单位代码 **10006**

学 号 **15061115**

分类号 **TP391.1**

****

毕业设计(论文)

110处警预案自动生成算法研究

|  |  |
| --- | --- |
| 院（系）名称 | 计算机学院 |
| 专业名称 | 计算机科学与技术 |
| 学生姓名 | 舒代昕 |
| 指导教师 | 丁嵘 |

**2019**年**5**月

110处警预案自动生成算法研究 舒代昕 北京航空航天大学

北京航空航天大学

**本科生毕业设计（论文）任务书**

Ⅰ、毕业设计（论文）题目：

110处警预案自动生成算法研究

Ⅱ、毕业设计（论文）使用的原始资料（数据）及设计技术要求：

本课题的目的是设计并实现一个基于BERT自然语言处理模型以及tf-idf相似度比对算法的110处警预案自动生成系统。该系统可以通过公安局的真实警情数据集进行分类模型训练，并且利用TF-IDF算法实现预案推荐优先级排序，最终给出系统推荐出的最优预案。在此基础上实现了人机交互功能，即对于某些案例，在分类模型判断效果不好的情况下进行人工评判，并且以此评判结果对模型进行优化。

Ⅲ、毕业设计（论文）工作内容：

接处警是公安系统中非常重要的一部分，全国公安系统每天都会处理非常庞大的接处警数据，在消耗相当数量的人力与物力的同时，如此庞大的数据量却只被储存在公安系统的数据库中，未能有效地利用起来。在这样的背景下，如何利用好这些数据，设计出能够协助民警进行接处警工作的系统，从而提高公安部门接处警的处理效率，使得人民警察能够更高效地为人民服务，是尚待深入研究的课题。本文为了解决这些难题，设计出一个实现处警预案自动生成的系统。本文完成了以下的工作：

1. 调研课题背景，了解国内外对于提高接处警效率的研究现状，学习有关NLP的知识，了解并学习NLP领域相关算法。

2. 为实验研究搜集相关数据，通过联系重庆市某公安分局，在他们的授权下获得了7000条接处警数据，覆盖了近90种常见的警情类别标签，较好地符合本项研究的需要，利用这些数据，我们建立了初步的警情数据库。再对所获得的原始数据进行人工标记等操作，使得该数据能够用于本课题实验。

3. 采用基于双向transformer的BERT语言模型进行警情分类模型的训练，采用基于稀疏向量相似度的TF-IDF模型进行警情相似度比对模型的训练。再根据以上两种模型的输出结果构建预案推荐结果评测模块。评测模块能够根据模型输出结果选择对于该案例相对最优的处警预案。

4. 搭建预案推荐系统，整合分类模型、相似度比对模型与结果评测模块，实现接受用户输入、分析并判断出相对最优解、向用户返还相对最优处警预案的功能。具体流程为：整个系统在前端接受用户通过文件输入或者以文本直接输入的方式输入案例后，对该案例进行警情分类，然后在该标签下的案例库中进行案例相似度比对并进行相似度排序，完成排序工作后将相对最优的处警预案返还给用户。

5. 优化系统，增添模型效果评测模块，采用人机交互功能来优化模型性能。评测模块通过对分类模块以及相似度比对模块的结果进行分析，能够得出本系统模型对于该类警情的处理效果。并对效果不佳的结果进行驳回处理，交付予专家进行人工评判，再将人工评判的结果作为训练数据对模型进行更新优化。

6. 整理整个实验中获得的数据，回顾所有材料，撰写论文

Ⅳ、主要参考资料：

1. Christopher D.Manning, Hinrich Schutze. 统计自然语言处理基础[M]. 电子工业出版社, 2005.

2. Daniel Jurafsky. 自然语言处理综论[M]. 电子工业出版社, 2005

3. 宗成庆. 统计自然语言处理[M]. 清华大学出版社, 2013

4. 郑捷. NLP汉语自然语言处理原理与实践[M]. 电子工业出版社, 2017.

5. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. CoRR abs/1810.04805 (2018)

6. Miguel Grinberg. Flask Web开发——基于Python的Web应用开发实战[M]. 人民邮电出版社, 2014

计算机 学院（系） 计算机科学与技术 专业类 150614 班

学生 舒代昕

毕业设计（论文）时间： 2018 年 10 月 29 日至 2019 年 5 月 22 日

答辩时间： 2019 年 5 月 30 日

成 绩：

指导教师： 丁嵘

兼职教师或答疑教师（并指出所负责部分）：

计算机 系（教研室） 主任（签字）：

注：任务书应该附在已完成的毕业设计（论文）的首页

**本人声明**

我声明，本论文及其研究工作是由本人在导师指导下独立完成的，在完成论文时所利用的一切资料均已在参考文献中列出。

作者：舒代昕

签字：

时间：2019年 6 月

110处警预案自动生成算法研究

学 生：舒代昕

指导老师：丁 嵘

摘要

在国家安防工作中，接处警是非常重要的一个环节，接处警的效果直接影响公安局民警的工作效率，也影响着国家和社会的长治久安。因此对于接处警智能系统的开发是今后较为重要的研究方向。实现接处警任务的智能处理，解放接警人力资源则是110 接处警智能系统的主要目标。本文基于重庆市某派出所提供的接处警数据，设计了一种110处警预案自动生成系统。该系统由前端网页模块、警情分析模块和结果评测模块组成，实现了对输入警情分类并进行处警预案推荐的功能。

首先，本文对警情分析模块进行了构造，该模块使用由BERT语言模型训练而成的警情分类模块和由TF-IDF模型训练而成的文本相似度比对模块组成，能够实现对输入警情的分类和推荐处警预案输出的功能。

其次，本文针对模型中仍然存在的问题设计出了结果评测模块，提供人机交互的功能，使用户能够对系统做出的错误判断加以修改，并且对模型进行更新优化。

最后，本文使用Flask框架搭建了处警预案自动生成系统，该系统包含警情分析模块，前端网页模块以及结果评测模块。

关键词：接处警，文本分类，文本相似度比对

The Research of Automatic Public-Security Response Algorithms

Author: Daixin Shu

Tutor: Rong Ding

**Abstract**

In the national security work, public-security response is a very important part. The effect of public-security response directly affects the work efficiency of the public security bureau, and also affects the long-term stability of the state and society. Therefore, the research of automatic public-security response system is an important direction in the future. The implementation of the automatic public-security response of the police task to liberate the security resources is the main part of the intelligent national security work system. Based on the data provided by Chongqing public security bureau, this paper designs an automatic public-security response system. The system consists of a case classification module and a case text similarity comparison module, which implements the function of classifying the input cases and recommending the pre-plan to response cases.

Firstly, this paper constructs the case analysis module, which uses the case classification module trained by the BERT language model and the text similarity comparison module trained by the TF-IDF model. The case analysis module can classify the input cases and recommend the pre-plan of these cases.

Secondly, this paper designs a result evaluation module, providing the function of human-computer interaction, which enables the user to modify the misjudgment made by the algorithms and optimize the model.

Finally, this paper uses the Flask framework to build the automatic public-security response system, which includes a case analysis module, a front-end web page module and a result evaluation module.

**Key words**: Public-Security Response，Text Classification，Text Similarity Comparison

目录

[1 绪论 1](#_Toc9846484)

[1.1 课题来源与意义 1](#_Toc9846485)

[1.1.1 课题来源 1](#_Toc9846486)

[1.1.2 选题的背景与意义 1](#_Toc9846487)

[1.2 研究现状 2](#_Toc9846488)

[1.2.1 国外研究现状 2](#_Toc9846489)

[1.2.2 国内研究现状 3](#_Toc9846490)

[1.3 本章小结 4](#_Toc9846491)

[1.4 论文结构安排 4](#_Toc9846492)

[2 研究目标与主要任务 6](#_Toc9846493)

[2.1 研究目标 6](#_Toc9846494)

[2.2 研究内容 6](#_Toc9846495)

[2.3 主要任务 6](#_Toc9846496)

[2.3.1 构建语料库 6](#_Toc9846497)

[2.3.2 搭建智能解决方案推荐系统 7](#_Toc9846498)

[2.3.3 推荐结果评测以及优化 7](#_Toc9846499)

[2.4 本章小结 7](#_Toc9846500)

[3 数据的处理与使用 8](#_Toc9846501)

[3.1 数据收集 8](#_Toc9846502)

[3.2 数据分类 9](#_Toc9846503)

[3.3 数据整理 11](#_Toc9846504)

[3.4 本章小结 12](#_Toc9846505)

[4 关键技术 13](#_Toc9846506)

[4.1 分类模型 13](#_Toc9846507)

[4.1.1 预训练模型与下游任务的应用 13](#_Toc9846508)

[4.1.2 BERT语言模型 13](#_Toc9846509)

[4.2 相似度比对模型 15](#_Toc9846510)

[4.2.1 独热编码 15](#_Toc9846511)

[4.2.2 余弦相似度 15](#_Toc9846512)

[4.2.3 TF-IDF算法 17](#_Toc9846513)

[4.2.4 利用TF-IDF和余弦相似度计算文本相似度 18](#_Toc9846514)

[4.3 本章小结 18](#_Toc9846515)

[5 预案推荐系统设计 20](#_Toc9846516)

[5.1 系统整体结构 20](#_Toc9846517)

[5.2 系统模块设计 22](#_Toc9846518)

[5.2.1 前端模块 22](#_Toc9846519)

[5.2.2 警情分析模块 24](#_Toc9846520)

[5.2.3 结果评测模块 28](#_Toc9846521)

[5.3 本章小结 29](#_Toc9846522)

[总结和展望 30](#_Toc9846523)

[致谢 31](#_Toc9846524)

[参考文献 32](#_Toc9846525)

图表目录

[图 1.1 接处警工作流程图 1](#_Toc9797860)

[表 3.1 警情主要信息 8](#_Toc9797861)

[图 3.1 警情分类 10](#_Toc9797862)

[图 3.2 警情处理结果分类 10](#_Toc9797863)

[图 3.3 警情样例及其人工标注 11](#_Toc9797864)

[图 3.4 每类所含警情数目 11](#_Toc9797865)

[图 3.5 每类标签所含警情个数统计 12](#_Toc9797866)

[图 4.1 三种模型的预训练模型架构 14](#_Toc9797867)

[表 4.1 单词编码表 16](#_Toc9797868)

[表 4.2 简单词向量 16](#_Toc9797869)

[表 4.3 词频词向量 17](#_Toc9797870)

[图 5.1 系统整体结构图 20](#_Toc9797871)

[表 5.1 实验环境软件配置 20](#_Toc9797872)

[表 5.2 实验环境硬件配置 20](#_Toc9797873)

[图 5.2 前端输入界面 21](#_Toc9797874)

[图 5.3 警情分析结果 23](#_Toc9797875)

[表 5.3 分类系数含义 24](#_Toc9797876)

[图 5.4 未经数据筛选训练结果评估 25](#_Toc9797877)

[图 5.5 模型准确率跟随每类警情数量下限的变化趋势 25](#_Toc9797878)

[图 5.6 模型精确率跟随每类警情数量下限的变化趋势 26](#_Toc9797879)

[图 5.7 模型召回率跟随每类警情数量下限的变化趋势 26](#_Toc9797880)

[图 5.8 模型f1值跟随每类警情数量下限的变化趋势 26](#_Toc9797881)

[图 5.9 模型准确率跟随每类警情数量上限的变化趋势 27](#_Toc9797882)

[图 5.10 模型精确率跟随每类警情数量上限的变化趋势 27](#_Toc9797883)

[图 5.11 模型召回率跟随每类警情数量上限的变化趋势 27](#_Toc9797884)

[图 5.12 模型f1值跟随每类警情数量上限的变化趋势 28](#_Toc9797885)

# 绪论

## 课题来源与意义

### 课题来源

本课题的来源是北京航空航天大学软件开发环境国家重点实验室的110处警预案自动生成算法研究项目，由丁嵘老师提出并负责。

在国家安防工作中，接处警是非常重要的一个环节，接处警的效果直接影响公安局民警的工作效率，也影响着国家和社会的长治久安。110接处警的主要工作是接警后直接将警务下派给辖区最近的派出所或巡逻执勤组，其内容包括接处警、信息研判、情报预警、情报信息汇总等等。110接处警工作流程图如图1.1所示：



图 1.1 接处警工作流程图

### 选题的背景与意义

当前110接处警工作存在很大的压力 ：第一、随着各种通讯工具的日益普及以及群众对于社会治安的需求不断提高，110电话呼入量也成倍增加。据统计，2017年北京市公安局日均接警量约2万个，一年就会产生700万条报警信息，在报警高峰时段甚至每两秒就会接到7个报警电话，接处警压力显著提升。然而如此庞大的报警数据中的绝大部分却仅仅作为档案存放在公安系统中，在某种程度上确实是一种浪费。第二、随着城市发展越来越快，导致群众很难准确地描述事发地点，再加上机动车数量呈现爆发式增长，道路上车辆“排长龙”几乎成为每天必然会发生的事情，这也导致处警民警寻找当事人愈发困难，客观上影响了110处警速度和服务质量，容易引起群众误解。第三、部分群众法制意识淡薄，在警察正常的执法工作中如果损害了其利益，就会恶意阻挠，甚至打骂警察，同时该类事件可能会被一些不明真相的群众和一些别有用心的人进行恶意炒作，严重影响了公安机关的形象，无形之中也给处警民警带来了巨大的心里压力。不仅如此，大量的无效报警信息加重着公安机关的工作负担。通常情况下，报案中心一天接到的骚扰来电、谎报警情数量占总报警电话的近15%。

但是，如果能够高效地做好110接处警工作，将会给社会带来巨大的贡献：第一、接处警工作相当于警方处理警情的线索收集和事前侦察过程，提高接处警效率也间接地提高了公安机关破案的效率，有利于维护群众的合法权益和社会的稳定。第二、接处警工作是警方与群众建立连接的第一个桥梁，高效的工作能够充分发挥联系人民群众的窗口作用。第三、接处警工作还能够反映出警方工作的缺陷与漏洞，根据接收到警情处理结果反馈，不断完善快速反应机制，有助于推动治安防控体系的建设。

110接处警工作，你报警我布警，看似简单，实际情况却非常复杂。该工作包括接警、指挥、处警、反馈等多个环节，涉及到各个警种、各个部门、上级指挥中心与基层派出所之间的相互协作，任何一个环节出问题都会影响整个工作。如果能够利用好公安系统中庞大的接处警数据，基于接处警规则和以往经验等对报警信息进行智能筛选与处理等操作，可以大大提高接处警效率，并且有效地减少接警失误的发生。

因此，新形势下做好110接处警工作应按照正规化建设的要求，应用大数据和人工智能技术，切实加强110报警服务台的智能化建设。

## 研究现状

### 国外研究现状

国外接处警服务最初可以追溯到19世纪20年代，当时欧洲和美国等西方发达国家已经出现了安保和报警相结合的行业。第二次世界大战以后，随着全球科学技术的快速发展，西方发达国家的接处警服务业也进入了一个快速发展的时期。二十世纪六十年代中期，美国提出：技术可能是警察工作的重要工具。美国第一个实时警察计算机系统于同期被安装在圣路易斯警察局。自此之后，警察对于计算机技术的使用迅速扩大[1]。上个世纪末，日本国家警察局就已经建立了卫星通讯系统辅助管理[2]。随着电子政务被视为促进政府，公民和民间社会组织之间的联系和沟通的重要因素，信息通信技术也被应用于公安领域[3]。2004年，印度特里凡得琅市的警察社区互动门户网站正式运作，迅速改善了该市在健康、文化和社区和谐方面的现状，这与该网站在加强警察与公民联系中所起到的积极作用是密不可分的[4]。2006年，意大利内政部长发起的在线警察局项目Ops正式启动，该项目使用虚拟现实交互界面，公民可以通过在线警察局查询信息、投诉犯罪以及与公安机关进行交流；而警察可以通过该系统监测犯罪趋势，以确保能够及时反应，并更有效地利用稀缺资源打击犯罪[5]。2013年，Greasley等人在英国警察部队指挥中心越来越大的资金限制的背景下，通过将指挥中心的运营方式与传统企业运营管理进行多方面的模拟研究，提出了一个既能有效提高指挥中心接警人员工作效率，又能使接警人员保持合理工作量的轮班制度模式[6]。2018年，Uzlov等人通过数据挖掘的方法，设计并开发了自动化智能技术工具，显著地提高了执法人员信息分析工作的效率[7]。

### 国内研究现状

我国早期的接处警服务只是一个简单的安防工程。经过近30年的发展，我国相关行业正朝着数字化、网络化、智能化的方向发展。Chen等人在2009年设计并实现了一种外勤警务位置感知系统，该系统可以使处警民警能够有效发送或接受警情的有关信息，从而有效提高处警民警的工作效率[8]。Lv等人基于SVG技术，设计开发了警用图形符号创作工具以及警用图形符号发布软件，并利用该软件实现了公安地理信息图形绘制系统，用于公安领域的指挥和动态调度，该系统已成为公安业务管理和信息共享可视化的重要建设内容[9]。在公安系统之外，Deng等人针对广州市消防指挥中心的指挥调度、战略决策机制设计并实现了消防指挥中心规范化综合报警系统，该系统已经实现了对消防人员接警和出警过程的视频监控，并可通过视频监控系统的指挥调度模块实时查看各消防中队的情况。基于3G互联网视频系统，指挥中心可以直接看到事故现场，实现在线指挥抢险救灾[10]。

## 本章小结

目前，我国现有的报警系统充分依托于高度发达的网络通信技术，能够明显加强警察与公民之间的联系。一旦发现报警情况，报警系统可以快速做出响应，及时准确地向平台传输报警信息。而接警人员可以根据具体警情来选择最佳的处警方案，提高处警效率。部分系统除了具备看、听、说的功能外，还具有强大的远程驱动功能。然而，已存在的系统只是实现了信息化，通过通信技术实现了多模态信息的全面交互，但是仍然称不上智能化，机器没有自主决策能力，也没有辅助决策能力，全部的指挥决策还是由人来完成。这会造成处置人员在业务不熟悉、经验不丰富的情况下，出现误判的可能性。

对于北京的接警员来说，虽然说记住北京每一处地方的名字，每一条胡同的位置，每一个地标建筑，都是必备技能。但是刚到岗位的青涩的接警员仍然会出现各种问题。例如：在一次老民警在指导新人接警时，发现接警员在记录中写下了“木樨地”，但老民警从接警员口中描述的报警地点特征判断出事发地点是“木樨园”。原来是接警员把“木樨园”和“木樨地”弄混了。幸亏老民警发现及时，如果因为这个失误出错了警，既耽误了处警，又浪费了珍贵的警力资源。因此，如何通过人工智能技术，减轻接警员工作压力，提高接警处警的准确率，是一个有待进一步研究的方向。

## 论文结构安排

本文主要介绍一种110处警预案自动生成算法系统的设计与实现。以下是本文的结构安排：

1. 绪论

本章主要针对本项目的课题来源，课题背景以及选题的意义进行了介绍。简单介绍了国内外对于提高接处警服务效率的研究及其成果，并指出了目前该领域不足之处以及未来研究方向。阐述了论文的大致结构安排。

1. 研究目标与主要任务

本章介绍了本系统的大致框架与主要模块，简要地介绍了系统主要模块的实现原理以及实现方案。

1. 数据的处理与使用

本章介绍了本次实验所使用的数据来源，并对数据的初步处理与标注方案进行了描述，对本系统所使用的数据格式做了较为详细的说明。

1. 主要算法介绍

本章详细描述了本系统中所使用的主要算法与模型的基本原理，并探讨了本系统选择该算法和模型的原因。

1. 系统架构与内部模块

本章首先对整个系统的结构进行了详细地介绍，描述了如何搭建系统前端与后端，并对系统中的各部分模块及其工作方式进行了较为详细地说明。

1. 总结与展望

对本文的内容进行总结概述，针对本系统提出今后的改进优化方案，并指出未来公安局智能接处警系统可能的发展方向。

# 研究目标与主要任务

## 研究目标

本项研究的目标是建立一个基于警情分类和文本相似度匹配的自动生成处警预案系统。接警员在接到报警后或通过语音识别，或直接记录报警信息的方式将这些文本信息输入该系统进行警情的抽取与分析，该系统再基于处警规则以及以往案例的处警经验知识库自动生成若干条处警预案并进行推荐，同时能够提供以往与此次报警情况相近的处警案例作为参考。

## 研究内容

接警、指挥这个过程其实相当复杂，一般需要多团队协作完成，但是具有一些一般性的流程与规律，我们拟采用基于警情分类和文本相似度匹配的框架。大体思路是通过特征分析、语义理解等技术对警情进行抽取与分析，然后通过接警规则匹配、接警知识库推理判断该警情具体类别以确定大致的处警预案，并计算相应的置信度。如果置信度高于某阈值，则转向该类别的指挥调度知识库，然后通过查询指挥调度知识库，找到当前条件与状态下的应该采取的动作集和处置人员编号，并且通过智能推荐算法对动作集进行排序，得出最高排序的动作，并推荐给该处置人员，以供参考，如果该推荐被处置人员拒绝并退回，则重新推荐；如果置信度低于某一阈值，则判断该警情无法通过系统进行有效分类，并把案情转交给相关专家进行审核，通过对专家智慧的融合，得出最终审核结果，并且将最终的审核结果存储到接警知识库；如果一个推荐被退回超过一定次数，则直接递交专家处理，并且把人工处理的结果录入知识库，为后续的反复迭代训练提供训练样本。通过上述的闭环过程，可以达到人机混合增强的效果。

## 主要任务

主要任务大致包括以下三个方面：

### 构建语料库

构建出高质量的模型是实现本系统最重要的任务之一，而高质量模型的前提条件就是构建高质量的语料库。虽然公安系统中接处警信息的数据量非常庞大，但对于国家安防工作而言，这些信息仍属于未公开信息。因此我们需要从公安局处获得部分接处警信息用于我们的研究。对此，对于本项目使用的实验数据，我们联系了重庆市某公安分局，并在他们的授权下获取了某段时间的七千多条数据构建我们项目实验所需的语料库。

在整个系统构建的过程中，我们的语料库主要用于BERT(Bidirectional Encoder Representations from Transformers)文本分类模型的训练以及TF-IDF(term frequency–inverse document frequency)文本相似度比对模型的训练。前者用于构建事件详情与事件类别的对应关系，从而使系统具备为输入警情分类的功能；后者则是判断事件与事件之间的相似度关系，从而为输入警情推荐相对最优处警预案。

### 搭建智能解决方案推荐系统

要实现处警预案推荐功能，我们首先需要搭建一个具备所有基础功能的处警预案推荐系统，以实现警情的输入以及与该警情所对应的相对最优处警预案的输出功能。此系统由四个基本模块组成：网页前端模块、警情分类模块、警情文本相似度比对模块以及结果评测模块。后端服务器通过Flask框架与前端网页链接，用来接收使用者上传的文本以及向用户推荐结果；警情分类模块通过基于双向Transformer的BERT语言模型实现，而警情文本相似度比对模块通过基于稀疏向量相似度的TF-IDF模型在同类警情中寻找相似案例并进行相似度排序；结果评测模块主要通过分类模块与相似度比对模块的输出结果对预案推荐质量进行评测。对于各模块的详细介绍将会在本文第五章中给出。整个系统在前端接受用户通过文件输入或者以文本直接输入的方式输入案例后，对该案例进行警情分类，然后在该类别的案例库中进行案例相似度比对并进行相似度排序，完成排序工作后将相对最优的处警预案返还给用户。

### 推荐结果评测以及优化

评测模块通过对分类模块以及相似度比对模块的结果进行分析，能够得出本系统模型对于该类警情的处理效果。并对效果不佳的结果进行驳回处理，交付予专家进行人工评判，再将人工评判的结果作为训练数据对模型进行更新优化。对于评测模块性能的评价指标，本文决定使用准确率、精准率、召回率和F1值作为分类模型的评价标准。

## 本章小结

本章是对本实验工作进行一个整体的概述。第一小节主要介绍了本实验的任务，以及需要达成的目标。第二小节详细地介绍了本系统的框架。第三小节对本系统的主要模块进行详细地描述，并简要地介绍了每个模块的实现方案。

# 数据的处理与使用

## 数据收集

对于本项目使用的实验数据，我们联系了重庆市某公安分局，并在他们的授权下获取了某段时间的七千多条数据用于我们项目的实验。

经过对数据的初步处理，包括对于无效信息的删除、对于涉密信息的修改与删除以及对于人员真实信息的修改与删除等操作，我们最终获取的数据主要信息的类型如表3.1所示：

**表 3.1 警情主要信息**

|  |  |
| --- | --- |
| 项目 | 内容 |
| 报警时间 | 报警人呼入电话的事件 |
| 报警类别 | 警情第一级分类 |
| 报警类型 | 警情第二级分类 |
| 报警细类 | 警情第三级分类 |
| 报警子类 | 警情第四级分类 |
| 事件详情 | 报警人报警时所描述的警情 |
| 反馈内容 | 处警人员赶往现场所核实的真实警情以及处警过程描述 |
| 处理结果 | 处警方案分类 |

在开始对数据进行分析之前，我们从公安局处了解了对于一个报警事件的整个接处警流程：首先，指挥中心接到报警人报警，经过简单地了解事件信息后，指挥中心接警员将事件简单地记录下来，之后根据事发地所在辖区将该报警事件下发到该辖区派出所。辖区派出所接警员接警之后记录事件信息，即“报警时间”和“事件详情”，同时根据该事件的初步分类以及其内容简要确定处警人员和处警预案。处警人员出发的同时联系报警人询问报警事件具体的信息，根据这些信息再次确定一个较为详细的处警方案。当处警人员处理结束进行反馈时，将时间、地点、人员、事件详情、处警方案、处警结果进行整合与总结，即为该事件的“反馈内容”，并根据“反馈内容”填写“报警类别”、“报警类型”、“报警细类”、“报警子类”和“处理结果”。由此我们可以得知：

报警类别、报警类型、报警细类、报警子类是公安局已给出的警情分类；

事件详情是报警人报给指挥中心的案情描述，指挥中心接警记录后将该案情下发到其发生地区的派出所，由于每日指挥中心接警量巨大，再加上报警人当时有可能情况紧急不能很好地描述具体警情，所以这个描述和真实的案情可能有出入；

反馈内容是派出所接到指挥中心发来的警情之后联系报警人并出警处理的具体过程，这其中所描述的警情是出警民警联系报警人或者赶往现场详细了解的真实警情，其可信程度比指挥中心下发的警情描述更高；

处理结果是公安局已给出的处理结果分类。

我们首先对公安局现有的警情类别标签进行了研究。公安系统中对于警情主要分为刑事类警情、民事类警情和行政类警情三大部分。其中对于刑事类警情区分得比较详细，门类也更复杂，相较于刑事类警情，民事类警情和行政类警情的区分并没有那么细致。但是刑事类警情的数目却远远比不上民事类警情与行政类警情，近五年来，有相当一部分的刑事类警情标签下的警情数目为0。民事类警情和行政类警情虽然分类较为粗糙，但几乎是每日必然会发生的，并且其数量相当庞大。由于每日警情数量巨大，为了提高工作效率，对于民事类警情与行政类警情，记录民警通常都将其报警类型和处理结果分为“其他”标签。从而导致大部分警情的分类区分度不高，所以不能直接使用公安局提供的分类。

## 数据分类

我们根据派出所所提供的警情分类和处理结果分类，并结合七千多条具体的警情提出了适合本项目实验的警情分类和处理结果分类。为了保证警情分类的真实性，我们将反馈内容中出警民警了解到的真实警情作为判断警情类别的主要依据，并且根据指挥中心下放的警情描述以及记录民警所记录的警情类别辅助判断；将反馈内容中民警根据具体警情所采取的行动用于判断处理结果类别的主要依据。

警情分类主要分为交通事故、纠纷、公民求助、刑事类警情（包括侵犯财产警情、侵犯公民人身权利警情以及其他刑事警情）、治安类警情、灾害事故等多个大类，其具体分类如图3.1所示：

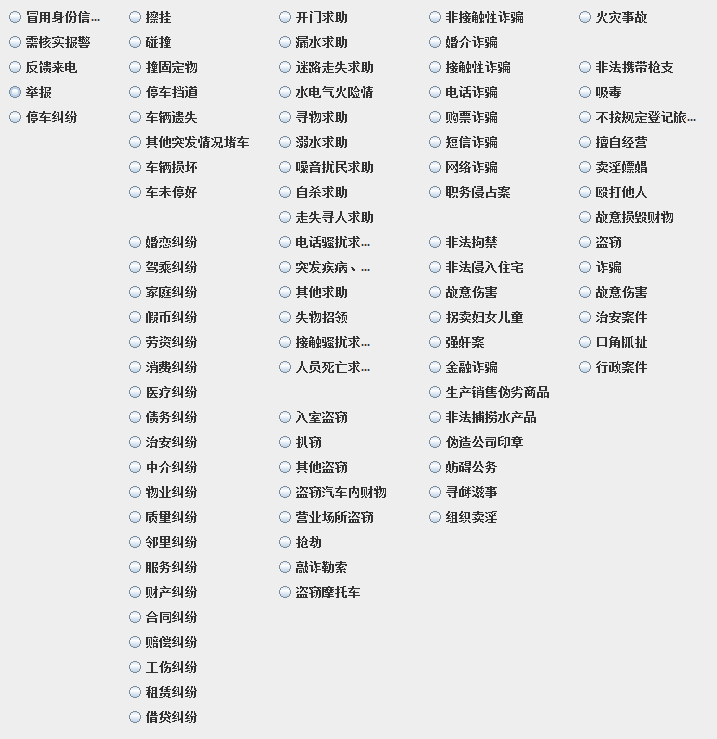


图 3.1 警情分类

我们以派出所提供的类别为主体，再根据人工标注时的具体情况与实验需求进行小规模的添加、修正与删除，最终得出了上述86条能够覆盖我们已获得所有警情的分类。

处理结果的分类，主要是由我们根据处理结果的描述所做出的分类，如图3.2所示：

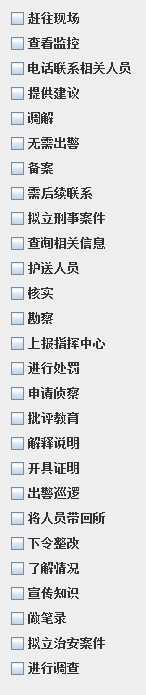
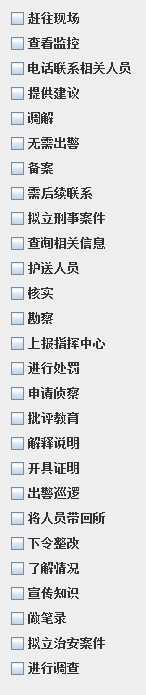


图 3.2 警情处理结果分类

对于一个确定的警情，它的警情分类也是确定的，不会出现给一个警情标注多个标签的情况。而处理结果则可以根据民警的处警行为动作附加多个动作标签。图3.3是几个警情样例以及其人工标注：



图 3.3 警情样例及其人工标注

其中，第一个与第二个警情由于公安局给出的报警类型具有足够的区分度，经过我对反馈内容的核实后决定沿用其分类；第三个和第四个的报警类型覆盖面过于广泛，不具有足够的区分度，于是我根据其反馈内容中的详细情况，结合事件详情，对这两个警情进行了标记。而处理结果的标记由于绝大部分都是“其他”类，所以没什么参考价值，因此我直接根据反馈内容中的民警的行动进行标注。

## 数据整理

在对整个数据进行初步处理之后，我对这部分数据进行了分析，在经过对类别标签数目的分析之后，发现本数据中的标签数目非常不均衡。具体标签中所含警情数目如图3.4所示：

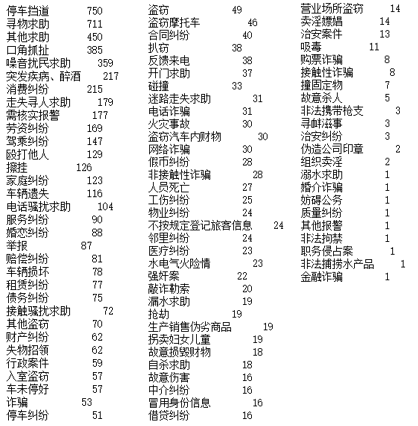


图 3.4 每类所含警情数目

经统计，在所有87种标签中，报警事件数目大于500个的只有2类，报警事件数目在200个到500个之间的只有5类，报警事件数目在100个到200个之间的有9类，其他的标签所包含的报警事件数目均小于100个，而报警事件数目不到10个的标签有18个。详细统计结果图3.5：

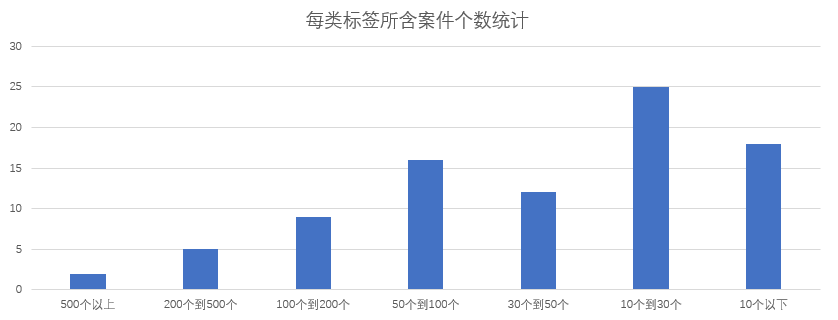


图 3.5 每类标签所含警情个数统计

因此，实验所用数据存在着非常严重的不均衡问题，如果直接使用原数据进行分类模型的训练，很容易发生过拟合的问题，需要对原数据进行整理和过滤。为此，需要做很多的尝试，这方面的工作将会在第五章进行详细的描述。

## 本章小结

由于本系统的实验数据来源于公安系统，数据的格式与内容和实验要求相比具有一定的差距，所以需要进行一定程度人工预处理。本章对于原始数据的人工预处理过程进行了详细的描述。第一小节对于数据的来源以及数据的内容进行了详细的描述，并分析了每种数据对于本实验的意义。第二小节主要介绍了对原始数据进行人工预处理的过程，包括对数据内容中敏感数据以及涉密数据的修改与删除，对数据进行人工标注等操作。第三小节主要介绍了数据中仍然存在并可能对模型产生影响的问题以及其解决方案。

# 关键技术

## 分类模型

在自然语言处理(Natural Language Processing, NLP)中，对文本进行分类是其中非常重要的一个分支。文本分类又分为二分类和多分类，本项目采用的是多分类模型。

### 预训练模型与下游任务的应用

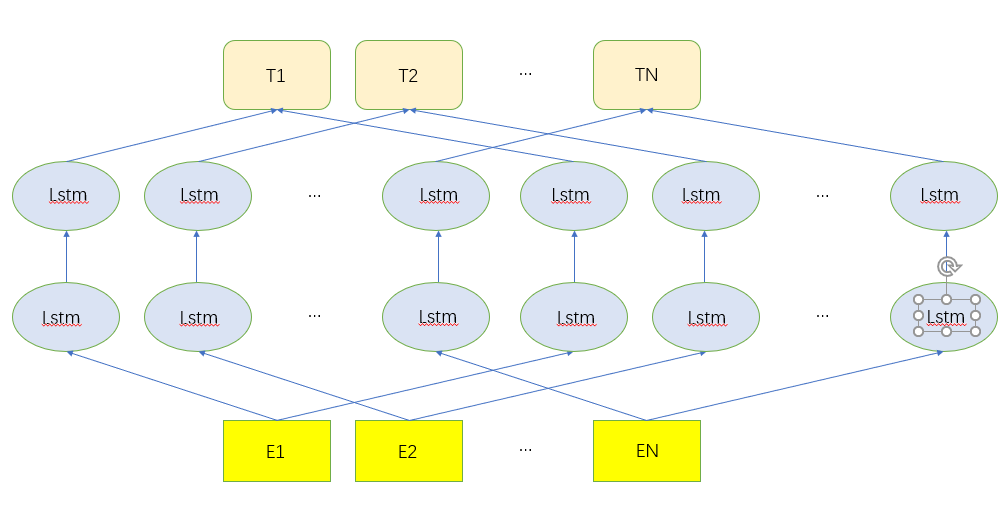
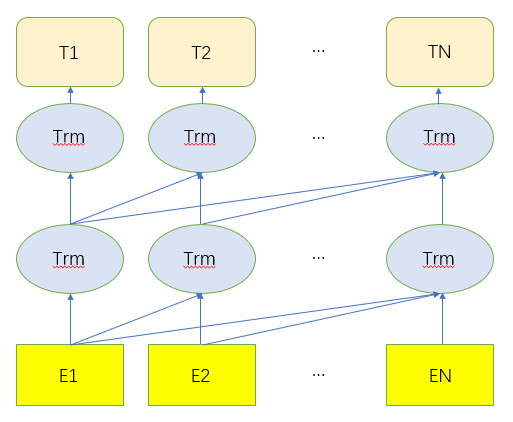
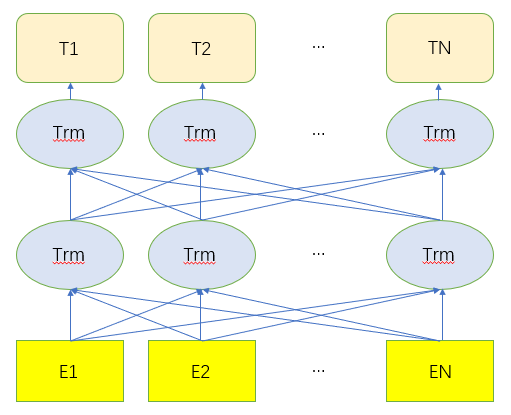
语言模型的预训练已经被多人证实有助于改进多种自然语言处理任务的效果，已成为现代NLP系统不可或缺的一部分[11][12]。相较于从零开始学习，在训练效果上，采用预训练的语言模型有着显著的提高。

而将预训练语言模型运用于下游任务目前有两种策略：一种是基于特征的方法，将预训练的结果作为附加特征加入到自然语言处理任务的特征中。这种预训练策略已经被推广到句子的映射[13]以及段映射[14]的学习中，预训练语言模型被用作这些学习里下游模型的特征；另一种是基于微调的方法，即引入最小的任务参数，并对预训练的参数进行微调来训练下游任务。近期语言模型迁移学习训练的一个方式就是在对一个语言模型目标进行预训练之前，先调整该模型参数对下游任务进行监督，这一方法只需要从头开始学习很少的参数就能够获得良好的效果[15]。往往两种方法在预训练期间共享相同的目标函数，并使用单向语言模型来进行学习。但是单向语言模型限制了在预训练期间对于任务特征的选择。这种限制对于句子级别的训练任务来说会降低训练效果，并且可能会极大地降低基于微调方法的性能。

### BERT语言模型

BERT模型是谷歌于2018年10月发布的基于双向Transformer实现的语言模型[16]。BERT模型在概念上简单，但是效果强大。它刷新了11个自然语言处理领域任务中的最好成绩。

BERT模型架构是基于Vaswani等人实现的多层双向Transformer编码器[17] ，相较于只使用单向Transformer的OpenAI GPT模型和使用分别从两个方向独立训练的LSTM相串联来生成下游任务的ELMo模型，只有BERT模型在训练中同时依赖上下文之间的关系，其结构比较如图4.1所示：



BERT OpenAI GPT ELMo

图 4.1 三种模型的预训练模型架构

显然，构建双向模型比构建单向模型更加严格。一般的语言模型只能单方向地（从左到右或从右到左）或者不完全双向（两个方向单独进行训练）来进行训练，因为采用双向条件会让模型在根据文本序列预测下一个单词对象(token)的时候出现该对象已经存在于序列中的问题，即单词“自己看见自己”(see itself)。

为了构建双向预训练模型，受到Taylor等人完型任务的启发[18]， BERT提出了构建一种“Masked Language Model”(Masked LM)。该模型采用对输入的文本随机屏蔽一定比例的单词，即用一个固定的单词（假设为A）替换掉被选中的单词。但是如果用A去替换掉所有被选中的单词，由于单词A在下游任务中可能是不存在的，所以会导致预训练模型与下游任务不匹配，因此对于被选中的单词采取：“80%用A代替，10%用任意单词代替，10%不变”的方式进行替换。之后仅通过上下文来预测被屏蔽单词的方式进行训练，这样就能够获得Masked LM。

除了Masked LM之外，BERT还引入“下一句预测”任务(Next Sentence Prediction，NSP)。由于许多重要的下游任务都是基于对两个句子之间关系的理解上，为了训练用于理解句子之间关系的模型，BERT首先构建一个用于训练该模型的预训练语料库，该语料库是从基础语料库中通过一定规则生成的，具体而言就是为每个预训练示例选择语句A和B，其中B有50%的概率是A的下一句话，50%的概率是来自基础语料库中的随机一句话。完成这项任务的训练之后，BERT能够在“下一句预测”任务中达到97％-98％的准确率。

在实际BERT模型的预训练中，Masked LM与NSP两个任务将会共同完成模型的预训练。经过后续对于BERT预训练模型的测试研究人员发现，去掉NSP任务，会严重损害判断语句之间关系任务的性能，而双向模型更是在所有任务上优于都单向模型。

BERT预训练所使用的语料库结合了BooksCorpus[19]和英文维基百科两个语料库，总共33亿个单词量，并且只提取维基百科的文本段落，使用基于文档级别的语料库而不是由若干句子所组成的语料库，这对于模型提取与识别较长的连续文本序列是至关重要的。模型以256个序列为一组进行训练，每个序列最多包含512个字符，总共12800个字符。而整个训练过程会进行1000000组上述的训练。

BERT模型性能的评估使用了General Language Understanding Evaluation(GLUE)[20]、Standford Question Answering Dataset(SQuAD)[21]、Named Entity Recognition(NER)和Situations With Adversarial Generations(SWAG)[22]等数据集及其评判标准。其中，GLUE是自然语言处理任务中比较全面的一个数据集合和任务集合，已经过多年的补充和完善。GLUE评判标准包含有MNLI、QQP、QNLI、SST-2、CoLA、STS-B、MRPC、RTE八类数据集，涵盖了包括判断语句之间关系、判断语句是否等价、判断语句所含情感等自然语言处理任务。与其他当前热门自然语言处理模型系统相比，BERT模型在所有任务上的表现均优于其他模型，平均准确率(Average Accuracy)提高了6.7%，对于GLUE任务中运用最广泛的MNLI任务，与之前最先进的技术相比，BERT有着4.7%的绝对精度提升。并且在其他数据集任务中，BERT模型的性能也大幅超越之前所有系统。

## 相似度比对模型

与现实当中的物品的相似度的比较相同，语句和语句之间也会有相似度，语句之间的相似度越高，就可以理解为这两句话在其表达意义上可能也有着相近之处。

### 独热编码

独热编码，即One-Hot编码，原本用于表示状态寄存器的状态，每一个状态都有它的独立有效位，并且在同一时间只会有一位是有效的。在自然语言处理的过程中，许多任务都会用到独热编码来进行文本向量化的工作。

比如“我爱北京天安门”这句话，在经过分词之后构造出来的单词表为[‘我’,’爱’,’北京’,’天安门’]。之后，我们就可以对这个单词表进行独热编码：‘我’=[0,0,0,1]，‘爱’=[0,0,1,0]，‘北京’=[0,1,0,0]，‘天安门’=[1,0,0,0]。完成对语料库中所有单词的独热编码之后，我们完成了将单词向量化的工作。

### 余弦相似度

余弦相似度，顾名思义，就是通过计算向量空间中两个向量夹角的余弦值来衡量这两个个体的相似度。同几何概念中的余弦值一样，余弦值越接近1，表明两个向量夹角越小，两个向量越相似；反之余弦值越接近0，两向量的差别越大。而要计算两个语句余弦相似度最重要的一点就是如何将这两句话映射成为两个向量。

由之前的介绍可知，我们可以通过独热编码将语料库中的单词向量化，但并不能做到将句子向量化。简单来说，我们可以将每一个单词的独热编码作为句子向量的一个元素来将该语句向量化，但在原本的语料库非常庞大，某些句子又很复杂的情况下，这种以独热编码向量为元素组成的向量其所占空间将会非常庞大，直接影响到整个系统的性能。如果将独热编码以简单的一个数字来表示，那么整个系统的负担将会小得多。以以下两个句子为例：

A：这条裤子号码大了，那条号码合适。

B：这条裤子号码不小，那条更合适。

经过对两句话中所出现的单词进行编码后得到单词表如表4.1所示：

**表 4.1 单词编码表**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 编码 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 单词 | 不小 | 了 | 合适 | 那 | 条 | 裤子 | 更 | 号码 | 这 | 大 |

使用这个单词编码表为两个句子编码如表4.2所示：

**表 4.2 简单词向量**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 语句 | 编码 | | | | | | | | | |
| A | 8 | 4 | 5 | 7 | 9 | 1 | 3 | 4 | 7 | 2 |
| B | 8 | 4 | 5 | 7 | 0 | 3 | 4 | 6 | 2 |  |

虽然这个方法成功地将两句话映射成为两个向量，但是由于这两个向量的维度不同，不能够去计算它们之间的相似度。所以需要寻找另外的方式去将句子映射成维度相同的向量。

考虑每个单词在句中出现的频率，若某个单词在一个句子中多次出现，说明对于该句子而言，这个单词可能非常重要；一个单词同时在两个句子中高频出现，那么这两个句子可能有着极高的相似度。仍以上面两个句子和其单词表为例，将单词的编码当作句子向量的下标，A与B的词向量如表4.3所示：

**表 4.3 词频词向量**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 单词 | 不小 | 了 | 合适 | 那 | 条 | 裤子 | 更 | 号码 | 这 | 大 |
| A句编码 | 0 | 1 | 1 | 1 | 2 | 1 | 0 | 2 | 1 | 1 |
| B句编码 | 1 | 0 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 0 |

由于单词表中的单词数量是一定的，所以该语料库中的任何句子所映射出来的向量维度都是相同的，于是句子之间的余弦相似度就可以被计算出来了：

由此可见，句子A和句子B对于该语料库中单词的词频向量相似，即两个句子的相似度很高。

### TF-IDF算法

TF-IDF是一种常用于信息检索与文本挖掘领域的技术，用以评估一个字或者单词对语料库中的一个语句或者文章的重要程度。其主要思想是：如果一个字或者单词在某个语句或者文章中出现的频率高，而在其他文章中很少出现，就可以认为这个字或者单词对该语句或者文章很重要，也可以说这个字或者单词具有很好的区分能力，反之则无关紧要。对于公安局接警员而言，为了提高接警效率，其记录警情通常都是记录与该警情有关的关键字或者关键词，因此TF-IDF这种以关键字词的出现频率来判断文本相似度的模型与本实验具有较高的适用性。

TF指词频，是指衡量某单词在文章中出现次数的指标，但是为了避免词频受到文章长度的影响而偏向较长文章，所以我们令TF的值等于该单词词频数除以文章总词数，若给定单词*ti*，其在文章*j*中的词频可以用公式(4.1)表示：

其中分子表示单词*i*在文章*j*中出现的次数，分母表示文章*j*的总词数。

而IDF指逆向文件频率，是衡量词语在整个语料库中普遍程度的指标，其主要思想是：如果在整个语料库中包含某个单词的文章越少，就说明该单词在整个语料库中具有很好的区分能力，这个单词的IDF指标就越大。给定一个单词*ti*，其IDF指标的计算由公式(4.2)给出：

其中表示语料库中的文章总数，表示语料库中包含单词*ti*的文章数目，但是由于可能出现该单词未在该语料库中出现而导致的情况，因此使用作为分母。而最终IDF的值采用该分数取对数的结果。

最终TF-IDF的值是将TF的值与IDF的值相乘得到的结果，即：

此公式也说明了一个单词A对于语料库中文章B的重要性与单词A在文章B中出现的频率称正比，与单词A在整个语料库中出现的频率称反比。

### 利用TF-IDF和余弦相似度计算文本相似度

TF-IDF算法本身并不能计算出两篇文章的相似度，但将TF-IDF算法与余弦相似度相结合计算文本相似度能够起到比直接使用余弦相似度更好的效果。现假设需要计算语料库中两篇文章的相似度，其计算步骤如下：

1. 使用TF-IDF算法，找出这两篇文章的关键词。
2. 按照单词的TF-IDF值，找出对于这两篇文章最重要的若干个关键词（假设每篇文章30个），并将它们合并成一个单词表，计算其中每个单词在两篇文章中的词频。为了减小文章长度对于词频的影响，先对统计出来的词频进行归一化，再将其结果作为单词在文章中词频。
3. 根据步骤2)中计算出的词频，生成两篇文章各自的词频向量
4. 根据两篇文章对于该单词表的词频向量，计算它们的余弦相似度，该相似度也是两篇文章的相似度。

## 本章小结

本章对实现整个系统的过程中涉及到的算法和模型进行了较为详细的介绍。第一小节主要介绍了本系统的警情分类模块，描述了预训练在自然语言处理任务中的重要地位，简要介绍了几个自然语言处理领域的预训练模型，分析了它们的优缺点，着重介绍了BERT语言模型的架构以及本系统选择它的原因。第二小节主要介绍了本系统对于文本相似度比对模型的选择，描述了独热编码和余弦相似度的相关概念，详细介绍了TF-IDF算法原理，以及利用TF-IDF算法和余弦相似度计算文本相似度的主要步骤。

# 预案推荐系统设计

## 系统整体结构

本章主要针对110处警预案自动生成系统的整体构架和各模块功能及其实现进行详细地介绍，本系统整体结构如图5.1所示，其中橙色部分为推荐系统主体框架。首先我们需要对警情初始数据进行分类并对每条数据进行人工标注，在进行分词、提取词向量等简单地预处理与特征工程生成语料库之后，利用语料库对BERT分类模型进行训练建立警情分类模块。预案推荐部分则利用TF-IDF模型对语料库进行训练建立警情文本相似度比对模块。而在分类与推荐的过程中，两个模块都会生成结果的置信度，该置信度可用于系统外部对推荐结果的人工评判。

整个系统的工作流程为：

1. 从前端网页输入待处理警情文本，上传至后端服务器中准备进行后续操作
2. 利用已训练好的BERT分类模型对该警情的类别进行判断，并检测其置信度。若置信度高于某一阈值，则将该类别输出，并将警情交付于文本相似度比对模块进行处理，否则驳回模型分类结果，将该警情交付于系统外人工处理。
3. 文本相似度比对系统接收到已分类的警情后，将该警情于数据库中同类别的警情进行相似度计算，并进行相似度排序，找出三个与该警情最相似的往期警情，并检测它们的相似度。若存在一个或多个往期警情的相似度高于某一阈值，则将其处理动作作为输入警情的推荐预案输出。否则驳回模型推荐结果，将该警情交付于人工处理。若该警情在分类阶段就因置信度不足而被驳回，仍可以进行文本相似度比对，此时相似度比对的对象就为整个警情库中的所有警情，其他步骤仍与正常比对流程相同。
4. 对于正常输出的分类类别与推荐预案，若用户认为系统得出的结果不符和警情描述，仍可以对该警情的类别与预案进行人工修改处理。人工处理的结果将作为新的标注数据（包括警情分类的标注以及处理方式的标注）输入到整个警情库中，作为对训练数据的扩充。而分类模型与相似度比对模型也会定期进行重新训练进行更新。

系统的整体结构如图5.1所示：

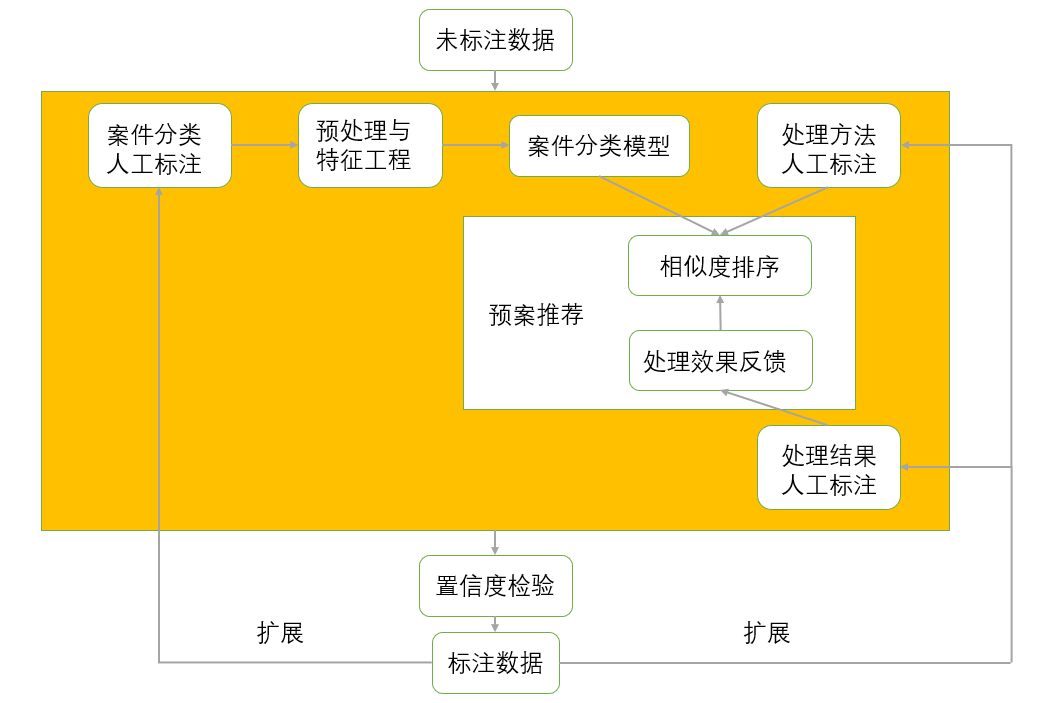


图 5.1 系统整体结构图

本文所介绍的110处警预案自动生成系统在Ubuntu 16 + Win 10双系统服务器下搭建，服务器长期运行的是Ubuntu 16系统。BERT文本分类模型和TF-IDF文本相似度比对模型均在Ubuntu 16系统环境下完成训练。实验环境软件版本信息如表5.1所示：

**表 5.1 实验环境软件配置**

|  |  |
| --- | --- |
| 项目 | 版本 |
| Python | 3.6.3 |
| tensorflow | 1.10 |
| CUDA | 8.0 |

实验环境硬件配置如表5.2所示 ：

**表 5.2 实验环境硬件配置**

|  |  |
| --- | --- |
| 项目 | 配置 |
| CPU | Intel Core i7-8700K at Stock |
| GPU | Gigabyte NVIDIA Geforce GTX1080Ti (Water Cooling) |
| Memory | Kingston HyperX 16G\*4 DDR4 |
| SSD | 256G SSD |
| HDD | 4T HDD（Format in NTFS） |

## 系统模块设计

根据本实验项目所要实现的目标，本系统主要由三个模块组成，分别是：网页前端模块、警情分析模块（包括警情分类模块和警情文本相似度比对模块）以及结果评测模块。

### 前端模块

本系统前端模块使用网页建立，通过浏览器访问，主要用于实现待处理警情的输入以及警情推荐结果的输出功能。当接警员接警之后可以直接以文本的方式将该警情输入系统进行处理，或者将多个警情汇总到一个文件中，上传该文件，对警情进行批量处理。前端模块接收到警情之后，将警情传递给后端模块来实现预案推荐功能，再根据后端模块的反馈将结果输出。浏览器的输入界面如图5.2所示：



图 5.2 前端输入界面

前端网页使用HTML实现，利用Flask框架与后端python相连。Flask框架是一种使用python编写的具有非常高可扩展性的轻量级Web框架，用户可自己安装所需的扩展包来精确地满足自身要求[23]。Web浏览器在使用网页中接口请求服务器响应时，浏览器会把请求信息发送给Web服务器，Web服务器再将该信息发送给Flask程序实例。程序实例保存了一个URL到python函数的映射关系，可以通过信息中的URL请求中获得信息知道下一步应该运行哪一个函数，从而实现Web浏览器对python代码的调用。

Flask框架只是实现了Web浏览器对本地python代码的调用，如果需要使用远端服务器中的案件分析模块功能，则需要与远端服务器建立连接。由于本地系统通过ssh协议与远端服务器建立连接，因此Web浏览器也需要通过ssh协议与服务器建立连接，从而调用服务器中警情分析系统。本系统采用的是paramiko模块，该模块由python语言编写的非交互式shell环境，可支持以ssh协议远程连接服务器并执行相关操作，该模块可以嵌入Flask框架。浏览器可通过调用Flask框架中视图函数的方式调用paramiko模块启动远程服务器。然而在使用paramiko远程登陆ssh，执行shell脚本时，远端只会加载少量的环境变量，因此需要临时模拟cmd命令来加载系统运行时所需的所有环境变量。

Web浏览器发送一个服务器请求并得到反馈的流程大致为：

1. Web浏览器在访问网页页面时，会同时触发与其绑定起来的视图函数。视图函数由route装饰器与网页URL绑定，在本系统中，index()函数的route装饰器为@app.route('/', methods=['GET', 'POST'])，表示其与主页绑定，当浏览器访问主页时，该函数就会被触发并执行。
2. index()函数会调用本地的analyze()函数，该函数设定了运行服务器上训练和测试程序所需的参数，包括输入输出文件地址，临时变量(load\_path)和程序运行命令(cmd)。analyze()函数会通过ssh = paramiko.SSHClient()建立一个paramiko模块实体对象ssh，该对象利用connect()函数与服务器建立连接并通过sftp = ssh.open\_sftp()建立文件传输实体对象sftp，该对象可以将本地文件上传到服务器，也能够将服务器上的文件下载到本地，此时本地需要将待分析警情文件上传到服务器相应地址。上传结束后通过stdin, stdout, stderr = ssh.exec\_command(load\_path+cmd)在服务器上执行本地设定的命令调用服务器上的程序。
3. 在服务器完成对警情的分析后，使用sftp对象将分析结果下载到本地并关闭ssh连接。analyze()函数将分析结果返回到index()视图函数的对象中，视图函数收到分析结果后返回Web服务器的响应，并打开结果展示页面，将分析结果展示给用户。

对于一个具体的警情，经过本系统的分析之后，输出的结果包括：警情信息、警情分类以及与该警情相似度前三的案例、相似度以及该案例的处警方案，其输出结果如图5.3所示：



图 5.3 警情分析结果

### 警情分析模块

警情分析模块包括警情分类模块和相似度比对模块两个子模块。用于接受网页端传送过来的待分析警情并进行分析，两个模块的本质均是计算该警情在模型中的置信度，并根据该置信度相用户推荐最佳结果。以下是对警情分析模块中两个子模块的详细介绍。

当接警人员接受到报警人描述的报警信息时，首先应该对该警情的类别有一个大致的判断，确定了警情的类别之后接警员再根据以往的经验以及接警员自身的判断，制定一个较为详细的处警方案。而警情分析模块首先需要进行的工作就是对输入警情进行分类。警情分类模型通过对经过人工标注和预处理后的数据训练得到。目前训练所用的有效数据共有7000条，将其按照6:2:2的比例划分为训练集、验证集、测试集进行训练。训练效果评测标准使用精确率，召回率，F1值，准确率等指标进行评测。首先我们需要明白四个常用二分类参数TP(True Positive)、TN(True Negative)、FP(False Positive)、FN(False Negative)的含义，如表5.3所示：

**表 5.3 分类系数含义**

|  |  |  |
| --- | --- | --- |
|  | 预测正例 | 预测负例 |
| 实际正例 | TP | FP |
| 实际负例 | FN | TN |

由于本实验中分类模型并不是二分类而是多分类，因此对于本数据集中的每一个类别都设定属于其类别的数据为正例，不属于其类别的数据为负例。以下是对上述评测标准的详细介绍。

准确率(Accuracy)，用来表示模型整体的分类性能，其计算公式如(5.1)所示：

精确率(Precision)，用来表示正确预测为正的占全部预测为正的比例，对于每个二分类任务，它的精确率的计算公式如(5.2)所示：

由于本实验是多分类任务，每个类别都存在一个精确率，所以整个模型的精确率为所有类别的精确率的平均值，如公式(5.3)所示：

召回率(Recall)，用来表示正确预测为正的占全部实际为正的比例。和精确率相同，每个类别都存在一个精确率，其计算公式如(5.4)所示：

模型的整体召回率的计算也与精确率的计算相似，如公式(5.5)所示：

而F1综合考虑了P和R两个指标，其计算公式如(5.6)所示：

经过训练，发现训练的效果并不好，只有0.6的准确率，如图5.4所示：

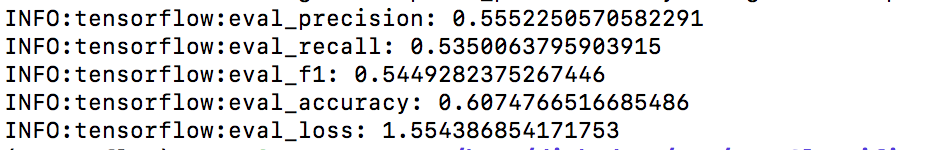


图 5.4 未经数据筛选训练结果评估

在第三章我们已经提出，现有数据存在数据类别不均衡的问题，容易发生过拟合的情况，因此需要对现有数据进行筛选。经过测试，我们决定以限制每一类别中案件数目上下限的方式测试模型性能，其主要方式是通过控制上限与下限两个变量逐步寻找模型性能最好的范围。其中下限指模型只取警情数目高于该值的类别进行训练，上限指模型对于警情数目高于该值的类别，随机选取上限数目个警情作为该类的警情库进行训练。经过测试，不设置上限，改变下限训练出模型的准确率、精确率、召回率、f1值如图5.5所示：

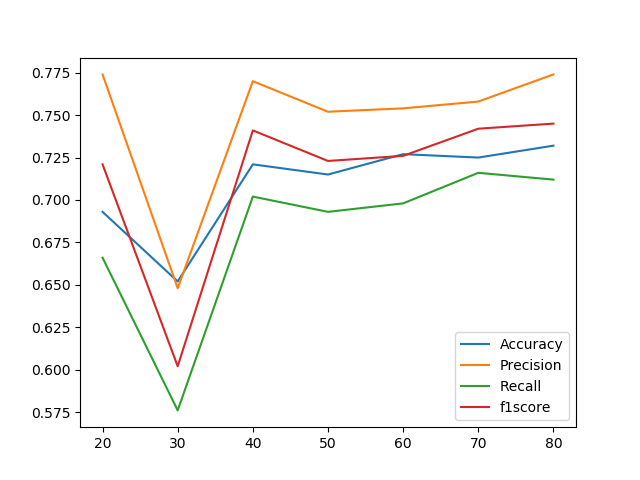


图 5.5 模型性能跟随每类警情数量下限的变化趋势

由图可得，模型的性能整体上是随着每类警情个数下限的提高而提高的，但是如果警情类别只选取警情个数高于60、70或者80的标签时，其所包含的警情总数分别为5199个、5005个、4703个，包含的警情类别个数分别为27个、24个、20个。虽然模型分类准确度有着些许提升，但是其泛用性却大幅下降。而如果选取警情个数高于40个的标签，其所包含的警情总数为5571，涵盖了75%以上的数据，并且所包含的警情类别个数为34个，具有一定程度的泛用性。因此，我们决定将40设定为最终模型警情数下限。

在此基础上，我们控制警情数量的下限为40不变，通过改变上限之后训练出来的模型准确率、精确率、召回率、f1值如图5.6所示：

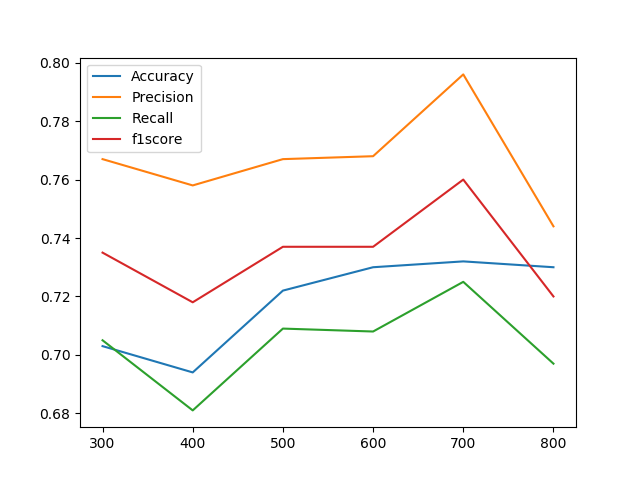


图 5.6 模型性能跟随每类警情数量上限的变化趋势

由图可知，当每类警情数量上限选择为700个的时候，模型的性能最好，并且40-700这个范围也包含了相对较多数量与类别的警情，因此，我们决定将40-700设定为最终用作训练模型的每类警情的数量。

当分类模块获取一个新的警情时执行分类预测脚本run\_predict.sh，在这个脚本中定义若干参数并调用run\_classifier.py文件开始对该数据进行分类预测。其预测结果将该警情对于所有类别标签的置信度输出到一个文件中，再从该文件中寻找出置信度最大的一个标签将其输出。

当得到一个警情的分类之后，接警员对于该警情的处理也有了一个大致的方案。而本系统也采用类似的方法。从警情分类模块出的到警情的分类结果之后，警情相似度比对模块并不与警情库中所有的警情进行比对，而是与相同类别中的所有警情进行比较。不仅提高了警情的推荐速度，还提高了预案推荐的精度。

相似度比对模型仍是以原始警情库为训练数据，通过对所有警情信息进行分词并建立词典，通过该词典对输入警情建立稀疏向量，再将该向量与案例库中所有案例的稀疏向量进行余弦相似度计算。建立词典、建立稀疏向量、计算余弦相似度的具体过程已经在第三章给出，这里便不再赘述。相似度比对模型的输出为输入警情与所有比对警情的相似度，我们一般采用相似度前三警情的处警动作序列作为该输入警情的推荐处警方案输出。

### 结果评测模块

分类模型与相似度比对模型均以预测数值进行判断，这也为人工对预测结果进行评测提供了可能性。由于本系统只能依据数值排序选出相对最优解，而这个解并不一定对于该案例是有效的，因此对于本系统中的分类模块与相似度比对模块，我们人工对其设定一个阈值，在结果评判中，若分类置信度或相似度高于该阈值，我们就可以认为模型预测得出的结果是可靠的，可以直接使用的。若警情最高的分类置信度或相似度未超过该阈值，我们则认定模型预测结果出现误差，需要转入人工评测阶段，此时网页不会出现模型预测的结果，而是提示用户该警情无法通过系统预测，需要进行人工评测。

具体评测过程为：

1. 在系统对输入警情做完分类工作后，检查其输出的置信度，若置信度高于分类模块阈值，则将该警情归为其所分类别，否则驳回系统分类，转为人工评测。
2. 若系统成功对该警情分类，则将该警情与警情库中相同类别的警情进行相似度比对，若存在相似度高于预设阈值的警情，就将其作为可信数据置入预输出警情列表，最后从预输出列表中输出相似度排名前三的警情的处警动作序列作为该警情的处警预案推荐方案。若系统未成功对该警情分类，即该警情系统分类被驳回，此时我们选择警情库中所有的案例与该警情进行相似度比对并将相似度与阈值进行比较，最后将相似度比对结果进行输出。若不存在相似度高于预设阈值的警情，就将所有预案驳回，并将该警情交付人工专家进行处理。
3. 前端网页在接收到警情分析模块的分析结果之后，将该结果返回给用户。网页会为用户提供一个人工评测窗口，在完成警情的人工评测之后，用户可以将评测结果输入到人工评测窗口内，将该警情的分类和处警预案以人工标注的训练数据的形式重新输入到系统之中保存。本系统中所有模型将会根据原数据集以及人工标注的新数据作为训练数据定期训练更新模型。

本系统将分类置信度的阈值设置为0.7，将相似度的阈值设置为0.6。

## 本章小结

本章主要对110处警预案自动生成系统的整体结构与内部模块进行了较为详细的介绍。第一小节介绍了系统的整体框架，并大致描述了整个系统的工作流程为警情文本输入、警情分类、相似警情比对、警情类型与推荐预案输出以及人机交互。最后对整个系统的运行环境以及硬件配置进行了简要地介绍。第二小节介绍了本系统的三大模块：前端模块、警情分析模块和结果评测模块。前端模块部分介绍了网页构造以及与后端进行交互的框架，警情分析模块部分主要介绍了该模块的构成，并且从警情分类模块与相似度比对模块两个方面介绍了警情分析的内部原理，并对第三章提出的数据不均衡的问题给出了解决方案。结果评测模块主要介绍了用户与该系统进行人机交互的方式，并提出了优化模型的有效方法。

# 总结和展望

在国家安防工作中，接处警是非常重要的一个环节，接处警的效果直接影响公安局民警的工作效率，也影响着国家和社会的长治久安。因此对于接处警智能系统的研究是今后较为重要的研究方向。实现接处警任务的智能处理，解放接警人力资源则是110 接处警智能系统的主要目标。本文首先对当前国内外的智能安防系统研究现状进行了简要地介绍，并指出其存在的不足之处。之后依次介绍了本实验数据来源与初步处理、本系统所使用模型及其原理、本系统整体框架及其内部模块三个方面对本实验进行了具体的描述，并从多方面描述了采用BERT模型与TF-IDF模型的优势。

本文的主要成果如下：

1. 搜集并构建了训练所用的接处警警情数据集，为将来本系统进行大量数据训练做好准备。
2. 搭建了前端网页和后端分析系统，并使用Flask框架连接前后端，实现了110接处警系统的整体搭建。
3. 使用BERT模型训练了警情分类模型，使用TF-IDF模型训练了警情文本相似度比对模型。
4. 对将来系统功能的优化以及模型性能的优化提供了可行的方向。

对于北京市而言，110接警平台日均接警量约2万个，一年就会产生700万条报警信息，在报警高峰时段甚至每两秒就会接到7个报警电话。然而到目前为止，仍未有一个具体可行系统方案能够有效地利用如此庞大规模的真实数据，这也意味着在将来，智能接处警系统有着非常长远的发展方向。在本系统实验中，我们只使用了一个派出所中四个月共7000条接处警数据，对于高精度模型训练来说，该警情库非常小。提高警情库的质量和体量，做到一个公安分局一个警情库甚至一个派出所一个警情库，也是未来模型性能改善的方向。

# 致谢

在本文结束之前，请允许我以最高的写意感谢曾经给予我帮助的单位及个人。

感谢丁嵘老师给予我的帮助和指导。在整个毕业设计期间，丁老师给我提出了很多非常有建设性的建议和意见，让我在设计与实现本项目期间少走了许多弯路。丁老师待人亲切，对待学术一丝不苟的精神对我为人处世的态度产生了深刻的影响。虽然这只是一个本科学生的毕业设计，丁老师仍然以一个实验室研究项目的态度对待它，这也使我必须以最严谨的态度和最大的努力去完成我的毕业设计。另外，丁老师也让我在自然语言处理领域有了更深的了解，让我的科研能力有了显著的提高。

感谢北京航空航天大学给我提供的优秀的学习环境与科研环境，在老师们的无私教导下，北航四年的学习生活中我学习到了很多基础与专业知识，使我真正有所成长，没有这四年在北航的学习经历，我是不可能完成本系统的设计与实现的。

感谢李磊学长与潘秋媚学姐，你们的帮助贯穿了我整个毕设过程。在学长和学姐的指导下，我很快便了解并掌握了本项目的具体内容，设计并完善了本项目的整体框架。并且他们在整个毕设过程中传授给我很多实验设计方面的经验，对我完成毕业设计有着很大的帮助。

感谢北航2015级6系全体成员，感谢周其林导员和杨文韬导员，你们为我四年的大学生活中提供了非常大的帮助，让我很快融入到大学生活之中。感谢1506大班的所有同学和150614小班的所有同学，有你们陪伴的大学四年学习和生活是我永远难忘的记忆。

感谢重庆市公安局的大力支持，你们提供的7000条接处警数据是开展本实验的基础。

最后要感谢我的父母，感谢我的家人，你们大力支持我的求学生活，不仅是在物质上，每当我遇到困难，你们还在精神上支持我，帮助我走出困境，为我提供前进的动力。没有你们的支持，我也不会走到现在这一步。你们给予我的期望也是我今后所奋斗的目标，我一定努力拼搏，以更好的成就来回应你们的期望，来报答你们的无私奉献。

# 参考文献

1. Colton K W. The impact and use of computer technology by the police[M]. ACM, 1979.
2. Takashi A. The usage of satellite communications at the police of Japan[C]// Institute of Electrical & Electronics Engineers International Carnahan Conference on Security Technology. IEEE, 2002.
3. Leary, M. L, Rappaport, M. Beyong the Beat. Ethical Considerations for Community Policing in the Digital Age[C]. The National Center for Victims of Crime, Washington D.C., November 2008.
4. Choppella V, Srivathsan K R. Fostering community interaction through the Trivandrum City Police Portal[C]// International Conference on Theory & Practice of Electronic Governance. ACM, 2009.
5. Neri F, Geraci P, Sanna G , et al. Online Police Station, A State-of-Art Italian Semantic Technology against Cybercrime.[C]// International Conference on Advances in Social Network Analysis & Mining. IEEE Computer Society, 2009.
6. Greasley A, Taylor T, Smith C. A simulation study of the shift system at a UK police communications centre[J]. 2013.
7. Uzlov D, Vlasov O, Strukov V. Using Data Mining for Intelligence-Led Policing and Crime Analysis[C]//2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). IEEE, 2018: 499-502.
8. Chen D, Du Y, Zhang Y. Design and implementation of a location awareness system for field police work[C]// International Conference on Geoinformatics. IEEE, 2009.
9. Lv D, Yu M , Song J , et al. Study on police graphic plotting technology based on web[C]// International Conference on Software & Computer Applications. ACM, 2017.
10. Deng L. Design and Implementation of the Normalized Comprehensive Alarming System in Guangzhou Fire Department Command Center[C]// International Conference on Intelligent Computation Technology & Automation. IEEE, 2014.
11. Dai A M, Le Q V. Semi-supervised Sequence Learning[J]. 2015.
12. Peters M E, Neumann M, Iyyer M, et al. Deep contextualized word representations[J]. 2018.
13. Logeswaran L, Lee H. An efficient framework for learning sentence representations[J]. arXiv preprint arXiv:1803.02893, 2018.
14. Le Q V, Mikolov T. Distributed Representations of Sentences and Documents[J]. 2014.
15. Radford A, Narasimhan K, Salimans T, et al. Improving language understanding with unsupervised learning[R]. Technical report, OpenAI, 2018.
16. Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.
17. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[J]. 2017.
18. Taylor W L. "Cloze Procedure": A New Tool For Measuring Readability[J]. The journalism quarterly, 1953, 30(4):415-433.
19. Zhu Y, Kiros R, Zemel R, et al. Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books[J]. 2015.
20. Wang A , Singh A , Michael J , et al. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding[J]. 2018.
21. Rajpurkar P , Zhang J , Lopyrev K , et al. SQuAD: 100,000+ Questions for Machine Comprehension of Text[J]. 2016.
22. Zellers R , Bisk Y , Schwartz R , et al. SWAG: A Large-Scale Adversarial Dataset for Grounded Commonsense Inference[J]. 2018.
23. Grinberg M . Flask web development : developing web applications with Python[M]// Flask Web Development: Developing Web Applications with Python. O'Reilly Media, Inc. 2014.