2021 年 第 21 卷 第 35 期
2021, 21(35): 15284-06

科 学 技 术 与 工 程
Science Technology and Engineering

ISSN 1671—1815
CN 11—4688/T

引用格式:李 忠, 杨百一, 李 莹, 等. ALBERT 与双向 GRU 的多标签灾情信息预测[J]. 科学技术与工程, 2021, 21(35): 15284-15289.
Li Zhong, Yang Baiyi, Li Ying, et al. Multi-tag disaster information prediction based on ALBERT and bidirectional GRU[J]. Science Technology and Engineering, 2021, 21(35): 15284-15289.

环境科学、安全科学

ALBERT 与双向 GRU 的多标签灾情信息预测

李 忠, 杨百一, 李 莹*, 李晓丽
(防灾科技学院应急管理学院, 廊坊 065201)

摘 要 针对灾害求助信息辨识不准确、响应能力不足的问题,提出了基于 ALBERT(a lite bidirectional encoder representation from transformers) 和双向 GRU(gate recurrent unit)的文本多标签分类模型。首先利用 ALBERT 预处理语言模型对文本信息进行编码,获取文本的动态词特征向量,并送入双向 GRU 神经网络进行训练,根据不同的单词赋予不同的权重,应用 Attention 机制进行解码。利用模拟退火算法求解最优阈值,以微平均值作为评价函数,确定样本的标签类别归属。与逻辑回归、朴素贝叶斯和长短期记忆 (long-short term memory,LSTM)神经网络等模型进行比较。结果显示,多标签分类模型具有更高的准确率,达到 95%,汉明损失仅到 0.05,能够更好地辨别灾情求助信息,提高救援效率。

关键词 动态词特征向量; 多标签排序; Attention 机制; 微平均值
中图法分类号 X922.2; **文献标志码** A

Multi-tag Disaster Information Prediction Based on ALBERT and Bidirectional GRU

LI Zhong, YANG Bai-yi, LI Ying*, LI Xiao-li
(Emergency Management College, Institute of Disaster Prevention, Langfang 065201, China)

[**Abstract**] Aiming at the problems of inaccurate identification and insufficient response ability of disaster help information, a text multi label classification model based on ALBERT and bidirectional GRU was proposed. Firstly, the text information was encoded by using the ALBERT preprocessing language model, and the dynamic word feature vector of the text was obtained. Then, it was sent to the bidirectional GRU neural network for training. According to different words, different weights are given, and the attention mechanism was used for decoding. The simulated annealing algorithm was used to solve the optimal threshold, and the micro average value was used as the evaluation function to determine the label category of samples. Compared with logistic regression, Naive Bayes and long-term and short-term memory (LSTM) neural network, the results show that the multi label classification model has a higher accuracy rate of 95%, and the Hamming loss is only 0.05, which can better identify the disaster information and improve the rescue efficiency.
[**Keywords**] dynamic word feature vector; multi label sorting; Attention mechanism; micro average value

自然灾害事件具有紧急性和突发性的特点,往往会造成人员伤亡和社会经济的巨大损失。正因为如此,在巨灾发生时,如何快速、有序、高效、准确地收集灾情信息是迫切需要解决的问题。社交媒体作为在灾难响应和恢复阶段的重要部分,及时响应受害者的请求可以极大提高受害者的生存概率[1]。目前已经有很多研究人员运用灾难中社交媒体发布的数据来获取灾区的情况,但往往会存在信息获取能力不足的问题,包括不能全方位多角度地获取信息和信息关联能力不强,不能对收集上来的信息做出准确的划分,以至于错失许多关键信息[2]。社交媒体可以收集大量与灾情有关的信息,可以帮助应急管理人员面对不同规模、范围和性质的灾情。但是,关键信息的错误传递,可能会导致救援效率降低。如何指定有效的信息传递战略,是目前亟待解决的问题。为此中外人员针对社交应急响应数据作了相关研究。国外研究人员主要在推特上收集灾难响应数据,及时发现灾区情况,如

收稿日期: 2021-06-29; **修订日期**: 2021-08-25
基金项目: 中央高校基本科研业务费专项(ZY20180121)
第一作者: 李 忠(1966—),男,汉族,山东诸城人,博士,教授,硕士研究生导师。研究方向:人工智能、信号处理。E-mail:lizhong@cidp.edu.cn。
* **通信作者**: 李 莹(1996—),女,汉族,内蒙古通辽人,硕士。研究方向:人工智能。E-mail:1169231950@qq.com。

Zahra 等[3]利用在推特上收集的信息,以识别灾害发生后的目击者类别为目的建立文本分类器,以此来提高在灾难中识别具有价值的目击者报告;Benitez 等[4]通过推特收集与灾害相关的信息对文本进行分类,针对数据量大、分类精度低等问题,利用遗传算法降低特征空间的维数,去除不相关的、冗余的特征以此来提高分类精度。中国研究人员主要以微博数据为依据,通过收集相关的灾难响应数据,进行文本分类以达到了解灾区灾情的目的。吴先华等[5]通过微博大数据获取有关暴雨内涝灾情的数据,利用 TF-IDF(term frequency-inverse document frequency)进行特征选择,建立暴雨灾情内涝表,以此来获取灾区的灾情以及公众情绪。徐敬海等[6]以带有位置的地震灾情信息作为数据源,通过应用反距离空间差值法,实现微博灾情信息面状化。王琳等[7]针对地震应急信息,通过"关键词分类"将地震应急信息分为震前基础信息、震中灾情信息和应急援助信息,以达到地震应急信息自动分类的目的,提高对应急响应信息处理的效率。

虽然上述方法可以实现在大量灾难响应信息中获取关键内容,但同一条灾难响应消息可能存在多个标签。如何从一条相关灾难响应的社交媒体文本中提取所有关键信息,并根据匹配度对标签排序,需要对灾难响应文本数据进行多标签数据挖掘。现针对含有多标签的灾难响应信息,利用深度神经网络模型提取相应的标签信息,有效提高对灾难响应信息的利用效率,达到尽快实施救援的目的。

1　文本分类模型

文本分类研究是数据挖掘技术应用的一个重要领域,涉及特征提取、算法设计、软件开发等多个方面,前人已经做了很多基础性研究工作。

1.1　词嵌入模型

所谓词嵌入模型,就是将文本信息转化为数字向量的模式。最原始的算法是将文本信息转化为 0-1 模型,即 one-hot 编码,用 0 和 1 两个数字来表示字符是否出现在文本中,但这种方法会因为数据量变大而导致计算变得复杂。Salton 等[8]提出 TF-IDF 模型,TF-IDF 是计算特征词在文本中的权重,其取值范围在[0,1]之间,其中 TF 指该特征词在这篇文档中的词频。虽然这种方法可以提取出文本中的特征词,但是如果把所有词都作为特征项,会导致特征向量过大而使得计算速度下降,那么如何在不损害文本特征信息的同时降低特征性向量的维度呢? Karen 等[9]提出了逆文档词频(inverse document frequency, IDF),表示区分文本的能力。Mikolov 等[10]提出了 Word2Vec 模型,一种利用神经网络进行词嵌入的学习方法,但该模型未能有效地利用所有文本信息。BERT(bidirectional encoder representation from transformers)模型,建立在 Transformer 之上,具有强大的语言表征能力和特征提取能力[11],但该模型可复现性差,模型收敛速度较慢,需要极高的算力才行。为改进 BERT 模型的收敛慢问题,Lan 等[12]提出了一种双向编码特征表示模型——ALBERT(a lite BERT)文本预训练语言模型,其先对大量文本数据进行预训练,再根据下游任务进行调整,达到提高下游预测效果的目的。ALBERT 预训练语言模型是基于 BERT 模型的双向 Trandformer 编码器的特征表示方式,可以在保证较少参数量的同时,提取文本的主要特征。

1.2　多标签分类模型

目前,关于多标签数据挖掘问题,主要分为两类:一类是多标签分类,另一类是多标签排序[12]。二者的不同在于后者可以返回类别标签的同时对标签匹配度进行排序。

从模型建立上看,文本的多标签分类一般有两种:一种是传统的机器学习模型,如 SVM、DT、Naïve Bayes、DT、Xgboost 等算法;另一种是深层网络机器学习模型,如 RNN、CNN、LSTM 和 GRU 等。从解决问题角度上看,文本的多标签分类包括问题转换法和算法适应法,其中问题转化法是将多标签文本分类转化为一个由多个单标签分类组合而成的问题,这样对多标签分类的问题求解就变成对单标签文本分类问题的求解,如 BR(binary relevance)二元关系法和 LP(label powerset)标签幂集法等[12]。而算法适应法则是通过对算法的改进,使二分类算法也可以解决多标签分类问题。本文中所使用的多标签分类算法是利用双向 GRU 神经网络对灾难响应信息的标签进行预测。

2　结合 ALBERT 和双向 GRU 的文本多标签分类

若要对文本信息进行多标签分类,需要先将文本进行预处理,再将预处理后的文本信息利用 ALBERT 预训练模型提取特征,得到词嵌入向量。根据提取的词嵌入向量,利用 Attention 机制的双向 GRU 神经网络进行训练,以到达多标签分类预测的目的。

2.1　ALBERT 模型

为了利用模型对文本进行多标签预测,需要将一条文本信息利用向量来表示,这可以利用词嵌入模型实现。为了解决 BERT 参数量大的问题,

ALBERT 采用双向 Transform 的 encoder 结果表示文本的特征,模型结构图如图 1 所示。

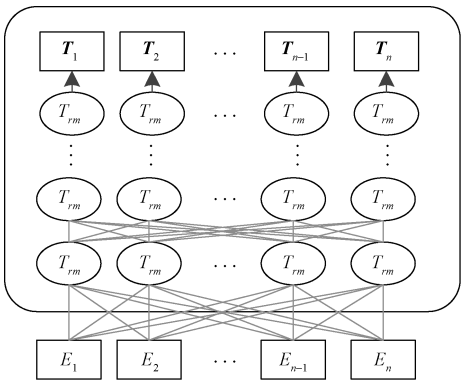

图 1 ALBERT 模型结构图

Fig. 1 Structure diagram of ALBERT model

$\boldsymbol{T}_i$ 表示句子中第 i 个词的特征向量,E_i 表示文中第 i 个词的序列化字符。在 BERT 的基础上,ALBERT进行了 3 个方面的改进[13]。

2.1.1 词嵌入向量因式分解

通过对大的词特征嵌入矩阵分解,得到两个相对较小的矩阵,从而将隐含层分离出来。利用因式分解,将模型的复杂度从 $O(VH)$ 降到 $O(VE + EH)$,复杂度变化为

$$O(VH) \rightarrow O(VE + EH) \tag{1}$$

式(1)中:V 表示词汇表长度;H 表示隐含层大小;E 表示词嵌入大小,当 $E \ll H$ 时参数量会明显减少。

2.1.2 跨参数共享

针对参数过多的问题,ALBERT 利用多层参数共享策略减少参数,包括全连接层和注意力层的参数共享。这种做法可以避免当模型网络深度增加时,参数数量也随之增加。

2.1.3 段落连续任务

ALBERT 对 BERT 中的 NSP(next sentence prediction)任务进行改进,为了消除主题识别的影响,提出了 SOP(sentence-order prediction)任务,可以在同一个文本中选取正负样本,不受句子顺序的影响。

2.2 双向 GRU 神经网络

传统的神经网络是单向的,状态从前向后传播。但是在多标签文本分类中,为了对文本进行更深层次的特征提取,让前一时刻的输出与后一时刻的输出联系在一起,采用双向 GRU 神经网络模型[14],其结构如图 2 所示。

从图 2 中可以看出,双向 GRU 可以看作两个单向 GRU 组合而成,在 t 时刻隐含层的状态由其前后时刻的隐含层状态决定,可表示为

$$\overrightarrow{h_t} = \mathrm{GRU}(x_t, \overrightarrow{h_{t-1}}) \tag{2}$$

$$\overleftarrow{h_t} = \mathrm{GRU}(x_t, \overleftarrow{h_{t-1}}) \tag{3}$$

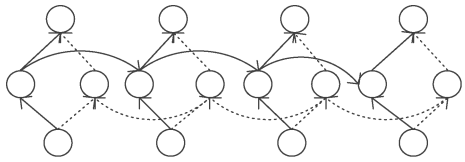

图 2 双向 GRU 神经网络结构图

Fig. 2 Structure diagram of bidirectional GRU neural network

$$h_t = w_t \overrightarrow{h_t} + v_t \overleftarrow{h_t} + b_t \tag{4}$$

式中:函数 GRU( ) 表示对输入词向量的非线性变换,通过词向量编码得出相应的 GRU 隐含层信息。x_t 表示双向 GRU 在 t 时刻的输入,w_t、v_t 分别表示 t 时刻双向 GRU 所对应的前向隐层状态 $\overrightarrow{h_t}$ 和反向隐含层状态 $\overleftarrow{h_t}$ 所对应的权重;b_t 表示 t 时刻隐含层状态所对应的偏置。

2.3 Self-Attention 机制

Self-Attention 机制是从大量的文本信息中提取出具有价值的信息,在自然语言问题中,需要对文本进行编码和解码[15]。编码器负责把文本转化为指定长度的词向量,即该文本的语义。解码器负责将词向量作为输入变量送入模型中,生成指定长度的序列。Self_Attention 机制就是在解码的过程中对不同的字符赋予不同的权重,计算公式为

$$u_t = v_t \tanh(w\boldsymbol{h} + w\boldsymbol{d}) \tag{5}$$

$$a_t = \mathrm{softmax}(u_t) \tag{6}$$

$$c = \sum_{l}^{L} a_{ll} h_l \tag{7}$$

式中:u_t 表示输出向量与解码时状态的相似度计算;l 表示向量的个数;w 表示权重;$\boldsymbol{h}$ 表示每一步的输出向量;$\boldsymbol{d}$ 表示解码时每一时刻的状态;c 是计算向量匹配度后的权值相加,计算后的向量用于解码器中每个时间步的输入;a 表示向量匹配度 softmax 后的匹配权重。

2.4 模型建立

令输入样本表示为 $X = \{X_1, X_2, \cdots, X_n\}$,标签样本表示为 $Y = \{Y_1, Y_2, \cdots, Y_m\}$,其中 n 表示有 n 条样本信息,m 表示有 m 个类别的标签。$Y_i = 1$ 时表示文本属于该标签,$Y_2 = 0$ 表示不属于该标签。

首先利用 ALBERT 预处理模型对文本进行向量转化,转化后的向量矩阵用 $\boldsymbol{E} = (\boldsymbol{E}_1, \boldsymbol{E}_2, \cdots, \boldsymbol{E}_n)$ 表示,即词嵌入向量,E_i 表示第 X_i 字符的向量表示。再将输出的词嵌入向量进行双向 Transform 编码,得到文本的特征向量 $\boldsymbol{T} = (T_1, T_2, \cdots, T_n)$,作为双向 GRU 网络的输入。结合 Attention 机制,进一步提取文本中的关键信息,最后,以交叉熵作为损失函数对神经网络模型进行编译。利用全连接层对结果进行归一化,得到响应信息的标签概率分布矩阵。计算流程如图 3 所示。

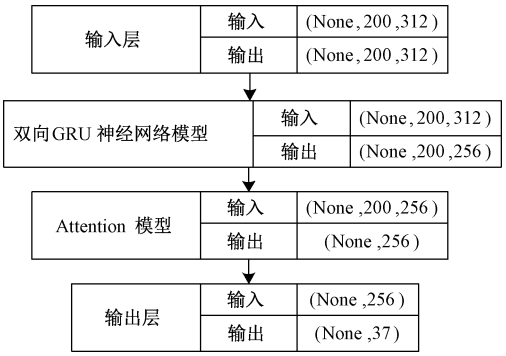

图 3　网络结构图

Fig. 3　Network structure diagram

3　实验结果与分析

3.1　数据源

本文中采用开源数据集(Appen)中的多语言灾难响应数据。该数据集共有 26 248 条数据,来自于若干年关于自然灾害的新闻报道,包括 2010 年海地地震、2010 年智利地震、2010 年巴基斯坦洪灾、2012 年美国超级风暴桑迪等。作为标准数据集,其中训练集、测试集与验证集的数据量如表 1 所示。

在该数据集中包含 37 个与应急响应信息相关的标签,每一条文本信息可以标记多个标签。根据数据集标签的属性,将其分为 3 个级别:一级标签为文本是否与灾情相关;二级标签将灾情信息分为五类,分别为援助相关、灾民相关、物资相关、灾况相关和灾区相关;三级标签为更具体的救灾信息分类。详细数据标签分类见表 2。

表 1　数据集划分

Table 1　Data set division

训练集	测试集	验证集	总数据量
21 046	2 629	2 573	26 248

表 2　数据标签分类详情

Table 2　Data label classification details

一级标签	二级标签	三级标签
灾情相关	援助相关	请求、提供、援助相关、医疗帮助、庇护所、搜索和救援、其他援助、援助中心
	灾民情况	失踪、灾民、伤亡
	物资相关	水、食品、衣服、钱、医疗产品
	灾况相关	安全、与天气相关、洪水、风暴、火灾、地震、寒冷、其他天气、直接报告
	设施相关	运输、基础设施相关、其他基础设施、建筑物、电力、工具、医院、商店

3.2　评价指标

针对文本多标签分类问题,评价指标主要有准确率(accuracy)、汉明损失(Hamming loss)、微平均 F_1 值(Micro_F_1)。

3.2.1　准确率

准确率 A 体现计算预测标签是否完全与实际标签相符合,其计算公式为

$$A = \frac{1}{m}\sum_{i=1}^{m}\delta(Z_i, Y_i) \tag{8}$$

式(8)中:m 为样本数量;Z_i 表示真实值;Y_i 表示预测值;$\delta(Z_i, Y_i)$ 表示真实标签是否完全匹配预测标签,若完全匹配,则 $\delta(Z_i, Y_i)$ 为 1,否则为 0。

3.2.2　汉明损失

汉明损失(L_{Ham})表示在所有预测的类别中错误样本的比例,数值越小,预测标签越准确。其计算公式为

$$L_{\text{Ham}} = \frac{1}{|N|}\sum_{i=1}^{|N|}\frac{\text{xor}(Z_i, Y_i)}{|M|} \tag{9}$$

式(9)中:N 为样本数量;M 为标签个数;Z_i 表示真实值;Y_i 表示预测值;xor ( )表示异或运算;Z_i 与 Y_i 相同为 0,不同为 1。

3.2.3　Micro_F_1

多标签前分类的 F_1-Score 分为微平均 Micro_F_1 和宏平均 Macro_F_1。其中微平均为计算所有类别标签的精确率和召回率后,再计算 F_1;宏平均为先计算出每一个类别标签的精确率和召回率,再计算平均的 F_1 值,该值越大,分类效果越好。为表示模型对整体文本的预测效果,本文中以微平均作为评价标准,其计算公式为

$$\text{Micro\_}P = \frac{1}{N}\frac{\sum_{i=1}^{N}\text{TP}_i}{\sum_{i=1}^{N}\text{TP}_i + \sum_{i=1}^{N}\text{FP}_i} \tag{10}$$

$$\text{Micro\_}R = \frac{1}{N}\frac{\sum_{i=1}^{N}\text{TP}_i}{\sum_{i=1}^{N}\text{TP}_i + \sum_{i=1}^{N}\text{FN}_i} \tag{11}$$

$$\text{Micro\_}F_1 = \frac{2\text{Micro\_}P\ \text{Micro\_}R}{\text{Micro\_}P\ \text{Micro\_}R} \tag{12}$$

式中:TP_i 表示真实值为 1、预测值也为 1 的样本数;FP_i 表示真实值为 0、预测值为 1 的样本数;FN_i 表示真实值为 0、预测值为 0 的样本数;Micro_P 表示预测的精确率;Micro_R 表示预测的召回率;Micro_F_1 即为微平均 F_1。

3.3　参数设置

部分双向 GRU 网络参数设置如表 3 所示。表中 Input_size 表示输入特征维数,Hidden_size 表示隐含层的维数,Embedding_size 表示词嵌入大小,Output_size 表示输出层维数。

3.4　结果与分析

训练结果如图 4 所示。从图 4 中可以看出,训

练集的损失和测试集的损失均不断下降,说明双向 GRU 网络仍在学习,且学习效果较好。在迭代次数(echo)为 10 时,测试集的损失值为 0.137,准确率达到 95.6%,说明模型预测的效果较好。

表 4 为不同词向量转化模型和文本模型的对比分析,其中词向量转化对比模型采用 TF-IDF,预测对比模型采用逻辑回归、朴素贝叶斯和 LSTM 长短期记忆神经网络。从表 4 中可以看出,本文中提出的文本多标签分类模型在准确率、F_1-Score、汉明损失上的效果均优于其他预测模型。

因本文中以 Sigmoid 作为双向 GRU 网络模型的激活函数,所以求得的结果为标签概率分布矩阵。若要达到标签自动分类的目的,需确定多标签分类阈值。本文中利用模拟退火算法对阈值求解,通过计算得出最优阈值为 0.34。将结果与阈值为 0.5 和 0.7 的结果进行对比,如表 5 所示。从表 5 中可以看出,利用模拟退火算法计算的阈值多标签分类效果优于其他阈值效果。

表 3 模型训练参数

Table 3 Model training parameters

参数	参数值
Input_size	128
Hidden_size	312
Embedding_size	128
Output_size	37
激活函数	Sigmoid
优化器	Adma
损失函数	交叉熵

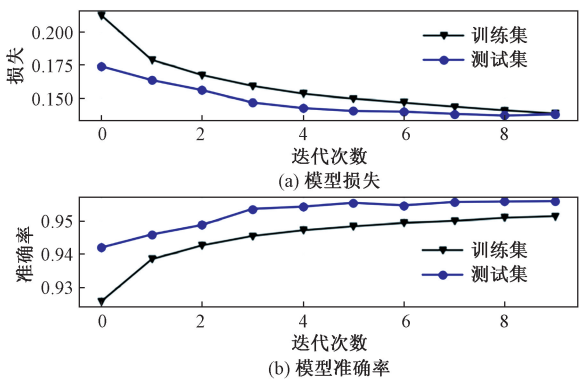

图 4 网络损失值与准确率变化趋势图

Fig. 4 Variationtrend of network loss value and accuracy

表 4 不同模型训练效果对比

Table 4 Comparison of training effects of different models

模型	Accuracy	Hamming loss	Micro-F_1
ALBERT-双向 GRU	0.95	0.66	0.037
ALBERT-双向 LSTM	0.94	0.65	0.053
TF_IDF + 逻辑回归	0.94	0.63	0.055
TF_IDF + 朴素贝叶斯	0.93	0.63	0.04

表 5 不同分类阈值预测效果对比

Table 5 Comparison of prediction effects of different classification thresholds

分类阈值	准确率	汉明损失	Micro-F_1
0.5	0.95	0.05	0.65
0.34	0.95	0.05	0.67
0.70	0.94	0.05	0.58

4 结论

在当今大数据时代,一旦有灾害发生,就会有若干人在社交媒体、数字网络等发布成千上万的灾难响应信息,如何识别出与灾情有关的信息并将请求帮助信息推送给救援组织,这对于提高救援质量和救援效率是非常重要的一项工作。从以下几点在分析了文本信息分类模型在灾后救援中的作用。

(1)针对大量的灾情响应信息,按照与灾害和救援相关的 37 个标签,提出了结合 ALBERT 和双向 GRU 神经网络的多标签文本信息分类模型,自动地对响应数据进行标签识别。

(2)以开源数据集(Appen)中的多语言灾难响应数据作为数据源进行计算,并将结果与逻辑回归、朴素 Bayes 和 LSTM 等方法的结果进行对比。

(3)实验结果显示,该模型可以有效地识别出文本的标签,准确率更高,汉明损失更小,能够更好地辨别灾情求助信息,提高救援效率。

参 考 文 献

[1] Klinsompus P, Nupairoj N. Critical message scheduling for disaster response and recovery phases[C]// International Conference on Information & Communication Technology Convergence. New York: IEEE, 2015: 65-70.

[2] 王恩雁, 李向阳. 大灾难应急响应中的跨组织信息共享与协调规划研究[C]//中国管理现代化研究会、复旦管理学奖励基金会. 第八届(2013)中国管理学年会——管理与决策科学分会场论文集. 上海: 中国管理现代化研究会, 2013: 9-18.
Wang Enyan, Li Xiangyang. Research on cross organizational information sharing and coordination planning in major disaster emergency response[C]//China Management Modernization Research Associ-ation and Fudan Management Award Foundation. The 8th (2013) China Management Academic Year Meeting-field Essays of Management and Decision Science Branch. Shanghai: China Management Moderni-zation Research Association, 2013: 9-18.

[3] Zahra K, Imran M, Ostermann F O. Automatic identification of eyewitness messages on twitter during disasters[J]. Information Processing & Management, 2020, 57(1): 13-21.

[4] Benitez I B, Sison A M, Medina R P. An improved genetic algorithm for feature selection in the classification of disaster-related twitter messages[C]// 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). Penang: IEEE, 2018: 238-243.

[5] 吴先华, 肖　杨, 王国复, 等. 基于微博大数据的城市内涝灾害的灾情及公众情绪研究——以南京市为例[J]. 灾害学, 2018, 33(3): 117-122.
Wu Xianhua, Xiao Yang, Wang Guofu, et al. Research on disaster and public sentiment of urban waterlogging disaster based on microblogging big data: take Nanjing as an example[J]. Journal of Catastrophology, 2018, 33(3): 117-122.

[6] 徐敬海, 褚俊秀, 聂高众, 等. 基于位置微博的地震灾情提取[J]. 自然灾害学报, 2015, 24(5): 12-18.
Xu Jinghai, Chu Junxiu, Nie Gaozhong, et al. Earthquake disaster information extraction based on location microblog[J]. Journal of Natural Disasters, 2015, 24(5): 12-18.

[7] 王　琳, 姜立新, 杨天青, 等. 地震应急信息自动分类方法研究[J]. 震灾防御技术, 2019, 14(4): 907-916.
Wang Lin, Jiang Lixin, Yang Tianqing, et al. Research on the method of automatic classification in earthquake emergency information[J]. Technology for Earthquake Disaster Prevention, 2019, 14(4): 907-916.

[8] Salton G, Yu C T. On the construction of effective vocabularies for information retrieval[J]. SIGPLAN Notices, 1975, 10(1): 48-60.

[9] Karen S J. A statistical interpretation of term specificity and its application to retrieval[J]. Journal of Documentation, 1972, 28(1): 11-20.

[10] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. Computer Science, 2013, 46(4): 23-30.

[11] 高李政, 周　刚, 罗军勇, 等. 基于 Bert 模型的框架类型检测方法[J]. 信息工程大学学报, 2020, 21(2): 214-220.
Gao Lizheng, Zhou Gang, Luo Junyong, et al. Bert-based frame identification method[J]. Journal of Information Engineering University, 2020, 21(2): 214-220.

[12] Lan Z, Chen M, Goodman S, et al. ALBERT: a lite BERT for self-supervised learning of language representations[J]. Computer Science, 2020: arXiv: 1909. 11942v5.

[13] 李思男, 李　宁, 李战怀. 多标签数据挖掘技术: 研究综述[J]. 计算机科学, 2013, 40(4): 14-21.
Li Sinan, Li Ning, Li Zhanhuai. Multi-label data mining: a survey[J]. Computer Science, 2013, 40(4): 14-21.

[14] 陈　虎, 高波涌, 陈莲娜, 等. 结合注意力机制与双向切片 GRU 的情感分类模型[J]. 小型微型计算机系统, 2020, 41(9): 1793-1799.
Chen Hu, Gao Boyong, Chen Lianna, et al. Sentiment classification model combining attention mechanism and bidirectional slice GRU[J]. Journal of Chinese Computer Systems, 2020, 41(9): 1793-1799.

[15] Lin L, Luo H, Huang R, et al. Recurrent models of visual coattention for person re-identification[J]. IEEE Access, 2019, 7(1): 8865-8875.