



武汉大学学报(信息科学版)
Geomatics and Information Science of Wuhan University
ISSN 1671-8860, CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

- 题目： 基于 BERT 迁移学习模型的地震灾害社交媒体信息分类研究
作者： 林森，刘蓓蓓，李建文，刘旭，秦昆，郭桂祯
DOI： 10.13203/j.whugis20220167
收稿日期： 2022-03-28
网络首发日期： 2022-09-05
引用格式： 林森，刘蓓蓓，李建文，刘旭，秦昆，郭桂祯. 基于 BERT 迁移学习模型的地震灾害社交媒体信息分类研究[J/OL]. 武汉大学学报(信息科学版).
<https://doi.org/10.13203/j.whugis20220167>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20220167

引用格式：

林森, 刘蓓蓓, 李建文, 等. 基于BERT迁移学习模型的地震灾害社交媒体信息分类研究[J]. 武汉大学学报·信息科学版, 2022, DOI: 10.13203/j.whugis20220167 (LIN Sen, LIU Beibei, LI Jianwen, et al. Social media information classification of earthquake disasters based on BERT transfer learning model[J]. *Geomatics and Information Science of Wuhan University*, 2022, DOI: 10.13203/j.whugis20220167)

基于 BERT 迁移学习模型的地震灾害社交媒体信息分类研究

林森¹ 刘蓓蓓¹ 李建文¹ 刘旭¹ 秦昆² 郭桂祯¹

1.应急管理部国家减灾中心, 北京 100124;

2.武汉大学遥感信息工程学院, 武汉 430079

摘要: 社交媒体数据具有现势性高、传播快、信息丰富、成本低、数据量大等优点, 已经成为分析突发灾害事件的重要信息源。但是, 社交媒体数据也存在质量各异、冗余而又不完整、覆盖不均匀、缺少统一规范、隐私与安全难以控制等缺点。为了能够利用社交媒体数据为灾害应急响应提供精准化依据, 迫切需要提出能够甄别社交媒体内容并进行有效分类的先进技术。针对此问题, 本文提出利用基于变换器的双向编码表征模型 (Bidirectional Encoder Representation From Transformers, BERT)的迁移学习方法, 面向灾后应急需求, 利用基于海量语料库的预训练模型, 对地震灾害事件后“黄金”72 小时内的微博数据进行多标签文本分类, 将其划分为致灾信息、损失信息、救援救助信息、舆情信息、无用信息 5 种类型, 提取出可用于灾情分析的精细化信息。本文模型在训练集和测试集上的分类准确率分别达 97% 和 92%, 有效提升了微博文本数据的分类精度。评估结果表明: 该模型能很好地提取社交媒体中地震灾害信息, 可应用于地震灾害事件的快速灾情研判, 弥补传统信息获取手段的滞后性。

关键词: BERT 模型; 迁移学习; 社交媒体; 地震灾害; 灾害应急响应; 多标签分类

收稿日期: 2022-03-28

第一作者: 林森, 博士, 助理研究员, 主要研究方向为灾害监测评估预警、灾害大数据挖掘。
linsen_ndrcc@126.com

Foundation support: 国家重点研发计划 (2018YFC1508806)。

1 引言

自然灾害每年对我国的国民经济和社会发展造成了重要影响。地震、暴雨洪

涝、台风、龙卷风等灾害一旦发生，往往具有广泛性、突发性、破坏性等特点，如汶川地震、河南郑州“7·20”特大暴雨、超强台风“利奇马”等，都对灾害影响范围内的人民生产生活产生了很大冲击。灾害事件发生后，信息收集是应急响应决策的第一步，有效的灾害应急响应很大程度上依赖于精准、实时、全面的灾害信息掌握^[1]。以重特大地震灾害为例，在世界各地历次大地震中，72 小时内的救援是最有效的，在此时间段内，救援队伍如果能获得较准确的救援范围信息，可能会救出更多的生命。然而，此时段也正是传统信息获取手段的盲点时段，基于现场的受灾范围评估、地方灾情上报、专业部门的现场调查等，往往在灾后 24 小时以后才能到达现场并逐步开展工作，而卫星遥感和航空遥感数据通常需要协调、资源调配和影像解译等过程。这些传统手段通常在短时间内难以发挥作用，很难完全满足突发灾害事件中高时效、高精度的应急决策需求^[2, 3]。

在移动互联网时代，社交媒体已成为突发灾害信息获取的重要来源^[4-6]，社交媒体数据具有时效性高、数据量大、信息多样的特点^[7]。特别是灾害发生初期，社交媒体中的灾害信息几乎与事件发生是同步的^[8]，而且信息粒度可直达社区层级。近年来，社交媒体已在多次突发灾害中发挥重要作用。例如，2010 年海地 7.0 级地震对当地造成了重创，研究人员利用软件快速收集社交媒体信息和短信，利用这些信息为迅速救助响应提供了积极支持^[9]。2021 年 7 月，河南多地发生特大暴雨，引发严重的城市内涝灾害，微博平台上迅速发起了志愿者、求救、物资需求等信息，通过此类信息实现了物资和救援力量的合理调度。社交媒体平台作为传统灾害信息获取手段的有效补充，弥补了灾害管理部门监测手段、感知体系的不足^[10]。

然而，突发灾害后，社交媒体相关内容关注度骤升，短时间内汇聚了大量重复和冗余信息^[11]，也存在错误甚至恶意的舆论信息^[12]。为此，社交媒体数据用于灾后应急工作之前必须进行内容分类，应面向应急需求进行内容甄别^[13]。当前研究主要集中在两方面：一类是判断信息是否与灾害有关；另一类是通过监督学习精细化地提取灾害相关内容^[14]。其中，后一类研究更适合现在越来越精细化的应急需求。例如，Qu 等在 2010 玉树地震的研究中设置了意见相关、实况更新、救济相关汇总、情绪相关、行动相关、偏离主题、其他等七个内容分类^[15]；Imran 等根据社交媒体的信息价值，将灾害相关信息分为仅个人关注、公众关心和无关 3 类，其中公众关心的信息中又分为注意和建议、伤亡和损失、捐赠、人

口丢失或发现、消息源头及其他等 6 类^[16]；Takahashi 等将台风灾害的推特信息分为 9 类，包括个人报道、二手报道、请求帮助、协调救济行动、提供精神抚慰、批评政府、祝愿和纪念表达、讨论原因和与社区成员重新取得联系^[17]；王艳东等在挖掘城市突发暴雨事件中，将微博数据标注为救援、伤亡损失、警示与天气、交通状况、灾害原因讨论等 5 类信息^[18]。从文本分类的输出来看，更多的研究主要聚焦在社交媒体信息的二分类或多分类上，即每条数据仅含一类信息。然而，实际中社交媒体的发布源很多，许多社交媒体信息可能不仅仅只包含某一类信息，单一数据可以包含多类信息，这种情况更符合真实的社交媒体信息。因此，多标签分类相对于二分类或多分类更适合用于灾害社交媒体信息的分类研究。

传统的文本分类方法多是基于数据挖掘技术，如线性判别分析、支持向量机、随机森林等，这些模型更多关注概率或信息量^[19-21]，依赖语料库的大小^[22]，但是文本的语义理解较弱，因此造成分类精度不稳定。随着机器学习快速发展，面向自然语言处理的深度学习技术逐渐用于社交媒体中突发灾害事件的信息提取。Nguyen 利用深度神经网络识别社交媒体信息可用于危机响应的信息，并进行主题分类^[23]；刘淑涵等利用卷积神经网络提取新浪微博中暴雨主题信息，并对信息进行了时空分析^[24]；杨腾飞等利用卷积神经网络提取地震灾害后社交媒体中的情感信息，并进行情感分类^[25]。虽然深度学习模型的精度比传统方法有了较大提升，但是前期构建词向量空间、参数训练和优化要花费大量时间，而且 word2vec 等词向量模型对于一词多义的问题尚无法很好解决^[26]。

基于大规模文本数据的预训练模型是自然语言处理的最新主流技术。2018 年，GPT、ELMo、BERT 等模型诞生^[27-29]，这些模型刷新了自然语言处理领域任务的表现。这些方法的共同点是依靠海量语料训练大量参数，使模型能从文本中捕捉到一词多义、词法、句法结构、现实知识等信息，对于普通应用者，只需通过微调模型就可以直接用于下游任务需求，免去了从零开始的过程。BERT 是谷歌团队于 2018 年提出的，由多层基于注意力机制(Attention)的转换器(Transformer)模型构成^[27]，非常适用于文本信息的特征提取。BERT 模型结构有利于迁移学习，对于下游文本分类^[30-33]、命名实体识别^[34, 35]、机器问答^[36, 37]等任务都有很好的性能。目前，利用 BERT 模型开展突发灾害信息分类的研究相对较少，迫切需要开展此类研究。

本文以地震灾害为例，收集震后 72 小时微博中发布的相关数据作为训练样

本，利用 BERT 预训练模型建立地震灾害社交媒体信息多标签分类模型，聚焦于快速提取微博中精细化的地震应急响应支撑信息，为灾后快速辅助响应决策提供科学依据。

2 研究方法

2.1 数据获取及预处理

本文选取新浪微博平台作为社交媒体代表，选择 2013—2022 年之间影响中国的 5 场重特大地震灾害事件为案例，基于 Python 相关网络爬虫技术，获取微博上以“地震”为关键词的相关数据，时间长度选取地震发生后“黄金”72 小时，抓取的文本信息包括发布者昵称、微博内容、发布时间、发布地址四个字段属性信息。

实验数据收集后，对数据进行预处理。本文仅限于微博短文本信息，不包含长文本、图片、视频等内容的研究，与这些相关的提示数据在分类任务中所起作用不大，另外还存在已丢失信息、表情信息、符号信息等，这些内容都可以根据“#”“@”“/”等标识符删除相应内容^[38]。微博发布地址通过地址匹配方法，查找出发布地址的经纬度信息。

2.2 多标签分类及数据标注

本文采用多标签分类，即每条微博数据可以含多个标签。根据文献研究和实际工作经验，面向地震灾害事件中的应急响应需求，设置致灾信息、损失信息、救援救助信息、舆情信息、无用信息 5 类标签，每一类标签的标注规则和部分关键词描述如表 1 所述。其中，致灾信息主要是地震、余震及次生灾害发生地点和强度的表达，可用于第一时间了解灾害大小和影响范围的评估，是灾害预评估的关键信息；损失信息是灾害已造成灾情的描述，包含具体承灾体（人、房屋、道路、电力、通信等）的受损数量或文字信息，可用于判断启动应急响应的级别和部署救援力量；救援救助信息主要是网络求助和社会行动，可精确辅助救援行动和管理救灾资源调配；舆情信息主要是民众在网络上的情绪表达，包括祈福、祝愿、恐惧、同情、无助、愤怒、谩骂、诋毁、探讨，有助于了解网络动态，及时关注舆情事件，积极引导事件讨论；无用信息是指与地震灾害无关的表达。

数据标注通过人工标注进行。每个标签用 1 和 0 表示，1 表示样本含该标签信息，0 表示样本不含该标签信息。所有标签中无用信息与其他 4 类信息互斥，

不会同时出现；其他 4 类标签可以同时出现。

表 1 地震灾害案例标注 5 类标签依据

Tab.1 Labeling criteria for the 5 labels in earthquake cases

标签类型	判别标准	关键词
致灾信息	震级信息；位置及震感描述；次生灾害信息	地震、级（震级）、余震、市（县、乡、村）、堰塞湖、地质灾害、水库、裂缝、断裂
损失信息	人员伤亡；房屋倒损；交通、电力、通信受损情况	死亡、重伤、房屋、道路、电力、通信、中断
救援救助信息	求救信息；政府、企业、社会组织、个人救援行动	被困、求救、食品、帐篷、物资、消防、部队、救援、捐款、通道
舆情信息	情绪表达；可能的谣言	祈福、希望、害怕、哀悼、难过、愤怒、造假、谣传
无用信息	蹭流量；与地震无关话题	——

2.3 BERT 模型

本文采用 BERT 中文预训练模型，该模型是基于海量的中文维基百科（包括简体和繁体）进行训练，具有很强的泛化能力。模型结构含有 12 层变换器（Transformer），隐藏层输出维度为 768，采用 12 个多头注意力机制(mutil-head attention)。BERT 是一个编码器（Encoder），非常适合文本任务的迁移学习，不需要再用大量的语料来进行训练，可以节约时间，效率高效。

BERT 的预训练输入是三个向量之和（图 1），包括词语向量、片段向量和位置向量，其中标记向量是将每一个词转换成词向量，片段向量表明这个词属于哪个句子，而位置向量是学习出来的位置信息。每个句子以[CLS] 标志开头，最后一层该位置对应向量可以作为整句话的语义表示，从而用于下游的分类任务等。[SEP] 标志用于分开两个输入句子。

BERT 预训练模型的输出结果是一个或多个向量。要利用好 BERT 的输出向量，下游任务有两种方式进行自己的训练：一种方式是通过微调，可以稍微改变预训练模型参数；另一种方式是特征抽取，不改变预训练模型参数，只是把预训练模型的输出作为特征输入到下游任务。

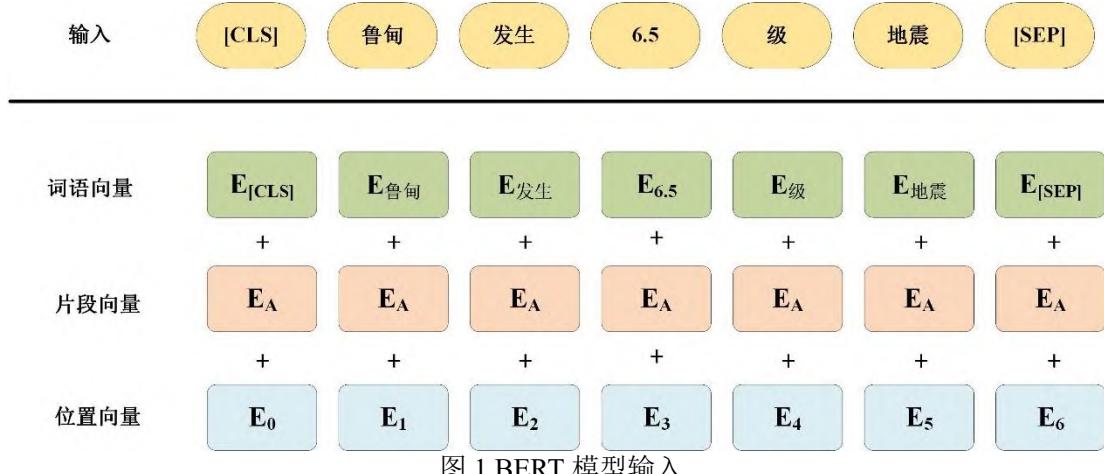


图 1 BERT 模型输入

Fig.1 BERT model inputs

2.4 多标签分类精度评估方法

多标签分类结果评估可以分为基于样本的评估和基于标签的评估^[39]。基于样本的评估先衡量单个样本的分类效果，再衡量整个数据集上效果；基于标签的评估先衡量单个类别的分类效果，再衡量所有类别上的均值。

(1) 基于样本的评估方法

基于样本的评估又分为基于分类的评估和基于排序的评估。基于分类的评估主要采用分类正确率指标(Accuracy)，即 5 个预测样本标签与真实样本标签完全一致，则该样本分类正确；如果有一个标签预测错误，则该样本分类错误；计算全部样本中分类正确的比例即为分类正确率；基于排序的评估主要用于评估标签的预测得分排序是否符合真实的标签排序，常用的评估方法有覆盖误差(Coverage error)、标签排序平均正确率(Label ranking average precision, LRAP)和排序损失(Ranking loss)：

$$coverage\ error = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \max_{j:y_{ij}=1} rank_{ij} \quad (1)$$

$$LRAP = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \frac{1}{\|y_i\|_0} \sum_{j:y_{ij}=1} \frac{|L_{ij}|}{rank_{ij}} \quad (2)$$

Ranking loss

$$= \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \frac{1}{\|y_i\|_0(n_{labels} - \|y_i\|_0)} | \{(k, l) : \hat{f}_{ik} \leq \hat{f}_{il}, y_{ik} = 1, y_{il} = 0\} | \quad (3)$$

$$\leq \hat{f}_{il}, y_{ik} = 1, y_{il} = 0 \}$$

式中：ij, ik, il 分别表示第 i 个样本的第 j, k, l 个标签， $n_{samples}$ 表示样本数量，

n_{labels} 表示标签类型数量， y 表示标签类型， \hat{f} 表示标签预测得分， $rank$ 表示标签排序， $rank_{ij} = |\{k: \hat{f}_{ik} \geq \hat{f}_{ij}\}|$ ， L 表示标签得分高于特定标签的统计， $L_{ij} = |\{k: y_{ik} = 1, \hat{f}_{ik} \geq \hat{f}_{ij}\}|$ ， $|\cdot|$ 表示集合中元素个数， $\|\cdot\|_0$ 表示 L_0 范式，即向量中非 0 元素个数。覆盖误差是计算要预测对所有真实标签需要包含的平均标签个数，这个指标越接近真实平均标签个数，说明预测越准确；标签排序平均正确率与分类正确率类似，是基于预测标签的得分排序计算平均准确率，取值在 0-1 之间，分值越高，标签预测效果越好；排名损失是计算错误标签对的平均数量，并通过正确标签和错误标签的倒数加权，取值越接近 0 越好。

(2) 基于标签的评估方法

基于标签的评估选用 AUC (Area Under Curve) 指标，该指标对标签不平衡的数据集非常适合。AUC 为受试者工作特征(Receiver Operating Characteristic, ROC)曲线下方的面积，而 ROC 曲线是以真阳性率(True positive rate, TPR)为纵坐标，假阳性率(False positive rate, FPR)为横坐标绘制的曲线，TPR 表示预测正类中实际正类比例，FPR 表示预测正类中实际负类比例，故 ROC 曲线越靠拢(0,1)点，越偏离 45 度对角线越好。相对应的 AUC 得分在 0-1 之间，分值越高，说明分类器越好。

3 数据及结果

本文的基本实验环境为：模型搭建和结果运算采用 TensorFlow2.3.0 框架，BERT 中文预训练模型基于 Hugging Face 公司的开源版本 (<https://huggingface.co/bert-base-chinese>)，精度评估主要采用 scikit-learn 的评估指标库，硬件方面利用 Tesla V100-PCIE-32GB GPU 做运算加速。

3.1 案例数据集

本文选取发生在 2013—2022 年间的芦山地震、鲁甸地震、尼泊尔地震、九寨沟地震、门源地震等 5 次地震作为研究案例。这 5 次地震强度较高，造成了较大损失，在社交媒体上有较大范围和较长时间的关注，容易获取到较多内容作为训练样本。通过数据清洗，收集到实验数据共计 54501 条（表 2），带地址数据有 2202 条，其中门源地震因为发生时间较近，收集到的带地址数据较多。每个案例中 90% 数据加入到训练集，10% 数据加入到测试集。

表 2 关于 5 个案例地震相关信息

Tab.2 Details about 5 earthquakes used in this study

地震名称	发生时间	数据收集时间段	数据量	带地址数据
2013 年芦山 7.0 级地震	4 月 20 日 8 时 02 分	4 月 20 日 8 时至 4 月 23 日 8 时	11127 条	167 条
2014 年鲁甸 6.5 级地震	8 月 3 日 16 时 30 分	8 月 3 日 16 时至 8 月 6 日 16 时	11072 条	218 条
2015 年尼泊尔 8.1 级地震 (我国西藏地区受影响)	4 月 25 日 14 时 11 分	4 月 25 日 14 时至 4 月 28 日 14 时	10982 条	183 条
2017 年九寨沟 7.0 级地震	8 月 8 日 21 时 19 分	8 月 8 日 21 时至 8 月 11 日 21 时	10646 条	314 条
2022 年门源 6.9 级地震	1 月 8 日 1 时 55 分	1 月 8 日 2 时至 1 月 11 日 2 时	10674 条	1320 条

3.2 标注样例和统计

数据标注小组根据表 2 的标注准则, 通过多轮标注和审核, 完成 54501 条数据样本的标注, 每个训练样本由 ID、微博内容和标签列表组成。如表 3 所示, ID 为“四川省地震局”的样本数据的标签列表为[1,1,1,0,0], 则表示微博内容中含有致灾信息、损失信息、救援救助信息。

表 3 训练数据样例

Tab.3 Examples of training data

ID	微博内容	致灾信息	损失信息	救援救助信息	舆情信息	无用信息
Hey-喵	我刚刚为“鲁甸县龙井小学 40 余人被困急需帐篷食物二【地震赈灾】”献出了自己的爱心, 每份微博的力量都可点亮被救助者的希望之光, 共同伸出援助之手, 让微博充满爱。[个人求助]鲁甸县龙井小学 40 余人被困急需帐篷食物二【地震赈灾】#四川雅安地震#截至 4 月 21 日 15 时, 四川雅安地震共累计造成 150 余万人受灾, 目前灾难造成的经济损失仍在不断统计中。据四川省雅安市人民政府新闻办公室腾讯官方微博公布, 全市近 40 万间房屋受损, 倒塌 12851 间, 34 座水库受损, 堤渠损毁 2668 处, 农林水经济损失 14.30 亿元, 全市需安置 48915 户。(新浪财经)	0	0	1	1	0
中国四达官方微博	#四川九寨沟 7.0 级地震# 我们距九寨沟县城还有 28 公里, 截至目前, 一路上路况良好, 就是网络信号极差, 还好还能通话@中国国际救援队 @中	1	1	0	0	0
四川省地震局		1	1	1	0	0

ID	微博内容	致灾信息	损失信息	救援救助信息	舆情信息	无用信息
国地震台网速报						
华为中国	#四川九寨沟 7 级地震#发生后，当地多处光缆损毁、基站中断，通信不同程度受阻。由于受灾地区滑坡不断，传统的应急通信手段难以奏效，华为配合四川移动紧急协调一套无人机高空基站赶赴灾区。	1	1	1	0	0
Ameliya_CY	#九寨沟地震# 帅爷说幸好我洗了个头，发型没有乱	0	0	0	0	1

所有训练样本中，不同标签数量并不平衡，如图 2 所示，5 类标签中致灾信息和舆情信息较多，是其他 3 类标签数量的 2-3 倍。其中，致灾信息较多是因为许多信息会引用地震震级话题，表示微博信息与当前地震的相关性；舆情信息较多是因为表达情绪是突发灾害发生后最常见的，如恐惧、担忧、紧张、祝福等。

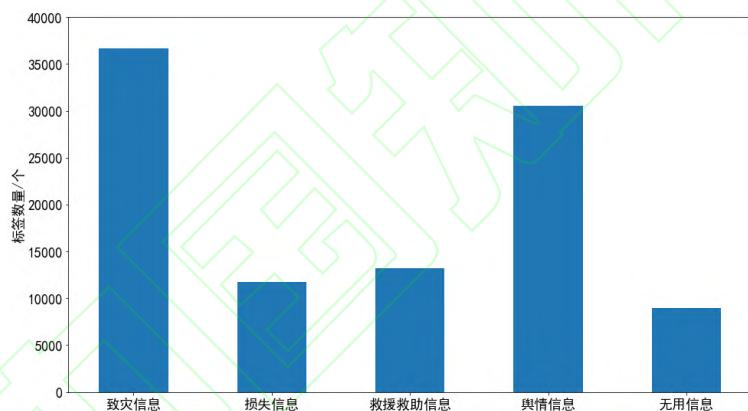


图 2 样本数据中 5 类标签的数量

Fig.2 The numbers of 5 labels in the sample data

在全部 54501 个样本数据中，含 2 个标签以上多标签样本数量为 32279 个，占全部样本数的 59.2%。这些多标签样本中，共有 11 种多标签组合方式，不同标签组合的联合概率分布反映了不同内容之间的相关性。如表 4 所示，不同标签组合的联合概率分布非常不均衡，（致灾信息，舆情信息）的组合概率非常高，说明信息发布者通常在关心地震致灾情况的同时，也会同时表达自己内心的意见或情绪；可以看出，凡是不含致灾信息的组合，出现概率相对偏低，而一旦加上致灾信息，组合出现的概率会提高 100 倍以上，这表明致灾信息在发布内容中有着非常频繁的出现，这与样本中致灾信息标签的数量较大有关系。致灾信息标签数量过大，会影响一些专题内容的有效提取，因为可能有大量的转发或二手消息，

属于信息冗余，未来的应用中可以对此标签消息进行再分解和提取。

表 4 不同标签组合的联合概率分布

Tab.4 The joint probability distribution of different label combinations

标签组合	标签数量	联合概率
(1,1,0,0,0)	2515	0.046
(1,0,1,0,0)	3516	0.065
(1,0,0,1,0)	14484	0.266
(0,1,1,0,0)	25	0
(0,1,0,1,0)	82	0.002
(0,0,1,1,0)	246	0.005
(1,1,1,0,0)	3746	0.069
(1,0,1,1,0)	2448	0.045
(1,1,0,1,0)	2291	0.042
(0,1,1,1,0)	10	0
(1,1,1,1,0)	2916	0.054

3.3 模型参数

本文以 BERT 特征提取的最后一层输出作为全连接层分类器的特征输入，同时利用 dropout 层防止过拟合。由于多标签分类每个标签其实属于二分类，因此模型激活函数采用更适合二分类的 sigmoid 函数，最终标签预测概率大于 0.5，标签为 1，否则为 0（图 3）。模型训练用微调方式，采用 Adam 优化器和二元交叉熵损失函数，梯度裁剪设为 1（防止梯度爆炸），输入最大长度设为 420（每条微博限制 140 字，UTF-8 编码中一个中文或标点占 3 个字节）。

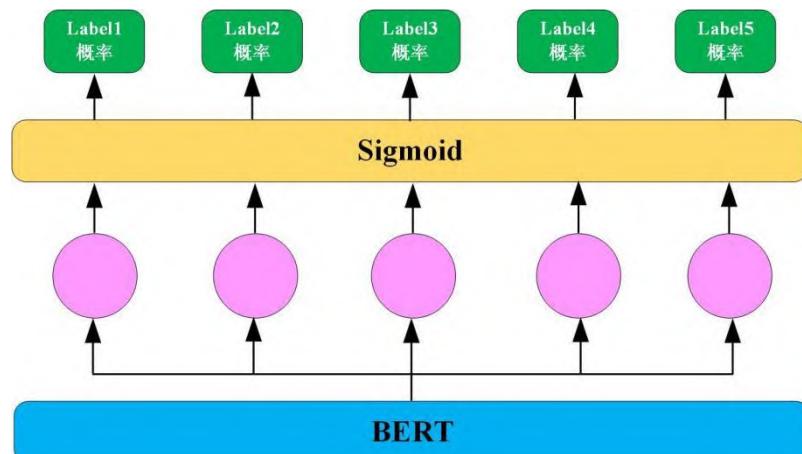
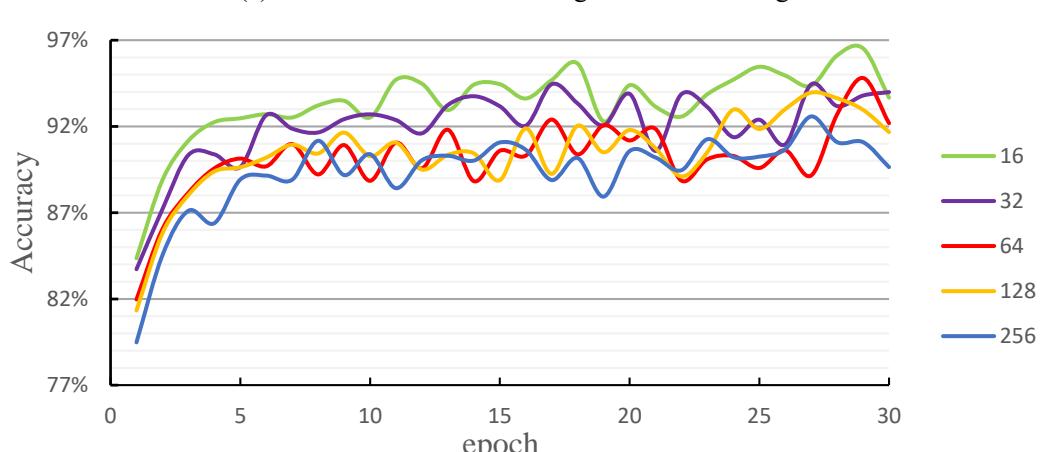
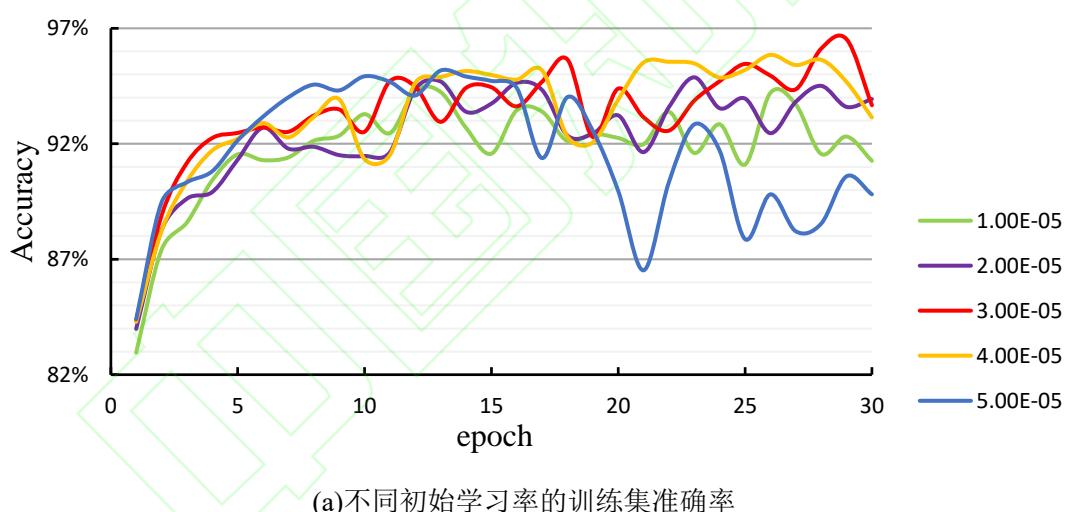


图 3 基于 BERT 预训练模型的多标签分类模型网络结构

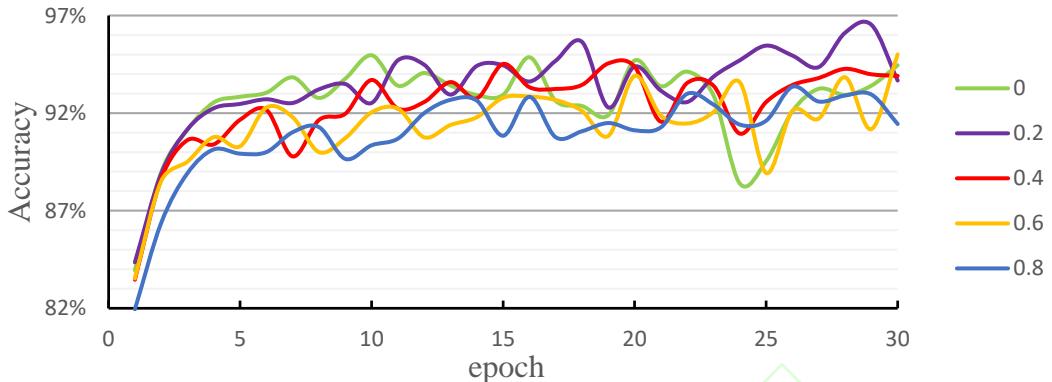
Fig.3 Multi-label classification model network structure based on BERT pre-trained model

在微调过程中，为找到最优参数组合，分别对初始学习率、批量大小和 dropout 比例等参数进行了对比实验，每个模型参数训练 30 轮(epoch)，取训练集上正确率(Accuracy)最高的参数为最优值。其中，学习率取 0.00001 到 0.00005 的 5 个数值（固定批量大小为 16，dropout 比例为 0.2），当学习率大于 0.00005 时，模型准确率快速下降，由图 4(a)可以看出，学习率为 0.00003 时性能最优；批量大小取 16、32、64、128、256 等 5 个数值（固定学习率为 0.00003，dropout 比例为 0.2），由于内存限制，批量大小为 128 和 256 时，输入最大长度分别调低为 360 和 270，由图 4(b)可以看出，批量大小越小，准确率越高，取批量大小为 16；dropout 比例取 0 到 0.8 等间隔的 5 个数值（固定学习率为 0.00003，批量大小为 16），由图 4(c)可以看出，dropout 比例为 0.2 时模型准确率最高。根据实验对比，模型的学习率参数取 0.00003，批量大小参数取 16，dropout 比例参数取 0.2。



(b)不同批量大小的训练集准确率

(b) Accuracies of train set using different batch sizes



(c) 不同 dropout 比例的训练集准确率

(c) Accuracies of train set using different dropout ratios

图 4 不同模型参数的性能对比

Fig.4 Performance comparisons of different model parameters

3.4 分类精度评估

对于训练集，评估指标仅选用基于样本的分类准确率和基于标签的 AUC 两项。模型经过 30 轮训练，模型在训练集上损失(Loss)为 0.026，分类准确率在第 29 轮取得最高值 96.5%，AUC 经过不到 10 轮训练即可达到最高得分 0.999(图 5)。

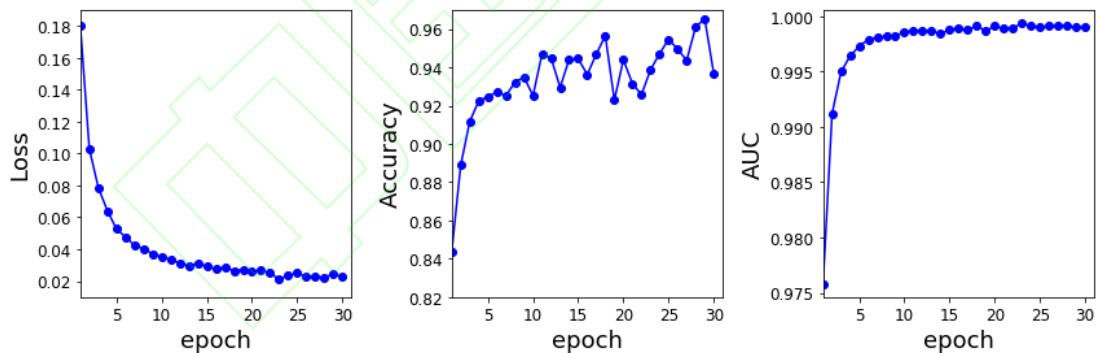


图 5 模型在训练集上的训练结果

Fig.5 The results of the mode on training dataset after 30 epochs

对于测试集，评估指标选用分类准确率、覆盖误差、标签排序平均正确率、排序损失、AUC 等。最终，模型在测试集上损失为 0.166，分类准确率为 91.6%，覆盖误差为 1.95 (测试集每个样本平均真实标签数 1.78 个)，标签排序平均正确率为 94.4%，排序损失为 0.036，AUC 得分为 0.988。另外，计算了测试集中每个标签的 AUC 得分 (表 5)，损失信息得分最低为 0.952，救援救助信息得分最高达到 0.998。以上指标表明，无论从样本角度，还是从标签角度，实验模型都

取得很好的效果，该模型非常可靠，在快速分类微博上突发灾害信息方面具有很高的可行性。

表 5 测试集中各个标签 AUC 得分

Tab.5 AUC scores for each single label in the test dataset

标签类型	致灾信息	损失信息	救援救助信息	舆情信息	无用信息
AUC 得分	0.991	0.952	0.998	0.967	0.983

3.5 标签的时空应用分析

突发灾害事件在社交媒体上具有生命周期的特点^[11, 40]。不同标签的时序变化，代表了灾后不同信息的时间发展规律，每个标签的差异反映了灾害应急响应中不同时间的关注点有所不同^[15]。本文对 5 个地震案例建立标签和时间关系，分析每个标签在震后 72 小时内的数量变化情况和每个标签分布在不同时间段的概率分布。如图 6 的热力图所示，地震发生后的 30 小时内是突发事件讨论主题最多的时间段，其中致灾信息和舆情信息内容最多，热度也持续较长，0-18 小时内信息呈爆发式出现，在其它时间段分布比较平均；损失信息和救援救助信息数量相对较少，周期性比较明显，损失信息一般在 0-18 小时内信息量较多，救援救助信息一般在 0-36 小时内偏多，48 小时之后两者都明显减少；无用信息数量最少，0-48 小时内占比较小，但 48 小时之后，占比明显增多。

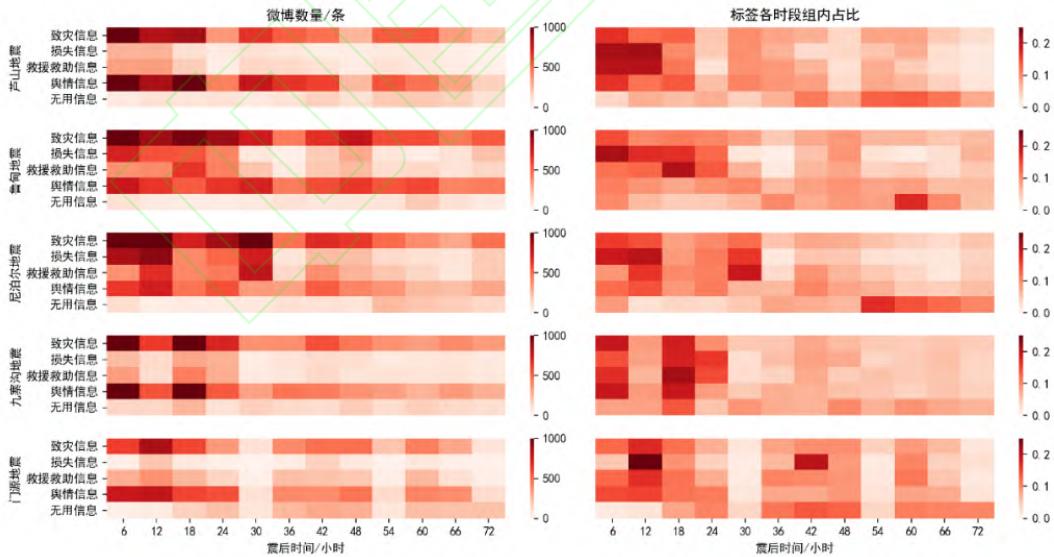
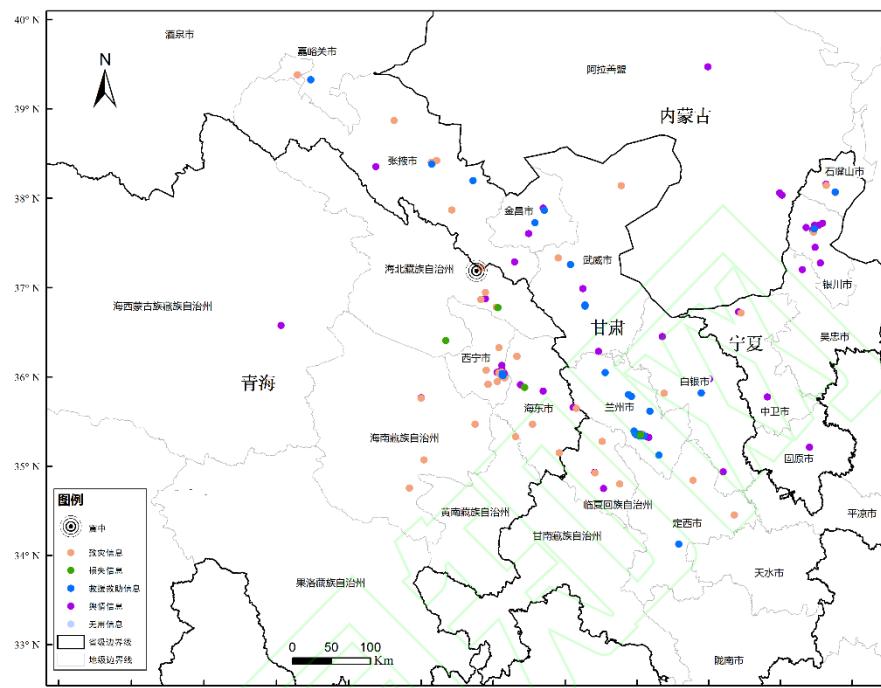


图 6 每个标签在不同时段的数量和概率分布热力图

Fig.6 Heat maps of the number and the probability distribution of each label over different time periods

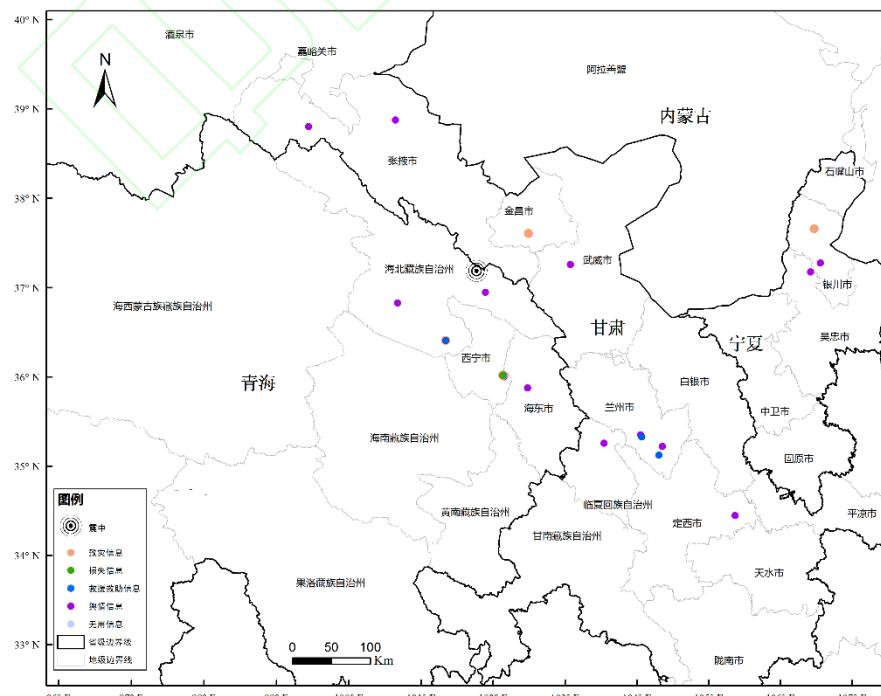
从可获得的带地址数据的时间和空间分布来看，带地址数据具有明显的时间周期。以门源地震为例，分别提取地震发生后 0-24 小时、24-48 小时、48-72 小

时内震中周边 500 公里的带地址数据(图 7), 可以发现, 0-24 小时内数据最多, 占 90%以上, 这个时间段是灾后信息获取的关键时间, 带地址的信息多为致灾信息、救援救助信息, 可用于灾后第一时间的灾情评估、应急决策。24 小时之后, 带地址的数据明显减少, 其中舆情信息的标签数量开始占据主导地位。



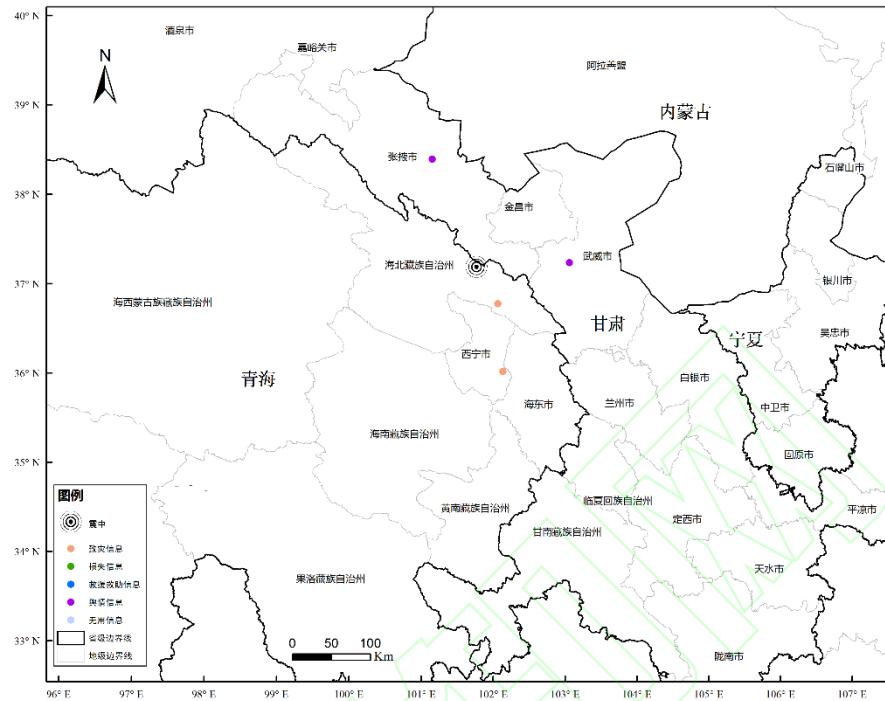
(a)0-24 小时不同标签分布

(a)Distribution of different labels during 0-24h



(b)24-48 小时不同标签分布

(b)Distribution of different labels during 24-48h



(c)48-72 小时不同标签分布

(c)Distribution of different labels during 48-72h

图 7 门源地震中不同时段内不同标签的空间分布

Fig.7 Spatial distribution of different labels in different time periods in the Menyuan earthquake

从图 6 和图 7 可知, 地震灾害发生后的短时间内应该重点关注某些主题内容, 这段时间的信息提取恰恰能够弥补传统手段的盲点。在灾害发生 36 小时之内, 社交媒体的焦点主要集中于人员伤亡、失踪人员搜寻、财产损失、救助需求等内容, 这对于应急响应中开展定向救援和物资调配有指示作用; 0-18 小时的致灾信息非常重要, 有助于专家对灾害影响范围和灾害造成的损失进行评估, 同时致灾信息中包含余震信息、次生灾害信息, 这类信息需持续关注, 防止次生灾害造成更大损失; 舆情信息具有“长尾”特征, 对于舆情管理需要持续跟踪, 甚至可能超出 72 小时的范围, 有统计显示, 较大地震的舆情可能持续 22 天^[40]; 无用数据的变化表明, 48 小时之后, 地震情况基本明朗, 平台进入相对沉寂期, 此时许多账号开始利用地震词条作为增加流量的手段, 这类数据要排除在分析之外。

4 总结

社交媒体信息近年来在突发灾害事件中逐步得到了应用，重特大灾害发生初期，灾区往往成为“信息孤岛”，有效利用好社交媒体信息能够辅助灾情的快速研判，及时做出灾害应急响应。目前，基于社交媒体的灾害信息分类主要是采用二分类或多分类方法，实际中每条社交信息可能包含丰富的内容，主题往往含有多个类别信息，因此，多标签分类更适用于灾害信息的主题分类。本文基于 BERT 预训练模型建立了突发灾害社交媒体应急信息的多标签分类模型，用于灾害发生后按应急需求进行信息分类。实验以地震灾害为例，从微博上获取 5 万多条相关样本数据并进行了人工标注，利用 BERT 预训练模型迁移学习，在训练集上通过微调训练建立模型，并在测试集上进行了模型验证。通过实验分析得到如下结论：

(1) 基于 BERT 预训练模型的迁移学习时效性高。与深度学习相比，BERT 预训练阶段已基于海量语料库提取了大量特征，基于 BERT 的迁移学习对于下游任务搭建非常便利，省去了特征提取、搭建网络、控制过拟合和参数调优等步骤，能够快速分类突发灾害事件的社交媒体信息。

(2) 基于 BERT 预训练模型建立的多标签分类模型对于社交媒体信息分类准确性好，信息提取完整。模型在训练集和测试集分类准确率上均高于 90%，数据集整体和单个标签的 AUC 得分都在 0.95 以上。精度评估结果证明模型对于社交媒体上的突发灾害信息有很好的区分度，完全可用于应急管理业务当中。另外，研究结果表明，近 60% 的微博数据包含 2 类以上信息，相对于单一标签分类，多标签分类更能体现社交媒体信息的价值。

(3) 不同标签在社交媒体中存在周期不同的问题。对于致灾信息、损失信息、救援救助信息等内容要在灾害发生后特定时间段内重点关注，0-18 小时致灾信息有效性高，0-36 小时是人员伤亡、救助需求发布较多的时段，详细数据可用于灾情评估和救援力量部署，而舆情信息有长时间保持热度的趋势，需要持续关注和分析。

(4) 标签分类可进一步精细化。从单标签 AUC 得分来看，损失信息和舆情信息相对于其他 3 个标签的得分稍低。通过预测结果发现，损失信息一般是数字信息，一旦通过文字描述，标签的准确性会有下降；对于舆情信息，网民的情绪信息较多，语言相对自由宽泛，结果中容易与无用信息分类混淆。下一步还可以将损失信息和舆情信息标签再细分成多个标签，会更好地捕捉细粒度突发灾害

应急信息。另外，致灾信息在所有信息中过多，这可能是大量转发导致，未来应用中可以考虑如何分辨原始信息和冗余信息。

参 考 文 献

- [1] Bai Hua, Lin Xunguo. Sina weibo disaster information detection based on Chinese short text classification[J]. Journal of Catastrophology,2016,31(2):19-23 (白华, 林勋国. 基于中文短文本分类的社交媒体灾害事件检测系统研究[J]. 灾害学,2016,31(2):19-23)
- [2] Velev D, Zlateva P. Use of Social Media in Natural Disaster Management[C]. International Economics Development and Research Center(IEDRC), Hong Kong, China, 2012
- [3] Bo Tao. Earthquake Disaster Data Mining and Application of Rapid Intensity Assessment based on Social Media[D]. Institute of Engineering Mechanics,China Earthquake Administration,2018 (薄涛.基于社交媒体的地震灾情数据挖掘与烈度快速评估应用[D]. 中国地震局工程力学研究所,2018)
- [4] Shan Jie,Qin Kun,Huang Changqing,et al.Methods of Crowd Sourcing Geographic Data Processing and Analysis[J].Geomatics and Information Science of Wuhan University,2014,39(4): 390-396 (单杰, 秦昆, 黄长青, 等. 众源地理数据处理与分析方法探讨[J]. 武汉大学学报(信息科学版),2014,39(4):390-396)
- [5] Goodchild M F. Citizens as sensors: the world of volunteered geography[J]. GeoJournal, 2007,69(4): 211–221
- [6] Schade S, Diaz L,Ostermann F O, et al. Citizen-based sensing of crisis events : sensor web enablement for volunteered geographic information[J]. Applied geomatics, 2013,5(1):3-18
- [7] Han Xuehua, Wang Juanle, Bu Kun, et al. Progress on disaster events information acquisition from web text[J]. Journal of Geo-information Science, 2018,20(8):1037-1046 (韩雪华,王卷乐,卜坤,等.基于 Web 文本的灾害事件信息获取进展[J]. 地球信息科学学报,2018,20(8): 1037-1046)
- [8] Cai Meizhu. Study On The Characters Of Internet Consensus To Emergent Natural Disaster [D]. Huazhong University of Science & Technology,2012 (蔡梅竹.突发自然灾害事件网络舆论特征研究[D].华中科技大学,2012)
- [9] Song Jiangong, Wang Zhixin, Li Qinyong, et al. Internet information processing for earthquake emergency response[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017,43(6):1155-1164(宋建功,王之欣,李勤勇,等.面向地震应急响应的互联网信息处理[J]. 北京航空航天大学学报,2017,43(6):1155-1164)
- [10] Goodchild M F, Glennon J A. Crowdsourcing geographic information for disaster response: a research frontier[J]. International Journal of Digital Earth,2010,3(3):231-241.
- [11] Shuai Xianghua, Hu Suping, Liu Qin,et al. Internet media-based acquisition and processing model of earthquake disaster situation[J]. Journal of Natural Disasters,2013,22(3):178-184(帅向华,胡素平,刘钦,等.地震灾情网络媒体获取与处理模型[J].自然灾害学报,2013,22(3): 178-184)
- [12] Heinzelman J, Waters C. Crowdsourcing Crisis Information in Disaster[R]. United States Institute of Peace, 2010
- [13] Chen Kefan, Yu Wei. Research on the ideas of local government emergency management information construction [J]. Communication & Information Technology,2020, (3):69-72 (陈科帆,余伟.地方政府应急管理信息化建设思路研究[J].通信与信息技术,2020,(3):69-72)
- [14] Zhang C, Fan C, Yao W L, et al. Social media for intelligent public information and warning in disasters: An interdisciplinary review [J]. International Journal of Information Management, 2019, 49:190-207

- [15] Qu Y, Huang C, Zhang P, et al. Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake[C]. Proceedings of the ACM 2011 conference on Computer supported cooperative work, Hong Kong, China, 2011
- [16] Imran M, Elbassuoni S, Castillo C, et al. Practical extraction of disaster-relevant information from social media[C]. World Wide Web, Social Web for Disaster Management (SWDM), Rio de Janeiro, Brazil, 2013.
- [17] Takahashi B, Tandoc EC, Carmichael C. Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines[J]. Computers in Human Behavior. 2015,50:392-398
- [18] WANG Yandong,LI Mengmeng,FU Xiaokang,et al.A New Method to Detect the Development Situation of Disasters Based on Social Media Co-word Network[J]. Geomatics and Information Science of Wuhan University, 2020, 45(5): 691-699(王艳东,李萌萌,付小康,等.基于社交媒体共词网络的灾情发展态势探测方法[J]. 武汉大学学报·信息科学版,2020,45(5):691-699)
- [19] Chakrabarti S, Roy S, Soundalgekar MV. Fast and accurate text classification via multiple linear discriminant projections[J]. The VLDB Journal,2003,12(2):170-185
- [20] He Jie. Random forest in application of text classification[D].South China University of Technology,2015 (贺捷.随机森林在文本分类中的应用[D].华南理工大学,2015)
- [21] Ping Yuan.Research on clustering and text categorization based on Support Vector Machine[D]. Beijing University of Posts and Telecommunications,2012 (平原.基于支持向量机的聚类及文本分类研究[D].北京邮电大学,2012)
- [22] Yu Tongrui, Jin Ran, Han Xiaozhen, et al. Review of pre-training models for natural language processing[J]. Computer Engineering and Applications,2020,56(23):12-22(余同瑞,金冉,韩晓臻,等.自然语言处理预训练模型的研究综述[J].计算机工程与应用, 2020,56(23):12-22)
- [23] Nguyen D T, Joty S R, Imran M, et al. Applications of Online Deep Learning for Crisis Response Using Social Media Information[J]. arXiv:1610.01030, 2016
- [24] Liu Shuhan, Wang Yandong, Fu Xiaokang. Extracting rainstorm disaster information from microblogs using convolutional neural network[J]. Journal of Geo-information Science, 2019, 21(7):1009-1017 (刘淑涵, 王艳东, 付小康. 利用卷积神经网络提取微博中的暴雨灾害信息[J]. 地球信息科学学报,2019,21(7):1009-1017)
- [25] Yang Tengfei, Xie Jibo, Yan Dongchuan,et al. Extracting sentiment information from social media based on deep learning and the research on disaster reduction[J]. Geography and Geo-Information Science, 2020, 36(2):62-68 (杨腾飞,解吉波,闫东川,等.基于深度学习的社交媒体情感信息抽取及其在灾情分析中的应用研究 [J].地理与地理信息科学,2020,36(2):62-68)
- [26] Li Zhoujun, Fan Yu, Wu Xian Jie. Survey of Natural Language Processing Pre-training Techniques[J]. Computer Science, 2020,47(3):162-173 (李舟军,范宇,吴贤杰.面向自然语言处理的预训练技术研究综述[J].计算机科学, 2020,47(3):162-173)
- [27] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. arXiv:1810.04805, 2018
- [28] Peters M, Neumann M, Iyyer M, et al. Deep Contextualized Word Representations[J]. arXiv:1802.05365, 2018
- [29] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[EB/OL].[2022-01-05].https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf
- [30] Heidari M, Jones J H. Using BERT to Extract Topic-Independent Sentiment Features for Social Media Bot Detection[C]. 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, USA, 2020
- [31] Mozafari M, Farahbakhsh R, Crespi N. A BERT-based transfer learning approach for hate speech detection in online social media[J]. arXiv:1910.12574, 2019

- [32] Jain P,Ross R,Schoen-Phelan B. Estimating Distributed Representation Performance in Disaster-Related Social Media Classification[C]. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Vancouver, Canada,2019.
- [33] Zahera H M. Fine-tuned BERT Model for Multi-Label Tweets Classification[C]. Text REtrieval Conference (TREC) ,2019
- [34] Liang C, Yu Y, Jiang H, et al. BOND: BERT-Assisted Open-Domain Named Entity Recognition with Distant Supervision[C]. KDD'20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, USA , 2020
- [35] Souza F, Nogueira R, Lotufo R. Portuguese Named Entity Recognition using BERT-CRF[J]. arXiv:1909.10649,2019
- [36] Wang Z, Ng P, Ma X, et al. Multi-passage BERT: A Globally Normalized BERT Model for Open-domain Question Answering[J]. arXiv:1908.08167,2019
- [37] Wei Y, Xie Y, Lin A, et al. End-to-End Open-Domain Question Answering with BERTserini[J]. arXiv:1902.01718,2019
- [38] Chen Zhiqun, Ju Ting. Research on Tendency Analysis of Microblog Comments Based on BERT and BLSTM[J]. Information studies: Theory & Application, 2020, 43(8):173-177(谌志群,鞠婷.基于 BERT 和双向 LSTM 的微博评论倾向性分析研究[J]. 情报理论与实践, 2020,43(8):173-177)
- [39] Tsoumakas G, Katakis I, Vlahavas I. Mining Multi-label Data[M]. Data Mining and Knowledge Discovery Handbook, 2009
- [40] Li G,Hai L,Chen J H.A survival analysis of periodic characteristics of China's emergent natural disaster event reported by network media:a case study on earthquake and typhoon disaster[J].Journal of Information Resources Management, 2015,5(3):18-24 (李纲,海岚,陈璟浩.突发自然灾害事件网络媒体报道的周期特征分析——以地震和台风灾害为例[J].信息资源管理学报,2015,5(3):18-24)

Social media information classification of earthquake disasters based

on BERT transfer learning model

LIN Sen¹ LIU Beibei¹ LI Jianwen¹ LIU Xu¹ QIN Kun² GUO Guizhen¹

1. National Disaster Reduction Center of the Emergency Management Department,
Beijing 100124, China

2. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan
430079, China

Abstract: **Objectives:** In recent years, extreme weather events have increased and sudden disasters have occurred frequently, which puts forward higher requirements for disaster emergency response. Once a disaster happened, information collection is the key to decision-making of response. With the rapid development of the Internet, social media platform has become an important source of emergency disaster information. However, social media platforms have a lot of duplication, errors and even malicious content in a short time. Social media content needs to be effectively screened through technical means to provide basis

for accurate disaster emergency response. **Methods:** The development of deep learning greatly boosts the accuracy and the efficiency of text task. This study took earthquake disasters as an example, over 50K microblog data in the 72 hours after 5 major earthquakes in China during 2013-2022 were obtained. A multi-label classification model was built by transfer learning based on BERT pre-trained model. Each sample was manually marked as one or more of five types of labels: hazards information, loss information, rescue information, public opinion information and useless information. **Results:** By fine-tune training, the classification accuracy of the model on the training set and the test set reached 95% and 91%, respectively. Single-label AUC score ranged from 0.952 to 0.998. **Conclusions:** Both metrics proved the model is of high reliability. The model can be applied to the emergency management in sudden disaster events, which is conducive to rapidly assisting disaster judgment.

Key words: BERT; transfer learning; social media; earthquake; disaster emergency response; multi-label classification

First author: LIN Sen, PhD, research assistant, specializes in natural disaster monitoring, assessment and early warning, disaster data mining. E-mail : linsen_ndrcc@126.com

Foundation support: The National Key Research & Development Program of China (2018YFC1508806).

网络首发:

标题: 基于BERT迁移学习模型的地震灾害社交媒体信息分类研究

作者: 林森, 刘蓓蓓, 李建文, 刘旭, 秦昆, 郭桂祯

DOI: 10.13203/j.whugis20220167

收稿日期: 2022-03-28

引用格式:

林森, 刘蓓蓓, 李建文, 等. 基于BERT迁移学习模型的地震灾害社交媒体信息分类研究[J]. 武汉大学学报·信息科学版, 2022, DOI: 10.13203/j.whugis20220167 (LIN Sen, LIU Beibei, LI Jianwen, et al. Social media information classification of earthquake disasters based on BERT transfer learning model[J]. *Geomatics and Information Science of Wuhan University*, 2022, DOI: 10.13203/j.whugis20220167)

网络首发文章内容和格式与正式出版会有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

面向非洲猪瘟疫情的社交媒体信息提取与舆情挖掘

林安琪, 吴浩, 韩磊, 岑鲁豫

武汉大学学报·信息科学版, doi: 10.13203/j.whugis20210406

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20210406>