

引用格式: 邵党国, 胡永健. 基于对抗训练与词性推理的文本情感分析[J]. 陕西理工大学学报(自然科学版), 2023, 39(3): 23-30.

基于对抗训练与词性推理的文本情感分析

邵党国^{1, 2*}, 胡永健¹

1. 昆明理工大学 信息工程与自动化学院, 云南 昆明 650504;
2. 昆明理工大学 云南省人工智能重点实验室, 云南 昆明 650504

摘 要: 针对现有的方面级情感分析方法忽略或没有充分提取句子长距离依赖关系和上下文信息, 以及模型鲁棒性和泛化能力较低等问题, 提出了一种基于投影梯度下降(PGD) 对抗训练和词性推理的方面级情感分析模型 PGDBD。首先利用 PGD 攻击让模型对抗训练, 进行正则化处理, 增强模型的泛化能力; 其次使用 BERT 生成的词向量取代深层金字塔卷积神经网络(DPCNN) 模型本身的词向量, BERT 能够捕捉句子中更长距离的依赖关系; 最后, 通过词性推理层提取方面词附近重要的局部特征, 利用 DPCNN 对全局特征进行提取, 获取更全面的上下文信息; 将局部特征和全局特征进行融合, 构成基于对抗训练和词性推理的情感分类模型。通过在 4 个公共数据集上进行实验, 结果表明该模型的性能相较于其他基线模型有明显的提升。

关键词: 情感分析; 对抗训练; 词性推理; BERT; 深层金字塔卷积神经网络

中图分类号: TP391 **文献标识码:** A **文章编号:** 2096-3998(2023) 03-0023-08

随着在线社交媒体的快速发展, 在线评论的数量呈直线增长, 大部分的网络在线评论包含评论者的情感表达^[1]。方面级情感分析的目的是提取评论句中的每个方面词并判断相应的情感极性^[2]。例如, 对于用户评论语句“这家餐厅的服务很好, 但是味道一般”, 可以判断用户对于“服务”这个方面的情感词是“很好”, 所以情感极性是正向的; 用户对于“味道”这个方面的情感词是“一般”, 所以情感极性是中性的。商家根据用户对各个方面的情感反馈, 可以对相应方面进行改善。同样, 政府部门可以根据群众对于某个热点事件的在线评论进行情感倾向分析, 从而制定相应的策略与方案。通过对用户评论进行方面级情感分析, 可以提取用户的情感信息, 对商家制定营销策略和政府掌控舆论十分重要^[3]。

在方面级情感分析领域, 首先是基于情感词典^[4]的方法, 即利用情感词典获取文本中情感词的情感值, 再通过加权计算来确定文本的整体情感倾向, 此方法没有考虑词与词之间的联系。其次是传统机器学习方法^[5], Vo 等^[6]通过特征抽取的方式判断句子的情感极性, Kirit-chenko 等^[7]提出基于 SVM 的方法进行方面级情感分析。机器学习方法存在依赖人工、特征向量稀疏和特征提取困难等缺点, 因此对情感分析的准确性有一定的影响。

近年来, 研究人员主要采用基于深度学习的方法来解决情感分析任务^[8], 该方法最大的优势就是不依赖人工, 通过构建网络模型对文本逐步分析、特征提取, 从而提高情感分类的准确率。常用的基于深度学习的方法有 CNN^[9]、LSTM^[10]等。VDCNN^[11]是第一个设计良好的字符级深层 CNN 用于 NLP 任务, 展示了深层网络的优势。Tang Duyu 等^[12]基于 LSTM 提出了 TD-LSTM 模型和 TC-LSTM 模型, 该模型利用了方面词和上下文之间的联系, 取得了更好的效果。2018 年谷歌提出 BERT^[13]模型, 它是一种

收稿日期: 2022-11-13 修回日期: 2023-02-12

基金项目: 国家自然科学基金项目(62266025)

* 通信作者: 邵党国(1979—), 男, 陕西咸阳人, 博士, 副教授, 主要研究方向为图像处理、文本处理。

基于双向 Transformer^[14] 的语言模型,可以捕获句子中的隐藏上下文信息和长距离依赖关系。

现有的情感分析模型已经能在情感分类任务中取得不错的效果,但模型的鲁棒性和泛化能力较低。Goodfellow 等^[15]发现对抗训练可以增强模型的鲁棒性。对抗训练的核心思想是使用对抗样本生成方法制作一个对抗样本集合,将对抗样本和原始样本混合放入训练集中,让模型进行对抗训练。针对模型鲁棒性和精确性之间的权衡问题,Zhang Hongyang 等^[16]提出 TRADES 算法来平滑决策边界,通过在损失函数中添加一个正则化损失减少原始样本和对抗样本之间的预测差异。Rice 等^[17]对现有的对抗训练算法进行了分析,研究表明投影梯度下降(Projected Gradient Descent, PGD)对抗训练是目前防御效果最好的对抗训练算法之一。

虽然上述情感分析模型取得了不错的效果,但是模型本身的鲁棒性和泛化能力比较一般,在外部有一定扰动的情況下,可能会导致模型的错误分类。并且现有的方面级情感分析方法不能充分提取上下文特征,无法充分利用文本中的语义信息,对文本中长距离依赖关系的捕获能力较弱。针对这些问题,本文提出一种基于 PGD 对抗训练和词性推理的情感分析模型 PGDBD(Projected Gradient Descent Adversarial Training, BERT and DPCNN)。

1 本文模型

本文提出的 PGDBD 模型整体结构如图 1 所示,主要由以下几部分构成:输入层、特征提取层和输出层。首先,利用 PGD 攻击生成对抗样本,然后将对抗样本与原始样本混合放在训练集中,让模型进行对抗训练,使模型的鲁棒性得到提高;其次, BERT(Bidirectional Encoder Representation from Transformers)层将文本中的每个单词转化成词向量;然后利用词性信息辅助模型调整方面词附近词的权重值,提取与方面词相关的隐藏特征,深层金字塔卷积神经网络(Deep Pyramid Convolutional Neural Networks for Text Categorization, DPCNN)经过多层等长卷积和池化,获得更全面的全局特征信息,然后将局部特征拼接融合到全局特征中。最后经过 Softmax 层进行情感分类。

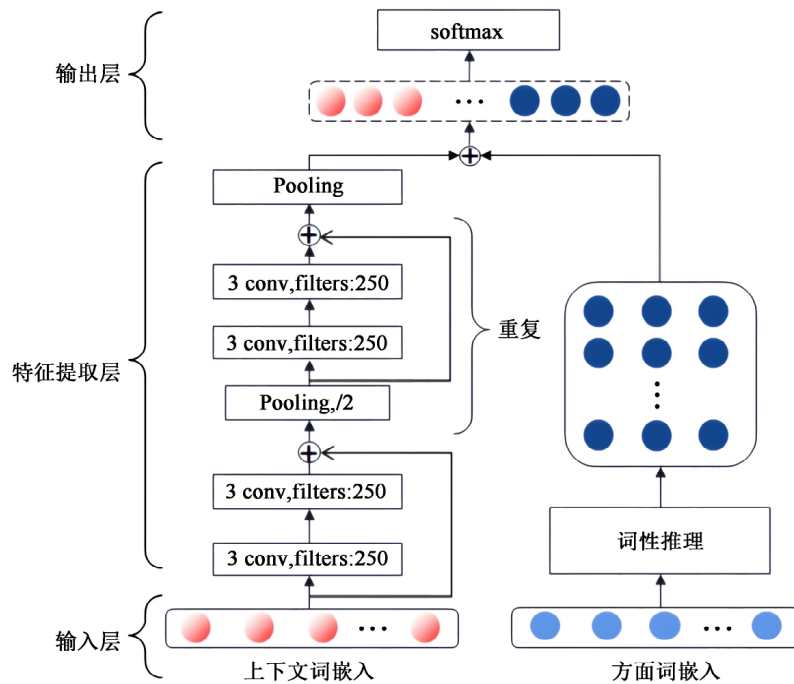


图 1 模型整体结构

1.1 输入层

1.1.1 PGD 对抗训练

对抗训练的核心点是构造扰动,使模型具备识别不同对抗样本的能力。对抗训练算法需要先利用对抗攻击生成扰动,再将扰动与原始样本结合生成对抗样本,然后再通过反向传播过程朝损失函数减小

的方向更新模型参数。这一过程可以被定义为一个最大-最小化的内外层优化问题,最大化问题是寻找使损失函数最大化的扰动来生成对抗样本的过程,而最小化问题则是最小化损失函数和更新模型参数的过程,即让模型具有一定的鲁棒性能够适应这种扰动,具体如下式:

$$\min_{\theta} E_{(x,y) \sim D} [\max_{\delta \in S} L(f_{\theta}(x + \delta), y)], \quad (1)$$

式中 δ 为添加的扰动, S 为扰动的约束空间, x 表示输入的原始样本, y 表示样本的标签, θ 是参数, f_{θ} 是神经网络函数, $L(f_{\theta}(x + \delta), y)$ 表示在输入 x 上加入扰动 δ , 再经过神经网络函数, 与标签 y 比较得到的损失。 $\max_{\delta \in S} L()$ 表示优化目标。

PGD 是一种迭代攻击, 相比于普通的 FGSM (Fast Gradient Sign Method) 和 FGM 仅做一次迭代, PGD 则是做多次迭代, 每次迭代计算得到的对抗扰动投影回指定的约束空间。其具体过程为: 将原始样本输入到模型参数初始化的模型中获得预测标签, 通过约束原始样本的预测标签与已知的真实标签之间的距离迭代地在原始样本中增加扰动, 并保持扰动量始终在设定的阈值范围内, 经过多次扰动迭代后得到对抗样本。PGD 生成的扰动可以表示为

$$\delta^k = \delta^{k-1} + \alpha \cdot \text{sign} [\nabla_x L(f_{\theta}(x + \delta^{k-1}), y)], \quad (2)$$

式中 α 为迭代的步长, k 为迭代的次数, sign 为符号函数, ∇_x 表示关于输入样本 x 的梯度。则 PGD 多次迭代后生成的对抗样本为

$$x^k = \prod_{x \in S} \{x^{k-1} + \alpha \cdot \text{sign} [\nabla_x L(f_{\theta}(x + \delta^{k-1}), y)]\}, \quad (3)$$

式中, $\prod$ 表示裁剪过程, 保证生成对抗样本的扰动量始终保持在设定的阈值范围内, x^k 表示第 k 次迭代后生成的对抗样本, 它取决于前一次扰动迭代后的样本 x^{k-1} 。

1.1.2 词嵌入

本文使用 BERT 生成的词向量取代 DPCNN 模型本身的词向量。在 DPCNN 中原始词向量模型是 region embedding, 但是 region embedding 的本质是使用 n -gram, 并不能充分地理解语义, 并且把词都压缩到低维的语义空间后, 相似意义的词可能出现共线问题, 即拥有相同的词向量, 当词向量模型调整参数时, 相似词的词向量会同时改变。所以本文使用 BERT 词向量来替代 DPCNN 模型原本的词向量。BERT 相比于普通的词向量模型更加有效, 利用 BERT 可以获得与原始词向量长度相同的增强语义向量, 并且能够捕捉更长距离的依赖关系和更全面的上下文信息。PGDBD 模型首先在输入层使用 BERT 将句子中的词转化为词向量, 然后将词向量进行拼接得到向量矩阵 X , 公式为

$$X_{1:n} = f_{\text{concat}}\{x_1, x_2, \dots, x_n\}, \quad (4)$$

式中, f_{concat} 表示拼接操作, x_i 是第 i 个词的词向量。

1.2 特征提取层

1.2.1 词性推理特征提取

本文通过词性信息辅助模型动态调整方面词附近词的权重, 赋予靠近方面词且重要词性的词更高的权重值, 赋予远离方面词且非重要词性的词更低的权重值。首先标注句子中词的词性, 使词向量融合词性信息, 获取隐藏语义信息。然后以方面词为中心, 将每个序列分成方面词区间 ($D_i \leq 0$)、附近区间 ($0 < D_i \leq \varphi$) 和远离区间 ($D_i > \varphi$), 如图 2 和下式所示:

$$D_i = |i - a_c| - \frac{l_a}{2}, \quad (5)$$

式中, i 表示句子中词的位置, a_c 表示方面词的中心位置, l_a 表示方面词的长度, φ 表示附近区间的长度。

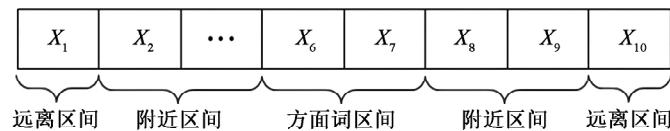


图2 序列划分

方面词区间的词, 权重系数为 1。对于附近区间的词, 如果存在形容词、副词等重要词性的词, 则权

重乘以系数 α 进行权重提升。对于远离方面词区间的远离区间的词,该区间词的权重乘以系数 β 进行权重衰减,具体如下式所示:

$$W_e = \{w_1, w_2, \dots, w_n\}, \quad (6)$$

$$w_i = \begin{cases} X_e, \\ [(1 - \frac{D_i}{n})\alpha] \cdot X_e, & \alpha > 1, \\ [(1 - \frac{D_i}{n})\beta] \cdot X_e, & 0 < \beta < 1, \end{cases} \quad (7)$$

$$L_e = X_e \cdot W_e, \quad (8)$$

式中 X_e 为单位列向量, w_i 为序列权重向量, W_e 为权重矩阵, L_e 为词性推理特征提取层的输出。

1.2.2 DPCNN 全局特征提取

通过 DPCNN 这种架构可以在增加网络层数的同时节约计算成本。DPCNN 首先将离散文本转换为连续文本,然后经过两层的等长卷积,卷积核 O 的大小为 h ,等长卷积可以挖掘句子中隐藏的语义信息。再重复通过 $1/2$ 池化层, $1/2$ 池化可以提高词位的语义,并且将序列长度压缩为原序列的 $1/2$,这样感知到的文本片段是之前的两倍。其中固定了特征图的数量进行下采样,这使网络内部的数据量和计算量逐层递减。随着网络层数的加深,模型可以提取更全面的全局特征信息,因此使用深层的 DPCNN 网络比使用浅层卷积神经网络有更好的效果, DPCNN 提取的特征如下式所示:

$$C_i = f(O \cdot x_{i:i+h-1} + b), \quad (9)$$

$$C = f_{\text{concat}}\{C_1, C_2, \dots, C_n\}, \quad (10)$$

式(9)中 b 为偏差, f 代表非线性变换函数。

1.3 输出层

将上述特征提取层中获得的局部特征与全局特征进行拼接融合。最后将句子表示传递给情感分类层,得到情感极性的分布:

$$P(y | \tilde{C}) = f_{\text{Softmax}}(W\tilde{C} + d), \quad (11)$$

式中, W 和 d 是情感分类层的可学习参数,通过最小化负对数似然来学习模型参数。

2 实验结果与分析

2.1 数据集

为了验证本文方法的有效性,本文在 4 个公共数据集上进行了实验,分别是 SemEval 2014 的笔记本电脑数据集(14Lap)和 SemEval 2014—2016 的餐厅数据集(14Rest、15Rest 和 16Rest),数据集的标签是正向、负向和中性的三分类标签。数据集的详细统计数据见表 1。

表 1 数据集统计

数据集	训练集			验证集			测试集		
	正向	中性	负向	正向	中性	负向	正向	中性	负向
14Rest	1 796	539	666	368	94	139	728	196	196
14Lap	824	383	717	161	72	149	340	167	128
15Rest	808	29	228	147	5	44	340	28	195
16Rest	1 106	54	406	191	9	60	474	29	127

2.2 参数设置

本实验基于 PyCharm 编程平台,使用 Pytorch 深度学习框架来搭建模型。Batch size 大小为 32,学习率为 5×10^{-5} ,最大序列长度为 64,dropout 为 0.1。BERT 采用 Google 的预训练模型,字向量维度为 768。所有实验采用 Adam 来优化模型。

2.3 评价指标

本文使用准确率(Accuracy, Acc)和 F1 值(F1)作为评价指标对模型性能进行评估,其定义为

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$f_{Precision} = \frac{TP}{TP + FP} \quad (13)$$

$$f_{Recall} = \frac{TP}{TP + FN} \quad (14)$$

$$F1 = \frac{2f_{Precision} \cdot f_{Recall}}{f_{Precision} + f_{Recall}} \quad (15)$$

式中, TP 表示正样本被预测为正样本, TN 表示负样本被预测为负样本, FP 表示负样本被预测为正样本, FN 表示正样本被预测为负样本。

2.4 对比方法

为了验证本文提出方法的有效性, 将本文方法与以下方法进行比较。

ATAE-LSTM 是基于 LSTM 的模型, 通过注意力机制获取更重要的上下文信息。

MemNet 是一种多层注意力机制模型, 采用多层注意力机制去捕获句子中词与词之间的关系, 获取更多的有效信息, 提高情感分类的准确率^[18]。

IAN 是一种基于 LSTM 和注意力的模型, 可以交互式地检测目标词和上下文的重要词^[19]。

SA-LSTM-P 采用具有多个正则化器的结构化注意力网络来捕获观点跨度, 基于注意力的 LSTM 捕获上下文信息, 来提高分类性能^[20]。

TNet-LF 能够将特定目标信息整合到单词表示中, 并且有一个 context-preserving 机制能够保留语义信息^[21]。

TNet-ATT 是 TNet 的扩展, 在 TNet 的基础上加入了注意力机制, 它将挖掘的监督信息纳入注意力机制^[22]。

ASCNN 和 ASGCN 分别使用 CNN 和图卷积网络 (GCN) 来捕获词之间长距离依赖关系和句法信息^[23]。

MCRF-SA 是基于多层 CRF 的结构化注意力模型, 它能够提取特定方面的观点跨度, 从而对目标对象的情感极性进行判断^[24]。

BERT 是一种基于双向 Transformer 的语言模型, 能学习句子和方面之间的深层关联。

2.5 结果分析

本文模型和基线模型的实验结果见表 2。

表 2 不同模型的实验结果

模型	%							
	14Rest		14Lap		15Rest		16Rest	
	准确率	F1	准确率	F1	准确率	F1	准确率	F1
ATAE-LSTM	77.20	—	68.70	—	—	—	—	—
MemNet	79.61	69.64	70.64	65.17	77.31	58.28	85.44	65.99
IAN	79.26	70.09	72.05	67.38	78.54	52.65	84.74	55.21
SA-LSTM-P	81.60	—	75.10	—	—	—	88.70	—
TNet-LF	80.42	71.03	74.61	70.14	78.47	59.47	89.07	70.43
TNet-ATT	81.53	72.90	77.62	73.84	—	—	—	—
ASCNN	81.73	73.10	72.62	66.72	78.48	58.90	87.39	64.56
ASGCN	80.86	72.19	74.14	69.24	79.34	60.78	88.69	66.64
MCRF-SA	82.86	73.78	77.64	74.23	80.82	61.59	89.51	75.92
BERT	82.97	73.91	78.74	76.14	86.23	73.02	90.12	76.45
PGDBD	83.52	74.41	80.78	76.90	88.74	74.43	91.42	77.84

由表 2 可知, 本文提出的 PGDBD 模型在 4 个数据集上的实验结果均优于基线模型。ATAE-LSTM 和 SA-LSTM-P 等基于 LSTM 的模型虽然捕获了评论句中的长距离依赖关系, 但忽略了与方面词相关的情感词。基于 CNN 的方法 ASCNN 使用 CNN 来提取特征信息, 但浅层 CNN 在卷积过程中会丢失大量

的重要信息,从而导致模型性能较差。ASGCN 模型使用了图卷积网络(GCN)来挖掘句子中的句法信息,但是实际效果比较一般。IAN、TNET-ATT 等基于注意力的模型忽略了不同方面之间的相互作用,从而导致情感极性的预测准确率较低。相比之下,本文模型使用 BERT 词向量代替 DPCNN 原有的词向量,更加有效。同时,本文方法将词性信息融入词向量中,动态调整序列中词的权重,提取方面词附近具有重要词性词的特征信息。相比于浅层的 CNN 模型,深层神经网络 DPCNN 能更充分地提取全局特征信息。而且本文方法结合了 PGD 对抗训练,让模型的鲁棒性和实际性能更好。PGDBD 模型可以捕捉更长距离的依赖关系,充分提取局部特征信息和全局特征信息,而且模型的泛化能力较强,所以情感分类效果更佳。

2.6 模型参数分析

为了研究 PGD 对抗训练中扰动约束空间 S 的约束半径,即 ϵ 值大小对模型性能的影响,本文分别设置 ϵ 的值为 0.001、0.01、0.1、1、2,在 14Lap 和 15Rest 两个数据集上进行实验。实验结果如图 3 和图 4 所示,从结果可以看出,在训练阶段加入对抗样本,会让模型抵抗攻击的能力有所提升。当 ϵ 值为 0.01 时,模型性能最优;当 ϵ 值过大时,模型的准确率和 F1 值呈下降趋势。产生这种情况的原因可能是添加的扰动值较大,在生成对抗样本时,生成了与原始样本差异较大的样本,但是样本的标签相同,模型识别这些对抗样本的能力下降,从而导致模型性能下降。

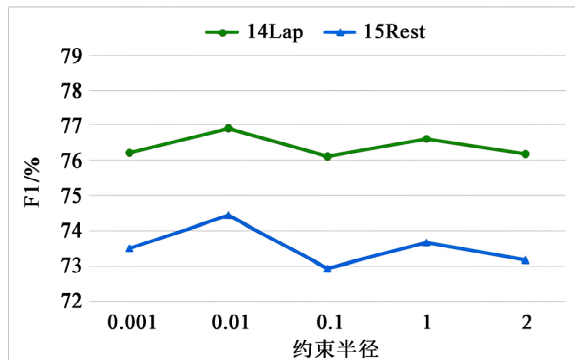
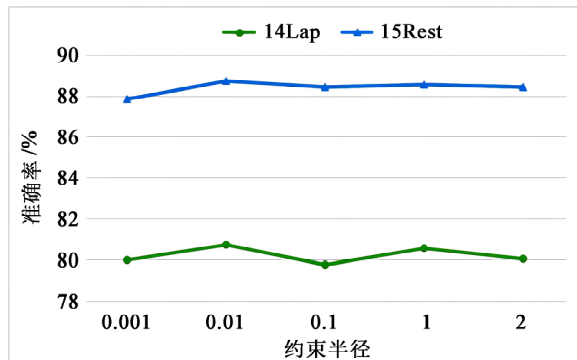


图 3 PGDBD 模型在不同约束半径下的准确率 图 4 PGDBD 模型在不同约束半径下的 F1 值

为了研究词性推理层中附近区间长度大小 φ 对模型性能的影响,本文分别设置 φ 的大小为{1, 2, 3, 4, 5, 6},在 14Lap 和 15Rest 两个数据集上进行实验。实验结果见表 3,从结果可以看出,当 $1 < \varphi < 4$ 时,模型性能随着 φ 增大呈上升趋势,说明此过程中模型对方面词附近重要词的特征提取能力逐渐增强。当 φ 值为 4 时,模型性能达到最优。当 $\varphi > 4$ 时,模型的准确率和 F1 值呈下降趋势,产生这种情况的原因可能是 φ 值过大,会引入冗余的信息和计算量,从而导致模型性能下降。

表 3 序列附近区间长度对模型性能的影响 %

φ	14Lap		15Rest	
	准确率	F1	准确率	F1
1	77.15	75.33	87.11	72.01
2	79.31	75.25	88.34	72.04
3	79.43	75.02	88.03	71.51
4	80.78	76.90	88.74	74.43
5	79.35	74.21	87.05	71.16
6	76.69	74.36	87.68	73.17

2.7 消融实验

为了验证 PGDBD 模型中不同模块对方面级情感分类效果的影响。本文在 14Lap 和 15Rest 数据集上进行消融实验,结果见表 4。从实验结果可以看出,PGD 对抗训练的加入使得模型对对抗样本的识别能力得到提升,增强了模型的泛化能力和实际性能。当不使用 BERT 生成的词向量取代 DPCNN 本身的词向量时,模型的性能明显下降,说明了 BERT 预训练模型能有效提高模型捕获文本语义信息和长距离

依赖关系的能力。当去除词性推理层时,模型无法准确地提取局部位置中与方面词相关的重要特征,导致模型性能下降。从实验结果来看,本文方法的改进是有效的。

表4 消融实验结果

模型	BERT	PGD	词性推理	14Lap		15Rest	
				准确率	F1	准确率	F1
DPCNN	—	—	—	78.02	74.23	86.10	71.74
BERT-DPCNN	✓	—	—	80.10	76.40	88.46	73.89
PGD-DPCNN	—	✓	—	78.36	74.65	86.33	72.12
W-DPCNN	—	—	✓	79.17	75.34	87.12	72.58
PGDBD	✓	✓	✓	80.78	76.90	88.74	74.43

注 “✓”表示加入 PGD 模块,“—”表示未加入该模块。

3 结论

本文提出了一种基于 PGD 对抗训练和词性推理的方面级情感分析模型 PGDBD。该模型在经过 PGD 对抗训练后,模型的鲁棒性和实际性能得到提升。使用 BERT 预训练模型生成的词向量取代 DPCNN 模型本身的词向量,能够捕获文本中的长距离依赖关系,通过词性推理层可以充分提取与方面词相关的重要词的特征,再经过 DPCNN 深层神经网络进行全局特征提取,能够获取更全面的上下文信息,在方面级情感分类任务上更有优势。实验结果表明,本文提出的 PGDBD 模型在情感分类任务上的性能优于现有的基线模型。下一步工作将包括研究其他对抗训练方法对方面级情感分析任务的效果,以及研究对抗训练应用于情感分析领域其他子任务的效果。

[参 考 文 献]

- [1] 王婷,杨文忠.文本情感分析方法研究综述[J].计算机工程与应用,2021,57(12):11-24.
- [2] 李然,林政,林海伦,等.文本情绪分析综述[J].计算机研究与发展,2018,55(1):30-52.
- [3] 张周彬,邵党国,马磊,等.一种循环互作用注意力的属性级情感分类模型[J].计算机应用与软件,2020,37(5):140-144.
- [4] BACCIANELLA S, ESULI A, SEBASTIANI F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining[C]//Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010.
- [5] 何炎祥,孙松涛,牛菲菲,等.用于微博情感分析的一种情感语义增强的深度学习模型[J].计算机学报,2017,40(4):773-790.
- [6] VO D T, ZHANG Yue. Target-dependent twitter sentiment classification with rich automatic features[C]//Twenty-fourth international joint conference on artificial intelligence, 2015.
- [7] KIRITCHENKO S, ZHU Xiaodan, CHERRY C, et al. Detecting aspects and sentiment in customer reviews[C]//Ottawa: 8th International Workshop on Semantic Evaluation (SemEval), 2014: 437-442.
- [8] 文志霄,梁志剑.基于词级和语义级注意力的细粒度情感分析模型[J].中北大学学报(自然科学版),2022,43(5):431-440.
- [9] CHEN Y. Convolutional neural network for sentence classification[D]. Waterloo: University of Waterloo, 2015.
- [10] WANG Xin, LIU Yuanchao, SUN Chengjie, et al. Predicting polarities of tweets by composing word embeddings with long short-term memory[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, 1: 1343-1353.
- [11] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for natural language processing[J]. ArXiv, 2016: 1606.01781.
- [12] TANG Duyu, QIN Bing, FENG Xiaocheng, et al. Effective LSTMs for target-dependent sentiment classification[J]. ArXiv, 2015: 1512.01100.
- [13] DEVLIN J, CHANG Mingwei, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. ArXiv, 2018: 1810.04805.

- [14] VASWANI A ,SHAZEER N ,PARMAR N ,et al. Attention is all you need [J]. NIPS17: Proceedings of the 31st International Conference on Neural Information Processing Systems 2017: 6000-6010.
- [15] GOODFELLOW I J ,SHLENS J ,SZEGEDY C. Explaining and harnessing adversarial examples [J]. ArXiv 2014: 1412. 6572.
- [16] ZHANG Hongyang ,YU Yaodong ,JIAO Jiantao ,et al. Theoretically principled trade-off between robustness and accuracy [C]// International conference on machine learning 2019: 7472-7482.
- [17] RICE L ,WONG E ,KOLTER Z. Overfitting in adversarially robust deep learning [C]//International Conference on Machine Learning 2020: 8093-8104.
- [18] TANG Duyu ,QIN Bing ,LIU Ting. Aspect level sentiment classification with deep memory network [J]. ArXiv 2016: 1605.08900.
- [19] MA Dehong ,LI Sujian ,ZHANG Xiaodong ,et al. Interactive attention networks for aspect-level sentiment classification [J]. ArXiv 2017: 1709.00893.
- [20] WANG Bailin ,LU Wei. Learning latent opinions for aspect-level sentiment classification [C]//New Orleans: Proceedings of the AAAI Conference on Artificial Intelligence 2018.
- [21] LI Xin ,BING Lidong ,LAM W ,et al. Transformation networks for target-oriented sentiment classification [J]. ArXiv 2018: 1805.01086.
- [22] TANG Jialong ,LU Ziyao ,SU Jinsong ,et al. Progressive self-supervised attention learning for aspect-level sentiment analysis [J]. ArXiv 2019: 1906.01213.
- [23] ZHANG Chen ,LI Qiuchi ,SONG Dawei. Aspect-based sentiment classification with aspect-specific graph convolutional networks [J]. ArXiv 2019: 1909.03477.
- [24] XU Lu ,BING Lidong ,LU Wei ,et al. Aspect sentiment classification with aspect-specific opinion spans [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing 2020: 3561-3567.

[责任编辑: 谢 平]

Text sentiment analysis based on adversarial training and part-of-speech-inference

SHAO Dangguo^{1,2} , HU Yongjian¹

1. Faculty of Information Engineering and Automation , Kunming University of Science and Technology ,
Kunming 650504 , China

2. Yunnan Key Laboratory of Artificial Intelligence , Kunming University of Science and Technology , Kunming 650504 , China

Abstract: Aiming at the problems that the existing aspect level sentiment analysis methods ignore or do not fully extract the long-distance dependency and context information of sentences , and the low robustness and generalization ability of the model , this paper proposes an aspect level sentiment analysis model PGDBD based on PGD adversarial training and part-of-speech-inference. First , PGD attack is used to make the model conduct adversarial training , regularize the model , and enhance the generalization ability of the model; Secondly , the word vector generated by BERT is used to replace the word vector of DPCNN model itself , and BERT can capture the longer distance dependencies in sentences; Finally , extracting important local features near aspect words through the part of speech inference layer , and the global features are extracted using the DPCNN deep neural network to obtain more comprehensive context information; The emotion classification model based on adversarial training and part-of-speech-aware is formed by fusing local features and global features. Experiments on four public datasets show that the performance of this model is significantly improved compared with other baseline models.

Key words: sentiment analysis; adversarial training; part-of-speech-inference; bidirectional encoder representation from transformers(BERT) ; deep pyramid convolutional neural networks for text categorization (DPCNN)