

金晨曦, 韩坤. 基于多特征融合和数据增强的警情文本智能分类方法[J]. 智能计算机与应用, 2025, 15(12): 88-93. DOI: 10.20169/j.issn.2095-2163.24110801

# 基于多特征融合和数据增强的警情文本智能分类方法

金晨曦, 韩坤

(中国人民公安大学 公安管理学院, 北京 100038)

**摘要:** 针对警情文本分类有效机制缺乏且样本体量小的问题, 本文提出了一种基于注意力机制的双通道特征提取方法用于警情智能分类。首先, 借助预训练模型 BERT 实现文本数据的向量化表征; 其次, 融合文本卷积神经网络 (Text Convolutional Neural Network, TextCNN) 与双向长短期记忆网络 (Bidirectional Long Short-Term Memory, BiLSTM) 以综合挖掘文本的深层次语义特征, 并使用注意力机制对特征进行加权处理; 最后, 为解决数据体量小的问题, 采用基于知识整合的增强表征模型 (Enhanced Representation through kNowledge IntEgration, ERNIE) 的数据增强策略对原始样本集进行有效扩充, 该数据增强策略提升了分类方法的效果。实验结果表明该方法在准确率、精确率、召回率以及  $F1$  值等评价指标上均优于当前主流分类方法, 证明了该方法在警情智能分类任务中的优越性。

**关键词:** 警情智能分类; 数据增强; 神经网络; BERT; ERNIE

中图分类号: TP391.1

文献标志码: A

文章编号: 2095-2163(2025)12-0088-06

## Research on alarm text classification based on multi-feature fusion and data augmentation

JIN Chenxi, HAN Kun

(School of Police Administration, People's Public Security University of China, Beijing 100038, China)

**Abstract:** To address the issue of the lack of an effective mechanism for classifying police incident texts and the small sample size, this study proposes a dual-channel feature extraction model based on attention mechanism. This model initially leverages the BERT model to achieve vectorized representation of text data, and then fuses TextCNN and BiLSTM to comprehensively mine deep semantic features of the text and the features are weighted by the attention mechanism. To solve the problem of small data volume, this study adopts a data augmentation strategy based on ERNIE to effectively expand the original sample set. Experimental results demonstrate that this strategy can significantly enhance model performance. The proposed model outperforms current mainstream classification models in terms of accuracy, precision, recall, and  $F1$  - score, validating the superiority of the proposed model in policing classification tasks.

**Key words:** policing classification; data augmentation; neural networks; BERT; ERNIE

## 0 引言

2023年11月, 全国公安厅局长座谈会议精神强调“着力推进智能化, 在深化智慧公安建设上聚焦发力”, 并正式提出“建立完善‘专业+机制+大数据’新型警务运行模式”。2024年5月, 全国公安工作会议强调“要进一步向‘机制’聚焦发力, 充分发挥‘情指行’一体化运行机制牵引作用”。其中, 警

情数据是构建情报主导警务、数据驱动警务的先导因素, 对警情数据的智能分析是深化智慧公安建设和发挥“情指行”一体化运行机制的重要手段。目前各地公安机关结合自身警情感知情况和大数据技术, 构建警情数据的收集、储存和应用平台, 取得一定进展<sup>[1-2]</sup>。但当前警情数据处理仍然主要依赖人工, 导致警情研判精细化程度不够, 缺少智能化分析应用<sup>[3]</sup>。以警情分类为例, 接警员对警情判断主要

**基金项目:** 中国人民公安大学公安管理学双一流创新研究专项(2023SYL04); 中国人民公安大学研究生课程建设项目(2022yjkskcs014)。

**作者简介:** 金晨曦(2000—), 女, 硕士研究生, 主要研究方向: 大数据分析, 自然语言处理。

**通信作者:** 韩坤(1999—), 男, 硕士研究生, 主要研究方向: 网络舆情治理与大数据分析。Email: hk1064223880@163.com。

收稿日期: 2024-11-08

依赖主观经验,可能导致警情分类分级存在偏差,造成警力资源调度不合理,乃至生命财产损失。

实践中,各类警情差异较大,警情文本数据长短文本并存。本文根据警情文本实际字符数量,将融合双向长短时记忆网络(BiLSTM)和文本卷积神经网络(TextCNN),形成双通道模式,并引入注意力机制,增强模型对文本特征的理解,从而提高警情分类的准确率;针对警情数据体量小、类别样本不均衡,提出基于 ERNIE 的数据增强策略,实现对少数类别进行扩充,弥补现有语料库中稀缺类别的数据不足。

## 1 相关工作

### 1.1 警情文本分类研究

文本分类是自然语言处理中的一项基本任务,旨在通过算法和模型将文本数据自动归类到预定义的标签或类别中,隶属于监督学习,常应用于情感分析、新闻分类以及主题分类等场景<sup>[4-5]</sup>。

目前,众多学者使用机器学习和深度学习为警情数据处理提供多种思路<sup>[6]</sup>。张齐等<sup>[7]</sup>以盗窃犯罪数据为样本,利用词频-逆文件频率(TF-IDF)提取样本特征,并使用极限梯度提升法 XGBoost(eXtreme Gradient Boosting)和 k 近邻法 KNN(K-Nearest Neighbo)等方法对样本数据进行分类测试,比较之后得出 XGBoost 的整体表现性能优于其他方法;刘冬等<sup>[8]</sup>针对警情文本存在极短文本和类别分布不均衡,提出 BERT-BiGRU-WCELoss 模型,在将

BERT(Bidirectional Encoder Representation from Transformers)和 BiGRU 结合,使用优化自适应权重损失函数 WCELoss 对警情中的少量样本赋予更多权重损失,在以某一自然月为单位的警情数据样本中分类准确率达 95.83%;李昀轩等<sup>[9]</sup>针对交通警情设计多任务迁移学习模型,中上游使用 BERT 的文本动态向量特征,下游采用 TextCNN、TextRNN 和 DPCNN 等组成分类模型,相较于传统模型准确率都有显著提升,但由于主要聚焦于交通警情,缺少其他类型警情分析。此外,传统方法中只关注文本本身的信息而无法理解深层语义信息,也未考虑标签之间的关系,刘新忠等<sup>[10]</sup>提出了融合 BERT-GAT(Graph Attention neTwork)-CorNet(Correlation Network)的多标签文本分类方法。在针对长文本的处理,王浩镔等<sup>[11]</sup>提出一种基于多级特征和混合注意力机制的 Seq2Seq(Sequence to Sequence)方法,通过词语级别和语义级别的特征提取,解决多标签分类中标签共生关系被忽略的问题,该方法的 F1 值和召回率均优于传统方法;叶瀚等<sup>[12]</sup>针对预训练语言模型 BERT 的输入长度限制,提出通过分句处理长文本,结合句向量平均池化与注意力机制加权法,优化分类特征编码,该方法在多个数据集上显著提升了分类准确率(最高提升 3.74%)和 F1-score。

本文使用的常见报警类别示例见表 1。

表 1 常见报警类别示例

Table 1 Examples of common alarm categories

类别	内容示例
纠纷	新城 * 号楼 * 单元 * 室,报警人称其跟其家人遭到陌生人无故辱骂等,影响到自己的生活和名誉。
求助	**镇 24-**-**,报警人称其妻子患有老年痴呆,身穿粉江西服外套,身高 150,体型偏胖,无联系电话,没有携带身份证,于今日下午 14 时左右出门,至今联系不到。
治安警情	**路东首一港务以南交叉口,报警人称自己的一部 vivo 手机被偷。
刑事警情	**港某园宿舍,报警人称:通过某软件共转账约 19 万余元,现在报警人联系不上对方,怀疑被骗。

### 1.2 数据增强方法

文本数据增强是旨在通过对现有文本数据进行变换与扩展,生成新的训练样本,缓解数据稀缺的难题。Jason 等<sup>[13]</sup>提出了简单数据增强(EDA)方法,依赖于简单的规则和词汇替换策略来扩充数据集;姜钰棋等<sup>[14]</sup>进一步提出 EdaDyFeFL 方法,对英文数据集 Cyber-Trolls 进行增强,引入了动态反馈机制,准确率提升了 7.28%;赵一鸣等<sup>[15]</sup>使用 SimBERT 对小样本数据进行增强,提出融合任务知识的 MQ-BERT 方法,接入 BiLSTM 对医学信息文本语料分

类;张小川<sup>[16]</sup>提出了一种惩罚生成式预训练语言模型的数据增强方法,在预训练模型 GPT-2 基础上,通过设置惩罚项和超参数,引导模型不断生成与原始数据相似的新文本,使用 BERT 模型作为过滤器完成数据增强。

现阶段已有多位学者尝试将注意力机制融入模型中,以扩大模型本身感受野。Li 等<sup>[17]</sup>提出量子自注意力神经网络(Quantum Self-Attention Neural Network, QSANN)架构,将量子计算与自然语言处理相结合,引入高斯投影量子自注意力挖掘高维量

子特征空间中单词间相关性,为量子计算在文本处理领域提供新路径;陈帅朴等<sup>[18]</sup>提出多重特征关联和图注意力网络融合的层次分类(Hierarchical Text Classification Network - Multi - Characteristic Graph Attention Network, HTCN-MCGAT)模型,重新设计BERT预训练模型微调阶段,使用图注意力网络对标签语义和层次结构关系建模,并构建层次融合注意力机制实现文献分类,在中文医学文献分类任务中  $F1 - score$  显著提高。

现有数据增强方法更多面向英文,而且容易改变原始语句的语义,因此亟需针对中文的文本增强方法,以有效应对中文语言的复杂多样性。

本文使用 ERNIR 对原始数据进行同义替换、随机插入、随机交换和随机删除。

## 2 模型构建

为提高警情文本分类的准确率,本文设计一基于注意力机制的双通道特征提取方法,具体结构如图 1 所示。

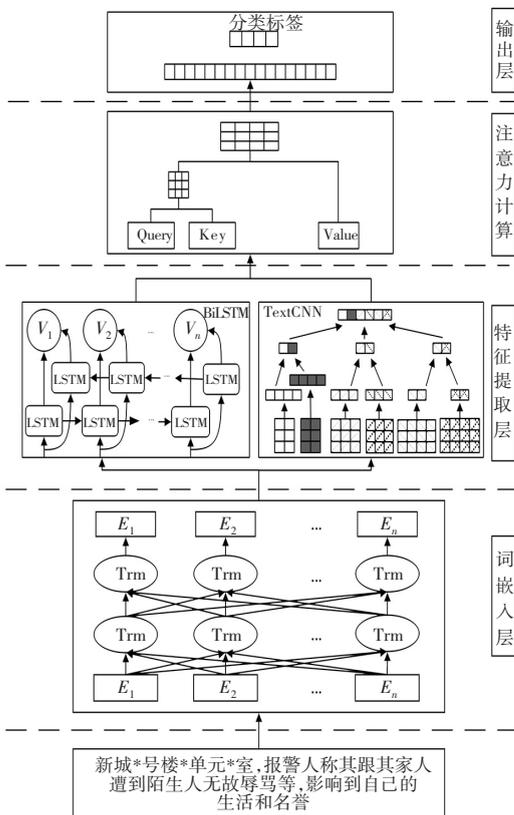


图 1 基于注意力机制的双通道特征提取方法

Fig. 1 Dual-channel feature extraction method based on attention mechanism

### 2.1 词嵌入层

谷歌团队<sup>[19]</sup>2018年提出预训练模型 BERT,不

仅能够捕捉文本中的双向语义关系而且具有强大的泛化能力。BERT 模型包括 3 个部分,具体结构如图 2 所示。

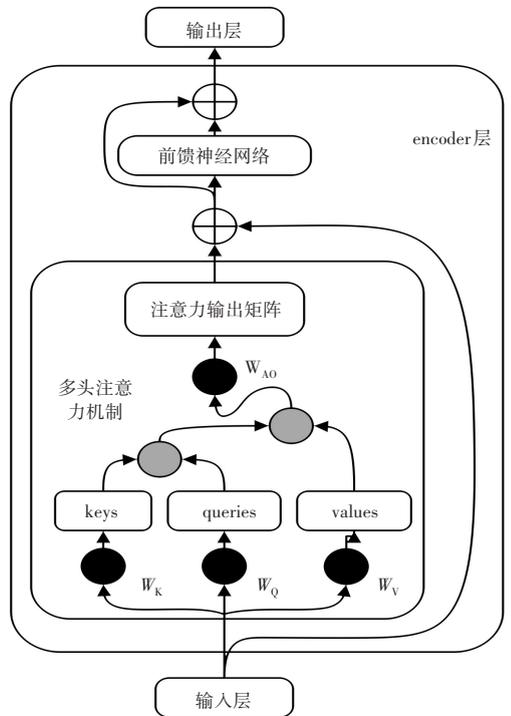


图 2 BERT 模型结构

Fig. 2 Structure of BERT

输入层融合了 3 种嵌入技术分别是 Token 嵌入(Token Embeddings)、段落嵌入(Segment Embeddings)和位置嵌入(Position Embeddings);三者互相配合,以融合多维度语义信息。

中间由多个编码块堆叠成为编码层。每个编码层接收输入层的向量序列并送入多头自注意力(Multi-Head Self-Attention, MHSA)机制进行深度处理。在 MHSA 中,输入向量通过 3 个不同的线性变换分别得到查询向量(Q)、键向量(K)和值向量(V),公式如下:

$$Q_i = XW_i^Q \quad (1)$$

$$K_i = XW_i^K \quad (2)$$

$$V_i = XW_i^V \quad (3)$$

其中, X 是输入向量,  $W_i^Q, W_i^K, W_i^V$  是第 i 个头的权重矩阵。

每个头计算查询向量与所有键向量之间的注意力分数,所得分数通过 Softmax 函数转化为注意力权重,并与值向量进行加权求和,得到每个头的加权输出向量,公式如下:

$$Attention\_Score_{ij} = \frac{Q_i \cdot K_j^T}{\sqrt{d_k}} \quad (4)$$

$$\text{Attention\_Weight}_{ij} = \text{Softmax}(\text{Attention\_Score}_{ij}) \quad (5)$$

$$\text{Output}_i = \sum_j \text{Attention\_Weight}_{ij} V_j \quad (6)$$

其中,  $d_k$  是键向量的维度。

所有头的输出向量被拼接起来,并通过一个额外的线性变换进行降维,得到 MHSA 的最终输出,经过残差连接和层归一化操作进入前馈神经网络(Feed Forward Neural Network)。前馈神经网络由两个线性变换和一个 ReLU 激活函数组成,公式如下:

$$\text{FFN\_Output} = \text{ReLU}(\mathbf{X}W_1 + b_1)W_2 + b_2 \quad (7)$$

其中,  $\mathbf{X}$  是 MHSA 的输出,  $W_1, b_1, W_2$  和  $b_2$  是 FFN 的权重和偏置项。

FFN 的输出也会经过一个残差连接和层归一化操作,作为编码层最终输出。

## 2.2 双通道特征提取层

### 2.2.1 BiLSTM 层

BiLSTM 由正向和反向长短期记忆网络(Long Short-Term Memory, LSTM)组成,二者对生成的隐藏状态进行拼接,作为该时间步最终输出,使 BiLSTM 能捕捉到文本中的长程依赖。BiLSTM 网络结构如图 3 所示。

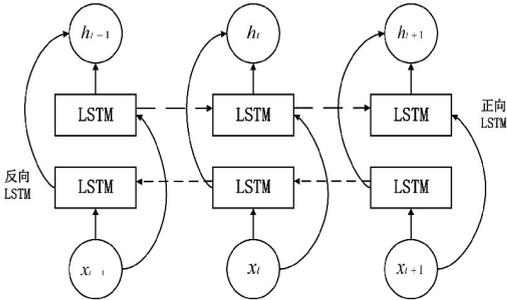


图 3 BiLSTM 网络结构

Fig. 3 BiLSTM network structure

### 2.2.2 TextCNN 层

TextCNN 层在抽取文本浅层特征方面表现出色,尤其擅长处理短文本数据,主要包括卷积层、池化层和全连接层。在卷积层中使用多个卷积核对输入的词向量矩阵进行卷积操作。卷积核的大小通常为  $h \times d$ ,高度为  $h$ ,宽度与输入矩阵中词向量的维度相同为  $d$ ,滑动窗口沿着文本的高度方向移动,以捕捉局部特征。

通过卷积得到的特征图被输入到池化层,采用最大池化(Max Pooling)策略,从每个特征图中提取最大值,以实现降维并保留句子中最关键的特征信息,防止过拟合现象的发生。

## 2.3 注意力机制

在神经网络模型中,注意力机制通过动态地调整输入数据各个部分的权重,使得模型在决策时能够更多地依赖于关键信息,而减少对不相关信息的关注<sup>[20]</sup>。输入包括查询向量(Query,  $\mathbf{Q}$ )、键向量(Key,  $\mathbf{K}$ )和值向量(Value,  $\mathbf{V}$ )。注意力机制通过点积计算查询向量与键向量的相似性得出注意力分数,公式如下:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (8)$$

其中,  $d_k$  是键向量的维度,  $\sqrt{d_k}$  是缩放因子。

得到的分数通过 Softmax 函数转换为概率分布,将权重分配给每个键对应的值,之后对值向量进行加权求和,得到最终的输出结果。

## 2.4 输出层

在输出层中将拼接后的特征向量输入至全连接层,通过非线性组合的方式对特征进行再次提炼,提取出对分类任务至关重要的信息;在全连接层之后衔接一个 Softmax 层,负责输出一个概率分布,判别输入样本隶属于各个类别的可能性,以实现警情文本数据的精确分类。

## 3 基于 ERNIE Bot 的数据增强方案

2022 年,OpenAI 推出 ChatGPT,开启预训练模型时代,随后国内也推出 ERNIE Bot 等大语言模型,此类大模型具有并行计算能力强、训练效率高,针对中文文本的处理效果强等优点<sup>[21]</sup>。

本文使用 ERNIE Bot 设计 4 种针对中文文本的增强方案:

(1) 同义替换(Synonym Replacement, SR):在已构建好的同义词词库中选取同义词代替原始单词,随机生成有别于原始语句的相似文本。

(2) 随机插入(Random Insertion, RI):在原始语句中随机插入词语,以增加句子的长度和复杂度。

(3) 随机交换(Random Swap, RS):随机交换原始词语的位置,生成词序不同的句子。

(4) 随机删除(Random Deletion, RD):以一定概率随机删除句子中的词,产生更短的句子变体。

ERNIE Bot 基于 ERNIE,可以很好处理输入的中文指令,而且相较于 EDA 增强手段,不会出现语句不通顺、歧义和关键信息缺乏等问题,ERNIE Bot 增强示例见表 2。

表2 ERNIE Bot 增强示例

Table 2 Example of ERNIE Bot Enhancement

内容	内容示例
原文本	** 港某 ** 宿舍,报警人称:通过某软件共转账约 19 万余元,现在报警人联系不上对方,怀疑被骗。
同义替换	** 港 * * 宿舍,报警人反映:通过某软件共转账约 19 万余元,现在报警人联系不上对方,怀疑被骗。
随机插入	** 港 * * 宿舍的报警人表示,他使用了某款软件,并进行了多笔转账,总额约为 19 万余元。不过,他现在竟然无法联系到对方了,这让他感到十分困惑。据他怀疑,自己可能被骗了。
随机交换	报警人在 ** 港 * * 宿舍,他称转账了约 19 万余元,通过某款软件。然而,现在他却联系不上对方,这让他开始怀疑是否被骗了。
随机删除	** 港 * * 宿舍报警,使用某软件转账约 19 万,现无法联系对方,疑被骗。

## 4 实验与分析

### 4.1 数据预处理

本文数据来自某市公安局 110 接处警平台。首先对警情数据进行预处理,确保每条警情标签与实际归属类别一致,剔除重复警情与无效警情。不同的场景对长短文本的界定不同,在本文中按照分类模型擅长处理的 512 个字符,将 4 类警情分为长文本和短文本,并且保证二者均衡。由于原始警情数据体量小,因而进行增强,增强后警情数据保持原始警情分布状态,要求增强后警情分布如图 4 所示。

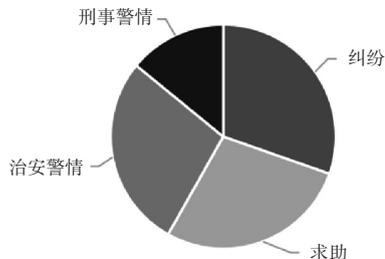


图4 增强后警情分布图

Fig. 4 Police data distribution chart after augmentation

采用基于 ERNIE Bot 的数据增强策略对数据集进行扩展,按照 8:2 的比例将数据集划分为训练集与测试集见表 3,ERNIE Bot 处理完毕的数据经过筛查之后保证增强后的数据与原始语句无歧义。

表3 数据集相关信息

Table 3 Information on datasets

数据集	训练集	测试集
增强前(原始警情)	1 951	487
增强后	7 767	1 942

### 4.2 实验环境

实验平台配置:CPU 为 AMD Ryzen 5 5600 G 处理器,内存为 16 G,GPU 为 RTX 3060。编程环境为 Python3. 9. 0,深度学习框架选用的是 Pytorch 1. 13. 1,机器学习库采用 Sklearn 1. 3. 2。模型参数

配置见表 4。

表4 模型参数配置

Table 4 Model parameters configuration

参数设置	含义	取值
Class_number	分类数	4
Epoch	遍历一次数据集的次数	5
Batch_size	每次训练所取得样本数量	64
Dropout	丢弃率	0.5
Learning_rate	学习率	1e-5

### 4.3 对比实验

为了验证所提方法的有效性,本文设计了两种类型的对比实验,选用准确率、精确率、召回率以及 F1 值作为评价指标。

第一种对比实验旨在评估增强效果见表 5。

表5 警情数据增强前后的模型效果

Table 5 Model performance before and after data augmentation

评价指标	增强前	增强后
准确率	0.610 1	0.816 5
精确率	0.669 0	0.872 8
召回率	0.610 1	0.816 5
F1 值	0.638 5	0.843 7

为进一步论证数据增强效果的可靠性,本文保持除分类数以外的参数不变的情况下,使用公开数据集 ChenSentiCrop、Wikipedia Detox、IMDB 进行对比实验。ChenSentiCrop 和 Wikipedia Detox 均以短文本为主,IMDB 长短文兼顾,更倾向于短文本,实验结果见表 6。实验结果表明基于 ERNIE 的数据增强方案能够有效应对数据不平衡问题。

第二种对比实验旨在验证本文所构建的文本分类方法的性能优势。利用经过增强的数据集,将本文方法与当前较为常见的主流分类方法 CNN、LSTM 和 GRU 进行对比实验,结果见表 7。本文所提出的模型表现最为均衡,既能保持较高的正类预测精确性,又能有效识别更多的正类样本,展示了出色的泛化能力,拥有更好的综合能力。

表 6 公开数据集增强前后对比

Table 6 Comparison of model performance on a public dataset before and after augmentation

数据集	增强阶段	准确率	精确率	召回率	F1 值
ChuSentiCorp	增强前	0.600 1	0.672 9	0.600 1	0.634 2
	增强后	0.621 0	0.720 6	0.621 0	0.667 1
IMDB	增强前	0.633 2	0.722 4	0.633 2	0.675 0
	增强后	0.633 7	0.773 5	0.633 7	0.695 3
Wikipedia Detox	增强前	0.758 1	0.764 9	0.758 1	0.761 6
	增强后	0.776 1	0.777 9	0.776 1	0.778 0

表 7 对比试验结果

Table 7 Result of comparative experiment

方法	准确率	精确率	召回率	F1 值
CNN	0.812 0	0.869 0	0.812 0	0.839 0
LSTM	0.760 2	0.894 9	0.760 2	0.820 6
GRU	0.721 2	0.894 6	0.721 2	0.788 3
本文方法	0.816 5	0.872 8	0.816 5	0.843 7

#### 4.4 消融实验

本文选取模型的多个关键组件进行了消融实验,并使用相同的数据集和训练策略来确保实验公正性,消融实验结果见表 8,可见本文所提出模型综合性能表现良好。

表 8 消融实验结果

Table 8 Results of ablation experiment

方法	准确率	精确率	召回率	F1 值
BERT	0.718 4	0.853 4	0.718 4	0.780 2
BERT-BiLSTM	0.783 8	0.804 1	0.783 8	0.793 6
BERT-TextCNN	0.807 4	0.856 5	0.807 4	0.831 2
BERT-BiLSTM-TextCNN	0.820 2	0.860 1	0.820 2	0.839 4
本文方法	0.816 5	0.872 8	0.816 5	0.843 7

## 5 结束语

本文针对警情文本中类别分布不平衡问题,首先采用 ERNIE 进行数据增强处理,通过 BERT 生成包括上下文信息的动态词向量,接入融合注意力机制的双通道模型对特征进行深度提取,优化警情分类效果。但当前研究聚焦的警情类别相对有限,未来研究可考虑纳入更多样化的警情类别,以拓宽模型的应用场景与实用性。

## 参考文献

- [1] 张武清. 用民意“引领”警务服务[J]. 民生周刊, 2022(19):68-71.
- [2] 宁辛雨, 孙国道, 金露, 等. 面向警情数据的交互式可视分析方法[J]. 计算机辅助设计与图形学学报, 2023, 35(7):1064-1076.
- [3] 戴佳静, 张艳, 蒋毅天. 关于警情感知大数据的一些思考[J]. 通信与信息技术, 2024(5):56-59.
- [4] 文飞. 传统与大模型并举:中文文本分类技术对比研究[J]. 智能计算机与应用, 2024, 14(6):88-94.
- [5] 张典, 刘畅, 陈雯柏, 等. 融合注意力机制的 MacBERT-DPCNN 农业文本分类模型[J]. 中国农机化学报, 2025, 46(8):83-89.
- [6] 王孟轩, 张胜, 王月, 等. 改进的 CRNN 模型在警情文本分类中的研究与应用[J]. 应用科学学报, 2020, 38(3):388-400.
- [7] 张齐, 李雪琛. 基于机器学习的多标签盗窃犯罪类型识别方法研究[J]. 中国人民公安大学学报(自然科学版), 2023, 29(1):88-93.
- [8] 刘冬, 翁海光, 陈一民. 一种处理严重不平衡数据的 BERT-BiGRU-WCELoss 短文本警情分类模型[J]. 计算机应用与软件, 2024, 41(9):217-223.
- [9] 李昶轩, 李萌, 陆建, 等. 基于多任务迁移学习的交通警情信息自动处理方法[J]. 中国公路学报, 2022, 35(9):1-12.
- [10] 刘新忠, 赵澳庆, 谢文武, 等. 基于 BERT-GAT-CorNet 多标签中文短文本分类方法[J]. 计算机应用, 2023, 43(S2):18-21.
- [11] 王浩镔, 胡平. 采用多级特征的多标签长文本分类算法[J]. 计算机工程与应用, 2021, 57(15):193-199.
- [12] 叶瀚, 孙海春, 李欣, 等. 融合注意力机制与句向量压缩的长文本分类模型[J]. 数据分析与知识发现, 2022, 6(6):84-94.
- [13] JASON W, ZOU Kai. EDA: Easy data augmentation techniques for boosting performance on text classification tasks [C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9<sup>th</sup> International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019:6381.
- [14] 姜钰祺, 侯智文, 王一帆, 等. 社交平台不平衡文本数据处理与应用研究[J]. 计算机科学与探索, 2024, 18(9):2370-2383.
- [15] 赵一鸣, 潘沛, 毛进. 基于任务知识融合与文本数据增强的医学信息查询意图强度识别研究[J]. 数据分析与知识发现, 2023, 7(2):38-47.
- [16] 张小川, 陈盼盼, 邢欣来, 等. 一种建立在 GPT-2 模型上的数据增强方法[J]. 智能系统学报, 2024, 19(1):209-216.
- [17] LI Guangxi, ZHAO Xuanqiang, WANG Xin. Quantum self-attention neural networks for text classification[J]. Science China Information Sciences, 2024, 67(4):303-313.
- [18] 陈帅朴, 钱宇星, 钱志强, 等. 多重特征关联和图注意力网络融合的文献分类方法研究——以中文医学文献为例[J]. 情报学报, 2024, 43(4):470-490.
- [19] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv, 1810. 04805, 2018.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv preprint arXiv, 1706. 03762, 2017.
- [21] SUN Yu, WANG Shuohan, LI Yukun, et al. ERNIE: Enhanced representation through knowledge integration [J]. arXiv preprint arXiv, 1904. 09223, 2019.