

# 融合BERT多层特征的方面级情感分析

李宁健, 方睿

成都信息工程大学, 计算机学院, 四川 成都  
Email: leeningjian@qq.com, fangrui@cuit.edu.cn

收稿日期: 2020年11月19日; 录用日期: 2020年12月2日; 发布日期: 2020年12月9日

## 摘要

BERT (Bidirectional Encoder Representation from Transformers) 预训练语言模型相较于传统词向量 (one-hot、word2vec等) 可以动态表示词语, 并且在11项下游任务中取得了最好的结果。基于BERT在特定语料中微调的方式逐渐被广泛应用于情感分析任务中并且取得了不错的效果, 然而仅仅使用BERT最后一个编码层的输出特征用于分类, 忽略了其他层学习到的语义特征。不同于以往基于BERT的分类模型, BERT-MLF融合了BERT每一个编码层输出的方面特征, 并通过卷积层提取层之间的关键语义特征, 减少冗余信息的影响, 充分利用了每个编码层学习到的信息。该方法在SemEval-2014 Task 4的Laptop和Restaurant数据集上进行了大量实验, 实验结果表明该方法实现了良好的分类性能。

## 关键词

深度学习, 方面级情感分析, BERT, 微调, 方面特征, 分类特征, 卷积神经网络

# Aspect-Based Sentiment Analysis Based on Multi-Layer Features of Bert

Ningjian Li, Rui Fang

School of Computer Science, Chengdu University of Information Technology, Chengdu Sichuan  
Email: leeningjian@qq.com, fangrui@cuit.edu.cn

Received: Nov. 19<sup>th</sup>, 2020; accepted: Dec. 2<sup>nd</sup>, 2020; published: Dec. 9<sup>th</sup>, 2020

## Abstract

Compared with traditional word vectors (one-hot, word2vec, etc.), BERT (Bidirectional Encoder Representation from Transformers) pre-trained language model can dynamically represent words and has achieved the best results in 11 downstream tasks. The method of fine-tuning in

specific corpus based on BERT has been widely used in sentiment analysis tasks and achieved good results. However, only the output features of the last coding layer of BERT are used for classification, ignoring the semantic features learned by other layers. Different from the previous classification model based on BERT, BERT-MLF integrates the aspect features of each coding layer of BERT, and extracts the key semantic features between the layers through the convolution layer to reduce the influence of redundant information and make full use of the information learned by each coding layer. A large number of experiments have been carried out on the Laptop and Restaurant datasets of SemEval-2014 Task 4, and the experimental results show that the method achieves good classification performance.

## Keywords

Deep Learning, Aspect-Level Sentiment Analysis, BERT, Fine-Tune, Aspect Feature, Classification Feature, Convolutional Neural Networks

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着经济和互联网的发展,网民数量急剧增长,人们在互联网上日益活跃,伴随着大量网络信息的产生。其中包括大量文本信息,如购物网站的用户评论、餐饮服务的用户点评等用户产生的文本信息。这些文本大多包含一定的感情色彩,人工分析大量文本中的情感信息往往费时费力,随着计算机和深度学习的发展,利用计算机分析文本中的情感信息能达到事半功倍的效果,其研究价值和商业价值越来越明显。

情感分析是自然语言处理中的一个热点任务,早期的句子级情感分析和文档级情感分析主要是对整个句子或文档进行情感分类,从一个句子或者文档中总结出一个情感,没有考虑句子中存在多个方面的情况。然而一个句子往往包含实体的多个方面,这无疑忽略掉了句子中丰富的情感信息。方面级情感分析(Asspect-Based Sentiment Analysis, ABSA [1] [2] [3])作为更加细粒度的情感分析,目的是对句子中多种方面的情感极性进行分析。比如,对于评论“这个手机的做工很精致,但是电池容量太小了。”,方面级情感分析旨在分析出对于“做工”和“电池容量”两个不同方面的情感倾向,得出<做工,积极>和<电池容量,消极>的结论。相较于句子级和文档级,方面级情感分析更加精准的对句子中不同方面的情感进行分析,没有忽略整句话或者整个文档中包含的多种情感信息。在实际应用中,如对电商平台中的用户评论或评价进行方面级情感分析,进一步分析出用户对产品不同方面的满意程度,有助于商家完善服务和产品的质量,有比较高的研究意义。

由于方面级情感分析需要从句子中分析出实体不同方面对应的情感极性,而不同实体方面往往有不同的情感极性,所以相比句子级和文档级要更加困难。现有的方法一般是对每个方面单独处理,为了准确分析出对应方面的情感,一般会将方面信息加入到模型中来监督模型提取方面对应的情感。在 BERT 等预训练语言模型被提出以前,多用 CNN 和 LSTM 提取特征,以及利用注意力机制关注句子的重要部分。BERT 被提出后,基于 BERT 微调的方式在分类效果上表现出优异的性能。而对于基于 BERT 微调的方面级情感分析,通常将最终方面特征和分类特征拼接后进行分类。也可以构造辅助句子将单句分类任务转换为 BERT 句子对分类任务。由于 BERT 独特的自注意力机制,使得其每个输出特征都关注了句

子中其他词语的信息,故本文的工作在 BERT 的基础上,提出基于 BERT 多层特征的模型(BERT-MLF),结合每一层的方面特征,并通过卷积神经网络提取层之间的语义特征,充分利用了每个编码层学习的语义信息,同时本文在现有数据集上进一步预训练,使得模型表现出良好的分类效果。

## 2. 相关研究

### 2.1. 语言模型

在自然语言处理任务中,最关键的一环是对字词的表示,计算机不能像人一样理解每个字词代表的含义,则需要用计算机能够识别的方式表示字词,即词向量。在语言模型还未被提出之前,主要以 one-hot 作为词向量,每个词向量都是词典长度大小的向量,向量中对应词的位置用 1 表示,其余位置用 0 表示。这种方式可以表示对应词典中的每个词语,但存在高维度和向量稀疏的问题。随后语言模型被提出用于训练出更加低维精准的词语表示,如 word2vec [4]通过 CBOW (Continuous Bag Of Words)和 Skip-gram 语言模型利用上下文预测中心词和中心词预测上下文作为训练目的进行无监督训练,得到固定长度的词向量可以表示词语间的关系,含义越相近的词语在向量空间中的距离也越近,且具有比较低的维度。类似 word2vec 这种低维词向量的还有 Global Vectors [5] (GloVe)。虽然通过这种语言模型可以得到低维向量,但是依然存在词语在不同的文本环境下有相同的词向量表示的问题。

不同于传统词向量,ELMO [6] (Embedding from Language Models)语言模型利用双向长短期记忆网络 [7] (Bi-Directional Long Short-Term Memory, BiLSTM)在大规模语料中学习词语的含义,在特定语料中微调可以使词向量适应不同语境中的词语表示,做到了动态表示词语。但是由于循环神经网络[8] (Recurrent Neural Network, RNN)在模型训练时存在难以并行化的问题,导致模型在大规模语料中训练非常的耗时。GPT [9]利用基于自注意力机制的单向 Transformer [10]编码器捕捉语义特征,并行能力和特征捕捉能力优于 ELMO,大大提升了训练速度,但是其单向模型忽略了上下文信息。BERT [11] (Bidirectional Encoder Representation from Transformers)考虑到句子的时序性,构建基于多层双向 Transformer 编码器的语言模型,在更大规模的数据中更加全面的学习句子中的语义信息和上下文信息,对词语的表示更加精确。目前,语言模型在大规模语料中预训练,特定语料下微调的方法被广泛应用于自然语言处理任务中,其效果突破了基于传统词向量的模型。

### 2.2. 基于传统词向量的模型

在模型结构方面,早期的机器学习方法通常需要人工构建大量特征工程,且分类效果一般。随着神经网络理论和应用的不断发展,各种神经网络模型不断涌现出来,深度学习模型在情感分析任务中展现了其强大的性能。

Kim [12]最早使用卷积神经网络(Convolutional Neural Networks, CNN)来对句子进行情感分类,并且对多种 CNN 变体进行了实验,如随机初始化词嵌入的 CNN-rand,使用固定的预训练词嵌入的 CNN-static,预训练和微调词嵌入的 CNN-non-static 和多通道的 CNN-multichannel。Huang [13]等人将方面信息加入到 CNN 模型中用于情感分类,监督模型捕捉不同方面的情感信息。随后针对 CNN 难捕捉长距离特征的缺点,Tang [14]等人提出了基于长短期记忆网络(Long Short-Term Memory, LSTM)的目标依赖 LSTM (TD-LSTM)和目标连接 LSTM (TC-LSTM)方法,该方法通过考虑方面信息来扩展 LSTM。他们将给定的目标视为特征并将其与上下文特征相联系,以进行方面情感分类。虽然 LSTM 具有建模方面和上下文关系的能力,但始终存在难并行化的诟病。随着注意力机制逐渐被应用到自然语言处理任务中,Wang [15]等人提出一种基于注意力机制的目标嵌入 LSTM 方法,这被证明是一种有效的强制神经模型处理句子相关部分的方法。注意力机制用于强制模型关注句子的重要部分,使模型具备了响应特定方面情感的能力。

Yang [16]等人考虑到句子的时序性, 也提出了两种基于注意力机制的双向 LSTM 以提高分类性能。Liu 和 Zhang [17]通过区分给定目标或者方面的左语境和右语境获得的注意来扩展注意力建模, 通过增加多个门进一步控制了注意力的作用。Tang [18]等人引入了端到端的记忆网络, 用于方面级别的情感分类, 该机制采用了带有外部记忆的注意力机制, 以捕获每个上下文词相对于给定目标方面的重要性。

### 2.3. BERT

BERT [11]是一种由双向 Transformer 编码器组成的新型预训练语言模型, Transformer 模型最早用于机器翻译, 由于其编码器独特的多头自注意力机制, 使其拥有优异的特征捕捉能力, 编码器结构如图 1 所示。

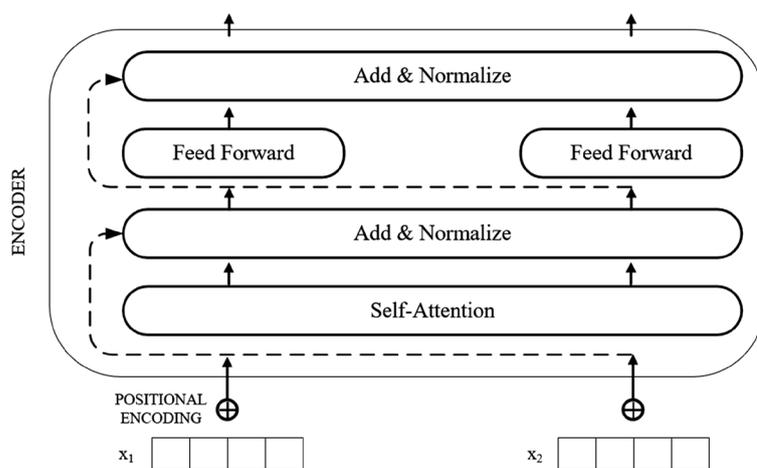


Figure 1. Transformer Encoder  
图 1. Transformer 编码器

Transformer 编码器主要由自注意力层、残差连接和归一化层以及前馈神经网络层组成, 核心部分为其多头注意力模块, 注意力机制可以计算出当前词语对句中其他词汇的关注程度和内部联系, 通过将输入向量与对应的权重矩阵相乘得到查询向量( $Q$ )、键向量( $K$ )和值向量( $V$ ), 并通过式 1 计算出注意力结果, 通过式 2 和式 3 拼接计算多头注意力的结果。

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (2)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \quad (3)$$

残差连接可以改善模型梯度消失问题并增强模型的泛化能力, 再通过归一化来加快模型的收敛速度。Transformer 模型结合了 CNN 和 RNN 的优点, 使其在可并行计算的同时也可以很好地提取特征。

BERT 通过构建多层双向 Transformer 编码器, 在 GPT 和 ELMO 的基础上, 结合两者的优点, 其 Transformer 编码器使其拥有优异的特征捕捉能力, 双向模型结构使其可以充分获取上下文信息, 模型结构如图 2 所示。同时, BERT 创新地构建了掩码语言模型(MLM, Masked Language Model)和下一句预测(NSP, Next Sentence Predict)两个无监督任务, MLM 任务类似于英文中的完形填空任务, 通过上下文信息去预测被遮蔽的单词, NSP 任务则是通过句子对分类任务判断前后两个句子是否连续。BERT 首先在大

规模语料上无监督训练, 预先学习好字符级、词级和句子级特征, 然后在下游任务中进行微调。在模型输入输出方面, 输入  $E_i$  为词向量(Token Embeddings)、分段向量(Segment Embeddings)和位置向量(Position Embeddings)相加得到的向量, 每经过一层 Transformer 编码器都会输出对应的特征  $T_i$ , 每个输出特征都包含了对每个输入的注意力信息。[SEP]为句子结束标记, 特殊标记[CLS]关注了整个句子的信息, 也被称为句向量, 常被用于句子级别的分类。

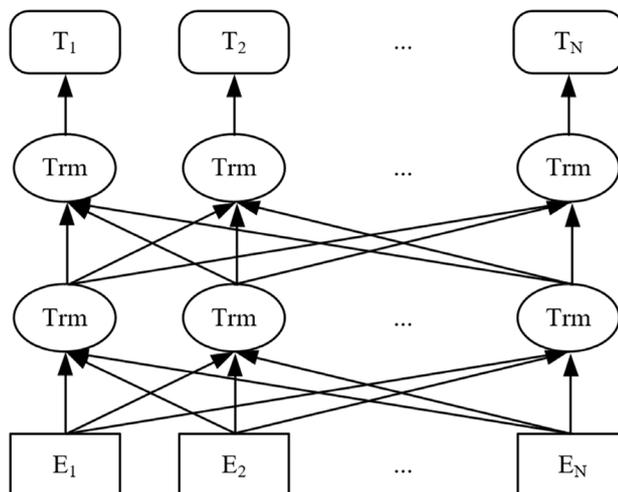


Figure 2. The structure of BERT language model  
图 2. BERT 语言模型结构

## 2.4. 基于 BERT 的模型

BERT 预训练语言模型在大规模语料中学习和理解语义的能力优于传统词向量, 解决了传统词向量静态向量的问题, 对词语的表示更加精确。Song [19]等人提出了一种避免重复的注意力编码网络(AEN), 该网络采用基于注意力的编码来建立上下文和目标之间的模型, 提出了标签不可靠性问题, 引入了标签平滑正则化, 并对比了以传统词向量为输入(AEN-GloVe)和以 BERT 生成的词向量为输入(AEN-BERT)的模型效果, 后者效果高于前者。其次, Song [19]通过将方面术语作为第二个句子, 将原句和方面术语构造为 BERT 句子对分类任务, 然后进行微调。Sun [20]等人在 BERT 的基础上通过为原始句子构建围绕方面术语的多种类型的辅助句子将方面级情感分析转化为句子对分类问题。Gao [21]等人认为简单的交换 BERT 的词表示并不能提高现有模型的效果, 然后通过一些简单的微调策略, 首先直接对 BERT 输出的方面特征进行分类, 然后将 BERT 输出中对应方面位置的信息与分类输出结合并且构造辅助句子用于分类。虽然 BERT 逐渐被用于方面级情感分析中, 但是几乎都只利用了最后一层输出特征, 忽略了其他层的输出特征, 造成了一部分语义信息的损失。

## 3. 微调策略

### 3.1. 研究问题定义

根据 SemEval-2014 Task 4 [22]对方面级情感分析任务的定义, 该任务包含四个子任务, 分别是方面术语抽取、方面术语的情感分类、方面类别检测和方面类别的情感分类。本文主要专注于解决子任务方面术语的情感分类。任务定义如下:

给定一个包含  $n + 1$  个词语的句子  $s$ :

$$s = \{w_0, w_1, \dots, w_{n-1}, w_n\} \tag{4}$$

和句中包含  $m$  个词语的方面术语  $a$ :

$$a = \{w_i, w_{i+1}, \dots, w_{i+m-1}\} \tag{5}$$

其中  $a$  为  $s$  的子序列, 可以有多个, 任务的目标就是要对句子  $s$  中的方面术语  $a$  进行情感{正向, 中性, 负向}的分类。比如给定句子 “Great food but the service was bad.” 以及方面术语 food 和 service, 任务目的就是分析出 food 和 service 分别对应的情感极性。

### 3.2. 融合 BERT 多层特征

BERT 作为预训练语言模型, 会预先在大规模语料中学习语义知识, 然后在下游任务中微调, 在以往基于 BERT 的分类模型中, 往往只利用最后一层的分类特征[CLS]进行分类。然而, Jawahar [23]等人通过一系列实验表明, BERT 在低层编码中学习到了短语级的信息, 在中层编码中学习到了句法信息, 高层编码中则学习到了丰富的语义特征。虽然在 BERT 的结构中, 每一层模型都会将上一层的输出特征作为输入去进行计算, 但是经过多次复杂的计算难免造成低层特征和中层特征的丢失。

其次由于 Transformer 编码器的自注意力机制会计算句子中所有单词与当前词汇的注意力信息, 并且通过模型的训练可以捕捉到特定单词间的内部联系, 因此相对于分类特征[CLS], 方面特征中更多的关注了与方面术语相关的情感信息。如在 Gao [21]的工作中, 首先将最终输出的分类特征[CLS]和方面特征进行拼接后分类, 然后直接将方面特征用于分类, 其实验结果表明后者的准确率高于前者, 直接利用方面特征进行分类更加有针对性地分析当前方面术语对应的情感信息。故 BERT-MLF 将每一层的方面特征结合起来, 使每一层输出的方面特征都发挥作用, 同时为了避免低层特征未彻底学习而带来冗余信息的问题, 再通过卷积神经网络提取层与层之间的多层特征, 在去噪的同时也能提取更多关键的特征, 模型结构如图 3 所示。

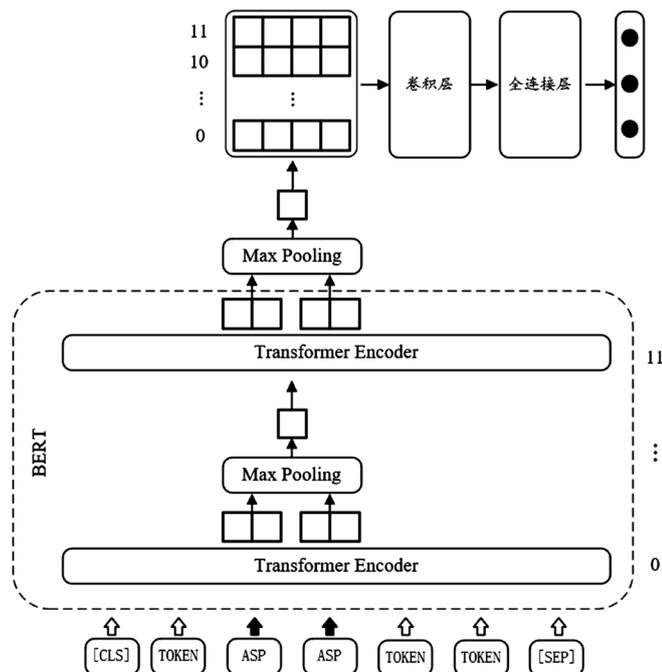


Figure 3. The structure of BERT-MLF  
图 3. BERT-MLF 模型结构

BERT 有两种预训练模型:

**BERT<sub>base</sub>**: 12 层 Transformer 网络, 768 个隐藏层单元, 12 个自注意力头, 总参数量 1.1 亿。

**BERT<sub>large</sub>**: 24 层 Transformer 网络, 1024 个隐藏层单元, 16 个自注意力头, 总参数量 3.4 亿。

由于 **BERT<sub>large</sub>** 需要比较高的显存, 所以本文选择 **BERT<sub>base</sub>** 作为预训练模型。

### 3.3. 卷积层和全连接层

卷积神经网络常被用于图像处理领域, 由于其在特征捕捉方面表现优异, 随后也被逐渐用于自然语言处理任务中, 如 Kim [12] 将卷积神经网络应用于句子级别的情感分类, 虽然卷积神经网络存在难捕捉长距离关系依赖的问题, 但是在本文中, 输入的 12 层方面特征为短序列, 因此卷积神经网络可以优雅地进行特征提取。卷积神经网络主要由输入层、卷积层、池化层组成。常规的输入层为单词的词向量, 如 Word2Vec 向量和 GloVe 向量, 在本文中输入层为最大池化后的 BERT 的 12 层方面特征。卷积层主要是在输入的向量矩阵(12 × 768)上滑动不同的卷积核并运行卷积运算, 不同卷积核大小赋予卷积核不同的感受视野, 同时也能捕捉不同的特征信息。本文使用的卷积核尺寸为  $h \times d$ ,  $h$  为卷积核高度分别为 3、4 和 5,  $d$  为特征维度大小, 在 BERT<sub>base</sub> 中  $d$  为 768, 每个尺寸的卷积核有 256 个。计算公式如式 6。

$$c = f(W * X + b) \tag{6}$$

式中  $c$  为卷积结果,  $W$  为卷积核权重矩阵,  $X$  为输入矩阵,  $b$  为偏置值,  $*$  表示卷积操作,  $f$  为激活函数, 本文使用 Relu 激活函数来优化学习。

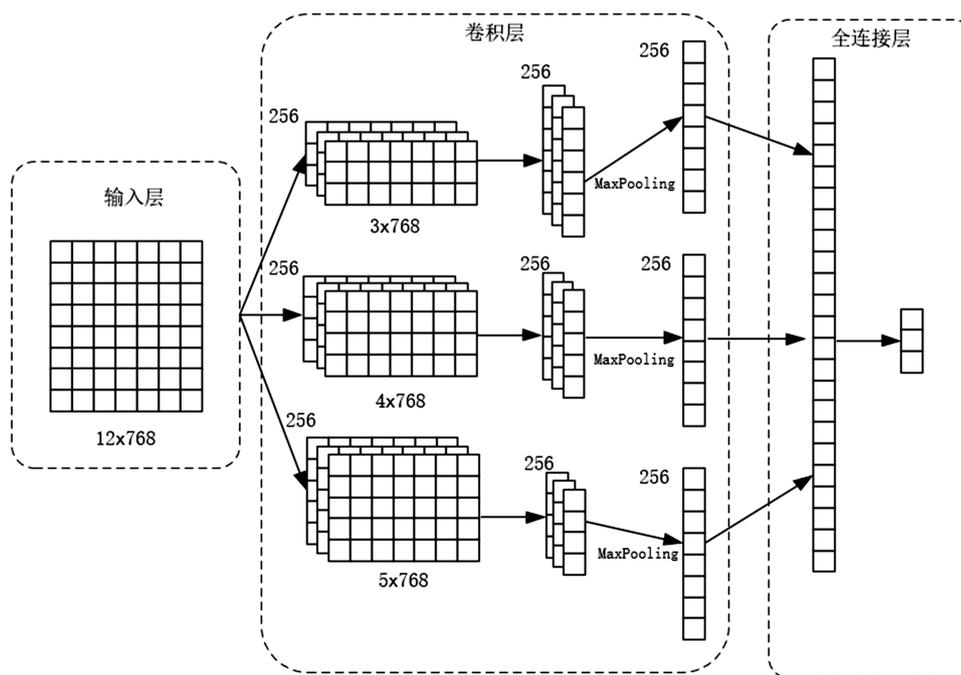


Figure 4. The structure of convolution layer and FC layer  
图 4. 卷积层和全连接层结构

池化层主要采用最大池化来保留关键特征, 减少特征的维度, 防止模型过拟合, 通过对不同尺寸的卷积核产生的卷积结果进行最大池化, 最终产生了  $3 \times 256 = 768$  个特征。全连接层主要是对经过卷积运算并最大池化后的特征进行全连接计算和 softmax 计算从而得到类别概率, 计算公式如式 7、式 8 所示, 卷积层结构如图 4 所示。

$$f = TW_f + b \tag{7}$$

$$s_i = \frac{e^{V_i}}{\sum_j e^{V_j}} \tag{8}$$

式中  $f$  为全连接计算结果,  $T$  为全连接输入特征,  $W_f$  为全连接权重矩阵,  $b$  为偏置值,  $s_i$  为 softmax 计算结果。

## 4. 实验分析

### 4.1. 数据集

实验数据选择的是被广泛使用的 SemEval-2014 Task 4 的 Laptop 和 Restaurant 数据集, 数据内容分别是电脑和餐厅的顾客评论, 情感类别包含正向、中性和负向。数据细节见表 1。

**Table 1.** Data sample distribution

**表 1.** 数据样本分布

数据集	训练集			测试集		
	Pos.	Neu.	Neg.	Pos.	Neu.	Neg.
Laptop	1002	464	870	343	169	128
Restaurant	2172	637	807	728	196	196

### 4.2. 实验设置

实验环境见表 2, 超参设置见表 3。

**Table 2.** Experimental environment information

**表 2.** 实验环境信息

实验环境	配置信息
操作系统	Ubuntu 16.04.6
GPU	Tesla V100
显存	16GB
内存	32GB
编程语言	Python 3.7.4
深度学习框架	PaddlePaddle 1.6.2

**Table 3.** Hyper parameter setting

**表 3.** 超参设置

参数	参数值
Learning rate	3e-5
Batch size	16
Training epochs	6
Max sequence length	128
Optimizer	Adam
Dropout rate	0.1

### 4.3. 评价指标

模型效果衡量标准选择的是分类准确率(Accuracy, Acc)和 Macro-F1 (MF1), 计算公式如下所示。

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{12}$$

$$MF1 = \frac{F_1^0 + F_1^1 + F_1^2}{3} \tag{13}$$

上式中  $TP$  表示正样本预测正确的样本数,  $TN$  表示负样本预测正确的样本数,  $FP$  表示负样本预测错误的样本数,  $FN$  表示正样本预测错误的样本数。 $P$  为精确率, 主要关注预测为正的样本数中有多少是真正的正样本。 $R$  为召回率, 主要关注所有的正样本中有多少被预测为正样本。 $F_1$  值为结合了精确率和召回率的综合评价指标, 由于本文的分类任务为三分类, 故选择 Macro-F1 值平均三个类别的  $F_1$  值。

### 4.4. 实验结果

模型对比方面首先对比了不进行卷积操作直接融合所有特征的情况, 然后与基于 BERT 的一些模型做了对比。进行对比的模型如下:

IAN [24]通过两个 LSTM 分别对方面实体和上下文建模, 然后通过注意力机制在方面和上下文间交互学习各自的特征表示。

BERT-SPC [19]在 BERT 的基础上, 将方面实体和上下文转换为句子对分类任务。

BERT-PT [25]受到机器阅读理解的启发, 利用 BERT 对电商平台的用户评论阅读理解, 这一方法同样适用于方面级情感分析。

AEN-BERT [19]是使用 BERT 生成的词向量为输入的注意力编码网络, 使用基于注意力机制的编码网络对方面和上下文建模, 并在分类时对标签进行平滑正则化。

TD-BERT [21]直接利用 BERT 输出的对应方面位置的特征, 进行情感分类。

TD-BERT-QA-CON [21]在构建辅助句子的前提下, 拼接 BERT 最后输出的分类特征[CLS]和方面特征, 进行分类。

本文主要进行情感正向、中性和负向三分类, 进行对比的模型结果中, “-”表示该指标未报告, 其他未标记的摘自相应的论文中。模型训练完毕后在测试集上进行测试, 最终结果为五次测试结果的平均值。实验对比结果见表 4、表 5, 各项最大值已加粗表示。

**Table 4.** Comparison before and after convolution

**表 4.** 卷积前后对比

卷积与否	Laptop		Restaurant	
	Acc	MF1	Acc	MF1
否	79.03	74.40	86.76	79.77
是	81.30	77.53	87.62	81.26

**Table 5.** Comparison of multi model effects  
**表 5.** 多模型效果对比

词嵌入	模型	Laptop		Restaurant	
		Acc	MF1	Acc	MF1
BERT	IAN	72.10	–	78.60	–
	BERT-SPC	78.99	75.03	84.46	76.98
	BERT-PT	78.07	75.08	84.95	76.96
	AEN-BERT	78.35	73.68	81.46	71.73
	TD-BERT	80.00	75.19	85.30	79.69
	TD-BERT-QA-CON	78.57	75.76	85.06	80.40
	BERT-MLF	81.30	77.53	87.62	81.26

实验对比结果表明, 相对于只利用 BERT 最终层的分类特征用于分类, BERT-MLF 全面利用 BERT 每一层方面特征, 结合 BERT 每层学习的不同特征可以有效提高分类效果。同时直接融合所有层的方面特征也同时融合了较多的冗余信息和噪音, 通过多个不同尺寸的卷积核提取不同层的关键特征, 减少冗余信息对分类效果的负面影响, 在分类效果上有一定的提升。此外, BERT 模型是在通用语料上进行预训练, 对于特定领域下的分类任务, 其数据分布与 BERT 通用语料不一致, 为了让模型参数更加契合特定领域的的数据, 本文还分别在 Laptop 和 Restaurant 数据集上进行进一步预训练。对于这两个数据集中的文本数据, 句子间的连续性不强和上下文关系较弱, 因此本文舍弃了 Next Sentence Predict 预训练任务, 保留 Masked Language Model 任务使模型在特定语料中继续学习。进一步预训练后模型效果对比如表 6 所示。

**Table 6.** Comparison of BERT-MLF further pre-training before and after  
**表 6.** BERT-MLF 进一步预训练前后对比

进一步预训练与否	Laptop		Restaurant	
	Acc	MF1	Acc	MF1
否	81.30	77.53	87.62	81.26
是	81.79	78.16	88.36	81.73

从对比结果中可以看出, 在特定领域中进一步预训练能帮助模型进一步提高分类效果, 但是由于参与进一步预训练的数据量不大, 分类效果提升不明显, 在预训练数据量充足的情境下, 模型深入学习特定领域的的数据, 进一步预训练后模型在分类性能上会有更好的表现。

## 5. 结束语

BERT 展现出优异的学习语义的能力, 基于 BERT 的模型在方面级情感分析中也有比较好的表现, 本文在 BERT 的基础上, 结合 BERT 每一层学习的语义特征用于分类, 有效提高了分类效果, 但是能否更好的利用特征, 对不同的特征进行不同的处理, 还需要进一步研究。其次, 目前对于句子中存在多个方面术语的情况, 往往将方面术语拆分出来单独处理, 忽略了方面术语之间的联系, 造成一些方面术语之间的信息损失, 后期会进一步探索如何解决这些问题, 提高分类效果。最后, 本文处理的数据为英文数据, 将来会就如何将模型应用在中文数据以及如何将具体的情感分析应用实现作进一步的探索。

## 基金项目

国家自然科学基金(61806029)。

## 参考文献

- [1] Hu, M. and Liu, B. (2004) Mining and Summarizing Customer Reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, 22-25 August 2004, 168-177. <https://doi.org/10.1145/1014052.1014073>
- [2] Liu, B. (2012) Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, **5**, 1-167. <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- [3] Pang, B. and Lee, L. (2008) Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, **2**, 1-135. <https://doi.org/10.1561/15000000011>
- [4] Mikolov, T., Chen, K., Corrado, G., et al. (2013) Efficient Estimation of Word Representations in Vector Space.
- [5] Pennington, J., Socher, R. and Manning, C.D. (2014) Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, October 2014, 1532-1543. <https://doi.org/10.3115/v1/D14-1162>
- [6] Peters, M.E., Neumann, M., Iyyer, M., et al. (2018) Deep Contextualized Word Representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long Papers), 2227-2237. <https://doi.org/10.18653/v1/N18-1202>
- [7] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [8] Elman, J.L. (1990) Finding Structure in Time. *Cognitive Science*, **14**, 179-211. [https://doi.org/10.1207/s15516709cog1402\\_1](https://doi.org/10.1207/s15516709cog1402_1)
- [9] Radford, A., Narasimhan, K., Salimans, T., et al. (2018) Improving Language Understanding by Generative Pre-Training. <https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf>
- [10] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, 4-9 December 2017, 5998-6008.
- [11] Devlin, J., Chang, M.W., Lee, K., et al. (2018) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding.
- [12] 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>
- [13] Huang, B. and Carley, K.M. (2019) Parameterized Convolutional Neural Networks for Aspect Level Sentiment Classification. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, October-November 2018, Brussels, 1091-1096. <https://doi.org/10.18653/v1/D18-1136>
- [14] Tang, D., Qin, B., Feng, X., et al. (2015) Effective LSTMs for Target-Dependent Sentiment Classification.
- [15] Wang, Y., Huang, M. and Zhao, L. (2016) Attention-Based LSTM for Aspect-Level Sentiment Classification. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, November 2016, 606-615. <https://doi.org/10.18653/v1/D16-1058>
- [16] Yang, M., Tu, W., Wang, J., et al. (2017) Attention Based LSTM for Target Dependent Sentiment Classification. *Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, California, 4-9 February 2017, 5013-5014.
- [17] Liu, J. and Zhang, Y. (2017) Attention Modeling for Targeted Sentiment. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Volume 2, Short Papers, 572-577. <https://doi.org/10.18653/v1/E17-2091>
- [18] Tang, D., Qin, B. and Liu, T. (2016) Aspect Level Sentiment Classification with Deep Memory Network. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, November 2016, 214-224. <https://doi.org/10.18653/v1/D16-1021>
- [19] Song, Y., Wang, J., Jiang, T., et al. (2019) Attentional Encoder Network for Targeted Sentiment Classification.
- [20] Sun, C., Huang, L. and Qiu, X. (2019) Utilizing Bert for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence.
- [21] Gao, Z., Feng, A., Song, X., et al. (2019) Target-Dependent Sentiment Classification with BERT. *IEEE Access*, **7**, 154290-154299. <https://doi.org/10.1109/ACCESS.2019.2946594>
- [22] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutopoulos, I. and Manandhar, S. (2014) SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Association for Computational Linguistics, Dublin, 27-35. <https://doi.org/10.3115/v1/S14-2004>

- [23] Jawahar, G., Sagot, B. and Seddah, D. (2019) What Does BERT Learn about the Structure of Language? *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, July 2019, 3651-3657. <https://doi.org/10.18653/v1/P19-1356>
- [24] Ma, D., Li, S., Zhang, X., *et al.* (2017) Interactive Attention Networks for Aspect-Level Sentiment Classification. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main Track*, Melbourne, 19-25 August 2017, 4068-4074. <https://doi.org/10.24963/ijcai.2017/568>
- [25] Xu, H., Liu, B., Shu, L., *et al.* (2019) BERT Post-Training for Review Reading Comprehension and Aspect-Based Sentiment Analysis. *Proceedings of NAACL-HLT 2019*, Minneapolis, 2-7 June 2019, 2324-2335.