependent sentiment dictionary for sentiment analysis[J]. *Neural Comput & Applic*, 2020, 32: 14719-14732.
- [6] ALEXANDER P, PATRICK P. Twitter as a corpus for sentiment analysis and opinion mining[C]//LREc. Valletta, Malta; ELRA, 2010: 1320-1326.
- [7] ABHILASHA S R, AMIT A, PREETI D. Comparative study of machine learning approaches for Amazon reviews [J]. *Procedia Computer Science*, 2018, 132: 1552-1561.
- [8] AIN, QURA, TUL, et al. Sentiment analysis using deep learning techniques; a review[J]. *International Journal of Advanced Computer Science and Applications*, 2017(8): 386-399.
- [9] 周湘贞, 李帅, 隋栋. 基于深度学习和注意力机制的微博情感分析[J]. 南京师大学报, 2023, 46(2): 115-121.
- [10] 李海峰, 周壁刚. 基于 CNN-BiLSTM 融合多头自注意力机制的电商评价情感分析探究[J]. 科技与产业, 2024, 24(2): 273-281.
- [11] PETERS M E, Mark N, Mohit I, et al. Deep contextualized word representations[J]. arXiv preprint arXiv, 2018, 1802: 05365.
- [12] DEVLIN J, C M W, LEE K, et al. Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv, 2018, 1810: 04805.
- [13] WANG S H, SUN Y, XIANG Y. ERNIE 3.0 Titan: exploring larger-scale knowledge enhanced pre-training for language understanding and generation [J]. arXiv preprint arXiv, 2021, 2112: 12731.
- [14] 沈彬, 严馨, 周丽华, 等. 基于 ERNIE 和双重注意力机制的微博情感分析[J]. 云南大学学报(自然科学版), 2022, 44(3): 480-489.
- [15] SUN Y, WANG S H, LI Y K, et al. Ernie: enhanced representation through knowledge integration[J]. arXiv preprint arXiv, 2019, 1904: 09223.
- [16] İRSOY O, CLAIRE C. Opinion mining with deep recurrent neural networks[C]//EMNLP. Doha, Qatar: Association for Computational Linguistics, 2014: 720-728.
- [17] 胡荣磊, 芮璐, 齐筱, 等. 基于循环神经网络和注意力模型的文本情感分析[J]. 计算机应用研究, 2019, 36(11): 3282-3285.
- [18] CHO K H, CHUNG J Y, CAGLAR G, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv, 2014, 1412: 3555.
- [19] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[J]. arXiv preprint arXiv, 2017, 1711: 05101.
- [20] KINGMA D P, BA J. Adam: a method for stochastic optimization[J]. arXiv preprint arXiv, 2014, 1412: 6980.
- [21] 孟天宇, 王欣, 许雅玺, 等. 一种基于 SKEP-ERNIE-BiGRU 网络模型结构的情感分类方法: CN115292485A [P]. 北京: 知识产权出版社, 2022.
- [22] WEI J, ZOUK. Eda: easy data augmentation techniques for boosting performance on text classification tasks[J]. arXiv preprint arXiv, 2019, 1523: 4300.

Sentiment Analysis of Civil Aviation Passenger Reviews Based on ERNIE-BiGRU and Attention Mechanism

XU Yaxi, LU Jianping

(School of Economics and Management, Civil Aviation Flight University of China, Guanghan 618307, Sichuan, China)

Abstract: In response to the insufficient information obtained from traditional text sentiment analysis, as well as the underutilization of emotional resources and other textual information, which leads to inadequate text information for determining the sentiment polarity of comments, an ERNIE-BiGRU-Attention model was proposed for sentiment classification of civil aviation passenger evaluations. Firstly, Easy Data Augmentation (EDA) was applied to enhance the dataset. Subsequently, based on Baidu's proposed pre-trained language model, Enhanced Representation through Knowledge Integration (ERNIE), emotional knowledge was extracted from the text. In terms of passenger evaluation feature extraction, Bi-directional Gate Recurrent Unit (BiGRU) was utilized and an attention mechanism was incorporated to extract features from the input. Finally, The results indicate that the proposed model performs excellently in classification, achieving a binary classification accuracy of 97.21%, a comprehensive F_1 average of 0.975 9, and a 0.73% improvement in accuracy compared to the baseline model.

Keywords: sentiment analysis; BiGRU; data augmentation; attention mechanism