

基于领域 BERT 模型的服务文本分类方法

闫云飞, 孙鹏, 张杰勇, 马钰棠, 赵亮

(空军工程大学信息与导航学院, 西安, 710077)

摘要 针对 BERT 模型领域适应能力较差, 无法解决训练数据类别数量不均衡和分类难易不均衡等问题, 提出一种基于 WBBi 模型的服务文本分类方法。首先通过 TF-IDF 算法提取领域语料中的词汇扩展 BERT 词表, 提升了 BERT 模型的领域适应性; 其次, 通过建立的 BERT-BiLSTM 模型实现服务文本分类; 最后, 针对数据集的类别数量不均衡和分类难易不均衡问题, 在传统焦点损失函数的基础上提出了一种可以根据样本不均衡性特点动态调整的变焦损失函数。为了验证 WBBi 模型的性能, 在互联网获取的真实数据集上进行了大量对比试验, 实验结果表明: WBBi 模型与通用文本分类模型 TextCNN、BiLSTM-attention、RCNN、Transformer 相比 Macro-F1 值分别提高了 4.29%、6.59%、5.3% 和 43%; 与基于 BERT 的文本分类模型 BERT-CNN、BERT-DPCNN 相比, WBBi 模型具有更快的收敛速度和更好的分类效果。

关键词 服务分类; 文本分类; BERT 模型; 双向长短时记忆网络(BiLSTM); 焦点损失函数

DOI 10.3969/j.issn.2097-1915.2023.001.015

中图分类号 TP391 **文献标志码** A **文章编号** 2097-1915(2023)01-0103-09

A Service Text Classification Method Based on Domain BERT Model

YAN Yunfei, SUN Peng, ZHANG Jieyong, MA Yutang, ZHAO Liang

(Information and Navigation School, Air Force Engineering University, Xi'an 710077, China)

Abstract Aimed at the problems that BERT model is poor in domain adaptability, and unable to cope with the problems of uneven number of training data categories and unbalanced classification difficulty, a service text classification method is proposed based on WBBi model. Firstly, the domain adaptability to the BERT model is improved by extending the BERT word list by extracting words from the domain corpus through the TF-IDF algorithm. Secondly, the service text classification is achieved by the established BERT-BiLSTM model. Finally, in view of the problems of unbalanced number of categories and unbalanced classification difficulty of the dataset, a zoom loss function is proposed which can be dynamically adjusted according to the characteristics of sample unbalance on the basis of the traditional focus loss function. In order to verify the performance of the WBBi model, a large number of comparative experiments are conducted on real datasets obtained from the Internet, and their experimental results show that the WBBi model improves the Macro-F1 values by 4.29%, 6.59%, 5.3%, and 43% respectively in comparison with the generic text classification models TextCNN, BiLSTM-attention, RCNN, and Transformer. Compared with

收稿日期: 2022-06-21

作者简介: 闫云飞(1999-), 男, 内蒙古包头人, 硕士生, 研究方向为服务计算方面研究。E-mail: yyf435382221@163.com

引用格式: 闫云飞, 孙鹏, 张杰勇, 等. 基于领域 BERT 模型的服务文本分类方法[J]. 空军工程大学学报, 2023, 24(1): 103-111. YAN Yunfei, SUN Peng, ZHANG Jieyong, et al. A Service Text Classification Method Based on Domain BERT Model[J]. Journal of Air Force Engineering University, 2023, 24(1): 103-111.

the BERT-based text classification models BERT-CNN and BERT-DPCNN, the WBBI model goes further at convergence rate and classifies still better results.

Key words service classification; text classification; bidirectional encoder representations from transformers (BERT) model; bidirectional long short-term memory networks (BiLSTM); focal loss

随着 SOA 架构的不断发展,网络平台中的服务数量出现爆炸式增长,服务管理的难度也日益增大,服务分类是服务管理的重要方法。通过对服务进行分类管理,可以快速依据服务类别实现服务检索和服务发现,从而保证面向服务架构(service-oriented architecture, SOA)中服务的高效利用。其中,服务文本分类是服务分类的实现方法之一,由于服务描述文本具有丰富的语义信息,且易于编辑和提取,服务文本分类目前已经成为服务分类领域的研究热点^[1]。

目前,服务文本分类方法可以分为基于特征工程的方法和基于深度学习的方法。基于特征工程的方法常通过人工的特征工程,产生文本特征向量,并通过分类算法实现文本分类。但是基于特征工程的分类效果直接取决于特征工程的效果,对于人工特征工程处理经验要求较高,难以快速达到最佳分类水平^[2-4]。基于深度学习的方法可以自动从数据中提取特征,无需人工提取与选择,操作难度大大降低,更加满足当前文本分类领域的研究需求^[5-7]。其中 BERT^[8] (bidirectional encoder representations from transformers, BERT)模型的提出为文本分类领域带来了里程碑式的变革,其拥有的多头注意力机制和便利的通用框架使得模型文本分类能力相比传统神经网络提升显著。目前,采用 BERT 预训练模型进行文本分类正逐渐成为研究热点^[9-11]。

然而,当前基于 BERT 模型的文本分类方法存在 2 个问题。

1) 基于 BERT 模型的文本分类算法常依据模型自带的词表进行词向量生成。尽管模型自带词表与模型契合度较高,但缺乏与相关专业领域词汇融合后进一步提升模型分类精度的考虑。

2) 针对服务文本语料库的不均衡性问题考虑不足。分类数据集常出现数据不均衡现象,传统数据均衡方法常基于采样方式调整、类别权重调整等方法进行数据均衡。没有从类别数量和样本分类难易性方面综合考虑。

为解决上述问题,本文提出了一种新的服务文

本分类框架。该框架分主要包括 3 个步骤:首先提取 Web 服务领域专有词汇加入 BERT 模型原始词表,扩充词表在 Web 服务领域词汇中的覆盖范围;其次,将服务文本语料输入词表扩充后的 BERT 模型进行训练;最后,依据服务数据集的特性和分类结果,选择最佳的损失函数以均衡数据集。为了在真实环境中检验模型性能,本文通过爬虫算法,在网络中获取了 437 类,共 22 205 个服务描述文本作为本文的实验数据集。在实验过程中,本文将提出的方法与现有的多个深度学习方法进行对比,结果表明本文提出的模型精确度更佳。

本文的主要贡献如下:

1) 设计了一种服务文本分类模型 WBBI(web services-based BERT-BiLSTM infrastructure, WBBI)。与传统服务文本分类方法相比,其优点在于拥有更高的领域适应性和针对数据集不均衡特点的适应性。

2) 提出了一种基于领域词汇增强的 BERT-BiLSTM 模型,该模型采用 TF-IDF 算法获取领域语料中的领域词汇并扩展 BERT 模型的原始词表,最后利用 BERT 模型生成的词向量结合 BiLSTM^[12] (bidirectional long short-term memory networks, BiLSTM)模型的上下文特征捕捉能力实现服务文本分类。

3) 提出了一种可以根据数据集不均衡性动态调整的变焦损失函数。其优势在于综合考虑了数据的类别不均衡和难易不均衡,相较于传统损失函数具有更强的灵活性。

4) 为了验证所提出方法的有效性,在互联网爬取的真实数据集上进行了大量实验。实验结果证明,相比于基线模型,本文的模型分类效果更佳。

1 WBBI 服务文本分类模型

WBBI 模型共分为 3 个部分:领域词汇提取、模型训练、变焦损失函数优化。WBBI 模型结构如图 1 所示。

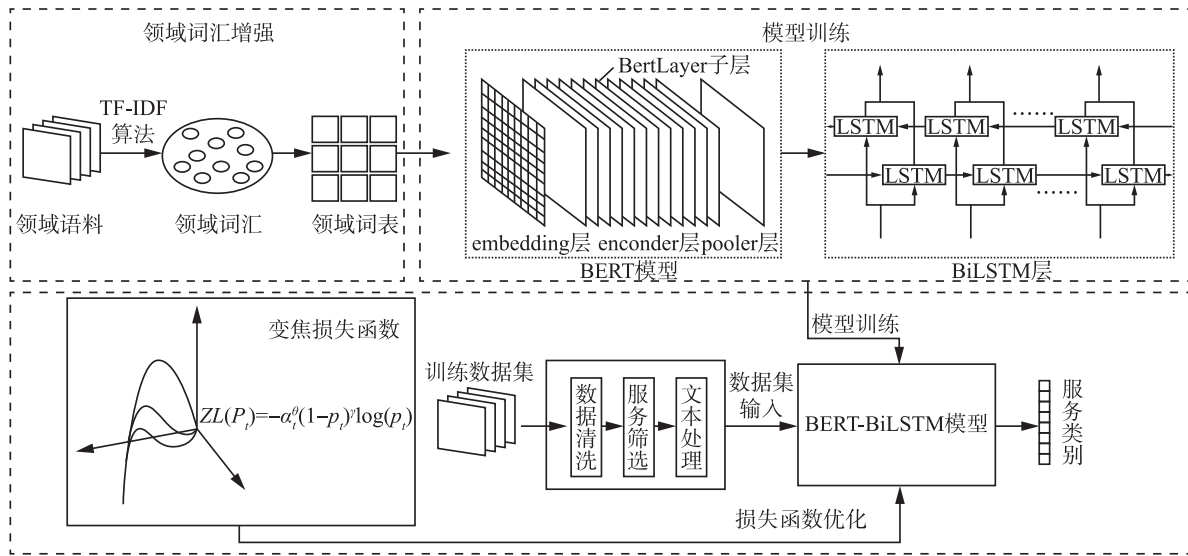


图 1 WBBi 模型结构

1.1 领域词汇增强

为获取领域词汇,本文采用词频-逆文档频率 (term frequency-inverse document frequency, TF-IDF)算法^[13]进行词汇提取。该算法基本思想是:若词汇在一篇文章中具有较高的出现频率,而在其他文章中很少出现,则认为该词汇具有很好的代表能力。其计算公式如下:

$$TF-IDF_{i,j} = TF_{i,j} \log\left(\frac{N}{n_i}\right) \quad (1)$$

式中: $TF-IDF_{i,j}$ 表示词汇 i 在文章 j 中的词频-逆文档频率; $TF_{i,j}$ 代表词汇 i 在文章 j 中出现的频率; N 表示语料库中文章的总个数; n_i 表示语料库中含有词汇 i 的文章个数。

本文将所爬取数据库中的文章进行分词,去除标点符号,统一小写处理,依据 TF-IDF 算法,计算所有词汇的权值,并进行排序。本文采用的 BERT 模型是基于谷歌公司公开发布的 BERT-Base-uncased 模型,其词表包含 28 996 个词汇,并提供了

101 个占位符 [unused]用以扩充词表。本文选取了词汇权值最大的前 101 个词汇对原始词表进行填充,最终得到扩充后的领域词表。

1.2 BERT-BiLSTM 模型结构

如图 1, BERT-BiLSTM 模型由 BERT 模型和 BiLSTM 模型组合而成,其中 BERT 模型通过嵌入层生成文本词向量,通过 encoder 层中的多头注意力机制和前馈神经网络层捕捉文本词汇特征,最后通过 pooler 层中的全连接层进入 BiLSTM 层, BiLSTM 层负责获取词向量的之间的上下文特征,最后通过全连接层进行分类。

1.2.1 BERT 模型输入

模型获得文本句子后,将句子中的单词转化为嵌入 (embedding)。BERT 的输入由字嵌入 (token embeddings)、段嵌入 (segment embeddings) 和位置嵌入 (position embeddings) 组成。其中,字嵌入代表文本单词的嵌入,其依照设定的领域词表按照贪心原则进行词汇匹配。句子输入方法见图 2。

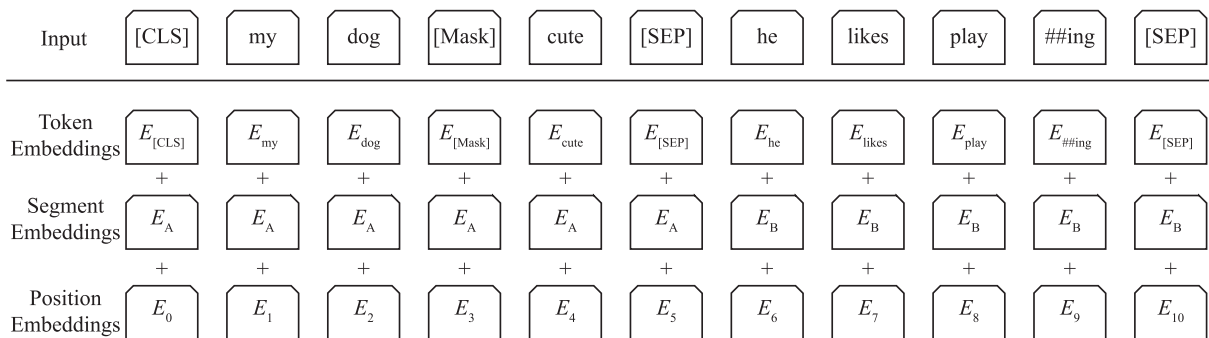


图 2 BERT 模型输入表示

1.2.2 Transformer Encoder 结构

BERT 模型中的 encoder 层由 transformer encoder 结构组成^[14],其基本单元结构如图 3 所示。

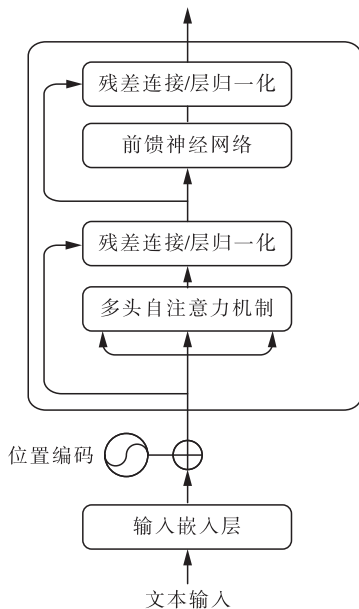


图 3 transformer encoder 单元结构

其中多头注意力机制层主要通过权重的更新来强调文本中的不同部分。单头注意力机制计算如下：

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (2)$$

式中:矩阵 \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 分别代表 Query、Key、Value 矩阵。 $\sqrt{d_k}$ 目的是将注意力矩阵转化为标准正态分布。

而多头注意力机制则为单头注意力机制的并行运算,其计算方式为:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) =$$

$$\text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)\mathbf{W} \quad (3)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}\mathbf{Q}_i, \mathbf{K}\mathbf{W}\mathbf{K}_i, \mathbf{V}\mathbf{W}\mathbf{V}_i) \quad (4)$$

式中: $\mathbf{W}\mathbf{Q}_i$ 、 $\mathbf{W}\mathbf{K}_i$ 、 $\mathbf{W}\mathbf{V}_i$ 、 \mathbf{W} 为系数矩阵; head_i 表示第 i 头的注意力结果。

残差连接和归一化层的功能是将上一层的输入和输出相加并归一化为标准正态分布。其目的为防止深层网络中存在的梯度消失问题。其计算公式如下:

$$\text{output} = \text{LayerNorm}(\text{input} + \text{Sublayer}(\text{input})) \quad (5)$$

式中: input 表示输入数据; output 表示残差连接和归一化层的输出; Sublayer 表示 input 输入的对应该网络层; LayerNorm 表示数据归一化操作。

前馈网络层是将输入数据通过两层线性映射和激活函数处理以提高网络的非线性拟合能力。其计算公式为:

$$\text{output} = \text{Relu}(\text{input} \cdot \mathbf{W}_1 \cdot \mathbf{W}_2) \quad (6)$$

式中: input 、 output 分别表示模型的输入和输出; Relu ^[15] 代表网络所采用的激活函数; \mathbf{W}_1 、 \mathbf{W}_2 分别为线性映射的权值矩阵。

1.2.3 BiLSTM 层

长短期记忆网络(long short-term memory networks, LSTM)模型是在循环神经网络(recurrent neural network, RNN)的基础上,为解决传统 RNN 模型出现的梯度消失和梯度爆炸问题,而产生的改进结构。LSTM 单元结构如图 4 所示。

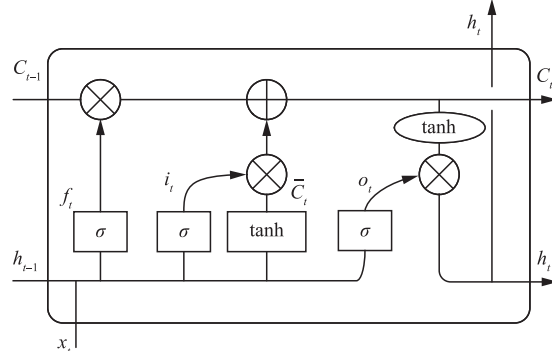


图 4 LSTM 结构图

在文本分类任务中,从当前词的上下文同时分析特征获取的特征信息更为准确。所以本文进一步采用 BiLSTM 模型,丰富模型的词向量特征获取能力,BiLSTM 由正向 LSTM 和反向 LSTM 组合而成,模型如图 5 所示。

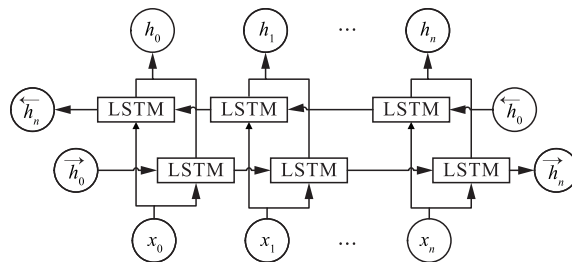


图 5 BiLSTM 结构图

由图可知,在 n 时刻 BiLSTM 网络的输出由 x_n 、 \vec{h}_n 、 \overleftarrow{h}_n 共同决定,网络更新公式为:

$$\vec{h}_n = \text{LSTM}(x_n, \vec{h}_{n-1}) \quad (11)$$

$$\overleftarrow{h}_n = \text{LSTM}(x_n, \overleftarrow{h}_{n-1}) \quad (12)$$

式中: \vec{h}_n 表示在 n 时刻 LSTM 网络的正向输出; \overleftarrow{h}_n 表示 n 时刻 LSTM 网络的反向输出; x_n 代表 BiLSTM 网络在 n 时刻的输入; h_n 代表 BiLSTM 网络在 n 时刻的输出。

$$h_n = \mathbf{w}_n \vec{h}_n + \mathbf{v}_n \overleftarrow{h}_n + b_n \quad (13)$$

式中: \mathbf{w}_n 、 \mathbf{v}_n 为各自的权重矩阵; b_n 为偏置。

1.3 变焦损失函数

1.3.1 现有损失函数问题分析

针对分类任务中数据集的不平衡性,神经网络通常采用加权交叉熵损失函数平衡数据集中的类

别,即将各个训练样本的交叉熵(cross entropy, CE)加权求和,加权交叉熵计算公式如下:

$$CE(p_i) = -\alpha_i \log(p_i) \quad (14)$$

式中: $p_i \in [0, 1]$ 代表样本正确分类的概率; $\alpha_i \in [0, 1]$ 代表类别数量平衡权重,其值通常根据逆类别频率确定。

然而,在分类任务中类别分类的难易性也常常对训练结果造成很大影响。如果增加难分类样本的权重,则会提高网络对于难分类样本的关注度,提升分类效果^[16]。

为了将难分类样本的权重增加,Lin^[17]提出了焦点损失函数(focal loss,FL),计算公式如下:

$$FL(p_i) = -\alpha_i (1-p_i)^\gamma \log(p_i) \quad (15)$$

式中: $\alpha_i \in [0, 1]$ 代表类别数量平衡权重, $p_i \in [0, 1]$ 代表样本正确分类的概率, $(1-p_i)^\gamma$ 被称为调制系数,通过该系数控制难易样本的权重,其中 $\gamma \in R^+$ 。具体来看,当 γ 值提高时,难分类样本($p_i < 0.5$)权重将增加,易分类样本($p_i > 0.5$)权重将减少,令模型更加专注于难分类样本的训练。CE 与 Focal loss 对比见图 6。

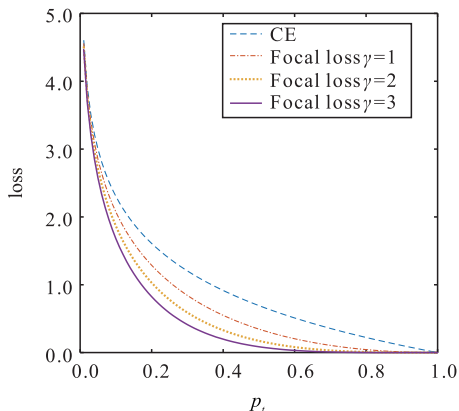


图 6 CE 与 Focal loss 对比图

图 6 中,横轴代表样本难易性指标 p_i ,纵轴代表当 $\alpha_i = 1$ 时,各类损失函数的值。可以看出,Focal loss 降低了易分类样本在损失中的权重,使得模型更加注重训练难分类样本。根据以上分析,我们可以将样本性质做如下归类,如表 1 所示:

表 1 样本性质与权值变化表

样本性质	权值变化	样本性质	权值变化
多	减少	少难	增加
少	增加	多易	减少
难	增加	多难	?
易	减少	少易	?

可以看出,综合样本的数量特征和难易特征,样本性质可以分为 4 种:即数量少且分类难度大(少难)、数量多且分类难度小(多易)、数量多且分类难度大(多难)、数量少且分类难度小(少易)。为达到

平衡数据集的目的,少难样本与多易样本权值变化方向明确,而多难样本与少易样本权值变化存在冲突,这使得损失函数难以为这两类样本赋予合适的权重。为直观分析损失函数在两类性质下的变化情况,可以画出类别数量平衡权重 α_i ,样本难易性指标 p_i 与损失函数变化的三维图像,见图 7。

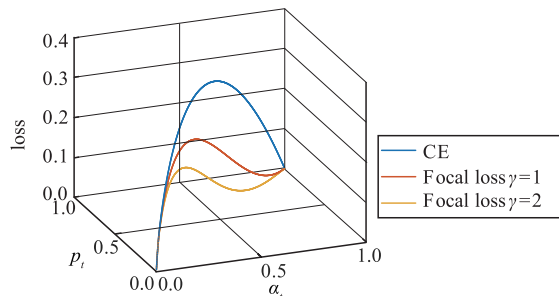


图 7 CE 与 Focal loss 三维对比图

图 7 中, x 轴代表类别数量平衡权重 α_i , y 轴代表样本难易性指标 p_i , z 轴为损失函数值。图像越接近原点,即代表样本的多难属性越强,越远离原点代表样本的少易属性越强。Focal loss 相较于 CE 波峰向原点移动,即相比于样本的多少,Focal loss 更趋向于关注样本的难易不均衡性,给难样本提供更高的权值。但若数据集的类别数量不平衡程度高于难易不平衡程度,Focal loss 将会给数量多的样本分配更高的权重,使得数据集平衡能力下降。

1.3.2 变焦损失函数

由上述分析看出,Focal loss 相对于 CE,其优势在于:随着 γ 的增加,曲线波峰向原点移动,即损失函数控制难易样本不均衡能力逐渐增加;但 Focal loss 缺少一个可变参数,使得损失函数针对样本类别数量不均衡性缺乏控制能力。这是因为一方面作为类别数量平衡权重 α_i ,其实际意义通常为数据的逆类别频率,是根据某一类别所含服务数量确定的固定值,另一方面,若直接通过改变每个类别所对应的 α_i 来实现对类别的平衡,在面对类别数量较多的多分类问题时,所要调整的参数数量将多到难以进行精确的对比控制。为了得到可以针对数据集的不平衡性质进行适应性变化的损失函数,本文在原有 Focal loss 的基础上提出了变焦损失函数(zoom loss,ZL),其函数形式为:

$$ZL(p_i) = -\alpha_i^\theta (1-p_i)^\gamma \log(p_i) \quad (16)$$

式中: α_i 为类别数量平衡权重; θ 为数量调制系数; p_i 代表样本属于真实样本的概率; γ 为难易调制系数。其特点在于,可以通过控制 θ 和 γ 以调整损失函数聚焦的样本不平衡性质,可以看出当 $\theta = 1, \gamma = 0$ 时,其函数形式等价于交叉熵损失函数,当 $\theta = 1, \gamma > 0$ 时,其函数形式等价于 Focal loss,图 8 展示了 Zoom loss 的函数图像:

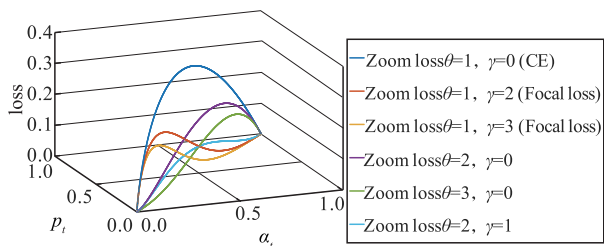


图8 变焦损失函数图像

可以看出,Zoom loss 的特性在于,可以根据数据集的数量不平衡和难易不平衡程度来调节损失函数的关注焦点,当数据集数量不平衡性强于难易性,则可增大 θ ,使得损失函数更趋向于为少样本分配较大的权重;当数据集难易性强于不平衡性,则可增大 γ ,使得损失函数更趋向于为难样本分配较大的权重。这使得 Zoom loss 可以针对数据集的不平衡特性进行适应性调整,提高模型的分类性能。

2 实验准备

2.1 实验数据集获取与处理

为获取服务描述文本的真实数据,本文使用爬虫工具从 ProgrammableWeb 网站上爬取了 437 类,共 22 205 个服务描述文本,用作本次实验的原始数据集。

为保证模型训练效果,需要对数据集进行无效服务清除。通过对数据集的分析,清除原始数据中存在的同一服务重复注册、服务描述文本相同、服务失效、服务标签为空或服务文本为空等问题的无效服务。对服务筛选后的数据进行分词,去除标点符号,去除停止词后,最终获得 51 类,共 13 694 个服务文本,作为本次实验的最终数据集。

2.2 实验环境搭建

本文实验采用的实验设备为 Intel(R) Core(TM) i7-10875H CPU@2.30 GHz;采用 GeForce GTX 2060 图形处理器,CUDA 版本 11.1.114。代码通过 pycharm 平台进行编程,采用 1.9.0 版本 pytorch 深度学习框架进行模型建立。BERT 模型版本采用了 Google 预训练的 BERT-Base-uncased 模型,模型包含 12 个多头注意力层,768 个隐藏层,每层 12 个注意力头数,共含有 1.1 亿个参数。

在数据集处理上,本文以 10:1 的比例切分训练集和测试集,batch_size 设置为 8,最大 epochs 为 17 次。模型学习率为 0.000 05,优化器为 Adam^[18]。

2.3 模型评估方法

分类问题中,常用精确率(precision)、召回率(recall)和 F1 值作为不均衡数据下多分类任务评估

方法。混淆矩阵如表 2 所示。

表 2 混淆矩阵表

	正类	负类
真类	真正类(TP)	真负类(TN)
假类	假正类(FP)	假负类(FN)

根据混淆矩阵表,各评估指标可表示如下:

1) 精确率:

$$precision = \frac{TP}{TP + FP} \quad (17)$$

2) 召回率:

$$recall = \frac{TP}{TP + FN} \quad (18)$$

3) F1 值:

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (19)$$

在多分类问题中,F1 值存在多种计算方法,本文采用 Macro-F1 值作为多分类问题中的衡量标准,计算方法为:

$$Macro-F1 = \frac{1}{n} \sum_{i=1}^n F1_i, \quad i = 1, 2, \dots, n \quad (20)$$

式中: n 代表分类类别数; $F1_i$ 代表第 i 类的 F1 值。

模型还常用准确率(Accuracy)作为评估模型分类性能的指标,其计算公式为:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (21)$$

3 实验设计及结果分析

3.1 领域词汇提取实验

为验证添加领域词汇后模型的优势,本文引入一组对照组 BERT-Random,该对照模型从数据集中随机获取数量相同的词汇以相同的方式增加进词表进行训练。通过对比 BERT 模型、BERT-Random 模型和本文使用的领域词汇增强 BERT 模型的分类效果,其结果如表 3 所示。

表 3 领域词汇增强结果

模型	准确率	Macro-F1	训练时间/min
BERT	0.719 23	0.694 74	138
BERT-Random	0.722 15	0.693 56	127
WBI	0.725 07	0.703 84	110

根据实验结果可以看出,本文使用的模型相比于原始 BERT 模型和 BERT-Random 模型,其准确率和 Macro-F1 值都有提升,且训练时间明显减少,根据实验对比分析可知,针对原始词表进行词汇增强可以有效提升模型性能,且针对领域词汇进行的增强取得了最好的效果。

3.2 损失函数改进实验

为验证 Zoom loss 的有效性,本文在 WBBI 模型上进行了实验,并与 CE 和 FL 进行了对比,本实验在 $\theta \in [1.0, 2.5]$, $\gamma \in [0.0, 2.0]$ 的范围内取多组值进行了实验,实验结果如表 4。

表 4 损失函数实验结果

损失函数	θ	γ	准确率	Macro-F1
CE	1.00	0.0	0.719 84	0.694 41
Focal loss	1.00	1.5	0.718 96	0.689 85
Focal loss	1.00	2.0	0.705 81	0.676 43
Zoom loss	1.50	0.0	0.711 07	0.686 93
Zoom loss	1.75	0.0	0.718 67	0.698 51
Zoom loss	2.50	0.0	0.721 00	0.702 46
Zoom loss	2.00	0.5	0.717 79	0.694 41
Zoom loss	2.00	1.0	0.718 08	0.699 63
Zoom loss	2.00	0.0	0.721 00	0.701 19

实验表明,Zoom loss 在 $\theta=2.0$, $\gamma=0.0$ 时效果高于 CE 和 Focal loss。同时也可以看出,CE 的实验效果高于 Focal loss,这是由于本数据集数量不平衡性较强,而 Focal loss 主要关注难分类样本,性能无法充分发挥。

同时也注意到 θ 的值不宜太高,过高的 θ 值会使得易分类样本的损失值增大,这将使得模型收敛时间大大增加,模型训练时间对比如表 5 所示。

表 5 损失函数训练时间对比

损失函数	θ	γ	训练时间/min
CE	1.0	0.0	199
Focal loss	1.0	1.5	201
Focal loss	1.0	2.0	125
Zoom loss	1.5	0.0	103
Zoom loss	1.75	0.0	137
Zoom loss	2.0	0.0	189
Zoom loss	2.5	0.0	299

由表 5 可以看出,当 γ 不变, θ 的值不断提高时,模型训练时间将会不断提高,反之当 θ 的值不变, γ 提高时,训练时间不断减少。这是由于在本文数据集中,当 Zoom loss 焦点逐渐向占有多数的易样本偏移时,易样本损失值权重增大,使得模型整体损失值增大,模型将难以收敛。

3.3 模型对比实验

3.3.1 通用文本分类模型对比实验

为验证 WBBI 服务分类模型的优越性,在相同

实验环境下同时与 TextCNN^[19]、BiLSTM-Attention^[20]、RCNN^[21]、Transformer^[22] 文本分类模型进行比较。

Transformer 模型整体由 encoder 和 decoder 组成,通过多头注意力机制,捕捉文本特征信息。模型学习率为 0.000 05,卷积核数量为 128,注意力头数为 8,Encoder 数量为 1,词向量维数为 200,优化算法为 Adam,词汇表大小为 10 000。

模型对比实验结果如表 6。总体来看,WBBI 模型实验结果各项评估指标上均优于基线模型,分类效果提升明显。其中,在测试集中的精确率、召回率和 Macro-F1 的实验结果下,相较于 TextCNN 模型分别上升了 4.94%、2.31% 和 4.29%。这是由于 TextCNN 模型只能获取文本局部特征,而 WBBI 的双向 LSTM 层可以更好地获取上下文信息,特征提取能力较强;相较于 BiLSTM-Attention 和 RCNN 模型精确率分别提高了 8.62% 和 5.42%;召回率分别提高了 3.45% 和 5.44%;Macro-F1 分别提高了 6.59% 和 5.3%。这是由于在词向量处理上,BERT 模型获取的词向量特征相比于其他深度学习模型具有更好的表征能力,同时可以看出 RCNN 的分类效果要优于 BiLSTM-Attention,这是由于 RCNN 在结合双向 LSTM 信息的同时也融合了原始词向量信息,信息特征获取更加全面;相较于 Transformer 模型,WBBI 模型的精确率、召回率和 Macro-F1 值则分别提高了 39.43%、43.06% 和 43%,这是由于 Transformer 模型需要更多的 epochs 来进行训练,在同等训练条件下,Transformer 模型的收敛速度要慢于 WBBI 模型,因此训练效果不佳。

表 6 对比实验结果

模型	精确度	召回率	Macro-F1
TextCNN ^[19]	0.661 8	0.688 3	0.662 7
BiLSTM-Attention ^[20]	0.625 0	0.676 9	0.639 7
RCNN ^[21]	0.657 0	0.672 4	0.652 6
Transformer ^[22]	0.316 9	0.280 8	0.273 5
WBBI	0.711 2	0.711 4	0.705 6

图 11 和图 12 分别展示了不同模型在迭代过程中的损失值和 Macro-F1 的变化情况。由图 11 可以看出相比于基线模型,WBBI 模型训练过程更加稳定且拥有更低的损失值,总体曲线波动较小。由图 12 可以看出 WBBI 拥有最高的 Macro-F1 值,且曲线波动较小,训练过程稳定,很好地体现了模型的优势。

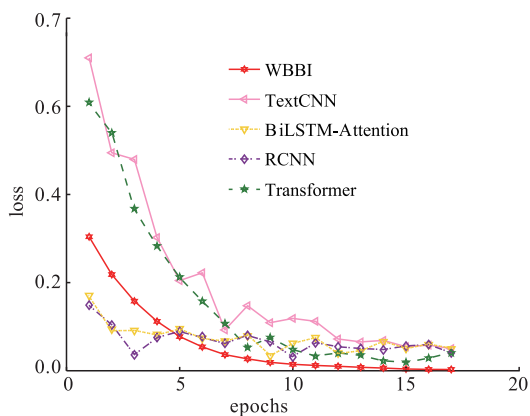


图 11 Loss 随 epochs 变化曲线

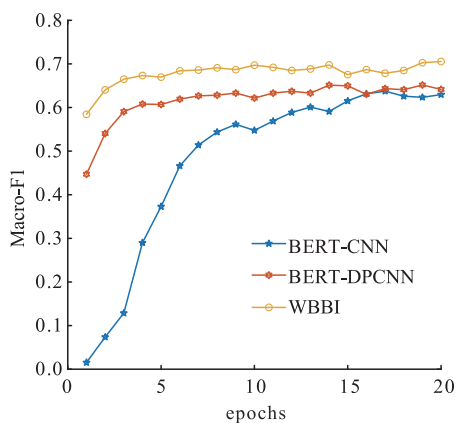


图 14 Macro-F1 随 epochs 变化曲线

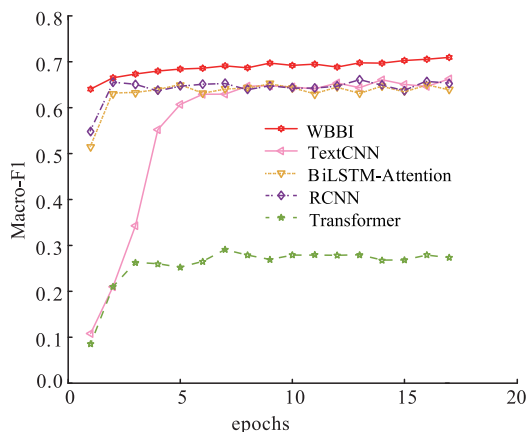


图 12 Macro-F1 随 epochs 变化曲线

3.3.2 基于 BERT 文本分类模型的对比实验

为进一步体现 WBBi 模型的性能，本文在相同实验条件下，进一步与采用 BERT 预训练模型的 BERT-CNN^[23]算法和 BERT-DPCNN^[24]算法进行对比。其中 BERT-CNN 算法是将 BERT 层输出到 2 层 CNN 层以获取文本的局部信息。本实验设定其卷积核大小为 3，卷积核数量为 250。BERT-DPCNN 算法是通过将 BERT 模型输出到拥有残差连接结构的深度 CNN 网络，这种结构可以在 BERT-CNN 的基础上获取更为丰富文本特征。本实验同样设定 DPCNN 中卷积层大小为 3，卷积核数量为 250。上述实验最大 epochs 为 20 次。模型学习率为 0.000 05，优化器为 Adam。实验结果见图 13 和图 14。

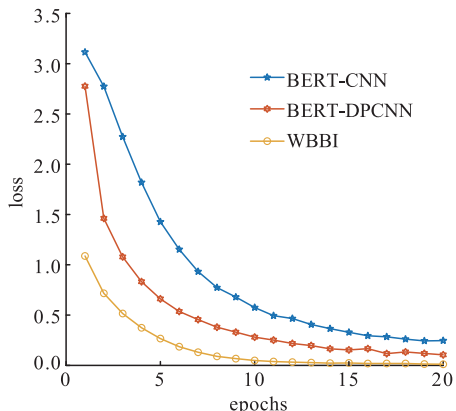


图 13 Loss 随 epochs 变化曲线

由图 13 和图 14 分析可知，相较于 BERT-CNN 算法和 BERT-DPCNN 算法，WBBi 模型收敛速度更快，模型训练过程更稳定，同时 WBBi 模型具有更低的 loss 和更高的 Macro-F1 值，即拥有更好的分类效果。具体来看：BERT-DPCNN 算法表现优于 BERT-CNN 算法，这是由于 DPCNN 模型相较于 CNN 模型拥有更深的网络结构，因此获得了较强的文本特征表征的能力。WBBi 模型拥有最佳的模型分类效果，这是由于一方面 WBBi 模型中的 BiLSTM 层拥有更强的捕捉长距离文本依赖关系的能力，其包含的双向 LSTM 结构也更适合捕捉文本的上下文关系，另一方面，WBBi 模型采用的领域词汇增强和变焦损失函数进一步增强了模型的训练效果。

4 结语

本文提出了一种新颖的服务文本分类模型以解决服务分类问题。该模型首先通过领域词汇增强的 BERT-BiLSTM 模型获取词向量上下文特征进行分类，最后提出了变焦损失函数，通过分析数据集的类别数量的不均衡性和难易不均衡性进行相应的参数调整，提升了模型的性能。经实验可知本文所提出的框架分类效果最优。下一步将继续针对 BERT 模型在新领域语料中的预训练和微调过程进行研究，进一步提升 BERT 模型的领域适应性。

参考文献

[1] TANG B, YAN M, ZHANG N, et al. Co-Attentive Representation Learning for Web Services Classification[J]. Expert Systems with Applications, 2021, 180: 115070.

[2] LIU X, AGARWAL S, CHEN D, et al. An LDA-SVM Active Learning Framework for Web Service Classification[C]// 2016 IEEE International Conference on Web Services. San Francisco, CA, USA:

- IEEE, 2016: 49-56.
- [3] 曹步清,肖巧翔,张祥平,等. 融合 SOM 功能聚类与 DeepFM 质量预测的 API 服务推荐方法[J]. 计算机学报,2019,42(6):1367-1383.
- [4] ALI S H, EL-ATIER R A, ABO-AL-EZ K M, et al. A Gen-Fuzzy Based Strategy (GFBS) for Web Service classification[J]. *Wireless Personal Communications*, 2020, 113(4): 1917-1953.
- [5] DU L, CAO D, LI J, et al. Classification and Extraction of Medical Clinical Trial Screening Standard Texts Based on Bi-LSTM and Attention Mechanism [J]. *IOP Conference Series Earth and Environmental Science*,2021,632:55-88.
- [6] WANG H, HE J, ZHANG X, et al. A Short Text Classification Method Based on N-Gram and CNN[J]. *Chinese Journal of Electronics*, 2020, 29 (2): 248-254.
- [7] XIE J, CHEN B, GU X, et al. Self-Attention-Based BiLSTM Model for Short Text Fine-Grained Sentiment Classification [J]. *IEEE Access*, 2019, 7: 180558-180570.
- [8] DEVLIN J, CHANG M, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. (2019-05-24)[2022-06-21]. <https://arxiv.org/pdf/1810.04805.pdf>.
- [9] WANG Y, SHEN X, YANG Y. The Classification of Chinese Sensitive Information Based on BERT-CNN [J]. *Artificial Intelligence Algorithms and Applications*, 2020, 1205: 269-280.
- [10] AMBALAVANAN A K, DEVARAKONDA M V. Using the Contextual Language Model BERT for Multi-Criteria Classification of Scientific Articles[J]. *Journal of Biomedical Informatics*, 2020, 112 (3): 103578.
- [11] GAO Z, FENG A, SONG X, et al. Target-Dependent Sentiment Classification with BERT [J]. *IEEE Access*, 2019, 7(1): 154290-154299.
- [12] SCHUSTER M, PALIWAL K K. Bidirectional Recurrent Neural Networks[J]. *IEEE Transactions on Signal Processing*, 2002, 45(11):2673-2681.
- [13] SALTON G. The SMART Retrieval System; Experiments in Automatic Document Processing[M]. Englewood Cliffs, USA: Prentice-Hall, 1971.
- [14] 黄伟,刘贵全. MSML-BERT 模型的层级多标签文本分类方法研究[J]. *计算机工程与应用*,2022,58(15): 191-201.
- [15] GLOROT X, BORDES A, BENGIO Y. Deep Sparse Rectifier Neural Networks[J]. *Journal of Machine Learning Research*, 2011, 15:315-323.
- [16] HOSSAIN M S, BETTS J M, PAPLINSKI A P. Dual Focal Loss to Address Class Imbalance in Semantic Segmentation[J]. *Neurocomputing*, 2021, 462(5): 69-87.
- [17] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J]. *Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42 (2):318-327.
- [18] VIEIRA J P A , MOURA R S. An Analysis of Convolutional Neural Networks for Sentence Classification [C]// 2017 XLIII Latin American Computer Conference (CLEC). Cordoba,Argentina: IEEE, 2017: 1-5.
- [19] ZHOU P, SHI W, TIAN J, et al. Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Stroudsburg: Assoc Computational Linguistics, 2016: 207-212.
- [20] CHOI K, FAZEKAS G, SANDLER M, et al. Convolutional Recurrent Neural Networks for Music Classification[C]// 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE, 2017: 2392-2396.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[C]// The 31st International Conference on Neural Information Processing Systems (NIPS 2017). Massachusetts, USA: MIT Press, 2017: 5998-6008.
- [22] KINGMA D,BA J. Adam: A Method for Stochastic Optimization[EB/OL]. (2017-01-30)[2022-06-21]. <https://arxiv.org/pdf/1412.6980.pdf>.
- [23] ZHOU G . Donation-Based Crowdfunding Title Classification Based on BERT+CNN[C]// 2021 International Conference on Bioinformatics and Intelligent Computing. [S. l.]:ACM,2021:291-296.
- [24] 路凯峰,杨溢龙,李智. 一种基于 BERT 和 DPCNN 的 Web 服务分类方法[J]. *广西师范大学学报(自然科学版)*, 2021, 39(6):87-98.

(编辑:徐楠楠)