基于“智慧政务”中的文本挖掘

摘要

随着大数据、云计算、人工智能等技术的发展，各行各业的生产效率得到了大幅度的提升，政务系统作为政府与人民交互的桥梁，引入新技术提高生产力迫在眉睫。近些年来，基于自然语言处理技术的新型智慧政务系统，由于其高效、准确、适应力强的特点，在群众留言划分、热点问题整理以及提供高质量答复方案的分析等政务工作方面具有十分广阔而光明的前景，极大的提高了为人民服务的效率。

在本文中，我们首先利用 java 和 Python 对给出的数据文本预处理，绘制词云图，然后基于文本预处理的操作，结合层次分析法和定性分析法的一致性指标和覆盖率指标，对留言进行归类，并得出排名前 5 的热点问题，最后对答复意见和留言进行最优化原理的动态规划算法得出答复意见评价方案。

问题一，首先对附件 2 的数据进行数据清洗、文本去重、停用词、结巴分词、TF-IDF 算法、贝叶斯算法等文本预处理操作，建立关于留言内容的一级标签分类的**词云图**模型，运用 **F-Score** 对分类方法进行评价，其中 F-Score 的计算值为 0.795。

问题二，基于问题一，对附件 3 使用**定性分析法**和**层次分析法**，将一级分类标签及影响留言成为热点问题的条件变量（事件类型、内容变量、细节信息变量、公众注意力）进行条件变量组合分析，其整体覆盖率为 0.95623，整体一致率为 1.000000。

问题三，基于问题一和问题二的分析，根据一级标签、二级分类标签、三级分类标签，将留言跟答复意见进行**最优化原理**的动态规划算法的处理，并结合层次分析法进行拟合，通过答复的相关性、完整性、可解释性等角度对答复意见的质量给出一套评价方案。

关键词：文本挖掘；词云图；F-Score；定性分析法；层次分析法；最优化原理

Text mining based on "intelligent government"

Abstract

With the development of big data, cloud computing, artificial intelligence and other technologies, the production efficiency of all walks of life has been greatly enhanced. In recent years, the new intelligent government affairs system based on natural language processing technology, due to its high efficiency, accuracy and strong adaptability, it has a very broad and bright prospect in the fields of the division of people's comments, the sorting of hot issues and the analysis of high-quality reply plans, which has greatly improved the efficiency of serving the people.

In this paper, we first use Java and Python on the given data text preprocessing, drawing word cloud map, message classification, and then based on the text preprocessing operations, combining the consistency index and coverage index of AHP and qualitative analysis, the paper classifies the message, and gets the top 5 hot questions. Finally, a dynamic programming algorithm based on the principle of optimization is applied to the reply opinion and message.

Question one,First of all, the data of Annex 2 are preprocessed by data cleaning, Text de-duplication, stop words, stutter segmentation, TF-IDF Algorithm, Bayes Algorithm, etc. , a word cloud model for first-level tag classification of message content was established. F-score was used to evaluate the classification method. The F-Score was 0.795.

Question two, based on question one, the use of qualitative analysis and analytic hierarchy process for Annex 3, the combination analysis of condition variables (event type, content variable, detail information variable, public attention) which influence the first-class classification label and the message become the hot topic, the overall coverage rate is 0.95623 and the overall consistency rate is 1.000000.

Question three, based on the analysis of question one and question two, according to the first-level label, the second-level label, and the third-level label, the comments and replies are processed by the dynamic programming algorithm based on the optimization principle, combining with analytic hierarchy process (AHP) , this paper gives a set of evaluation scheme for the quality of the response opinions from the angles of relevance, completeness and Interpretability of the response.

Keywords: Text Mining; Word Map; F-Score; qualitative analysis; analytic hierarchy process; optimization principle;

目录

1. 研究意义

1.1 研究背景

随着城市治理理念与方式的转变，传统的电子政务模式依然无法跟上信息时代的脚步，智慧政务悄然兴起，备受青睐，是政府信息化发展的高级阶段，成为提高政府管理水平和有效实施相关举措的重要形态之一。近年来，随着微信、微博、市长信箱、阳光热线等网络问政平台逐步成为政府了解民意、汇聚民智、凝聚民气的重要渠道，各类社情民意相关的文本数据量不断攀升，给以往主要依靠人工来进行留言划分和热点整理的相关部门的工作带来了极大挑战。同时，随着大数据、云计算、人工智能等技术的发展，建立基于自然语言处理技术的智慧政务系统已经是社会治理创新发展的新趋势，对提升政府的管理水平和施政效率具有极大的推动作用。

智慧政务是在建设智慧城市的大背景下，发展信息经济和智慧经济，实现经济和社会转型升级的必由之路。智慧政务广泛应用于物联网、云计算、移动互联网、人工智能、数据挖掘等现代信息技术，通过资料整合、流程优化、业务协同，提高政府办公、政务、监管、决策的智能化水平，从而形成高效、集约、辨明的服务型政府运营模式。可以实现各职能部门的各种资源的高度整合，提高政府的业务办理和管理效率；加强职能监管，形成高效敏捷便民的新型政府，保持城市可持续发展，为企业和公众建立一个良好的城市生活水平。

1.2 研究意义

在处理网络问政平台的群众留言中，留言信息量过大，通过对留言文本内容的挖掘分析，检测出文本关键分类标签，不仅方便快捷的帮助政府相关部门的工作，而且有利于建设“智慧政务”，特定时间特定地点的热点事件的挖掘对于政府处理人民群众的生活问题具有针对性，同时提高了政府的管理水平和施政效率，结合留言内容给予相应的合理性回复。

2. 分析方法与过程

2.1 总体流程

图 1 总体流程

本论文的分析过程大致分为以下几个步骤：

第一步：获取分析用的原始数据（群众留言文本）并导入 java 及 Python，对文本数据进行随机抽取；

第二步：对获取的数据进行文本挖掘，包括数据清洗、结巴分词、停用词过滤等文本预处理操作，并绘制一级标签分类的词云图模型；

第三步：文本数据处理后，采用 F-score 对分类方法进行评价；

第四步：对群众留言数据处理，运用定性分析法和层次分析法获取热点问题和评价指标；

第五步：通过对留言内容的关键词匹配和动态规划算法，结合最优化原理和层次分析法对相关留言答复给予评价方案。

2. 2 具体步骤

2.2.1 数据介绍

本文使用的数据为题目所给的附件 1、附件 2、附件 3、附件 4 的分类标签、群众留言及相关部门答复意见的数据。从总体上来说，“智慧政务”作为新兴的高效化政务工具，在该平台进行留言的人们群众比较多，而留言内容的主题

也较为明显。

2.2.2 基于问题一的文本留言预处理

根据附件 2 中的 9211 条的群众留言文本数据内容，我们需要先对留言详情以及留言主题这两部分进行文本预处理，留言文本中存在大量价值含量很低的甚至无价值的留言，我们将这些文本数据进行数据清洗、文本去重、结巴分词、停用词过滤、词频统计、TF-IDF 算法、贝叶斯算法及 F-Score 评价模型等，对文本进行预处理，把大量的无价值的文本留言内容去除就显得很有必要。

我们运用 java1.7 与 Python2.8 相结合的方式对文本内容进行以下操作：数据清洗（去除特殊符号及空格）、文本去重、结巴分词、停用词过滤、词频统计等，按照各自处理的特性，我们按顺序进行留言文本预处理 [1]。

2.2.2.1 数据清洗

由于在附件 2 中群众留言文本存在大量的空格以及无关的特殊符号，下面利用正则表达式的方法将附件 2 的内容导入至 java 中，并提取其中的群众留言内容列表下的留言文本数据，引入相关的库，编写相关的代码去除其中的空格以及特殊字符等，其局部代码及结果如下图所示：

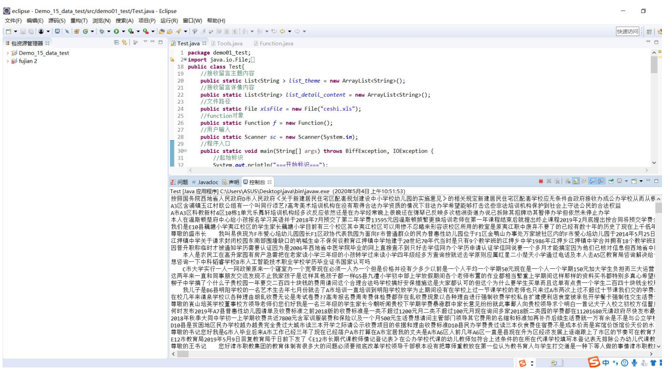


图 2 数据清理代码

2.2.2.2 文本去重

文本去重即是去除相同或相似的群众留言文本。由于语言的特点，在附件 2 中，对于不同的群众所写的留言内容存在相同或相似的意思，同一名群众有可能重复留言的现象，因此对文本进行去重也显得必要。

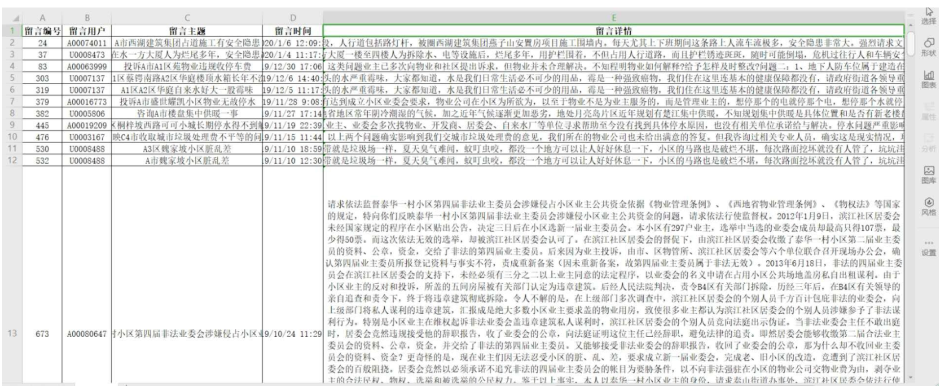


图 3 文本去重后的留言

如上图中留言编号为 319 和 379、530 和 532 的用户所写的留言内容是一致或相似的，对于此类情况便需要进行文本去重。由于相近的语义存在有价值的

留言内容，为保留更多的有用语料，我们采用简单的比较删除法，两两比较，完全相同的就去除，仅保留其中一条即可。

2.2.2.3 结巴分词

对留言文本进行数据清洗后，每一条留言都已完整的列出，在进行文本挖掘的时候，首先应该对文本进行分词，采用中文分词将连续的字符列按照一定的规范重新组合词序列。

分词结果的准确性对后续文本的挖掘算法具有重大的影响，如果分词不佳，即使后期的算法很优秀也无法达到预期的效果，不同的分词效果，直接影响词语在文本中得重要性，从而影响特征文本向量的选择。

本文才用的是 Python 的中文分词中的结巴分词[2]，结巴分词采用了基于前缀词典实现的高效词图扫描，生成句子中汉字所有可能成词情况所构成的有向无环图（DAG），采用动态规划查找最大概率路径，找出基于词频的最大切分组合，使得分词效果更佳，jieba 分词系统提供分词、词性标注，支持用户自定义词典，关键词提取等功能。

在分词的同时，采用了 TF-IDF 算法及贝叶斯算法，提取每条群众留言内容的前 5 个关键词，这里采用 jieba 自带的词义库。

部分分词结果示例如下图：

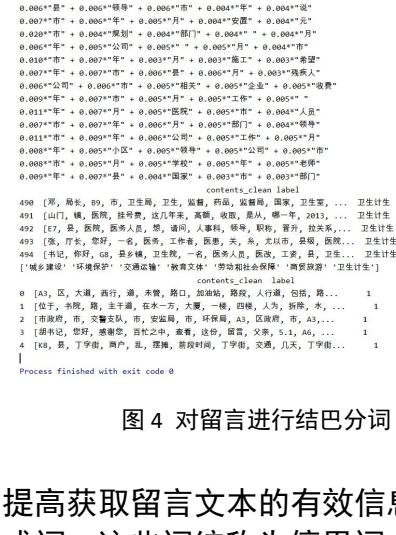


图 4 对留言进行结巴分词

2.2.2.4 停用词过滤

为节省存储空间和提高获取留言文本的有效信息效率，在处理文本之前会自动过滤掉无意义的字或词，这些词统称为停用词（stopwords），停用词的两个特征为：一是极为普遍、出现频率高；二是包含信息量低，对文本标识无意义。例如中文中的“的”、“呢”、“了”、“啊”等，停用词的介入可能会造成选出的特征几乎是停用词，从而影响结果的分析，在进行停用词过滤时保留了对否定词“不”、“没有”等词语，自定义在 stopwords 中删除或增添停用词[7]。

本文采用基于停用词库并自定义进行词表的增添删改操作的基础上对附件 2 中得群众留言文本进行停用词过滤，将分词结果中的词语进行匹配，若匹配成功，则进行删除结果，结果示例如下：

0	A3 区 大道 西行 道 未管 路口 加油站 路段 人行道 包括 路灯 杆 圈 西湖 建筑 ...
1	位于 书院 路 主干道 在水一方 大厦 一楼 四楼 人为 拆除 水 电等 设施 烂尾 多年 ...
2	尊敬 领导 A1 区苑 小区 位于 A1 区 火炬 路 小区 物业 市程明 物业管理 有限公...
3	A1 区 A2 区华庭 小区 高层 二次 供水 楼顶 水箱 长年 不洗 自来水 龙头 水 毒...
5	2015 年 购买 盛世 耀凯 小区 17 栋 楼 楼 两层 共计 平方 足额 缴纳 物业费...
6	西地省 地区 常年 阴冷 潮湿 气候 近年 气候 恶劣 地处 月亮 岛 片区 近年 规划 楚...
7	尊敬 胡书记 您好 家住 市 A3 区 桐梓 坡 西路 可可 小城 居民 停水 小区业主 业...
8	梅家田 社区 辖区 小区 居民 依法 依规 小区 物业公司 交纳 城市 垃圾处理 费 环卫局...
9	尊敬 市政府 领导 您们好 市 A3 区 魏家坡巷 业主 多年 小区 脏 乱 差 社区 得不...
11	请求 依法 监督 泰华 一村 小区 第四届 非法 业主 委员会 涉嫌 侵占 小区业主 公共 ...
12	住 梅 溪湖 壹号 御湾 楼 2019 年 月份 住 每天晚上 停水 白天 水 很小 用水 ...
13	尊敬 领导 你们好 市 A4 区 捞刀河 镇 彭家巷 社区 鸿涛 翡翠 湾 一名 业主 一名...
14	地铁 号线 施工 导致 万家 丽路 锦楚 国际 星城 小区 三期 一个月 停电 10 来次 ...
15	尊敬 领导 你好 A6 区润 紫 郡 业主 今年年初 小区 周边 竖起 一道道 高压线塔 筑...
16	市 A5 区 朝晖路 锦楚 国际 新城 三区 月份 一共 停电 次 每次 说 原因 停电 线...
17	肯定 选择 A9 市 西南角 支持 A9 市 西南角 设 站 A8 县 南 北 中三向 融城...
18	尊敬 领导 A6 区 几年 发展 突飞猛进 城市道路 绿化 建设 却显 落后 滨水 新城 高...
19	A5 区楚府 线 包括 森林 雅苑 楚府 十城 天际 山庄 多个 小区 停电 短短 一周 停...
20	涂 愈 一名 建筑业 从业者 求助 请求 相关 部门 调查 西地省 建望 集团 及西地 省辉...
21	市 A2 区 黄谷路 368 号 山水 嘉园 栋 单元 706 房 改建 成 户 出租 人员...
22	市政府 市 交警支队 市 安监局 市 环保局 A3 区政府 市 A3 区 杜鹃 文苑 小区 ...
23	市 B 市 C 市要 融城 交通 基础 长株 潭 城铁 开通 串联 市 河东 主城区 B 市...
24	尊敬 胡书记 现有 民生问题 近两年 A5 区 嘉华 路 嘉兴 路 交叉路口 西北角 处有 ...
25	市 地铁 号 线西 延 二期 建议 暂缓 修建 一是 造价 高 建设周期 长 二是 速度慢 ...
26	修建 市 火车站 城铁 站 市 火车 南站 市 黄花 机场 高铁 联络线 优势 已建长 株 ...
27	尊敬 胡书记 现有 两个 民生问题 近两年 A5 区 嘉华 路 两旁 人行道 状况 脏乱差 ...
28	胡书记 冬天 市 湿冷 冬天 受不了 诶 太冷 被子 感觉 潮潮 洗 衣服 难 干 早上 起...
29	A3 区银盆 南路 华韵 城市 海岸 群 租房 泛滥 小区 居民 生活 得不到 保障 投诉无...
30	尊敬 市委 市政府 市是 一座 历史 名城 一座 幸福感 城市 幸福感 体现 市委 市政府 ...

图 5 停用词过滤后分词结果

1.2.2.5 TF-IDF 算法

TF-IDF 算是中文分词中较为简单、基础而通用的一种算法，其核心思想即文本关键词提取在对附件 2 中留言内容进行过滤分词后，需要将这些词语转换成向量，以供挖掘分析使用，这里采用 TF-IDF 算法，把群众留言内容转化为权重向量，TF-IDf 算法的基本原理如下[2]：

第一步：计算词频，即 TF 权重：词频（TF）=某个词在文本中出现的次数考虑到留言文本的长短不同，为了与不同留言内容进行比较，进行词频标准化，除以文本的总词数或者除以该留言文本中出现的次数即：

$$\text{词频（TF）}=\frac{\text{某个词在文本中出现的次数}(N)}{\text{文本的总字数（M）}} \quad (1)$$

或 $$\text{词频（TF）}=\frac{\text{某个词在文本中出现的次数}(N)}{\text{该文本出现次数最多的词的出现次数（O）}} \quad (2)$$

第二步：计算 IDF 权重，即逆文档频率，IDF 越大，此特征性在文本中越集中，说明该分词在该文本中的标签属性能力越强。

$$\text{逆文档频率（IDF）}=\log\left(\frac{\text{词料库的文本总数（D）}}{\text{包含该词的文本数+1（}D_w\text{）}}\right) \quad (3)$$

第三步：计算 TF-IDF 值，TF-IDF=词频（TF）X 逆文档频率（IDF）

实际分析得出 TF-IDF 值与一个词在群众留言文本中出现的系数成正比，某个词在文本的重要性越高，TF-IDF 值越大。在进行计算没歌词的 TF-IDF 中，需要进行排序，次数最多的即为群众留言文本内容的关键词。

使用 TF-IDF 算法，找出每条留言内容的前 10 个关键词，并将关键词合并为集合，计算每条留言内容的在整个集合中的词的词频，如果没有就值为零，并生成每条留言内容的 TF-IDF 权重向量。计算公式为：

TF-IDF=词频（TF）X 逆文档频率（IDF） (4)

根据词频统计结果及算法，将群众留言内容的特征词提取出来，大致分为七类：

表 1 一级分类标签特征词

一级分类	特征词
城乡建设	市、业主、小区、年、领导、月、县、部门、开发商、政府
环境保护	污染、市、村民、居民、县、环保局、年、领导、生产、生活
交通运输	出租车、市、县、公司、快递、元、司机、年、说、车辆
教育文体	学校、市、教师、学生、教育、年、老师、孩子、县、教育局
劳动和社会保障	年、月、市、工作、领导、职工、县、劳动、单位、工资
商贸旅游	市、县、年、公司、元、电梯、月、说、部门、业主
卫生计生	医院、医生、年、月、说、市、领导、日、患者、生育

由于上述词频统计得出的特征词提取中，均含有“市”、“年”二词，即对此词进行删除，再进行 TF-IDF 算法，此时算法所提取的关键词降为 8 个，具体情况如下表：

表 2 去除相同关键词后的一级标签分类

一级分类	特征词
城乡建设	业主、小区、领导、月、县、部门、开发商、政府
环境保护	污染、村民、居民、县、环保局、领导、生产、生活
交通运输	出租车、县、公司、快递、元、司机、说、车辆
教育文体	学校、教师、学生、教育、老师、孩子、县、教育局
劳动和社会保障	月、市、工作、领导、职工、劳动、单位、工资
商贸旅游	县、公司、元、电梯、月、说、部门、业主
卫生计生	医院、医生、月、说、领导、日、患者、生育

2.2.2.6 词云图绘制

词云图，主要是将文本数据中出现频率较高的关键词以可视化的形式战线出来，使人一眼看出文本数据的主要表达意思。

本文通过将附件 2 的留言文本内容导入 Python，进行词频统计之后，按照词频降序排列画出排在前 100 热词的词云图，紧接着对词进行大致分类，此时将留言数据进行一级标签分类。

2.2.2.7 贝叶斯算法

根据附件 2 中群众留言内容的 TF-IDF 权重向量后，根据每一类留言标签的特征向量，采用朴素贝叶斯算法进行模型检测评估[5]。

这里用 Python 中的 sklearn 中的 GaussianNB 模块。先将附件 2 的数据分为“城乡建设”和“非城乡建设”的数据随机选取中 80%的数据作为“训练数据”进行模型训练，20%的数据作为“测试数据”进行进行预测。预测出“测试数据”的结果后，计算出其对应的 F-Score。接着如果将预测的结果为“非城乡建设”的数据进行进行下步预测，即将预测结果为“非城乡建设”进行预测是否为“环境保护”，则会导致错误率不断增加。为此，我们不将预测结果的数据进行下一步预测，而是将附件 2 中的所有数据分为“环境保护”和“非环境保护”，这样可以使单项数据结果的准确率更高，也不会对后面的其他数据造成影响。

原始的朴素贝叶斯只能处理离散数据，当 $x_1,...,x_n$ 是连续变量时，我们可以

使用高斯朴素贝叶斯完成分类任务。

当处理连续数据时，一种经典的假设是：与每个类相关的连续变量的分布是基于高斯分布的，故高斯贝叶斯的公式如下：

$$\mathrm{P}（x_i = v \mid y_k）= \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(v-\mu_{y_k})}{2\sigma^2}) \tag{5}$$

其中 $\mu_y, \sigma_{y_k}^2$，表示全部属于类 y_k 的样本中变量 x_i 的均值和方差

2.2.2.8 F-Score 评价

信息检索、分类、识别等领域的两个最基本的指标就是召回率（Recall Rate）和准确率（Precision Rate），其中召回率也叫查全率，准确率也叫查准率，概念公式[3]：

$$\text{召回率（Recall）}= \frac{\text{预测为真实正例}}{\text{所有真实正例样本的个数}} \tag{6}$$

$$\text{准确率（Precision）}= \frac{\text{预测为真实正例}}{\text{所有被预测为正例样本的个数}} \tag{7}$$

由于召回率与准确率是相互影响的，理想情况下两个指标都高当然最好，但一般情况下 Precision 高 Recall 就低，Recall 高，Precision 就低，为保证精确率的情况下，F-score 是一种衡量机器学习模型性能的指标，F-Score 可看作为准确率和召回率的一种调和平均，它的最大值是 1，最小值为 0，计算公式为：

$$\mathrm{F_1} = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} \frac{2P_i R_i}{P_i + R_i} \tag{8}$$

P 表示 Precision，R 表示 Recall，与此同时，设根据检验结果判定正负样本，其中 TP 表示 True Positive，被判定为正样本，事实上也是正样本；FP 表示 False Positive,被判定为正样本，但事实上是负样本；TN 表示 True Negative，被判定为负样本，但事实上是正样本；FN 表示 False Negative,被判定为负样本，但事实上是正样本，此时引入精确度（Accuracy）表示预测结果的精确度，TP、TN、FP、FN 之间和精确度、召回率、准确率的关系解释说明。

表 3 TP、TN、FP、FN 关系说明

真实情况	预测结果	
	正例	反例
正例	TP（真正例）	FN（假反例）
反例	FP（假正例）	TN（真反例）

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + TN + FP + FN} \tag{9}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + FP} \tag{10}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{11}$$

根据对附加 2 中的 9211 条群众留言文本进行词频以及词云图的统计，测试获取各留言文本的一级分类标签，将测试获取的一级分类标签与附件 2 中各群

众留言文本的真实一级分类标签进行比较，得出如下结果：

表 4 不同类型留言的准确率、精确率、召回率以及 F-Score 值

留言一级类别	Accuracy	Recall	Precision	F-score
城乡建设	0.755	0.850	0.707	0.772
环境保护	0.800	0.700	0.875	0.778
交通运输	0.750	0.800	0.727	0.762
教育文体	0.775	0.750	0.852	0.798
劳动和社会保障	0.850	0.895	0.886	0.890
商贸旅游	0.800	0.700	0.883	0.781
卫生卫计	0.765	0.750	0.819	0.783

2.2.3 基于问题二的热点问题挖掘

根据附件 3 中的 4327 条群众留言文本，我们需先应用问题一的 TF-IDF 算法得出的词云图分类方法对附件 3 的留言进行归类，并统计词频，挖掘文档中的前 20 个采用定性分析法，通过结合群众的点赞数和反对数定义合理的热度评价指标，挖掘特地时间特定地点的热点问题[4]。

2.2.3.1 热点问题的定义

随着通信业务的发展和移动通信技术的日益普及，互联网智能平台越来越成为人们日常沟通和交流的主要渠道。一般人群留言的内容是复杂多样的，需要对留言文本中隐含的处于“未点燃”的热点问题进行挖掘和分析，并运用多种信息分析技术，对留言中的热点问题及时作出反应，以便掌握处理该类问题的最佳时机，从而提高政府人员对处理热点问题的能力和监测能力。因此如何在留言文本信息对象上进行热点问题的汇集和发现、研究热点的形成过程和发展规划、准确把握留言信息热点，对于建设“智慧政务”具有十分重要的现实意义。

与留言中出现的一般性问题相比，热点问题一般不易发现，而且发生的原因不太明确，而一般性问题发生概率较高并且趋于稳定状态，诱发原因往往可以根据经验来确定，热点问题比一般问题的具有更高的价值，但发现热点问题比发现一般问题更困难，这也是本题需要解决的问题，若能及时了解问题的发展状况，将对政府工作气道重大的指导作用。

结合对大量留言信息的分析和研究，概括出了目前在附件 3 中留言热点问题所表现出来的特点：①广泛性：群众留言问题广泛存在于不同的一级留言分类标签，设计城乡建设、教育文体、交通运输等七大类型。②复杂性：群众留言的信息中并未直接明了的说明问题，大篇幅的留言中存在关键词在最后留言处才涉及的情况，有时甚至未出现任何关于一级、二级、三级留言分类标签的关键字。

通过定性分析法[6]和层次分析法[8]以及留言信息的特点分析，对热点问题进行定义：某一时段内反映特定地点或特定人群收到关注和评价较多的一些活动时事件的总称。本文将热点问题氛围两个类别：自定义的热点问题分析和未知主题的热点挖掘。

图 6 热点问题挖掘过程

2.2.3.2 fsQCA 定性分析法

定性比较分析（ Qualitative Comparative Analysis，QCA）是由美国社会学家查尔斯·拉 金（ Charles Ragin）所提出的一种基于布尔代数和集合理论的、以案例定量分析为导向的研究方法。简而言之， QCA 方法旨在分析变量间的因果关系和什么样的原 因组合在一起会促成某一结果的产生。

QCA 分析法既不同于定性研究中强调对单案例或少数案例进行深描的研究方式，也区别于定量研究中大规模样本的统计分析，而是试图探索规避传统定量研究与定性研究局限性的中间路径。QCA 认为，复杂社会问题是多因素相互结合、共同作用下的结果，因而，它强调条件组态（案例生成特定结果的各要素之间的组合）对社会问题最终结果的作用机制。QCA 以组态分析思维取代由各单变量效应叠加的定量分析思维，同时以集合之间的逻辑关系替代定量研究中的相关关系。

QCA 具有以下几个显著特点：第一，不同于以往的单案例分析，QCA 采用了跨案例的研究方法，既能有效区分不同案例的异质性，又能观察不同案例的共性；第二，不同于传统的定量分析，QCA 对案例的数量要求通常为 10 到 80 个，这也使得 QCA 在分析中小样本时会更具优势；第三，是 QCA 假定导致某种社会现 象的因果关系是复杂多元且非线性的，其运算得到的 因果关系也是以原因组合的形式来呈现的；第四，是 QCA 还聚焦于引致某一结果的充分和必要条件，这也是传统回归分析所无不能及的；第五，QCA 包括清晰集（csQCA）、模糊集（ fsQCA）与多集值（ mvQCA）定性比较分析等三种模式，鉴于邻避问题中的政策工具选择涉及复杂的政府内外部各要素的组合分析，且条件变量本身具有一定的主观性和模糊性，需要对部分变量进行更细致的测度。因此，运用 fsQCA 分析方法，探索政策工具选择的约束条件具有合理性和适当性[6]。

（1）留言数据选取

QCA 对案例的数量要求（10 到 80 个）和质量要求(案例的典型性、多样性、资料全面性和结果确定性）来进行案例选取。其中典型性指该案例具有一定社会影 响和较为广泛的关注；多样性指案例在类别、时间和地 点等要素上所具有的差异性；资料全面性指同一案例 要具有多种类型的资料来源加以支撑；结果确定性是 指结果变量能够被观察和定性。

首先，使用 Python 中的 jieba 分词对附件 3 中的群众留言内容进行分词并统计词频等预处理操作，得到词频统计最高的关键词所在的前 25 条留言数据，如下图所示：

留言编号	留言主题	留言时间
199379	A2区丽发新城附近修建搅拌厂，严重污染环境	2019/11/25 10:17
192337	A3区梅溪湖看云路一师润芳园小区临街门面油烟扰民	2019/9/5 12:23
270086	A3区青青家园公共消防生命通道架设油烟管道，持续半年	2019/5/8 9:45
284828	A3区兰亭湾畔小区违规开餐厅	2019/6/27 19:55
282978	A3区奥园城市天骄楼盘半夜泵车浇灌混凝土严重扰民	2019/10/21 18:00
240551	A3区西湖街道茶场村五组什么时候能拆迁	2019/7/9 17:09
236401	A3区洋湖街道工地施工噪音扰民归谁管？	2019/9/11 20:41
209549	A4区天健盛世A1区工地长期深夜施工扰民	2019/7/15 19:01
276660	A1区马王堆扰民现象还是存在	2019/4/19 15:16
278948	A1区老地方美食广场没有消防证却能正常营业	2019/12/20 22:55
218417	A1区火星镇兴和社区没有提供供老百姓休闲的地方	2019/3/11 8:20
289930	A1区辉煌国际城二期居民楼下商铺违法开饭店，维权近三	2019/1/27 12:18
289808	A1区朝阳街道解放东路二里牌向韶村马路市场死灰复燃	2019/3/20 17:42
270505	A1区A2区华庭负一楼车库又成垃圾场了	2019/12/6 14:29
313964	12123上申请驾驶证期满换证，一个星期了都无人受理	2019/4/26 15:28
360102	A5区劳动东路魅力之城小区底层餐馆油烟扰民	2019/9/10 6:13
222385	A7县楚龙街道楚瑞家园小区白蚁防治治标不治本	2019/3/22 16:46
223653	A7县黄星大道海伦春天小区附近辅道上违停车辆太多了	2019/4/2 11:40
209074	A7县江背镇阳雀新村几十年仍然是泥巴路	2019/3/22 16:55
229903	A7县新国道107距我家仅3米，相关政府部门为何不同意拆	2019/7/22 17:05
238391	A7县星沙中贸城欺诈业主、不退还业主购房资金	2019/3/25 10:26
248367	A市205路公交车经常不按时发车	2019/10/24 20:31
281546	丽发新城小区附近搅拌站粉尘大,无法呼吸	2019/11/29 14:19
230554	投诉A市伊景园滨河苑捆绑车位销售	2019/8/19 10:22
246362	A市魅力之城小区底层商铺营业到凌晨，各种噪音好痛苦	2019/8/26 1:50

图 7 词频较高的留言

（2）解释变量的设计

根据已有的研究成果，并结合选定的 25 条群众留言主题，本研究最终确定了 4 个具体的条件变量：事件类型、内容变量、细节信息变量、公众注意力

① 事件类型：根据题目所述以及上述留言主题，将群众留言内容筛选划分为城乡建设、环境保护、交通运输。

② 内容变量：在群众留言内容文本中有“噪音”、“垃圾”、“扰民”、“公交车”等词语，反映诸多群众的真实想法，扩大留言内容的影响力。

③ 细节信息变量：群众留言内容会详细说明时间地点以及人物事件，不但能够增强可信度，同时有利于群众留言内容的高热度传播。

④ 公众注意力：在附件 3 中所有的留言内容均有公众的投票，即点赞票和反对票。

表 5 变量选择和赋值说明

类型	变量名称	类型	判断说明	权重	赋值
条件变量	事件类型	城乡建设	城乡间是否扰乱人们正常生活的建设	33.3%	1
		环境保护	人们居住的环境是否受到污染	33.4%	0
		交通运输	人们的交通是否遇到不便之处	33.3%	0
	内容变量	有信源	有关于事件类型的相关词汇出现	70%	1
		无信源	无出现事件类型的相关词汇	30%	0
	细节信息变量	有时空细节	留言有详细的时间地点	68%	1
		无时空细节	留言无详细的时间地点	32%	0
	公众注意力	有投票	有人们对留言点赞或反对或多次留言	70%	1
		无投票	无人进行点赞或反对	30%	0
结果变量	热点问题评定	热点问题	多次出现的特定时间特定地点的留言	70%	1
		并非热点问题	留言内容出现次数较少	30%	0

（3）结果分析

①单变量必要性分析

首先，在完成变量操作化与编码后，运用 fsQCA 3.0 软件，生成真值表，即可得到条件变量的所有组合情况。其次，在构成真值表后，便可继续进行单变量必要性分析，该步骤主要是用以判断单个原因变量和结果变量之间是否存在充分关系或必要关系，并通过一致性指标（Consistency）进行判断，一致性指标的运算公式如下所示：

$$\text{Consistency}（\mathrm{X_i} \le Y_i) = \sum[\min(X_i, Y_i)] - \sum X_i \qquad （12）$$

当一致性指标大于 0.8 时，则认为改条件变量（X）为结果变量（Y）的充分条件，即改条件变量的出现能够导致结果的发生，当一致性指标大于 0.9 时，则认为 X 为 Y 的必要条件。

在 QCA 中，除了一致性指标外还采用覆盖率指标（Coverage）来判断条件变量或变量组合对于结果的解释力，例如：当覆盖率为 0.9 时，则说明该条件变量或原因变量的组合能够解释 90% 的案例，其计算公式如下：

$$\text{Coverage}（\mathrm{X_i} \le Y_i) = \sum[\min(X_i, Y_i)] - \sum \mathrm{Y} \qquad （13）$$

表 6 热点问题的单变量必要性分析

变量名称	一致性	覆盖率
事件类型	0.514292	0.792593
内容变量	0.740741	0.740741
细节信息变量	0.629630	0.739130
公众注意力	0.855556	0.837500

在群众留言文中筛选出来的 25 条留言中，通过单变量必要性分析发现：在四个变量中，仅有“公众注意力”，可以成为留言数据文本中的热点问题的充分条件，同时，其覆盖率 0.83，然而其余变量均低于 0.8，这也说明热点问题并非由单一条件变量决定的，因此，在必要条件下需进行条件变量组合分析去提取热点问题的必要条件：特定时间、特定地点、留言诸多且公众关注度较高的留言。

②条件变量组合分析

在完成上述操作后，可进一步进行原件变量组合分析，即分析不同的条件变量组合是否也对结果变量具有良好的解释力，同时也是通过抑制率和覆盖率指标来判断。

运行 fsQCA 3.0,可得到条件变量的组合分析结果（复杂解、中间解、简单解），其中，复杂解是完全遵循变量设置而得到的结果，因此，本文选取复杂解进行分析。结果显示：复杂解的整体一致性和整体覆盖率超过 0.9，表明复杂解的整体结果对于所选举的 25 个案例具有较强的解释力。

表 7 条件变量组合分析

序号	条件变量组合	原覆盖率	净覆盖率	一致性
1	~事件类型*内容变量*细节信息变量*公众注意力	0.111111	0.074074	1.000000
2	事件类型*~内容变量*细节信息变量*公众注意力	0.074074	0.0370370	1.000000
3	事件类型*内容变量*~细节信息变量*公众注意力	0.074074	0.074074	1.000000
4	事件类型*内容变量*细节信息变量*~公众注意力	0.037037	0.037037	1.000000

5	~事件类型*~内容变量*细节信息变量*公众注意力	0.074074	0.074074	1.000000
6	~事件类型*内容变量*~细节信息变量*公众注意力	0.148148	0.111111	1.000000
7	~事件类型*内容变量*细节信息变量*~公众注意力	0.074074	0.037037	1.000000
8	事件类型*~内容变量*~细节信息变量*~公众注意力	0.037037	0.037037	1.000000
9	事件类型*内容变量*~细节信息变量*~公众注意力	0.074074	0.074074	1.000000
10	事件类型*~内容变量*细节信息变量*~公众注意力	0.037037	0.037037	1.000000
11	~事件类型*~内容变量*~细节信息变量*公众注意力	0.037037	0.037037	1.000000
12	~事件类型*~内容变量*细节信息变量*~公众注意力	0.037037	0.037037	1.000000
整体覆盖率：0.95623		整体一致性：1.000000		

其中原覆盖率表示该条件变量组合能够解释的留言占总留言的比重，净覆盖率表示仅能被该条件变量组合所揭示的留言占总留言的比重，“*”表示“且”，即条件变量必须同时存在，“~”表示“非”，即表示该原因变量不存在。

据此，可以得出群众留言文本中的热点留言设置了三种围观条件变量组合：

组合一：~事件类型*内容变量*细节信息变量*公众注意力。该条件变量组合的内涵就是在进行留言文本数据分类时，对留言一级分类标签的缺失的情况下，根据文本内容，加上存在同类留言的叠加，同时细节信息变量的出现，将会引起公众的注意力，上述变量的组合将会使得热点留言更加显著。

组合二：~事件类型*~内容变量*细节信息变量*公众注意力。该条件变量组合指的是在仅有细节信息变量完善的情况下，公众注意力将会是直接得出热点留言的渠道。

组合三：~事件类型*内容变量*~细节信息变量*公众注意力。留言文本内容标签缺失，对留言文本内容进行分析并结合公众注意力去评定留言是否为热点留言。

2.2.3.3 层次分析法（APH 法）

APH 法是一种解决多目标的复杂问题的定性与定量相结合的决策分析方法。该方法将定量分析与定性分析结合起来，用决策者的经验判断各衡量目标能否实现的标准之间的相对重要程度，并合理地给出每个决策方案的每个标准的权数，利用权数求出各方案的优劣次序，比较有效地应用于那些难以用定量方法解决的问题[8]。

（1）建立层次结构模型

将决策的目标、考虑的因素（决策准则）和决策对象按它们之间的相互关系分为最高层、中间层和最低层，绘出层次结构图。

最高层（决策目的、要解决的问题）：将某一时段内反映特定地点或特定人群问题的留言进行归类，定义合理的热度评价指标，并给出评价结果，按格式给出排名前 5 的热点问题，还有具体留言信息。

最底层（决策时备选方案）：Ⅰ：从众多群众留言中识别出相似留言，并以此为依据进行热度分类

Ⅱ：按照特定留言时间进行归并分类

Ⅲ：定义合理的热度评价指标和计算方法，对指标进行排名（如群众对同一问题的反映程度）

中间层（考虑的因素、决策准则）：①地点、人群的识别（表达多样化）
②相似的计算复杂（特征对，两两之间计算相似计算量大）

：

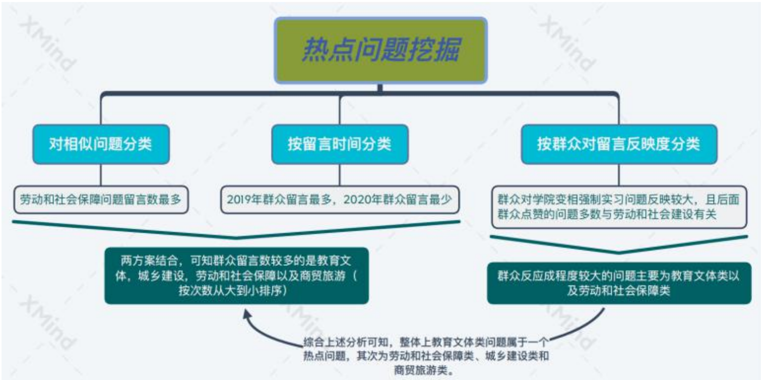


图 8 热点问题挖掘流程图

Ⅰ：对相似问题分类

留言时间	留言详情	一级分类
劳动和社会保障 计		104
	城乡建设 计数	101
	商贸旅游 计数	48
	教育文体 计数	96
	卫生计生 计数	58
	交通运输 计数	55
	环境保护 计数	33
	总计数	495

图 9 相似留言的分类情况

Ⅱ：按留言时间分类（以 1 年一个间隔）

	A	B	C	D	E	F
1	留言编号	留言用户	留言主题	留言时间	留言详情	一级分类
2	15525	A00034751	敬请您过问没有及时公布A市最低工资标准信息的现象	2011/10/1 10:16:20	份，那是公布最低工资	劳动和社会保障
3	26115	A00080334	A市泉塘的道路绿化问题需要重视	2011/10/26 21:41:35	和盼盼路的樟树和其他	城乡建设
4	110243	A00058711	提高导游的素质，让来K8县旅游者有尊严	2011/11/13 18:40:38	我们要凡是来K8县的旅	商贸旅游
5	103534	A00058711	K市的文化要超强经济的发展速度	2011/11/14 22:28:36	四抓人才，发展	教育文体
6	46855	A000103488	请问违反计划生育的人员是否再就业的权力也一并取消了？	2011/11/15 8:04:47	教，违反计划生育的人	劳动和社会保障
7	15471	A0008749	对A市出租车管理的建议	2011/11/18 15:48:58	本地人还如此，那么外	交通运输
8	15463	A00082886	A市楚华宾馆后棚户改造建议	2011/11/23 14:53:43	有改造拆迁计划？有的	城乡建设
9	44078	A00036344	C市C2区教师的社保基金哪里去了？	2011/12/15 19:39:15	这么多的教师近几年的	劳动和社会保障
10	123778	A00088194	提高L市民整体素质，比如免费职业培训班	2011/5/2 15:00:19	我渴求学习……增加社	劳动和社会保障
11	166015	A00048447	西地省公共卫生拨的款去哪了？	2011/5/22 20:21:31	贴，他们却跑路油钱都	卫生计生
12	67317	A000108720	政府能否提高工资福利	2011/5/24 10:20:22	年的财政报告都是增长	劳动和社会保障
13	13641	A00051719	退休政策调整的问题有话向市长说	2011/5/28 20:47:17	而社保局又遇政策调整	劳动和社会保障
14	166008	A00029725	乡村医生公共卫生政府补贴钱不应被拖欠	2011/5/31 21:19:07	每个卫生室就领了五六	卫生计生
15	103704	A00057207	蓝宁道路新加工贸易走廊该如何搞？	2011/6/15 12:06:43	加工贸易走廊到底怎么	商贸旅游
16	125278	A00054893	由市长大人对L市发展建议引发的几点补充建议	2011/7/16 10:26:47	小组，建议选用年纪稍	教育文体
17	160378	A00085164	关于优秀运动员退役安置办法的补偿规定	2011/8/17 12:24:25	介，再说一个运动员在	教育文体
18	48251	A00037114	请问工资领据和工作证明是证明不存在劳动关系的吗？	2011/9/28 16:21:56	劳动者和用人单位不存	劳动和社会保障
19	183987	A000111488	I市食品安全心系民生，常抓不懈不容忽视	2011/9/28 8:35:16	我们电话询问疾控中心	卫生计生
20	37042	A00034667	望B市职工医疗保险费能在网上查询	2012/1/19 17:49:30	金那样，让人民能在网	城乡建设
21	37042	A00034667	望B市职工医疗保险费能在网上查询	2012/1/19 17:49:30	金那样，让人民能在网	劳动和社会保障
22	110216	A00055421	K8县九亿广场的公厕要安装照明灯	2012/10/10 16:50:40	。黑黑的，在晚就在	城乡建设
23	134484	A000103696	L4县的经济寻求发展请从消除贸易壁垒开始	2012/10/17 17:17:22	已从事的行业每年各个	商贸旅游
24	59487	A0006465	请帮我解决晨七宝山煤矿因职业病退休工人的抚恤金问题	2012/10/19 20:30:44	金，已办手续，但单位	劳动和社会保障
25	13026	A00028291	60年代出生的享受独生子女待遇安置，这合理吗？	2012/10/23 10:59:18	15平米建房，拆迁安	卫生计生
26	180378	A00087073	还我尊严，给我们一个自食其力的机会！	2012/10/28 20:15:31	，国家号召全社会帮	劳动和社会保障
27	21066	A000103913	关于新农合医保不合理的追问	2012/11/20 9:38:14	刑满释放、新生儿等	劳动和社会保障
28	37107	A00014157	B9市卫生局和药品监督局乱收费，请邓局明查！	2012/12/15 18:39:45	的前提下，给全体乡	卫生计生
29	89933	A00056518	1813路有人私建收费站	2012/12/16 15:35:29	到J市J2区桥口镇下渡	交通运输
30	174438	A00016633	为何日月星城KTV扰民这么久，相关部门却不闻不问？	2012/12/16 3:39:15	果，每次环保部门来	环境保护

图 10 按留言时间的分类情况

Ⅲ：按群众对留言认可度即公众注意力进行分类（即群众对留言的点赞数）

留言编号	留言用户	留言主题	留言时间	留言详情	点赞数	反对数
360114	A0182491	A市经济学院体育学院变相强制实习	2017-06-08 17:31:20	公司签了合同，并且公司也要和我们	9	0
360108	A0283523	A5区劳动东路魅力之城小区一楼的夜宵摊严重污染附近的空气，急需处理！	2019-08-01 16:20:02	是得要维护社会和谐稳定，合法维	6	0
360101	A324156	A5区劳动东路魅力之城小区油烟扰民	2019-07-28 12:49:18	清洗也没有，每天油烟直排，要死	4	0
360107	A0283523	A5区劳动东路魅力之城小区一楼的夜宵摊严重污染附近的空气	2019-07-21 10:29:36	是得要维护社会和谐稳定，合法维	3	0
286572	A23525	请求A市地铁2#线在梅溪湖CBD处增设一个站	2018-10-27 15:13:26	桃花岭和梅溪湖，傍晚半2#线间	3	0
360100	A324156	魅力之城小区临街门面油烟直排扰民	2019-09-05 12:29:01	人，一天24小时都是烟，请政府关	3	0
360113	A3352352	A市经济学院强制学生外出实习	2018-05-17 08:32:04	们都不知道！学校很小但是这几年	3	0
353426	A0098773	J4县供销合作社在岗失业职工追缴社保	2020-01-06 10:20:31	中国人民共和国宪法》和《劳动法》	2	0
360103	A0012425	A5区劳动东路魅力之城小区临街门面烧烤夜宵摊	2019-09-25 00:31:33	烧烤夜宵更加扰民，油烟24小时	1	0
360105	A120356	A5区魅力之城小区一楼被搞成商业门面，噪音扰民严重	2019-08-26 08:33:03	，影响我们的晚年生活，架空层被	1	0
321736	A0092521	A市能不能提高医疗门诊报销比例	2019-06-12 08:23:01	又不让，小孩体弱多病各种开支，	1	0
360111	A1204455	A市经济学院组织学生外出打工合理吗？	2019-11-05 10:31:38	十个小时以上，（晚班时间是20:	1	0
289408	A0012413	在A市人才app上申请购房补贴为什么通不过	2018-11-15 16:07:12	我也购买了灵活就业人员养老保险	0	0
336608	A0005623	希望西地省把抗癌药品纳入医保范围	2019-09-08 21:01:59	贵，明明有药可治，就是买不起药	0	0
323149	A1241141	请给K3县乡村医生发卫生室执业许可证	2019-06-20 20:38:47	新村的是证件下来啊，有些老村医	0	0
343955	A108051	A市能否设立南塘城轨公交站？	2019-10-31 21:19:59	之区，南塘小学，A市一中城南中学	0	0
316619	A235259	请问A市什么时候能普及5G网络？	2019-05-14 11:22:13	全程盖的城区建成，看消息介绍，	0	0
360110	A110021	A市经济学院寒假过年期间组织学生去工厂工作	2019-11-22 14:42:14	多难过！虽说不是强制性的，但不	0	0
323034	A012414	L市物业服务收费标准应考虑居民的经济承受能力	2019-06-19 17:46:24	场垃圾清运费按不超过30元/平方	0	0
319659	A023956	A市江山帝景新房有严重安全隐患	2019-05-30 17:34:02	，雨雪天气后过道全部是水和雪，	0	0
360106	A235367	A市魅力之城小区底层商铺营业到凌晨，各种噪音好痛苦	2019-08-26 01:50:38	不说还严重影响健康，大家对此	0	0
313964	A108906	12123上申请驾驶证期满换证，一个星期了都无人受理	2019-04-26 15:28:42	申请的，都快一个星期了都没人受	0	0
337458	A078325	能否分层单独补交超面积房款？	2019-09-16 18:45:29	地，得知该房占地属划拨用地面	0	0
360104	A012417	A市魅力之城商铺无排烟管道，小区内到处油烟味	2019-08-18 14:44:00	出都搞得业主一身油烟味，而且	0	0
360109	A0080252	万科魅力之城小区底层门店深夜经营，各种噪音扰民	2019-09-04 21:00:18	着吆喝声，划酒声，炒菜烧烤的	0	0
304503	A011244	A市什么时候能实行独生子女护理假？	2019-02-26 15:22:03	生子女多，照顾四个老人真的力不从	0	0
360112	A220235	A市经济学院强制学生实习	2019-04-28 17:32:51	求学生必须去学校安排的几个点实	0	0
360102	A1234140	A5区劳动东路魅力之城小区底层餐馆油烟扰民	2019-09-10 06:13:27	进屋内，窗户长期不能打开，晚上	0	0
351074	A001241415	对A市参保记录的几点疑问	2019-12-19 17:46:04	字，真是讽刺，以上，我再问A市	0	0

图 11 按公众注意力分类的留言情况

2.2.4 基于问题三的答复意见评价方案

针对附件 4 的 2817 条相关部门对留言的答复意见，从留言文本标签分类，提取留言文本及答复意见的关键词，运用最优化原理从答复的相关性、完整度、可解释度等角度，建立一套答复意见的评价方案。

2. 2. 4. 1 最优化原理[9]（最优子结构性质）

最优性原理是指多阶段决策过程的最优决策序列具有这样的性质：不论初始状态和初始决策如何，对于前面决策所造成的某一状态而言，其后各阶段的决策序列必须构成最优策略，简而言之，一个最优化策略的子策略总是最优的。一个问题满足最优化原理又称其具有最优子结构性质，是动态规划的基础。

在实际生活中，寻找最优解的问题时常出现，动态规划帮助我们寻找解决问题的最佳方案。然而，它在为我们提供便利的同时也给我们带来了极大的考验。本文采用动态规划算法将待求解问题划分成若干个子问题分别求解，最终得到原问题的最优答案，以寻求留言答复的最优方案。

2. 2. 4. 2 动态规划算法[9]

动态规划算法的基本思想是将待求解问题划分成若干个相互关联的子问题，通过对子问题的求解以自底向上的方式计算出最优值。动态规划算法通过提高程序的空间复杂度来达到降低时间复杂度的目的。由此可见，使用动态规划算法进行求解的问题具有两个性质，即最优子结构性质和重叠子结构性质。

最优子结构性质：即待求解问题的最优解包含了其子问题的最优解，因此，通过求解子问题的最优解可逐步构成原问题的最优解。

重叠子问题性质：即求解由原问题分解出的子问题后，其结果可用于求解其他的子问题，做到对每一个子问题只求解一次，避免产生不必要的重复计算，以获得更高的解题效率。

对于动态规划算法的具体实现，可分为以下步骤：

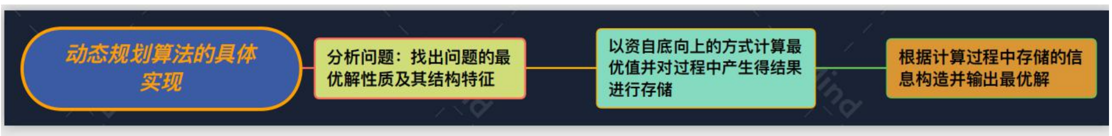


图 12 动态规划算法流程

①贝尔曼递推方程

设 n 维离散系统为 $x(k+1)=a[x(k),u(k)], k=0,1,...,N-1$,要求确定最优控制，

使性能指标 $J(u)=h[x(N)]+\sum_{k=0}^{N-1}g_k[x(k),u(k)]$ 最小。式中,N 是固定的; $u(k)$ 不受限制（或） $u(k)\in U$ 。

记 J_N 为达到终端状态 x(k)的末级性能指标，即有 $J_N=h[x(N)]$.当 $k=N-1$ 时，得到 $J_{N-1}=J_N+[x(N-1),u(N-1)]$ 。

由于控制变量 $u(N-1)$ 对当前状态想 $x(N-1)$ 无作用，只能改变后续状态变量 $x(N)$ 的值，而且控制变量必然是当前状态的函数，记最优控制得到的性能指标为 $J_k^*[x(k)]=\min_{u(k)}\{J_{k+1}^*[a(x(k),u(k))]+g_k[x(k),u(k)]\}$ 。

上式称为贝尔曼递推方程，它是动态规划的基本递推关系式。由已知的 k+1 级过程的最优性能指标 $J_k^*[x(k+1)]$，根据递推公式确定最优控制，就可得到 k 级过程的最优性能指标 $J_k^*[x(k)]$。

②计算最优值

采用二维数组 length 记录最长公共子序列的长度，使用二维数组 s 记录两个字符串的公共元素。当两个字符相同时，即 x[i]=y[j]时，其最长公共子序列长度为 length[i-1][j-1]加 1，表示该元素为相同元素。当两个字符不相同时，需判断 length[i-1][j]与 length[i-1][j-1]的大小关系，选取较大的数作为 length[i][j]的值，并将 s[i][j]置为相应的值;若 length[i-1][j]>length[i][j-1]，将 s[i][j]的值置为 2，否则置为 3。使用伪代码实现如下。

```
for( int i = 1; i <= x.length( ); i++)
length[i][0] = 0; //当 y 序列为空时 ,最长公共子序
列的长度为 0
for( int i = 1; i <= y.length ( ); i++)
length[0][i] = 0; //当 x 序列为空时 ,最长公共子序列
长度为 0
for( int i = 1; i <= x.length ( ); i++){
for( int j = 1; j <= y.length ( ); j++){
if( x[i] == y[j]){
//x 中字符与 y 中字符相同时 ,最长公共子序列长度增
加 1 ,该字符是子序列中的元素 ,将标记数组的值置为 1
length[i][j] = length[i-1][j-1]+1; s[i][j] = 1;
}else{
//该字符不是最长公共子序列中的元素时 ,根据其具体
情况置为不同的标记
if( length[i-1][j] > length[i][j-1]){
length[i][j] = length[i-1][j]; s[i][j] = 2;
}else{
length[i][j] = length[i][j-1]; s[i][j] = 3;
}
}
}
}
```

图 13 计算最优值得代码

③构造最长公共子序列

最优值计算完成后，根据计算过程中存储的结果构造最优解。采用递归的方式从右下到左上遍历标记数组 s，寻找最长公共子序列并输出。从数组的最后一个元素（即 s[m][n]）开始进行判断:若其值为 1，则表明该元素是最长公共子序列中的元素，将其行数和列数分别减 1 进行下一次判断;否则该元素不是最长公共子序列中的元素，若其值为 2 则表示 length[i-1][j]>length[i][j-1]，需要将其行数减 1 继续进行判断;若其值为 3，需将其列数减 1 继续进行判断。采用递归方法构造如下。

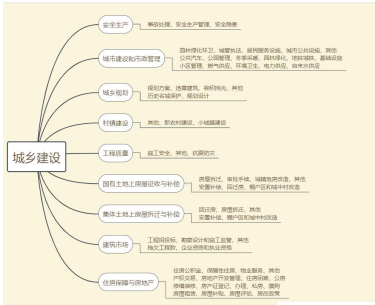
```
voidGouZao( int m ,int n ,char x [],int s [][]) //构造递
归函数
{
if( m = = 0||n = = 0) return 0; //判断完成 ,结束递归
过程
if( s [m][n] = = 1){
//是最长公共子序列中的元素时 ,打印输出该元素
print( "%c" ,x [i ]);
GouZao( m -1 ,n -1 ,x ,s); //递归判断其他元素
}else{
//该元素不是子序列中元素 ,根据标记值将 m 或 n 值减
1 继续进行判断
if( s [m][n] = = 2){
GouZao( m -1 ,n ,x ,s};
}else{
GouZao( m ,n -1 ,x ,s};
}
}
}
```

图 14 构造最长公共子序列代码

采用动态规划算法求解最优解问题，将原留言问题分解为若干个具有相互关联关系的子问题，对每一个子问题只进行一次运算，并存储其结果用以构造原问题的最优解，避免了大量的重复计算，降低了程序的时间复杂度，在提高程序效率的基础上较为准确的得出问题的最终答案，为解决问题提供了极大的便利。因此，在解决实际问题中，动态规划算法起到了极为重要的作用。

2.2.4.3 层次分析法

通过层次分析法将不同的一级分类标签与二级分类标签和三级分类标签进行匹配汇总，层次分析结果如下：



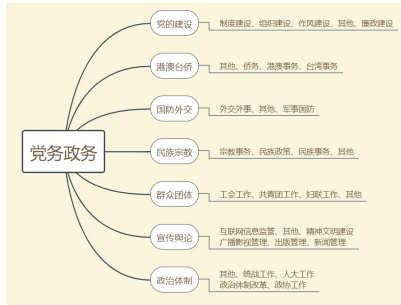


图 15、16 城乡建设和党务政务的各级分类标签

图 17、18 国土资源和卫生计生的各级分类标签

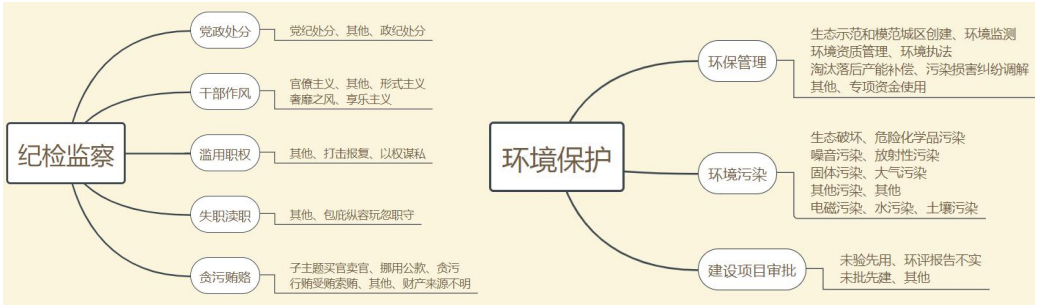


图 19、20 纪检监察和环境保护的各级分类标签

图 21、22 交通运输和经济管理的各级分类标签

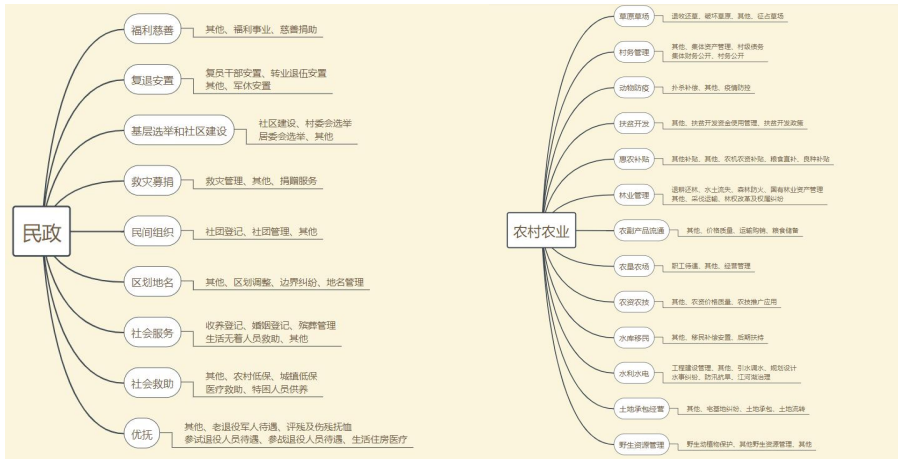


图 23、24 民政和农村农业的各级分类标签

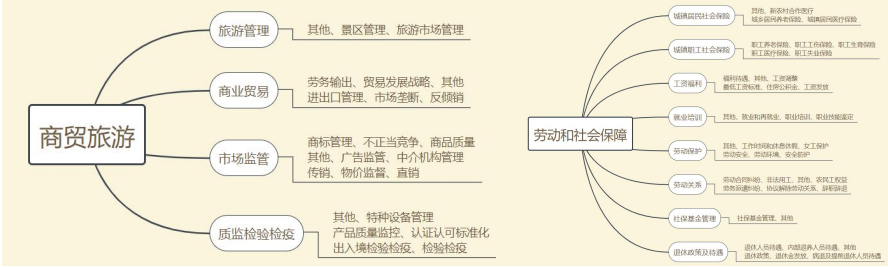


图 25、26 国贸旅游和劳动和社会保障的各级分类标签

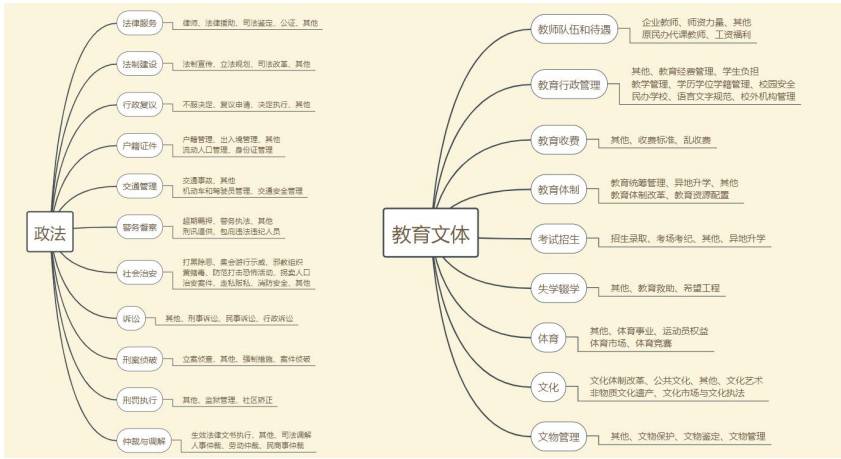


图 27、28 政法和教育文体的各级分类标签

结合以上以及分类标签的层次分析拟定为一套关于留言答复的评价方案、

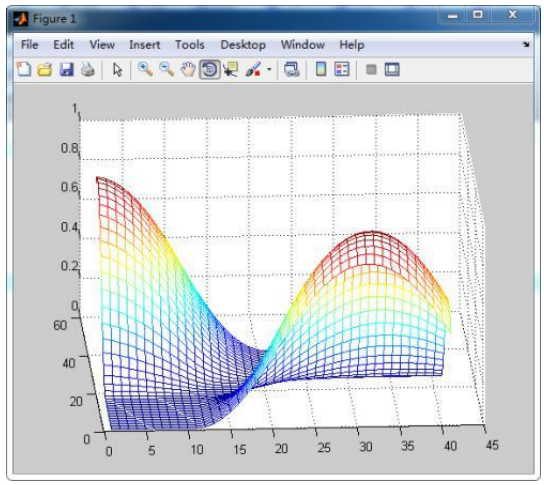


图 29 留言与答复关键词的拟合情况

通过在 MATLAB 上进行留言与答复的关键词一级分类标签的拟合分析，由于每一条答复意见均具有相关问题的处理方案，可知留言与答复具有相关性、完整性和可解释性。

2.3 分析结果

2.3.1 基于问题一的词云图模型

通过对附件 2 的群众留言文本的预处理，进行数据清洗、文本去重、停用词过滤、结巴分词、词云图绘制、TF-IDF 算法、贝叶斯算法、F-Score 评价的操作，

通过词云图建立关于留言文本内容的一级标签分类模型，结果如下表所示：

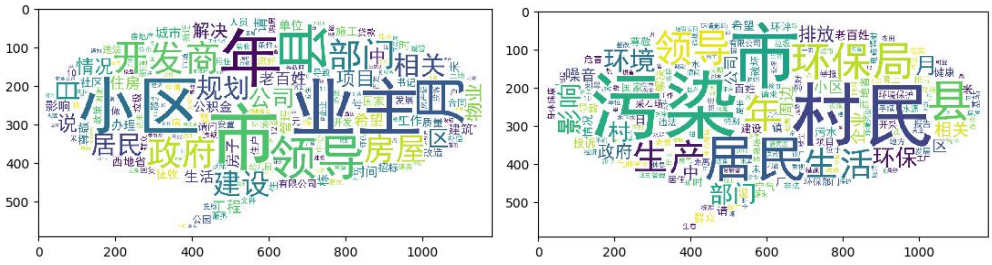


图 30、31 城乡建设和环境保护的词云图

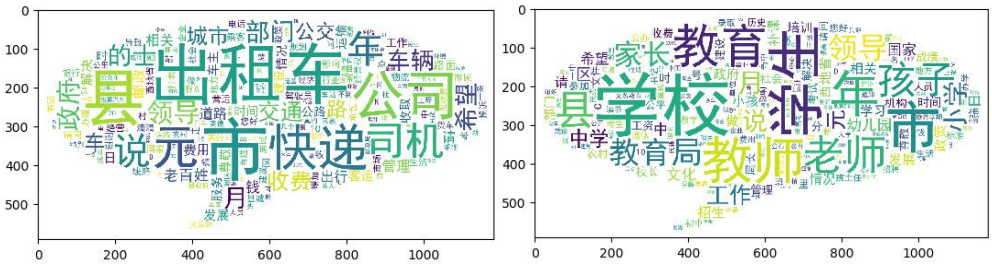


图 32、33 交通运输和教育文体的词云图

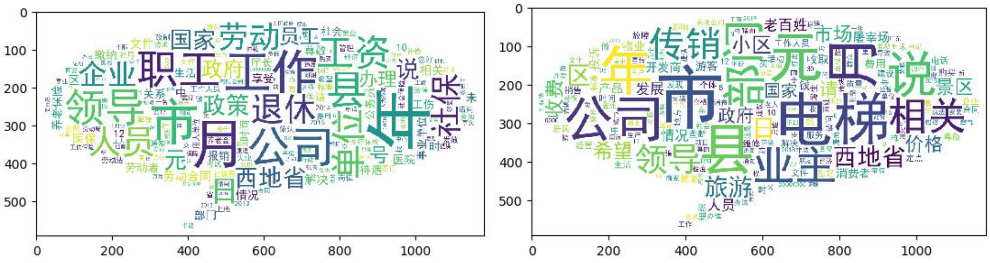


图 34、35 劳动和社会保障和商贸旅游的词云图

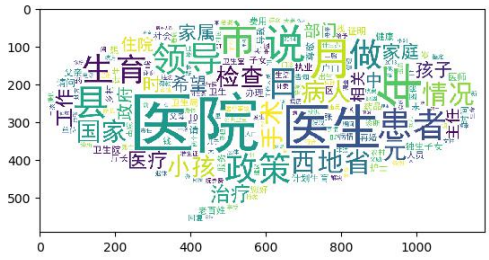


图 36 卫生计生词云图

通过对附件 2 中的 9211 条群众留言文本建立的词云图模型，通过 F-score 对其分类方法进行评价：

$$F_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{2P_iR_i}{P_i+R_i} \tag{14}$$

$$F_1 = \frac{1}{7}*(0.771+0.778+0.762+0.798+0.890+0.781+0.783) \approx 0.795$$

2.3.2 基于问题二的定性分析法和层次分析法

对于附件 3 的留言文本数据进行定性分析法（fsQCA），通过单变量必要性

分析、条件变量组合分析，根据一致性指标和覆盖率指标来对定性分析法的结果进行分析，得出三项条件变量组合：~事件类型*内容变量*细节信息变量*公众注意力；~事件类型*~内容变量*细节信息变量*公众注意力；~事件类型*内容变量*~细节信息变量*公众注意力。

表 8 排名前 5 的热点问题

热度排名	问题 ID	热度指数	时间范围	地点/人群	问题描述
1	1	45	2019-11-02 14:23:11 到 2020-01-26 19:47:11	A2 区丽发新城	A2 区丽发新城附近建搅拌站噪音扰民
2	2	40	2019/7/7 7:28:06 到 2019/8/31 6:33:25	A 市伊景园滨河苑	A 市伊景园滨河苑捆绑车位销售
3	3	21	2019-07-21 10:29:36 到 2019/12/4 16:25:06	A 市 A5 区魅力之城小区	小区油烟扰民
4	4	14	2019/2/25 23:24:54 到 2019/11/27 23:14:33	A 市涉外经济学院生	学校强制学生去定点企业实习
5	5	11	2019/1/6 11:02:20 到 2019/5/22 10:05:15	A 市辉煌国际城	A 市辉煌国际城二期商铺非法营业

2.3.3 基于问题三的最优化原理的评价方案模型

针对附件 4 的 2817 条留言内容及相关部门的答复，分别求解出留言及答复所属分类标签，并运用最优化原理，对文本数据进行动态规划算法以及层次分析，得出关于答复意见质量的评价方案。

表 9 评价方案分析

角度	评价说明
相关性	每条留言以及答复意见的关键词提取，拟合两者关键词的分类标签，结果是一致的
完整性	答复意见含有相关留言问题的解决方案
可解释性	对于留言内容所反馈的问题能够给予对应的回复

3. 结论

群众留言处理是建设“智慧政务”的重要一步，在研究探讨过程中，必须对留言文本进行预处理、找出热点问题，并对群众的留言的相关性、完整性、可解释性等角度进行答复分析，才能实现高效便捷智能化的“智慧政务”效果。

由于在进行答复的过程中影响答复效果的因素有很多，数据较为庞大，所以为了后续的研究，必须对留言数据进行文本预处理，结合定性分析法和层次分析法提取热点问题，排除无价值的文本内容的干扰，将属性减少，减少误差并降至最低以达到最优化。

基于文本的预处理，我们对附件 2,、附件 3 和附件 4 中的留言文本均进行一级标签分类，包含着数据清洗、文本去重、结巴分词、停用词过滤、TF-IDF

算法等步骤，得出一级分类标签的关键词词云图模型，根据 F-Score 对分类模型进行评价，F-Score 越大，分类情况更加精确。

基于对热点问题的定义，分析在诸多留言信息中词频较多且反馈问题较多的 25 条留言，通过一致性指标及覆盖率指标进行定性分析法以及层次分析法，探测出特定时间特定地点/人物的留言，得出排名前 5 名的热点问题。

对留言进行答复，首先在基于问题一和问题二的解决问题的方向上，通过最优化原理建立动态规划算法，依据各级留言分类标签的层次分析法，匹配关键词并拟合文本进行答复。

无论如何，群众留言问题处理不仅能快速的发现并解决人们所反映的问题，而且对于建设“智慧政务”具有重大的作用。同时人民群众应该关注如何留言才能更加简洁明了的说明自己所要反馈的问题，使政府工作人员一眼就能看出所要反馈的信息，政府工作人员也应关注提高自身修养和综合实力以便解决群众问题，这是今后人们所共同奋斗努力的。

4. 参考文献

[1]崔志刚.基于电商网站商品评论数据的用户情感分析[D].北京:北京交通大学，2014.
[2]朱少杰.基于深度学习的文本情感分类研究[D].哈尔滨:哈尔滨工业大学，2014.
[3]秦彩杰,管强.一种基于 F-Score 的特征选择方法[J].宜宾学院学报.2018,18(6):4-8.
[4]时志芳.移动投诉信息中热点问题的自动发现与分析[D].北京:北京邮电大学，2013.
[5]张帆.贝叶斯算法在校园留言板垃圾过滤中的应用研究[D].郑州大学.2016.
[6]吴海荣,罗月颖.健康类谣言高热度传播的组合因素影响研究——基于清晰集定性比较分析.文化与传播[J].2019,2:91-99.
[7]闫鹏,郑雪峰,朱建勇,等.一种基于嵌入式特征选择的垃圾邮件过滤模型[J]. 小型微型计算机系统,2009,30(8):1616-1620.
[8] 王晓陵,陆军主编. 最优化方法与最优控制[M]. 哈尔滨工程大学出版社，2016.
[9]何思瑶,沈樾,辛琰钰.动态规划算法求最优解问题[J].电声科技,2019，43(3)：42-44