图 学 学 报 （preprint）
JOURNAL OF GRAPHICS

基于 BERT 预训练模型的灾害推文分类方法

林佳瑞，程志刚，韩　宇，尹云鹏

(清华大学土木工程系，北京 100084)

摘　　要：社交媒体已成为当前发布和传播突发灾害信息的重要媒介，有效识别并利用其中的真实信息对灾害应急管理具有重要意义。针对传统文本分类模型的不足，提出一种基于 BERT 预训练模型的灾害推文分类方法。经数据清洗、预处理及算法对比分析，在 BERT 预训练模型基础上，研究构建了基于长短期记忆-卷积神经网络(LSTM-CNN)的文本分类模型。在 Kaggle 竞赛平台的推文数据集上的实验表明，相比传统的朴素贝叶斯分类模型和常见的微调模型，该分类模型性能表现优异，识别率可达 85%，可以更好地应对小样本分类问题。有关工作对精准识别真实灾害信息、提高灾害应急响应与沟通效率具有重要意义。

关　键　词：文本分类；深度学习；BERT；预训练模型；微调；灾害；应急管理

中图分类号：X43　　**DOI：**10.11996/JG.j.2095-302X.2022030000

文献标识码：A　　**文 章 编 号：**2095-302X(2022)03-0000-00

Disaster tweet classification method based on machine learning

LIN Jia-rui, CHENG Zhi-gang, HAN Yu, YIN Yun-peng

(Department of civil engineering, Tsinghua University, Beijing 100084, China)

Abstract: Social media has become an important medium for the release and dissemination of disaster information. It is of great significance for disaster emergency management to effectively identify and use the real information. Given the shortcomings of the traditional text classification model, a disaster tweet classification method based on BERT pre-trained model is proposed. After data cleaning and preprocessing, this study constructed a text classification model based on Long Short-Term Memory-Convolutional Neural Network (LSTM-CNN) through comparative analysis, based on BERT. Experiments on the tweet data set of the Kaggle competition platform showed that the proposed classification model performs better than the traditional Naive Bayesian classification model and the common fine-tuning model, and the recognition rate can reach 85%. This study is of great significance to accurately identify the real disaster information and improve the efficiency of disaster emergency response.

Keywords: text classification; deep learning; BERT; pre-trained model; fine-tuning; disaster; emergency management

1　研究背景

突发灾害是目前全球面临的重大问题之一，严重威胁人民的生命财产安全和社会发展。据应急管理部发布的2020全国自然灾害基本情况报告，中国全年各种自然灾害共造成1.38亿人次受灾，直接经济损失3 701.6亿元。在灾害发生时，能够有效地开展应急管理工作至关重要。应急管理的核心之一是有效的信息沟通和传递，及时准确的信息共享能够辅助管理部门精准高效决策，降低突发灾害的损

收稿日期：2021-10-10；**定稿日期：**2021-11-22

Received：10 October 2021；**Finalized**：22 November 2021

基金项目：国家自然科学基金项目(72091512, 51908323)

Foundation items: National Natural Science Foundation of China (72091512, 51908323)

第一作者：林佳瑞(1987－)，男，助理研究员，博士。主要研究方向为智能建造、韧性城市，E-mail：lin611@tsinghua.edu.cn

First author：LIN Jia-rui (1987-)， research assistant professor， Ph.D. ,His main research interests cover intelligent construction and resilient city. E-mail: lin611@tsinghua.edu.cn

失[1-2]。

近年来，推特等社交网络的高速发展，为公众利用社交媒体自主发布和传播应急信息、提出应急响应建议带来了新的机会[3]。丰富而庞杂的社交信息也为细粒度下城市韧性的研究提供了重要数据支撑[4]。

然而，社交媒体信息发布的便捷性也为谣言的产生和传播提供了条件。由于政府等权威机构信息发布的滞后性与个人信息发布的随意性，社交媒体数据真假混杂、往往难以有效利用。某些不实谣言的传播甚至可能导致巨大的资源浪费和社会恐慌。因此，对灾害推文的有效识别与分类对及时甄别谣言信息、维护社会稳定[5]、提升应急管理效率具有重要意义。

针对该问题，本研究旨在利用深度学习技术提出一种灾害推文自动分类方法，实现灾害相关推文的准确分类和识别，对控制灾害谣言传播、提升真实灾害信息传播效率并辅助应急决策具有重要价值。

2 研究现状

2.1 文本分类

文本分类是自然语言处理的一个分支，目前常用的文本分类经典算法包括朴素贝叶斯、决策树、支持向量机(support vector machine, SVM)等，但往往存在分类精度较低、适应性与鲁棒性不足等问题[6-8]。对这些传统方法进行适当地改变与整合，可改善其性能。例如，李蓉等[9]通过研究 SVM 分类器的错误样本点分布，将 SVM 与 K-近邻方法结合，提高了分类器的分类精度。近年来，随着深度学习的兴起与大规模语料库构建，基于深度学习的文本分类算法飞速发展。陈翠平[10]将深度信念网络应用于文本分类领域，证明了深度信念网络相较于传统 BP 神经网络的优越性。DEVLIN 等[11]于 2018 年提出了新的自然语言表示模型 (bidirectional encoder representations from transformers, BERT)，该模型基于双向 Transformer 进行大规模预训练，用户可在其基础上通过迁移学习微调应对不同文本处理任务，被广泛应用和关注。

2.2 BERT 模型

BERT 模型是以双向 Transformer 为基础，面向掩码模型(masked language model)和下一句判断(next sentence prediction)任务构建的深度学习模型。当前，采用大量文本作为数据集，预训练而成的 BERT 模型已成为处理多项自然语言处理 (natural language processing, NLP)任务的通用架构。该模型具有以下典型特征。

(1) Attention机制。相比传统的卷积神经网络（convolutional neural networks, CNN）和循环神经网络（recurrent neural network, RNN）语言模型，BERT引入了注意力(Attention)机制，能更直接地处理词间关系[12]。该机制由MNIH等[13]于2014年首次提出并运用至图像处理任务，效果良好。此后，注意力机制又被BAHDANAU等[14]引入NLP领域，并逐渐在NLP领域的Seq2Seq(sequence to sequence)任务中得到广泛应用。

如图1所示，在Seq2Seq模型中，该机制不但将Encoder生成的单个向量提供给Decoder，而且将其文本处理过程中的每个状态向量都输入Decoder，令后续步骤自行提取信息。

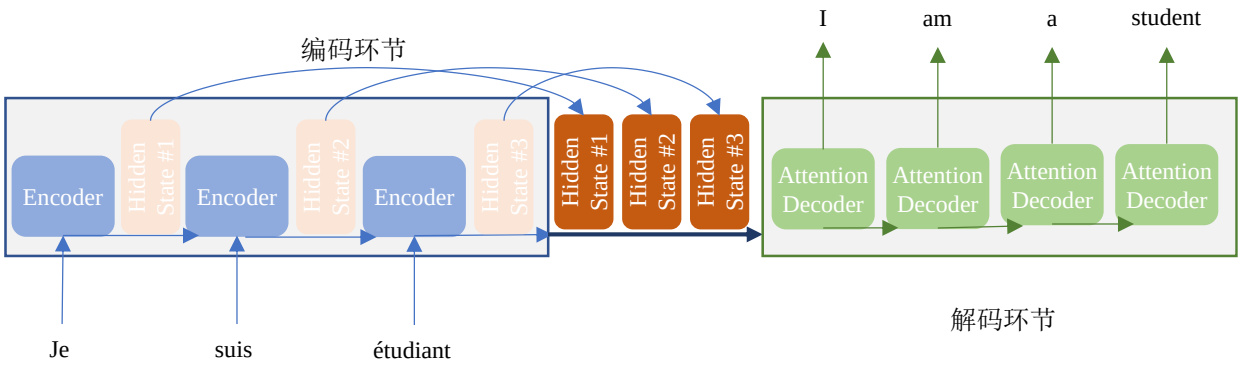

图1 机器翻译中Seq2Seq示意图

Fig. 1 Schematic diagram of Seq2Seq in machine translation

传统的 Encoder 和 Decoder 都是基于 RNN 实现的，难以有效进行并行运算。为了提高计算效率实现并行运算，谷歌团队在 2017 年将注意力机制运用到 Encoder 和 Decoder 中，提出了自注意力机制(Self-Attention mechanism)[15]。

如图 2 所示，自注意层中首先对输入序列I进行编码得到$a_1, a_2, a_3, \cdots, a_n$，再对每个$a_i$做 3 次线性变换得到矢量$\boldsymbol{q_i}$，$\boldsymbol{k_i}$，$\boldsymbol{v_i}$，分别计算$\boldsymbol{q_i}$和整条文本中每个矢量$k$的相似度$Similarity$，经过 *Softmax* 运算后得到权重，分别与对应的$\boldsymbol{v}$相乘得

$$o_i = \sum_{m=1}^{m=n} Softmax\left(Similarity(\boldsymbol{q_i}, \boldsymbol{k_m})\right) \boldsymbol{v_m} \quad (1)$$

其中，计算矢量相似度$Similarity$的函数有多种形式，该机制中通常采用向量点积。

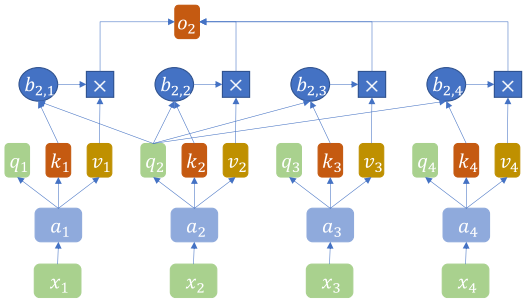

图2 Seq2Seq中的自注意力计算流程

Fig. 2 Self-Attention calculation process in Seq2Seq

而多头注意力机制(multihead self-attention)则是多次计算每个输入矢量a_i的q，k，v，例如，q_i由$q_i^1, q_i^2, \ldots, q_i^h$组成，从而将式(1)重复$h$次，将$h$次的结果拼接并线性变换后得到输出结果。多头注意力机制可以从多个空间提取文本信息，从而全面学习到文本的各种特征。

(2) Transformer模型。以自注意层替换Seq2Seq模型中的RNN层来构建Encoder和Decoder，从而形成Transformer模型。在运算过程中，由于没有先后顺序，可以进行高效运算。同时，需要引入位置编码(Positioning Encoding)表示顺序信息。

(3) BERT整体架构。基于Transformer，BERT模型主要由BERTLARGE和BERTBASE 2个参数大小不同的模型组成。BERTLARGE由每个包含16个头的24个自注意力层组成，中间向量维度是1024；BERTBASE则有12个自注意力层，每层有12个头，输出的向量维度是768。此外，Facebook还提出了改进的RoBERTa模型，由于改进了训练方法，此模型能更好地表示语义特征[16]。

3 研究方法

3.1 整体框架

考虑到BERT模型的优越性，针对灾害推文分类问题，本研究提出如图3所示的研究框架，主要包括以下4个步骤。

(1) 数据清洗。由于数据集中除了文本之外，还包含大量其他信息，如标点符号、表情包、停词(指没有实际含义的代词、冠词、数词、感叹词等，使用停词越多，文本越口语化)和网址链接等，因此首先需要对无关文本信息进行清理，删除停词、表情符号等对文本意义无明显贡献的符号。

(2) 经典方法基准构建。针对推文特点，提取网址链接、hashtag标签等文本参数特征，并通过文本向量化表示和卡方特征优化，构建经典的朴素贝叶斯分类模型，作为文本分类的经典方法对照基准模型。

朴素贝叶斯分类法的基础是贝叶斯定理及条件独立假设。设训练样本中的每一个实例x由m个属性值及1个类标签构成，即$\{x_1, x_2, \cdots, x_m\}$，其中类标签$c$取值自有限集$C\{c_1, c_2, \cdots, c_n\}$，模型旨在计算某个测试样本$X$=$x$的最可能的类标签$c$。贝叶斯公式是一种由先验概率、似然概率得到后验概率的方法。

(3) BERT模型迁移学习。BERT预训练采用的语料库主要是BooksCorpus(800 M words)，English Wikipedia(2 500 M words)，另外，训练过程只提取了Wikipedia的文本段落，忽略列表、表格和标题，为了提取长连续序列，语料库主要使用文档级语料库，而不是无序的句子级语料库[11]。这些语料库数据量大且较全面，一般包括了所有的日常生活用语，虽然不是专门的灾害相关语料库，但由于其适用性较广也适合灾害场景。

以基于大规模语料库预训练的BERT模型为基础，通过迁移学习优化参数，并引入不同的后续分类器，建立一系列基于BERT的深度学习模型。具体而言，后续分类器主要采用线性分类器、CNN网络、LSTM网络以及LSTM+CNN组合等4种形式。

(4) 模型性能评估对比.引入正确率、AUC、F1值等评价指标对不同模型进行性能评估，选择最优模型。其中，正确率衡量了模型分类的正确性；AUC是接收者操作特征(receiver operating characteristic, ROC)曲线包围的面积，能有效评价不同类别样本数目不均衡时的分类性能[17]；F1值则通过统计召回率和精确率来综合评估模型性能，作用与AUC相似。考虑到本研究采用的数据集中

正负样本数量相近，用正确率作为模型性能主要评价指标。

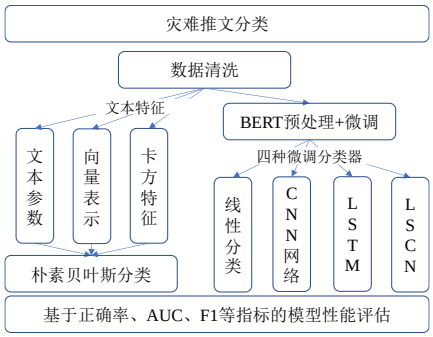

图3 整体框架

Fig. 3 Framework of the classification method

3.2 文本特征提取与表示

文本表示是将文字表示为向量作为模型输入数据的方法，典型的文本表示方法包括不考虑词序、语法的词袋模型[18]、词频-逆文档频度 (term frequency-inverse document frequency, TF-IDF)[19]以及基于无监督学习的Word2vec(Word to Vector)模型[20]等。不同分类模型采用的Tokenization方法并不完全一致，朴素贝叶斯分类模型主要采用词袋法和TF-IDF方法，神经网络分类模型采用keras的embedding层进行词嵌入，在含有BERT预处理的分类模型中，主要依靠BERT模型的基于WordPiece的词嵌入方法。

对于朴素贝叶斯模型的输入向量，本文主要采用词袋模型和TF-IDF算法进行特征提取，其2种方法提取到的向量矩阵为稀疏矩阵，需要利用卡方特征提取合适的特征数量以提高朴素贝叶斯模型的性能。

而在基于BERT微调的分类方法中，可直接采用BERT模型的输出作为文本的向量表达，随后结合后端深度神经网络模型训练分类器。BERT模型主要由transform层构成，如图2所示，在自注意力计算过程中，每个单词创建3个不同的向量$\boldsymbol{Q}$，$\boldsymbol{K}$和$\boldsymbol{V}$，输出的是根据$\boldsymbol{Q}$和$\boldsymbol{K}$处理后的$\boldsymbol{V}$向量[15]。

3.3 基于 BERT 的迁移学习

在对输入数据进行预处理的过程中，除了需要对文本词语、位置进行编码外，还需要加入段落向量表征句子对信息。序列首位一般需要添加特殊标记[CLS]，表示一条文本或一对句子，且句子中的分隔处也要加上分隔符[SEP]。

使用者可以针对不同任务自行对BERT模型进行微调，即采用不同结构和参数的深度神经网络进行训练。本模型整体是一个encoder，没有decoder部分，因为模型需要输出一个固定长度的向量，依此向量与label之间的关系对文本进行分类，整个模型包含BERT和神经网络(由CNN、LSTM、池化层、全连接层组成)，每个单词编码之后的向量维度为768。

由于本文的分类对象是文本序列，而处理序列的2种基本的深度学习算法分别是RNN和一维CNN[21]，且长短时记忆网络(long short-term memory-convolutional neural network, LSTM)能有效地避免梯度消失和梯度爆炸，有更好的适用性。因此，对于微调模型，本实验采用了线性分类层、CNN、LSTM以及CNN+LSTM相结合的深度神经网络进行训练。线性分类层结构比较简单，可用于初步分类；由于输入的数据是单通道的向量，所以CNN模型采用一维卷积结构，由5层卷积核为4的卷积层和3层池化层组成；而LSTM分类器则是由12层隐藏层神经元个数为240的LSTM层构成，后续接RELU激活层和线性输出层；LSTM和CNN结合的网络层则由4层LSTM层和4层卷积层组成。

4 结果分析

4.1 数据集概况

本文选择了Kaggle提供的数据集，共包括10 872条数据，其中，正负样本数分别为6 530和4 342，每条数据包含有keyword、location、text(推文文本)、target(1表示灾难，0表示非灾难)等信息。部分数据见表1。训练集和测试集数量分别为6 090和1 523，比例约为4∶1，句子最大长度为84，平均长度为15.27。本实验环境包括：CPU为32核Intel(R) Xeon(R) Silver 4215R，显卡为Tesla T4，操作系统为：Ubuntu 18.04，采用的框架是Tensorflow，PyTorch和Keras。

表 1 示例文本

Table 1 Text sample

ID	Keyword	Location	Text	Target
1	-	-	Our Deeds are the Reason of this #earthquake	1
52	ablaze	Philadelphia	Crying out for more! Set me ablaze	0
4424	electrocute	-	Why does my phone electrocute me	0

4.2 模型影响因素分析

在基于 BERT 预训练模型的微调过程中，对学习率、预训练模型种类、batchsize、dropout 层、输入特征和分类算法进行分析，实验结果如下：

(1)学习率和预训练模型种类。为探究学习率对模型训练效果的影响，为模型设置了从0.000 04到0.000 15的共计12个学习率数值，对照组的预训练模型是“bert-base-uncased”，另外3组的模型分别是“bert-base-cased”，“bert-large-uncased”和“RoBERTa”，前2个是基于$BERT_{BASE}$预训练的模型，且第1个模型将所有字母转化为小写，而第2个则保留了大小写；最后2个分别是基于$BERT_{LARGE}$和RoBERTa预训练的模型。模型的批量尺寸均为32，dropout层参数为0.5，加入的隐藏层第二维大小均为50。

每个模型训练4个epoch，将模型在每个epoch的训练集和验证集的正确率和损失值平均后可得到训练集损失值、验证集损失值和验证集正确率随学习率变化的图像(图4)。

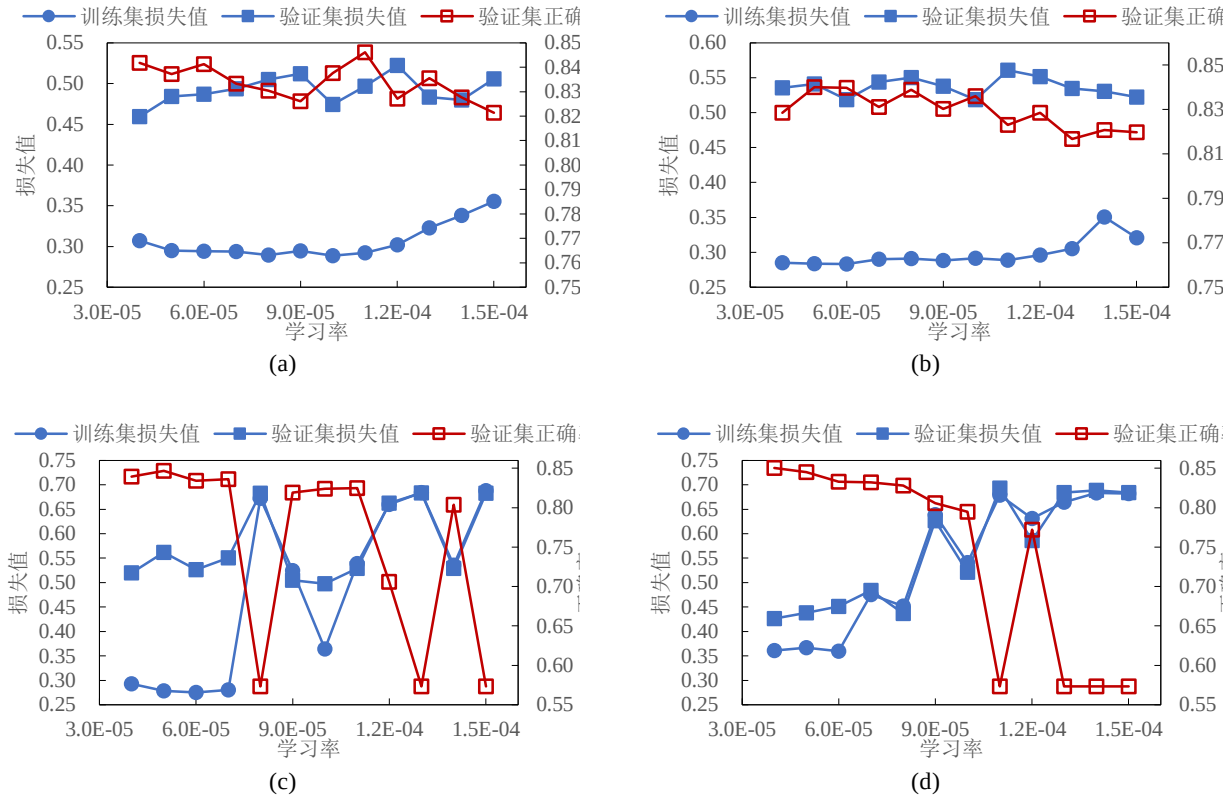

图4 不同预训练模型在验证集和训练集上的性能((a)“bert-base-uncased”预训练模型；(b)“bert-base-cased”预训练模型；(c)“bert-large-uncased”预训练模型；(d)“RoBERTa”预训练模型)

Fig. 4 Performance of different pre-trained models on the validation set and training set ((a) bert-base-uncased pre-trained model; (b) bert-base-cased pre-trained model; (c) bert-large-uncased pre-trained model; (d) RoBERTa pre-trained model)

由图4可以看出，基于$BERT_{BASE}$预训练的微调模型受学习率变化影响较小；而基于$BERT_{LARGE}$预训练的微调模型则受学习率数值影响较大，且学习率数值较小的模型效果更优。从验证集正确率角度，性能排序为RoBERTa，$BERT_{LARGE}$和$BERT_{BASE}$，且保留大小写对模型性能影响较小。另外，这几种方法的训练集损失值和验证集的损失值相近，说明经过BERT预处理后，模型基本不会出现过拟合。由表2可知，较于普通的传统模型(朴素贝叶斯模型的最优识别正确率为82.96%)，基于BERT预训练加微调后的模型正确率提升了1%以上。

表2 不同预训练模型的测试效果

Table 2. Test results of different pre-trained models

预训练	学习率	AUC	Accuracy	F1
bert-base-uncased	0.000 04	0.886 8	**0.828 0**	**0.828 2**
	0.000 06	0.892 1	0.824 0	0.824 5
	0.000 10	**0.893 4**	0.820 7	0.821 1
bert-base-cased	0.000 04	**0.881 6**	0.808 3	0.808 8
	0.000 08	0.876 0	0.821 4	0.821 5
	0.000 10	0.877 7	**0.822 1**	**0.821 9**
bert-large-uncased	0.000 04	**0.884 5**	0.821 4	0.822 0
	0.000 06	0.884 1	**0.826 0**	**0.826 2**
	0.000 09	0.830 2	0.819 4	0.818 4
RoBERTa	0.000 01	**0.891 8**	0.841 1	0.840 8

0.000 02	0.891 3	**0.843 7**	**0.843 5**
0.000 07	0.880 9	0.831 9	0.830 7

(2)批量尺寸。模型微调过程中的批量大小也会对模型训练产生影响，本次实验对基于“bert-base-uncased”预训练的模型分别设置了3种批量，即16，32和64，测试集结果见表3。

表3 不同批量尺寸的测试结果(学习率0.000 04)

Table 3 Test results of different batch sizes

(learning rate: 0.000 04)

批量	AUC	Accuracy	F1
16	0.890 6	**0.827 3**	0.827 5
32	0.886 8	**0.828 0**	0.828 2
64	0.890 5	**0.833 9**	0.833 9

测试结果表明，批量尺寸越大，模型的预测正确率越高。

(3)dropout层。在微调过程中是否加入dropout层也会影响模型性能，本实验对基于“bert-base-uncased”预训练的模型进行了探讨，测试结果见表4。

表4 dropout层对测试结果的影响(学习率0.000 04)

Table 4 Impact of dropout layer on test results

(learning rate: 0.000 04)

dropout	AUC	Accuracy	F1
0.5	0.886 8	**0.828 0**	0.828 2
0.8	0.893 1	**0.828 6**	0.828 7
无	0.885 4	**0.835 2**	0.835 4

测试结果表明：在微调过程中，用户设计的网络层中加入dropout层会对模型性能造成不利影响。

(4)文本输入特征。一般情况下的文本分类多使用BERT词向量化后的第一个token[CLS]作为后续模型的输入，为了比较不同输入的识别效果，本文比较了第一个和最后一个token及句子整体向量3种不同输入的模型性能(批量尺寸为32)。

表5在本任务中，将最后一个token向量和句子整体向量作为输入的识别效果更好。这可能是由于本任务中，句子的整体含义更具代表性。

表5 不同输入向量的测试结果(学习率0.000 04)

Table 5 Test results of different input vectors

(learning rate: 0.000 04)

输入	AUC	Accuracy	F1
首个token	0.886 8	**0.828 0**	0.828 2
末尾token	0.890 3	**0.833 2**	0.833 4
句子向量	0.885 5	**0.834 5**	0.834 6

(5)不同模型的识别效果对比。研究对比了不同模型后端的分类算法的识别效果，除简单的线性分类层，还引入了CNN和LSTM网络以及LSTM+CNN的串联网络(LSCN)进行分类。

通常情况下，朴素贝叶斯分类器正确率一般在78%左右。通过衡量特征词与分类标签的关联程度，经过卡方特征筛选文本数据特征可以对模型进行优化。优化后，当卡方特征数量保留在5 000左右时，朴素贝叶斯模型的分类效果达到最佳，识别正确率可达83%。而只利用LSTM和CNN，没有BERT预处理的神经网络分类模型，测试准确率只能达到80%(此模型采用的embedding词汇表大小为5 000，对每条文本的前55个单词编码，batchsize为96，学习率为0.001，CNN卷积层和池化层与微调模型结构基本一致)。

以该模型为基准，表6为各模型的分类结果。由表6可知，基于BERT微调的分类器精度较高，性能与基准NB模型相当；当结合LSTM和CNN后，其识别效果则显著优于最佳NB模型。实验表明：将LSTM+CNN的串联网络“LSCN”模型作为微调分类器，学习率设为0.000 02，不设dropout层的分类识别模型最优，正确率近85%，且由于数据集较小，采用更大的batchsize不会显著提升模型性能，不同BERT模型预处理后的准确率差别不大。

表6 不同分类模型的测试结果

Table 6 Test results of different classification models

模型	学习率	AUC	Accuracy	F1
线性层	0.000 04	0.886 8	**0.828 0**	0.828 2
CNN	0.000 04	0.889 0	**0.838 5**	0.838 3
LSTM	0.000 02	0.881 3	**0.840 4**	0.840 1
LSCN	0.000 02	0.897 1	**0.845 7**	0.844 9
NB	-	0.875 3	**0.829 6**	0.782 5

此外，BERT预训练过程涉及Masked Language Modeling和Next Sentence Prediction 2个主要任务。其中，前者类似于完形填空，利用全文的所有词句信息进行预测，这也体现了BERT模型的双向性；而后者则类似于语义推理，用来推测上下文的逻辑关系，适用于长文序列。针对本文灾害信息分类问题，前者更有价值，可以挖掘深层信息，而后者则适用性较低。因此，未来针对本文的分类场景，可在训练过程中进一步提高掩码的随机性，提升预训练模型的专业性。

5 结 束 语

针对灾害推文分类问题，本研究提出了基于BERT预训练和“LSTM+CNN”的分类模型，并以Kaggle竞赛中的灾难推文数据进行了实验分析与性能对比。由于训练所用数据规模较小，数据有限，这也是解决小样本文本分类的一种有效方法。结果表明：

(1) 相比经典的朴素贝叶斯分类方法和神经网络模型，基于BERT预训练模型的深度学习方法性能更加优异，将正确率提升了1.6%。

(2) 在微调和迁移学习过程中，是否加入dropout层及批量大小对本任务的模型性能有一定影响。针对BERT输出的词向量种类进行实验，结果表明用代表整个句子的词向量作为后续分类算法的输入效果最优。

(3) 本文还对比了CNN和LSTM等不同网络与BERT预处理模块结合的分类模型，结果表明两者均优于经典线性分类器，且两者相结合的LSTM+CNN模型最优。

综上所述，基于BERT预训练模型的深度文本分类方法可以有效解决小样本分类问题，可快速识别灾害相关推文、减少谣言传播，对提升灾害应急响应与沟通效率、建设韧性城市具有重要意义。尽管本研究过程中采用了英文文本数据集，但基本方法流程同样适用于中文灾害文本分类问题，只需针对性更换中文的预训练BERT模型即可。未来可以进一步引入跨语言分类算法[22-23]，构建统一的算法模型。

参考文献（References）

[1] PLOTNICK L, TUROFF M, WHITE C. Partially distributed emergency teams: considerations of decision support for virtual communities of practice[M]//Supporting Real Time Decision-Making. Boston: Springer US, 2010: 203-220.

[2] CHEN Y D, WANG Q, JI W Y. Rapid assessment of disaster impacts on highways using social media[J]. Journal of Management in Engineering, 2020, 36(5)：04020068.

[3] ACAR A, MURAKI Y. Twitter for crisis communication: lessons learned from Japan's tsunami disaster[J]. International Journal of Web Based Communities, 2011, 7(3): 392-402.

[4] WANG Y, TAYLOR J E, GARVIN M J. Measuring resilience of human–spatial systems to disasters: framework combining spatial-network analysis and fisher information[J]. Journal of Management in Engineering, 2020, 36(4): 04020019.

[5] 夏志杰，吴忠，栾东庆. 基于社会化媒体的突发事件应急信息共享研究综述[J]. 情报杂志, 2013, 32(10): 122-126, 121.

XIA Z J, WU Z, LUAN D Q. The literature review of information sharing research for emergency respond based on social media[J]. Journal of Intelligence, 2013, 32(10): 122-126, 121 (in Chinese).

[6] 何锴. 基于自然语言处理的文本分类研究与应用[D]. 南京: 南京邮电大学, 2020.

HE K. Research and application of text classification based on natural language processing[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2020 (in Chinese).

[7] YANG Y M, LIU X. A re-examination of text categorization methods[C]//The 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 1999: 42-49.

[8] 张华伟，王明文，甘丽新. 基于随机森林的文本分类模型研究[J]. 山东大学学报: 理学版, 2006, 41(3): 139-143.

ZHANG H W, WANG M W, GAN L X. Automatic text classification model based on random forest[J]. Journal of Shandong University: Natural Science, 2006, 41(3): 139-143 (in Chinese).

[9] 李蓉，叶世伟，史忠植. SVM-KNN分类器: 一种提高SVM分类精度的新方法[J]. 电子学报, 2002, 30(5): 745-748.

LI R, YE S W, SHI Z Z. SVM-KNN classifier—A new method of improving the accuracy of SVM classifier[J]. Acta Electronica Sinica, 2002, 30(5): 745-748 (in Chinese).

[10] 陈翠平. 基于深度信念网络的文本分类算法[J]. 计算机系统应用, 2015, 24(2): 121-126.

CHEN C P. Text categorization based on deep belief network[J]. Computer Systems & Applications, 2015, 24(2): 121-126 (in Chinese).

[11] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language uUnderstanding[EB/OL]. [2021-09-08]. https://arxiv.org/abs/1810.04805v1.

[12] 王楠禔. 基于BERT改进的文本表示模型研究[D]. 重庆: 西南大学, 2019.

WANG N (T /Z). Research on improved text representation model based on BERT[D]. Chongqing: Southwest University, 2019 (in Chinese).

[13] MNIH V, HEESS N, Graves A, et al. Recurrent models of visual attention[C]//The 27th International Conference on Neural Information Processing Systems. New York: ACM Press, 2014: 2204-2212.

[14] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. [2021-08-12]. https://arxiv.org/abs/1409.0473.

[15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. [2021-08-10]. https://arxiv.org/abs/1706.03762v5.

[16] LIU Y H, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized BERT pretraining approach[EB/OL]. [2021-07-22]. https://arxiv.org/abs/1907.11692 .

[17] 陈慧灵. 面向智能决策问题的机器学习方法研究[D]. 长春: 吉林大学, 2012.

CHEN H L. Research on machine learning methods for intelligent decision-making[D]. Changchun: Jilin University, 2012 (in Chinese).

[18] 黄春梅，王松磊. 基于词袋模型和TF-IDF的短文本分类研究[J]. 软件工程, 2020, 23(3): 1-3.

HUANG C M, WANG S L. Research on short text classification based on bag of words and TF-IDF[J]. Software Engineering, 2020, 23(3): 1-3 (in Chinese).

[19] 施聪莺，徐朝军，杨晓江. TFIDF算法研究综述[J]. 计算机应用, 2009, 29(S1): 167-170, 180.

SHI C Y, XU C J, YANG X J. Study of TFIDF algorithm[J]. Journal of Computer Applications, 2009, 29(S1): 167-170, 180 (in Chinese).

[20] 江大鹏. 基于词向量的短文本分类方法研究[D]. 杭州: 浙江大学, 2015.

JIANG D P. Research on short text classification based on word distributed representation[D]. Hangzhou: Zhejiang University, 2015 (in Chinese).

[21] 弗朗索瓦·肖莱. 张亮译. Python深度学习[M]. 北京: 人民邮电出版社, 2018: 147.

Deep learning with Python[M]. Beijing: Posts & Telecom Press, 2018: 147 (in Cjinese).

[22] 刘星佐. 跨语言文本分类技术研究[D]. 长沙: 国防科学技术大学, 2016.

LIU X Z. Research on cross-language text classification technology[D]. Changsha: National University of Defense Technology, 2016 (in Chinese).

[23] 高影繁，王惠临，徐红姣. 跨语言文本分类技术研究进展[J]. 情报理论与实践, 2010, 33(11): 126-128, 104.

GAO Y F, WANG H L, XU H J. Progress in research on cross-language text categorization technology[J]. Information Studies: Theory & Application, 2010, 33(11): 126-128, 104 (in Chinese).