

文章编号: 2096-1618(2024)04-0397-07

# 基于 BERT 的多信息融合方面级情感分析模型

张孝峰, 陶宏才

(西南交通大学计算机与人工智能学院, 四川 成都 611756)

**摘要:** 方面级情感分析任务旨在判断文本语句不同方面的情感极性, 是自然语言处理领域的热点任务之一。当前基于 BERT 的方面级情感分析方法大多仅将其作为预训练词嵌入工具, 没有充分利用其本身的语义提取能力和任务处理能力, 在利用句法及方面词位置等信息时, 常忽略了不同外部信息的关联性。针对上述问题, 提出一种基于 BERT 的多信息融合网络模型。首先, 构建基于 BERT 词嵌入的主路径和辅路径, 辅路径可以避免对词嵌入特征产生干扰; 其次, 在辅路径下, 根据文本语句的句法依存树和多信息融合算法让不同词语与方面词的句法距离、位置距离等信息产生交互; 最后, 使用卷积神经网络和注意力机制将主路径与辅路径的特征融合。3 个数据集的实验结果表明, 提出的模型是有效的。

**关键词:** 方面级情感分析; 句法依存树; BERT; 多信息融合

**中图分类号:** TP391.1

**文献标志码:** A

**doi:** 10.16836/j.cnki.jcuit.2024.04.001

## 0 引言

随着大数据时代的到来, 网上每时每刻都在产生海量评论文本数据。通过文本情感分析技术, 可挖掘用户的整体评价与各个方面的情感倾向, 对评价对象展开细致分析以辅助决策。

按照研究对象的粒度来划分, 文本情感分析任务分为粗粒度和细粒度两类。其中, 粗粒度情感分析任务包含篇章级情感分析和句子级情感分析任务。

篇章级情感分析是对整个文档或篇章进行情感极性分类的任务。Tang 等<sup>[1]</sup>使用 CNN (convolutional neural network) 来获取用户和产品的关联语义特征, 将其作为全局信息来促进文档级的嵌入学习过程, 从而获得更好的文本表示。Yang 等<sup>[2]</sup>考虑到文档中可能存在不同层次的特征, 使用多层注意力网络来捕捉文档不同粒度的特征。但是, 当文档中不同句子涉及不同对象或不同情感的表述时, 整个文档所表达的情感信息并不一致, 这类情况更适合句子级情感分析。相对于文档, 句子文本更短, 包含的对象和情感信息更少, 句子级情感分析是对整个句子所表达的情感极性进行分类<sup>[3]</sup>。Wang 等<sup>[4]</sup>使用双层 CNN 框架, 先通过学习带噪声的标签数据实现输入数据去噪, 然后使用第二层 CNN 实现句子级情感极性分类, 从而降低噪声影响。但是, 当句子中存在多个实体属性, 且各实体具有不同的情感倾向时, 整个句子的情感倾向也是不一致的。因此, 无论是篇章级还是句子级情感分析都不

适用于评估此类文本。而更细粒度的方面级情感分析 (aspect-based sentiment analysis, ABSA) 可解决这一问题。

方面级情感分析指判断文本语句在不同方面的情感极性, 是自然语言处理的热点任务之一。本文针对方面级情感分析任务, 提出基于 BERT (bidirectional encoder representation from transformers) 的多信息融合网络模型 BERT-MIFN (BERT-based multi-information fusion network)。主要工作如下: (1) 提出一种多信息融合算法将文本语句的句法信息和方面词位置等信息交互融合; (2) 使用基于 BERT 的双路径结构来避免信息融合对词嵌入特征产生干扰; (3) 3 个公开数据集上的实验结果表明本文模型的有效性。

## 1 相关工作

当前, 已有许多学者就方面级情感分析任务展开研究。早期, 他们提出一系列基于情感词典的方法, 根据 SentiWordNet<sup>[5]</sup>、SenticNet<sup>[6]</sup>、NRC<sup>[7]</sup> 提前构建好的情感词典中不同词语或词组的情感极性实现情感判定。郁圣卫等<sup>[8]</sup>提出一种基于领域情感词典特征表示的细粒度意见挖掘方法, 使用领域情感词典在观点文本上构建特征表示并将其加入序列标注模型的输入部分, 有效利用情感词典资源。但基于情感词典的方法难以解决词形词性变化问题, 且扩充情感词典工作量大, 难度高。为提高情感分类的准确性, 采用基于机器学习的方法进行情感分类研究并在当时取得较好效果。但基于传统机器学习的方法仍然存在一定局限

性。于是,目前越来越多的研究者对基于深度学习的方法进行研究与改进,并取得更好效果。

Hochreiter 等<sup>[9]</sup>针对循环神经网络(recurrent neural network, RNN)存在梯度消失、梯度爆炸和长距离依赖等问题,提出长短期记忆网络(long short-term memory, LSTM)。Socher 等<sup>[10]</sup>将 RNN 应用于情感分析任务,并取得较好效果。Tang 等<sup>[11]</sup>考虑目标与上下文的语义关联,在 LSTM 基础上添加不同形式的方面表示来处理方面级情感分析任务。Kim 等<sup>[12]</sup>将卷积神经网络应用于情感分析任务中,证明卷积运算也可以有效捕获文本中的语义结构信息。Zhang 等<sup>[13]</sup>认为基于注意力机制和 RNN 的模型缺乏捕获文本中句法关系和长距离依赖的能力,而基于 CNN 的模型无法处理单词不连续的情况,因此提出 ASGCN (aspect-specific graph convolutional network)。该模型利用多层图卷积网络(graph convolutional network, GCN)来融合文本句法信息及词语位置信息。

Bahdanau 等<sup>[14]</sup>首次将注意力机制应用于自然语言处理任务后,基于注意力机制的方面级情感分析模型逐渐被提出。Wang 等<sup>[15]</sup>提出的 ATAE-LSTM (attention-based LSTM with aspect embedding)模型在输入层或隐藏层拼接额外的方面词向量,并通过注意力机制,对隐藏状态加权求和,以更好地表示方面词与上下文的语义交互及全局特征。Ma 等<sup>[16]</sup>认为之前的研究大都是针对方面词建模,忽略了上下文词的重要性,因此提出 IAN (interactive attention networks)。该模型对方面词和上下文词分别建模,并通过建立交互注意力机制让两者分别交互,从而能同时考虑到方面词的影响以及上下文与方面词的关联信息。针对之前使用注意力机制的方法仅使用粗粒度的注意力,没有考虑到方面词与上下文词每一个词之间的双向交互信息问题, Fan 等<sup>[17]</sup>提出 MGAN (multi-grained attention network)。该模型结合粗粒度和细粒度的注意力机制,构建了多粒度注意力网络,以捕获词级的方面词和上下文的交互信息。

在使用深度学习模型处理情感分析任务中,首先要对输入文本进行向量化表示。因此,词向量的好坏,直接决定模型提取语义信息的质量。初期,词嵌入表示多使用 Word2Vec<sup>[18]</sup>和 Glove<sup>[19]</sup>等。其中,预训练词向量 Glove 在 Word2Vec 的基础上,额外考虑了语料库中的先验信息,将局部上下文表示和全局语料信息结合在一起。之后,DevLin 等<sup>[20]</sup>提出基于 Transformers 的预训练语言表示模型 BERT。BERT 生成的词向量通过大量语料预先训练得到,具有丰富的语言知识和语义信息。仅需要一个额外的输出层,就可以对模型

进行微调。BERT 提出后,迅速被应用到方面级情感分析任务中。Song 等<sup>[21]</sup>提出 BERT-SPC (bert-based model for sentence pair classification),通过对 BERT 进行微调来处理方面级情感分析任务。Zeng 等<sup>[22]</sup>提出 LCF-BERT (local context focus BERT),分别对全局上下文和局部上下文进行建模,取得了更优效果。Phan 等<sup>[23]</sup>则在前者基础上改进,提出了 LCFS-ASC (local context focus on syntax ASC)。该模型通过句法依存树构建句法连接图,计算文本词与方面词的句法距离。

也有许多研究者通过设计针对性的网络结构来捕捉文本词位置、句法及词性等外部信息,以提高模型性能。Zhang 等<sup>[13]</sup>提出的 ASGCN 使用多层图卷积网络来融合语法、句法和位置特征。此外,还有研究者从句法剪枝等角度来改进模型。Wang 等<sup>[24]</sup>提出 R-GAT (relational graph attention network)。该模型对普通的句法依存树进行重塑和修剪,构建了一个新的以目标方面词为根节点的句法依存树结构,并提出一个关系图注意力网络来编码新的依存树。Zhang 等<sup>[25]</sup>提出了 SSEGCN (syntactic and semantic enhanced graph convolutional network)模型,提前构建文本词性信息、句法依存关系等外部知识,使用带有自注意力的方面感知注意力机制来获取文本的得分矩阵,在捕捉与方面相关语义信息的同时,学习句子的全局上下文信息。

上述研究中,结合词语位置和句法等信息对 BERT 微调来处理方面级情感分析任务,忽略了不同外部信息之间的交互;或是仅将 BERT 作为动态词嵌入工具,而没有充分利用其全局文本语义提取能力。因此,本文提出基于 BERT 的多信息融合网络 (BERT-MIFN)模型,在融合外部信息的同时考虑到不同信息的相互影响。

## 2 本文模型

### 2.1 模型框架

本文提出的方面级情感分析模型为基于 BERT 的多信息融合网络 BERT-MIFN,整体框架成双路径结构如图 1 所示。首先,使用 BERT 对文本进行向量化,在辅助路径中,利用多信息提取层实现文本词与方面词的句法距离、位置距离等信息的特征融合与交互,并生成辅助路径文本表示的权重;其次,在路径融合层使用卷积神经网络提取主路径与辅助路径文本表示的融合特征,并通过注意力层实现上下文语义交互;最后,经过池化层和全连接层,得到维度与任务情感极性类别数量相同的向量。

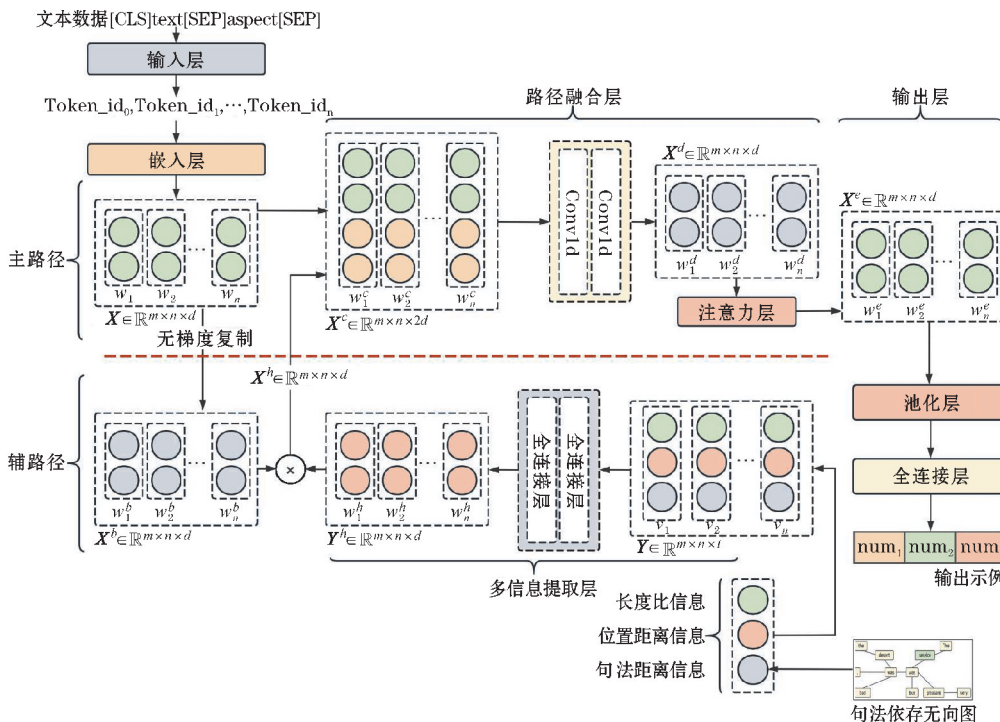


图1 BERT-MIFN 模型框架

## 2.2 输入层与嵌入层

输入层将文本数据分词并转换成对应的 Token\_id, 嵌入层的主要作用是通过 BERT 模型根据文本数据的 Token\_id 得到每个 Token 的嵌入向量, 从而将输入内容的表现形式由 Token\_id 转换为文本向量矩阵  $X \in \mathbb{R}^{m \times n \times d}$ 。其中  $m$  代表 Batch\_size,  $n$  代表文本长度。单条文本的嵌入向量为  $W = \{w_1, w_2, \dots, w_n\}$ ,  $w_j$  表示第  $j$  个 Token 对应的嵌入向量, 维度为  $d$ , 后文将以单条文本嵌入向量  $W$  为例展开阐述。

## 2.3 双路径结构

模型将内部结构分为主路径和辅助路径两部分, 将文本嵌入向量的处理过程隔开。主路径使用 BERT 模型转换的文本向量矩阵  $X$ , 单条文本的嵌入向量为  $W = \{w_1, w_2, \dots, w_n\}$ ; 对  $X$  进行无梯度复制得文本向量矩阵  $X^b$ , 单条文本的嵌入向量表示为  $W^b$ 。辅助路径在  $W^b$  上进行后续处理。 $W$  和  $W^b$  共享初始的词嵌入向量, 但不共享 BERT 权重, 从而避免对词嵌入特征产生干扰。

## 2.4 多信息提取层

在辅助路径中, 多信息提取层使用包含位置、句法等信息的矩阵  $Y \in \mathbb{R}^{m \times n \times t}$  作为输入, 其中  $m$  表示 Batch\_size,  $n$  表示文本长度,  $t$  表示信息维度。单条文本的词信息矩阵为  $V = [v_1, v_2, \dots, v_n]$ , 通过 2 个全连接层, 使不同外部信息交互并融合不同特征。

在单条文本中, 第  $i$  个词语的词信息向量为  $v_i = [v_p^i, v_s^i, v_r^i]$ , 其中  $v_p^i$  表示文本中第  $i$  个词语与方面词的位置距离信息;  $v_s^i$  表示文本中第  $i$  个词与方面词的句法距离信息;  $v_r^i$  表示实际文本长度  $l_{\text{text}}$  与总文本长度  $l_{\text{max}}$  的比值信息。

首先, 在计算不同词语与方面词的句法距离前, 须对文本进行句法分析, 生成句法依存树, 如图 2 所示。

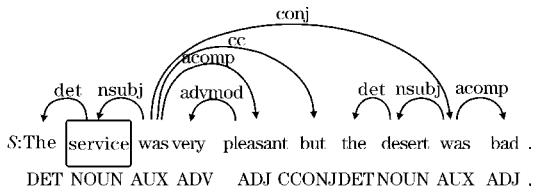


图2 文本句法依存树

根据句法依存树计算节点和边信息, 生成句法依存无向图, 如图 3 所示。

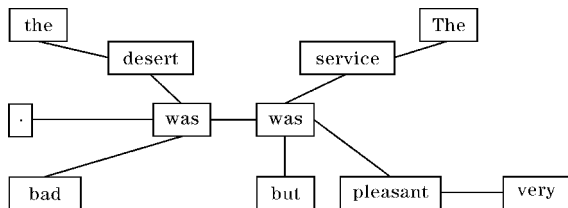


图3 文本句法依存无向图

其次, 计算文本词信息向量。不同词语与方面词之间的句法距离  $s_i$  表现在无向图中对应节点的最短路径, 如果无法连通, 则等于文本长度。不同词语与方面词的位置距离  $p_i$ , 即词语与方面词在句中所处位置



下标差值的绝对值。 $l_{\text{term}}$ 表示方面词长度, $p_{\text{term}}$ 表示方面词位置下标,不同词的词信息向量 $v_i$ 计算式如下:

$$v_i = \begin{cases} \left[ \frac{1}{l_{\text{text}}}, \frac{1}{l_{\text{text}}}, \frac{l_{\text{term}}}{l_{\text{max}}} \right], & i=0 \text{ 或 } i=l_{\text{text}}+l_{\text{term}}+2 \text{ 或 } i=l_{\text{text}}+1 \\ \left[ \frac{l_{\text{term}}}{\text{abs}(i+1-p_{\text{term}})}, \frac{l_{\text{term}}}{s_i+1}, \frac{l_{\text{term}}}{l_{\text{max}}} \right], & i>0 \text{ 且 } i<l_{\text{text}}+1 \\ \left[ 1, 1, \frac{l_{\text{term}}}{l_{\text{max}}} \right], & i>l_{\text{text}}+1 \text{ 且 } i \leq l_{\text{text}}+l_{\text{term}}+1 \\ [0, 0, 0], & i>l_{\text{text}}+l_{\text{term}}+2 \end{cases}$$

然后,将所有句子文本词信息矩阵作为多信息提取层的输入,通过2个全连接层,将 $t$ 个维度的词信息转化为1个维度,并将最后一个维度扩展为与词向量维度 $d$ 相同,得到新的文本词信息矩阵 $\mathbf{Y}^h \in \mathbb{R}^{m \times n \times d}$ ,计算过程如下:

$$\mathbf{Y}_a^h = \text{Linear}(\text{Relu}(\text{Linear}(\mathbf{Y})))$$

$$\mathbf{Y}^h = \text{Expand}(\mathbf{Y}_a^h, (1, d))$$

最后,将 $\mathbf{Y}^h$ 作为权重,与辅路径的文本向量矩阵 $\mathbf{X}^b$ 逐元素相乘,得到加权文本向量矩阵 $\mathbf{X}^h$ 。

## 2.5 路径融合层

路径融合层实质上由一个双层卷积神经网络和一个自注意力机制构成。首先,将辅助路径的加权文本向量矩阵 $\mathbf{X}^h$ 与主路径的文本向量矩阵 $\mathbf{X}$ 拼接,得到新的文本向量矩阵 $\mathbf{X}^c \in \mathbb{R}^{m \times n \times 2d}$ 。

其次,将其作为输入,经过双层卷积神经网络,提取细粒度的语义特征,得到文本向量矩阵 $\mathbf{X}^d \in \mathbb{R}^{m \times n \times d}$ ,计算如下:

$$\mathbf{X}^d = \text{CNN}(\text{Relu}(\text{CNN}(\mathbf{X}^c)))$$

最后,使用BERT的BertSelfAttention函数对 $\mathbf{X}^d$ 进行自注意力权重计算,返回加权后的文本向量矩阵

表1 数据集标记分布详情

Datasets	Positive		Negative		Neural	
	Train	Test	Train	Test	Train	Test
Restaurant	2164	728	807	196	637	196
Laptop	994	341	870	128	464	169
Twitter	1561	173	1560	173	3127	346

## 3.2 参数设置与评价指标

实验中,使用BERT生成词嵌入,默认嵌入维度768,使用AdamW作为优化器,批处理大小为32。初始学习率设置为 $1 \times 10^{-4}$ ,L2正则项系数设置为 $1 \times 10^{-5}$ ,Epoch为20。同时,本文采用准确率(Accuracy)和宏平均F1(Macro-F1)值作为情感极性预测的评价指标。

## 3.3 对比实验

为证明模型的有效性,在3个公开数据集上,选用

$\mathbf{X}^e \in \mathbb{R}^{m \times n \times d}$ ,并使用Tanh函数激活。

## 2.6 输出层

本层的目标是将文本向量矩阵池化,并将特征维度从词嵌入维度降到与情感极性类别数量一致。首先,使用BERT的BertPooler函数,对文本向量矩阵 $\mathbf{X}^e$ 进行池化。因为在BERT中,句子的第一个标记“[CLS]”的隐藏状态,汇总了整个序列的隐藏信息,所以BertPooler函数实质上是取所有文本向量第一个Token对应的隐藏状态 $\mathbf{D}_{\text{cls}} \in \mathbb{R}^{m \times d}$ 。其次,经过全连接层和Tanh函数激活,返回池化结果 $\mathbf{X}^{\text{pooler}} \in \mathbb{R}^{m \times d}$ 。然后,再次经过全连接层,将 $\mathbf{X}^{\text{pooler}}$ 的特征维度由词嵌入维度 $d$ 降低到情感极性类别数量3,并使用softmax函数得到极性判定的概率分布。最后,在模型训练时使用交叉熵损失函数计算损失。

## 3 实验与分析

### 3.1 实验环境与数据集

实验所涉及的环境:Windows Server 2019 操作系统,CPU为AMD EPYC 7453,内存512G,GPU为NVIDIA A800 80GB。另外,使用的开发语言为Python,开发工具为Pycharm,深度学习框架torch1.13,CUDA版本为11.8。

实验数据集选用方面级情感分析任务常用的公开数据集,包括国际语义评测竞赛SemEval-2014 Task4<sup>[26]</sup>提供的Restaurant14和Laptop14数据集及ACL14 Twitter<sup>[27]</sup>数据集。3个数据集包含的情感极性分为3类:积极、消极和中性,其分布如表1所示。

以下模型与本文的BERT-MIFN模型进行比较:

ATAE-LSTM<sup>[15]</sup>。在隐藏层拼接额外的方面词向量,并通过注意力机制,让模型更加关注方面词和上下文的交互特征。

IAN<sup>[16]</sup>。对方面词和上下文词分别建模,通过建立交互注意力机制来同时捕获上下文和方面词的关联信息。

MGAN<sup>[17]</sup>。结合粗粒度和细粒度的注意力机制,构建多粒度注意力网络,来捕获方面词与上下文的交

互信息。

ASGCN<sup>[13]</sup>。通过句法依存树构建句法邻接矩阵,并利用多层图卷积网络来融合词语位置信息和句法信息。

BERT-SPC<sup>[21]</sup>。通过设置特定掩码并微调 BERT 实现方面级情感分析任务。

LCF-BERT<sup>[22]</sup>。利用双结构 BERT 分别建模,并使用上下文特征动态加权机制作用于文本向量矩阵,从而降低远离局部上下文的单词权重。

LCFS-ASC<sup>[23]</sup>。对 LCF-BERT 进行改进,在前者基础上,引入句法相对距离概念并以此生成权重。

R-GAT<sup>[24]</sup>。对普通的句法依存树进行重塑和修剪,构建了一个新的以目标方面词为根节点的树结构。

SSEGCN<sup>[25]</sup>。提前构建句法依存关系和词性等外部知识,并使用注意力机制获取文本得分矩阵。

实验结果如表 2 所示,标“\*”的为本文训练得到的结果,其他引自原文。

表 2 不同模型对比实验结果

单位:%

Model	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
ATAE-LSTM	77.20	—	68.70	—	—	—
IAN	78.6	—	72.1	—	—	—
MGAN	81.25	71.94	75.39	72.47	72.54	70.81
ASGCN	80.77	72.02	75.55	71.05	72.15	70.40
BERT-SPC	84.46	76.98	78.99	75.03	73.55	72.14
LCF-BERT-CDW *	86.79	80.94	80.61	77.09	75.39	74.11
LCFS-ASC-CDW	86.71	80.31	80.52	77.13	—	—
R-GAT+BERT	86.60	<u>81.35</u>	78.21	74.07	76.15	74.88
SSEGCN+BERT	<b>87.31</b>	81.09	<u>81.01</u>	<u>77.96</u>	<u>77.40</u>	<u>76.02</u>
BERT-MIFN	<u>87.23</u>	<b>81.58</b>	<b>82.13</b>	<b>79.02</b>	<b>77.60</b>	<b>76.16</b>

由表 2 知,本文提出的 BERT-MIFN 模型在 Laptop 和 Twitter 数据集上的准确率和宏平均 F1 指标均取得最优结果,在 Restaurant 数据集上的准确率为次优。上述对比模型分为两类,即:使用 Glove 等作为词嵌入的模型(ATAE-LSTM、IAN、MGAN、ASGCN)和其余使用 BERT 生成词嵌入的模型。实验结果表明 BERT 模型生成的词嵌入向量相对于 Glove 等语义信息更加丰富,效果更好。BERT-MIFN 与 BERT-SPC、LCF-BERT-CDW 和 LCFS-ASC-CDW 对比,结果表明融合文本句法距离信息和位置距离信息比只考虑一种信息效果更好。BERT-MIFN 与 R-GAT+BERT 和 SSEGCN+BERT 对比,实验结

果表明使用 BERT 对文本进行整体语义提取要优于部分仅使用 BERT 作为词嵌入的模型效果。

3.4 消融实验

为证明本文所提模型整体结构的有效性,将不同结构去掉后进行消融实验,实验结果如表 3 所示。w/o DP 表示不使用双路径结构,w/o MIFN-PI 表示在多信息融合层中去掉位置距离信息,w/o MIFN-SI 表示在多信息融合层中去掉句法距离信息,w/o MIFN-RI 表示在多信息融合层中去掉实际文本长度与总文本长度的比值信息。

表 3 消融实验结果

单位:%

Model	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
BERT-MIFN	<u>87.23</u>	<b>81.58</b>	<b>82.13</b>	<b>79.02</b>	<b>77.60</b>	<b>76.16</b>
w/o DP	86.96	81.13	81.81	78.25	75.29	73.57
w/o MIFN-PI	<b>87.32</b>	81.15	80.72	77.51	75.43	74.68
w/o MIFN-SI	86.79	80.68	80.56	77.39	75.29	74.48
w/o MIFN-RI	86.88	80.78	80.56	76.57	75.87	74.87

由表 3 知,本文模型 BERT-MIFN 的整体效果优于去掉不同结构后的模型。对比 w/o DP 和本文模型,结果表明双路径结构能够减少多信息提取层对文本嵌入特征的干扰。对比 w/o MIFN-PI、w/o MIFN-SI、w/o MIFN-RI 和本文模型,结果表明文本位置距离信息、句

法距离信息和文本比值信息,可以让模型更加关注与方面词相关的上下文。但 w/o MIFN-PI 在 Restaurant 数据集上的准确率指标反而略高于本文模型,本文推断是由于不同数据集对文本与方面词位置距离信息的敏感程度不同导致。对比 w/o MIFN-PI 和 w/o MIFN-SI,去掉

文本与方面词的位置距离信息和句法距离信息后,两组实验在不同数据集不同指标上表现不一致,表明不同数据集语言特性不同,对句法距离信息和位置距离信息的敏感性存在差异。

## 4 结束语

所提出的基于 BERT 的多信息融合方面级情感分析模型 BERT-MIFN,使用双路径结构分别进行多信息融合和整体语义提取,减少信息融合对文本嵌入特征的干扰;使用多信息提取层对文本句法距离和位置距离等信息进行融合,获得更丰富的特征。将本文模型与其他模型在 3 种公开数据集上进行对比实验,结果表明,相对于其他仅使用注意力机制或仅将 BERT 作为词嵌入的部分模型,本文的 BERT-MIFN 模型整体上取得更优的效果。此外,对模型的消融实验结果也证明模型不同结构的有效性。下一步工作将考虑融合文本词性等信息,同时进一步探究不同信息之间的交互影响。

## 参考文献:

- [1] Tang D, Qin B, Liu T. Learning semantic representations of users and products for document level sentiment classification [C]. Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing, 2015: 1014–1023.
- [2] Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification [C]. Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, 2016: 1480–1489.
- [3] 汤凌燕,熊聪聪,王嫒,等.基于深度学习的短文本情感倾向分析综述[J].计算机科学与探索,2021,15(5):794–811.
- [4] Wang H, Liu B, Li C, et al. Learning with noisy labels for sentence-level sentiment classification [C]. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019: 6286–6292.
- [5] Baccianella S, Esuli A, Sebastiani F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining [C]. International Conference on Language Resources and Evaluation, 2010: 2200–2204.
- [6] Cambria E, Poria S, Hazarika D, et al. SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings [C]. Proceedings of the AAAI conference on artificial intelligence, 2018: 1795–1802.
- [7] Mohammad S, Turney P. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon [C]. Proceedings of the NAACLHLT 2010 workshop on computational approaches to analysis and generation of emotion in text, 2010: 26–34.
- [8] 郁圣卫,卢奇,陈文亮.基于领域情感词典特征表示的细粒度意见挖掘[J].中文信息学报,2019,33(2):112–121.
- [9] Hochreiter S, Schmidhuber J. Long Short-Term Memory [J]. Neural Computation, 1997, 9(8): 1735–1780.
- [10] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C]. Proceedings of the 2013 conference on empirical methods in natural language processing, 2013: 1631–1642.
- [11] Tang D Y, Qin B, Feng X C, et al. Effective LSTMs for Target-Dependent Sentiment Classification [C]. Proceedings of 26th International Conference on Computational Linguistics, 2016: 3298–3307.
- [12] Kim Y. Convolutional Neural Networks for Sentence Classification [C]. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014: 1746–1751.
- [13] Zhang C, Li Q C, Song D W. Aspect-based Sentiment Classification with Aspect specific Graph Convolutional Networks [C]. Proceedings of Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019: 4568–4578.
- [14] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. arXiv preprint arXiv:1409.0473, 2014.
- [15] Wang Y, Huang M, Zhu X, et al. Attention-based LSTM for aspect-level sentiment classification [C]. Proceedings of the 2016 conference on empirical methods in natural language processing, 2016: 606–615.
- [16] Ma D H, Li S J, Zhang X D, et al. Interactive at-

- tention networks for aspect-level sentiment classification[C]. Proceedings of the 26th International Joint Conference on Artificial Intelligence, 2017: 4068–4074.
- [17] Fan F, Feng Y, Zhao D. Multi-grained attention network for aspect-level sentiment classification [C]. Proceedings of the 2018 conference on empirical methods in natural language processing, 2018: 3433–3442.
- [18] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]. Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013: 3111–3119.
- [19] Pennington J, Socher R, Manning C. GloVe: Global Vectors for Word Representation[C]. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014: 1532–1543.
- [20] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019: 4171–4186.
- [21] Song Y W, Wang J, Jiang T, et al. Attentional encoder network for targeted sentiment classification [J]. arXiv preprint arXiv:1902.09314, 2019.
- [22] Zeng B, Yang H, Xu R, et al. LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification[J]. Applied Sciences, 2019, 9 (16): 3389.
- [23] Phan M H, Ogunbona P O. Modelling context and syntactical features for aspect-based sentiment analysis[C]. Proceedings of the 58th annual meeting of the association for computational linguistics, 2020: 3211–3220.
- [24] Wang K, Shen W, Yang Y, et al. Relational Graph Attention Network for Aspect-based Sentiment Analysis [C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 3229–3238.
- [25] Zhang Z, Zhou Z, Wang Y. SSEGCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis [C]. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022: 4916–4925.
- [26] Pontiki M, Galanis D, Pavlopoulos J, et al. Semeval-2014 task 4: Aspect based sentiment analysis [C]. The 8th International Workshop on Semantic Evaluation, 2014: 27–35.
- [27] Dong Li, Wei F R, Tan C Q, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification [C]. The 52nd Annual Meeting of the Association for Computational Linguistics, 2014: 49–54.

## Aspect-Level Sentiment Analysis Model based on BERT with Multi-Information Fusion

ZHANG Xiaofeng, TAO Hongcai

(School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China)

**Abstract:** Aspect-level sentiment analysis aims to determine the sentiment polarity of different aspects of a text sentence, which is one of the hot tasks in the field of natural language processing. Most of the current BERT-based aspect-level sentiment analysis methods only use it as a pre-trained word embedding tool, without making full use of its own semantic extraction capabilities and task processing capabilities, and often ignore the correlation of different external information when using information such as syntax and aspect word position. To solve the above problems, this paper proposes a multi-information fusion network model based on BERT. Firstly, the main path and auxiliary path based on BERT word embedding are constructed, and the auxiliary path could avoid interference with the word embedding features. Secondly, under the auxiliary path, the syntactic distance, position distance and other information between different words and aspect words are interacted according to the syntactic dependency tree of the text sentence and the multi-information fusion algorithm. Finally, convolutional neural network and attention mechanism are used to fuse the features of the main path and the auxiliary path. Experimental results on three datasets show that the proposed model is effective.

**Keywords:** aspect-level sentiment analysis; syntactic dependency tree; BERT; multi-information fusion