消防实战指挥智能辅助决策的研究及应用

摘 要

随着我国城市化水平的不断提高，人们的生活和工作方式随之改变，社会和人民群众对消防救援工作提出了更高的要求。超高层建筑、大型地下建筑和新型建筑层出不穷，建筑物分布过于密集，使得城市火灾发生机率不断上升，为消防安全工作带来了新的挑战。研究消防实战指挥智能辅助决策可以对消防救援指挥工作迅速提供科学合理的力量的调派，对减小人民生命财产损失和保障城市安全、社会稳定具有十分重大的意义。

面对日益增多的灭火救援工作，人工智能时代对公安消防部队的数据分析和力量派遣提出了新的要求。本文在总结和研究传统的决策模型的基础上，建立了一个基于深度学习和案例推理的智能辅助决策模型，主要包括预案管理和推理决策两个过程。在预案管理阶段，结合Attention机制和Bi-LSTM自然语言处理方法建立预案特征模型，这是智能决策的重要部分。决策推理部分主要是根据消防案件的特征属性，运用基于Relief-F权值更新算法计算特征权重，并将特征权重应用于案例推理（CBR）的相似度检索上对案例库进行匹配，选取与目标案例最相似的原案例对应的解决方案运用于新的案例上，从而为消防救援进行迅速的指挥调度提供辅助决策方案。

关键词：文本分类；案例推理；消防指挥决策；Attention机制；Bi-LSTM模型

Abstract

With the continuous improvement of the level of urbanization in China, people's life and work style also change, society and the people put forward higher requirements for fire rescue work. Super high buildings, large-scale underground buildings and new-type buildings emerge in endlessly, and the buildings are too densely distributed, which makes the probability of urban fire rising, and brings new challenges to the fire safety work. The research on intelligent assistant decision-making of fire rescue command can provide scientific and reasonable power dispatch for fire rescue command work quickly, which is of great significance to reduce the loss of people's lives and property and ensure the safety and social stability of the city.

In the face of increasing fire rescue work, the era of artificial intelligence puts forward new requirements for data analysis and force dispatch of public security fire forces. On the basis of summarizing and studying the traditional decision-making model, this paper establishes an intelligent decision-making model based on deep learning and case-based reasoning, which mainly includes two processes: plan management and reasoning decision-making. In the plan management stage, the plan feature model is established by combining attention mechanism, Bi LSTM and other natural language processing algorithms, which is an important part of intelligent decision-making. In the decision-making reasoning part, according to the characteristics of fire-fighting cases, the feature weight is calculated by using the updating algorithm based on the Relief-F weight, and the feature weight is applied to the similarity retrieval of case-based reasoning (CBR) to match the case base, and the solution corresponding to the original case which is most similar to the target case is selected to apply to the new case, so as to quickly point out the fire-fighting rescue. Volatilization scheduling provides auxiliary decision-making scheme.

**Key Words：Text classification; case-based reasoning; fire command decision; attention mechanism; bi LSTM mode**

目录

[第一章 绪论 4](#_Toc22150702)

[1.1 选题背景及研究意义 4](#_Toc22150703)

[1.1.1 研究背景 4](#_Toc22150704)

[1.2.2 研究意义 5](#_Toc22150705)

[1.2 研究现状 6](#_Toc22150706)

[1.2.1国外消防指挥智能决策研究现状综述 6](#_Toc22150707)

[1.2.2国内消防指挥智能决策研究现状综述 7](#_Toc22150708)

[1.3 本文主要工作和贡献 8](#_Toc22150709)

[1.4 本论文章节安排 9](#_Toc22150710)

[1.5 本章小结 9](#_Toc22150711)

[第二章 相关基础理论知识及背景综述 10](#_Toc22150712)

[2.1 深度学习相关知识 10](#_Toc22150713)

[2.1.1 深度神经网络概述 10](#_Toc22150714)

[2.1.2 几种典型的深度神经网络 11](#_Toc22150715)

[2.1.3 深度神经网络训练过程 15](#_Toc22150716)

[2.2 统计语言模型概述 16](#_Toc22150717)

[2.2.1 n-gram模型 16](#_Toc22150718)

[2.2.2 神经概率语言模型 17](#_Toc22150719)

[2.2.3 循环神经网络语言模型 18](#_Toc22150720)

[2.3 本章小结 19](#_Toc22150721)

[第三章 文本分类过程及模型研究 20](#_Toc22150722)

[3.1 本文分类过程及方法 20](#_Toc22150723)

[3.1.1 文本分类一般过程 20](#_Toc22150724)

[3.1.2 词嵌入及Word2Vec模型 21](#_Toc22150725)

[3.1.3 文本特征提取与表示方法 25](#_Toc22150726)

[3.1.4 文本分类模型 28](#_Toc22150727)

[3.2 基于Bi-LSTM的文本提取和分类器设计 29](#_Toc22150728)

[3.3 Attention机制 30](#_Toc22150729)

[3.4 基于Attention Bi-LSTM的文本特征提取和分类器设计 33](#_Toc22150730)

[3.5 实验结果与分析 35](#_Toc22150731)

[3.5.1 实验环境 35](#_Toc22150732)

[3.5.2 语料数据集介绍 36](#_Toc22150733)

[3.5.3 实验设计 36](#_Toc22150734)

[3.5.4 实验细节 39](#_Toc22150735)

[3.5.5 实验结果分析 40](#_Toc22150736)

[3.6 本章小结 42](#_Toc22150737)

[第四章 案例推理基研究 44](#_Toc22150738)

[4.1 CBR（Case-Based Reasoning）案例推相关知识概述 44](#_Toc22150739)

[4.1.1 CBR基本理论 44](#_Toc22150740)

[4.1.2 CBR一般流程 44](#_Toc22150741)

[4.1.3 案例的检索方法 45](#_Toc22150742)

[4.1.4 案例学习和评估 46](#_Toc22150743)

[4.1.5 案例库的维护 47](#_Toc22150744)

[4.2 案例推理模型设计及改进 47](#_Toc22150745)

[4.2.1 Relief-F特征加权算法及改进 48](#_Toc22150746)

[4.2.2 案例调整 50](#_Toc22150747)

[4.2.3 案例库维护的改进方法 51](#_Toc22150748)

[4.3 本章小结 54](#_Toc22150749)

[第五章 模型实验及评估 55](#_Toc22150750)

[5.1 应用背景及数据描述 55](#_Toc22150751)

[5.1.1 应用背景 55](#_Toc22150752)

[5.1.2 数据描述 55](#_Toc22150753)

[5.2 案件处置 56](#_Toc22150754)

[5.2.1 案件分析与数据预处理 57](#_Toc22150755)

[5.2.2 案例检索与匹配 57](#_Toc22150756)

[5.3 应急方案及评估 60](#_Toc22150757)

[5.4 小结 61](#_Toc22150758)

[第六章 总结及展望 62](#_Toc22150759)

[6.1 全文总结 62](#_Toc22150760)

[6.2 工作展望 62](#_Toc22150761)

[参考文献 64](#_Toc22150762)

[硕士期间参加的研究项目和发表论文及专利 67](#_Toc22150763)

[致谢 67](#_Toc22150764)

第一章 绪论

* 1. 选题背景及研究意义

1.1.1 研究背景

城市是以人为主体，由社会、经济、资源、环境等要素之间相互作用、相互制约构成的复杂空间地域系统。现代城市化发展越来越快，人口分布稠密，建筑物密集，商业、社会活动集中，存在诸多火灾突发事件发生的社会因素。随着我国城市建设步伐和城市化发展不断加快，造成人口急剧膨胀，由认为因素和自然因素造成的火灾发生事故频率和灾害程度迅速增加，使得城市消防安全面临空前挑战。

党的十九大报告提出了总体国家安全观，报告指出：统筹发展和安全，增强忧患意识，做到居安思危，是我们党治国理政的一个重大原则。必须坚持国家利益至上，以人民安全为宗旨，以政治安全为根本，统筹外部安全和内部安全、国土安全和国民安全、传统安全和非传统安全、自身安全和共同安全，完善国家安全制度体系，加强国家安全能力建设，坚决维护国家主权、安全、发展利益。

根据国家统计局2019年1月最新发布的数据显示，2018年我国城镇常住人口83137万人，占总人口比重（城镇化率）为59.58%，我国正处于城镇化率快速发展的区间，城市发展进入重要转型期。随着城市化水平的不断提高，城市居民的生活方式也发生了巨大的变化，原有的消防安全问题依旧存在：日常生活使用的用电器越来越多，一些常用的如电热毯、烤箱、煤气灶等，都会增加火灾发生的可能性；生活中的装饰材料、易燃纺织物、塑料等也会增加火灾的危险性；随着经济发展，生产需要和消防安全的矛盾也越来越突出，一些商家为了追逐利益，常常会擅自改动防火隔离带、违章拆卸消防设施等，都大大增加了火灾发生的风险；由于城市建筑物密集，人员高度集中，一旦发生火灾事故，人员疏散和救援都比较困难，容易造成大量人员伤亡和财产损失。这些因素决定了我国城市火灾发生的可能性和事故的危害性将不断增加，对城市发展和社会稳定的影响越来越强。尤其是一些重大的火灾事故，不仅会给人民群众的生命财产带来巨大损失，而且会给城市经济建设和可持续发展以及社会稳定带来严重危害。以2010年为例，全国共发生火灾13.2万起，造成的死亡人数为1108人，受伤人数为573人直接经济损失为17.7亿元。全国消防部队接警58.9万起，是建国以来消防救援任务数量最多的一年，其中火灾扑救达13.2万 起，抢险救援达18.3万 起，社会各方救助达14.8万起，公务执勤近１万起，其他出动11.6万起，总共出动人员达615.9万人次，出动灭火救援车达96.8万辆次，营救遇险人数达159965人，挽回经济损失达430多亿元。仅2018年1-8月，1至8月份，全国共接报火灾16.61万起，亡933人，伤560人，直接财产损失20.53亿元，为有效遏制火灾高发势头，确保民众生命财产安全，国家各部委和政府纷纷开展电气火灾治理工作。

1.1.2 研究意义

城市消防救援指挥系统一般包含消防指挥中心，消防力量调派（如消防中队、消防车辆等）、现场的应用系统等，其核心系统是指挥中心的指挥调度系统。工作流程通常分为：火灾报警、消防接警、处警、执行和反馈、监控及记录、数据统计等步骤。特别是近年来，随着大数据时代的到来，消防信息化建设已成为全国消防部队全面实施科技强警、改革警务机制、提高部队战斗力的一项重要战略性举措。随着智慧城市建设和智慧消防工作的深入推进，消防应急管理和指挥已成为现代城市建设和管理的重要内容。消防指挥系统建设的现状集中反映了一个行业、部门和城市乃至国家的消防救援水平。目前我国各地各部门在消防指挥系统建设方面，由于受技术条件限制、行政管理职能划分不同等诸多因素，系统的开发和建设五花八门，没有太多的参考意义，远达不到智慧消防工作的要求。传统的消防作战指挥运行模式多依赖指挥员根据掌握的警情数据人工的决策调配力量，缺乏高科技手段，决策手段单一，决策率低下，耗费时间长，并且在应急决策中，决策者处于一定的时间和心理压力之下，决策的合理性具有一定的盲目性和随意性。不仅会耗费大量时间，人力物力成本，还会错过最佳救援时间，容易造成大量财产损失和人员伤亡。

因此，可以通过计算机辅助决策来提高消防实战应急决策的准确性和效率，提高城市消防指挥决策水平。基于人工智能的消防救援指挥系统技能发挥计算机迅速处理突发事件的优势，又能够根据人工智能中的案例推理方法对案例库中的源案例进行检索和匹配，找出最合适的预案，从而辅助决策者进行决策，大幅缩短警力调配时间，进一步提升消防预警速度和现场指挥水平。通过机器学习推理产生的预案结合决策者更新后的案例又可以作为新的案例加入到案例库中，不断更新不合理和不完善的预案，实现系统的自我学习和自我完善功能。

人工智能技术的引入，将贯穿到接警、处警、决策、指挥调度等实战任务中，大大节省了决策成本，优化分配任务，节约决策时间，遏制火灾势态的发展，提高决策效率。实现城市消防指挥调度迅速、科学、高效等功能，提高城市消防应急指挥管理现代化和智能化水平，最大程度减少城市火灾等突发事件造成的损失和影响，减小人们的生命财产损失，具有重大的经济价值和社会意义。

* 1. 研究现状

1.2.1国外消防指挥智能决策研究现状综述

目前西方各国已相继建立起各自的消防应急指挥系统，早在2000年前，欧洲就将112作为欧盟各国成员的统一报警电话。以美国为例，在2001年“911”事件发生后，美国成立了国土安防部，各城市都成立了“911应急中心”，在处理火灾、洪水等重大消防事件中发挥了巨大作用。美国芝加哥消防应急指挥中心大厦共有1000多人，其中接处警人员100余人，警察500余人，机房200多人，消防200多人，每天处理20000多报警量。一旦接到报警电话，电子屏幕就会立即显示报警人的相关信息和他附近的地图，接警员按照报警信息将事件传递给大厅内相应的警察或消防处理，处警人员通过计算机辅助指挥调度，将指令传给各街区消防中队。

随着大数据时代的到来，各国也将大数据技术和消防指挥结合起来，目前在智慧消防领域应用较为领先的是美国。智慧消防起源于2012年，是由美国标准技术研究院（NIST）提出，将信息物理系统（CPS）应用在消防装备和灭火器材领域。近年来，美国纽约消防部利用大数据技术计算出城市里33万建筑的危险系数，包括建筑物使用年限、用电器使用情况等。当地消防部门利用这一点对火灾发生的概率和发生火灾时的危险程度进行预测，并按照危险程度的高低进行一一排查，对危险性较大的建筑进行标注提示，做好相应的预警工作。通过这种方式，有效提高灭火救援的准确率和效率，保障人民的生命财产安全。

关于智能指挥的研究，国外有学者提出基于GIS的支持决策系统模型[1]，并验证了其在资源调度上的优势，促进消防车辆和资源有效规划和控制。

1.2.2国内消防指挥智能决策研究现状综述

我国正处于经济高速发展的时期，而城市建筑物分布密集，人口稠密，火灾的发生会严重危害人民的生命财产安全和城市经济发展。为此国务院在广西南宁建立了我国第一个应急联动指挥中心，这一系统集公安、消防、交通、急救、防洪等保障措施于一体，是政府对应急救助手段的一次跨越。

大数据为人工智能的应用提供了海量的数据，随着大数据和人工智能技术的不断成熟，应用场景也越来越广泛。智能消防的概念提出以来，不断有学者投入这方面的研究之中，为智能消防不断的注入新的思想、新技术和新方法：

（1）北京的一体化指挥调度工程实验室在2017年提出规则推理的方法[2]，根据事件的描述信息，通过建模计算筛选出历史与此事件相似的典型案例以及其所采用的决策。经过类比的思想，把相似案件的知识应用到新事件中，从而得到新事件的解决方案。

（2）应急管理和应急指挥已成为现代化城市管理的重要内容，与上文类比的思想相似，在应急指挥系统中，广东工业大学的学者在2007年运用了基于案例推理的方法（CBR）并结合基于规则推理的方法（RBR），建立了机器学习的城市应急指挥智能决策模型[3]，并制定了一系列的效益评估模型。根据人工智能中案例推理的方法对历史案例进行检索匹配，找出最匹配的预案知识，从而辅助指挥员迅速决策。在检索匹配中主要使用的是决策树中的ID3算法，使用信息熵作为选择属性的标准对训练集进行分类并构造决策树。运用基于规则的推理机，提取案情描述中的核心关键词及描述值，然后采用正向推理的方法，推理出该案件所属的案件类别。同时，他还使用了关联规则的Apriori算法和聚类的方法进行数据挖掘，抽取案例的特征。并且由于Apriori无法对稀有信息进行分析，基于此他们提出了改进的Apriori-DD算法，减少数据库内不起作用的数据，提高算法效率。

（3）Agent的概念由Minsky在其1986年出版的《思维的社会》一书中提出。Minsky认为社会中的某些个体经过协商之后可求得问题的解，这些个体就是Agent。他还认为Agent应具有社会交互性和智能性。西藏消防支队提出了一种基于Agent技术的智能接处警系统[4]（2018），这是一种软件设计思想。由多个Agent组成系统（MAS）进行交互，协同完成更为复杂的任务。其各Agent结构主要为：接警、电话定位、警力分布、路径寻优、决策和处警6个Agent，具有分布性和并行性的优点。

（4）孔凡全等人[5]提出了基于机器学习技术的消防调度指挥工作，通过RBF径向基网络建模，并使用SPSS软件分析特征指标，设计了基于RBF神经网络的自动化力量调派模型。但由于各地区部门的数据标准不一致，他们团队仅选取了物资仓储场所类型的火灾为研究对象，并且数据特征维度较少，并不具有代表性。

（5）针对非平衡的警情数据，李卫红[6]等人提出，现有的警情数据在时空分布上不均匀，对非平衡数据的训练建模属于弱学习器，预测精度较低。他们采用集成学习算法来提升非平衡数据弱学习器的性能。这是一种基于K-Means均值聚类的Boosting算法，它以Boosting为基础，应用GA-BP神经网络生成基分类器，借助K-Means均值聚类算法进行基分类器的集成，从而实现将弱学习器提升为强学习器。这一研究为消防智能决策提供了新思路。

* 1. 本文主要工作和贡献

本文的主要研究内容和研究成果总结如下：

本文学习研究了案例推理的一般过程，包括案例表示、案例检索、案例库的构建和案例评估等。在案例的表示中，针对消防警情的本文信息数据，本文又对文本分类的常见过程进行了研究，包括文本表示模型、特征抽取和分类器等部分。

本文在前人研究的基础上，设计了Attention+Bi-LSTM模型用于特征提取，解决了传统RNN文本分类的梯度消失问题。对于文本分类中数据的高维度问题，本文研究并采取了Word Embedding机制，并使用训练好的Word2Vec来代替Embedding层，大大减少了需要训练的参数，减少训练时间。避免了高维度的输入导致Bi-LSTM模型产生维度灾难问题。同时，通过Embedding层的weights参数载入预 训练好的词向量，得到的是包含语义的Word Embedding，提高分类器的性能；而通过Embedding随机初始化的Word Embedding不表达语义。

在案例的检索过程中，应用较广泛的是KNN的检索策略，而KNN一般使用均权法，不同的属性对案例的表示通常具有不同的权重，会影响计算结果的可靠性。本文优化了Relief-F特征选择算法，在现有Relief-F算法上对抽样方法做出了改进，通过计算对占有权重较大的属性赋予较合理的权重，并应用于消防火灾救援案例中的相似度检索上，解决了上述存在的问题。

* 1. 本论文章节安排

本文正文部分共分为六章，具体内容如下：

第一章：绪论部分。主要介绍国内外在消防救援指挥智能决策领域的研究背景及研究现状以及本文的主要工作和贡献；

第二章：主要介绍相关领域的技术综述。主要包括深度学习相关理论、深度神经网络训练过程和统计语言模型等知识。并简要介绍了几种典型的深度神经网络。

第三章：文本分类相关算法及改进方法。首先介绍文本分类过程和相关分类算法，并针对传统文本分类算法存在的问题提出改进算法，引入Attention机制和Word Embedding机制，并设计了基于Attention机制的Bi-LSTM文本特征提取和分类模型，最后通过实验验证了该模型的优势。

第四章：CBR基本原理及改进方法。主要介绍了案例推理的基本原理和相关流程，并针对现有的方法存在的问题提出了相关改进措施。

第五章：实验结果分析和评估。通过实例验证了案例推理和文本分类模型在消防实战指挥决策中应用的可行性。并对实验结果进行了分析，证明了该模型在消防实战指挥决策中的有效性。

第六章：总结及展望。主要总结了文本的工作和贡献，并对下一步的研究方向进行了展望。

* 1. 本章小结

本章对本文的研究背景和意义进行了总结，并对国内外在智能消防领域的研究现状进行了综述。指出了本文的主要研究内容和贡献，最后给出了论文的章节安排。

第二章 相关基础理论知识及背景综述

深度学习（Deep Learning）模型是一种具有深层非线性结构的深度神经网络模型，能够模拟人脑分析和学习的过程，被广泛应用于非结构化数据的处理中。在自然语言处理（Natural Language Process，NLP）中最基础的任务就是统计语言模型，它能够生成富含语义信息的词向量，并且统计语言模型的本质也是一个分类过程。本章节主要介绍深度学习相关知识和统计语言模型的相关模型。

2.1 深度学习相关知识

2.1.1 深度神经网络概述

深度学习是在人工神经网络（ANN）的基础上衍生出来的一种方法[7]。人工神经网络是一种最简单的神经网络，由输入层、隐藏层和输出层构成。在神经网络中，每个神经元被称之为一个节点，不同的节点之间的连线表示连接该节点信号的加权值，称之为权重。如下图2-1所示，是一个简单的前馈神经网络（Feedforward Neural Network）



图2-1 基本前馈神经网络

向量是由*M*个输入变量构成的，神经网络隐藏层有*N*个神经元，输出层有一个最终的输出变量，在输入层进入隐藏层的时候，神经网络的计算流程如下：





其中，矩阵*w*是连接输入层和隐藏层的桥梁，称为weight。常量是一个*N*维向量，称为bias，隐藏层神经元接收到向量*z*之后，会进行激活函数计算。



这里称为激活函数，常用的激活函数有sigmoid，tanh等。

完成了神经元上的计算后，继续向输出层移动。从隐藏层到输出层，也有一个weight向量，图中记为，输出层计算如下：



与在隐藏层类似，输出层也有一个偏置，它与输出层的维度一致。

基于上述基本的神经网络结构，通过增加隐藏层数量和使用权值共享等方法，即可发展为深度神经网络。深度学习模型的本质就是多隐藏层多感知层的神经网络结构。传统的机器学习方法，无论是监督学习还是非监督学习，在进行模型训练之前都需要利用特征工程的方法对训练数据进行相应的预处理。而深度学习模型是一种end to end的模型，既能够完成信息的降维，剔除冗余信息，又能够提取到丰富的特征向量，从训练数据到结果特征进行端到端提取，不需要额外的特征工程。利用深度学习既能够提高学习效率，也使得整个机器学习过程集成化。

2.1.2 几种典型的深度神经网络

本小节主要介绍几种典型的深度学习模型，包括卷积神经网络（CNN），循环神经网络（RNN）和长短时记忆模型（LSTM）。

近几年，深度学习算法在图像处理和自然语言处理领域取得了十分出色的成果，其中卷积神经网络（Convolutional Neural Network，CNN）[8]充分利用多层感知器结构，具备很好的学习复杂、高维和非线性映射关系的能力，在图像识别任务和语音识别中得到了广泛的应用。[9][10]不同于传统的全连接神经网络结构，CNN中每一层的神经元只与上一层部分神经元连接，很大程度上减少了网络中参数的数量，这种结构的优点在于在充分考虑原始数据结构的情况下，极大程度的减少了参数数量，提高计算效率，同时避免过拟合现象。

CNN之所以被广泛应用，其优势总结下来有以下几点：

由于计算机存储数据多以矩阵结构的形式存储，而CNN的网络结构在处理矩阵上有着独到的优势；

1. 深度学习模型具有端到端处理数据的特点，能够使得数据的特征提取过程与实际任务相结合，使整个过程更加集成化，形成一个完整的数据处理结构；
2. CNN中权值共享机制能够极大程度的减少网络参数，使神经网络的适应性更强，同时也能够避免过拟合现象的产生。

循环神经网络（Recurrent Neural Network，RNN）[11]也是深度学习代表性算法之一，能够把近邻信息进行有效整合，是一个随时间变换重复循环的结构。通常情况下，人类理解和思考问题都不是孤立的，而是会结合当前问题的前后环境深入思考，为了解决思考时序性，RNN也就应运而生了。在RNN网络中，输入都是具有时间连续性的序列结构，随着时间的向前推移，历史数据经过计算之后信息并不会被抛弃，而是归入到下一次循环中，因此该结构的神经网络被称为循环神经网络。

经典的RNN结构如下图2-2所示



图2-2 RNN循环结构

类似于CNN的权值共享机制，RNN的每一层参数也是被重复利用的，在RNN模型中也能够极大程度的减小模型参数，降低模型的复杂度，这也是所有深度学习算法模型所共有的优点。正如前面所提到过的，RNN是为了处理时序性的数据而生的，但由此也带来了它的局限性，它无法处理长时间依赖问题[12]，具体表现为，RNN网络对最后输入的信息记忆最深，即长时间依赖。产生原因主要是由于当传播时间超过一定的限制之后，神经网络本身会出现梯度消失问题，所带来的后果就是历史信息在传播若干层之后会消失，无法对后面的结果产生影响。

如上图2-2所示，假设时间序列只有三段，为已知值，神经元没有激活函数，即最简单的RNN传播模型，传播过程如下：

，

，

，

在时刻，损失函数



对于每次训练的损失函数都是每一时刻损失值的叠加，即



使用随机梯度下降法训练RNN模型就是对参数求偏导，不断调整参数值使得损失函数达到最小值的过程。对时刻的求偏导如下（其他时刻类似）：







由式2.10，2.11，2.12可以看出，对并没有长期依赖，但对求偏导，会随着时间的推移而产生长期依赖。由于是与相关的函数，根据上述偏导过程，可以得出任意时刻对求偏导的公式：





若再加上激活函数，如tanh，则





由于激活函数*tanh*的导数，且大部分情况下是小于1的，若，那么当时，，这就是RNN中出现梯度消失的原因，而LSTM能够缓解梯度消失这个问题。

长短时记忆模型（Long-Short Term Memory，LSTM）是一种特殊的RNN，能够学习长期依赖关系。它们由Hochreiter＆Schmidhuber（1997）提出，并在随后的工作中被许多人提炼和推广[13]。



图2-3 LSTM单元模块结构图

与传统的RNN模型结构不同的是，LSTM单元在内部增加了若干个“门”结构，分别是遗忘门，输入门和输出门，它们由sigmoid函数（）和点乘操作组成。如图2-3所示，三个乘积操作代表三个门，sigmoid函数的值介于0至1之间，0代表流入不能通过门，1代表能通过。LSTM的第一步是要决定“遗忘掉”哪些信息，由遗忘门的sigmoid层控制，的计算如下：



其中表示遗忘门权重，表示遗忘门偏置。为更新值，为新的候选值，它们的计算如下：





这里，分别为更新门权重和偏置，为候选值权重和偏置。将旧的状态与相乘，丢弃需要丢弃的信息，得到新的存储信息和最后的输出。







其中分别为更新输出值的权重和偏置，为最终输出。

LSTM为什么能够缓解梯度消失呢？类似于传统的RNN，当前状态



将代入有：



如果加上激活函数tanh，即



与RNN中式2.16求偏导的过程相似，LSTM中



令，可知，回顾RNN中梯度消失的原因，是由于式2.16，而在LSTM中，即，这样就解决了传统RNN中导致的梯度消失问题[19]。

2.1.3 深度神经网络训练过程

早期的神经网络模型复杂程度较低，需要训练的参数规模小，通常使用感知机算法训练神经网络。随着深度神经网络结构的越来越复杂，需要训练的参数越来越多，感知机无法满足这种需求，进而产生了BPTT(Back Propagation Through Time)算法[14][15]。深度学习的训练过程就是通过误差向前传播，对模型中的参数进行更新优化。

为了对深度学习模型进行优化，首先需要定义代价函数表示真实值与模型计算结果之间的误差，误差值越小，说明模型的参数越优，模型的拟合程度越好，越趋近于实际情况。对于给定的包含*n*个样本的训练集，对于任意样本，其误差代价函数为：



为了得到最小的代价函数值，模型不断地对参数*w*，*b*进行更新。而在实际应用中，需要先初始化参数，进而对代价函数进行优化求解。常用的优化算法包括随机梯度下降、遗传算法和模拟退火等。梯度下降算法进行权值更新的公式如下：





其中为学习率（Learning rate），表示每次更新的步长，深度神经网络的训练过程就是以代价函数的值达到最小为目标，不断地更新迭代模型参数，从而使得深度学习模型的计算结果更加接近真实值，实现对模型参数的优化。

2.2 统计语言模型概述

自然语言处理的关键问题在于如何表示句子，不同于传统的图像、语音等可以转化为计算机编码元素，自然语言需要复杂的编码才能被计算机理解。并且这种编码是简单的排列，本身不带有任何语义层面的信息，只是为了解决自然语言的存储问题，因而会丢失很多语义上重要的信息。统计语言模型正是为了解决这种问题[16]。

统计语言模型是这样定义的：

假设一个句子*s*由*n*个词按顺序排列组成，表示为：



那么该句子出现的概率表示该句子的合理性，用概率对语言建模，不需要进行语义层面的分析，这也是机器学习解决问题的典型思路，自然界的事物规律可以由概率表达。是*n*个词出现的联合条件概率，可表示为：



其中的概率由决定。目前构建统计语言模型的方法有很多，最典型的有n-gram模型、神经概率语言模型和循环神经网络语言模型。

2.2.1 n-gram模型

考虑到的计算，利用贝叶斯公式可得：



根据大数定律，当语料库足够大时上式可转化为：



其中和分别表示词和词在语料中出现的次数，可知当很大时，和统计起来会多么耗时。由式2.31可知，一个词出现的概率与他前面出现的所有词都有关。如果假设一个词出现的概率只与前面固定数目的词有关呢？这就是n-gram的基本思想，基于马尔可夫的思想，认为一个词出现的概率只与前*k*-1个词有关，式2.32可以变为：



于是式2.33就可以变为



这样简化，不仅使得参数综述变少，也使单个参数的统计变得简单了。

2.2.2 神经概率语言模型

神经概率语言模型（Neural Probabilistic Language Model，NPLM）是在2003年被Bengio等人首次提出的一种通过人工神经网络来训练的语言模型[16]。一个基本的神经概率语言模型网络结构如下图2-4所示，主要包含输入层、投影层、隐藏层和输出层。



图2-4 神经概率语言模型结构图

对于语料中的任意词*w*，表示前面的*n*-1个词，类似n-gram。因此训练样本



于是有：





这里隐藏层激活函数，*W*和*U*分别为投影层和隐藏层，隐藏层和输出层之间的权重矩阵，*p*，*q*分别为隐藏层和输出层上的偏置矩阵向量。上述得到的是一个长度为*N*的向量，它的分量不能表示概率。因此要得到当上下文为时下一个词恰好为词典*D*中第*i*个词的概率，需要做一个softmax归一化：



其中*i*表示词*w*在词典*D*中的索引，*j*表示中的单词在单词表中的索引。

与n-gram模型相比，神经概率语言模型不需要平滑处理，因为softmax函数决定了函数值一定会处于**（0，1）**区间中。另一方面，通过神经概率语言模型训练得到的词向量能够体现词和词之间的语义关系，但是也存在一个很大的问题：引入n-gram思想后，中心词的长度是固定的，没有完整的包含一句话的历史信息，一定程度上会丢失一部分信息。其次，隐藏层的存在会使得网络参数较复杂，计算量较大。

2.2.3 循环神经网络语言模型

循环神经网络语言模型（Recurrent Neural Network Language Model，RNN-LM）是一种基于循环神经网络的语言模型，它是由Mikolov等人在2010年提出的。[17][18]由于RNN是随时间变化不断重复循环的结构，利用RNN训练出的语言模型，能够包含丰富的语义信息，极大程度上满足语言模型任务本身。

由于循环神经网络语言模型本身也属于一种概率模型，表示为中心词与中心词前的词和历史状态信息组成的条件概率：



这里*T*表示训练样本的数量，分别表示*t*时刻下输入数据、预测目标和历史状态信息。



图2-5 循环神经网络语言模型结构图

图2-5所示为循坏神经网络语言模型结构示意图，分为输入层、隐藏层和输出层。与一般的神经网络语言模型结构不同的是该模型没有投影层，由于隐藏层是被反复利用的，所以这种模型结构比传统的神经网络语言模型更优。

前面提到过，RNN独特的网络结构对于序列结构的数据处理上有天然的优势，文本信息则是典型的序列结构数据。同时，相比于神经概率语言模型训练样本中给出的上下文长度是固定的，循环神经网络语言模型由于利用了RNN独特的循环结构，在隐藏层中能够将出现过的历史信息全部保存，可以使用任意长度的上下文作为条件。

2.3 本章小结

本文主要介绍了深度学习相关知识，包括深度学习的起源及演化、深度学习的训练过程，并介绍了几种典型的深度神经网络及其优缺点。由于本文的重点在于文本分类和案例推理，因此在本章的后半部分介绍了统计语言模型相关方法，为下一章节介绍自然语言处理和文本分类算法做铺垫。

第三章 文本分类过程及模型研究

在消防案例库中，对案件的描述通常包含案件类型这一字段，通常情况下是接警员根据当时报警人说明的信息，人工判断后填写上去的。在实际工作中，消防局每天要处理成千上万的消防案件，由于情况紧急且种类繁多，人工处理这些信息不仅效率低，还有可能错误的识别警种，导致做出错误的决策，所带来的后果是不可估量的。因此使用自然语言处理的方法，对报警人说明的文本信息自动识别和分类，不仅可以提高工作效率，还能够避免接警员误判所带来的麻烦和损失。

文本分类是自然语言处理中最基本的任务，文本信息作为一种非结构化数据，难以用固定的特征进行表示，将文本信息转化为机器可以理解的结构化信息是文本分类任务中的一大难点。词语是自然语言处理中具有表达能力的基本单元，在进行自然语言处理时常常以词语为基本单位首先表示。文本分类的任务通常可以分为词语表示和文本特征提取，本章首先介绍文本分类的一般过程，随后介绍词嵌入方法、文本特征提取方法和常用的文本分类模型，并在分类器构造进行了研究。

3.1 本文分类过程及方法

3.1.1 文本分类一般过程

文本分类的主要任务是将文本数据集中的文本通过分类器映射到标签集合，文本分类主要任务如图3-1所示，主要包括数据清洗、分词、词向量构造、文本特征提取与表示、分割数据集、训练分类器等几个步骤[20]。



图3-1 文本分类流程图

1. 数据清洗主要是针对文本中出现的异常字符和标点符号进行删除。对于一些无实际意义，对文本语义没有什么影响的字符进行删除，例如中文的“的”、“或”等停顿词；英文中的“to”、“and’”等。进行文本数据清洗后能够使文本最大程度保留有用信息，提高文本的精度，也能够降低词向量的维度，降低分类器计算的复杂度。
2. 词语是表达句子语义的基本单元，由于中文的特殊结构，词和词之间没有明显的间隔，对于中文语料需要对文本进行分词处理，对于英文则不需要分词。虽然在中文中汉字是最小的单位，但是词语才能表达独立的含义，因此对中文文本信息的分词处理是必要的。
3. 词向量的构造和训练是文本分类中的一项重要任务。由于在自然语言中词语是表达语义的基本单元，为了便于计算机理解和处理，需要对词语进行向量化表示。
4. 文本特征提取和表示是文本分类的重要步骤。由于文本数据是非结构化的，没有预定义的数据模型，因此从文本中提取有效特征并对该文本进行表示是自然语言处理的关键。
5. 分割数据集。将数据集划分为训练集和验证集是深度学习的常规方法，训练集用来训练模型和分类器，测试集则用来测试分类器的性能。
6. 分类器构造。利用分割出来的训练集进行训练和拟合，不断调整模型参数，将模型代价函数训练到最低，并用测试集数据进行测试。

3.1.2 词嵌入及Word2Vec模型

词语是自然语言处理中具有表达能力最基本的单元，因此在进行自然语言处理时常常以词为基本单位进行优先表示，即使用一连串的数字或者向量表示某个词语，也被称为词向量表示[21]。常用的词向量构造方法有：独热编码（one-hot）和分布式表示（Distributed representation）[22]。传统的独热表示方法仅仅对词进行符号化处理，使用*N*位寄存器（单词表的长度）来对*N*个状态进行编码，使用这种方式的明显缺点在于，一是词向量维度很大且稀疏，二是任意两个词向量之间相互独立，没有任何语义信息关联。而分布式表示方法是一种基于深度学习和机器学习的Word Embedding方法。该方法利用统计语言模型对中心词及其上下文进行建模，通过神经网络对大量的语料信息进行训练，输出指定维度的特征向量。用这种方法的好处在于，生成词向量维度可控，并且包含了丰富的语义信息，同时也起到了特征降维的作用。

Word2Vec是Google2013年开源的一个词向量训练工具[23]，可以将单词映射成实数值向量，其模型核心思想是：利用深度学习模型训练出指定维度的词向量，这些词向量之间的相似度，即词语之间的距离，表示了词语之间的相似度。得益于Hinton在1986年提出的Distracted representation[22]方法，Word2Vec模型也是用该方法来表示词向量。Word2Vec主要有CBOW和Skip-Gram两种模型。它们都是一种三层的神经网络结构，但是Skip-Gram模型是利用中间词去计算周围词语为词典中某一词的概率，而CBOW模型则是利用上下文去计算当前词为词典中某一词的概率。它们都有两种训练策略：基于层次化softmax的网络结构（Hierarchical Softmax）来训练和基于抽样方法来训练，下文会详细介绍。

1. Skip-Gram模型

Skip-Gram模型通过给定的输入词来预测其上下文，其中*k*为上下文窗口的大小，即左右选取词的个数。Skip-Gram可以计算出周围词基于中间词的条件概率，对于语句*S*利用Skip-Gram计算*S*为自然语言的概率的公式为：



其中表示的上下文语境。模型的训练目标就是最大化。对于整个文本*T*，可以得到文本的概率表示模型：



为了得到最大条件概率，模型的最大似然函数表示为：



取对数后为：



Skip-Gram的训练目标就是使上式3.4最大。模型如下图3-2所示



图3-2 Skip-Gram模型

正如前面所说，Skip-Gram的训练策略有基于层次结构和基于抽样结构两种。基于层次结构softmax算法首次由Morin[24]等人用于神经网络语言模型中。Word2Vec使用该种结构在输出层可以提高计算性能，所谓层次化，就是利用树结构的思想将输出层利用softmax解决多分类问题转化为二分类问题。在时间复杂度上对比，利用softmax计算概率的时间复杂度为*O(n)*，而利用层次化结构的softmax计算概率的时间复杂度大约为的数量级。与CBOW模型不同的是，该模型没有隐藏层，可以看作一个2层的结构，输入层与输出层的哈夫曼树相连，每个叶子节点唯一编码，表示语料库中的词语。输入层与哈夫曼树的每个非叶子节点相连，代表词语的类别，每条连接都有以单词的词频表示的权重。另一种基于抽样结构的策略如下图3-3所示



图3-3 基于抽样的Skip-Gram模型

基于抽样结构的Skip-Gram模型思想是Mikolov等人[22]提出的，可以看作是Gutmann等人提出的NCE方法（Noise Contrastive Estimation）的改进版，将抽样函数与softmax结合起来的一种方法，只需要将各个词按照其对应的词频来随机选取并进行反例更新。具体如下：统计词典中每个词的词频占总词语数的比例，归一化到长度为1的区间，然后在0-1间随机抽取一个数，对应区间中的单词并更新该单词。将句子*j*对应的单词的对数似然函数用下面的公式3.5代替完成抽样过程：



其中表示服从词频分布，表示上下文信息相加之后的词向量。使用该模型的优点在于只需要遍历抽样的一个样本，不需要遍历整个样本库，能够大大减少计算量。

1. CBOW模型

与Skip-Gram模型不同的是，该模型具有隐藏层并且它是利用单词的周围词预测中间词的概率的模型。对于中间词，利用上下文计算属于词典中某个词的概率。

对于语句*S*利用CBOW模型计算其为自然语言的概率为：



其中为文本中单词的联合概率。同样的，对于整个文本*T*，构建如下似然函数：



模型的目标函数就是使最大，因此得到模型的对数似然函数为：



CBOW模型的训练目标就是使得该函数取得最大值。模型的主要结构如下图3-4（图片摘自文献[23]，下同）：



图3-4 CBOW模型一般框架

与Skip-Gram模型一样，CBOW模型也有两种训练策略，具体方法与skip-gram类似，这里不多赘述。基于抽样的CBOW模型结构如下图3-5



图3-5 基于抽样策略的CBOW模型结构

基于抽样的CBOW模型的目标是最大化正样本概率的同时最小化负样本的概率，由此进行词向量的训练。

3.1.3 文本特征提取与表示方法

文本特征提取与表示就是用合适的矩阵或者向量来表示文本，以便计算机能够对文本进行理解和计算。由于文本数据是非结构化的，数据特征表现不明显，想要对文本进行分类或是后续的处理，就需要对文本按一定的规律进行转化，因此对于文本的特征提取是很有必要的。常用的文本特征提取方法有TF-IDF方法、深度学习方法和Doc2Vec方法等。

1. TF-IDF方法

TF-IDF（Term Frequency-Inverse Document Frequency，TF-IDF）词频-逆文档频率统计方法[25]是一种在文本挖掘中被广泛使用的特征向量化方法，TF-IDF算法的主要思想是：如果某个词在某一篇文档中的出现频率越高，并且在其它文章中很少出现，则认为该词或具有很好的类别区分能力，适合用来分类。它体现了在文档中一个词语在语料库中的重要程度。

在介绍TF-IDF算法之前，首先来了解以下词袋模型（word bag model）[26]。它是一种经典的文本表示方法，在实际应用中常常会使用到。词袋模型的思想是：对于集合，*D*表示一个文档，表示文档中出现的词语。从它的思想可以看出，词袋模型的缺点在于：对原文的语序信息不能很好的表达，导致对文本上下文的语境的表示会造成一定的影响。在词袋模型中，词语特征的元素最简单的表示方法是数值特征，即对一篇文档中词语出现的频率进行统计，再利用这些数值的集合进行文本表示。TF-IDF是常用的数值特征表示方法，对于每一条文本，不仅要考虑每个词在该文本中出现的次数，还需要考虑包含这个词的文档数目，利用其导数找出对文本内容贡献更高的词，然后利用这些词表示该条文本。之所以把过程弄得这么复杂，是因为对于一些常见的词，如果这些词在几乎每条文本中都会出现，那么可以认为这些词是常用词汇，对文本的区别度不大，因此利用TF-IDF算法可以在一定程度上提高词语特征的代表性。

基于该方法计算某个词*w*的区分能力大小的公式如下：



其中n(w)表示词*w*在当前文档中出现的频率，*N*表示训练语料库中文本的总数，N(w)表示训练语料库中包含词*w*的总文本数。

1. 基于深度学习的特征提取方法

通过前面章节中对深度学习的介绍可知，深度学习中的循环神经网络模型在处理非结构化数据上有着天然的优势，而文本数据正是这种非结构化序列数据，一篇文档可以看成是一系列词语组成的序列，所有利用深度神经网络提取文本特征是可行的。本小节主要介绍如何利用循环神经网络提取文本特征和表示文本。

由于组成文本的基本单元是词语，因此提取文本特征也应由词语开始，而词语是由词向量组成，因此使用循环神经网络提取特征，只需要将组成该文本词语的词向量依次输入模型中去，通过神经网络不断循环，即可得到关于该文本的特征向量。前面提到过利用Word2Vec方法得到的词向量具有丰富的语义信息，因此词向量是可以直接输入到神经网络模型中去训练的，而不需要做额外的处理。相对于词袋模型，利用深度神经网络提取和表示文本具有以下优势：（1）可以通过调整网络结构控制特征向量的维度，为后续文本分类任务减小训练时间；（2）循环神经网络能够考虑到词语的先后顺序，我们知道一篇文章中语序和上下文环境对于理解这篇文章是有很大作用的。综上，使用循环神经网络对自然语言中文本特征提取和表示是一种比较理想的方法。

1. Doc2Vec模型提取文本特征

Doc2Vec模型又称为Paragraph Vector，是Mikolov等人在2014年提出的基于Word2Vec模型的表示文本和句子的分布式方法[27]。Doc2vec与上面提到的Word2Vec模型网络结构上基本相同，具有以下优点：在样本训练时，句子长度不需要固定；克服了词袋模型忽略语序和句法等缺点。

Doc2Vec模型是在Word2Vec模型的基础上建立起来的，Word2Vec预测出来的词向量是包含语义信息的，同样的，Doc2vec中也应用了类似的结构，因此也是包含语义的。同样的，Doc2Vec也有两种训练方式，一种是PV-DM（Distracted Memory Model of paragraph vectors），类使用Word2Vec中的CBOW模型，如图3-6所示：



图3-6 PV-DM模型结构

由图可以看出，PV-DM与CBOW类似也是一个三层结构，除了增加了句子向量作为输入外，其他的部分都相同。PV-DM训练段向量的基本思想是通过句子以及句子中某个中心词的上下文来预测该中心词。按照这种方法遍历整篇文本便可以得到每个词的词向量和整篇文章的句子向量。

另一种是PV-DBOW（Distracted Bag of Words of paragraph vector），类似于Word2Vec中的Skip-gram模型，如图3-7所示：



图3-7 PV-DBOW模型结构

每次从一句话中滑动采样固定长度的词，取其中一个词作预测词，其他的作输入词。输入词对应的词向量word vector和本句话对应的句子向量Paragraph id作为输入层的输入，将本句话的向量和本次采样的词向量相加求平均或者累加构成一个新的向量***X***，进而使用这个向量***X***预测此次窗口内的预测词。

Doc2vec相对于Word2Vec不同之处在于，在输入层，新增了一个句子向量Paragraph id，句子向量可以被看作是另一个词向量，它扮演了一个记忆。使用词袋模型每次训练只会选取到句子中一小部分词训练，这样会忽略除了本次训练词以外该句子中的其他词，这样单独训练每个词的向量表达，将向量组累加在一起组成句子，这正是词袋模型的缺点，忽略了文本的词序问题。而Doc2vec中的段落向量则弥补了这方面的不足，它每次训练也是滑动截取句子中一小部分词来训练，句子向量在同一个句子的若干次训练中是共享的，所以同一句话会有多次训练，每次训练中输入都包含句子向量。它可以被看作是句子的主旨，有了它，该句子的主旨每次都会被放入作为输入的一部分来训练。这样每次训练过程中，不光是训练了词，得到了词向量。同时随着一句话每次滑动取若干词训练的过程中，作为每次训练的输入层一部分的共享句子向量，该向量表达的主旨会越来越准确。

3.1.4 文本分类模型

经过上述几种方法得到文本特征向量化表示之后，接下来要做的工作就是对他们进行分类。文本分类器的性能决定了文本分类的效果，常见的文本分类器有朴素贝叶斯分类器、支持向量机分类器、FastText、TextCNN、TextRNN等。本小节将介绍几种典型的文本分类器。

1. 朴素贝叶斯分类器

朴素贝叶斯（Naïve Bayes，NB）分类器是一种基于统计学概率的分类器[30]，这种算法的思想就是通过贝叶斯公式计算出文本中每个特征属性属于该类别的概率，在所有类别的结合中找出使条件概率最大的标签作为分类结果，算法的前提是每个特征之间相互独立。对于特征

和类别集合，条件概率计算公式如下：



其中表示文本*j*属于类别*i*的概率，表示文本*j*占训练样本数的比例。设表示单个文本*d*的特征集合且这些特征相互独立，根据贝叶斯公式有：



利用式3.11可以计算出使条件概率最大的标签的类别。贝叶斯算法原理简单，算法效率高，在新闻分类、垃圾邮件识别等领域应用广泛，但是缺点是默认特征之间相互独立，忽略了文本特征之间的关联关系。

1. 支持向量机分类器

支持向量机（Support Vector Machine，SVM）也是一种应用广泛的分类器[31]，最初用于解决二分类问题，后来被发展为解决多分类问题。它的思想是将特征向量利用核函数映射到高维空间，找到一个最佳分类超平面，使得对于不同类别的样本之间的间隔最大。它的优点在于将分类过程作为一个学习过程，能够防止分类器过度训练，并且无论训练样本大小如何，模型只需要能够找到支持向量就可解决问题，能够大大减少计算量，且分类效果表现良好。

1. FastText分类器

FastText是Facebook2016年开源的一个词向量计算和文本分类工具[28]，它的优点非常明显：在文本分类任务中FastText的训练时间比深度网络小很多量级，却能够取得和深度网络相媲美的精度。

FastText的核心思想就是：将整篇文档的词及n-gram向量叠加平均得到文档向量，然后使用文档向量做softmax多分类。这中间涉及到两个技巧：字符级n-gram特征的引入以及分层Softmax分类。模型结构与Word2Vec的CBOW非常相似，如图3-8。



图3-8 FastText模型结构

可以看出，和CBOW一样，FastText也只有三层结构，输入层是多个向量表示的词语，输出都是一个特定的标签，隐藏层是对多个词向量的叠加平均。所不同的是，CBOW的输入是目标词语的上下文，而FastText的输入是多个单词和它的n-gram特征，这些特征用来表示单个文档；CBOW的输入词向量是被one-hot编码过的，FastText的输入是被embedding过的；CBOW的输出目标是词语，FastText的输出是文档对应的类别标签。

另外，FastText的训练本质与CBOW也类似，值得注意的是，FastText在输入时将单词的字符级别的n-gram向量作为额外的特征，即使用Bag of n-gram作为额外特征获取关于局部词顺序的部分信息；输出时使用分层softmax代替传统的softmax，大大降低了模型训练时间。

3.2 基于Bi-LSTM的文本提取和分类器设计

LSTM适用于处理序列数据的深度学习模型，通过网络的不断循环对历史信息和当前信息加权运算得到基于全局语义信息的特征向量。利用LSTM对文本特征进行提取能够保证最后得到的特征向量保留文本的语序和语义信息，并且维度可控。在前面的章节中，我们在介绍几种典型的深度神经网络的时候详细介绍过LSTM，并给出了它为什么能够缓解梯度消失的原因。本文在基于LSTM的基础上利用Bi-LSTM设计分类器和特征提取模型。

在解决自然语言处理问题时，LSTM模型仅考虑上文信息，没有考虑到下文信息，一定程度上会丢失部分语义信息。不同于LSTM，Bi-LSTM[29][34]是一个双层模型，一层从左往右传播，一层从右往左传播，每一层都是传统的LSTM模块，不仅具备了LSTM的优点，还考虑到了上下文的整体语境，真正做到了利用上下文提取文本特征。基于Bi-LSTM的文本特征提取和分类模型基本结构如图3-9所示。



图3-9 Bi-LSTM模型结果图

Bi-LSTM模型结构与LSTM结构基本相同，区别在于LSTM的隐藏层时单向传播的，而Bi-LSTM模型由两个相互独立的隐藏层构成，且传播方向相反。对于相同的输入数据，可以得到两个不同的隐藏层特征向量输出，然后将两个特征向量通过一定的方法拼接经过softmax得到最终的分类结果。相对于LSTM模型，基于Bi-LSTM的文本特征提取和分类模型能够更加充分的考虑隐藏在输入数据中的整体信息，即上下文环境。从前后两个维度获取一个更加完整的特征，最后通过特征向量计算得到输入文本的类别概率分布。

3.3 Attention机制

Attention机制用于自然语言处理问题最早出现在Encoder-Decoder中[33]，它的核心思想是模拟人脑的注意力机制，人脑在特定情况下对于事物的注意力会集中在某一特定的地方，而忽略其他部分的特点。原理是对于较为关键的部分，投入更多的注意力，而对于其他部分分配较少的注意力，合理的利用大脑的资源，并且还可以减少不关键部分对关键因素的影响。Attention模型经常结合Encoder-Decoder模型一起使用，本章节通过介绍Encoder-Decoder的Attention机制来说明注意力机制的原理。

编码-解码（Encoder-Decoder）模型是一种多对多的序列转化模型，本质与前面提到的统计语言模型的思想类似，与统计语言模型不同的是，在Encoder-Decoder中，主要是通过转化模型，将一组输入序列通过编码解码操作预测另一组序列作为输出。在Encoder-Decoder过程中，编码模块会将输入序列通过计算提取其中一个中间语义编码，所有输入数据的信息都将会被包含在中间语义编码内。然后解码器进行解码操作，每次解码都会基于中间语义编码和输入序列的历史信息，拼接得到一个新的序列作为输出结果。Encoder-Decoder也是以RNN结构为基础，如下图3-10所示。



图3-10 基于RNN的Encoder-Decoder模型

如图3-10所示，输入为一个序列化数据，经过RNN单元计算得到中间语义，解码过程通过历史输出信息预测新的输出信息，同时还加入解码中的中间语义。对于输入数据，输出序列为，*C*为编码后的中间语义，那么Encoder过程可以表示为：



Decoder过程是对中间语义C的解码操作：



可以看出，在计算输出时，需要用到解码生成的历史输出。在进行解码操作时，解码源都是同一个中间语义编码，这样带来的后果是模型输入序列中每一个元素对于输出序列中每个元素的影响力相同，然而实际情况却并非如此。此外，中间语义*C*的编码结果也跟输入序列***X***相关，输入序列对*C*的影响力会随着序列传播逐渐增强。因此序列先后顺序对传统的Encoder-Decoder模型影响较大，依赖序列的输入顺序决定输出与实际情况往往不同。

由于以上问题，Attention机制被提出来了，Attention能够模拟人类大脑的思维方式，在处理问题时将大部分注意力放在主要的位置，在Encoder-Decoder表现为对输入序列的注意力概率分布。利用注意力概率分布控制输入序列中元素对输出序列中元素的影响力大小，减小无关信息对输出结果的影响。图3-11表示了一个基于Attention机制的Encoder-Decoder模型。



图3-11 基于Attention机制的Encoder-Decoder模型

输出数据的计算公式为：



与传统的Encoder-Decoder模型中的中间语义编码不同，是长度等于输出元素个数的不同的中间语义编码，是通过对输入序列进行非线性变换得到的：



这里，表示输入数据经过变换后得到的数值，*T*代表输入序列的长度，表示输入元素对输出元素的注意力概率分布。通过以上方式就可以得到全部中间语义编码。基于注意力机制的概率分布计算方法核心思想如下图3-12所示。



图3-12 Attention概率分布计算示意图

图中，输出数据对于输入数据的注意力概率的计算方法为：



这里表示输入向量在Encoder中隐藏层的状态，表示输出向量在Decoder中的隐藏层状态，函数F用于计算两种状态相似的概率。

Attention模型对于传统模型有着显著优化的作用，利用Attention机制处理文本，可以提取出性能更好的文本特征，通过Attention中的数值识别出文本中更加重要的部分，为模型提供了更好的解释性。

3.4 基于Attention Bi-LSTM的文本特征提取和分类器设计

基于Attention机制的思想，将它与Bi-LSTM结合运用到文本特征提取和分类中，通过计算注意力概率分布，可以突出关键输入对输出的作用，优化文本特征向量，提高分类性能。

本文设计的基于Attention Bi-LSTM的文本特征提取和分类器模型的基本思想是：利用Bi-LSTM模型计算各个时刻下的状态并结合最终状态，求出各个时刻状态对于最终状态的概率分布，从而利用注意力分布对最终状态进行进一步优化。本文设计的整个文本分类模型主要包括三个部分：词向量构造、特征提取和分类器设计。

前面的章节中讲到过Word2Vec技术，本模型词向量构造部分使用的正是Word2Vec，通过训练得到富含语义信息的词向量；而对于分类器的设计，直接在输出端连接softmax函数，概率最大的输出即为该文本的类别。特征抽取部分是模型设计的重点，根据前面的介绍可知，Bi-LSTM模型能够有效地处理文本序列信息，同时能够考虑到文本上下文信息，在此基础上结合Attention机制使模型能够在考虑上下文语境的情况下有选择性的保留重要信息。

由于利用softmax作为输出层归一化计算分类，本文选择交叉熵损失函数作为目标函数，因此文本特征提取和分类的目标函数定义为：



其中*T*表示文本数量，表示当前文本类别的真实概率向量，表示当前模型预测的当前文本*i*的概率分布向量，通过最小化目标函数就能够训练得到分类模型。基于Attention Bi-LSTM的文本分类模型如图3-13所示。



图 3-13 基于Attention机制的Bi-LSTM文本分类模型

图中，*F*表示Bi-LSTM前后方向最终状态隐层值之和，即最终状态，*a*代表所有时刻下隐藏层单元状态对于最终状态的注意力概率分布，例如表示*n*时刻下Bi-LSTM的状态对于最终状态*F*的注意力概率，使用softmax函数求得。由n时刻下各独自方向状态相加而来，表示经过注意力加权后的文本特征向量。

其中，*n*时刻下输出数据对于最终状态*F*的注意力概率计算方法如下：





其中*N*代表输入数据中元素的个数，***U***为权重向量矩阵，*F*表示Bi-LSTM中各独自方向最终隐藏层状态之和，表示*n*时刻下各独自方向状态之和，的计算方法在LSTM的介绍中已经详细讲解过，这里不再赘述。在上图所示的模型中，经过注意力加权后的最终特征的计算过程如下：



表示*n*时刻下Bi-LSTM的状态对于最终状态*F*的注意力概率,*N*为输入序列元素的个数，表示*n*时刻下Bi-LSTM各独自方向状态之和。得到注意力加权后的最终文本特征之后，通过softmax层，计算文本分类标签的概率分布，过程如下：





其中*T*表示类别标签数量，***V***为模型输出层的权值矩阵，表示向量中的第*i*个分量，向量长度与*T*相等。经过softmax函数后，可以得到文本类别概率分布*y*，与真实概率分布*Y*代入式3.17求得交叉熵损失。

使用Attention机制结合Bi-LSTM建模进行本分类得优点如下：

1. 相对于Bi-LSTM，加入了Attention机制，加入上下文的同时考虑输入序列对最终编码额注意力概率分布，更加强调关键信息对文本分类得作用，减小次要因素对分类的影响。比传统方法更能够突出重点，对于潜在语义特征具有更强的挖掘能力。
2. 使用Bi-LSTM模型相对于传统的LSTM分类模型，生成的特征向量可以提供更加丰富的语义信息，并且LSTM只考虑了上文信息，Bi-LSTM理论上比它的分类效果更好。

3.5 实验结果与分析

3.5.1 实验环境

本实验是基于深度学习框架Keras上进行实现的。Keras是一个高级API，可以工作在Theano和TensorFlow之上，如果说TensorFlow或者Theano是神经网络方面的巨人，那么Keras就是站在巨人肩膀上的人。作为一个兼容Theano和TensorFlow的神经网络高级包，Keras来组建一个神经网络更加快速，只需要几行代码就可以实现，而且广泛的兼容性能使Keras在Windows和MacOS或者Linux上完美运行。

本文设计的消防警情文本分类实验具体环境如表3-1所示：

表3-1 开发环境

|  |  |
| --- | --- |
| 操作系统 | Windows10 |
| 开发工具 | PyCharm，python3.6 |
| 开发平台 | Keras |
| CPU | Inter Core（TM） i5-4590 |
| RAM | 16G |

3.5.2 语料数据集介绍

本文设计的文本分类框架目的是为了对消防警情文本进行分类，为后面案例推理和指挥调度决策做铺垫。本文所采用的实验数据来自某市2018年的消防接警数据，经过脱敏处理后有效数据约60000多条，数据类型共分为火灾、社会救助、抢险救援、反恐排爆和其他共五类，由于反恐排爆和其他类这两类的数据非常少，本文仅选取了前三类数据作为实验数据。

3.5.3 实验设计

实验分别使用CNN，LSTM，Bi-LSTM和Attention Bi-LSTM模型在相同数据集上进行对比，并分别与使用Word2Vec预训练和不使用Word2Vec预训练加载词向量对比，实验的学习率（learning rate）均设置为0.01，优化算法采用RMSprop()算法，综合比较分类效果。

1. CNN模型结构

CNN模型的结构图和各参数设置如下图3-14



图3-14 CNN文本分类模型结构图

首先将一个文本处理成向量的embedding层，每条警情文本被处理成一个50\*200的二维向量。其中50是每条文本的固定长度，200是这个词在空间中的词向量，通过一层卷积和池化缩小向量长度，再加一层flatten层将2维向量压缩到1维，最后通过2层Dense将向量长度收缩到3，对应警情文本的3类。

实验仅使用了一个卷积层和64个卷积核，由于警情数据长度只有50个词，即50维特征，卷积池化后只剩下16维，再增加多的卷积没有什么意义。

1. LSTM模型结构

LSTM的模型结构如下图，每条文本经过embedding层后得到一个50\*200的二维向量，通过将这50个词向量逐个输入LSTM中，得到一个长度为128的1维向量，最后连接一个全连接层将长度压缩到3。如图3-15：



图3-15 LSTM文本分类模型结构

1. Bi-LSTM模型结构

双向LSTM的结构与LSTM结构基本一致，只是Bi-LSTM的输出层变成了256维。如图3-16所示：



图3-16 Bi-LSTM文本分类模型结构

1. Attention Bi-LSTM模型结构

与Bi-LSTM不同的是，在Bi-LSTM层处理后，输出向量变成了50\*32的二维向量，Attention层在Dropout层之后，并增加了一个Flatten平滑层，将长度变为1600的一维向量，结构如图3-17所示。



图3-17 Attention Bi-LSTM文本分类模型

3.5.4 实验细节

警情文本数据中存在大量噪声，例如电话号码，地址和标点符号等，这些噪声会干扰模型对文本信息的挖掘，对分类结果产生影响。由于中文以字为单位，与英文不同，单独的字不能表达意思，因此需要对文本分词处理。本文使用python的结巴（Jieba）分词工具包，通过停用对研究没有价值的介词代词和标点符号等后，对文本进行分词。按6：2：2将数据集分为训练集、验证集和测试集。

本文采用Word2Vec中的Skip-Gram模型，通过输入来预测上下文，通常来说，词向量维度设置为100~300较为合适，本文所有的词向量维度均设置为200，窗口大小为2（过滤掉出现2次以下的低频词）。使用Word2Vec训练词向量的好处在于：

1) 通过Embedding层的weights参数载入预训练好的词向量，得到的是包含语义的Word Embedding；而通过Embedding随机初始化的Word Embedding不表达语义

2) 使用预训练好的Word2Vec来代替Embedding层，大大减少了需要训练的参数，减少训练时间。

3.5.5 实验结果分析

为验证Attention+Bi-LSTM在消防警情短文本的分类性能，本文对上述设计的实验模型在消防警情语料库上进行实验，并做出训练历史的准确率和损失函数变化，如图3-18，3-19所示：

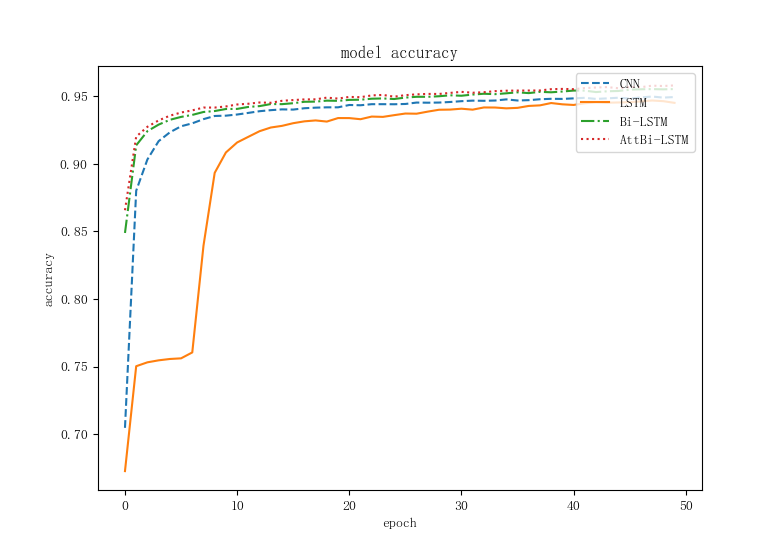


图3-18 4种模型准确率

图3-18展示了基于CNN、LSTM、Bi-LSTM和Attention Bi-LSTM在消防文本信息分类的准确率，可以看出：这四种模型在文本分类上都能取得不错的效果，但是基于Attention Bi-LSTM的文本分类模型的分类准确率更好一些，说明在注意力机制综合考虑上下文信息对于提高文本分类准确率有更好的效果。

从图3-19可以看出所有模型在利用梯度下降迭代训练时，函数的损失值是逐渐下降并最终趋于一个很低的稳定值，模型都取得了较好的收敛效果。其中Attention Bi-LSTM的loss随着epoch的增加最终的loss在四个模型中最低。

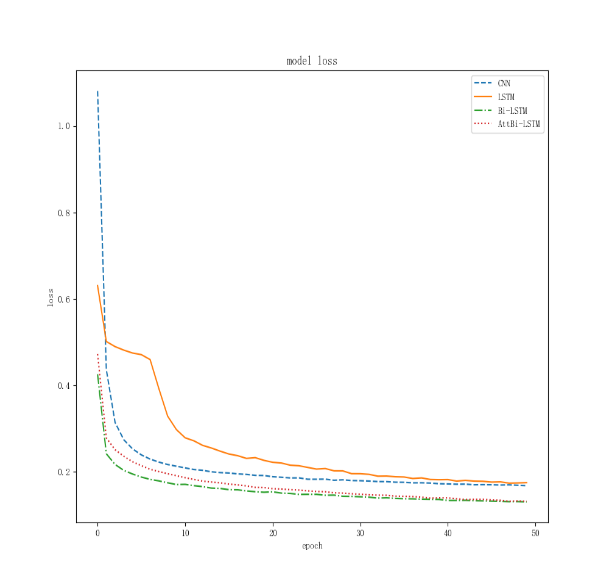


图3-19 4种模型loss

表3-2是不同模型在测试集上的准确率、召回率、F1值和训练每个epoch耗费的时间（秒），并分别使用Word2Vec和不使用Word2Vec对比。

表3-2 不同分类模型结果比较

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **分类模型** | **Precision** | **Recall** | **F1** | **Time/Epoch/s** |
| CNN | 0.9437 | 0.9437 | 0.9436 | 59 |
| LSTM | 0.9547 | 0.9543 | 0.9540 | 61 |
| Bi-LSTM | 0.9506 | 0.9509 | 0.9507 | 139 |
| Attention Bi-LSTM | 0.9602 | 0.9600 | 0.9600 | 87 |
| CNN+ Word2Vec | 0.9437 | 0.9437 | 0.9436 | 33 |
| LSTM+Word2Vec | 0.9533 | 0.9537 | 0.9534 | 39 |
| Bi-LSTM+Word2Vec | 0.9573 | 0.9571 | 0.9572 | 101 |
| Attention Bi-LSTM+Word2Vec | 0.9688 | 0.9676 | 0.9657 | 61 |

从表2结果可以看出，在相同的数据集上，使用Word2Vec训练的词向量作为文本特征输入，训练时间在所有模型中都有明显减小，并且训练准确率相对于不使用Word2Vec预训练稍有提升。引入Attention机制后的模型的准确率在所有模型中准确率最高，但是会牺牲一部分的训练时间作为代价。由此可见，使用Word2Vec训练词向量，可以很好的降低向量维度和高稀疏性，有效减少训练参数；引入Attention机制结合Word2Vec能够达到良好的文本分类效果。

下面是对实验结果的分析：

1. 在基于Bi-LSTM的特征提取和分类模型的基础上增加Attention机制可以明显的提分类准确率，使用Word2Vec训练词向量后在消防警情数据集上精度相对于LSTM或是Bi-LSTM提升了1.5%。实验结果也说明Bi-LSTM综合上下文全局信息能够在一定程度上消除传统方法中信息输入顺序对最终状态的额外影响，加入Attention机制后更关注对分类结果重要的信息，能够有效缓解历史信息对分类结果影响力减弱的问题。
2. 横向比较4中模型在消防警情数据集上的实验结果可以看出，4中分类模型都能够取得不错的分类效果，一定程度上也说明了深度学习算法在自然语言处理中的优势。但是基于Attention Bi-LSTM的文本分类模型的分类准确率更好一些，从理论分析来讲，基于Attention Bi-LSTM的文本分类模型能够存储更多的语义环境信息，还能够提取更深层次的文本特征，为文本分类提供更多的依据。
3. 使用Word2Vec训练词向量，可以很好的降低向量维度和高稀疏性，有效减少训练参数，大大减小训练时间。

3.6 本章小结

本章主要介绍了文本分类过程及方法，随后介绍了基于Attention Bi-LSTM文本特征提取和分类模型，并通过实验对比对模型的性能进行了分析。在本章中，首先介绍了文本分类的一般过程，主要包括词向量构造、文本特征提取和表示和分类器设计。在词向量构造中介绍了词嵌入word Embedding和Word2Vec模型的原理；在文本特征提取方法中列举了TF-IDF文本特征提取方法、基于深度学习的文本特征提取和表示方法以及Doc2vec文本特征提取和表示方法，并说明了几种方法的优缺点；在分类器设计中也介绍了几种典型的分类模型如朴素贝叶斯、支持向量机和FastText等。随后介绍了基于Bi-LSTM的文本特征提取和分类模型，这是一种端到端的深度学习模型，还考虑了文本上下文的信息，将文本特征提取和分类器训练相结合，一方面利用特征提取的结果进行分类，另一方面利用分类标记对特征提取过程进行反馈，优化特征向量，提高分类准确率。模型的优点在于不依赖人工进行特征提取工程，减少工作量，另一方面避免仅从词频的角度进行文本特征提取的局限性。紧接着对Attention机制进行了详细介绍，包括Encoder-Decoder模型，并将Attention机制的思想因应用于文本分类中，结合Bi-LSTM设计了基于Attention Bi-LSTM的文本特征提取和分类模型。最后将四个模型在消防警情数据集上进行实验对比，分析模型的性能。结果表明基于Attention的Bi-LSTM模型能够消除输入数据对输出数据的不合理影响力，强调了关键信息对于文本分类结果的作用，结果还表明Bi-LSTM能够结合上下文语境信息，缓解了由于输入顺序不同对结果造成的影响不同的问题。并且使用Word2Vec训练词向量，可以很好的降低向量维度和高稀疏性，有效减少训练参数，大大减小训练时间，一定程度上也能够提高分类准确率。

第四章 案例推理基研究

基于案例的推理（CBR），是人工智能领域新崛起的一项技术，它是基于过去的经验的推理方法，是人类认知科学的体现。日常生活中，人们为了解决一个新的问题，首先会根据这个问题的特征进行回忆，在脑海中寻找与该问题类似的已解决的问题，并将解决方法应用于新的问题的求解之中。目前基于案例的推理方法已经被广泛的应用于应急管理和医疗等领域。本章介绍了案例推理基本原理和相关流程，并对现有案例推理中的存在的问题进行了改进。

4.1 CBR（Case-Based Reasoning）案例推相关知识概述

基于案例的推理（Case-Based Reasoning，CBR）首先源于认知科学，是由AI衍生而来的一项技术，它的工作方式是案例的推理和学习[33]，这个与人类思维和处理问题的方式一致，从一个已知案例到另一个新案例的类比过程。

4.1.1 CBR基本理论

案例推理是人类解决问题的一种常用方式，即以往解决问题所积累的经验会影响以后在相同或是相似的问题上的思考与判断，它属于类比推理的一种方式，是人类认知科学的体现。从认知科学的角度来说，CBR系统的构建基于两个前提假设：

1. 相同或相似的情况有相同或相似的解法
2. 相同或相似的状况会重复发生

由于CBR的诞生解决了人工智能领域中专家系统（Expert System，ES）存在的弊端：难以获取规则、专业知识难以表达清楚等，并且随人工智能的蓬勃发展，CBR越来越受到大家的重视，已被广泛应用于各个领域之中。

4.1.2 CBR一般流程

1994年，Aamodt和Plaza提出经典的R4模型，CBR的研究大致被分为4个过程[35]，即Retrieve（检索），Reuse（再利用），Revise（修正）和Retain（保留），并且后面的大多数关于CBR的研究都是基于这个模型或是在这个模型上改进的模型。

1. 案例检索：运用相似性算法，将当前案例的特征属性对比从数据库中检索与过去最相似的案例。
2. 再利用：运用相似案例的解决方案处理当前问题。
3. 修正：如果以往的案例解决方法不能完全满足当前案例中出现的问题，可以以过去的案例解决方法为基础进行适当的调整和修正
4. 保留：将修正后的案例解决方法连同案例保存于数据库中存为新的案例。

整个推理流程如下图4-1所示



图4-1 CBR流程图

4.1.3 案例的检索方法

案例的检索就是从案例库中寻找一个或多个与目标案例最相似的案例。CBR系统的优点在于它能够快速而准确的检索出相似案例，案例检索是CBR系统中的关键技术，它的检索策略和过程将直接影响到整个CBR系统的效率和性能。

CBR的检索结果需要满足以下几个性能要求：

1. 有效性：即检索出来的案例尽可能的少；
2. 精确性：即检索出来的案例尽可能的与当前案例的特征属性相关和相似；
3. 易修改性：即检索出来的案例尽可能易于修改，修改后能够解决当前问题；
4. 快速性：即检索时间短，速度快。

目前在CBR系统中常用的检索技术有最近邻法和归纳索引法。

1. 最近邻法

最近邻法是通过计算两个对象在特征空间中的距离来获得两个案例之间的相似度，在CBR案例检索中，大多使用的是该方法，使用这个方法首先要确定案例间距离的定义，距离越小，相似度越高。

设案例，其中是*x*的特征值，是权重，对于*n*维特征空间上的点*X*，*Y*，它们在*D*上的距离为[36]：



其中：



如果连续，当*r*=1时，为曼哈顿距离；当*r*=2时为欧式距离；当*r*>0时，称为闵科夫斯基距离。

最近邻法是一种比较容易实现和理解的检索策略，但是它并不完美。用来检索的目标案例和案例库中的源案例的特征值都是相同的，它们的权重也是确定的。但实际情况并非如此，特征权重是根据情况不同动态变化的。这一点也是本文研究的重点之一，在后面的算法改进部分会提到。

1. 归纳索引法

归纳索引法是不断从案例的各组成部分抽取最能将案例与其他区别开来的成分，类似于决策树，可以有效地组织案例。通过归纳索引能够确定案例的最佳特征，利用归纳的方式安装案例层级关系，将案例数据库建立起来。当要讨论一个新的案例的决策方案时，只需要根据案例所属的类别，即可找出最相似的案例。整个归纳检索的过程，关键在于数据层次性关系的表达，该方法的缺点是：当案例属性无阶层关系或者属性不完整时，无法进行归纳检索。

4.1.4 案例学习和评估

将检索匹配到的相似案例进行适当的调整，使其能适应新的问题。实际情况中需要对调整后的决策方案进行评估，可以通过向专家咨询或者在实际情况中应用来实现。

通过评估后的新案例，如果效果良好，具有使用价值，则把它加入到案例库中作为新案例，并进行案例学习。案例学习过程如下：

设案例库，为案例库中某源案例，，目标案例与的相似度，对于专家定义的阈值，学习策略有以下几种情况[37]：

1. 如果调整后的新案例与案例库中所有的案例均不相似，即，则该新案例可以直接加入到案例库；
2. 如果调整后的新案例与案例库中某案例相一致，即，则新案例方案在案例库中已经存在，无需再加入到案例库；
3. 对于任意的，调整后的新案例可以加入到案例库；
4. 对于任意的，将与案例库中最相似的源案例更新为调整后的新案例。

案例的学习就是不断更新案例库和获取新知识和解决方案的过程，从而动态的扩充案例库，案例推理是一个增量的学习过程，不断地丰富知识有效的适应各种新案例的情况。

4.1.5 案例库的维护

在案例推理系统中，系统的增量式的学习会使得案例库不断的增加。虽然案例库的增加会使得案例知识变得丰富，能够更好的解决问题，但是如果放任案例库一直增加下去，将会导致案例检索时间变长，所带来的负荷越来越大，影响了CBR系统的性能，降低了解决问题的效率，这种情况被称之为“沼泽问题”。因此需采取一定的措施对案例库进行维护，使案例库维持在一定的规模，既不影响其性能，又能够使得案例更好的适应新的情况。

常用的案例库维护方法有[36]：

1. 随机删除法

当案例库中的案例数超出一定限制时，每新增一条案例到案例库，系统会随机选择一条旧案例进行删除。这种方法虽然能够使案例库维持在一定的规模，但由此带来的后果是，系统随机地删除案例可能会导致系统解决问题的能力下降，甚至会出现有些问题无法解决的情况。

1. 基于案例分类的删除策略

通过对案例进行分类，找出核心案例和非核心案例，并将非核心案例按照一定的方式排序。当案例库超出一定的规模时，按照非核心案例到核心案例的排序方式进行案例删除。这种方法的后果是：没有考虑案例的性能度量，可能将高性能的核心案例删除，而低性能的案例还维持在案例库中，不仅没有提高CBR系统的性能，反而会使该系统决解决问题的性能下降。

1. 基于模式归纳的案例库维护方法

通过抽象和泛化来提取案例间的共性，对于案例库中相似度非常接近的案例进行删除保留其一。

4.2 案例推理模型设计及改进

传统的案例推理模型的检索方法都是基于KNN的检索策略，最近邻法是一种比较容易实现和理解的检索策略，但是它并不完美。用来检索的目标案例和案例库中的源案例的特征值都是相同的，它们的权重也是确定的。但实际情况并非如此，特征权重是根据情况不同动态变化的。针对上面存在的问题，我们提出了一种新的特征权重计算方法，在现有Relief-F算法上对抽样方法做出了改进，通过计算对占有权重较大的属性赋予较合理的权重[40]，并应用于消防火灾救援案例中的相似度检索上，解决了上述均权法存在的问题。并且在案例库的维护中，本文采用了聚类分析方法和离群分析法对案例进行维护，一定程度上缓解了传统案例维护方法会影响案例推理性能的缺点。

4.2.1 Relief-F特征加权算法及改进

Relief算法最早由Kira提出[38]，是一种特征权重算法，根据各个特征和类别的相关性赋予不同的权重，权值小于某个阈值的特征认为是对该分类无效的特征而被移除，算法比较简单且运行效率高，但局限于两类数据的分类问题。Relief算法中特征和类别的相关性可理解为：如果该特征对近距离样本的区分能力很好，则说明该特征具有很好的分类能力，并赋予该特征较大的权值。算法思想如下：

从样本训练集*D*中随机选择一个样本*R*，然后从样本*R*的同类样本中寻找最近邻样本*H*，称为Near Hit；从和*R*不同类样本中寻找最近邻样本*M*，记为Near Miss，然后根据以下规则更新每个特征的权重：如果样本*R*和分对样本Near Hit在某个特征上的距离小于样本*R*和Near Miss上的距离，则说明该特征对于分类是有益的，并增加该特征的权重；反之，如果样本*R*与Near Hit在某个特征的距离大于*R*和Near Miss上的距离，说明该特征对分类起到负面作用，并降低该特征的权重。以上过程重复*m*次，最后得到各特征的平均权重，特征权重越大，说明该特征的分类能力越强，反之说明该特征的分类能力越弱。Relief算法的运行时间随着样本的抽样次数*m*和原始特征个数*N*的增加线性增加，因此运行效率非常高。

由于Relief算法只能对两类数据进行分类，1994年Kononeil对Relief算法进行了扩展，得到了Relief-F算法处理多分类问题[39]。

Relief-F算法的基本思想如下：

每次从样本集*D*中随机取出一个样本*R*，然后从与*R*同类的样本中找出*k*个近邻样本*H*，从*A*的不同类样本中找出*k*个近邻样本*M*，最后按照公式更新每个特征的权重。

算法步骤如下：

1. 将所有特征权重初始化为0
2. 随机抽取*m*次 for i=1 to m，do

2.1 随机选择一个样本*R*

2.2 从样本*R*的同类样本中找到*R*的*k*个最近邻样本，从*R*的每个不同类样本中找到*R*的*k*个最近邻样本

2.3 对于*N*个特征 for A =1 to N，do





其中，表示类中的第*j*个最近邻样本，表示样本和在特征*A*上的差，为类*C*所占的比例，为随机抽取的样本*R*的比例，计算公式如下：



其中为样本在特征A上的取值，为特征A的最大值，为特征A的最小值。

该算法存在的缺点：每次随机抽样可能会抽到同一类，计算出的特征权重在该类的属性上可能会被赋予较大的值，这对于分类来说显然是不合理的。本文在该算法的基础上设计了一种新的权值更新算法：在训练集的*m*类样本中每类随机抽样一个计算特征权重，重复抽取*n*次后求平均，这样做的好处在于降低了样本分布差异对特征权重的影响，从而得到更有效的特征权重。

改进后的整个算法权值更新流程如下图4-2所示。



图4-2 基于改进的Relief-F权值更新流程图

样本特征权重的计算方法如下：

1. 从每类样本中随机抽取一个样本，抽取*n*次，样本记为
2. 对于第*i*类的第*t*次抽样，调用一次Relief-F算法，求出的样本特征权重，记为
3. 将上述求得的特征权重相加并求取均值，即为最终特征权重***W***



以上，就完成了样本特征权重的更新，相对于Relief-F算法更新特征权重，改进后的算法考虑到了抽样的随机性，避免了Relief-F算法中每次可能会在同一类样本中抽样，降低了样本分布差异对特征权重计算的影响，计算出来的权重更具有合理性。

4.2.2 案例调整

将优化后的权值更新算法应用于案例推理的案例检索中，但是检索到的案例不一定能够完美解决当前问题，需要对检索到的最相似的案例的解决方案进行适当调整并评估。通常是相关人员根据领域知识来对检索到的案例进行相应的适配调整，调整后的案例需专家对其进行评价，如果评价复用效果理想，则将新案例方案投入使用并对案例进行保存和学习，案例的学习过程在4.1.4小节中已经详细介绍，这里不过多赘述；如果评价复用效果不理想，则需要重新进行案例调整。案例的调整、复用和评价是一个反复迭代的过程。

对于其他复杂案例的调整，通常与领域直接相关，通过启发式规则指导来完成，案例调整的启发式规则是CBR求解问题的重点和难点，关于案例调整策略，很多学者也对此有不同的观点[41][42],包括在CBR系统中一如KDD技术发现案例改写规则、CBR和RBR（基于规则推理）混合推理以及神经网络技术等。

4.2.3 案例库维护的改进方法

前面提到过，在案例推理系统中，系统的增量式学习会使得案例库不断扩大，但这个扩大并不是良性的。随着新案例的不断加入，案例库会变得十分庞大，这将导致案例的检索时间增加，影响案例推理系统的求解效率和性能。所以必须采取合理的办法对案例库进行维护使案例库维持在一定的规模，但是通常的案例库维护方法仅仅进行了案例的随机删除来减小案例库规模，这种删除方法无异于饮鸩止渴，表面上使案例库维持在一定的大小，实际上随机删除案例可能会将高性能的案例删除，导致系统解决问题的能力急剧下降，甚至还会导致一些问题没有相似案例而无法求解。

基于上述问题，本文采用了离群数据分析和聚类分析方法对案例库进行维护。首先对案例库全部案例进行离群分析，找出案例库中的离群案例，这些案例可能是无用的噪声案例，也可能是有用的特殊案例，需要专家通过领域知识确定。然后对案例库进行聚类，对每个小类中找到几个包含此类案例聚类属性的典型案例，组成该小类新的案例库，以此方法将所有小类整合在一起形成新的案例库，在新的案例库上进行案例检索和调整。

1. 离群数据分析

离群数据是与其他常规数据偏离，不满足一般数据特征，与其他数据不一致的数据。离群数据有可能是噪声数据，也有可能是包含极具价值的数据。离群数据分析就是将数据集中的异常数据提取出来，常用的方法有基于距离的方法和基于模型的方法和基于密度的方法。

本文使用的算法是孤立森林（Isolation Forest，IF），它是由周志华等人提出的一种无监督学习算法[43]，属于组合决策数模型，适用于高维数据集，并且被证明是一种非常有效的异常检测方法。

算法的思想如下：假设我们用一个随机超平面来切割数据空间,切一次可以生成两个子空间，然后再继续用一个随机超平面来切割每个子空间，循环下去，直到每子空间里面只有一个数据点为止。我们可以发现那些密度很高的簇是可以被切割很多次才会停止切割，但是那些密度很低的点很容易很早的就停到一个子空间了。

怎么来切割这个数据空间是iForest的设计核心思想，本文仅介绍最基本的方法。由于切割是随机的，所以需要用组合的方法来得到一个收敛值（蒙特卡洛方法），即反复从头开始切，然后平均每次切的结果。iForest 由*t*棵孤立树（Isolation Tree，iTree）组成，每个iTree是一个二叉树结构，其实现步骤如下：

1. 从训练数据集中随机选择个样本点作为树的根节点；
2. 随机指定一个特征维度，在当前节点数据中随机产生一个切割点*p*，*p*点产生于当前节点数据指定特征维度的最大值和最小值之间
3. 以切割点*p*生成一个超平面，将当前数据空间划分为2个子空间
4. 在子空间中递归步骤*b*和*c*，不断构造新的子节点，直到无法再分割为止。

获得*t*个孤立数之后，孤立森林训练结束，开始用生成的孤立森林评估测试数据。对于训练样本*x*，让它遍历每一颗孤立树，统计*x*落在每棵树的第几层，即*x*在树的高度。通过此方法我们可以得到*x*在每棵树的高度平均值，设置阈值，平均高度低于此阈值的数据即为异常数据。

如图4-3所示为4个测试样本*X*，*Y*，*Z*，*H*遍历产生一颗孤立树：



图4-3 4测试样本遍历一颗孤立树

可以看到，样本*X*和*Y*的高度是3，*Z*的高度是2，*H*的高度是1，因此样本*H*最有可能是离群点，因为它最早就被孤立（isolated）了。

1. 聚类分析

找到离群数据后，需要对案例库样本进行聚类，使得类内数据对象相似，类间数据对象相异，并选取每一聚类内的部分数据代表该聚类，以减小案例库规模。

常用的聚类算法有：

1. K-means聚类算法

K-means聚类算法[44]是机器学习中常用的聚类算法，它的优势在于只需要计算样本点和群中心点之间的距离就行了，速度非常快。但缺点也很明显，它需要知道有多少类，不同的参数*k*可能会产生不同的聚类结果。

1. 均值偏移聚类算法

均值偏移（Mean shift）聚类算法[45]是一种基于滑动窗口（sliding window）的算法，通过该方法可以找到密集的数据点。同时，它还是一种基于中心的算法，它的目标是定位每一类的中心点，通过更新中心点的候选点来实现滑动窗口中的点的平均值。这些候选窗口在后期处理阶段被过滤，以消除几乎重复的部分，形成最后一组中心点及其对应的组。

与K-means聚类算法相比，均值偏移算法不需要选择聚类的数量，但是缺点是对参数*r*选择窗口的大小的选取非常重要。

1. DBSCAN聚类算法

DBSCAN(Density-Based Spatial Clustering of Applications with Noise)[46]是一个比较有代表性的基于密度的聚类算法，类似于均值转移聚类算法，但它有几个显著的优点：不需要事先设定聚类数量，相比于均值偏移算法，它可以将异常值识别为噪声，此外该算法还可以很好的找到任意大小和任意形状的聚类。

1. 高斯混合模型聚类

高斯混合模型（Gaussian Mixture Model，GMM）[47]比K-Means更具灵活性。使用高斯混合模型，我们可以假设数据点是高斯分布的。这样，我们就有两个参数来描述聚类的形状：均值和标准差。以二维的例子为例，这意味着聚类可以采用任何形式的椭圆形状（因为在*x*和*y*方向上都有标准差），不仅限于园。因此，每个高斯分布可归属于一个单独的聚类。但它的缺点依然是需要自定义聚类参数*k*。

本文使用的是DBSCAN基于密度的聚类算法，算法的思想是：同一类别的样本之间是紧密相连的，即该类别样本周围一定存在同类样本。

输入：样本集，邻域参数，描述邻域样本分布的紧密程度。其中表示某一点的邻域距离半径，MinPts表示半径为的邻域中数据点个数的最小值。领域和核心对象的定义如下：

邻域：对于，邻域表示样本集D与的距离不大于的样本集，即，集合个数记为

核心对象：对于任意样本，如果，则是核心对象。

算法流程如下：

1. 初始化核心对象集合，初始化类别k=0;
2. 遍历D中的元素，如果是核心对象，则将其加入到集合中;
3. 如果集合中全部元素都已被访问，则结束算法，否则进入下一步;
4. 在核心对象集合中随机选择未访问的核心对象，将标记为类别k并标记为已访问，将的邻域中未访问的数据存入集合S中;
5. 如果集合S为空，则当前类生成完毕，k=k+1，返回第三步。否则从集合S中选择一个样本点，标记为类别k和已访问，判断是否为核心对象，如果是，将中未访问的点加入到S中，返回第5步。

经过上述步骤就可以得到所有分类样本簇

与传统的K-means聚类算法仅局限于凸样本的聚类相比，DBSCAN的优势在于对所有类型的样本均可聚类，不需要输入聚类数*k*，还可以找出样本中的异常点。

通过聚类分析，我们可以选取每一聚类簇内的部分典型代表数据对象来表示该聚类，实现案例库的精简，起到减少案例库数据规模，提高检索性能的作用。

4.3 本章小结

本章主要介绍了CBR的基本流程，包括案例的检索、案例的学习和评估以及案例库的维护方法。随后针对现有案例推理方法中存在的问题提出了改进方法，主要包括权值更新算法的改进，使用每一类样本中抽样代替整个案例库样本随机抽样，降低了样本分布差异对权值计算带来的影响；对于现有案例库维护方法存在的会降低案例推理系统性能的缺点，设计了离群点检测和聚类分析算法相结合的方法，使用孤立森林算法代替传统方法进行离群点检测，算法效率高并且适用于高维度数据。在聚类算法上，利用DBSCAN密度聚类算法解决了传统的K-means聚类算法无法处理非凸样本和需要确定聚类数的短板。对样本进行离群分析和聚类分析后，选取每一聚类簇内的部分典型代表数据对象来表示该聚类，实现案例库的精简，可以提高检索性能。

第五章 模型实验及评估

5.1 应用背景及数据描述

5.1.1 应用背景

由于案例已经存储在案例库系统中，因此发生公共安全类事件时，如何快速准确的在案例库中检索出与当前案例最相似的案例是应急指挥工作的核心内容。对于已经存储在案例库中的海量案例，如何通过有效手段提取案例中隐藏的有用信息，进而指导我们指定更合理的决策方法，也是体现消防实战指挥辅助决策系统性能的重要方面。

因此，本文通过从案例库中选取若干案例样本进行分析和挖掘，并将案例特征应用于案例库的检索中，以验证本文设计的系统的有效性，为避免检索到的案例与选取的案例相同，本实验将选取若干样本之后剩余的案例作为案例库。由于案例检索结果的准确度和合理性无法直接定性分析，必须投入到实际应用中，根据案件处置的结果反馈判断推理的合理性，因此本文在于将推荐方案与原方案进行对比分析，在此基础上确定推荐案例的准确性。

5.1.2 数据描述

目前，消防指挥中心数据主要来源于案例库各历史案例，本文所选的案例库来自某市消防局2018年的数据，经过筛选和清洗后剩余有效数据约60000条样本。收集的案件信息特征主要包括下面几个：

1) 案件类型：反恐排爆、火灾、救援抢险、社会救助和其他

2) 建筑结构：砖混结构、钢结构、钢筋混凝土结构、砖木结构、木结构和其他结构

3) 烟雾状况：有火光、有火星、有爆炸声、有烟、有浓烟、有烟有火、其他

4) 案件等级：0~5级，2+~5+级

5) 气象信息描述：风向、风力、温度、天晴等

6) 处置对象：大跨度厂房、在建工程、改建工程、脚手架、隧道、列车、汽车、救助人员、高空排险、放射性物质等

7) 楼房层数

8) 燃烧楼层

9) 人员被困数

10) 伤亡人数

数据相关属性与说明见表5-1

表5-1 案件特征字段

|  |  |  |  |
| --- | --- | --- | --- |
| **属性名称** | **属性说明** | **属性类型** | **备注** |
| 案件类型 | 5个类型 | Varchar |  |
| 案件等级 | 5个等级 | Varchar |  |
| 建筑结构 | 多种结构 | Varchar |  |
| 处置对象代码 | 多种类型 | Varchar |  |
| 烟雾情况 | 多种类型 | Varchar |  |
| 受伤人数 | 整数 | int |  |
| 燃烧楼层 | 整数 | int |  |
| 楼房层数 | 整数 | int |  |
| 气象信息描述 | 多种类型 | Varchar | 案发当天的天气情况 |
| 处置方案 | 多种类型 | Varchar | 消防车辆的调派决策 |
| 主要情况 | 文本信息 |  | 描述案件经过 |

5.2 案件处置

下面给出了案例库中3条不同类型的案例样本，其中案件类型是接警员根据当时报警人说明的信息，人工判断后填写上去的。在实际的接处警任务中，消防局每天要处理成千上万的消防案件，由于情况紧急且种类繁多，人工处理这些信息不仅效率低，还有可能错误的识别警种，导致做出错误的决策，所带来的后果是不可估量的。为此本文设计了基于Attention Bi-LSTM的文本分类模型，直接对报警人描述的文本信息即案件特征中的“主要情况”进行分类，不仅可以减少接警员工作量，提高工作效率，还能够避免上述错误判断带来的损失。案例原始样本数据特征如下：

案件1：案件等级：1，主要情况：报警人称上址积水，影响交通，防汛办无法处理，称交警已在现场，请民警到场处理。（电告消防），建筑结构：Nan，烟雾状况：Nan，处置对象代码：26090A，案件等类型：2，立案时间：2018/9/17 7:30:25，燃烧面积：0，气象信息描述：天气：小雨 温度：24℃ 风向：东南风 风力：0级，人员被困数：0，调派车辆方案：器材消防车（QC）

案件2：案件等级：2，主要情况：仓库着火，火势较大，建筑结构：Nan，烟雾状况：Nan，处置对象代码：12105A，案件等类型：1，立案时间：2018/9/17 7:15:31，燃烧面积：500，气象信息描述：天气：小雨 温度：24℃ 风向：东南风 风力：0级，人员被困数：0，调派车辆方案：通讯指挥消防车（TZ），抢险救援消防车（JY）\*2，压缩空气泡沫消防车（YS）\*2，泡沫消防车（PM）\*3

案件3：案件等级：1，主要情况：处理马蜂窝，建筑结构：Nan，烟雾状况：Nan，处置对象代码：50200，案件等类型：1，立案时间：2018/9/16 14:16:21，燃烧面积：0，气象信息描述：天气：多云 温度：26℃ 风向：东南风 风力：0级，人员被困数：0，调派车辆方案：泡沫消防车（PM）

5.2.1 案件分析与数据预处理

研究数据库我们发现，建筑结构和烟雾状况这两个特征为空，案例中并未给出具体的描述，为便于案例检索和匹配，这里我们将这两个字段删除，待后续数据库对这两个特征进行补充完善，再重新进行预处理。

对于数值类型的特征，根据数值大小划分为“多”“少”“高”“低”等，对于案发时间的处理，由于季节不同白天和夜间的时间也不同，根据季节性相关规律，这里将时间划分为白天和夜间，并使用Label Encoding方式进行编码。

数值划分规则如下：

1. 楼房层数：高（>=10），中（5-15），低（<=4）
2. 人员被困数：少（<=5）,中等（5-10），多（>=10）
3. 燃烧楼层：高（>=15），中（5-15），低（<=4）
4. 燃烧面积：小（<=50m2），中（50-1000m2），大（>=1000m2）

气象信息由于包含天气、温度、风向、风力等描述，这里仅选取天气信息并按照类别分为晴、雨、雾、雪、多云等，同样使用Label Encoding编码，处置对象编码也如此。

对于“主要情况”特征字段的处理，通过第三章设计的文本特征提取和分类模型进行分类，整个数据语料库中有反恐排爆、火灾、救援抢险、社会救助和其他5中案件类型，对样本案例中的文本信息进行分类并编码，处理后的数据示例如下表5-2，特征字段用首字母表示，分别为案件序号、案件等级、处置对象、案件类型、气象信息、燃烧楼层、燃烧面积、楼房层数、立案时间和调派车辆，下同。

表5-2 处理后的数据示例

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **AJXH** | **AJDJ** | **CZDX** | **AJLX** | **QXXX** | **RSLC** | **RSMJ** | **LFCS** | **LASJ** | **DPCL** |
| **0** | **1** | **51** | **2** | **2** | **0** | **0** | **0** | **0** | **QC\*1** |
| **1** | **2** | **21** | **1** | **2** | **0** | **1** | **0** | **0** | **TZ\*1,JY\*2,YS\*2,PM\*3** |
| **2** | **1** | **26** | **3** | **1** | **0** | **0** | **0** | **0** | **PM\*1** |

5.2.2 案例检索与匹配

使用改进的Relief-F权值更新算法，从案例库中随机抽取5次，计算各个属性的特征权重结果如下表5-3，最后一行为均值：

表5-3 特征属性权重

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **抽样次数** | **特征权重** | | | | | | | |
| **AJXH** | **AJDJ** | **CZDX** | **AJLX** | **QXXX** | **RSLC** | **RSMJ** | **LFCS** |
| **1** | **0.6522** | **0.2105** | **0.3952** | **0.0338** | **0.0357** | **0.2381** | **0.0707** | **0.0859** |
| **2** | **0.5556** | **0.1756** | **0.3179** | **0.0071** | **0.0238** | **0.1548** | **0.0341** | **0.1056** |
| **3** | **0.6152** | **0.2457** | **0.3592** | **0.0403** | **0.0238** | **0.2976** | **0.0539** | **0.0698** |
| **4** | **0.4751** | **0.2637** | **0.3786** | **0.0564** | **0.0357** | **0.2381** | **0.1016** | **0.1014** |
| **5** | **0.5326** | **0.2069** | **0.4019** | **0.0215** | **0.0238** | **0.3214** | **0.0292** | **0.0985** |
| **w** | **0.5661** | **0.2205** | **0.3706** | **0.0318** | **0.0286** | **0.2500** | **0.0579** | **0.0922** |

从表5-3可以看出，在案例的相似度检索中，案件等级、案件类型、处置对象和燃烧面积所占的权重较大，这与实际情况相符合。

由于案例推理的准确度跟案例库的大小有关，并且根据推理结果无法直接定量确定推理的准确度与合理性，必须经过案例修改投入实际应用中根据处置反馈结果判断推理的合理性，因此只能从原方案的基础上确定推荐方案的合理性。本文使用60条案例作为目标案例进行测试，并使用剩下的全部样本数据作为案例库案例（6万余条），分别使用基于改进的Relief-F的特征加权案例推理方法和传统特征加权的方法（权重全为1），计算目标案例和原案例的相似度，输出与目标案例最相似的原案例对应的派车方案，将推荐方派车案转换为矩阵，转换方法如下：

根据数据库中派车种类及数量，对车种排序，排名前七的车辆依次为泡沫消防车（PM）,压缩空气泡沫消防车（YS），抢险救援消防车（JY），器材消防车（QC），云梯消防车（YT），水罐消防车（SG）和通信指挥消防车（TZ），它们的总和占到了车辆总数的95%以上。

例如，推荐方案1的派车方案为：[泡沫消防车（PM），泡沫消防车（PM），抢险救援消防车（JY），压缩空气泡沫消防车，通信指挥消防车（TZ）]。则对应的解决方案矩阵为，以此类推。

测试样本原调派方案结果与使用权值更新算法的案例推理方法和传统的KNN检索策略的案例推理方法检索的结果见下表5-4，5-5，5-6，由于篇幅有限，这里仅统计其中10条样本。

表5-4 测试样本原调派方案

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **样本** | **调派方案** | | | | | | |
| **PM** | **YS** | **JY** | **QC** | **YT** | **SG** | **TZ** |
| **1** | **1** | **0** | **0** | **0** | **0** | **0** | **0** |
| **2** | **2** | **3** | **2** | **1** | **0** | **0** | **1** |
| **3** | **1** | **0** | **1** | **0** | **0** | **0** | **0** |
| **4** | **1** | **0** | **0** | **0** | **0** | **0** | **0** |
| **5** | **2** | **2** | **0** | **0** | **0** | **0** | **0** |
| **6** | **1** | **0** | **0** | **0** | **0** | **0** | **0** |
| **7** | **1** | **0** | **0** | **0** | **0** | **0** | **0** |
| **8** | **2** | **2** | **0** | **0** | **0** | **0** | **0** |
| **9** | **1** | **0** | **0** | **0** | **0** | **0** | **0** |
| **10** | **2** | **1** | **1** | **0** | **1** | **0** | **0** |

表5-5 基于改进权值更新算法的案例推理推荐的调派方案

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **样本** | **调派方案** | | | | | | |
| **PM** | **YS** | **JY** | **QC** | **YT** | **SG** | **TZ** |
| **1** | **1** | **0** | **0** | **0** | **0** | **0** | **0** |
| **2** | **6** | **1** | **1** | **0** | **1** | **0** | **1** |
| **3** | **1** | **0** | **1** | **0** | **0** | **0** | **0** |
| **4** | **0** | **1** | **0** | **0** | **0** | **0** | **0** |
| **5** | **3** | **1** | **0** | **0** | **0** | **0** | **0** |
| **6** | **1** | **0** | **0** | **0** | **0** | **0** | **0** |
| **7** | **1** | **1** | **0** | **0** | **0** | **0** | **0** |
| **8** | **2** | **1** | **0** | **0** | **0** | **0** | **0** |
| **9** | **1** | **0** | **1** | **0** | **0** | **0** | **0** |
| **10** | **3** | **1** | **1** | **0** | **1** | **0** | **0** |

表5-6 传统案例推理推荐的调派方案

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **样本** | **调派方案** | | | | | | |
| **PM** | **YS** | **JY** | **QC** | **YT** | **SG** | **TZ** |
| **1** | **0** | **1** | **0** | **0** | **0** | **0** | **0** |
| **2** | **6** | **0** | **0** | **0** | **1** | **0** | **0** |
| **3** | **0** | **1** | **1** | **0** | **0** | **0** | **0** |
| **4** | **0** | **1** | **0** | **0** | **0** | **0** | **0** |
| **5** | **3** | **0** | **0** | **0** | **1** | **0** | **0** |
| **6** | **1** | **0** | **0** | **0** | **0** | **0** | **0** |
| **7** | **1** | **1** | **0** | **0** | **0** | **0** | **0** |
| **8** | **1** | **2** | **0** | **0** | **0** | **0** | **0** |
| **9** | **0** | **0** | **1** | **0** | **0** | **0** | **0** |
| **10** | **3** | **1** | **1** | **0** | **0** | **0** | **0** |

可以看出，两种方法计算出来的推荐调派方案存在着一定的差异，尤其是在派车比较多的情况下，这种差异愈发明显。

消防救援事关重大，在抢险救援工作中，多派一辆车会造成资源浪费，但是少派一辆车可能会导致灭火效果不佳，救援不到位等恶劣后果，甚至还会导致重大财产损失及人员伤亡，因此对于车辆调派方案的研究，即使在性能上有一点点的提升，由此带来的价值却是不可估量的。为更好地比较两种方法的性能，我们推荐案例的解决方案与原测试样本的派车方案进行对比打分，方法如下：

对于多派的车辆，会造成资源浪费，每多派一辆，扣分；对于少派的车辆，可能会造成灭火效果不理想甚至财产损失及人员伤亡，因此每少一辆，扣分。分别将原方案*Y*和特征加权计算推荐方案、传统案例推理方法推荐的方案做差值，则总扣分数



这里我们设=2，=10，总分为100分，扣分之后各样本得分情况如下表5-8：

表5-7 两种方法得分

|  |  |  |
| --- | --- | --- |
| **样本** | **特征加权+CBR分数** | **传统CBR分数** |
| 1 | 100 | 88 |
| 2 | 50 | 20 |
| 3 | 100 | 88 |
| 4 | 88 | 88 |
| 5 | 88 | 76 |
| 6 | 100 | 100 |
| 7 | 98 | 98 |
| 8 | 90 | 90 |
| 9 | 98 | 88 |
| 10 | 98 | 88 |

可以看出，特征加权算法推荐方案的得分普遍高于传统CBR的得分，即基于Relief-F权值更新算法的特征加权案例推理系统更优于传统案例推理系统。在推荐案例的合理性研究中，我们还可以定义得分阈值，例如本文中我们定义得分阈值=90，即得分大于等于90才被认为是合理的推荐方案。

5.3 应急方案及评估

从以上实验结果可以看出，本文设计的Attention Bi-LSTM消防文本分类模型结合改进的Relief-F特征加权案例推理模型检索出的决策方案相对于传统方法在原样本方案上对比更佳，本文设计的CBR系统推荐的解决方案更接近于真实情况。由于本实验选取的案例库较小，系统中样本数据不够丰富，数据结构不完整，并且案例种类较少，这些缺点一定程度上会造成系统经验不足，在案例检索与匹配中可能会找不到合适案例的情况，导致某些案例无法找到合适的解，例如本实验中的样本2，无论是传统案例推理方法还是本文设计的案例推理方法，推荐出的解决方案都不合适。另外，本文设计的扣分表，仅是理论上的，具体数值设计需要在应用实战中结合消防局指标要求进行测试，收集更多的测试样本进行分析，确定扣分参数。

5.4 本章小结

本章通过选取案例库中部分样本案例，对消防实战指挥智能辅助决策系统设计的合理性与有效性进行了验证。对本课题主要研究的两个关键技术：Attention Bi-LSTM消防文本分类模型和基于改进的Relief-F权值更新算法的案例推理技术的检索和匹配进行了详细阐述，文本分类模型从源头上将报警人提供的信息自动进行警种识别，能够减小接警员的工作量，提高工作效率，并将分类结果作为案例特征输入检索系统中；案例推理系统对案例库样本进行分类随机抽样计算特征权重，并应用于案例的相似度检索上。在实验对比上，由于使用案例推理方法推荐的方案需要投入实际应用并根据反馈结果才能判断推荐方案的合理性，实验仅从测试样本的原方案和两种推荐方案进行对比，并结合实际情况设计了案例合理性的评估方案，实验结果表明本文设计的系统相比于传统方法效果更佳。本文设计的系统只是从理论上证明比传统系统更优，在今后的研究中，作者将在实际应用中收集更多的实际应用的数据来验证和提高系统的性能。

同时，实验结果还表明，若案例库案例不足，缺乏多样性，也会导致案例推理系统性能不理想，某些新发生的案件在案例库中找不到适合的解决方法。这也说明，案例推理系统只起到了一个辅助决策的作用，为决策者在应急决策中提供参考方向，并不能代替最终实施方案。在消防救援工作中，当发生突发事件时，消防人员就可以结合辅助决策系统的推荐方案进行决策的调整和适配，大大减少决策时间，提高决策效率，为消防救援工作争取宝贵的救援时间。

第六章 总结及展望

6.1 全文总结

城市消防指挥决策是一个新兴领域，研究者对于这方面的研究工作还很浅显，而根据调查显示，我国每年因城市火灾造成了大量的人员伤亡和财产损失，因此对于消防指挥智能决策系统的研究具有重大的现实意义和理论价值的。文本在前人研究的基础上，根据消防实战指挥智能决策系统的特点和需求，提出了一套消防实战指挥智能辅助决策系统，应用基于Attention Bi-LSTM文本分类模型对报警人提供的信息进行分类，提高接警员工作效率，同时将分类结果应用于案例推理中，接下来在案例推理模块中，提出了一种改进特征加权算法，并通过实验证明该方法相对于传统方法检索效果更好。本文应用基于特征加权的案例推理方法检索相似案例的解决方案，应用于新案例的辅助决策中，为消防救援提供了一套合理的辅助决策系统，对决策者在消防抢险救援决策中提供了有力的支撑。

本文主要在以下方面进行了创新：

1. 在详细分析消防指挥辅助决策需求的基础上，结合深度学习模型设计了一种基于Attention Bi-LSTM的消防文本分类模型；
2. 对现有Relief-F权值更新算法做出了改进，降低了样本分布差异对特征权值的影响，改进算法求出的权重矩阵更具有说服性；
3. 将案例推理方法应用于消防救援中车辆的调派上，并将特征加权算法应用于案例推理的相似度检索中，实验表明，相对于传统的案例推理方法，检索性能更好。
4. 对于案例库的维护，提出并设计了离群分析和聚类分析的方法。

6.2 工作展望

消防实战指挥智能辅助决策系统功能复杂，内容较多。这样一个庞大的决策支持系统作者不可能在毕业论文设计的算时间内完成从开发、设计到实现的全部工作，只是针对上述几个问题进行了初步的技理论性技术探讨。该系统并不是完美的，并没有投入实际使用中去，许多问题还有待发现，需要在实际使用中进一步完善和研究。因此，作者今后的研究主要包括：

1. 对报警人提供的文本信息详细分析，如命名实体识别，为案例分析提供更多的特征；
2. 研究该领域相关技术的发展趋势，优化案例检索和相似度匹配算法问题，本文使用的是特征加权算法，这只是对消防案例检索的初步尝试，更好的方法有待研究；
3. 丰富案例库案例类别。CBR系统在解决非结构化、知识获取困难等复杂环境下的决策问题具有很大的优势，这一切的前提都在与案例库的多样性，只有能够从案例库中检索到相似案例，才能够进行下一步工作；
4. 将神经网络模型与案例推理相结合，设计新的应急决策指挥系统，提高决策效率和性能。

参考文献

1. Andersson Granberg T , Lundberg J , Ulander A , et al. Supporting dispatch decisions for the fire and rescue services[C]// IEEE International Conference on Intelligent Transportation Systems. IEEE, 2015.
2. 李亚东, 夏耘, 宋婷婷. 大数据助推人工智能在公安指挥调度上的应用[J]. 数字通信世界, 2017(9):24-26.
3. 刘浠. 面向城市应急指挥的智能决策系统的研究与实现[D]. 广东工业大学, 2007.
4. 彭桥. 人工智能技术在消防接处警系统中的应用研究[J]. 自动化与仪器仪表, 2018(7).
5. 孔凡全, 杜光, 田赫. 消防调度指挥工作发展方向探索——基于机器学习技术的消防调度指挥解决方案研究[C]// 2017中国消防协会科学技术年会论文集. 2017.
6. 李卫红, 童昊昕. 针对非平衡警情数据改进的K-Means-Boosting-BP模型[J]. 中国图象图形学报, 2017(9).
7. Kohonen T . An introduction to neural computing[J]. Neural Networks, 1988, 1(1):3-16.
8. Lecun Y , Boser B , Denker J S , et al. Backpropagation Applied to Handwritten Zip Code Recognition[J]. Neural Computation, 1989, 1(4):541-551.
9. Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[C]// NIPS. Curran Associates Inc. 2012.
10. Lawrence S , Member, IEEE, et al. Face Recognition: A Convolutional Neural-Network Approach[J]. IEEE Transactions on Neural Networks, 2002, 8(1):98-113.
11. Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211
12. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE transactions on neural networks, 1994, 5(2): 157-166.
13. Hochreiter S, Schmidhuber J. Long short-term memory[M]// Supervised Sequence Labelling with Recurrent Neural Networks. Springer Berlin Heidelberg, 1997:1735-1780.
14. Werbos P J. Backpropagation through time: what it does and how to do it[J]. Proceedings of the IEEE, 1990, 78(10): 1550-1560
15. Boden M. A guide to recurrent neural networks and backpropagation[J]. the Dallas project, 2002.
16. Bengio Y, Ducharme R, jean, et al. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, 3(6):1137-1155
17. Mikolov T, Kombrink S, Burget L, et al. Extensions of recurrent neural network language model[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE,2011:5528-5531.
18. Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model[C]// INTERSPEECH 2010, Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September. DBLP, 2010:1045-1048.
19. Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv:1409.2329, 2014
20. Joachims T. Text Categorization with Suport Vector Machines: Learning with Many Relevant Features[C]// European Conference on Machine Learning. Springer, Berlin,Heidelberg, 1998:137-142.
21. Lai S, Liu K, He S, et al. How to Generate a Good Word Embedding[J]. IEEE Intelligent Systems, 2015, 31(6):5-14.
22. Turian J, Ratinov L, Bengio Y. Word representations: a simple and general method for semi-supervised learning[C]// ACL 2010, Proceedings of the, Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden. DBLP, 2010:384-394.
23. Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in neural information processing systems. 2013: 3111-3119.
24. Morin F, Bengio Y. Hierarchical Probabilistic Neural Network Language Model[C]//Aistats. 2005, 5: 246-252.
25. Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3(Jan): 993-1022.
26. Wallach H M. Topic modeling: beyond bag-of-words[C]//Proceedings of the 23rd international conference on Machine learning. ACM, 2006: 977-984.
27. Le Q, Mikolov T. Distributed representations of sentences and documents[C]//International Conference on Machine Learning. 2014: 1188-1196.
28. Joulin A, Grave E, Bojanowski P, et al. Bag of Tricks for Efficient Text Classification[J]. 2016:427-431.
29. Graves A , Jaitly N , Mohamed A R . Hybrid speech recognition with Deep Bidirectional LSTM[C]// Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on. IEEE, 2013.
30. Ghorpade T, Ragha L. Featured based sentiment classification for hotel reviews using NLP and Bayesian classification[C]// International Conference on Communication. 2012.
31. MORAES, Rodrigo, Valiati F, et al. Document-level sentiment classification: An empirical comparison between SVM and ANN[J]. Expert Systems with Applications, 2013, 40(2):621-633.
32. Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014
33. Kolodner J L . An introduction to case-based reasoning[J]. Artificial Intelligence Review, 1992, 6(1):3-34.
34. Yao Y, Huang Z. Bi-directional LSTM Recurrent Neural Network for Chinese Word Segmentation[C]// International Conference on Neural Information Processing. Springer, Cham,2016:345-353.
35. Aamodt, Agnar, Plaza, et al. Case-based reasoning: foundational issues, methodological variations, and system approaches[J]. Ai Communications, 1994, 7(1):39-59.
36. 杨善林, 倪志伟. 机器学习与智能决策支持系统[M]. 科学出版社, 2004.
37. 刘健. 基于案例推理的知识系统的设计与实现[D]. 南京航空航天大学, 2004.
38. Kira K, Rendell L A. The feature selection problem: traditional methods and a new algorithm[C]// Proceedings of the tenth national conference on Artificial intelligenceAAAI Press, 1992:129-134.
39. Kononenko I. Estimating attributes: analysis and extensions of RELIEF[C]// Proceedings of the European conference on machine learning on Machine LearningSpringer-Verlag New York, Inc., 1994:171-182.
40. 杜渂, 侯忠辉, 刘亮亮. 基于特征加权FCM算法的消防火灾等级研究[J]. 网络安全技术与应用, 2016(4):46-48.
41. ANAND, S. S, HUGHES, et al. Hybrid Data Mining systems : The next generation[C]// Pacific-asia Conference on Research & Development in Knowledge Discovery & Data Mining. 1998.
42. 韩小妹, 韩景倜. 基于CBR应急保障物流体智能决策支持系统研究[J]. 计算机工程与应用, 2007, 43(20):204-206.
43. Liu F T, Kai M T, Zhou Z H. Isolation Forest[C]// Eighth IEEE International Conference on Data Mining. 2009.
44. Hartigan J A, Wong M A. Algorithm AS 136: A K-Means Clustering Algorithm[J]. Journal of the Royal Statistical Society, 1979, 28(1):100-108.
45. 康青杨, 吴尽昭, 沈强,等. 一种基于路径的快速均值偏移算法[J]. 计算机应用与软件, 2013(12):1-3.
46. Uncu O , Gruver W A , Kotak D B , et al. GRIDBSCAN: GRId density-based spatial clustering of applications with noise[C]// IEEE International Conference on Systems. IEEE, 2007.
47. Huang Y, Englehart K B, Hudgins B, et al. A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses[J]. IEEE Trans Biomed Eng, 2005, 52(11):1801-1811.

硕士期间参加的研究项目和发表论文及专利

致谢

本篇论文是在彭明喜导师和杜渂导师的亲切关怀和耐心指导下完成的，感谢二位导师近两年来对我在学习和科研工作中给予的支持和帮助。他们作为研究生导师，在学术研究和科研工作上有着高深的造诣，并且在教育学生方面，言传身教，让我受到了很大的启发。另外，还要感谢他们在我研究生阶段给了我很多机会参与众多科研项目，不仅使我在学术上有很大的进步，更使我积累了很多丰富的项目经验。再次感谢他们在这两年多的时间里给予我的关怀和帮助。

其次，要感谢与我一起学习，一起进步的学长和同学朋友们。感谢他们在学习工作中对我的照顾和帮助，在他们的指导和帮助下，我积累了丰富的项目经验，为以后的工作打下了基础。感谢他们这两年的陪伴，让我更加坚定了自己求学的方向。

最后，特别感谢我的父母和家人，感谢你们无私的奉献和默默付出，为我提供了良好的学习条件和无数的鼓励，感谢你们无微不至的关怀。