

基于 CNN-BiLSTM 融合多头自注意力机制的 电商评价情感分析

李海峰, 周壁刚

(大连交通大学经济管理学院, 辽宁 大连 116028)

摘要: 针对传统单一的基于机器学习的情感分析方法在特征提取以及语义理解方面效果不尽如人意的的问题。构建一种基于 CNN-BiLSTM 融合多头自注意力机制的电商评价情感分析模型, 能够更好地处理文本中的长距离依赖关系和捕捉情感信息的语义关系, 从而提高模型的鲁棒性和泛化能力, 进而提高商家对消费者评论的情感理解和评价准确性。基于一个中文电商公开数据集对模型进行了实验, 并将其与其他模型进行了比较。实验结果表明, 该模型的精确度、准确度、召回率和 F1 值等指标均优于其他模型。

关键词: 情感分析; 神经网络; 注意力机制; 深度学习

中图分类号: TP183; TP391.1 **文献标志码:** A **文章编号:** 1671-1807(2024)02-0273-09

2022 年, CNNIC 发布了第 50 次《中国互联网络发展状况统计报告》, 报告显示^[1], 截至 2022 年 6 月, 我国网民规模为 10.51 亿人, 互联网普及率达 74.4%; 较 2021 年 12 月新增网民 1 919 万人, 互联网普及率提升 1.4 个百分点。截至 2022 年底, 我国网络购物用户规模达 8.45 亿人, 较 2021 年 12 月增长 319 万人, 占网民整体的 79.2%。随着互联网和移动互联网的普及, 电子商务已成为日常生活中不可或缺的一部分, 人们越来越多地借助电商平台购买商品和服务, 同时越来越多的消费者也在购物过程中依赖于在线评价来做出决策。电商评价不仅为消费者提供了其他购买者的意见和反馈, 也成为电商企业了解用户满意度、改进产品和服务的重要依据。然而, 电商平台上的商品评价数量巨大, 其质量和可信度也参差不齐, 消费者需要花费大量的时间和精力才能获取到有价值的信息。因此, 情感分析技术在电商评价中的应用变得至关重要。

由于中文语言的特殊性, 情感分析面临着系列的挑战。首先, 文本的情感表达往往较为复杂, 同一个词语在不同的上下文中可能具有不同的情感倾向, 难以捕捉到词语之间的复杂语义关系, 限制了情感分析的性能。此外, 情感分析还受

到文本的主观性和领域特定性的影响, 传统方法往往难以适应不同领域或特定情感类别的分析需求。

近年来, 深度学习作为一种强大的机器学习方法, 已经在情感分析中得到了广泛的应用。例如, 递归神经网络(recursive neural network, RNN)和卷积神经网络(convolutional neural network, CNN)可以有效地处理序列数据和文本数据, 并在情感分析中取得了显著的性能提升。作为循环神经网络(recurrent neural network, RNN)的特殊形式, 双向长短时记忆网络(bidirectional long short-term memory network, BiLSTM)不但能够缓解 RNN 的梯度爆炸问题, 还能够捕捉到文本中的时序信息, 通过同时考虑过去和未来的上下文, 提供了更全面的语义理解。注意力机制(attention)能够帮助模型在处理文本时集中关注于重要的词语或短语。它可以给予不同的词语不同的权重, 使模型能够专注于与情感相关的关键信息。自注意力机制(self-attention)允许模型在输入序列的不同位置之间建立关联, 并根据每个位置的重要性来加权组合信息。相比传统的固定权重的方法, 自注意力机制能够动态地学习每个位置的相关性, 使模型能够更好地理解序列中不同元素之间的依赖关系。此外,

收稿日期: 2023-10-24

基金项目: 辽宁省教育厅基本科研项目(LJKMZ20220834)

作者简介: 李海峰(1979—), 男, 黑龙江牡丹江人, 教授, 博士, 硕士研究生导师, 研究方向为大数据管理、知识管理、系统工程; 周壁刚(1997—), 男, 安徽滁州人, 硕士研究生, 研究方向为情感分析、数据挖掘、深度学习。

预训练的深度学习模型,如 Word2vec(词向量模型)和基于 transformer 的双向编码器表示模型(bidirectional encoder representations from transformer, BERT),通过在大规模文本语料库上进行预训练,能够学习到更丰富的语义和上下文信息。基于此,本文所采取的 Word2vec、CNN、BiLSTM 和多头自注意力机制(multi-head-self-attention)相融合的模型,能够显著聚焦上下文依赖关系,提高情感分析的准确性。

情感分析又称意见挖掘^[2],指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。其作为一种自然语言处理技术,旨在自动识别和提取文本中表达的情感信息,将其分类为正面、负面或中性等。目前,情感分析技术已成为自然语言处理领域的重要研究方向。它被广泛应用于评价分析、舆情分析、品牌管理等方面,取得了一定的成果和效果。同时随着智能客服和智能助理的发展,情感分析在自然语言处理领域中的应用也越来越广泛。按照所处理文本的颗粒度不同,情感分析可以分为词语级、句子级及篇章级等几个方面;按照分析方法的不同,情感分析可以分为基于情感词典、基于机器学习和基于深度学习三个方面。基于情感词典的情感分析(指文本情感分析,下同),是指旨在通过使用情感词典(也称为情感词汇表或情感字典)来确定文本中的情感倾向。情感词典是一个包含单词或短语的词汇表,每个单词或短语都与情感类别(如积极、消极、中性等)相关联。在情感分析过程中,将文本与情感词典进行匹配,计算文本中出现积极和消极情感词的频率或得分,并根据这些分数来判断整体情感倾向。常见的方法包括计算文本的情感得分(如积极得分减去消极得分)或应用阈值来判断情感类别(如果积极得分大于消极得分,则判断为积极情感)。Esuli 和 Sebastiani^[3]提出了一种基于 SentiWordNet 的情感分析方法,它使用了 WordNet 词汇数据库和情感极性值来为单词分配情感极性。该方法可以根据词汇在 SentiWordNet 中的情感极性值计算文本的整体情感得分。Hutto 和 Gilber^[4]提出了效价感知词典与情感推理机(valence aware dictionary and sentiment reasoner, VADER)模型,它是一个基于规则的情感分析模型,使用一个预先构建的情感词典和一组规则来推断文本的情感极性。周咏梅等^[5]提出了构建基于 HowNet 和 SentiWordNet 的中文情感词典方法。将词语自动分解为多个义元后计算其情感倾

向强度,并且使用词典校对方法对词语情感倾向强度进行优化。杜嘉忠等^[6]提出构建特征-情感词,在对专用情感词和通用情感词进行分类管理,在特定领域取得了很好的效果。周哲和商琳^[7]提出了一种基于动态词典的文本分类算法,根据训练集的不同生成不同的动态情感词典以此来对特征进行提取,并融入三支决策思想辅助情感分析。

在机器学习研究方面,Pang 等^[8]首先利用朴素贝叶斯,支持向量机和最大熵三种机器学习算法对文本情感倾向性进行分析,得出了支持向量机分类效果最好的结论;Ye 等^[9]使用了朴素贝叶斯(Naive Bayes, NB)、支持向量机(support vector machine, SVM)以及加入特征的 N-gram 技术进行了文本分类训练,并取得了不错的效果;Mullen 和 Malouf^[10]研究使用朴素贝叶斯分类器对推特文本进行情感分析,并尝试预测股市的涨跌情况。结果表明,情感分析对于预测股市趋势具有一定的准确性;Wang 和 Manning^[11]研究使用词袋模型结合 NB 分类器和逻辑回归(logistic regression, LR)进行情感分类和主题分类。结果表明,这种简单而有效的方法在情感分析任务中取得了良好的表现。

深度学习作为近年来新兴的方法,在多个领域取得了突破性的应用成果。深度学习能够自动从原始数据中学习到高层次的抽象特征表示,通过多层次的非线性变换来捕捉数据中的复杂关系,从而提高模型的表达能力和预测性能。在情感分析领域,梁斌等^[12]提出一种基于多注意力卷积神经网络(multi-attention convolution neural networks, MATT-CNN)的特定目标情感分析方法,可以获取更深层次的情感特征信息,有效识别不同目标的情感极性。程艳等^[13]提出一种基于注意力机制的多通道 CNN 和双向门控循环单元(MC-AttCNN-Att-BiGRU)的神经网络模型。该模型能够通过注意力机制关注到句子中对情感极性分类重要的词语,提高了模型的文本特征提取能力。曾诚等^[14]提出一种结合 ALBERT 预训练语言模型与卷积神经网络(convolutional recurrent neural network, CRNN)的弹幕文本情感分析模型 ALBERT-CRNN,可以有效判断文本极性。许雪晨和田侃^[15]提出一种基于金融文本情感分析的指数预测模型 SA-BERT-LSTM,能够有效提高模型预测的准确率。徐绪堪和周泽聿^[16]构建了一种基于 BiLSTM-CNN 模型的微信推文情感分类模型,能够很好地获取上下文信息,从而提高模型准确率。

从过往的研究可以看出,依赖人工构建的情感词典往往受限于标注词典的数量和准确度,并且构建情感词典是一个复杂的任务,往往需要大量的人力和时间。此外,情感词典可能无法涵盖新出现的词语或特定领域的术语,导致对于这些词语的情感分析准确性下降;而传统的机器学习算法,往往将文本视为独立的特征,忽略了词语之间的上下文关系,然而,上下文信息对于情感分析至关重要,因为同一个词语在不同的上下文中可能具有不同的情感倾向。传统方法难以捕捉到这种复杂的语义关系,并且处理长文本时困难重重。

1 CNN-BiLSTM 融合多头自注意力机制的情感分析模型

针对上述问题,本文构建所提出的模型,能够对电商评价文本中的情感进行情感倾向分类。首先,利用 Word2Vec 训练词向量,建立索引-文本映射字典,利用预训练 Word2vec 向量来构建字典集对应的 Word2vec 向量,将原始文本(包括标签和文本)里的每个词转为对应的索引数字;然后,将经过处理的向量通过 CNN 层,利用卷积层和池化层提取文本中的局部特征,接着使用 BiLSTM 层建模上下文联系,然后通过自注意力机制捕捉整个句子全局关系,设置多个注意力头捕捉更丰富的语义信息。最后,将得到的表示传入全连接层进行情感分类,使用 softmax 函数将分类结果映射到 $(0, 1)$ 区间,最终输出对应的情感类别或情感得分。模型总体结构如图 1 所示。

1.1 词嵌入模型

词嵌入模型通常使用向量表示来表示单词或词语,其中最常用的模型是 Word2Vec。最早由 Tomas 等^[17-18]提出并提供了详细的模型描述和实验结果,Word2Vec 模型有两种实现方法:连续词袋模型(continuous bag-of-words, CBOW)和跳词模型(skip-gram)。CBOW 模型的目标是通过上下文

单词来预测目标单词。它的基本思想是在给定上下文单词的情况下,预测目标单词。skip-gram 模型则恰恰相反,利用中心词来预测上下文单词。本文采用的模型为 skip-gram。

假设目标单词为 w_0 ,使用一个 One-Hot 编码向量来表示,记为 x_0 。 x_0 是一个长度为 V 的向量,其中 V 是词汇表的大小。 x_0 中只有第 w_0 个元素为 1,其余元素为 0。

假设上下文单词为 w_1, w_2, \dots, w_n ,对于每个上下文单词,同样使用 One-Hot 编码向量表示,记为 x_1, x_2, \dots, x_n 。下面给出其公式表达式为

$$h_0 = W^T x_0 \quad (1)$$

$$y_i = \text{softmax}(U^T h_0) \quad (2)$$

$$L = - \sum (t_i \ln y_i) \quad (3)$$

矩阵 W 和 U 分别为隐藏层和输出层的参数; h_0 为目标单词的隐藏层表示向量;softmax 为一个分类器; y_i 和 t_i 分别为上下文单词 w_i 的预测概率向量和真实标签向量; L 为损失函数,用来衡量模型的预测结果与真实上下文单词的差异。模型的参数 W 和 U 可以通过随机梯度下降法(stochastic gradient descent,SGD)或其他优化算法来更新。

1.2 CNN-BiLSTM 融合多头自注意力机制模型

1.2.1 卷积神经网络

CNN 是一种常用于处理具有网格结构数据(例如图像、语音、文本)的深度学习模型。CNN 广泛应用于计算机视觉任务,如图像分类、目标检测和图像生成。随着深度学习的快速发展,CNN 也被应用于自然语言处理领域,而 TextCNN 则是 CNN 在文本分类任务中的应用。它是一种用于文本分类和文本相关任务的卷积神经网络模型。通过卷积操作和池化操作来捕捉文本中的局部特征。

(1)卷积层。卷积层在神经网络中具有重要作用,主要用于实现特征的初步提取。在图像领域,常使用正方形卷积核对图像矩阵进行上下左右四

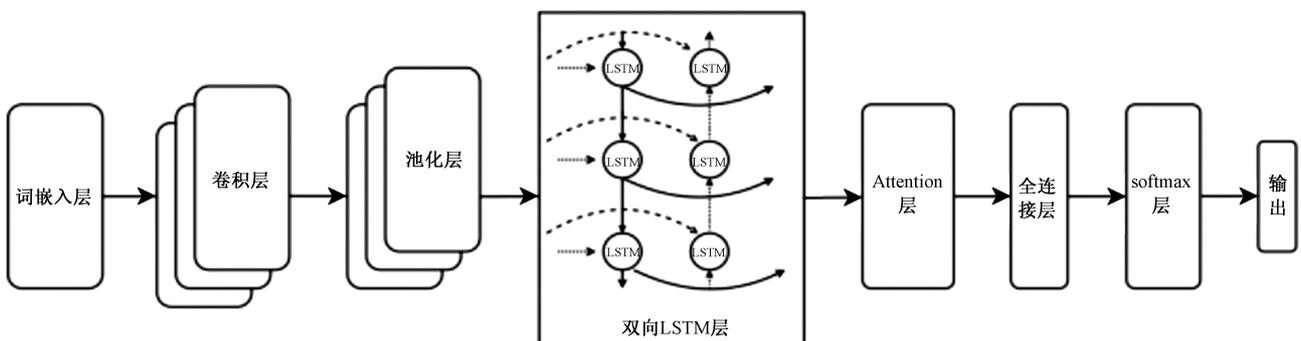


图 1 模型总体结构

个方向的扫描。同传统 CNN 不同的是,在文本领域,为了保留词向量作为文本表示的最小单元,文本的卷积核的宽度通常设置为词向量的维度。因此,文本的卷积操作是在文本行的方向进行的,即沿着语序顺序进行从上往下的滑动。这样可以通过卷积运算捕捉到文本中的局部特征和上下文关系,为后续的任务提供更丰富的信息。

本模型将文本数据表示为词向量矩阵,每个词对应一个词向量。这里使用上文所说的预训练的词向量模型(Word2Vec)得到。假设文本数据有 L 个词,每个词的词向量维度为 D ,则输入数据表示为一个 $L \times D$ 的矩阵。

类似于图像数据,文本数据也可以视为具有网格结构,其中每个词向量对应一个位置。本文中使用了多个不同大小的卷积核对文本进行卷积操作,以捕捉不同长度短语的局部特征。每个卷积核的宽度为 K ,即卷积窗口的尺寸为 $K \times D$,假设使用 N 个不同大小的卷积核,分别为 K_1, K_2, \dots, K_N 。对于每个卷积核,进行卷积操作后得到的特征图大小为 $L - K_i + 1$ 。

对于第 K_i 个卷积核,卷积操作可以表示为

$$Z_i = \text{ReLU}(W_i \otimes X_{i:i+n-1} + b_i) \quad (4)$$

式中: $Z_i \in \mathbb{R}^{(L-K_i+1)}$; $L - K_i + 1$ 为第 i 个卷积核的输出特征图; $W_i \in \mathbb{R}^{(K_i \times D)}$; $K_i \times D$ 为第 i 个卷积核的权重矩阵; \otimes 为卷积操作; $i:i+n-1$ 为句子向量的第 i 到第 $i+n-1$ 个词, $b_i \in \mathbb{R}$ 为偏置项;ReLU 为非线性激活函数,其定义如下:

$$f(x) = \max(0, x) \quad (5)$$

式中: x 为输入值; $f(x)$ 为输出值;ReLU 函数的作

用是将负数输入映射为 0,而将非负数输入保持不变。相比于其他激活函数,如 Sigmoid 和 Tanh 函数,ReLU 函数的计算速度更快。并且可以有效缓解梯度消失问题,使得反向传播时能够更好地传递梯度。

(2)池化层。由于经过卷积操作后所得到的特征图的维度和复杂度依旧较高,所以在进行完卷积操作后需要进行池化操作。池化层的作用是对输入数据进行降采样,从而减少模型的参数数量,降低计算复杂度,并提取输入数据的关键特征。池化操作一般使用最大池化(max pooling)或平均池化操作(average pooling)。本文模型使用最大池化。最大池化的原理是在输入数据的每个子区域中选择最大的特征值作为输出。具体而言,最大池化会将输入数据分割成不重叠的区域(通常是矩形区域),然后在每个区域中选择最大的特征值。这样就可以减少数据的维度,保留最显著的特征。

最大池化的公式可以表示为

$$\text{maxPooling}(X_{ij}) = \max(X_{ij}[p, q]) \quad (6)$$

式中: X_{ij} 为输入数据的第 i 行、第 j 列的子区域; $X_{ij}[p, q]$ 为子区域中的第 p 行、第 q 列的元素。

1.2.2 双向长短期记忆网络

长短期记忆网络(LSTM)是一种特殊的循环神经网络(RNN),通过使用门控机制来解决 RNN 在处理长期依赖问题时的梯度消失和梯度爆炸等问题。LSTM 包含一个细胞状态(cell state)和输入门(input gate)、遗忘门(forget gate)和输出门(output gate)三个门(图 3)。

在 BiLSTM 中,将输入序列按照时间步骤拆分

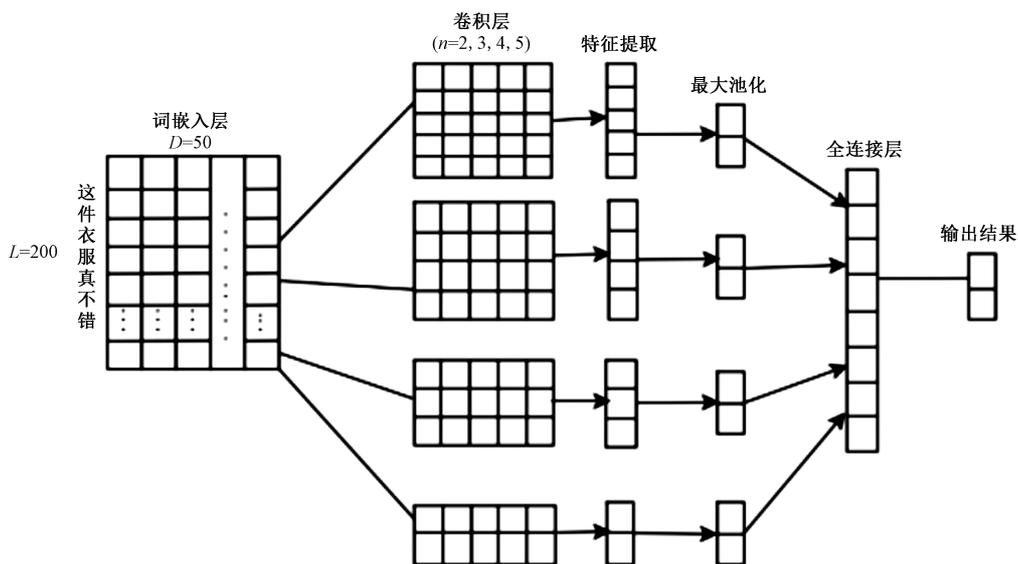


图 2 TextCNN 原理

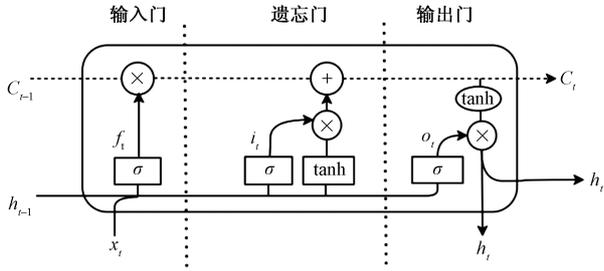


图 3 LSTM 结构

为正向和反向两个方向,并在每个方向上构建一个 LSTM 单元。这样可以同时考虑上下文信息,提高模型的性能。在本模型中把 TextCNN 的输出特征拼接在一起,形成一个更长的特征向量,然后将其传递给 BiLSTM 层。BiLSTM 模型的结构如下公式表示。

$$i_t = \sigma(\mathbf{W}_{x_i}x_t + \mathbf{W}_{h_i}h_{t-1} + b_i) \quad (7)$$

$$f_t = \sigma(\mathbf{W}_{x_f}x_t + \mathbf{W}_{h_f}h_{t-1} + b_f) \quad (8)$$

$$c_t = f_t \odot c_{t-1} \odot \tanh(\mathbf{W}_{x_c}x_t + \mathbf{W}_{h_c}h_{t-1} + b_c) \quad (9)$$

$$o_t = \sigma(\mathbf{W}_{x_o}x_t + \mathbf{W}_{h_o}h_{t-1} + b_o) \quad (10)$$

$$h_t = o_t \odot \tanh(c_t) \quad (11)$$

式中: x_t 为当前时间步的输入; h_{t-1} 为上一个时间步的隐藏状态; \mathbf{W}_{x_i} 、 \mathbf{W}_{x_f} 、 \mathbf{W}_{x_c} 、 \mathbf{W}_{x_o} 和 \mathbf{W}_{h_i} 、 \mathbf{W}_{h_f} 、 \mathbf{W}_{h_c} 、 \mathbf{W}_{h_o} 分别为输入/输出和隐藏状态的权重矩阵; i_t 、 f_t 、 o_t 分别为输入门、遗忘门、输出门的输出; c_t 为细胞状态的更新; h_t 为隐藏状态的更新; \odot 为逐元素相乘; \tanh 和 σ 为激活函数; b_i 、 b_f 、 b_c 、 b_o 为偏置项。

1.2.3 自注意力机制

注意力机制最早由 Bahdanau^[19]提出,通过学习对图像中不同区域的注意力权重,将视觉注意力引入到图像描述生成模型中,从而提升了模型的质量和准确性。后注意力机制被广泛应用于自然语言处理领域,考虑到传统的序列到序列模型(如基于循环神经网络的模型)在生成每个输出时只能依赖于固定长度的上下文信息。注意力机制允许模型根据输入的不同部分自动选择相关的上下文信息,以便更好地生成每个输出。并且通过动态计算注意力权重来适应不同的输入情况。这种机制使得模型能够处理更长的输入序列,可以更好地捕捉序列中的依赖关系。自注意力机制在传统注意力机制的基础上进行了进一步改良,通过并行计算,同时计算一句话中所有元素之间的相似性得分,从而获取全局的信息而非单一上下文的信息,这使得自注意力机制能够更全面地理解整个序列的语义,并更好地捕捉元素之间的复杂关系,使用谷歌

Transformer 自注意力机制,其相关公式如下所示。

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_q \quad (12)$$

$$\mathbf{K} = \mathbf{X}\mathbf{W}_k \quad (13)$$

$$\mathbf{V} = \mathbf{X}\mathbf{W}_v \quad (14)$$

$$\text{attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (15)$$

式中: \mathbf{W}_q 、 \mathbf{W}_k 、 \mathbf{W}_v 为线性变化矩阵; \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 分别为查询向量(query)、键向量(key)和值向量(value); d_k 为查询向量和键向量的维度。最后,将注意力权重和值向量进行加权求和,得到自注意力机制的输出。

先求出 $\mathbf{Q}\mathbf{K}^T$ 也就是注意力矩阵,注意力矩阵被用来给 \mathbf{V} 加权,功能是进行标准正态化,目的是防止出现梯度问题。

在下式运算中,每个批次里面的句子长度不一样,对于短的句子按照最长的句子自动补0。softmax的计算公式如下。

$$\sigma(Z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (16)$$

当 z 和 j 任意一个为0时,分母都有意义,当无用处用mask代替,不需要参与计算过程,从而加快整体的速度。这一过程使得softmax计算结果基本为0,避免了无意义情形。

加权过后得到 \mathbf{V} ,即 $\text{attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$,训练不是直接反传回最初始层,而是做残差连接后再反传:

$$\mathbf{X} + \text{SubLayer}(\mathbf{X}) \quad (17)$$

归一化后求均值和方差:

$$\mu_i = \frac{1}{m} \sum_{j=1}^m x_{ij} \quad (18)$$

$$\sigma_i^2 = \frac{1}{m} \sum_{j=1}^m (x_{ij} - \mu_i)^2 \quad (19)$$

求完均值和方差后再进行归一化操作运算

$$\text{LayerNorm}(\mathbf{X}) = \alpha \frac{x_{ij} - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} \quad (20)$$

多头自注意力机制通过引入多个注意力头(attention head)。每个注意力头都是一个独立的自注意力机制,它学习到一组不同的权重和表示。在多头自注意力机制中,输入序列首先通过线性变换映射到多个不同的query、key和value空间。每个注意力头都会对这些映射后的查询、键和值进行独立的注意力计算,得到每个位置的表示。最后,将模型中每个注意力头的表示通过线性变换和拼接操作来合并,就得到最终的输出表示(图4)。

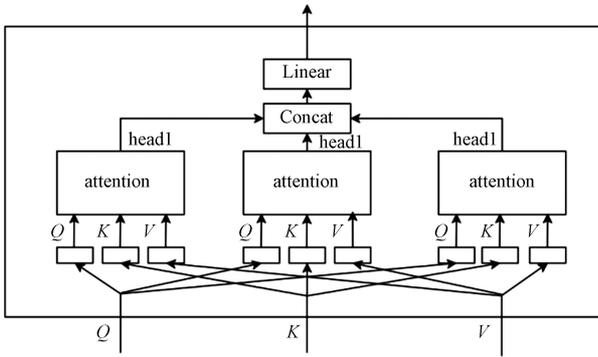


图 4 多头自注意力机制结构

2 模型实验

2.1 数据集介绍

数据集选自公开电商数据集“online_shopping_10_cats”，数据类别共有 10 种，数据量超过 6 万条，具体的数据格式如表 1 所示。

2.2 参数设置

利用 Pytorch 深度学习框架搭建模型，开发语言为 Python，优化器选择 Adam，损失函数为交叉熵损失函数 (cross entropy loss)，CNN 层卷积核的数量 (kernel_sizes) 设置为 [2, 3, 4, 5]，用于提取不同文本的有效特征。模型默认参数值如表 2 所示。

针对数据集，采用 Python 中的 sklearn 库按照 8:1:1 的比例将数据集划分为训练集，测试集和验证集，并统计了数据集中的句子长度，如图 5 所示。

表 1 数据类型

数据类别	总体数量	正例数量	负例数量
书籍	3 851	2 100	1 751
平板	10 000	5 000	5 000
手机	2 323	1 165	1 158
水果	10 000	5 000	5 000
洗发水	10 000	5 000	5 000
热水器	575	475	100
牛奶	2 033	992	1 041
衣服	10 000	5 000	5 000
计算机	3 992	1 996	1 996
酒店	10 000	5 000	5 000
总数	62 774	31 728	31 046

表 2 默认参数

kernel_sizes(卷积核宽度)	[2,3,4,5]
embedding_dim(词向量维度)	50
n_epoch(训练迭代次数)	50
Lr(学习率)	0.000 1
batch_size(批处理尺寸)	64
num_filters	48
hidden_dim	64
dropout	0.3
num_head	4

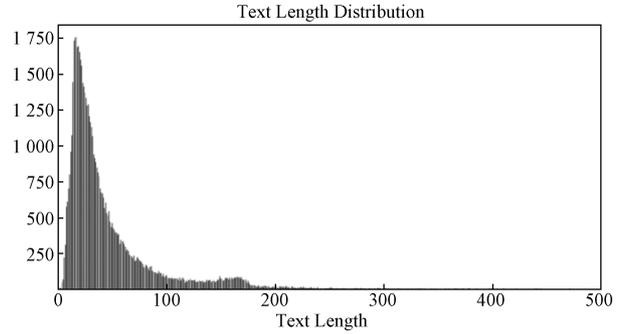


图 5 句子长度统计

根据图 5 统计，将句子最大长度 (max_sen_len) 设置为 200 h，可以有效获取数据集中数据信息。

2.3 评价指标

混淆矩阵 (confusion matrix) 是评估分类模型性能的常用工具，它以实际类别和预测类别为基础，将样本进行分类统计。混淆矩阵可以提供更详细的分类结果信息，从而计算得出准确率、精确率、召回率和 F_1 值等评价指标。假设有两个类别：正例 (positive) 和负例 (negative)。混淆矩阵是一个 2×2 的矩阵，如表 3 所示。

准确率 (accuracy) 是分类模型预测正确的样本数量与总样本数量的比例，其公式为：

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

精确率 (precision) 指分类模型预测为正例的样本中，真正为正例的样本数量与预测为正例的样本数量之比，公式表示为

$$precision = \frac{TP}{TP + FP} \quad (22)$$

召回率 (recall) 指分类模型正确预测为正例的样本数量与真实正例样本数量之比。公式为

$$recall = \frac{TP}{TP + FN} \quad (23)$$

F_1 值 (F_1 -score) 是精确率和召回率的调和平均值，用于综合考虑模型的准确性和召回性能。公式表示为

$$F_1 = \frac{2 \times (precision \times recall)}{precision + recall} \quad (24)$$

表 3 混淆矩阵

真实结果	预测结果	
	正例	反例
正例	TP	FN
反例	FP	TN

2.4 实验结果

为了验证不同超参数对实验结果产生的最终

影响,对于卷积核数量(num_filters)(16,32,48,64,80),隐藏层节点数(hidden_dim)(16,32,64,128,256),dropout 率(0.1,0.2,0.3,0.4,0.5),注意力头数(num_heads)(1,2,4,8,16)分别予以调整以获取最优模型的参数。结果如图 6~图 9 所示。

从实验结果可以看出,当卷积核数量较少时,可能导致模型欠拟合,无法捕捉数据中的重要特征。而较多的卷积核数量可以增加模型的特征表达能力,每个卷积核都可以学习不同的特征;隐藏层节点较多时显著增加了模型的性能,但同时也增

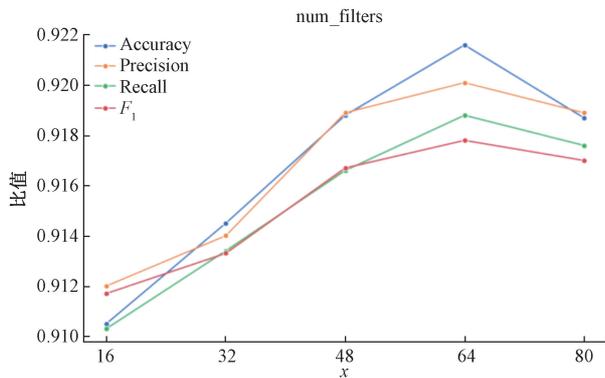


图 6 卷积核数量影响

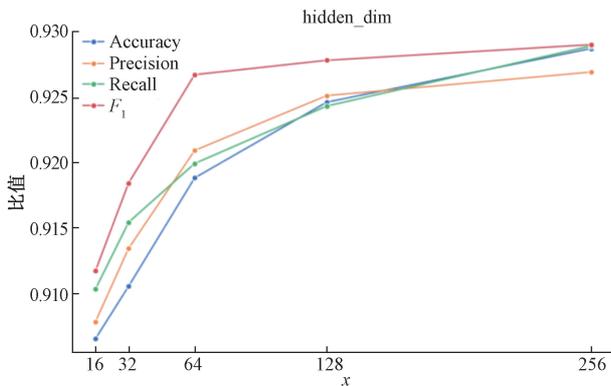


图 7 隐藏层节点影响

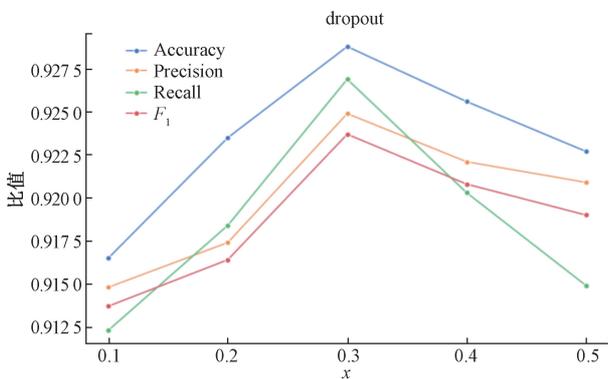


图 8 dropout 率影响

加了过拟合风险,而较少的隐藏层节点数可以加快模型的训练速度和收敛,减少了计算的复杂性;dropout 率指训练过程中随机地将每个隐藏层的输出置为零(即将其激活值乘以零),从而减少神经网络的复杂性和冗余性。较大的 dropout 率可以减少模型的复杂度,但也会导致信息损失过多,而 dropout 率较低时,可以有效提取模型特征,但也会过度拟合训练集导致性能下降;多个注意力头可以提高模型的计算效率以对输入序列进行更全面的建模,而较少的注意力头数可能导致模型对输入序列中的变化和噪声不够敏感,从而降低模型的泛化能力。

为了更进一步验证 CNN-BiLSTM-Muti-Head-self-Attention 模型的性能,在同一数据集上使用了其他几种深度学习算法作对比实验,实验结果如表 4 所示。

从实验结果可以看出,BiLSTM 相较于 CNN, accuracy、precision、recall、F₁ 四项指标分别提升了 1.445%、1.665%、1.660%、0.853%,这归功于 BiLSTM 可以从正向和反向两个方向处理输入序列,结合过去和未来状态的信息,从而使得模型能够全面理解上下文;而在引入注意力机制后,BiLSTM-Att 相较于单 BiLSTM 模型,在各项指标方面也有显著提高,这主要因为注意力机制有选择性

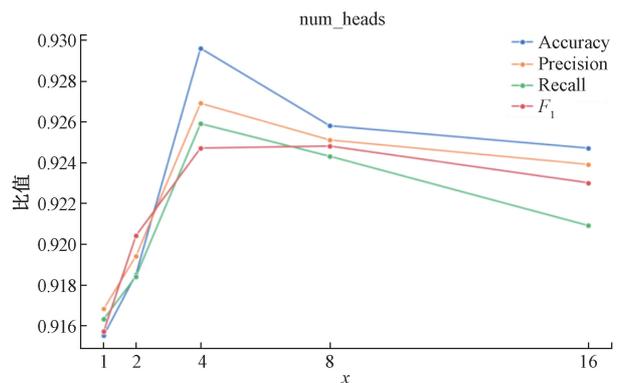


图 9 注意力头数影响

表 4 实验结果

变量	accuracy/ %	precision/ %	recall/ %	F ₁
CNN	87.181	87.098	85.102	86.833
BiLSTM	88.596	88.773	86.662	87.686
BiLSTM-Att	89.889	89.981	90.984	90.072
BiLSTM-self-Att	90.033	89.989	90.889	90.768
CNN-BiLSTM-Att	91.209	91.022	91.365	91.132
CNN-BiLSTM-self-Att	91.878	91.784	91.864	91.805
CNN-BiLSTM-Muti-Head-self-Att	92.665	92.886	93.012	93.003

地关注输入中的特定部分,将更多的注意力集中在与当前任务相关的重要信息上;通过结合 CNN、BiLSTM 的优点,并先后引入注意力机制,自注意力机制,多头自注意力机制,可以看出融合了多头自注意力机制的 CNN-BiLSTM 模型在 accuracy、precision、recall、 F_1 指标方面均优于其他模型,引入多个注意力头部,可以让模型学习到更多不同的关注权重和表示空间,从而增强了模型的表达能力。通过多头的并行计算,模型可以从多个视角综合得到更全面和丰富的表示。

为了验证模型的泛化能力,在测试集上再次进行了实验,实验结果如表 5 所示,可以看出模型在测试集上依然具有较好的表现,这说明对于超参数进行了合理地选择和调整,并充分学习了样本的特征,模型既能够适应训练数据的复杂性,又能够泛化到未见过的测试数据。

表 5 测试集实验结果

变量	accuracy	precision	recall	F_1
CNN-BiLSTM-Multi-Head-self-Att	91.481	91.874	92.065	92.233

3 结语

本文在电商评价情感分析任务中引入多头自注意力机制的效果,结合卷积神经网络(CNN)和双向长短期记忆网络(BiLSTM),并利用多头自注意力机制来增强模型对输入序列的关注能力。通过实验验证和分析,得出了以下结论。

首先,相较于传统深度学习方法,本文构建的模型具有显著的性能提升。通过引入多个注意力头部,模型能够从不同的视角关注评价文本中的不同部分,捕捉更丰富的语义信息。这种多头关注机制使得模型在情感分类任务中取得了更好的准确性和泛化能力,并且通过并行运算,大大减少了训练时间,提高了实验效率

其次,动态调整了多个超参数,以提高模型性能,并比较了引入多头自注意力机制的 CNN-BiLSTM 模型与其他经典模型的性能差异。结果显示,本文构建模型在电商评价情感分析任务中取得了优异的结果,并且超过了其他模型的表现。

最后,将模型在测试集进行测试以进一步验证,在测试集上本文构建的模型依然表现优异,这说明该模型具有很强的泛化能力和鲁棒性。

综上所述,在进行电商评价情感分析的任务中,引入多头自注意力机制的 CNN-BiLSTM 模型

具有显著的优势。这一研究为情感分析领域的研究和实践提供了新的启示,但同时,该模型也具有不足之处,比如 Word2vec 预训练模型在处理一词多义的问题上效果不是很好,以及调整过多的超参可能会使模型过拟合的可能性加大,这为进一步改进和扩展模型提供了有益的参考。

参考文献

- [1] “数字”点亮美好生活——透视第 50 次《中国互联网络发展状况统计报告》[J]. 信息系统工程, 2022 (10): 4-5.
- [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] ESULI A, SEBASTIANI F. SentiWordNet: a publicly available lexical resource for opinion mining [C]//Proceeding of LREC, 2006: 417-422.
- [4] HUTTO C J, GILBER E. VADER: a parsimonious rule-based model for sentiment analysis of social media text [C]//In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. ICWSM, 2014: 216-225.
- [5] 周咏梅, 杨佳能, 阳爱民. 面向文本情感分析的中文情感词典构建方法[J]. 山东大学学报(工学版), 2013, 43 (6): 27-33.
- [6] 杜嘉忠, 徐健, 刘颖. 网络商品评论的特征-情感词本体构建与情感分析方法研究[J]. 现代图书情报技术, 2014 (5): 74-82.
- [7] 周哲, 商琳. 一种基于动态词典和三支决策的情感分析方法[J]. 山东大学学报(工学版), 2015, 45(1): 19-23.
- [8] PANG B, LEE L L, SHIVAKUMAR V. Thumbs up? sentiment classification using machine learning techniques [J]. Association for Computational Linguistics, 2002, 10: 79-86.
- [9] YE Q, ZHANG Z Q, ROB L. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches[J]. Expert Systems With Applications, 2008, 36(3): 6527-6533.
- [10] MULLEN T, MALOUF R. Sentiment analysis of tweets for the prediction of stock market movement. Proceedings of the Workshop on Language in Social Media (LSM), 2013: 9-17.
- [11] WANG S, MANNING CD. Baselines and bigrams: Simple, good sentiment and topic classification [C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL), 2012: 90-94.
- [12] 梁斌, 刘全, 徐进, 等. 基于多注意力卷积神经网络的特定目标情感分析[J]. 计算机研究与发展, 2017(8): 1724-1735.
- [13] 程艳, 尧磊波, 张光河, 等. 基于注意力机制的多通道 CNN 和 BiGRU 的文本情感倾向性分析[J]. 计算机研究与发展, 2020(12): 2583-2595.
- [14] 曾诚, 温超东, 孙瑜敏, 等. 基于 ALBERT-CRNN 的弹

- 幕文本情感分析[J]. 郑州大学学报(理学版), 2021(3): 1-8.
- [15] 许雪晨, 田侃. 一种基于金融文本情感分析的股票指数预测新方法[J]. 数量经济技术经济研究, 2021, 38(12): 124-145.
- [16] 徐绪堪, 周泽聿. 基于多尺度 BiLSTM-CNN 的微信推文的情感分类模型及应用研究[J]. 情报科学, 2021, 39(5): 130-137.
- [17] TOMAS M, ILYA S, CHEN K, et al. Distributed representations of words and phrases and their compositionality[J]. *Advances in Neural Information Processing Systems*, 2013, 26(2): 3111-3119.
- [18] TOMAS M, ILYA S, CHEN K, et al. Efficient estimation of word representations in vector space[J]. *arXiv preprint arXiv: 1301.3781*, 2013, 3(1): 1-12.
- [19] Dzmitry B, Cho K, Bengio Y H. Neural machine translation by jointly learning to align and translate[J]. *arXiv preprint arXiv: 1409.0473*, 2014, 4(12): 1-15.

Research on Emotional Analysis of E-commerce Evaluation Based on CNN-BiLSTM Fusion Multi-Head-Self-Attention Mechanism

LI Haifeng, ZHOU Bigang

(School of Economics and Management, Dalian Jiaotong University, Dalian 116028, Liaoning, China)

Abstract: In response to the shortcomings of traditional single-machine learning-based sentiment analysis methods in feature extraction and semantic understanding, a novel e-commerce sentiment analysis model was developed. This model integrated a combination of CNN-BiLSTM and multi-head self-attention mechanisms, aiming to better address long-distance dependencies in the text and captured the semantic relationships of emotional information. This enhanced the model's robustness and generalization capabilities, consequently improving merchants' understanding of consumer sentiments and the accuracy of evaluations. Experiments were conducted on a publicly available Chinese e-commerce dataset, and the model was compared with other existing models. The experimental results indicate that this model outperforms others in terms of precision, accuracy, recall, and F1 score, among other metrics.

Keywords: emotional analysis; neural networks; attention mechanisms; deep learning