

# 群众问政留言及政府答复的文本挖掘分析

## ——基于自然语言处理技术

### 摘要

随着信息技术的发展,微博、微信、博客等平台以及政府设立的网络问政平台成为互联网时代群众问政留言和意见表达的重要载体。政府管理者能否及时可靠的给予反馈回复,这与他们是否真正理解公众所表达的信息息息相关。然而,互联网上与政务相关的非结构化文本数据集呈现井喷式增长,形成了海量的文本信息空间。庞大的信息量使得政府管理者全面了解用户信息、真正解决群众所关心的问题变得愈发困难。

借助于网络问政、自然语言处理以及文本挖掘技术,对群众问政留言及政府答复文本进行研究,可以智能的将群众留言分类、对热点话题进行搜索并对相关部门的答复意见进行评价,从而提升为民服务效率,促进经济社会的发展。

本文主要研究内容和成果如下:

(1) 建立群众留言分类模型。本文利用设计的朴素贝叶斯分类模型和卷积神经网络模型对预处理后的留言文本进行分类,通过比较得到一个更优的自动分类模型。能够实现将留言内容输入模型后,自动输出该留言内容所属的类别,从而将群众留言分派至相应的职能部门进行处理,提高政务处理的效率。

(2) 提取热点话题。本文先将留言按区域进行划分,然后利用 LDA 主题概率模型和网络语义图对各区域留言主题进行分析,若描述同一件事情的留言主题多次出现,则将其视为可能出现的热点问题并根据留言主题写出热点描述,然后利用正则表达式的寻找功能,将符合热点描述的留言提取,最后根据留言数、点赞数、反对数、持续时间计算出热度指数,并对热度指数进行排名。从而帮助政府工作者对群众集中反映的问题给予及时的回应。

(3) 答复意见质量的评价。本文将相关性、完整性和可解释性作为评价答复意见质量的三个指标。其中相关性用 Word2Vec 模型计算的余弦相似度乘以 100 作为答复意见的相关性评分。完整性根据政府相关文件结合答复意见详情制定了一套标准的网络留言答复格式以及评分规则,使用正则表达式将答复意见与答复标准进行比较,按照评分规则进行打分。可解释性是以答复意见详情的字数、是否引用法律条文、是否对问题的发生进行解释或采取什么措施进行处理三方面按照评分规则进行打分。然后利用层次分析法计算出这三个指标的权重,分别为: 0.581, 0.309, 0.11。最后将三个指标的评分乘以对应的权重后再累加的结果作为该答复意见的综合评分。

**关键词:** 卷积神经网络; 朴素贝叶斯文本分类模型; Word2vec 模型; LDA 主题模型; 层次分析法

# **The Text Mining Analysis of People's Political Message and Government's Reply**

## **--Based on Natural Language Processing Technology**

### **Abstract**

With the development of information technology, platforms such as weibo, WeChat and blog, as well as the network platform set up by the government, have become important carriers for people to make political comments and express their opinions in the Internet era. Whether government managers can provide timely and reliable feedback is closely related to whether they truly understand what the public is saying. However, the unstructured text data set related to government affairs on the Internet shows a blowout growth, forming a huge amount of text information space. The huge amount of information makes it more and more difficult for government managers to fully understand user information and truly solve the problems concerned by the masses.

With the help of web politics, natural language processing, and text mining, to study the message of the masses and the text of government reply, It can intelligently classify people's comments, search for hot topics and evaluate the replies from relevant departments, So as to enhance the efficiency of serving the people and promote the development of the economy and society.

The main research contents and achievements of this paper are as follows:

- (1) Establish the classification model of public comments. This paper uses the designed naive bayesian classification model and convolutional neural network model to classify the pre-processed message text. Obtaining a better automatic classification model by comparison. After the message content is input into the model, the category of the message content is automatically output, so as to assign the public message to the corresponding functional departments for processing, and improve the efficiency of government affairs processing.
- (2) Extract hot topics. Firstly, dividing the message by regions. Then, using the LDA topic probability model and network semantic graph to extract the topic of the message content in each region .If the subject of a message describing the same thing appears multiple times, consider it as a possible hotspot issue and write a hotspot description based on the subject of the message, and then use the regular expression search function, extract messages that match the hotspot description. Finally, calculate the heat index based on the number of comments, likes, oppositions, and duration, and rank the heat index. In order to help government workers to respond to the problems of the masses in a timely manner.
- (3) Evaluation response comments on quality. Taking the correlation, completeness and interpretability as three indexes to evaluate the quality of response comments. Using Word2Vec model to calculate the cosine similarity. Cosine similarity times 100 as the correlation score of the response comments. According to the relevant government documents combined with the details of the comments make a set of standard network

message reply format and scoring rules, Then using regular expressions to compare response comments with response criteria. Finally, getting the score of completeness according to the scoring rules. Interpretability is scoring by the number of words in reply to the details of the opinion, whether to quote legal provisions, whether to explain the occurrence of the problem or what measures to take to deal with. Then the using the analytic hierarchy process to calculate the weight of these three indicators, which are 0.581, 0.309 and 0.11 respectively. Finally, the score of the three indicators is multiplied by the corresponding weight and then the summation result is taken as the comprehensive score of the reply. The sum of the scores of the three indicators multiplied by the corresponding weights as a composite score of the responses

**Keywords:** Convolutional Neural Network; Naive Bayesian Text Classification Model; Word2vec Model; LDA Thematic Model; Analytic Hierarchy Process

# 目 录

<b>1 引言</b>	<b>6</b>
1.1 挖掘背景	6
1.2 挖掘目的及意义	6
<b>2 问题分析</b>	<b>6</b>
2.1 问题一分析	6
2.2 问题二分析	8
2.3 问题三分析	9
<b>3 文本挖掘方案设计</b>	<b>9</b>
3.1 数据预处理	9
3.1.1 数据去重	9
3.1.2 分词	9
3.1.3 去停用词	10
3.1.4 Word2vec 模型	10
3.2 群众留言分类模型方案设计	11
3.2.1 朴素贝叶斯文本分类模型	11
3.2.2 卷积神经网络模型（CNN）	12
3.2.3 模型评价指标	14
3.2.4 模型选择	14
3.3 热点问题挖掘方案设计	14
3.3.1 发现热点	15
3.3.2 提取热点	15
3.3.3 热度指数	16
3.4 答复意见质量评价方案设计	16
3.4.1 相关性评分	16
3.4.2 完整性评分	17
3.4.3 可解释性评分	17
3.4.4 评价指标权重	18
3.4.5 综合得分	19
<b>4 实验及结果分析</b>	<b>20</b>

4.1 Word2vec 模型检验.....	20
4.1.1 相关性.....	20
4.1.2 类比性.....	20
4.2 群众留言分类模型实验及结果分析 .....	21
4.2.1 朴素贝叶斯分类模型评价.....	21
4.2.2 卷积神经网络分类模型评价.....	22
4.3 热点问题挖掘及结果分析 .....	23
4.3.1 LDA 主题模型结果及分析 .....	24
4.3.2 网络语义图展示及分析.....	24
4.3.2 热点展示及分析.....	25
4.4 答复意见质量评价及结果分析 .....	26
4.4.1 相关性评分展示.....	26
4.4.2 完整性评分展示.....	26
4.4.3 可解释性评分展示.....	27
4.4.4 综合评分展示.....	27
<b>5 总结与展望.....</b>	<b>28</b>
5.1 总结 .....	28
5.2 未来展望 .....	28
5.2.1 改进使用工具.....	28
5.2.2 使用深层神经网络.....	28
5.2.3 全自动化提取热点.....	29
5.2.4 答复意见评价优化.....	29
<b>致谢.....</b>	<b>30</b>
<b>Acknowledgement .....</b>	<b>30</b>
<b>参考文献.....</b>	<b>31</b>

# 1 引言

## 1.1 挖掘背景

近年来,随着物联网、云计算、大数据等信息技术的兴起与发展,政府的运作方式和创新模式发生转变。现今,国务院的各个部门与地方各级政府都不同程度的推行了政务公开。自 2006 年以来,全国有近 90%的县级以上政府或政府部门建立了政府网站,大多数的县级以上地方政府和中央国家机关建立了行政服务中心。目前,全国已建立综合性行政服务中心近 2.8 万个。由于传统城市和政府是按业务、管理职责分别设定的,各个部门各司其职,存在严重的部门壁垒,城市基本运行数据孤立的存在于不同的“烟囱”中。从而造成传统电子政务信息孤岛、数字鸿沟、网络安全威胁、互联网治理水平低以及行政效率不高等问题,正是因为传统的电子政务不足以应对新的运作方式。所以,更为先进的政务系统应运而生。

智慧政务系统的应用为上述问题提供了解决方案。智慧政务系统是利用物联网、云计算、移动互联网、人工智能、数据挖掘和知识管理等技术,提高相关机构办公、监管、服务、决策的智能化水平,形成高效、敏捷和便民的新型政务系统。目的就是要实现上述职能的数字化、网络化、智能化、精细化和社会化。而构建智慧政务系统的核心技术之一是自然语言处理,自然语言处理是计算机对人类所使用的口头或书面形式的自然语言进行各种处理、加工的技术,这种技术能够将自然语言翻译为计算机语言,以便计算机更容易地接受和执行人类以自然语言形式输入的指令,是发展人工智能技术的基础。

## 1.2 挖掘目的及意义

基于对智慧政务系统和自然语言处理技术的认识,本文将自然语言处理技术应用用于处理公众在各类问政平台反应的社情民意,智能的将群众留言分类、搜索热点话题并对相关部门的答复意见进行评价。首先利用词袋模型、TF-IDF 算法<sup>错误!未找到引用源。</sup>、Word2vec 模型、朴素贝叶斯文本分类模型和卷积神经网络(CNN)对群众留言文本进行分类;然后利用 LDA 主题概率模型<sup>错误!未找到引用源。</sup>和网络语义图对群众反应的热点话题进行挖掘,最后利用层次分析法从相关性、可解释性和规范性的角度对相关部门答复意见的质量进行综合评价。研究内容在一定程度上能给相关部门的工作带来便利,从而提升为民服务效率,促进经济社会的发展。

# 2 问题分析

## 2.1 问题一分析

根据群众留言建立自动分类模型。首先对留言文本进行预处理,进而得到各一级

标签频数分布情况如表 1 所示。

表 1 一级标签频数分布表（单位：条）

城乡 建设	劳动和社会 保障	教育 文体	商贸 旅游	环境 保护	卫生 计生	交通 运输
1989	1961	1566	1159	921	874	598

由表 1 可知，各一级标签类别数存在着明显的差异，其中留言数量最多的“城乡建设”与留言数量最少的“交通运输”类相差了三倍之多，可见数据具有不平衡性。为解决这一问题，可以考虑将数据集划分为多个二分类任务，或者一个二分类加两个多分类的形式。经过试验，将较好区分的“城乡建设”、“教育文体”、“卫生计生”三个一级分类对应的数据归纳为种类 1，剩余四个一级分类对应的数据归为种类 2。虽然后期在单独对种类 1 或种类 2 内的一级指标时进行测试时， $F1-Score$  会有明显提高（超过 96%），但是在划分为种类 1 还是种类 2 时的准确率约为 91%，各一级分类留言数都有损失，因而后期的  $F1-Score$  虽然高但一个二分类加两个多分类的真实效果并无明显提高，所以最后决定利用一个多分类进行建模。

文章尝试了多个基于传统机器学习和基于深度学习的循环神经网络（RNN）和卷积神经网络模型（CNN），最终选择了效率更高的朴素贝叶斯分类器和卷积神经网络模型。通过实验，当各一级标签类别数均超过 200 之后，使用 CNN 模型进行分类效率会明显高于朴素贝叶斯文本分类模型。

将分类语料进行分词和去停用词后，对各语料长度进行统计得到图 1 所示的预料长度频数分布图。其中，预料平均词长约为 128，大部分语料词长小于 200。因而考虑将 128 或 200 作为 CNN 建模时的长度参数。经进一步实验，发现以 200 作为建模长度参数准确率并无明显提高，反而需花费更长的时间，所以最终决定使用 128 作为 CNN 建模时的长度参数。

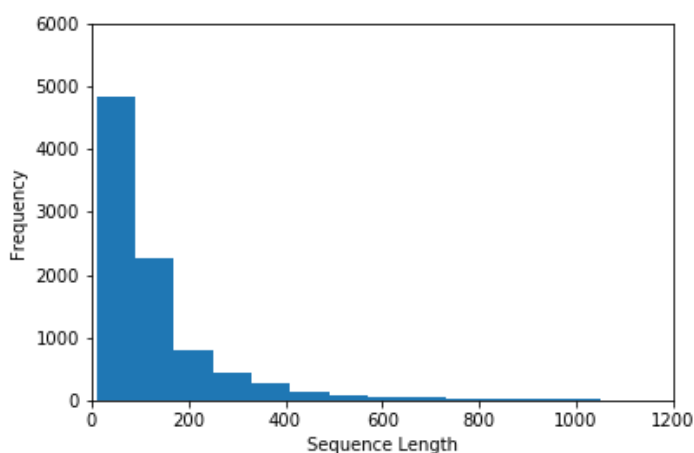


图 1 预处理后语料长度分布

2.2 问题二分析

提取留言中的热点问题。对附件 3 中的 4326 条留言数据进行统计分析，可得如图 2 所示的留言时间频数分布图和图 3 所示的区域留言数分布图。在时间上，留言主要分布在 2019 年 1 月至 2020 年 1 月；在区域上，留言主要为西地省 A 市的留言，且可将留言进一步细分到 A 市中的 9 个区（县）。

所以，本文考虑按月份或按地区提取热点问题，但考虑到持续时间较长的热点问题若按单月进行提取，该热点问题很有可能会被忽略。经过测试发现，按区域提取热点问题的效果还算理想，故最终选择按区域划分后再对数据进行挖掘，从而提取区域内的热点话题。

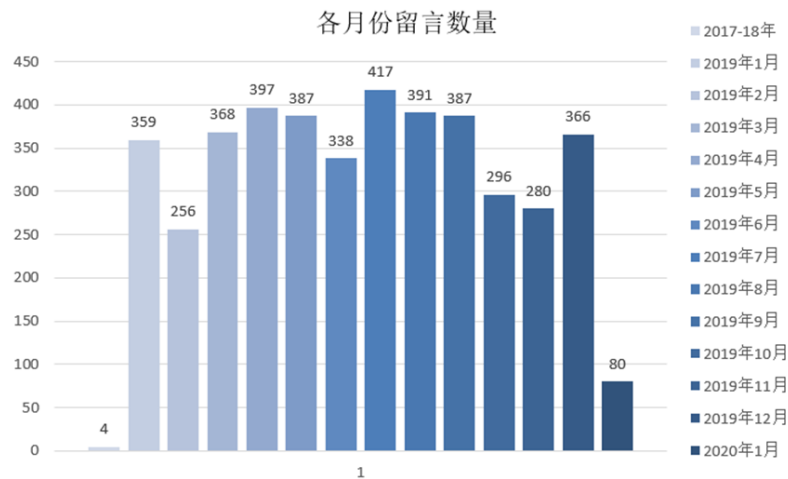


图 2 各月份留言频数图

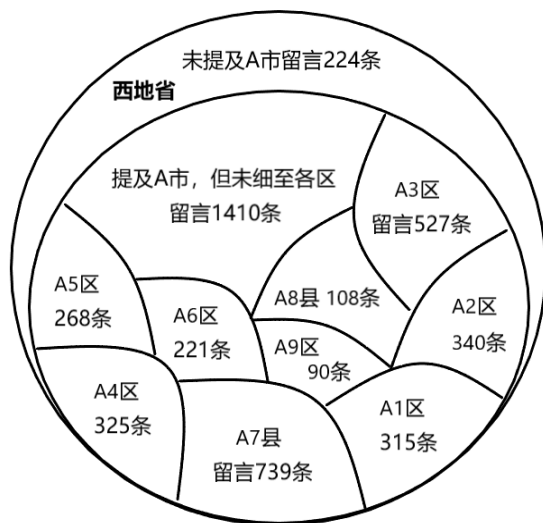


图 3 各区域留言数分布图



## 2.3 问题三分析

从相关性、完整性和可解释性等角度对答复意见的质量给出一套评价方案。

根据题目要求,并结合答复意见详情,本文首先将相关性、完整性和可解释性作为评价答复意见的指标,并进行量化,量化方法如下。

相关性是指答复意见的内容是否和留言内容相匹配。留言内容主要是提出问题,答复意见主要是谈论如何解决这个问题。若两者相关,则两个文本中的关键词词义基本相似,故对相关文本进行预处理并得到词向量后,通过计算留言内容和答复意见词向量的余弦相似度来衡量答复意见与留言内容之间的相关性。

完整性是指答复意见是否具有一定的规范性。本文通过陕西省生态环境厅参考《关于规范政府信息依申请公开答复工作的通知》<sup>[4]</sup>文件和江西省抚州市临川区人民政府《政府信息公开申请答复规范文本》<sup>[5]</sup>文件,并根据题目所给的答复意见详情进行适当的调整,制定了一套答复意见文本格式作为规范文本,进而根据答复意见符合规范文本的程度进行自动打分,根据得分来衡量答复意见与留言内容之间的完整性。

可解释性是指答复意见中是否引用相关的理论或者法律条文、行为规范、意见政策等对该答复进行解释。且认为答复意见的解释内容通常会占有一定的字数,故本文以答复意见中的正文内容的文本长度作为评分基础,以是否引用法律条文、行为规范、意见政策、提及解决方案等作为加分项来进行评分,根据得分来衡量答复意见与留言内容之间的可解释性。

然后,利用层次分析法的思想计算三个指标的权重,最后将指标值与对应的权重加权求和计算出综合得分,以此来衡量答复意见的质量。

## 3 文本挖掘方案设计

### 3.1 数据预处理

#### 3.1.1 数据去重

由于重复数据会对文本分类的精度造成影响,所以在分析之前需对留言文本进行去重处理。附件2中共有9210条记录,本文将留言用户、留言时间和留言详情相同的数据归为重复数据,按照这样的规则将重复数据予以剔除后得9068条记录。

#### 3.1.2 分词

中文分词是指利用分词算法将一个汉字序列切分成一个一个单独的词,使得该段汉字序列能被电脑自动识别语义。分词是文本挖掘的基础,直接影响挖掘的效率。常见的分词工具有jieba分词、pkuseg分词、哈工大语言云、清华THULAC等<sup>[6]</sup>。表

2 为 jieba 分词、pkuseg 分词、哈工大语言云分词的示例效果。

由表 2 可知，哈工大语言云分词效果最佳，但由于其商用性需要付费，难以普及。所以本文主要使用应用广泛的 jieba 分词工具进行分词，为了提高分词效果，在分词词典中加入政务领域的特有词汇，如：“A 市”、“西湖建筑集团”等。从而构成新的分词词典对文本进行分词。此外，在分词过程中，还动态添加特有词汇，以便实现更好的分词效果。

表 2 分词效果对比

示例句子：A 市西湖建筑集团占道施工有安全隐患	
jieba	A 市 西湖 建筑 集团 占 道 施工 有 安全隐患
pkuseg	A 市 西湖 建筑 集团 占道 施工 有 安全 隐患'
哈工大	A 市 西湖建筑集团 占道 施工 有 安全隐患

3.1.3 去停用词

对文本进行分词后的文本中常常会带有许多标点符号或没有具体意义的词，如“什么”，“一个”等，这些没有意义的词称为停用词。若保留停用词，容易对有效信息造成干扰<sup>[7]</sup>。本文以公开的中文停用词词典为基础，在分析过程中动态添加停用词，以便高效的得到较为规范化的分词文本。

3.1.4 Word2vec 模型

为从数字化的角度对留言文本进行分析，需要将文本进行词向量化<sup>[8]</sup>。本文利用 Word2vec 模型对文本进行词向量化处理，具体步骤如下。

(1) 语料准备

① 首先，将附件 2、附件 3、附件 4 中的“留言主题”，“留言详情”，“答复意见”三项内容合并，并保存 txt 文档；

② 然后，下载维基百科（wiki）公开数据中最新的中文语料。该网站提供的中文语料数据为未经处理的 xml 格式数据，要对其进行预处理才能正式使用。首先将 xml 格式数据转为 txt 格式，然后进行繁体转为简体、分词、去除非中文词操作后保存为 txt 文档；

③ 最后，利用网络爬虫，以 7 个一级分类为关键词，在微博平台上爬取相关数据，去重后保存为 txt 文档。微博爬取数据量表 3 所示。

表 3 微博爬取文本数量表（单位：条）

一级分类	环境 保护	教育 文体	城乡 建设	商贸 旅游	交通 运输	卫生 计生	劳动和 社会保障	合计
数量	694	687	685	682	666	634	613	4661

## (2) 模型训练

基于上述语料，使用 Python 中的 gensim 工具包中的 Word2vec 函数历时 6 个小时，训练出一个 50 维的词向量模型，并保存为 model 格式和 txt 格式两种格式。

## 3.2 群众留言分类模型方案设计

对问题一进行分析后，设计出群众留言分类模型。首先将留言主题和留言详情合并作为分类语料，将一级标签作为分类的标签。接着对数据进行去重、分词和去停用词的预处理操作预处理。然后将数据以 8: 2 的比例构建训练集与测试集，训练集中以 20% 的比例构建验证集。利用训练集建立多分类朴素贝叶斯模型和卷积神经网络模型，并利用验证集来确定模型复杂程度的参数。最后利用测试集对分类模型的效果进行评价。

### 3.2.1 朴素贝叶斯文本分类模型

#### (1) 理论介绍

朴素贝叶斯文本分类模型是一个简单、有效且实用性很强的文本分类模型<sup>[9]</sup>。

设有变量集  $U = \{A_1, A_2, \dots, A_n, C\}$ ，其中  $A_1, A_2, \dots, A_n$  是实例的属性变量， $C$  是取  $m$  个值的类变量。假设所有的属性都条件独立于类变量  $C$ ，即每一个属性变量都以类变量作为唯一的父结点，就得到朴素的贝叶斯分类器<sup>[10]</sup>。

使用朴素贝叶斯分类器进行分类的做法是通过概率计算，从待分类的实例的属性值  $a_1, a_2, \dots, a_n$  求出最可能的分类目标值。即计算各类  $c_i \in C$  对于这组属性的条件概率  $P(c_i | a_1, a_2, \dots, a_n)$ ，其中  $i = 1, 2, \dots, m$ ，输出条件概率最大的类标签作为目标值。

例如： $P(c_1 | M)$  为待分类文本  $M$  属于第一类文本的概率， $P(c_2 | M)$  为待分类文本  $M$  属于第二类文本的概率，只要  $P(c_1 | M) > P(c_2 | M)$ ，我们就可以判定邮件  $M$  属于第一类文本，反之则属于第二类文本。

#### (2) 模型设计

本文设计的朴素贝叶斯分类模型对留言详情进行分类的流程如图 4 所示，具体步骤如下。

- ① 输入附件 2 中预处理后的文本。
- ② 利用词袋模型对词向量进行计数处理，将留言文本转换为词频矩阵。
- ③ 计算词袋模型中词语的 TF-IDF 权重值，提取出最显著的 2500 个特征词作为特征向量。
- ④ 利用得到的训练样本的特征向量构建朴素贝叶斯模型对文本进行分类。
- ⑤ 对模型效果进行评估。



图 4 朴素贝叶斯分类模型流程图

### 3.2.2 卷积神经网络模型（CNN）

#### （1）模型介绍

卷积神经网络（CNN）一直广泛应用于图像处理领域，当 Yoon Kim 在 2014 年提出将 CNN 应用于文本分类<sup>[11]</sup>，之后 CNN 在自然语言处理方面得到进一步发展，主要应用包括情感分析、文本分类和问答系统等。一个通用的面向自然语言处理的 CNN—NLP 模型结构主要由嵌入层、卷积层、池化层、全连接层构成<sup>[12]</sup>，如图 5 所示。不同层相关定义以及功能如下。

##### ① 嵌入层

训练集中的每个单词可以由词向量表示，由所有单词向量组成嵌入矩阵  $M \in R_{n \times d}$ ，其中  $n$  表示词语的个数， $d$  表示单词向量的维度。训练经过数据预处理后表示成索引数字后，通过嵌入层映射将索引转化为词向量表示。假设单个样本经过数据预处理后由  $k$  个词语组成，那么对于每个词语，通过与嵌入矩阵  $M$  相乘，得到该词语对应的词向量表示。该样本词向量矩阵便可由  $k$  行  $d$  列的矩阵  $A \in R_{s \times d}$  表示<sup>[13]</sup>。

##### ② 卷积层

卷积层是 CNN 的重要组成部分，用于提取样本向量矩阵特征。在图像中，卷积核通常是对图像的一小块像素区域进行计算；在自然语言处理中，卷积核通常是对文本所构成的词向量进行计算。

##### ③ 池化层

池化层是降低卷积层的输出结果以及提取更深层次的特征表达。由于一个区域有用的特征极有可能在另一个区域同样适用，池化的过程实际上是对卷积后的某块特征区域求最大值或者求平均值的过程。

##### ④ 全连接层

全连接层（也叫前馈层），用来将最后的输出映射到线性可分的空间上，在整个卷积神经网络中充当“分类器”的作用。

在构建卷积神经网络模型时，本文采用 dropout 方法，按一定比例将隐藏层中的神经元权重设置不工作来减少模型的计算量，同时能在一定程度上避免模型出现过拟合现象。

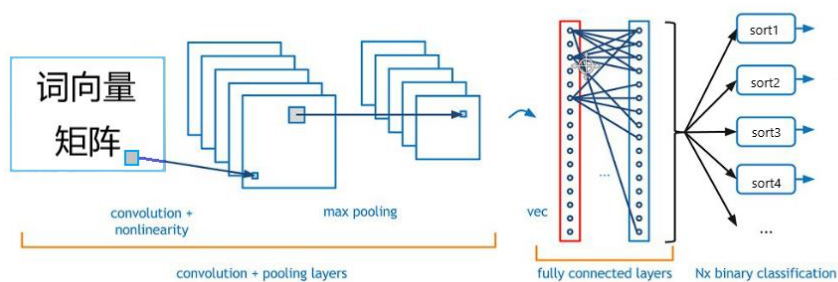


图 5 CNN 模型示意图

## (2) 模型设计

本文设计的 CNN 分类模型对留言详情进行分类的流程如图 6 所示，该模型分为 7 层，每层功能详细介绍如下。

① **Input:** 将附件 2 中分词和去停用词后的训练文本作为输入。

② **Embedding:** 对输入进行词嵌入，从而构建词向量矩阵。该层可以用 word2vec 模型预训练的词向量矩阵初始化 embeddings；也可以随机初始化 embeddings，然后基于语料通过模型网络对 embeddings 进行更新和学习。经过试验随机初始化 embeddings 的模型测试效果较为理想，因而这里选择随机初始化 embeddings。

③ **Conv3/Conv4/Conv5/Conv6:** 对样本向量矩阵进行卷积核大小为 3、4、5、6 的卷积操作，提取样本的特征。

④ **ReLU+Maxpooling:** 经过卷积之后的数据必须经过激活才能进行池化，以防止梯度消失等问题，本文使用 ReLU 函数作为激活函数，然后对其进行最大池化，降低数据维度。

⑤ **Concat:** 将经过池化后的数据以头尾相接的方式合并，拉成一条长长的特征向量。

⑥ **Dropout:** 训练时筛选出其中一定比例的神经元使它们停止工作，从而降低模型的计算，同时也一定程度上避免了模型在训练过程中出现的过拟合现象。

⑦ **Fully Conncted:** 将特征向量和各类别连接起来，得到最后的输出即为分类结果。

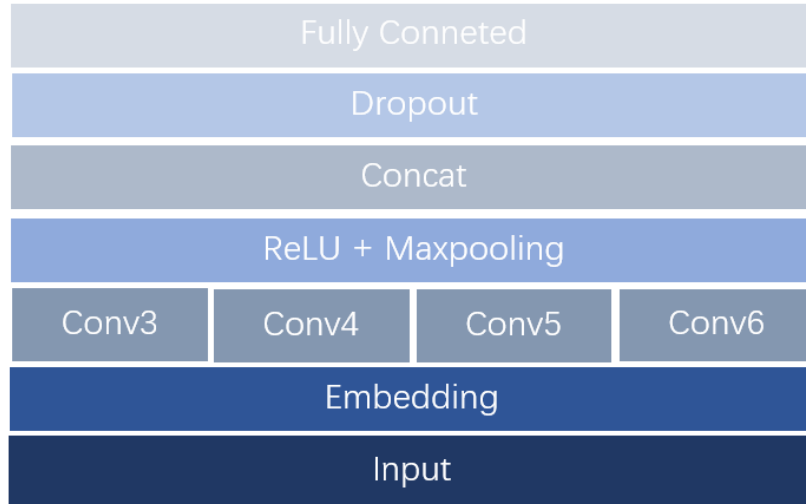


图 6 CNN 模型设计图

### 3.2.3 模型评价指标

本文主要采用准确率以及  $F1-Score$  得分对分类模型的效果进行评价。准确率（Accuracy）为正确分类样本数与总样本数之比，通常情况下可以有效的反映一个分类模型的好坏，但若遇到极不平衡的样本，效果就不是理想。

在准确率的基础上，为了更好的评价分类模型，引入基于查准率、查全率得到的  $F1-Score$  来评估模型的好坏。多分类  $F1-Score$  计算公式如式 1 所示。

$$F = \frac{1}{7} * \left( \sum_{i=1}^7 \frac{2 * P_i * R_i}{P_i + R_i} \right) = \frac{1}{7} \left( \sum_{i=1}^7 F_i \right), \text{ 其中 } F_i = \frac{2 * P_i * R_i}{P_i + R_i} \quad (1)$$

$P_i$  为第  $i$  类的查准率， $R_i$  为第  $i$  类的查全率。

### 3.2.4 模型选择

一般情况下，基于深度学习的模型会优于基于传统机器学习算法的模型，但深度学习的模型无法发现未知样本的特征，因而在数据量较少的情况下效率往往不如传统机器学习的算法。在示例数据中利用  $F1-Score$  对模型进行检验发现，朴素贝叶斯分类器会优于 CNN 模型，而在全部数据中，CNN 模型效率和精确度会更高。因而，在各一级标签类别较少的情况下建议使用朴素贝叶斯分类器；各一级标签类别数量均超过 200 后使用 CNN 模型。

## 3.3 热点问题挖掘方案设计

经过对问题二的分析，从发现热点、提取热点以及计算热度指数三个方面设计出热点问题挖掘方案。

### 3.3.1 发现热点

(1) 首先将附件 3 导入 Python, 将“留言主题”和“留言详情”两列数据合并作为语料, 利用正则表达式寻找功能, 以市、区、县、镇作为关键词寻找出所有多次出现的地名词 (市、区、县、镇级), 保存到自定义词典。再以街道、社区、小区、路等作为关键词寻找可能出现的路名、小区名等。由于与街道、社区等词语衔接的词语众多, 存在很多“假词”, 本文仅提取出现超过一定次数的词语, 例如: 长度为 2 的词语若超过 30 次才提取, 否则不提取。提取后需要人为判断是否为可能的地名词, 若判断为地名词则保存进自定义词典。

(2) 按照留言中提到的区域将留言划分为 A1~A9 区 (县)、A 市、非 A 市 11 个区域, 并将各区域的留言单独保存。

(3) 依次导入已按区域划分好的留言数据。使用 LDA 主题模型提取各区域“留言主题”的主题 (主题数根据留言数量大小决定), 打印各个主题中排名前十的关键词, 并基于“留言主题”制作网络语义图。然后根据主题关键词和网络语义图中的词语动态调整分词词典和停用词词典, 直至可大致判断各个主题的内容。

若该区域留言数量众多, 且直接使用 LDA 模型无法大致判断是否存在热点问题, 那么使用 Word2Vec 模型计算留言主题之间的余弦相似度, 并将余弦相似度大于 0.8 的留言主题聚为一类, 然后再对各类留言使用 LDA 模型和网络语义图进行分析。

(4) 根据各主题关键词在网络语义图上的关系, 若具有强关联性, 则利用正则表达式寻找功能, 打印前八条符合条件的“留言主题”, 若多条“留言主题”反映同一件事, 则将该事件归纳为可能的热点, 并大致写出热点描述。

(5) 多次出现的地名很有可能热点发生地, 点赞数众多的留言也有可能属于热点留言。若第 (1) 步中提取的地名词和点赞数超过 50 的留言未出现于可能热点中, 那么就利用正则表达式寻找功能, 打印相关留言, 若反映同一件事的留言多次出现, 则归纳为可能的热点, 并大致写出热点描述。

### 3.3.2 提取热点

得到热点问题以及热点的大致描述后, 利用正则表达式以各热点的发生地点和事件内容为关键词, 在附件 3 所有留言文本数据中进行查找, 若地点和事件都和关键词匹配上, 则判断留言属于该热点问题, 最后将该热点的所有留言在数据中索引出来, 并单独保存。

为防止有留言因描述不一致被忽略, 在使用正则表达式提取时采取多条件匹配发<sup>[14][15]</sup>, 例如: 使用“辉煌.{0,3}小区”作为关键词, 辉煌小区、辉煌国际小区、辉煌国际城小区都会被查找出来。

### 3.3.3 热度指数

某一时段内群众集中反映的某一问题称为热点问题，是群众对自己关心或与自身利益相关的各种事务在情绪、意愿、态度等方面的线上表达。并由群众关系延伸对线上留言的群众行为，例如对留言的点赞和反对等。通过热点指数，及时发现热点问题的热度，有助于相关部门及时处理民众最关心的事件，提高政府服务能力。

热点指数（ $S$ ）由留言数、点赞数、反对数的占比总和所得，基本公式为

$$S = \alpha \frac{\text{留言数}}{\text{总留言数}} + \beta \frac{\text{点赞数}}{\text{总点赞数}} + \gamma \frac{\text{反对数}}{\text{总反对数}} \quad (2)$$

其中  $\alpha, \beta, \gamma$  分别为留言数、点赞数、反对数的评分系数，且  $\alpha + \beta + \gamma = 1$ 。

$S$  越高，表明该条留言成为热点问题的概率越大，留言数反映了群众对某一问题的重视度，留言越多表明了问题事件的影响程度越大。点赞数和反对数可以反映群众对该留言反映问题的关注度，点赞数越高代表其他群众赞成问题事件的反映，反对数越高代表其他群众反对问题事件的反映。

热点问题事件会随时间的增加进而发酵，因此应该及时发现问题并进行解决<sup>[16]</sup>。为了满足时效性特征，使用幂函数对上述热点指数  $S$  值进行修正，得到公式

$$S_t = S \frac{1}{t^\varepsilon} \quad (3)$$

其中  $t$  为时间范围，能反映留言内容的新鲜程度； $\varepsilon$  为衰减因子，是大于 1 的正整数，它控制着  $S_t$  的衰减速度。

此时对所得的基础表进行热点数据分析，一般来说，留言操作比点赞和反对操作要较为繁琐和耗时，因而只有严重影响到自身利益才会使群众进行留言操作，因此赋予留言、点赞、反对不同的权重，其中  $\alpha = 0.4, \beta = 0.3, \gamma = 0.3, \varepsilon = 2$ 。

## 3.4 答复意见质量评价方案设计

经过对问题三的分析，设计出答复意见质量评价方案。首先需计算出相关性、完整性和可解释性这三个评价指标的得分并作为该指标的取值；然后计算出三个指标对应的权重；最后将指标得分与对应的权重进行加权求和得到综合得分，利用综合得分对答复意见质量进行评价。

### 3.4.1 相关性评分

本文利用余弦相似度对答复意见的相关性进行评价。首先将留言内容（留言主题加留言详情）和答复意见，进行预处理，然后利用 Word2vec 模型计算出留言内容的词向量及其对应的答复意见的词向量，分别将词向量标准化之后计算留言内容和



答复意见之间的相似度，值越靠近 1 代表该留言内容的答复意见的相似度越强。然后，将留言内容与答复意见的余弦相似度乘以 100 作为它们在相关性方面的得分（若低于 60 分，最终会返回 60 分），记为  $S_{\text{相关}}$ 。

### 3.4.2 完整性评分

经过前文分析，本文认为政府部门网络答复意见规范文本可用如下格式表示：

XXX（回复对象）：XXX（问候语）！您的 XXX 已收悉，XXX 回复如下：XXX  
（答复详情）XXX。XXX（感谢语），X 年 X 月 X 日（答复时间）。

如上所示，各下划线为一部分，可分为称呼、问候语、留言已读、答复详情、感谢语、答复时间 6 个部分。根据此格式规范，利用 python 中的 jieba 分词工具的词性标注功能以及正则表达式对答复意见进行分析处理。主要分为三个步骤。

（1）利用 jieba 分词工具对答复意见进行分词和词性标注处理。

（2）由于动词范围太广，用词性标记法无法高效率判断是否符合网络留言答复规范，所以利用正则表达式寻找每条答复开头部分是否存在“收悉”，“回复”等内容，寻找每条答复结尾部分是否存在感谢语。

（3）根据答复意见的完整程度进行自动评分并保存。

完整性评分规则：满分一百。若答复详情长度低于或等于 15 字符，为 60 分，不再处理；若答复详情长度高于 15 字符，低于等于 50 字符，为 70 分，不再处理；高于 50 字符的答复详情基础分为 70 分。另根据答复意见在称呼、问候语、留言已读、感谢语、答复时间上是否满足网络留言规范进行加分，若满足其中一项规范则在基础分上加 6 分，例如：答复详情开头文本词性中带有“I”（习语词，例如：您好），则在基础分上加 6 分。将完整性得分记为  $S_{\text{完整}}$ 。

### 3.4.3 可解释性评分

本文主要以答复意见中的正文内容的文本长度作为可解性的评分基础，另附加两个方面的加分项：一是判断答复意见中是否提到过的法律法规、政策条文等，若提到则加分；二是若答复意见中按条理回答，或提到为什么会出现这个问题，抑或将来的会采取什么行动去解决问题等，都对其进行加分。实现步骤如下。

（1）使用正则表达式对答复意见进行简单的处理——剔除答复意见开头结尾未涉及答复详情的内容。

（2）计算经过简单处理后的答复意见的字符串长度，根据长度按照制定好的规则给予基础分。

（3）使用正则表达式判断是否引用了法律法规、政策条文等，若有法律法规等出现，按照制定好的规则加分。

（4）用正则表达式判断是否按照条理进行回答、是否出现因果关系的连词（出现的

问题和发生原因是因果关系)、是否出现“计划”,“采取...措施”等表明怎么解决的词语,若其中一部分存在,则按照制定好的规则加分。

评分规则:满分一百。基础分最低为 60 分(50 字符以下),按照字数每增加 50 字符增加 5 分,最高为 80 分(大于或等于 200 字符)。加分项每项 10 分,若满足其中几个条件即可加分。例如:答复意见经步骤(1)处理后长度为 178 字符,引用了法律并且按照条理进行回答,那么该答复意见的得分为:75+10+10=95 分。将可解释性得分记为  $S_{\text{可解释性}}$ 。

### 3.4.4 评价指标权重

#### (1) 构造判断矩阵

利用表 4 所示的方法构造判断矩阵<sup>[18]</sup>。对多人进行简单采访并参考多方面资料,认为相关性比可解释性略微重要,比完整性明显重要,可解释性比完整性稍微重要,则判断矩阵为:

$$A = \begin{pmatrix} 1 & 2 & 5 \\ 1/2 & 1 & 3 \\ 1/5 & 1/3 & 1 \end{pmatrix} \quad (4)$$

表 4 标度法

重要性标度	含义
1	表示两个元素相比,具有同等重要性
3	表示两个元素相比,前者比后者稍微重要
5	表示两个元素相比,前者比后者明显重要
7	表示两个元素相比,前者比后者强烈重要
9	表示两个元素相比,前者比后者极端重要
2, 4, 6, 8	表示上诉判断的中间值
以上各数的倒数	两个元素反过来比较

#### (2) 根据判断矩阵确定权重

利用规范列平均法求权重,首先对列向量归一化得到对应的矩阵  $B$ :

$$B = \begin{pmatrix} 0.5880 & 0.6 & 0.5556 \\ 0.2941 & 0.3 & 0.3333 \\ 0.1176 & 0.1 & 0.1111 \end{pmatrix} \quad (5)$$

对矩阵  $B$  按行相加,并归一化,得到所求特征向量  $W$ :

$$W = (0.581 \quad 0.309 \quad 0.110)^T \quad (6)$$

计算最大特征根:

$$AW = \begin{pmatrix} 1 & 2 & 5 \\ 1/2 & 1 & 3 \\ 1/5 & 1/3 & 1 \end{pmatrix} * \begin{pmatrix} 0.581 \\ 0.309 \\ 0.110 \end{pmatrix} = \begin{pmatrix} 1.747 \\ 0.929 \\ 0.329 \end{pmatrix} \quad (7)$$

$$\lambda_{\max} = \sum_{i=1}^3 \frac{(AW)_i}{n * W_i} = 3.0014 \quad (8)$$

### (3) 一致性检验

对构造的判断矩阵进行一致性检验，用来确定权重分配是否合理。

一致性指标：

$$CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{3.0014 - 3}{3 - 1} = 0.0007 \quad (9)$$

检验系数：  $CR = \frac{CI}{RI}$ ，其中  $RI$  是平均一致性指标，如表 5 所示。

$$CR = \frac{CI}{RI} = \frac{0.0007}{0.52} = 0.0013 < 0.1 \quad (10)$$

表 5 RI 系数表

阶数	1	2	3	4	5	6	7
RI	0	0	0.52	0.89	1.12	1.26	1.36

一般来说，当  $CR < 0.1$  时，可认为判断矩阵具有满意的一致性。由计算结果可知判断矩阵具有满意的一致性。

### (4) 指标评价权重

将相关性、完整性和可解释性对应的权重记为  $W_{\text{相关}}$ 、 $W_{\text{完整}}$  和  $W_{\text{可解释}}$ 。经过以上计算，可得答复意见相关性、可解释性和完整性三个指标的权重如表 6 所示。

表 6 答复意见评价指标权重

指标	相关性	可解释性	完整性
权重	0.581	0.309	0.110

### 3.4.5 综合得分

基于指标得分和指标权重，得到综合得分（ $S_{\text{综合}}$ ）的计算公式，如式 11 所示。

$$S_{\text{综合}} = S_{\text{相关}} * W_{\text{相关}} + S_{\text{完整}} * W_{\text{完整}} + S_{\text{可解释}} * W_{\text{可解释}} \quad (11)$$

## 4 实验及结果分析

### 4.1 Word2vec 模型检验

对第3章训练出的 Word2vec 模型的效果从相关性(relatedness)和类比性(analogy)两方面进行。

#### 4.1.1 相关性

相关性是指对于一对相关度比较高的词，它们的词向量余弦相似度会非常接近。由图7可以看到，与“A市”词向量余弦相似度最接近10个词语中，在真实词义中，除了“西地省”外，都是市级地名词，与“A市”相关度高。由此可见，本文使用 Word2vec 训练出的词向量模型的效果还是非常不错的。

```
In [7]: [i for i in model.wv.similar_by_word('A市', topn = 10)]
Out[7]: [('B市', 0.9146615862846375),
          ('K市', 0.8932117819786072),
          ('L市', 0.8822261095046997),
          ('C市', 0.8821918964385986),
          ('西地省', 0.8585730195045471),
          ('J市', 0.8570210337638855),
          ('G市', 0.8549147248268127),
          ('M市', 0.8384421467781067),
          ('F市', 0.8205327987670898),
          ('E市', 0.8193401098251343)]
```

图7 相关性示例

#### 4.1.2 类比性

类比性是指对给定的词  $y$ ，能否找到一个对应的词  $x$ ，使  $x$  与  $y$  的关系能够类比另外两个已知相同关系的词  $a$  与  $b$  之间的关系。在类比性上，比较经典的是“king-queen”与“man-woman”之间的关系，但由于本文训练词向量的语料中已去除非中文字符，无法使用这个例子，因而本文利用“男人-男孩”与“女人-女孩”之间的关系来进行说明。由图8可知，这两组词语之间是非常接近的，侧面验证了这个 Word2vec 模型的可靠性。

```
In [5]: model.similarity("男人", "男孩") - model.similarity("女人", "女孩")
E:\python-3.7.0\Lib\site-packages\ipykernel_launcher.py:1: Deprecati
0.0, use self.wv.similarity() instead).
"""Entry point for launching an IPython kernel.
Out[5]: 0.010892034
```

图8 类比性展示

4.2 群众留言分类模型实验及结果分析

利用附件 2 预处理后的训练数据集建立按照分类模型方案建立朴素贝叶斯分类模型和卷积神经网络模型，并利用测试集对分类效果进行评价。

4.2.1 朴素贝叶斯分类模型评价

由表 7 可知，得到的朴素贝叶斯分类模型在测试集上准确率为 87.68%，模型的  $F1-Score$  为 0.87。其中“劳动和社会保障”、“教育文体”、“环境保护”三个一级分类的  $F1-Score$  超过 90%，分类效果较好，而“交通运输”的  $F1-Score$  最低，仅为 0.78。由支持数据可看到，“交通运输”类别的留言数量最少，很有可能是因为留言少导致特征提取不全面，因而召回率低，进而导致了  $F1-Score$  不理想，但也并不是很差。

由图 9 所示的朴素贝叶斯文本分类模型预测结果与真实结果的混淆矩阵可知， $F1-Score$  最低的“交通运输”，有很大一部分被误判为了“商贸旅游”和“城乡建设”，因为在特征词上与“交通运输”类别的特征词较为接近，而后者的数量少，可能在训练未能完全提取出该类别的特征，因而造成了误判。从混淆矩阵上，无论是横向分析，还是竖向分析，各类别的误判数量还都是较高的，这很有可能是因为词袋模型只是记录词语是否出现，不考虑词语的具体意思而造成的。

表 7 朴素贝叶斯分类模型评价

Test Acc: 87.68%				
名称	Precision	Recall	F1-score	Support
交通运输	0.85	0.72	0.78	120
劳动和社会保障	0.90	0.91	0.91	386
卫生计生	0.93	0.84	0.88	164
商贸旅游	0.82	0.80	0.81	227
城乡建设	0.83	0.87	0.85	422
教育文体	0.92	0.94	0.93	324
环境保护	0.90	0.93	0.91	168
Accuracy			0.88	1811
Macro avg	0.88	0.86	0.87	1811
Weighted avg	0.88	0.88	0.88	1811

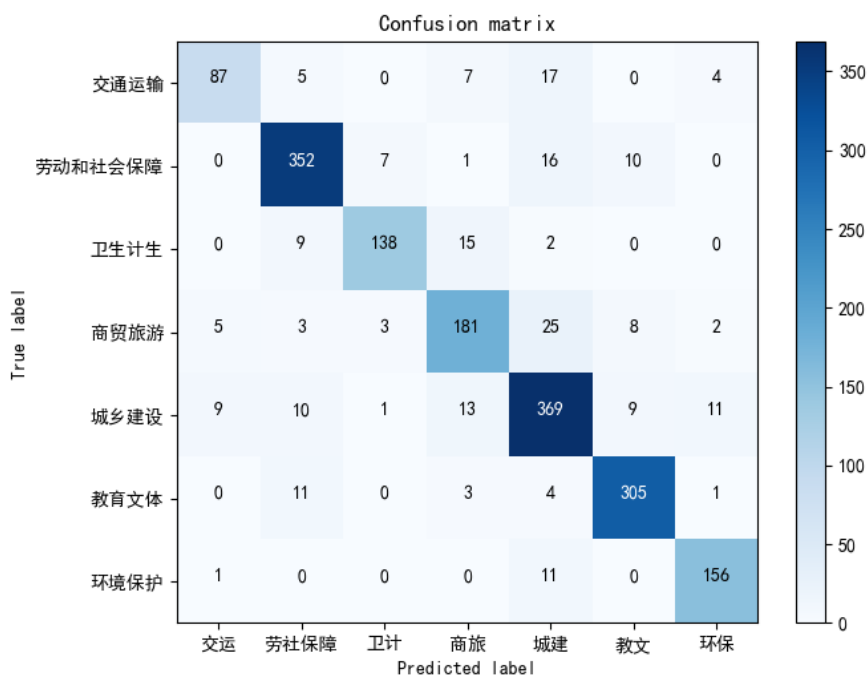


图 9 真实分类与预测结果的混淆矩阵

#### 4.2.2 卷积神经网络分类模型评价

由表 8 可知，CNN 模型的预测结果相对朴素贝叶斯分类器有明显的提高。7 个一级分类中，教育文体、劳动和社会保障的  $F1-Score$  达到了 0.95， $F1-Score$  最低的分类也达到了 85%。由于使用了全部数据划分出来的训练集进行训练，数据量相对较大，因而无论是从准确率还是从模型  $F1-Score$  上进行比较，基于深度学习的 CNN 模型都要优于基于传统机器学习的朴素贝叶斯分类器。

由图 10 所示的 CNN 模型预测结果与真实结果的混淆矩阵可知，“城乡建设”不仅容易被误判成其他类别，其他类别也容易误判为“城乡建设”，若去掉“城乡建设”这个分类，混淆矩阵是非常理想的。经进一步分析，发现“城乡建设”包含的三级分类最多，涉及范围也非常广，这是导致“城乡建设”分类不够理想的主要原因之一，另外一部分原因是可能题目所提供数据中存在争议分类（归纳为两个一级分类都合理）。

表 8 CNN 测试结果

Test Loss: 0.28		Test Acc: 91.22%		
名称	Precision	Recall	F1-score	Support
城乡建设	0.90	0.88	0.89	422
环境保护	0.93	0.92	0.92	168
交通运输	0.86	0.84	0.85	120
教育文体	0.95	0.96	0.95	324
劳动和社会保障	0.95	0.94	0.95	386
商贸旅游	0.84	0.89	0.87	227
卫生计生	0.84	0.90	0.90	164
Accuracy			0.91	1811
Macro avg	0.91	0.90	0.91	1811
Weighted avg	0.91	0.91	0.91	1811

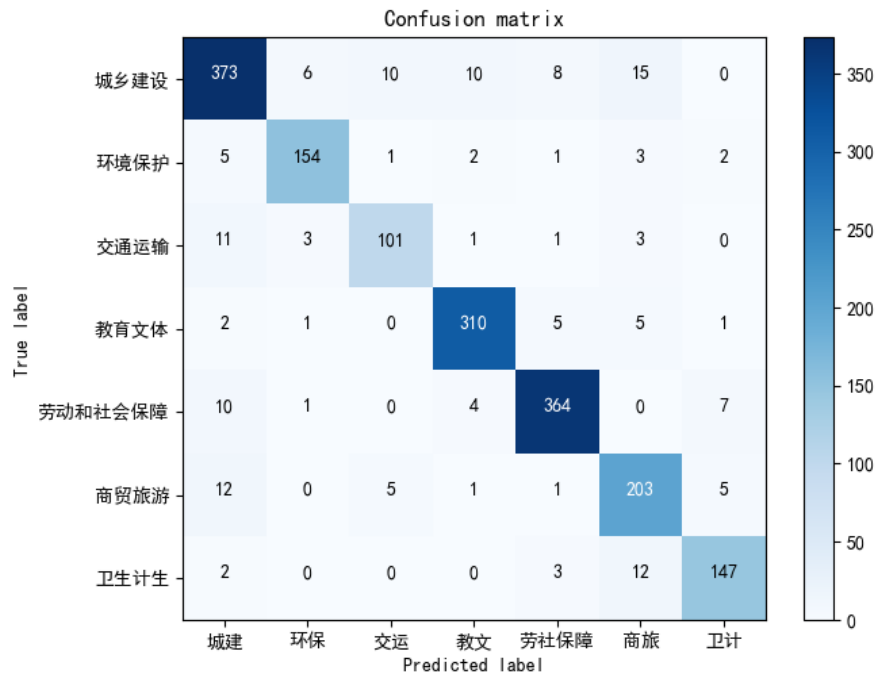


图 10 真实分类与预测结果的混淆矩阵

### 4.3 热点问题挖掘及结果分析

4.3.1 LDA 主题模型结果及分析

表 9 是以 A5 区留言主题作为语料使用 LDA 模型后得到的 3 个主题和各个主题中最常见的 10 个词，每个词语后面的数值可反映词语出现的概率。通过三个主题特征词的归纳可知：①某有限公司涉嫌诈骗；②保利（中航）违建，存在安全隐患；③劳动东路（魅力之城）夜宵摊扰民。当然，这里并不能直接断定以上三个主题是该区域存在的热点问题，还需要结合网络语义图，并输出与上述三个主题有关留言主题，然后做一进步判断。

表 9 A5 区 3 个主题及 10 个主要词语的概率分布

Topic1	Topic2	Topic3
有限公司 0.010	严重 0.008	劳动东路 0.012
诈骗 0.009	业主 0.007	魅力之城 0.011
涉嫌 0.008	保利 0.007	物业 0.009
投资 0.006	中航 0.006	严重 0.008
业主 0.006	违建 0.006	中学 0.007
违规 0.005	安全隐患 0.006	街道 0.007
解决 0.005	未 0.005	小学 0.007
巨额 0.005	停车场 0.005	扰民 0.006
资金 0.005	新 0.005	夜宵摊 0.006
拖欠 0.005	街道 0.005	苑 0.006

4.3.2 网络语义图展示及分析

图 11 是使用 A5 区留言主题作为语料所得的网络语义图。词语与词语之间的连线颜色越深代表两个词语之间的关系越强，词语背景的圆圈越大代表词语出现的概率越高。

图 11 可看到，“有限公司”、“涉嫌”、“诈骗”之间的连线颜色非常深，在 LDA 主题模型中这三个词语也在同一个主题中，那么代表着这三个词语同时出现与多个留言主题，利用正则表达式寻找并输出带有以上三个词其中两个词语的留言主题，若多个留言主题描述同一件事，即可将该事件归纳为可能的热点问题，并大致写出热



点描述；网络语义图中，“劳动东路”、“魅力之城”、“严重”出现次数多且靠得非常近，并结合 LDA 模型主题三中的关键词，大致可知：劳动东路魅力之城小区的夜宵摊严重扰民。利用正则表达式寻找功能，以“魅力之城|劳动东路”，“夜宵摊|扰民”作为关键词，输出符合条件的留言主题即可判断是否有多人反映该事件，若是，则归纳为可能的热点问题，否则忽略。以此方法对各区域留言进行分析，即可大致得到留言中的热点问题，最后使用正则表达式寻找功能将符合条件的留言提取和保存，并按照规则给出热点指数即可。

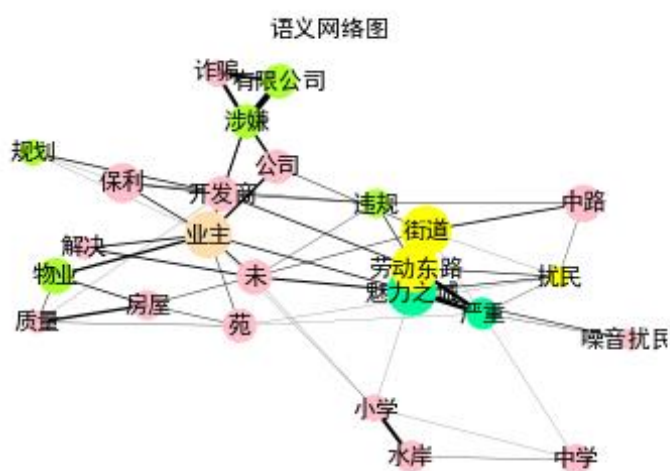


图 11 A5 区留言主题网络语义图

#### 4.3.2 热点展示及分析

图 12 展示附件 3 中提取出的一部分热点，对于所挖掘出的热点，本文从留言数、点赞数、反对数、持续时间四方面计算得到了各热点的热度指数。从图中我们可以发现，虽然有些热点留言数多，但是由于持续时间过长，热度被削减了；也有部分热点留言不多，但其点赞数众多，点赞数也侧面反映了问题的热度，因而热度指数会相对较高。最终我们按热度指数大小选出了前 5 个热点问题，并对各热点问题的留言进行保存。

	问题描述	留言数	点赞数	反对数	开始时间	截止时间	时长	热度指数
5	A市58车货 诈骗案恶性退出案件	7	798	1	2019-01-08	2019-02-25	47	0.024718
9	A6区月亮岛路 沿线架设高压线, 威胁居民的安全	9	201	3	2019-03-26	2019-04-15	20	0.015185
1	丽发新城搅拌站 噪音扰民、污染严重	55	51	2	2019-11-02	2020-01-26	85	0.012194
12	A市伊景园滨河苑 捆绑销售车位	51	9	0	2019-07-07	2019-09-01	56	0.011782
7	A5区劳动东路魅力之城 小区楼底烧烤夜摊油烟扰民	21	18	18	2019-07-21	2019-12-04	136	0.010628
16	A市 有关人才购房补贴政策问题	29	46	7	2018-11-15	2019-12-12	392	0.004595
15	A3区天顶街道青山新村 店铺违规改商	14	0	2	2019-04-22	2019-08-05	104	0.003252
13	涉外经济学院 强制学生定点实习	10	1	13	2017-06-08	2019-11-27	902	0.002612
6	A4区天健盛世A1区三期 夜间施工严重扰民	7	4	3	2019-03-15	2019-07-15	122	0.002425
8	西地省聚利(利聚)投资有限公司 涉嫌诈骗巨额资金	7	33	2	2019-03-09	2019-07-23	136	0.002368

图 12 热点展示图

## 4.4 答复意见质量评价及结果分析

### 4.4.1 相关性评分展示

图 13 展示了相关性强、未明确答复、答非所问的答复意见余弦相似度，相对而言还是较为合理的，但是这种评价方法也存在十分明显的缺点：若两个文本的长度不一致，余弦相似度就会出现偏差。为避免这种现象发生，本文已对答复意见进行去头去尾处理，虽然并不能完全准确的评价答复意见与留言内容的相关性，但对于大多数问答来说，评价都是较为合理的。答复意见的相关性评分见附件。

```
In [21]: # 相关性强的答复意见(留言编号: 3720)
text1 = '2016年下半年新开铺路全线开通后给市民行车带来了极大的方便, 但在新开铺路与披塘路交叉路口'
text2 = '网友“UU008444”您好! 您的留言已收悉。现将有关情况回复如下: 经查, 新开铺路(绕城高速—'
# 未给出明确的答复意见(留言编号: 11985)
text3 = '''尊敬的市委市政府领导:
我想反映的是福元西路沿线的公交线路太少的问题。这条路与A1区北路接轨, 所在的小区包括珠江花城/
807原来的路线是从A市大学那边开到福元西路再绕道A1区北路上然后拐到省妇幼, 要是想到五一广场还
沿线小区上十万人出门只能靠步行至A1区北路上的马厂站坐公交了! 请问, 这合理吗? 年轻的人上下班?
这种情况从2009年一直持续到了今天, 已经有5年多了。从没有一辆公车可以让小区居民从家门口到达城
不远处的楚江世纪城, 之前也是公交线路不发达, 现在有好几路公车经过, 方便多了! 希望市委市政府
text4 = '网友: 您好! 留言已收悉'
# 答非所问(留言编号: 4133, 答复意见编号: 4901)
text5 = '''本人小孩明年即将读小学, 孩子和母亲的户口在A市, 而我本人孩子的父亲因为工作关系户口在
text6 = '''网友“UU008761”您好! 您的留言已收悉。现将有关情况回复如下: 万家丽高架桥全路段入口均
print('相关性强的答复意见(留言编号: 3720), 余弦相似度为: %.6f'%model.n_similarity(yu_chuli(text
print('未给出明确的答复意见(留言编号: 11985), 余弦相似度为: %.6f'%model.n_similarity(yu_chuli(t
print('答非所问(留言编号: 4133, 答复意见编号: 4901), 余弦相似度为: %.6f'%model.n_similarity(y
<
相关性强的答复意见(留言编号: 3720), 余弦相似度为: 0.934144
未给出明确的答复意见(留言编号: 11985), 余弦相似度为: 0.400181
答非所问(留言编号: 4133, 答复意见编号: 4901), 余弦相似度为: 0.653729
```

图 13 相关性评价示例

### 4.4.2 完整性评分展示

由图 14 可知，对于未作出明确答复的留言，缺少了最主要的留言详情部分，因而只可得到 60 分。对于满足本文提出的政府部门网络留言答复规范的答复意见，在完整性方面将会得到 100 分。对于不完全符合规范的留言，会依据满足程度进行打

分。各答复意见的完整性详细得分见附件。

```
In [7]: # 留言编号: 12451
text1 = '''网友: 您好! 留言已收悉'''
# 留言编号: 2759
text2 = '''网友“A00077538”: 您好! 针对您反映A3区含浦镇马路卫生很差!
# 留言编号: 23770
text3 = '''    一、罗巷集镇于2013年进行了基础设施建设, 公路两边修建了
print('未作出明确答复, 完整性得分: %.2f'%auto_score2(text1))
print('规范答复, 完整性得分: %.2f'%auto_score2(text2))
print('不完全符合规范答复, 完整性得分: %.2f'%auto_score2(text3))

未作出明确答复, 完整性得分: 60.00
规范答复, 完整性得分: 100.00
不完全符合规范答复, 完整性得分: 82.00
```

图 14 完整性评价示例

#### 4.4.3 可解释性评分展示

由图 15 可知, 未作出明确答复的答复意见并不具备可解释性, 因而得分为 60 分。而对于进行了一定的解释, 但未引用到法律条文等, 由于解释内容没有法律作为支撑点, 所以未能获得满分。依据法律条文等作出解释的答复意见将获得满分。各答复意见的可解释性详细得分见附件。

```
In [6]: # 留言编号: 12451
text1 = '''网友: 您好! 留言已收悉'''
# 留言编号: 153290
text2 = '''“UU0081023”您好! 您在网上反映“M9县星城小区一天物业合同到期拒绝交接
# 留言编号: 9225
text3 = '''网友“UU0081831”    您好! 您的留言已收悉。经A5区教育局调查了解, 现将
print('未作出明确答复, 可解释性得分: %.2f'%auto_score3(text1))
print('有一定的解释, 但没引用法律条文, 可解释性得分: %.2f'%auto_score3(text2))
print('引用政策并详细解释的答复意见, 可解释性得分: %.2f'%auto_score3(text3))

未作出明确答复, 可解释性得分: 60.00
有一定的解释, 但没引用法律条文, 可解释性得分: 90.00
引用政策并详细解释的答复意见, 可解释性得分: 100.00
```

图 15 可解释性评价示例

#### 4.4.4 综合评分展示

由图 16 可知, 综合评分是根据相关性、完整性、可解释性而得到的, 相关性得分在综合评分中占比最高, 因为对于留言者来说, 若答复意见与留言内容并没直接联系, 那么即使答复留言完全满足答复规范, 解释得十分细致都不会感到满意。在满足相关性的情况下, 答复意见若具有规范性、可解释性, 那么留言者看起来会感到十分舒服, 因而答复意见的综合评分会更高。

```
In [8]: # 计算综合得分
df['综合得分'] = df['相关性得分']*0.581 + df['完整性得分']*0.11 + df['可解释性得分']*0.309
df.loc[:, ['相关性得分', '完整性得分', '可解释性得分', '综合得分']].head()
```

Out[8]:

	相关性得分	完整性得分	可解释性得分	综合得分
0	93.54	94	90	92.49674
1	73.02	100	90	81.23462
2	89.59	94	90	90.20179
3	76.41	94	100	85.63421
4	86.26	100	75	84.29206

图 16 答复意见综合评分

## 5 总结与展望

### 5.1 总结

本文根据题目要求，基于自然语言处理技术（NLP），运用了 Word2vec 词向量模型、卷积神经网络模型（CNN）等工具，构建了一个智慧政务系统，其基本功能主要有：文本分类、热点提取、答复意见评价。该系统已基本满足政府部门处理网络留言的需求，能有效提升政府部门的管理水平和施政效率。虽然该系统很大程度上使用了智能化技术，但是在热点提取时，还是需要一定的人工操作，热点也需要人为判断，后续还有待改进。

### 5.2 未来展望

#### 5.2.1 改进使用工具

（1）在中文文本分词工具中，经过测试，认为哈工大语言云的分词效果最佳，但其商用化是需要付款的，若项目有经费支持，建议使用哈工大语言云；

（2）使用数据库装载数据。对于新获取的留言，先使用现有模型进行分类，划分好后分发到各部门处理，若一定时间内没反馈（分类正确）则加入到数据库；分类模型以数据库的留言（带标签）为语料，定时更新，这样能不断挖掘各分类新出现的特征，以得到更好的分类结果。

#### 5.2.2 使用深层神经网络

在深度学习中，虽然并不是层数越多，分类效率就越高，但是增加卷积层在一定程度上确实会提取出更多的特征。15 年由中国学者提出了 Resnet 分类网络结构有效的避免了多层卷积网络中低效率层的影响，能有效提高深层神经网络的分类效率，但是通常情况下，使用 Resnet 网络结构的模型进行训练，训练时间都是以天来计算的，花费时间会较长，但分类会明显提高。若设备配置支持，也有充足的训练时间，

使用 Resnet 分类网络进行训练，能有效提高分类效率。

### 5.2.3 全自动化提取热点

本文中，提取热点时借助人工作业主要是因为按区域划分不够细致，因而经常需要人为判断。对提取热点功能进一步优化可以从留言划分入手，首先导入该城市的地图，按照留言中提及的地名词将留言归纳到地图中的各个区域（需要细致某小区、某道路等），为各区域设置一个留言数阈值，若留言数超过阈值，该地区将会被自动点亮。对于被点亮区域的留言，利用 CNN 模型对其进行分类（若已有分类这步可忽略），再按分类将留言划分，留言数达到一定数量的分类将自动使用 Word2vec 词向量两两计算留言主题的余弦相似度，余弦相似度达到一定阈值归为同一个问题，自动提取并保存。

### 5.2.4 答复意见评价优化

在相关性评价方面，由于留言内容和答复意见在长度上并不是一致的，词语之间也存在的顺序关系，因而使用 Word2Vec 模型得到的词向量计算两文本的余弦相似度来衡量答复意见的相关性不能说十分理想，后期若有办法将答复意见的内容归纳为一句简短的文本（主题语句），然后再使用 Word2vec 词向量模型计算留言主题与答复意见主题语句的余弦相似度，或许能更为合理的评价答复意见的相关性。

## 致谢

时光荏苒，在赛题论文即将完成之际，比赛也即将完结。在此，首先我要代表全体队员向指导老师致以最真诚的感谢。在过去的一个月里，我们一步一步努力学习，在我队导师的指导下，慢慢深入分析，最终成功建立智慧政务系统。

其次，要感谢泰迪杯组委会给我们提供的参赛机会，给予我们将理论知识应用到实际生活事例上的机会，提高我们的实操能力。

最后，要感谢队伍中每一位队员的付出与努力，以及感谢队员亲友对成员的关心与支持，我们将会更加努力，变得更好！

## Acknowledgement

Time flies, and the competition is about to end when the thesis paper is about to be completed. Here, first of all, on behalf of all the team members, I would like to extend my most sincere thanks to the instructor. In the past month, we have worked hard step by step to learn. Under the guidance of our team's mentors, we slowly and deeply analyzed and finally successfully established an Smart government system.

Secondly, I would like to thank the Teddy Cup Organizing Committee for providing us with the opportunity to participate, giving us the opportunity to apply theoretical knowledge to practical life examples, and improve our practical ability.

Finally, we would like to thank every member of the team for their dedication and efforts, as well as for their care and support for the members, we will work harder and become better!

## 参考文献

- [1] 李锐,张伟彬.基于 TF-IDF 算法和 LDA 主题模型数据挖掘技术在电力客户抱怨文本中的应用[J].自动化技术与应用,2018,37(11):46-50.
- [2] 程海琪.基于情感分类的酒店评论短文本主题挖掘[D].浙江工商大学,2020.
- [3] 董悦,王梦.基于情感分析与 LDA 模型的网络舆情案例研究[J].价值工程,2019,38(34):169-172.
- [4] 陕西省生态环境厅办公室.关于规范政府信息依申请公开答复工作的通知[OL].2019-2-22.<http://sthjt.shaanxi.gov.cn/service/files/shbh/2019-02-22/38121.html>.
- [5] 人民政府.政府信息公开申请答复规范文本[OL].2018-06-12. <http://www.guixi.gov.cn/Item/56906.aspx>.
- [6] 齐凯凡.基于卷积神经网络的新闻文本分类问题研究[D].西安理工大学,2018.
- [7] 孙媛媛.基于深度学习的文本分类研究[D].哈尔滨工业大学,2019.
- [8] 张彬城,陈杰彬,林越.一种基于潜在语义索引和卷积神经网络的只能阅读模型[D].泰迪杯优秀论文,2019.
- [9] 张航.基于朴素贝叶斯的中文文本分类及 Python 实现[D].山东师范大学,2018.
- [10] 姜天宇,王苏,徐伟.基于朴素贝叶斯的中文文本分类[J].电脑知识与技术,2019,15(23):253-254+263.
- [11] Kim Y. Convolutional Neural Networks for Sentence Classification[J],Computer Science,2014(03):189.
- [12] 杨锐,陈伟,何涛,等.融合主题信息的卷积神经网络文本分类方法研究[J].现代情报,2020,40(04):42-49.
- [13] 万磊,张立霞,时宏伟.基于 CNN 的多标签文本分类与研究[J].现代计算机,2020,(08):56-59+95.
- [14] Shufeng X, Kuiyi W, Donghong J, et al. A Short Text Sentiment-Topic Model for Product Reviews [J].Neurocomputing,2018(08):94-102.
- [15] Hui H, Xueke X, Dayong W, et al. A Sentiment Classification Method Based on Sentiment-Specific Word Embedding[J].Journal of Chinese Information Processing,2017,31(3):170-196.
- [16] 应毅,李晓明,栾晶.时间敏感的微博热点爬取与发现模型研究[J].淮海工学院学报

(自然科学版),2019,28(02):25-28.

[17]李净桦.基于神经网络的社交话题热度预测模型研究[D].重庆邮电大学,2019.

[18]张燕,杜红乐.基于层次分析法的乡村旅游影响因素研究[J].微型电脑应用,2020,36(04):54-56.