

# 结合混淆数据增强和关系图注意网络的 方面级文本情感分析

陈越昆<sup>1</sup>,马宗泽<sup>1</sup>,郭敬松<sup>1</sup>,李剑锋<sup>1</sup>,张云港<sup>2</sup>

(1. 云南中烟工业有限公司 营销中心,云南 昆明 650231; 2. 云南师范大学 信息学院,云南 昆明 650500)

**摘要:** 基于方面的文本情感分析,也称为方面级情感分类,是自然语言处理中的一项常见任务,其目的在于分析给定目标中方面术语的情感极性。基于深度神经网络的模型在文本情感分析的过程中,存在训练样本不足,或只使用注意力机制而忽略了句法关系,不能充分利用方面节点的上下文语义信息等问题,导致效果不佳。针对样本数据量和多样性不足,以及语义语法信息利用度不高所导致的深度学习模型准确率不高的问题,提出一种新的文本情感分类模型,该模型结合混淆(Mixup)数据增强技术和关系图注意神经网络的优势,采用混淆增强对上下文特征向量进行插值混合,使用关系图注意网络集成类型化的句法依赖信息,从而可更有效地利用语义语法树的特征。在3个公开数据集上的实验结果表明,使用混淆数据增强和关系图注意网络的模型与其他方法相比具有较好效果。

**关键词:** 方面级文本情感分析;混淆数据增强;关系图注意网络

**中图分类号:** TP391.1 **文献标志码:** A **文章编号:** 1672-8513(2024)06-0767-08

近年来,电子商务、网络直播购物平台已为人们购物的重要渠道。在各种互联网购物平台上,消费者通常会对某个所购买产品进行评论,这些评论信息对购物平台或商户来说具有重要意义,利用这些评论,可以对其所售卖的产品、客户甚至相关市场有更深入的了解和分析。这些评论的信息处理从自然语言处理的角度来说可归结为基于方面的情感分析(aspect based sentiment analysis, ABSA),也称为方面级情感分类(aspect level sentiment classification, ALSC),目前已成为自然语言处理领域的研究热点之一。

方面级情感分析的目的是判断语句内容所包含的情感极性。例如,有如下的一句话:“from the responding speed to the graphical user interface, this operating system outperforms Linux easily”。在此用户评论中,蕴含两个方面:operating system 和 Linux,这2个方面的情感极性可判断为:前者为正向(积极的),后者为负向(消极的)。通常,方面级情感分析是以实体为中心,并从相关文本中提供更为细粒度的情感意见信息。

在文本情感分析任务中,早期的工作是通过机器学习中的传统方法来实现的,如支持向量机(SVM)、决策树、最大熵等,这些方法都试图通过模型训练来获取与任务相关的知识,但其结果远不能令人满意。如今,随着硬件计算能力的提高,基于深度学习的方法在情感分析任务上明显优于传统的机器学习技术。例如,CNN、LSTM等深层神经网络被用于文本情感分类,与传统方法相比取得了很大的改进。近年来,图神经网络(Graph Neural Network)的研究得到了极大的发展,许多图神经网络例如图卷积神经网络(GCN)<sup>[1]</sup>,图注意网络(GAT)<sup>[2]</sup>被引入到各种自然语言处理任务中。图神经网络的优势在于可以处理依赖树等句法结构<sup>[3-4]</sup>。

采用深度神经网络解决文本情感分析相关问题时,如果训练数据不充分则会引起模型的欠拟合,

收稿日期:2023-11-19.

基金项目:云南中烟工业有限责任公司科技项目(2018QT03).

作者简介:陈越昆(1978-),男,本科,经济师。主要从事卷烟智能化营销方法及应用研究。

通信作者:李剑锋(1977-),男,本科,二级营销师。主要从事烟草智能化营销研究。

而通过人工对大量数据进行标注显然代价很高,而且容易出现标注错误.如能引入自动的数据增强方法应对深度神经网络的训练数据量及数据多样化问题,将在很大程度上缓解模型因学习不足导致的性能下降问题,且可以大大减少数据标注的成本.文献[5]提出了一种基于插值对数据集进行增强的方法混淆(Mixup)数学增强,该方法可将规模较小的数据集动态地扩充为规模较大的数据集,且可在合理范围内适当增加数据扰动,从而可以用于在训练数据缺乏和样本数据近似时仍可保证模型的训练和学习效果.

本文针对混淆数据增强以及关系图注意网络(relational graph attention networks)在训练数据扩展和语法语义学习中的优势,提出一个结合两者的文本情感分析模型(RGAT - Mixup).该模型对上下文的特征向量施加 Mixup 技术,来增强上下文的特征向量信息;利用关系图注意网络(RGAT)将依赖标签特征信息引入到模型中,使模型能够更好地捕捉单词之间的关系.在3个公共数据集上的实验结果表明,该模型在文本情感分析上可均取得较好效果.

## 1 相关研究工作

在基于深度学习的方法引起大多数研究员的注意之前,许多传统的机器学习方法应用于方面级情感分析任务中.大多数传统的机器学习方法都非常依赖高质量的人工特征,然而获取这些特征既耗时又费力.此外,这些传统模型的泛化能力较弱.由于深度学习的普及,传统的机器学习方法在方面级情感分析已很少使用.在深度学习模型中,CNN、RNN<sup>[6]</sup>、LSTM等神经网络首先被应用于方面级情感分析任务中.而后,注意力机制在计算机视觉领域取得成功,其逐渐从视觉领域被推广应用于各类自然语言处理(Natural Language Processing, NLP)任务中.

最近,研究人员提出了预训练模型和“预训练+微调”的范式,并且在各类 NLP 任务中都取得了不错的效果,一些方面级情感分析任务的相关研究工作中开始应用了许多预训练模型.例如, Li 等<sup>[7]</sup>证明,仅仅使用简单的全连接层,在方面级情感分析任务中利用 BERT 预训练模型所得出的效果也是非常有效的. Liao 等<sup>[8]</sup>提出使用基于深度双向 transformer 的 RoBERTa,从文本和方面标签中提取特征,并应用交叉注意力机制使模型更加关注方面词及其相关词.随着知识图谱等概念的提出,一些研究员引入领域知识对文本模型进行微调,例如, Xu 等<sup>[9]</sup>提出了一种后训练方法,该方法引入领域知识,然后对 BERT 进行微调,在文本情感分析上获得了更为精确的结果.

图神经网络(GNN)最早由 Gallicchio<sup>[10]</sup>提出,随后因其良好性能得到了许多学者的关注. GNN 已用于许多 NLP 任务,如机器翻译<sup>[11]</sup>、文本分类<sup>[12]</sup>、关系提取<sup>[13]</sup>等.对于方面级情感分析任务, Zhang 等<sup>[14-15]</sup>提出了 ASGCN 和 PWCN 模型,利用特征或节点距离的依赖树来解决方面级情感分析任务. Wang 等<sup>[16]</sup>提出了 R - GAT 模型,该模型将依赖树的语法特征与节点距离相结合. Bai 等<sup>[17]</sup>提出 RGAT 模型将依赖树目标添加到词向量表示中.施荣华等<sup>[18]</sup>提出了基于图注意力网络的方面级别文本情感分析在 SemEval - 2014 公开数据集上取得了较好的效果.已有的方面级情感分析研究工作中没有将混淆数据增强与图神经网络相结合的相关工作,这促使本文在利用图神经网络深化语义和语法学习的同时,通过混淆增强技术从训练数据中获得更多的特征信息,使文本情感分析模型获得更强的泛化能力.

## 2 所提模型

本文提出的 RGAT - Mixup 模型的组成如图 1 所示,该模型主要由 3 个部分构成.

1) 采用 Mixup 的上下文特征提取; 2) 采用 RGAT 的语句语义依法特征提取; 3) 模型输出模块.

### 2.1 上下文特征提取器模块

RGAT - Mixup 中的上下文特征提取模块工作原理是,将句子对用单词嵌入序列来表示并送入编码器中,再经过全连接层就可以得到上下文特征.模块的核心是采用 Mixup 对生成的语句嵌入进行增强,生成混淆语句样本.

该模块工作原理:模块接收 2 个句子作为输入.句子是包含  $N$  个单词的序列,由  $\mathbf{B}, \mathbf{B} \in \mathbf{R}^{d \times N}$ . 矩阵  $\mathbf{B}$  表示,该矩阵的每一行为该句子中的一个单词.运用如 RoBERTa<sup>[27]</sup>等自然语言理解中的预训练模型,即可得到

矩阵  $B$  的词嵌入表示. 将其输入编码器获得句子的编码. 若预训练模型没有编码器, 则可采用长短时神经网络 LSTM 进行编码; 否则采用预训练模型自带的编码器进行编码.

此后, 采用 Mixup 技术对输入的 2 个句子的嵌入表示做混合增强.

$$\tilde{x} = \lambda F(B^i) + (1 - \lambda) F(B^j). \quad (1)$$

$$\tilde{y} = \lambda y^i + (1 - \lambda) y^j. \quad (2)$$

其中,  $F(\cdot)$  为编码器, 句子嵌入表示为  $F(B)$ ,  $\lambda \in [0, 1]$  为混合系数. 通过 Mixup 增强后可得到混淆样本  $\tilde{x}$  及其标签  $\tilde{y}$ . 混淆样本  $\tilde{x}$  经全连接层获得语句的上下文特征向量.

$$h = \{h_1, \dots, h_n\}. \quad (3)$$

其中,  $\{h_1, \dots, h_n\}$  为输入语句. 对  $h$  进行平均池化, 则可得到训练集上的上下文特征.

$$h_{con} = pool(h_0, h_1, \dots, h_n). \quad (4)$$

## 2.2 语法语义的特征提取

图神经网络是近年来广泛使用的提取语义语法特征的有效工具, 其天然具备分析图/树结构信息的能力. 然而, 大多数图神经网络在提取语法语义特征时未对信息的依赖进行处理. 因此, 本文采用了文献[16]中提出的关系图注意网络 (relational graph attention network, RGAT), 其可将语句间的依赖关系纳入网络学习, 具备了更强的语法语义表示能力.

本文的语法语义特征提取模块分为 3 部分: 1) 语义语法树; 2) 关系图注意网络; 3) 池化.

### (1) 语义语法树

在语义语法树的生成中, 文本采用文献[3]中的 DBP (deep baffle parsing) 解析器. DBP 的基本工作流程为: 首先获取语句间的依存关联种类. 由于关联种类的数量预先无法得到, DBP 对多层感知机 (5) 式进行了改造, 从而使其可以处理不确定类别的分类问题.

$$s_i = Wr_i + b. \quad (5)$$

DBP 的输入为词嵌入 (word embedding) 和位置嵌入 (pos embedding) 的连接向量, 采用双向长短期神经网络 (BiLSTM) 得到输入特征向量的语法特征  $r_i$ , 之后使用两个改造后的多层感知机分别对依存词 (dep) 和核心词 (head) 特征进行编码.

$$h_i^{(arc-dep)} = MLP^{(arc-dep)}(r_i). \quad (6)$$

$$h_j^{(arc-head)} = MLP^{(arc-head)}(r_j). \quad (7)$$

式(6)和式(7)中,  $h_i^{(arc-dep)}$  和  $h_j^{(arc-head)}$  分别表示依存词-弧和核心词-弧的特征. 此后, 弧分数矩阵  $S^{(arc)}$  即可由(8)得到.

$$S^{(arc)} = H^{(arc-head)} U^{(arc)} h_i^{(arc-dep)} + H^{(arc-head)} U^{(arc)}. \quad (8)$$

式(8)中,  $H^{(arc-head)}$  是多层感知机对输入语句词嵌入表示的特征编码后组成的编码矩阵; 矩阵  $U^{(arc)}$  是仿射变换矩阵. 式(8)把句子中的每个词分别当做依存词和核心词, 与剩余词进行内积运算, 从而得到弧矩阵  $S^{(arc)}$ , 进而对词和词之间的依存关系进行判别. 式(9)用于得到依存弧上的标签类别, 其综合了考虑了依存关系的先验概率以及词作为核心词或依存词处于某种依赖关系的后验概率.

$$s_i^{\{label\}} = r_{y_i}^T U^{(1)} r_i + (r_{y_i} \oplus r_i)^T U^{(2)} + b. \quad (9)$$

式(9)中  $r_{y_i}^T U^{(1)} r_i$  是将  $i$  作为依存词、 $y_i$  作为核心词情况下的后验概率,  $(r_{y_i} \oplus r_i)^T U^{(2)}$  是将  $i$  或  $y_i$  作为依存弧的两端的后验概率,  $b$  为偏置项, 其表示依存类别为 label 的先验概率.

使用 DBP 获取了语法语义依赖关系后, 本文采用关系图注意网络提取语法特征.

### (2) 关系图注意网络

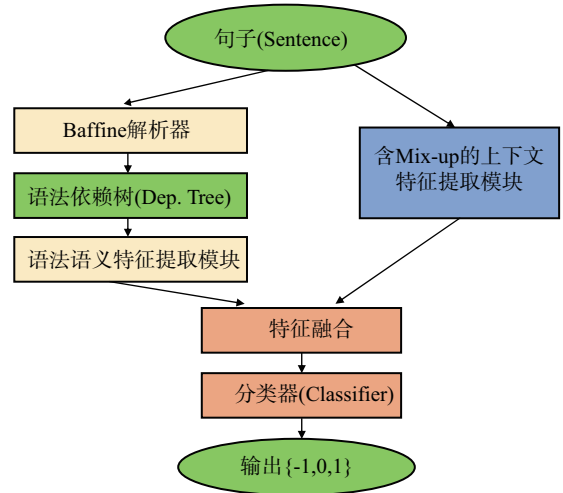


图1 RGAT-Mixup 模型架构图

同前期的图注意网络<sup>[1]</sup>相比较,关系图注意网络(RGAT)能够充分利用词之间的依存关系, RGAT可以把语法关系融合到网络的焦点计算中,从而输出更精确的语法表示.

从DBP获取的依赖树中保存了词语关系,在RGAT的前期研究工作中<sup>[19-20]</sup>,RGAT致力于获取节点感知注意和关系感知注意的概率分布,两者的概率和作为完整的注意力权重.考虑任意图 $G=(V, M, R)$ ,  $V$ 表示节点(单词)集合,  $M$ 为邻接矩阵,  $R$ 为边,即节点间的关系集合.

对于输入输出, RGAT将一组词嵌入表示 $\{x_1, x_2, \dots, x_n\}$ 作为初始隐藏状态,记为 $\{h_1^0, h_2^0, \dots, h_n^0\}$ ,随着RGAT的层数 $l(l \in [1, L])$ 的增加,渐进式迭代生成更高层的语义特征 $\{h_1^l, h_2^l, \dots, h_n^l\}$ ,即第 $i$ 层采用前一层第 $i-1$ 层的特征表示 $\{h_1^{i-1}, h_2^{i-1}, \dots, h_n^{i-1}\}$ 及邻接矩阵 $M$ 作为输入去得到第 $i$ 层的特征表示 $\{h_1^i, h_2^i, \dots, h_n^i\}$ .

对任意一个词 $w_i$ ,其相邻词的索引矩阵为 $N(i)$ ,用 $r_{ij}$ 代表2个词 $w_i$ 与 $w_j$ 之间的关联,关系图注意网络的一大优点为可获取节点感知注意 $s^N$ 和关系感知注意 $s^R$ <sup>[17]</sup>.网络中第 $l$ 层节点感知注意 $s^N$ 的计算如公式(10)所示.

$$s_{ij}^N = \begin{cases} f(h_i^{l-1}, h_j^{l-1}), & j \in N(i); \\ -\inf, & j \notin N(i). \end{cases} \quad (10)$$

公式(10)中, $f(\cdot)$ 是一个注意力权重,这里使用缩放的点积注意力函数<sup>[14]</sup>.

$$f(h_i^{l-1}, h_j^{l-1}) = \frac{(W_Q^l h_i^{l-1})^T (W_K^l h_j^{l-1})}{\sqrt{\frac{d}{Z}}}. \quad (11)$$

公式(11)中, $W_Q^l, W_K^l \in \mathbb{R}^{\frac{d}{Z} \times d}$ 为第 $l$ 层GAT层的 $z^h$ 头的参数矩阵.关系感知的注意权重 $s^R$ 如公式(12)所示.

$$s_{ij}^R = \begin{cases} f(h_i^{l-1}, r_{ij}), & j \in N(i); \\ -\inf, & j \notin N(i). \end{cases} \quad (12)$$

将式(10)和式(12)中得到的节点感知注意和关系感知注意进行融合,得到最终的注意力分数,其计算如式(13)所示.

$$\hat{\alpha}_{ij} = \begin{cases} \frac{\exp(s_{ij}^R + s_{ij}^N)}{\sum_{j' \in N(i)} \exp(s_{ij'}^R + s_{ij'}^N)}, & j' \in N(i); \\ 0, & j' \notin N(i). \end{cases} \quad (13)$$

Bai等<sup>[17]</sup>证明了关系依赖标签在特征聚合过程中也很重要,关系图注意网络采用逐点卷积变换(point-wise convolution transformation, PCT),将语法树依赖标签嵌入特征向量中,其定义如式(14)所示.

$$\text{PCT}(H^l) = \delta(h^l \odot W_{p_1} + b_{p_1}) \odot W_{p_2} + b_{p_2}. \quad (14)$$

公式(14)中, $\delta$ 为ReLU激活函数, $\odot$ 表示卷积运算, $W_{p_1}, b_{p_1}, W_{p_2}, b_{p_2}$ 表示两个卷积核的权值和偏差.关系图神经网络采用多头注意力机制对特征向量进行更新,如式(15)所示.

$$h_i^l = \bigoplus_{k=1}^Z \sigma \left( \sum_{j \in N(i)} \hat{\alpha}_{ij}^k (W_V^k h_j^{l-1} + W_{V_r} r_{ij}) \right). \quad (15)$$

公式(15)中, $\bigoplus$ 表示向量的拼接操作, $Z$ 为注意头的数量, $\sigma$ 为激活函数, $W_V^k \in \mathbb{R}^{\frac{d}{Z} \times d}$ 为 $l$ 层的 $z^h$ 头参数矩阵, $W_{V_r} \in \mathbb{R}^{\frac{d}{Z} \times d}$ 为参数矩阵.本文的模型采用了多层的RGAT,设 $\{\hat{h}_0, \hat{h}_1, \dots, \hat{h}_n\}$ 为模型得到的语义语法特征,采用全局平均池化,则如式(16)所示可获取最终的语义语法特征.

$$h_{syn} = \text{pool}(\hat{h}_0, \hat{h}_1, \dots, \hat{h}_n). \quad (16)$$

### 2.3 模型输出

本文模型的输出模块包含2个组件,第一个组件为特征融合器,其负责对RGAT获取的上下文语义特征及语法语义特征做融合处理;第二个组件则是一个分类器,根据融合的特征给出情感判定类别.

#### (1) 特征融合器

本文采用基于门控机制的特征融合方式,对由(4)式得到的上下文特征和由(16)式得到的语法语义特征进行融合,得到融合特征 $h_f$ ,如式(17)所示.

$$h_f = g \circ h_{syn} + (1 - g) \circ h_{con}. \quad (17)$$

其中,  $\odot$  表示矩阵对应元素相乘.  $g$  为门控函数<sup>[20]</sup>, 其定义为式(18)所示.

$$g = \sigma(W_g[h_{syn}; h_{con}] + b_g). \tag{18}$$

式(18)中,  $W_g, b_g$  为模型参数,  $[h_{syn}; h_{con}]$  为矩阵的连接操作.

(2) 分类器

采用简单的全连接层作为分类器, 对特征融合器输出的融合特征进行分类, 分类器的输出为语句属于某种情感分类的可能性, 其计算如式(19)所示.

$$P(y = c) = \frac{\exp(Wh_f + b)_c}{\sum_{c' \in C} \exp(Wh_f + b)_{c'}}. \tag{19}$$

公式(19)中,  $W$  和  $b$  为可调模型参数. 给定包含  $N$  个训练样本的训练集, 模型的损失函数见式(20).

$$l = - \sum_{i=1}^N \sum_{c \in C} I(y = c) \log(P(y = c)) + \lambda \|\theta\|^2. \tag{20}$$

其中,  $I(\cdot)$  是指示函数,  $\theta$  为模型参数, 可看出, 模型损失函数采用了交叉熵损失及 L2 正则化.

### 3 实验及结果

#### 3.1 数据集

本文实验在 3 个公开的英文数据集上进行, 具体为: SemEval 2014 – task4 Restaurant (REST14)<sup>[21]</sup>, SemEval 2014 – task4 Laptop (Laptop14)<sup>[21]</sup>, 以及 Twitter 评论数据集<sup>[22]</sup>. 3 个数据集均将文本情绪划分为 3 类: 正面、中性及负面. 表 1 列出了 3 个实验用数据集中样本的分布情况.

表 1 实验用数据集中各类样本的分布情况

数据集	Training set			Testing set		
	正类	中性	负类	正类	中性	负类
Rest14 <sup>[21]</sup>	2 164	637	807	728	196	196
Laptop14 <sup>[21]</sup>	994	464	870	341	169	128
Twitter <sup>[22]</sup>	1 561	3127	1 560	173	346	173

#### 3.2 实验环境及设置

本文实验用工作站主要硬件配置为: 一块 Intel i7 – 9700K CPU, 64GB 内存, 一块 RTX 2080Ti GPU. 实验采用 PyTorch 为主要的代码平台, 采用分类准确率和 Micro – F1 分数作为实验评估指标.

本文模型在实验中对上下文特征提取模块实验了 2 种方式, 一种为采用额外的 LSTM 的为编码器的 GloVe 模型(命名为: RGAT – Mixup – LG), 另一种为直接采用预训练好的 BERT 的提取模型, 其已经内置了编码器(命名为: RGAT – Mixup – BERT), 2 种上下文特征提取模块实验中的主要参数可见表 2.

模型的输出模块中编码器接收的输入按语法语义特征提取模块不同方法向量维数可取 200 或 868, 输出维数为 3 的结果向量. 整个模型训练时批大小设置为 16, epoch 设置为 20.

表 2 模型上下文特征提取模块主要参数

模块/参数	RGAT – Mixup – LG	RGAT – Mixup – BERT
丢弃率(dropout)/%	50	50
惩罚项因子 $\lambda$	0.2	0.2
GloVe 向量维数	300	—
网络优化器	Adam	Adam
训练学习率	0.000 2	0.000 2
输入向量维数	—	768
LSTM 向量维数	100	—

#### 3.3 比较实验及结果

本文模型与多个现有的文本情感分类方法进行了比较, 实验结果可见表 3.

从表 3 可以看出, 本文模型通过关系图注意网络引入了语法语义依存关系, 与其他模型相比, 在 2 个实验指标上都有不同程度的效果提升, 尤其与同样采用关系图神经网络

络的 RGAT - BERT 相比较有优势,说明本文采用混淆 (Mixup) 数据增强可使模型学习的有效性增加,从而提升模型的泛化能力. 与对抗学习增强<sup>[16]</sup>相比,本文的混淆数据增强获得了更好的性能. 同时,从表 3 的结果还可以看出,使用大规模预训练模型如 BERT 的模型能得到较好的实验结果,证明预训练模型对文本情感分类性能有明显提升作用.

表 3 不同模型在 3 个数据集上的实验结果 (精度和 Micro - F1 分数)

比较模型	Rest14 <sup>[21]</sup>		Laptop <sup>[21]</sup>		Twitter <sup>[22]</sup>	
	精度/%	Micro - F1	精度/%	Micro - F1	精度/%	Micro - F1
SVM <sup>[17]</sup>	80.44	—	70.74	—	63.75	63.79
LSTM <sup>[23]</sup>	78.61	67.78	69.74	63.54	76.64	68.02
AOA <sup>[24]</sup>	80.28	70.65	72.66	67.85	72.32	70.66
ASGCN <sup>[14]</sup>	81.27	72.33	75.6	71.55	72.46	70.55
BAT <sup>[25]</sup>	86.03	79.24	79.35	76.50	N/A	N/A
RGAT <sup>[16]</sup>	83.73	76.16	78.36	74.11	75.63	74.35
RGAT - BERT <sup>[16]</sup>	86.11	81.23	80.97	78.42	76.46	75.73
RGAT - Mixup - LG	86.48	80.31	79.34	79.34	75.92	75.05
RGAT - Mixup - BERT	87.04	81.49	81.43	79.65	76.80	75.89

### 3.4 数据增强对模型性能的影响评估

在实际应用中,时常难以获得大量的训练数据,或因为计算条件限制而必须丢弃部分训练数据,当训练数据不足时有可能导致模型无法得到充足训练. 数据增强技术是一类可以用来扩展训练数据的方法,本文使用的混淆数据增强对文本情感分类的效果在前述实验中已得到验证. 本节将进一步验证在使用部分训练数据情况下对模型性能的影响. 实验中,从 3 个数据集按 10% 的间隔随机选取 10% 到 100% 的数据集,用每个比例的数据集分别进行模型训练,每个模型的测试集均一致使用原数据集的测试数据. 实验中对比了使用/不使用混淆增强的两类模型,每类模型包含分别使用/不使用 BERT 预训练模型 2 个模型.

图 2 ~ 4 分别给出了所评估模型在 3 个数据集上的实验结果. 从图中结果可看出,当训练数据量较少时,如低于 50%,相比于不使用数据增强的模型,使用数据增强可显著提升模型性能,即用较少的训练数据带来了较大的模型提升. 随着训练数据量的加大,提升幅度逐渐降低,模型性能整体趋势为随训练集的增大而提高. 此外,从实验结果还可以得出和前述对比实验相同的一个结论,即使用预训练模型可有效提升文本情感分类结果.

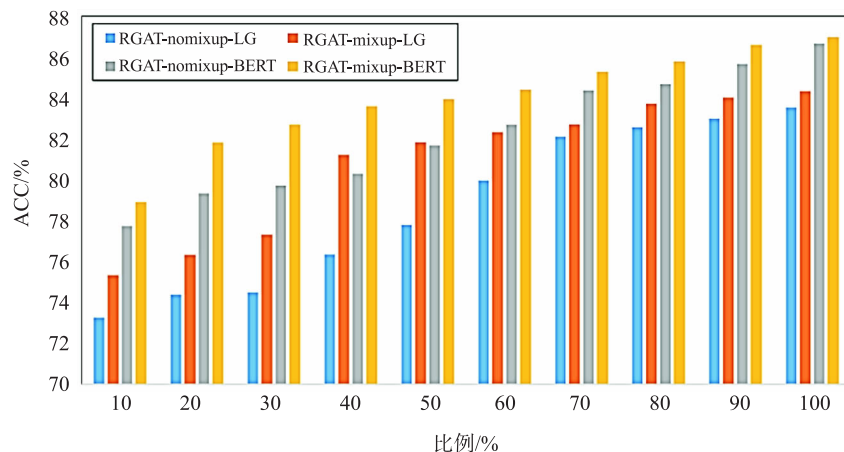


图 2 不同训练集规模下在 REST 数据集的分类结果

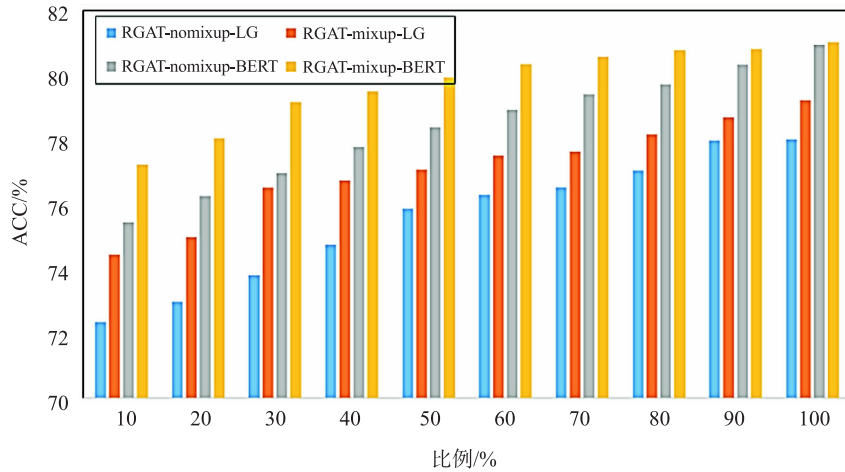


图 3 不同训练集规模下 LAPTOP 数据集的分类结果

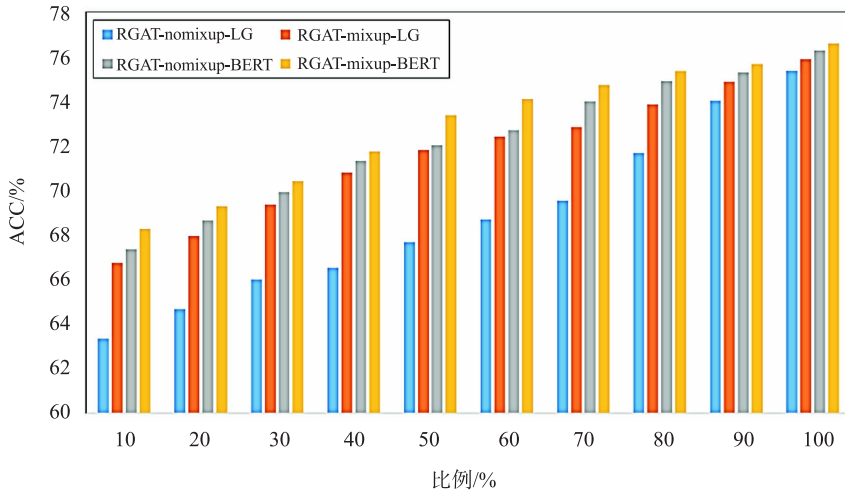


图 4 不同训练集规模下在 Twitter 数据集的分类结果

### 3.5 预训练模型对模型性能的影响评估

大规模文本预训练模型的有效性已在以往诸多自然语言处理相关应用中得到了验证,本文实验评估了常见预训练模型对本文方法的影响,实验中选取了 BERT<sup>[26]</sup>, RoBERTa<sup>[27]</sup> 和 ERNIE<sup>[27]</sup> 3 个大规模预训练模型分别加入本文模型与不包含预训练模型的方法(RM - LG)进行评估比较.表 4 中,模型名称中的 RM 为 RGAT - Mixup 的缩写.

表 4 预训练模型对 RAT - Mixup 模型的性能影响评估结果

预训练模型	Rest14 <sup>[21]</sup>		Laptop <sup>[21]</sup>		Twitter <sup>[22]</sup>	
	精度/%	Micro - F1	精度/%	Micro - F1	精度/%	Micro - F1
RM - LG	86.48	80.31	79.34	79.34	75.92	75.05
RM - BERT	87.04	81.49	81.43	79.65	76.80	75.89
RM - RoBERTa	87.13	81.98	81.54	79.72	76.87	75.95
RM - ERNIE	86.43	81.27	80.96	79.03	76.42	75.39

由表 4 可看出,嵌入预训练模型的 3 个模型 RM - BERT, RM - RoBERTa, RM - ERNIE 在实验数据集上的性能均优于不使用预训练模型的 RM - LG,这验证了大规模预训练模型对于提升本文模型性能是有效的.采用 ERNIE 的模型性能相比 BERT 和 RoBERTa 稍低,推测是由于该模型训练数据中中文样本较多,而本文所用数据集为英文文本,因此该模型表现稍差.

## 4 结语

针对文本情感分类任务,提出结合混淆数据增强和关系图注意网络的模型,混淆样本用于改善模型的训练效果,同时,关系图注意网络增强了对词语依赖关系的表示能力.实验证明,所提方法能提升文本情感分类效果,尤其在训练数据量较小的情况下,模型提升能力更为显著.大规模预训练模型的加入,可更进一步带来性能的改善.未来将关注于模型在实际卷烟营销环境中的推广使用,同时,将针对中文语境继续研究语句中语法语义特征的提取方法.

### 参考文献:

- [1] LI Q, WU X M, LIU H, et al. Label efficient semi-supervised learning *via* graph filtering[C]//IEEE/CVF CVPR. 2019:9582-9591.
- [2] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//ICLR. 2018:1-12.
- [3] DOZAT T, MANNING C D. Deep biaffine attention for neural dependency parsing[C]//ICLR. 2017:1-8.
- [4] KARIMI A, ROSSI L, PRATI A. Adversarial training for aspect-based sentiment analysis with BERT[C]//IEEE ICPR. 2021:8797-8803.
- [5] ZHANG H, CISSE M, DAUPHIN Y N, et al. Mixup:beyond empirical risk minimization[C]//ICLR. 2018:1-13.
- [6] LIU P, JOTY S, MENG H. Fine-grained opinion mining with recurrent neural networks and word embeddings[C]//EMNLP. 2015:1433-1443.
- [7] LI Z, WEI Y, ZHANG Y, et al. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification[C]//AAAI. 2019,33(1):4253-4260.
- [8] LIAO W, ZENG B, YIN X, et al. An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa[J]. Applied Intelligence, 2021, 51(6):3522-3533.
- [9] XU H, LIU B, SHU L, et al. BERT post-training for review reading comprehension and aspect-based sentiment analysis[C]//NAACL-HLT. 2019:2324-2335.
- [10] GALLICCHIO C, MICHELI A. Graph echo state networks[C]//IEEE IJCNN. 2010:1-8.
- [11] BASTINGS J, TITOV I, AZIZ W, et al. Graph convolutional encoders for syntax-aware neural machine translation[C]//EMNLP. 2017:1957-1967.
- [12] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification[C]//AAAI. 2019,33(1):7370-7377.
- [13] LI Y, JIN R, LUO Y. Classifying relations in clinical narratives using segment graph convolutional and recurrent neural networks (Seg-GCRNs)[J]. Journal of the American Medical Informatics Association, 2019, 26(3):262-268.
- [14] ZHANG C, LI Q, and SONG D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. [C]//EMNLP. 2019:4568-4578.
- [15] ZHANG C, LI Q, SONG D. Syntax-aware aspect-level sentiment classification with proximity-weighted convolution network [C]//ACM SIGIR. 2019:1145-1148.
- [16] WANG K, SHEN W, YANG Y, et al. Relational graph attention network for aspect-based sentiment analysis[C]//ACL. 2020:3229-3238.
- [17] BAI X, LIU P, ZHANG Y. Investigating typed syntactic dependencies for targeted sentiment classification using graph attention neural network[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 29:503-514.
- [18] 施荣华, 金鑫, 胡超. 基于图注意力网络的方面级别文本情感分析[J]. 计算机工程, 2022, 48(2):34-39.
- [19] SHAW P, USZKOREIT J, VASWANI A. Self-attention with relative position representations[C]//NAACL-HLT. 2018:464-468.
- [20] CHO K, VAN M B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//EMNLP. 2014:1724-1734.
- [21] PONTIKI M, GALANIS D, PAVLOPOULOS J, et al. SemEval-2014 task 4: aspect based sentiment analysis[C]//International Workshop on Semantic Evaluation. 2014:27-35.
- [22] JIANG L, YU M, ZHOU M, et al. Target-dependent Twitter sentiment classification[C]//ACL-HLT. 2011:151-160.
- [23] TANG D, QIN B, FENG X, et al. Effective LSTMs for target-dependent sentiment classification[C]//COLING. 2016:3298-3307.

- [24] HUANG B,OU Y,CARLEY K M. Aspect level sentiment classification with attention – over – attention neural networks[C]//Social, Cultural, and Behavioral Modeling: 11th International Conference. 2018:197 – 206.
- [25] CHEN M,WU W,ZHANG Y,et al. Combining adversarial training and relational graph attention network for aspect – based sentiment analysis with BERT[C]//IEEE CISP – BMEI. 2021:1 – 6.
- [26] DEVLIN J,CHANG M W,LEE K,et al. BERT: pre – training of deep bidirectional transformers for language understanding [C]//NAACL – HLT. 2019:4171 – 4186.
- [27] LIU Y,OTT M,GOYAL N,et al. Roberta:A robustly optimized BERT pretraining approach[C]//ICLR. 2020:1 – 15.
- [28] SUN Y, WANG S, LI Y,et al. ERNIE: Enhanced representation through knowledge integration[J]. 2019. DOI:10.48550/arXiv.1904.09223.

## Integrating mix-up data augmentation and relational graph attention network for text sentiment analysis

CHEN Yue-kun<sup>1</sup>, MA Zong-ze<sup>1</sup>, GUO Jing-song<sup>1</sup>, LI Jian-feng<sup>1</sup>, ZHANG Yun-gang<sup>2</sup>

(1. Marketing center of China Tobacco Yunnan Industrial Co., Ltd., Kunming 650231, China;

2. School of Information Science, Yunnan Normal University, Kunming 650500, China)

**Abstract:** Aspect – based sentiment analysis ( ABSA ), also called aspect level sentiment classification ( ALSC ), is a com – mon task in natural language processing ( NLP ). In the ABSA task, a new ABSA model is proposed to solve the problem that the accuracy of the deep learning model is not high due to the lack of sample data and the low utilization of semantic syntax trees. The model combines the mix – up data enhancement technology with the relational graph attention network ( RGAT ). The model uses mix – up technology to interpolate the context feature vectors, and uses RGAT to integrate typed syntactic dependency information, so as to make more effective use of the features of the syntax tree. The experimental results on three public datasets show that the model combined with Mixup and RGAT is competitive with other state of the art methods.

**Key words:** aspected – based sentiment analysis; mix – up data augmentation; relational graph attention network

(责任编辑 段 鹏)

---

### 版权声明

本刊已加入清华光盘版(中国知网)、超星期刊出版平台、万方数据——数字化期刊群、重庆维普中文科技期刊数据库、华艺线上图书馆、中国科技论文在线、长江文库等多家权威数据库,凡在本刊发表文章的同时会被以上网站转载,作者向本刊所投稿件,均视为愿意进入上述各期刊数据库,作者著作权使用费已含在本刊所付的稿酬中,凡有异议者,请另投他刊或特别声明.

本刊编辑部