

基于文本筛选和改进BERT的长文本方面级情感分析

王 昆, 郑 毅, 方书雅, 刘守印*

(华中师范大学 物理科学与技术学院, 武汉 430079)

(* 通信作者电子邮箱 syliu@mail.ccnu.edu.cn)

摘 要: 方面级情感分析旨在分类出文本在不同方面的情感倾向。在长文本的方面级情感分析中, 由于长文本存在的冗余和噪声问题, 导致现有的方面级情感分析算法对于长文本中方面相关信息的特征提取不够充分, 分类不精准; 而在方面分层为粗粒度和细粒度方面的数据集上, 现有的解决方案没有利用粗粒度方面中的信息。针对以上问题, 提出基于文本筛选和改进BERT的算法TFN+BERT-Pair-ATT。该算法首先利用长短时记忆网络(LSTM)和注意力机制相结合的文本筛选网络(TFN)从长文本中直接筛选出与粗粒度方面相关的部分语句; 然后将部分语句按次序进行组合, 并与细粒度方面相结合输入至在BERT上增加注意力层的BERT-Pair-ATT中进行特征提取; 最后使用Softmax进行情感分类。通过与基于卷积神经网络(CNN)的GCAE(Gated Convolutional Network with Aspect Embedding)、基于LSTM的交互式注意力模型(IAN)等经典模型相比, 该算法在验证集上的相关评价指标分别提高了3.66%和4.59%, 与原始BERT模型相比提高了0.58%。实验结果表明, 基于文本筛选和改进BERT的算法在长文本方面级情感分析任务中具有较大的价值。

关键词: 方面级; 情感分析; 预训练模型; 长短时记忆神经网络; 注意力机制

中图分类号: TP389.1 **文献标志码:** A

Long text aspect-level sentiment analysis based on text filtering and improved BERT

WANG Kun, ZHENG Yi, FANG Shuya, LIU Shouyin*

(College of Physical Science and Technology, Central China Normal University, Wuhan Hubei 430079, China)

Abstract: Aspect-level sentiment analysis aims to classify the sentiment of text in different aspects. In the aspect-level sentiment analysis of long text, the existing aspect-level sentiment analysis algorithms do not fully extract the features of aspect related information in the long text due to the redundancy and noise problems, leading to low classification accuracy. On the datasets with coarse and fine aspects, existing solutions do not take advantage of the information in the coarse aspect. In view of the above problems, an algorithm named TFN+BERT-Pair-ATT was proposed based on text filtering and improved Bidirectional Encoder Representation from Transformers (BERT). First, the Text Filter Network (TFN) based on Long Short-Term Memory (LSTM) neural network and attention mechanism was used to directly select part sentences related to the coarse aspect from the long text. Next, the related sentences were associated with others in order, and after combining with fine aspects, the sentences were input into the BERT-Pair-ATT, which is with the attention layer added to the BERT, for feature extraction. Finally, the sentiment classification was performed by using Softmax. Compared with the classical Convolutional Neural Network (CNN) based models such as Gated Convolutional network with Aspect Embedding (GCAE) and LSTM based model Interactive Attention Network (IAN), the proposed algorithm improves the related evaluation index by 3.66% and 4.59% respectively on the validation set, and improves the evaluation index by 0.58% compared with original BERT. Results show that the algorithm based on text filtering and improved BERT has great value in the aspect-level sentiment analysis task of long text.

Key words: aspect-level; sentiment analysis; pre-trained model; Long Short-Term Memory (LSTM) neural network; attention mechanism

0 引言

情感分析(Sentiment Analysis, SA)是自然语言处理(Natural Language Processing, NLP)中一个非常重要的子领域。近些年在学术界获得广泛的关注, 在现实中也具有重要的应用前景^[1]。

实际生活中, 用户在互联网上产生的文本中包含的情感是复杂的、多方面的, 用户在一段文本中会表达出在不同方面的不同情感倾向。基于这个事实, Liu^[2]提出了方面级情感分析(Asspect-Based Sentiment Analysis, ABSA)任务, 旨在分类出文本在不同方面所包含的情感, 实现文本的多维度情感理解

收稿日期: 2020-02-19; 修回日期: 2020-04-09; 录用日期: 2020-04-17。

作者简介: 王昆(1996—), 男, 湖北黄冈人, 硕士研究生, 主要研究方向: 自然语言处理、情感分析; 郑毅(1993—), 男, 湖北武汉人, 博士研究生, 主要研究方向: 深度学习、时序信号处理; 方书雅(1995—), 女, 湖北襄阳人, 硕士研究生, 主要研究方向: 人脸识别、目标检测; 刘守印(1964—), 男, 河南周口人, 教授, 博士, 主要研究方向: 无线通信、物联网、机器学习。

功能。但是,因为 ABSA 任务的复杂性,人工提取文本中的方面及其表达的情感非常困难,所以相关数据集稀缺,初期的工作主要集中在如何提取方面的研究^[3]。近几年, Pontiki 等^[4-5]在 SemEval 比赛中增加了基于 ABSA 的子任务,同时给出了短文本 ABSA 数据集,于是开始有大量的学者针对 ABSA 任务中方面信息利用和情感上下文问题提出各种解决方法。目前 ABSA 任务的主流解决方案包括早期复杂的特征工程方法^[6-7]以及近期的神经网络算法^[8-11]。已有研究表明,神经网络算法优于特征工程方法^[1]。在神经网络算法中, Wang 等^[8]首次基于注意力机制和长短时记忆网络(Long Short-Term Memory, LSTM)提出 ATAE-LSTM(ATtention-based LSTM with Aspect Embedding)模型,利用方面语义来搜寻文本中与之相关位置的内容,在短文本 ABSA 数据集上性能优于单纯的 LSTM^[12]。Ma 等^[9]注意到方面及其上下文的协调性,在 ATAE-LSTM 的基础上进一步提出了交互式注意力模型(Interactive Attention Network, IAN),考虑了文本与方面语义之间的交互,进一步提升了情感分类正确率。Huang 等^[10]注意到 IAN 的池化层忽略了文本与方面中的每个词之间的相互关系,提出了没有池化层的 AOA-LSTM(Attention Over Attention LSTM)模型,取得了比 IAN 更优的分类结果。Xue 等^[11]首次提出结合卷积神经网络(Convolutional Neural Network, CNN)^[13]和门控机制^[14]的 GCAE(Gated Convolutional network with Aspect Embedding)模型,利用门控单元从文本中筛选出与方面相关的信息,取得了更高的正确率。

虽然关于方面级情感分析的带标签数据稀缺,但是非方面级情感分析的标记数据集很丰富, Chen 等^[15]提出将转移胶囊网络(Transfer Capsule Network, TransCap)用于将迁移非方面级情感分析中的知识到方面级情感分类,自此该领域的研究开始倾向于使用迁移学习算法。同时,随着语言模型的不不断发展, Devlin 等^[16]提出的文本预训练模型 BERT(Bidirectional Encoder Representation from Transformers)在处理文本分类任务中表现出色,摒弃了卷积和循环神经网络的结构,使用双向 Transformer^[17]搭建整体网络,预训练过程中能够从浅入深地学习到语言中包含的语法语义^[18]。BERT 同样适用于 ABSA 任务, Sun 等^[1]利用方面来构建疑问句,并结合 BERT 模型提出 BERT-pair-QA-M 等模型,实验结果表明在短文本 ABSA 数据集上比使用神经网络算法的分类结果更优。Gao 等^[19]在 BERT 的基础上提出 TD-BERT 模型,通过抽取 BERT 编码层特定位置的特征进行方面级情感分类,进一步提升分类性能,表明 BERT 在特征提取方面具有优越性能。

在长文本 ABSA 任务上,由于在方面上表达的情感信息只存在于该方面相关的语句中,因此如何挖掘文本中与方面有关的上下文是一个非常重要的挑战。袁丁等^[20]基于 GCAE,提出了 CNN-GCAE 模型,利用字词向量联合模型对数据进行预处理,克服了 GCAE 模型在数据预处理方面的不足,提高了在长文本 AI Challenger2018 细粒度情感分析比赛数据集上的召回率^[20]。

以上现有的神经网络算法仍然存在如下问题:首先,基于循环神经网络的模型仍然存在一定的梯度消失,而基于卷积神经网络的模型没有考虑到文本的时序关系,在长文本 ABSA 任务上的泛化能力较差;其次,因为 BERT 的最长输入序列长

度固定,超过这个长度将被直接截断,被截断的文本可能仍然具有没有被利用的信息。所以,对于长文本还需要进一步的数据处理,神经网络模型的性能还有进一步提升的空间。同时, AI Challenger2018 细粒度情感分析比赛数据集上将方面分为粗粒度方面和细粒度方面两个层次,对于粒度方面的利用研究目前还在初步阶段,如何把现有的模型应用在这个数据集上是一个挑战。

针对以上问题,本文提出文本筛选和改进 BERT 的算法来处理长文本的 ABSA 任务。该算法首先利用文本筛选网络(Text Filter Network, TFN)对文本中的每个语句进行粗粒度方面的分类,得到部分语句,再对部分语句按次序进行组合,输入到 BERT-Pair-ATT 情感分类模型中进行训练和预测。BERT-Pair-ATT 模型使用句子对的方式来处理输入。因为原始 BERT 模型直接将每个位置点的输出进行平均得到特征向量,没有体现出每个位置的重要程度,所以本文在 BERT 的顶层增加一个注意力层,对 BERT 的输出特征赋予不同的权重。实验结果表明,使用本文算法在长文本方面情感分析数据集上的结果比原始 BERT 高 0.58%。本文主要工作为:

1)设计了文本筛选网络 TFN,利用 LSTM 来提取层次方面中粗粒度方面的文本特征,同时结合了注意力机制来对文本中每个部分进行特征加权,最终筛选出和粗粒度方面相关的文本(部分语句),对数据进行充分的预处理。

2)先利用 TFN 对原始文本进行一次文本筛选,再将带有方面信息的文本(部分语句)连同细粒度方面同时输入至 BERT-Pair-ATT 中进行特征提取, BERT-Pair-ATT 模型充分提取方面中包含的语义特征,提高了长文本 ABSA 任务的性能。

1 长文本方面级情感分析算法设计

1.1 任务定义

本文重点研究长文本在粗粒度和细粒度两个层次下所包含的情感。对于一段文本,需要分类出在预定义的粗粒度方面下的每个细粒度方面所包含的 4 种情感趋势。情感趋势的集合表示为 $E = \{\text{“未提及”}, \text{“积极”}, \text{“中性”}, \text{“消极”}\}$ 。统一起见,本文将文本表示为 S , S 可以由单个字符 $char$ 的集合来表示,即 $S_c = \{char_1, char_2, \dots, char_{c_num}\}$; 经过 jieba 分词软件分词可以在词 $word$ 的角度表示为 $S_w = \{word_1, word_2, \dots, word_{w_num}\}$; 经过分句之后在句子 $sent$ 上表示为句子的组合 $S_s = \{sent_1, sent_2, \dots, sent_{s_num}\}$, 其中 c_num , w_num 和 s_num 分别为文本的包含的字符总数、单词总数和句子总数。 S_s 经过一定方法筛选之后的集合表示为 $S_f = \{sf_1, sf_2, \dots, sf_m\}$, m 为 S_f 中句子总数, S_f 为 S_s 的子集。粗粒度方面(Coarse Aspect)表示为 $CA = \{CA_1, CA_2, \dots, CA_p\}$, 细粒度方面(Fine Aspect)表示为 $FA = \{FA_1, FA_2, \dots, FA_q\}$, 其中 p, q 代表的集合中方面的总数, FA 集合中的元素是将 CA 中元素进一步细粒度划分得到的。

1.2 算法框架

本文的算法框架如图 1 所示,分为两个步骤:第一步首先对数据集中的原始文本进行分句算法处理得到句子组合 $S_s = \{sent_1, sent_2, \dots, sent_{s_num}\}$ 。为保证句子的次序,依次将每个句子与每个粗粒度方面的组合 $\{sent_i, CA_r\}$ 输入至 TFN 模型中判断这个句子是否与此粗粒度方面相关,相关则输出 1,反之输出 0,其中 r 为 CA 中第 r 个粗粒度方面。最终得到所有相关语句的集合 $S_f = \{sf_1, sf_2, \dots, sf_m\}$, 其中 $m \leq s_num$; 当所有语句都

与粗粒度方面无关时,选取原始文本,此时 $m=s_num$ 。这一步能够压缩数据集中文本的长度,增强数据在第二步的信息效率。第二步利用基于 BERT 的 BERT-Pair-ATT 模型进行方面级情感分类,其输入采用句子对的方式,第一个文本为 S_f 中句子按照次序组合起来的新文本,第二个文本为 CA_f 层次下的每个 FA ,例如“位置”层次下的“交通是否便利”,输出为情感趋势 E 中的类标签。

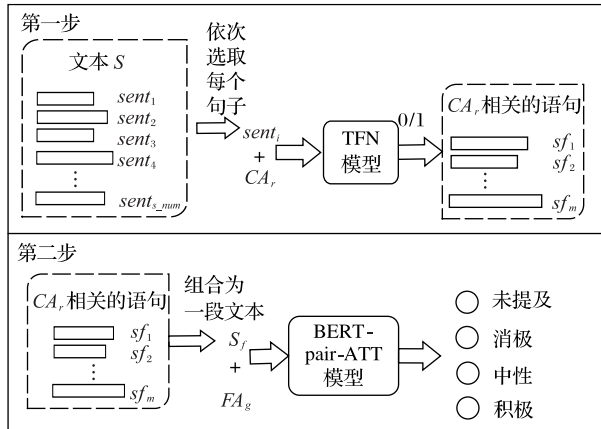


图1 本文算法框架

Fig. 1 Framework of proposed algorithm

1.3 TFN 模型设计

TFN 模型的任务是对数据集的文本中的每个句子按照与之相关的粗粒度方面进行正确分类,筛选出对于第二步有分类价值的部分语句。本文设计的 TFN 模型整体框架如图 2 所示。主要包含了词嵌入层、两个 LSTM 编码器^[12]、注意力层和二分类器模块。本文利用词嵌入模型 Word2Vec^[21-22]在原始语料库上预先训练出一个领域相关的词嵌入模型作为词嵌入层的模型参数,大小为 $v \times d$, v 为词汇表的大小, d 为词嵌入的维度。通过 LSTM 模型分别对句子和粗粒度方面文本建模,分别得到大小为 $h \times l$ 和 $h \times c$ 的对应嵌入向量,其中 h 为 LSTM 的隐藏层的维度, l, c 分别为分句后语句和粗粒度方面在词层面的长度。利用注意力机制,得到原始文本与粗粒度方面的注意力权重。最后用权重与句嵌入向量相乘的结果作为分类器的特征向量输入,分类出这个句子是否和粗粒度方面是相关的。以下将具体介绍单独模块细节。

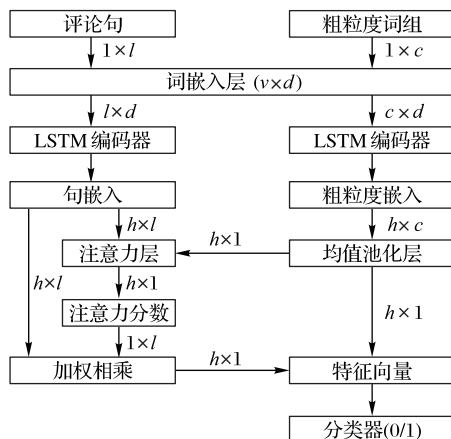


图2 TFN 模型架构

Fig. 2 TFN model architecture

1.3.1 LSTM 编码器

如图 3 所示的 LSTM 结构^[12],分别对语句以及粗粒度方面建模。输入的语句由 l 个词组成 $\{w_s^1, w_s^2, \dots, w_s^l\}$,粗粒度方面由 c 个词组成 $\{w_a^1, w_a^2, \dots, w_a^c\}$ 。输入至词嵌入层,得到每个词的向量表示 $w^i \in \mathbb{R}^d$,对于没有出现在词表中的词,本文使用 $U(-0.25, 0.25)$ 的均匀分布随机初始化,然后通过 LSTM 网络得到语句的张量表示 $S_h = [h_s^1, h_s^2, \dots, h_s^l]$, $S_h \in \mathbb{R}^{h \times l}$ 以及粗粒度方面的张量表示 $F = [h_a^1, h_a^2, \dots, h_a^c]$, $F \in \mathbb{R}^{h \times c}$,其中 h^k 为 LSTM 的隐藏层在每一个时间点 k 上的输出, LSTM 的隐藏层的输出具体的计算公式如式(1)~(6)所示:

$$i^k = \sigma(W_i[h^{k-1}, w^k] + b_i) \quad (1)$$

$$f^k = \sigma(W_f[h^{k-1}, w^k] + b_f) \quad (2)$$

$$o^k = \sigma(W_o[h^{k-1}, w^k] + b_o) \quad (3)$$

$$\bar{cell}^k = \tanh(W_c[h^{k-1}, w^k] + b_c) \quad (4)$$

$$cell^k = f^k \odot cell^{k-1} + i^k \odot \bar{cell}^k \quad (5)$$

$$h^k = o^k \odot \tanh(c^k) \quad (6)$$

其中: i^k, f^k, o^k 分别为输入门、遗忘门和输出门在 k 时刻的输出; $\bar{cell}^k$ 和 $cell^k$ 分别为 k 时刻的中间量和状态量; W_i, W_f, W_o, W_c 为权重矩阵; b_i, b_f, b_o, b_c 为偏置向量; σ 为 Sigmoid 函数; “ $\odot$ ” 符号为矩阵对应元素相乘。

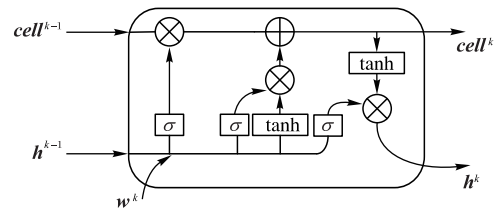


图3 LSTM 结构

Fig. 3 LSTM structure

1.3.2 注意力模块

注意力机制在情感分析中广泛应用,在情感分析任务中有很多模型中都嵌入了注意力机制^[8-10]。本文提出一种基于注意力机制的架构,将粗粒度方面的信息与原始文本同时考虑进来,共同计算对于语句中每个词的注意力分数。在方面情感分析中, Wang 等^[8]证明了方面对于相关任务的重要性,并且本文认为除了文本整体的语义,一些和粗粒度方面相关的词或者词组能够有效地帮助分类相关方面,所以本文提出使用注意力机制来学习文本中不同词对于粗粒度方面的权重。粗粒度方面的句嵌入 F 通过均值池化层进行处理得到 $f_{avg} \in \mathbb{R}^{h \times 1}$, 即如式(7)所示:

$$f_{avg} = \sum_{i=1}^m h_a^i / m \quad (7)$$

通过将 S_h 和 f_{avg} 作为注意力模块的输入,可以得到 f_{avg} 对于 S_h 中的每个时间点输出的注意力权重 $weight \in \mathbb{R}^{n \times 1}$, 本文注意力模块的相似度计算准则为尺度点乘,按照式(8)计算权重:

$$weight = \text{softmax} \left(\frac{S_h^T \cdot f_{avg}}{\sqrt{d_h}} \right) \quad (8)$$

其中 softmax 函数为指数归一化函数,定义如式(9)所示:

$$\text{softmax}(x_i) = \frac{x_i}{\sum_j x_j} \quad (9)$$

利用权重向量 $weight$ 对句嵌入向量 S_h 进行加权并与粗粒度方面的语义嵌入向量 f_{avg} 进行拼接 (concatenate) 得到最终的特征表示 $y \in \mathbb{R}^{2h \times 1}$, 如式 (10) 所示:

$$y = \text{concatenate}(S_h \cdot weight, f_{avg}) \quad (10)$$

1.3.3 分类器

分类器包含了两层前馈神经网络: 第一层使用大小为 h 的全连接层学习参数 $W_1 \in \mathbb{R}^{2h \times h}$; 第二层使用大小为 2 的全连接层学习另一组权重 $W_2 \in \mathbb{R}^{h \times 2}$ 。 y 输入至这两层前馈网络得到输出特征向量 v , 最终对 v 进行 softmax 运算得到输出 $output$, 公式如式 (11)~(12) 所示:

$$v = W_2^T (W_1^T \cdot y + b_1) + b_2 \quad (11)$$

$$output = \text{softmax}(v) \quad (12)$$

1.4 BERT-Pair-ATT 模型设计

BERT 是一种语言表征模型, 在大型的语料库上进行预训练, 然后利用预训练模型的模型参数在其他的 NLP 任务上进行微调。 BERT-Pair-ATT 模型基于 BERT 原始模型进行改进, 具体的模型结构如图 4 所示。

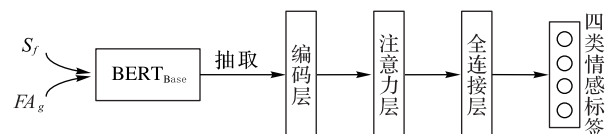


图 4 BERT-pair-ATT 模型架构

Fig. 4 BERT-pair-ATT model architecture

Sun 等^[1]提出构造疑问句的方式来处理 ABSA 任务, 并且得到了性能提升, 同时本文数据集中的细粒度方面都具有疑问句的形式, 例如“交通是否便利”, 因为这种数据形式, 所以 BERT-Pair-ATT 将文本语句作为第一个文本, 细粒度方面作为第二个文本作为输入。 BERT-Pair-ATT 的输出与标准的 BERT_{BASE}^[16] 结构不同, BERT_{BASE} 的结构将首位的特征作为分类特征, 利用这个分类特征输入至全连接层来分类至具体的类别, 而本文利用注意力机制, 对所有的位置的特征进行处理, 在其上添加注意力层, 随机初始化两个矩阵 $W_3 \in \mathbb{R}^{b \times d}$, $W_4 \in \mathbb{R}^{b \times 1}$ 来学习所有位置上的哪些特征对于分类的任务更加有价值, 其中 b 为 BERT 模型隐层维度, W_3 、 W_4 在训练过程中不断调整学习所有位置的输出中与分类任务更为相关的部分。注意力

层的分数计算见式 (13), 其中, E_B 为 BERT 的编码层的输出, $E_B \in \mathbb{R}^{L \times b}$ 。式 (14)~(15) 将 $score$ 权重与编码器的输出相乘得到加权的特征 $feature$, 再将每个位置的特征加权平均得到最终的分类型特征, 并输入到全连接层进行分类。全连接层的权重为 $W_3 \in \mathbb{R}^{b \times 4}$, 偏置为 b_3 , 最终得到预测的情感标签 $label$ 。

$$score = \text{softmax}(\tanh(E_B \cdot W_3) \cdot W_4) \quad (13)$$

$$feature = \frac{1}{L} \sum_{i=1}^L E_B^i \cdot score^i \quad (14)$$

$$label = \text{softmax}(feature \cdot W_3 + b_3) \quad (15)$$

2 实验与结果分析

2.1 数据集

实验采用 AI Challenger 2018 比赛中的细粒度情感分析数据集。此数据集为目前国内最大的公开中文方面情感分析数据集, 其中训练集和测试集分别含有 10.5 万和 1.5 万条餐饮评论, 文本为段落级别的长文本, 并且每段文本给出了 6 个粗粒度方面 (“位置” “服务” “价格” “环境” “菜品” “其他”) 和总计 20 个细粒度方面的 4 类情感 (“未提及”: -2, “中性”: 0, “积极”: 1, “消极”: -1)。作为模型验证, 实验选取其中的 “位置” 粗粒度下的 “交通是否便利” “距离商圈远近” “是否容易寻找” 3 个细粒度方面进行分析。由于任务是多标签分类, 最终数据集的数量按照细粒度方面的数目扩展 3 倍至 31.5 万和 4.5 万条, 训练集的数量统计如表 1 所示。

表 1 训练集上的标签统计

Tab. 1 Label statistics on training set

粗粒度	细粒度	积极	中性	消极	未提及
位置	交通是否便利	21 254	1 024	1 318	81 383
	距离商圈远近	20 201	533	586	83 680
	是否容易寻找	17 947	2 472	3 976	80 605

表 2 中展示了从训练集中随机选取一个评论文本及其在 “位置” 粗粒度方面下的情感倾向值。对于其中的评论文本样例, 在 “位置” 粗粒度方面下, 仅提及其中关于商圈的信息, 所以在其层次下的各个细粒度方面下的情感只在 “商圈是否远近” 上表达一种积极的情感, 并未提及其他方面。样例具体的情感倾向标签如表 2 所示。

表 2 数据集评论文本样例

Tab. 2 Sample of review text in dataset

评论样例	粗粒度方面	细粒度方面	情感倾向值
和闺蜜相约欧亚达鱼酷, 三个人竟然喜欢不同口味, 最后只得放弃我最爱的黄金酸辣, 吃的油泼香辣口味的, 之前觉得这个口味一般, 没有黄金酸辣那种酸爽的惊艳感, 不过这次吃感觉也不错, 鱼烤的特别好, 味道也很香滑, 非常好吃, 香辣口味满满的都是辣椒, 视觉感觉特别辣, 但是辣度只是微辣水平, 吃的不是很辣, 对于不是特别能吃辣可是又希望有点微辣口味的朋友来说是个不错的选择。看我的名字就知道我是鱼酷的粉, 欧亚达开业之前去过以前大成那边的店, 那个基本上都得等座, 现在的空间大了一些, 7 点到的基本没等就有座挺好的, 下次还会来这家顺便可以逛逛。现在 3 个口味都可以打 9 折, 蛮合适的, 大家可以选择。	位置	交通是否便利	-2
		距离商圈远近	1
		是否容易寻找	-2

2.2 TFN 模型的训练数据

TFN 模型的训练数据来源于 AI Challenger 2018 比赛中的细粒度情感分析数据集。评论文本由多个评论句组成。为了利用原始数据集中的粗粒度方面的信息来提取出与之相关的评论句, 假设一段评论文本在一个粗粒度下的任意一个细粒度方面上的情感标签不为 “未提及”, 则表明这段评论文本中

至少有一个评论句在提及这个粗粒度方面。基于这个假设, 本文算法利用词向量来计算原文中所有词与粗粒度词之间的余弦相似度来筛选出最相关的词, 提取出该词所在的句子作为正类, 同时随机选取该粗粒度下的细粒度方面下全部都为 “未提及” 的评论中的评论句作为负类, 这一类语句并没有在提及这个粗粒度方面。经过上述算法的处理, 最终得到针对

“位置”粒度的检测数据集,汇总的数据集的标签统计如表3。为了保证数据集的完整性,采用了大约相同数目的标签分布。

表3 TFN模型数据集标签统计
Tab. 3 Label statistics of TFN model dataset

数据集类型	提及	未提及	总计
训练集	44 169	53 801	97 970
验证集	6 195	7 810	14 005

2.3 评价指标

由表1可见数据集的标签具有分布不平衡的特性。因为在不平衡数据集上正确率不能准确衡量数量少的类别上的分类性能,而 $macro-F1$ 能够客观准确计算每一个类别上的分类性能,所以实验采用 $macro-F1$ 值来对模型的分类性能进行评价。首先在四类情感标签上分别计算出 $F1$ 值,然后对每个标签上的 $F1$ 进行加权平均得到最终模型在本文数据集上的 $macro-F1$ 值,这样可以有效抑制数据不平衡对评价价值的影响。 $F1$ 的计算公式如式(16)~(18)所示:

$$R_{(i)} = \frac{TP_{(i)}}{TP_{(i)} + FN_{(i)}} \quad (16)$$

$$P_{(i)} = \frac{TP_{(i)}}{TP_{(i)} + FP_{(i)}} \quad (17)$$

$$F1_{(i)} = 2 \times \frac{R_{(i)} \times P_{(i)}}{R_{(i)} + P_{(i)}} \quad (18)$$

其中: i 表示在第 i 个情感标签上的计算的评价指标; R 为召回率(Recall); P 为精准率(Precision); TP 为正类情感标签被模型预测为正类的样本数量(True Positive); FN 为正类情感标签被模型预测为负类的样本数量(False Negative); FP 为负类情感标签被模型预测为正类的样本数量(False Positive)。最终的 $macro-F1$ 指标计算公式如式(19):

$$macro-F1 = \frac{1}{4} \sum_{i=1}^4 F1_{(i)} \quad (19)$$

2.4 相关模型对比

实验将TFN+BERT-Pair-ATT与如下相关模型进行对照实验:

1)GCAE^[11]。它是一种卷积网络结合门控机制的深度学习网络,通过门控机制来模拟注意力机制筛选出属性相关的信息作为特征。

2)ATAE-LSTM^[8]。该模型将属性的词嵌入分别添加至评论文本作为LSTM的输入以及LSTM的隐藏层作为注意力函数的输入,最终将注意力层的输出与LSTM隐藏层的加权和作为分类器的特征。

3)IAN^[9]。利用交互式的注意力网络处理ABSA任务,作为对比,本文将任务中的在原文本出现的属性替换为预定义的类别属性,即“交通是否便利”“距离商圈远近”“是否容易寻找”,从而使得IAN能够适用于本文任务。

4)AOA-LSTM^[10]。该模型包含了两个双向长短期记忆网络(Bi-LSTM)以及AOA模块,借鉴了注意力机制的思想。同样的本文将属性修改为预定义的属性以适用于本文任务作为对比。

5)BERT-Single。只考虑原始文本,采用全原始连接层进行微调,因为文本对应了多个标签,针对每个标签进行训练与验证。

6)BERT-Pair。输入评论文本以及细粒度方面作为句子对分类任务,并只采用原始的全连接的方式进行微调。

前四类模型的文本嵌入层采用Word2Vec算法来预训练参数,后两类模型采用BERT文本表示法。

2.5 相关超参数

实验中所有的模型都基于PyTorch^[23]深度学习框架,使用4块1080TiGPU进行训练推理。Word2Vec的词嵌入维度 d 为300,词典的大小 v 为20 000。TFN使用的LSTM隐层维度为300,在训练中使用Adam优化器(Adaptive Moment Estimation),学习率设置为0.001, $\beta_1=0.9$, $\beta_2=0.99$,训练过程中的批处理大小为256,全局迭代5次,为了防止过拟合,在网络的每一层加入Dropout^[24],丢弃概率(Dropout rate)为0.2。前四类对比模型都使用Adam优化器,学习率为0.001,批处理大小设置为32,迭代次数设置为10。BERT-Single、BERT-Pair以及BERT-Pair-ATT都使用BERT_{BASE}作为基础模型架构,表4为基于BERT_{BASE}的文本表示法的相关模型的训练参数设置,且在预训练初期使用warm-up技巧进行预热^[16]。以上所有的模型要优化的损失函数全部为负指数的似然函数 $-\log(P(y|x))$,其中 (x,y) 为(输入,目标标签)。

表4 BERT参数设置

Tab. 4 BERT parameter setting

参数	值
优化器	Adam
学习率	2E-5
最大长度(max length)	350
批处理大小(batch size)	32
迭代次数(epoch)	4

2.6 结果与分析

TFN训练过程中在训练集和测试集上的top1正确率如图5所示,由于加入Dropout,训练集上的正确率低于测试集,最终TFN在测试集上的最高正确率为92.82%。

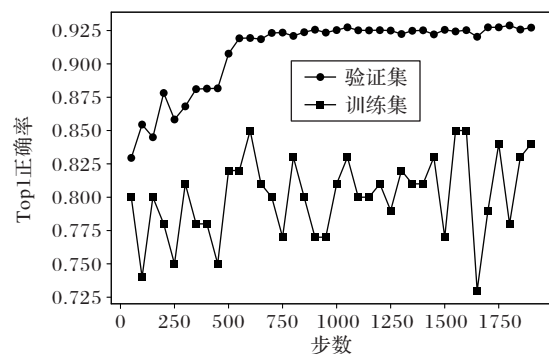


图5 TFN训练中top1正确率曲线

Fig. 5 Top1 accuracy curves in TFN training

实验的相关模型都经过10次实验,取10次中最好的结果保存得到表5实验结果。由表5中的数据可以在本文数据集上得到一些结论:1)除了BERT-Single模型,其他使用BERT为基础文本表示法的模型比CNN或者循环神经网络

(Recurrent Neural Network, RNN)文本表示法的模型表现更加优异;2)本文提出的BERT-Pair-ATT模型比BERT-Pair模型在 *macro-F1* 上高出 0.27%,并且在每个标签下的 *F1* 都有提升;3)本文提出的文本筛选和BERT-Pair-ATT的模型架构能够提升传统的单模型的性能,其中TFN+BERT-Pair-ATT模型取得了最高的结果,使单模型BERT-Pair-ATT的 *macro-F1* 再次提升0.31%。

由于BERT-Single没有利用方面中的信息,导致最终的分类结果不佳。在考虑利用方面之后,BERT-Pair模型在 *macro-F1* 上提高了4.14%。BERT预训练语言模型因为预先在大型语料库上进行充分训练,已经学习到了语言中的相关信息,因此最终的训练结果基于都优于基于Word2Vec表示法的相关模型;并且使用了句子对方式用于长文本方面情感分析任务上的BERT模型,只需要训练一个模型参数就能够带来性能上的提升。对于长文本的处理,本文算法采取的是神经网络来策略截短,而不是使用注意力机制来自适应地选取,此方式能够降低BERT模型的计算开销,将输入控制在BERT的限制长度内,同时也能够保留充足的信息,所以使用TFN模型来预处理数据集,增强了BERT的输入文本信息的有效性,使性能进一步提升。

表 5 不同模型的实验结果对比

Tab. 5 Experimental results comparison of different models

文本表示法	算法模型	F1/%					<i>macro-F1</i> /%
		未提及	消极	中性	积极	<i>F1</i> /%	
Word2Vec	GCAE	94.97	61.77	23.63	81.06	65.36	
	ATAE-LSTM	94.76	58.94	19.30	79.59	63.15	
	IAN	94.84	60.37	22.83	79.67	64.43	
	AOA-LSTM	94.92	61.20	21.96	80.81	64.72	
BERT	BERT-Single	95.30	61.34	17.67	82.87	64.30	
	BERT-Pair	95.30	68.63	27.17	82.67	68.44	
	BERT-Pair-ATT	95.38	68.88	27.58	83.00	68.71	
	TFN+BERT-Pair	95.44	68.96	26.98	83.03	68.60	
	TFN+BERT-Pair-ATT	95.42	69.32	28.14	83.19	69.02	

3 结语

本文利用文本筛选和改进BERT算法来处理长文本方面情感分析任务。首先基于原始数据集的标签构造新的数据集用来训练文本筛选模型TFN,其次,利用TFN模型筛选出评论中粗粒度方面相关的部分语句,然后使用句子对方式作为BERT的输入方式,并且在输出层进一步添加注意力层。通过与相关深度学习模型的对比分析,本文模型能进一步提升长文本ABSA任务的性能。

本文通过TFN模型对语句进行筛选,其正确率还需要进一步提升,接下来的工作中将结合主题模型进一步提高长文本处理能力,得到更加准确的文本筛选模型,从而提升整体算法的性能。

参考文献 (References)

- [1] SUN C, HUANG L, QIU X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence [C]// Proceedings of the 2019 Conference of the North American Chapter

- of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2019: 380-385.
- [2] LIU B. Sentiment Analysis and Opinion Mining [M]. Synthesis Lectures on Human Language Technologies. San Rafael, CA: Morgan and Claypool Publishers, 2012: 1-167.
- [3] 陈苹,冯林.情感分析中的方面提取综述[J].计算机应用,2018,38(S2):84-88,96.(CHEN P, FENG L. Review of aspect extraction in sentiment analysis [J]. Journal of Computer Applications, 2018, 38(S2): 84-88, 96.)
- [4] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2016 Task 5: aspect based sentiment analysis [C]// Proceedings of the 10th International Workshop on Semantic Evaluation. Stroudsburg, PA: Association for Computational Linguistics, 2016: 19-30.
- [5] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2015 task 12: aspect based sentiment analysis [C]// Proceedings of the 9th International Workshop on Semantic Evaluation. Stroudsburg, PA: Association for Computational Linguistics, 2015: 486-495.
- [6] WAGNER J, ARORA P, CORTES S, et al. DCU: aspect-based polarity classification for SemEval Task 4 [C]// Proceedings of the 8th International Workshop on Semantic Evaluation. Stroudsburg, PA: Association for Computational Linguistics, 2014: 223-229.
- [7] KIRITCHENKO S, ZHU X, CHERRY C, et al. NRC-Canada-2014: detecting aspects and sentiment in customer reviews [C]// Proceedings of the 8th International Workshop on Semantic Evaluation. Stroudsburg, PA: Association for Computational Linguistics, 2014: 437-442.
- [8] 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. Stroudsburg, PA: Association for Computational Linguistics, 2016: 606-615.
- [9] MA D, LI S, ZHANG X, et al. Interactive attention networks for aspect-level sentiment classification [C]// Proceedings of the 26th International Joint Conferences on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 4068-4074.
- [10] HUANG B, OU Y, CARLEY K M. Aspect level sentiment classification with attention-over-attention neural networks [C]// Proceedings of the 2018 International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, LNCS 10899. Cham: Springer, 2018: 197-206.
- [11] XUE W, LI T. Aspect based sentiment analysis with gated convolutional networks [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2018: 2514-2523.
- [12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [13] PATTANAYAK S. Convolutional neural networks [M]. Pro Deep Learning with TensorFlow. Berkeley, CA: Apress, 2017:

- 153-221.
- [14] DAUPHIN Y N, FAN A, AULI M, et al. Language modeling with gated convolutional networks [C]// Proceedings of the 34th International Conference on Machine Learning. New York: JMLR. org, 2017: 933-941.
- [15] CHEN Z, QIAN T. Transfer capsule network for aspect level sentiment classification [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2019: 547-556.
- [16] DEVLIN J, CHANG M, 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: Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2019: 4171-4186.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc. , 2017: 6000-6010.
- [18] TENNEY I, DAS D, PAVLICK E. BERT rediscovers the classical NLP pipeline [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2019: 4593-4601.
- [19] GAO Z, FENG A, SONG X, et al. Target-dependent sentiment classification with BERT [J]. IEEE Access, 2019, 7: 154290-154299.
- [20] 袁丁,章剑林,吴广建. 基于方面级的餐厅用户评论细粒度情感分析[J]. 软件, 2019, 40(8): 181-189. (YUAN D, ZHANG J L, WU G J. Fine-particle sentiment analysis based on aspect level restaurant customer comment [J]. Computer Engineering and Software, 2019, 40(8): 181-189.)
- [21] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [EB/OL]. [2019-11-12]. <https://arxiv.org/pdf/1301.3781.pdf>.
- [22] 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. New York, NY: Curran Associates Inc. , 2013: 3111-3119.
- [23] NYPASZKE A, GROSS S, MASSA F, et al. PyTorch: an imperative style, high-performance deep learning library [C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. New York, NY: Curran Associates Inc. , 2019: 8024-8035.
- [24] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc. , 2012: 1097-1105.
- WANG Kun**, born in 1996, M. S. candidate. His research interests include natural language processing, sentiment analysis.
- ZHENG Yi**, born in 1993, Ph. D. candidate. His research interests include deep learning, sequential signal processing.
- FANG Shuya**, born in 1995, M. S. candidate. Her research interests include face recognition, object detection.
- LIU Shouyin**, born in 1964, Ph. D., professor. His research interests include wireless communication, internet of things, machine learning.