

“智慧政务”中的文本挖掘应用

摘要

由于互联网技术的快速发展,智慧政务系统中积累了海量的业务信息。为适应移动端碎片化的使用场景,以微信、微博、市长信箱、阳光热线等形式出现的短文本政务内容信息成为文本数据在互联网内容中的主要呈现形式。面对着日益庞大的政务信息量,迫切需要应用人工智能的相关技术,对数据进行深层次的分析并研究各类信息的规律和关系,以更好地提高政府的管理水平和施政效率,促进官民生活和谐发展。19 大后,公民与政府在网络上的互动日益频繁,网络问政平台上政府与群众互动不仅受到网络渠道特征的影响,还受到政府国际关系条块因素的影响。因此,将数据挖掘技术有效地应用于各类社情民意分析是目前政府工作的迫切需要。文本挖掘技术是近几年来数据挖掘领域的一个新兴的分支。而在海量政务信息中,除了规范化程度很强的数据库数据外,还有大量的政务叙述性文本描述。对这些大量的政务文本进行相关文本挖掘技术研究和应用是非常有意义的。

本文主要针对政府部门中大量的政务文本信息,对其中的政务文本针对不同业务问题构建相应挖掘模型方案。本文的研究问题包括以下几点:

(1) 在群众留言分类问题方面。结合实际应用需要,对政务业务中的一些术语建立专业词库;同时针对政务文本的特征,提出了具有针对性的特殊预处理方法;根据关键特征对政务文本进行 $TF-IDF$ 方法计算,分别利用贝叶斯算法、基于 $TF-IDF$ 的贝叶斯模型和基于 $TF-IDF$ 的支持向量机 (SVM) 模型来构建留言内容的一级标签分类模型。

(2) 在热点问题挖掘方面。在留言内容的一级标签分类模型和政务属性信息抽取方法相结合基础上,针对附件 3 政务文本的特征先进行一级标签的分类,根据分类类别再利用 K-means 聚类分析算法对相近类别进行聚类,以聚焦热点问题。其中 K-means 聚类中类别 k , 利用层次聚类的计算方法进行确定。

(3) 在答复意见评价问题方面。网络问政平台的涌现带来了政府与群众互动模式的革新,以附件 4 为研究对象,通过群众留言和政府回复的内容在及时性、相关性、完整性、可解释性角度进行分析,实现了一套评价方案。

关键词: 文本挖掘; 支持向量机; K-means 聚类分析; 网络问政

Abstract

Due to the rapid development of Internet technology, massive business information has accumulated in the smart government system. In order to adapt to the fragmented usage scenarios of mobile terminals, short-text government content

information in the form of WeChat, Weibo, mayor's mailbox, and sunshine hotline has become the main form of text data in Internet content. Faced with the increasing amount of government information, there is an urgent need to apply artificial intelligence related technologies, perform in-depth analysis of data and study the laws and relationships of various types of information to better improve the government's management level and governance efficiency. Harmonious development of official and civilian life. After the 19th National Congress of the Communist Party of China, the interaction between citizens and the government on the Internet became more and more frequent. The interaction between the government and the masses on the Internet platform was not only affected by the characteristics of the network channels, but also by the factors of the government's international relations. Therefore, the application of data mining technology to the analysis of various social conditions and public opinion is an urgent need for government work. Text mining technology is an emerging branch in the field of data mining in recent years. In the massive government information, in addition to the database data with a strong degree of standardization, there is also a large amount of narrative text description of government affairs. It is very meaningful to conduct research and application of relevant text mining technology on these large amounts of government affairs texts.

This article mainly aims at a large amount of government affairs text information in government departments, and constructs corresponding mining model solutions for different business problems in the government affairs text. The research questions in this article include the following:

(1) The problem of classification of messages in the masses. Combined with actual application needs, a professional vocabulary was established for some terms in the government affairs business; at the same time, a special special pre-processing method was proposed for the characteristics of the government affairs text; the TF-IDF method was used to calculate the government affairs text according to the key features, using support Vector machine (SVM) algorithm to build a first-level label classification model of message content.

(2) In mining hot issues. Based on the combination of the first-level label classification model of the message content and the government attribute information extraction method, the first-level label is classified according to the characteristics of the government text in Annex 3, and the K-means clustering analysis algorithm is used to classify similar categories according to the classification category. Perform

clustering to focus on hot issues. Among them, the category k in K-means clustering is determined by the calculation method of hierarchical clustering.

(3) In response to comments and evaluation questions. The emergence of the online questioning platform has brought about the innovation of the interaction model between the government and the masses. Taking Annex 4 as the research object, the content of the mass messages and government replies are analyzed from the perspective of timeliness, relevance, completeness, and interpretability to achieve A set of evaluation plans.

Keywords: text mining technology; support vector machine; K-means cluster analysis; network politics

目录

1. 研究背景..... 1

2. 挖掘目标及问题分析..... 1

 2.1 挖掘目标..... 1

 2.2 问题分析..... 2

3. 数据预处理..... 3

 3.1 分词..... 3

 3.2 去停用词..... 5

4. 群众留言分类模型..... 7

 4.1 文本数据分析..... 7

 4.2 群众留言分类..... 8

 4.3 分类模型评估..... 9

5. 热点问题挖掘模型..... 10

 5.1 文本数据分析..... 11

 5.2 聚类热点问题挖掘..... 15

 5.3 聚类性能指标分析..... 19

6. 针对答复意见建立评价方案..... 19

 6.1 文本数据分析..... 19

 6.2 数据研究..... 24

 6.3 评价方案..... 28

7. 总结..... 28

参考文献..... 28

附录..... 29

1. 研究背景

近年来,随着经济的发展和信息技术的深入应用,智慧政务系统中积累了海量业务信息。其中微信、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短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。

文本挖掘(Text Mining)作为近几年来数据挖掘领域的一个新兴的分支,涵盖多种技术,包括信息抽取、信息检索、自然语言处理和数据挖掘技术。主要着力于帮助用户从来源于 web 或者数据库中的具有丰富语义的文本数据进行结构化分析。文本挖掘已经逐步成为数据挖掘任务中一个日益流行而重要的研究领域,其大多数任务的目标是对数据集的文本内容进行一些列的操作,这些操作步骤通常包括文本预处理、特征提取、结构分析、文本摘要、文本分类、文本聚类等。

目前,文本挖掘在已经在很多领域中得到了成功的应用,例如在科技情报、信息检索、互联网信息统计、专利文献信息、医学等领域都已经得到了广泛的应用。面对海量的政务文本信息,文本挖掘技术是非常有用的技术,它能从这些海量的政务文本信息中挖掘中隐藏的、对政府部门工作人员有用的信息。例如:通过文本分类技术从群众留言信息中能够按照一定的划分体系进行政务信息的分类,以便职能部门任务的分派,进而提高信息处理效率;通过文本聚类技术能够挖掘出热点问题,有助于相关部门进行有针对性地处理,提升服务效率;通过信息抽取技术能够从文本中自动抽取时间、地点、事件等信息。因此,文本挖掘能够为政府业务提供有效的决策支持,不仅能提高各类社情民意信息分析的质量和效率,还能有效支持智慧政务系统从大量群众留言记录中有效和快速地进行留言分类、发现留言趋势等,为职能人员更好地降低业务工作量,提高工作效能。

2. 挖掘目标及问题分析

2.1 挖掘目标

根据群众问政留言记录和政府提供的政务划分体系,其中留言记录主要包括留言主题和留言详情,挖掘能够代表政府数据的独特特征的文本构成专业特征词

库，以此构建留言内容的一级标签分类模型。

结合留言内容的一级标签分类模型，针对留言文本信息利用数据挖掘技术挖掘出热点问题及相应热点对应的留言信息。

结合留言内容的一级标签分类模型和热点问题，从答复的相关性、完整性、可解释性、及时性等角度针对留言答复意见建立一套评价标准。

2.2 问题分析

在文本挖掘的过程中, 机器学习、模式识别、数据库、自然语言处理甚至语言学等领域的技术会被利用从特定的非结构化文本中抽取特定的信息, 之后生成的结构化信息被存储到数据库中以便在后续的应用中被查询和使用。文本挖掘的流程框架示意图如 1 所示。

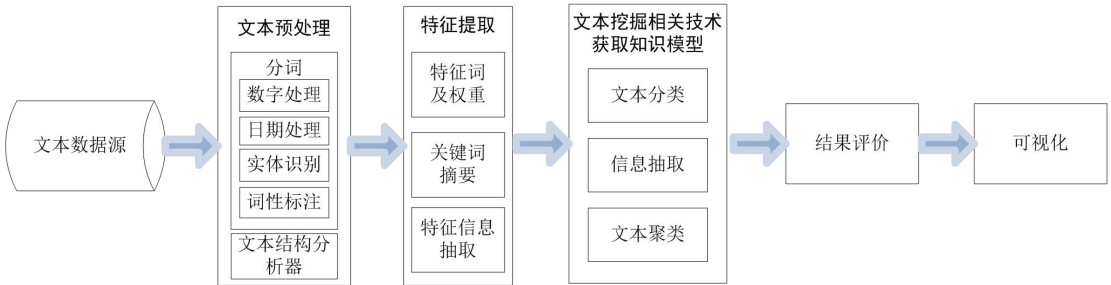


图 1 文本挖掘处理流程示意图

2.2.1 问题 1 分析

根据已知群众留言数据和级别划分体系对留言进行分类，建立关于留言内容的一级标签分类模型。附件 1 中给出了一种内容分类的三级标签体系，附件 2 给出了互联网上公开的群众留言信息，包括留言编号、留言用户、留言主题、留言时间、留言详情和一级分类。结合附件 1 的二级分类和三级分类构建的语料库，对附件 2 中留言主题和留言详情文本特征进行分析，得到能够代表留言信息独特特征的统计量形成特征矢量，以此构建语料库。基于有效特征词库建立留言内容的一级标签分类模型。

2.2.2 问题 2 分析

根据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。附件 3 中给出了互联网上公开的群众留言信息。首先，建立针对附件 2 中留言主题和留言详情特征矢量的统计量提取关键语料

库，根据问题 1 模型对附件 2 数据进行分类；其次提取“留言主题”属性文本对每个分类类别的语料库进行相似度匹配，不同类别的的匹配群为不同的热点问题。该问题的难点在于，运用了 K-means 聚类的相似度匹配算法，该算法中的 k 值难以确定，尤其是针对不同分类类别，每一种类别的 k 值也不相同。为此，在构建聚类模型时利用层次聚类算法进行 k 值的自动确定。

2.2.3 问题 3 分析

根据相关部门对留言的答复意见，从答复的及时性、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，并尝试实现。附件 4 中给出了群众留言信息及答复意见数据，运用问题 2 中的方法对附件 4 中的数据进行分类，并对每类数据从答复的及时性、相关性、完整性、可解释性等角度进行分析；运用问题 3 的方法将附件 4 的数据分为热点问题和非热点问题，将这两大类数据也从答复的及时性、相关性、完整性、可解释性等角度进行分析，在上述方法技术上建立评价分析报告。

3. 数据预处理

3.1 分词

本文中的“留言主题”和“留言详情”多以成句、成段的形式出现, 这种形式难以被计算机直接学习与利用, 利用 jieba 分词对中文文本的字符串进行匹配和词频统计，进而将文本拆分为可被识别与学习的具有独立意义的词语。例如附件 2 数据，如表 1 前十项所示。

表 1 jieba 分词表

序号	词量	留言主题+留言详情
0	75	['A', '市', '西湖', '建筑', '集团', '占', '道', '施工', '有', '安全隐患', '。', 'A3', '区', '大道', '西行', '便', '道', '未管', '所', '路口', '至', '加油站', '路段', '人行道', '包括', '路灯', '杆', '被', '圈', '西湖', '建筑', '集团', '燕子', '山', '安置', '房', '项目',]
1	61	['A', '市', '在水一方', '大厦', '人为', '烂尾', '多年', '安全隐患', '严重', '位于', '书院', '路', '主干道', '的', '在水一方', '大厦', '一楼', '至', '四楼', '人为', '拆除', '水', '电等', '设施', '后', '烂尾', '多年', '用', '护栏', '围着', '不但', '占用',]
2	617	['投诉', 'A', '市', 'A1', '区苑', '物业', '违规', '收', '停车费', '。', ' ']

		尊敬', '的', '领导', ':', 'A1', '区苑', '小区', '位于', 'A1', '区', '火炬', '路', '小区', '物业', 'A', '市程明', '物业管理', '有限公司', '未经', '小区业主', '同意', '利用', '业主', '的', '公摊',]
3	87	['A1', '区', '蔡锷', '南路', 'A2', '区华庭', '楼顶', '水箱', '长年', '不洗', '。', 'A1', '区', 'A2', '区华庭', '小区', '高层', '为', '二次', '供水', '楼', '水箱', '长年', '不洗', '现在', '自来水', '龙头', '的', '水', '严重', '霉味', '大家', '都', '知道',]
4	85	['A1', '区', 'A2', '区华庭', '自来水', '好大', '一股', '霉味', '。', 'A1', '区', 'A2', '区华庭', '小区', '高层', '为', '二次', '供水', '楼', '水箱', '长年', '不洗', '现在', '自来水', '龙头', '的', '水', '严重', '霉味', '大家', '都', '知道', '水是', '我们',]
5	124	['投诉', 'A', '市', '盛世', '耀凯', '小区', '物业', '无故', '停水', '。', '我', '在', '2015', '年', '购买', '了', '盛世', '耀凯', '小区', '17', '栋', '3', '楼', '4', '楼', '两层', '共计', '2', '千', '平方', '一直', '以来', '我们', '按时', '足额', '缴纳', '物业费', '及其',]
6	50	['咨询', 'A', '市', '楼盘', '集中', '供暖', '一事', '。', '由于', '西地省', '地区', '常年', '阴冷', '潮湿', '的', '气候', '加之', '近年', '气候', '逐渐', '更加', '恶劣', '地处', '月亮', '岛', '片区', '近年', '规划', '有', '楚江', '集中', '供暖', '不知', '规划',]
7	88	['A3', '区', '桐梓', '坡', '西路', '可可', '小城', '长期', '停水', '得不到', '解决', '。', '尊敬', '的', '胡书记', ':', '您好', '!', '家住', 'A', '市', 'A3', '区', '桐梓', '坡', '西路', '可可', '小城', '的', '居民', '长期以来', '经常', '停水', '。', '小区业主', '业委会',]
8	278	['反映', 'C4', '市', '收取', '城市', '垃圾处理', '费', '不', '平等', '的', '问题', '。', '我们', '是', '梅家田', '社区', '辖区', '内', '的', '小区', '居民', '我们', '每年', '都', '依法', '依规', '向', '小区', '物业公司', '交纳', '了', '城市', '垃圾处理', '费', '。',]
9	314	['A3', '区', '魏家坡', '小区', '脏乱差', '。', '尊敬', '的', 'A', '市政府', '领导', ':', '您们好', '!', '我', '是', 'A', '市', 'A3', '区', '魏家坡巷', '的', '业主', '多年', '以来', '我们', '小区', '的', '脏', '乱', '差', '多次', '向', '社区', '反映',]

表 1 可以看出有些词切分的不恰当，如“A”，“市”等。这些词本应该被分在一起，在 jieba 分词器中并不含有这些政府业务上特殊的词汇，因此针对政府领域中大量的政务文本的特征，指定自定义的词典，加入到 jieba 词库里。结合实际应用需要，对政府业务中的一些术语建立专业词库。虽然 jieba 分词器有新词识别能力，但是自行添加新词可以保证更高的正确率。例如附件 2 数据，如表 2 前十项所示。表 2 中可以发现有些词汇已经被正确的切分出来了，说明添加自定义的词典对文本数据的分词效果较好。

表 2 自定义词典分词表

序号	词量	留言主题+留言详情
0	69	['A 市', '西湖', '建筑', '集团', '占道', '施工', '有', '安全隐患', '。', 'A3 区', '大道', '西行', '便', '道', '未管', '所', '路口', '至', '加油站

1	60	<p>，'路段'，'，'，'人行道'，'包括'，'路灯'，'杆'，'，'，'被'，'圈'，'西湖'，'建筑'，'集团'，'燕子山'，'安置房'，.....]</p> <p>['A市'，'在水一方'，'大厦'，'人为'，'烂尾'，'多年'，'，'，'安全隐患'，'严重'，'。'，'位于'，'书院'，'路'，'主干道'，'的'，'在水一方'，'大厦'，'一楼'，'至'，'四楼'，'人为'，'拆除'，'水'，'、'，'电等'，'设施'，'后'，'，'，'烂尾'，'多年'，'，'，'用'，'护栏'，'.....]</p>
2	611	<p>['投诉'，'A市'，'A1区'，'苑'，'物业'，'违规'，'收'，'停车费'，'。'，'尊敬'，'，'的'，'领导'，'：'，'A1区'，'苑'，'小区'，'位于'，'A1区'，'火炬'，'路'，'，'，'，'小区'，'物业'，'A市'，'程明'，'物业管理'，'有限公司'，'，'，'未经'，'小区业主'，'同意'，'，'，'利用'.....]</p>
3	85	<p>['A1区'，'蔡锷'，'南路'，'A2区'，'华庭'，'楼顶'，'水箱'，'长年'，'不洗'，'。'，'A1区'，'A2区'，'华庭'，'小区'，'高层'，'为'，'二次'，'供水'，'，'，'楼顶'，'水箱'，'长年'，'不洗'，'，'，'现在'，'自来水'，'龙头'，'的'，'水'，'严重'，'霉味'，'，'，'大家'，.....]</p>
4	83	<p>['A1区'，'A2区'，'华庭'，'自来水'，'好大'，'一股'，'霉味'，'。'，'A1区'，'A2区'，'华庭'，'小区'，'高层'，'为'，'二次'，'供水'，'，'，'楼顶'，'水箱'，'长年'，'不洗'，'，'，'现在'，'自来水'，'龙头'，'的'，'水'，'严重'，'霉味'，'，'，'，'大家'，'都'，'知道'，'，'，'.....]</p>
5	121	<p>['投诉'，'A市'，'盛世耀凯'，'小区'，'物业'，'无故'，'停水'，'。'，'我'，'在'，'2015'，'年'，'购买'，'了'，'盛世耀凯'，'小区'，'17'，'栋'，'3'，'楼'，'，'，'，'4'，'楼'，'两层'，'共计'，'2'，'千'，'平方'，'，'，'一直'，'以来'，'我们'，'按时'，'足额'，'缴纳'，'物业费'，.....]</p>
6	48	<p>['咨询'，'A市'，'楼盘'，'集中'，'供暖'，'一事'，'。'，'由于'，'西地省'，'地区'，'常年'，'阴冷'，'潮湿'，'的'，'气候'，'，'，'加之'，'近年'，'气候'，'逐渐'，'更加'，'恶劣'，'，'，'地处'，'月亮岛'，'片区'，'近年'，'规划'，'有'，'楚江'，'集中'，'供暖'，'，'，'.....]</p>
7	83	<p>['A3区'，'桐梓坡'，'西路'，'可可'，'小城'，'长期'，'停水'，'得不到'，'解决'，'。'，'尊敬'，'的'，'胡书记'，'：'，'您好'，'！'，'家住'，'A市'，'A3区'，'桐梓坡'，'西路'，'可可'，'小城'，'的'，'居民'，'长期以来'，'经常'，'停水'，'。'，'小区业主'，'、'.....]</p>
8	276	<p>['反映'，'C4市'，'收取'，'城市'，'垃圾处理'，'费'，'不'，'平等'，'的'，'问题'，'。'，'我们'，'是'，'梅家田'，'社区'，'辖区'，'内'，'的'，'小区'，'居民'，'，'，'我们'，'每年'，'都'，'依法'，'依规'，'向'，'小区'，'物业公司'，'交纳'，'了'，'城市'，'垃圾处理'.....]</p>
9	310	<p>['A3区'，'魏家坡'，'小区'，'脏乱差'，'。'，'尊敬'，'的'，'A市'，'政府'，'领导'，'：'，'您们好'，'！'，'我'，'是'，'A市'，'A3区'，'魏家坡巷'，'的'，'，'业主'，'，'，'多年'，'以来'，'，'，'我们'，'小区'，'的'，'脏'，'，'，'乱'，'，'，'差'，'多次'，'向'，'社区'.....]</p>

3.2 去停用词

文本数据进行分词处理后,会存在如“这”、“的”、“唉”、“吧”等的干扰词汇,这些词汇被称为停用词,这些词汇在中文文本数据中大量出现,会对后

续的文本特征的构建造成极大的干扰,进而影响到后续文本分类过程。因此使用停用词表(根据群众留言信息进行停用词表制作)对分词后的数据进行词汇的停用处理。例如附件 2 中数据,如表 3 前十项所示。从表 3 部分展示可以看出,文本处理的还是比较好的。

表 3 停用词文本处理表

序号	词量	留言主题+留言详情
0	30	['西湖', '建筑', '占道', '施工', '安全隐患', '大道', '西行', '未管', '路口', '加油站', '路段', '人行道', '路灯', '杆', '圈', '西湖', '建筑', '燕子山', '安置房', '项目',]
1	29	['在水一方', '大厦', '人为', '烂尾', '安全隐患', '书院', '主干道', '在水一方', '大厦', '一楼', '四楼', '人为', '拆除', '水', '电等', '设施', '烂尾', '护栏', '围着', '占用',]
2	238	['投诉', '物业', '违规', '停车费', '火炬', '物业', '程明', '物业管理', '未经', '小区业主', '利用', '业主', '公摊', '公共', '面积', '业主', '滥收', '停车费', '收费', '车辆',]
3	35	['蔡锷', '南路', '华庭', '楼顶', '水箱', '长年', '不洗', '华庭', '高层', '二次', '供水', '楼顶', '水箱', '长年', '不洗', '自来水', '龙头', '水', '霉味', '水是', '日常生活',]
4	33	['华庭', '自来水', '好大', '一股', '霉味', '华庭', '高层', '二次', '供水', '楼顶', '水箱', '长年', '不洗', '自来水', '龙头', '水', '霉味', '水是', '日常生活', '必不可少',]
5	39	['投诉', '盛世耀凯', '物业', '无故', '停水', '购买', '盛世耀凯', '两层', '共计', '平方', '足额', '缴纳', '物业费', '费用', '入住', '物业公司', '为所欲为', '物业', '业主',]
6	20	['咨询', '楼盘', '供暖', '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', '供暖', '规划', '供暖', '具体位置', '楼盘']
7	25	['桐梓坡', '西路', '可可', '小城', '停水', '得不到', '家住', '桐梓坡', '西路', '可可', '小城', '停水', '小区业主', '物业', '开发商', '居委会', '寻求', '停水', '承诺', '给与',]
8	87	['垃圾处理', '费', '平等', '梅家田', '社区', '辖区', '依法', '依规', '物业公司', '交纳', '垃圾处理', '费', '环卫局', '垃圾处理', '费是', '合法', '合规', '有利于', '环境',]
9	91	['魏家坡', '脏乱差', '政府', '魏家坡巷', '业主', '脏', '乱', '差', '社区', '得不到', '停车', '不容乐观', '停车', '车辆', '乱', '停', '乱放', '消防通道', '火灾', '小区业主',]

4. 群众留言分类模型

4.1 文本数据分析

针对附件 2 中“留言主题”和“留言详情”属性数据特点构建了自定义词典和停用词表，其中停用词表中包含了“A 市”、“A3 区”、“您们好”等大量不需要的名词、动词、形容词词汇和数值。运用词典和停用词表进行上述预处理后，进一步进行词频统计，查看高频次词语。如图 2 词频词云展示。



图 2 词频词云展示图

从词云中可以看出预处理效果还是比较好的,把关键名词提炼出来了。为了更好地对群众留言分类,对附件 1 中二级分类和三级分类属性进行词语切分,加入附件 2 的分词数据中,利用向量空间模型(VSM)进行文本特征的构建,即将非结构化的文本内容表达为向量式的、可用于机器学习与理解的表达形式的过程。VSM 模型中,文档与特征词汇均被假设为一个 t 维空间向量的一部分,文本内容中共有词汇个数 t 个,在 t 维空间下,文档 D_i 被表征为该空间下的一个向量,如式(1)所示:

$$D_i=(d_{i1},d_{i2},d_{i3},\cdots,d_{ij}) \quad (1)$$

d_{ij} 表示为第 j 个词项的权值, 则如表 4 所示的 $n \times t$ 阶的矩阵可用于表示文本集和, 其中 n 为文本数据集中文本的数量。

表 4 VSM 模型矩阵示例表

	词项 1	词项 2	词项 3	词项 t
文档 1	d_{11}	d_{12}	d_{13}	d_{1t}
文档 2	d_{21}	d_{22}	d_{23}		d_{2t}
...	...				
文档 n	d_{n1}	d_{n2}	d_{n3}	d_{nt}

4.2 群众留言分类

在群众留言的一级分类模型设计中，选用了三种方法，分别是贝叶斯算法、基于 $TF-IDF$ 的贝叶斯分类模型和基于 $TF-IDF$ 的 SVM 分类模型。

4.2.1 贝叶斯算法

在文本数据进行向量式后对文档中各个单词的词频 TF 进行统计（即每个单词在文档中出现的次数），为每个词项赋予权值。然后含有权重的数值矩阵经过贝叶斯算法进行分类。

贝叶斯算法（Naive Bayes, NB）的主要思想是在给定待分类文本的条件下，计算其属于各个类别的条件概率，然后选择其中条件概率最高的类别作为该文本所属的类别，算法过程如下：

- 1) 对于某个测试文本 d ，计算该文本属于类别 C_j 的概率：

$$p(C_j|d) = \frac{p(C_j)p(d|C_j)}{p(d)} \quad (2)$$

$$p(d) = \sum_{j=1}^{|C|} p(C_j)p(d|C_j) \quad (3)$$

其中 $p(C_j)$ 表示类别 C_j 的概率。 $p(d|C_j)$ 表示文档 d 属于类别 C_j 的概率。

- 2) 最后，将文本分到概率 $p(C_j|d)$ 最大的那个类别中。

$$V_{max} = \arg \max p(C_j|d) \quad (4)$$

4.2.2 基于 $TF-IDF$ 的贝叶斯分类模型

本文使用的文本分类算法是在 $TF-IDF$ 基础上进行贝叶斯方法分类的。

$TF-IDF$ 是特征词项权重的计算中的最具代表性的模型，该模型将词汇的表征性与词汇在本条数据与其它数据中的出现频率联系起来，主要由两部分组成，

计算如式 (5) 所示:

$$TF-IDF(w_{MN}) = TF(w_{MN}) \cdot IDF(w_{MN}) \quad (5)$$

其中 $TF(w_{MN})$ 表示词汇 w_n 在文档 Doc_m 中的出现频率, $IDF(w_{MN})$ 可表示为式 (6) 所示:

$$IDF(w_{MN}) = \log(N/n_k) \quad (6)$$

N 为文本集合中的文本数量, n_k 是指包含词汇 w_{MN} 的文本频数。

4.2.3 基于 $TF-IDF$ 的 SVM 分类模型

在 $TF-IDF$ 基础上利用支持向量机对政务留言文本进行分类。支持向量机 (SVM) 是由 Vapnik 所提出经典机器学习模型。在文本分类过程中, SVM 模型将输入的训练文本作为几何空间中的一个点, 通过对训练样本的学习构建一个可分离不同样本的超平面, 并根据测试样本在落在超平面两侧的位置确定其类别。SVM 模型在规模较小的数据上的分类效果优良, 由于其具有的核函数特性, 在某些高维情景下同样适用。

4.3 分类模型评估

在使用文本分类器对未标记的政务留言文本数据进行分类后, 需要对不同分类器的分类效果进行评估, 主要的评估指标为准确率 ($Precision, P$)、召回率 ($Recall, R$) 与 F_1 值, 其中准确率与召回率可表示为:

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$F_1 = \frac{1}{n} \sum_{i=1}^n \frac{2P_i R_i}{P_i + R_i} \quad (9)$$

P 是指被分类器分类为某一类的样本集中, 正确标记的样本数据所占的比例; R 是指分类器所标记正确的样本数据占有所有该类别下真实样本数据集的比例, 两者从不同的角度衡量了分类任务的分类效果; F_1 值则是对两者的加权平均, 综合两者对分类效果进行评估。其中 P_i 为第 i 类的查准率, R_i 为第 i 类的查全率。

分别利用 7 种一级分类类别的准确率、召回率和 F_1 值对 3 种分类方法进行评估。如表 5 所示。

表 5 一级分类类别运用 3 种分类方法的准确率、召回率和 F_1 值

类别	贝叶斯算法			基于 TF-IDF 的贝叶斯模型			基于 TF-IDF 的 SVM 模型		
	准确率	召回率	F_1	准确率	召回率	F_1	准确率	召回率	F_1
	/%	/%	/%	/%	/%	/%	/%	/%	/%
城乡建设	0.89	0.81	0.85	0.80	0.94	0.87	0.84	0.92	0.87
环境保护	0.85	0.96	0.90	0.91	0.92	0.92	0.93	0.95	0.94
交通运输	0.67	0.91	0.77	0.91	0.66	0.77	0.89	0.83	0.86
教育文体	0.92	0.88	0.90	0.92	0.87	0.89	0.95	0.91	0.93
劳动和社会保障	0.92	0.92	0.92	0.87	0.95	0.91	0.93	0.94	0.93
商贸旅游	0.88	0.79	0.83	0.92	0.79	0.85	0.90	0.85	0.87
卫生计生	0.88	0.91	0.89	0.98	0.83	0.90	0.95	0.87	0.91

根据表 5 中 3 种方法的准确率、召回率和 F_1 值来看每一个类别分类效果，可以大致了解到基于 $TF-IDF$ 的 SVM 模型效果最好。为了整体了解 3 种分类算法的好坏，绘制了贝叶斯分类方法、基于 $TF-IDF$ 的贝叶斯分类方法和支持向量机（SVM）分类方法的平均准确率、平均召回率和平均 F_1 值的柱状图。如图 3 所示。从平均准确率、平均召回率和平均 F_1 值可以明显的看出基于 $TF-IDF$ 的 SVM 模型分类精度是最高的。



图 3 3 种方法的平均准确率、平均召回率、平均 F_1 值比较

5. 热点问题挖掘模型

某一时段内群众集中反映的某一问题可称为热点问题，如“XXX 小区多位业主多次反映入夏以来小区楼下烧烤店深夜经营导致噪音和油烟扰民”。及时发现热点问题，有助于相关部门进行有针对性地处理，提升服务效率。

由于附件 3 文本数据量较大，且未给出分类类别，整体进行热点分析，识别效果较差。为了减少聚类数量和提高聚类效果，针对热点问题分析的思路分两个步骤。第一步是在上一问的群众留言分类的基础上对附件 3 中“留言主题”和“留言详情”合并文本数据进行类别分类。第二步在有类别的基础上，对每一个类别进行聚类分析，挖掘热点问题。由于“留言详情”属性数据量较多，存在干扰词汇会影响后续模型辨别精度，因此主要采用“留言主题”属性数据进行热点问题挖掘。

5.1 文本数据分析

针对第一步，在文本分类前，由于附件 3 文本与附件 2 中文本数据有差异，因此针对附件 3 文本中的特殊词汇制作出新的词典和停用词分别导入 jieba 分词器和停用词表中，进行文本的预处理。由于基于 TF-IDF 的 SVM 分类模型效果较好，利用该模型对附件 3 中“留言主题”和“留言详情”合并文本数据进行分类，共分 11 类（标签分别为 1, 2, 3, 4, 5, 6, 7, 9, 11, 14, 15），其分类部分结果如图 4 所示。其中一级级别类别标注：{“城乡建设”: 1, “环境保护”: 2, “交通运输”: 3, “教育文体”: 4, “劳动和社会保障”: 5, “商贸旅游”: 6, “卫生计生”: 7, “党务政务”: 8, “国土资源”: 9, “纪检监察”: 10, “经济管理”: 11, “科技与信息产业”: 12, “民政”: 13, “农村农业”: 14, “政法”: 15}。全部分类结果（已按照类别标注顺序排列好）保存在 2_classification_result.csv。

Index	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数	分词	类别
0	188006	A000102948	A3区一米阳光婚纱摄影是否合法的税了？	2019/2/28 11:25:05	--	0	0	A3区 一米阳光 婚纱摄影 合法 纳税 座落在 A市 A3区 联丰路 米兰 春天 一家 名叫 一米阳光 婚纱摄影 影像 年里 工作室 营业额上百万	6
1	188007	A00074795	咨询A6区道路命名规划初步成果公示和城乡门牌问题	2019/2/14 20:00:00	--	0	1	咨询 A6区 道路 命名 规划 成果 公示 城乡 门牌 A市 A6区 道路 命名 规划 成果 公示 文件 转化 正式 成果 希望 加快 命名 规划 道路 安装 门牌 名	1
2	188031	A00040066	反映A7县香华镇金鼎村水泥路、自来水到户的问题	2019/7/19 18:19:54	--	0	1	A7县 香华镇 金鼎村 水泥路 自来水 到户 系 香华镇 金鼎村 七里 组 村民 不知 相关 水泥路 到户 政策 自来水 到户 政策 政府 主导 投资 村民 集资	1
3	188039	A00081379	A2区黄兴路步行街大古道巷住户卫生间装便外排	2019/8/19 11:48:23	--	0	1	A2区 黄兴路 步行街 大古道巷 住户 卫生间 装便 外排 靠近 黄兴路 步行街 城 南路 街道 大古道巷 一步 搭桥 小区 停车场 东面 围墙 第一 单元 住户	1
4	188059	A00028571	A市A3区中海国际社区三期与四期中间空地夜间施工噪音扰民	2019/11/22 16:54:42	--	0	0	A市 A3区 中海国际社区 三期 四期 空地 夜间 施工 噪音 扰民 A市 A3区 中海国际社区 三期 四期 蓝天 瑞 洲 幼儿园 旁边 空地 三不管 状态 物业	1
5	188073	A909164	A3区麓泉社区单方面改变麓谷明珠小区6栋架空层使用性质	2019/3/11 11:40:42	--	0	0	A3区 麓泉社区 单方面 改变 麓谷明珠小区 架空层 使用性质 麓泉社区 麓谷明珠小区 居民 感觉 震惊 伤心 购房 签合同 产权 提及 架空层 改建 老年	1
6	188074	A909092	A2区富绿新村房产的性质是什么？	2019/1/31 20:17:32	--	0	0	A2区 富绿 新村 房产 性质 二高 一部 发出 非法 集资 打击 通知 中是 金融 犯罪 通知 特征 欺骗性 强 涉及 群众 损失 富绿 新村 欺骗 业主 西地省	1
7	188119	A00035029	对A市地铁连横用工问题的质疑	2019/5/27 16:04:44	--	0	0	A市 地铁 连横 用工 质疑 一名 A市 地铁站 上班 安检员 中介公司 介绍 上班 安检员 岗位 两个 班次 一个班 早上 六点半 一个班 二点 晚上 一个	5
8	188170	A88011323	A市6路公交车随意变道通行	2019/12/23 8:50:24	--	0	0	A市 公交车 随意 变道 通行 分两 公交车 司机 座位 旁边 汽车 编号 A3区 大道 金星 路口 西往东 辅道 左 拐 通行 司机 并未 地面 车辆 指示 导向	3
9	188249	A00084085	A3区保利麓谷林语桐梓坡路与麓松路交汇处地铁凌晨2点施工扰民	2019/9/17 4:25:00	--	0	0	A3区 保利麓谷林语桐梓坡路 与麓 松路 交汇处 地铁 凌晨 施工 扰民 保利麓谷林语桐梓坡路 与麓 松路 交汇处 地铁 凌晨 施工 扰民 家里 老人	1
10	188251	A00013092	A7县特立路东四路口晚高峰太堵，建议调整信号灯配时	2019/10/19 11:02:40	--	0	0	A7县 特立路 东四 路口 晚高峰 太堵 建议 调整 信号灯 配时 晚高峰 五点 特立路 东四 路口 东 往西方 向车 越来越 实在 堵 情况 至少 两到 三个	1
11	188260	A00053484	A3区青青家园小区乐果果零食炒货公共通道摆放空调扰民	2019/5/31 17:06:13	--	0	0	A3区 青青家园 小区 乐果果 零食 炒货 公共通道 摆放 空调 扰民 宁静 复习 迎考 大半 年底 商 空调 冰箱 外机 轰鸣 扰民 A3区 天顶 街道 青山 新村	1
12	188396	A00047580	关于拆除聚美龙楚在西地省商学院宿舍旁安装变压器的请求	2019/4/15 16:23:09	--	2	1	拆除 聚美龙楚 西地省 商学院 宿舍 安装 变压器 桐梓坡 589 白鹤 咀 停车场 由聚 美龙楚 新能源 公司 建 商学院 新能源 汽车 充电桩 项目 告知 情	1
13	188399	A00097934	A市利保壹号公馆项目夜间噪声扰民	2019/7/3 6:23:25	--	0	0	A市 利保壹号公馆项目 夜间 噪声 扰民 举报 A市 利保壹号公馆项目 夜间 噪声 扰民 A市 利保壹号公馆项目 位于 A市 A3区 威嘉苑 麓谷桐梓坡 开发商 西	1
14	188409	A00032274	A市地铁3号线星沙大道站地铁出入口设置极不合理！	2019/6/19 10:14:39	--	0	4	A市 地铁 星沙大道 地铁 出入口 设置 不合理 领导 不合理 领导 A7县 星沙 街道 两个 孩子 上有老下有 小 居民 居民 直斥 A市 地铁 言论 过激 强烈 不再 关注 关	1
15	188414	A00096844	A4区北辰小区非法住改商问题何时能解决？	2019/8/1 7:20:31	--	0	0	A4区 北辰小区 非法 改商 何时能 解决 北辰 一名 业主 小区 改商 数量 达 上 百家 安全隐患 发生 火情 幸 扑灭 A市 业委会 物业 业主 改商 治理	1
16	188416	A00029753	请给K3县乡村医生卫生室执业许可证	2019/06/20 20:38:47	--	0	0	K3县 乡村 医生 卫生室 执业 许可证 K3县 乡村 卫生室 无证 行医 状态 情况 原因 卫生室 医疗机构 许可证 早就 过期 两年 卫生局 迟迟 不发 新执	7
17	188451	A00013004	A7县香华镇石塘铺村有党员家开麻将馆	2019/4/11 17:54:25	--	0	2	A7县 香华镇 石塘铺村 党员 家开 麻将馆 香华镇 一名 村民 接到 政府 通知 要求 开 麻将馆 警察 上门 通知 通知 关门 石塘铺村 胡 海山 党员 营业	1
18	188455	A00035902	咨询异地办理出国签证的问题	2019/5/16 15:20:43	--	0	0	咨询 异地 办理 出国 签证 书记 外地人 毕业 落户 A市 申请 短期 出国 旅游 签证 回到 户籍 所在地 办理 A市 办理 A市 领馆 告知 办理 地址 需 村	5
19	188467	A00050188	投诉A市温斯顿英语培训学校拖延退费！	2019/3/28 19:57:19	--	0	1	投诉 A市 温斯顿 英语 培训学校 拖延 退费 退费 之日起 之内 退还 学习 费用 今已 经过 合同 退费 每次 学校 推脱 退费 每次 退费 退费 退费 退费	4

图 4 基于 TF-IDF 的 SVM 模型分类结果部分展示

从图 4 中可以观察到“留言主题”文本中对应的{“城乡建设”: 1, “环境保护”: 2, “交通运输”: 3, “教育文体”: 4, “劳动和社会保障”: 5, “商贸旅游”: 6,

“卫生计生”：7,}分类效果还是不错的。

为了更好地观察到分类类别数量，根据 11 个分类类别，分别对每一个类别进行数量统计，如表 6 所示。

表 6 分类类别数量统计表

类别	1	2	3	4	5	6	7	9	11	14	15
数量	2545	475	378	330	329	189	74	3	1	1	1

由表可以看出 7, 9, 11, 14, 15 这 5 个类别的数量是非常少的, 则分类个数较少。为了进一步直观的观察到每一个类别的关键词, 进行了词云的绘制。如图 5 中的 1 类, 2 类, 3 类, 4 类, 5 类, 6 类, 7 类, 9 类, 11 类, 14 类, 15 类所示。展示可以看出: 1 类中主要的关键词是“A 市, 小区、业主、开发商、建设”, 2 类中主要关键词是“小区、居民、噪音、扰民、油烟”, 3 类中主要关键词是“A 市、道路、车辆、通行、交通”, 4 类中主要关键词是“幼儿园、学校、教育局、老师、学生”, 5 类中主要关键词是“西地省、工资、公司、员工、社保”, 6 类中主要关键词是“A 市、公司、业主装修、政府、电梯”, 7 类中主要关键词是“A 市、西地省、医院、患者、医疗”, 从这较多词量的前 8 类中可以发现其分类的类别还是比较准确的, 有较少识别错误的。再看从 9, 11, 14, 15 这后 4 类的关键词中可以更加直观地发现其分类结果较好。



1 类



2 类



3 类



4 类



5 类



6 类



7 类



9 类



11 类

14 类



15 类

图 5 11 种分类类别词云展示

针对第二步,提取分词后的“留言主题”属性文本,发现有两行数据丢失,与分词前文本的“留言主题”、“留言详情”属性对照,“留言详情”属性数据正常,则“留言主题”属性不具有代表性。如表 7 所示。采取了把相对应序号的两行中的“留言详情”属性数据添加到“留言主题”中进行去停用词处理措施。部分处理结果如表 8 所示。

表 7 “留言主题”文本异常数据

序号	留言编号	留言用户	留言主题	留言时间	点赞数	反对数	留言详情
65	189856	A00073717	）	2019/7/3 11:53:00	1	0	“A 市保利麓谷林语小区，是 A 市的大型小区，居住人口有一万多人。地下一层墙体严重开裂，并且地下水 and 泥沙大量溢出，特别是逢下雨天 严重，小区还算比较新，才建几年，楼高 33 楼……”
	287458	A909241	环境污染	2019/12/20 14:06:00	0	0	丽发新城小区旁建了一个大型搅拌站，运渣车吵得人精神崩溃，灰尘满天飞，我们该怎么办？

表 8 附件 3 文本数据处理结构

序号	词量	留言主题
0	6	['A3 区', '一米阳光', '婚纱', '艺术摄影', '合法', '纳税']
1	9	['咨询', 'A6 区', '道路', '命名', '规划', '成果', '公示', '城乡', '门牌']
2	6	['A7 县', '春华镇', '金鼎村', '水泥路', '自来水', '到户']
3	8	['A2 区', '黄兴路', '步行街', '大古道巷', '住户', '卫生间', '粪便', '外排']
4	10	['A 市', 'A3 区', '中海国际社区', '三期', '四期', '空地', '夜间', '施工', '噪音', '扰民']
5	7	['A3 区', '麓泉社区', '单方面', '改变', '麓谷明珠小区', '架空层', '使用性质']
6	5	['A2 区', '富绿', '新村', '房产', '性质']
7	5	['A 市', '地铁', '违规', '用工', '质疑']
8	5	['A 市', '公交车', '随意', '变道', '通行']
9	9	['A3 区', '保利麓谷林语桐梓坡路', '与麓', '松路', '交汇处', '地铁', '凌晨', '施工', '扰民']

表 8 中可以很明显的看到词汇别处理的很完整。为了对全部词汇整体的词量了解,对“留言主题”属性文本预处理后,又进行了词频的统计,其统计结果用词云展示,如图 6 所示。



图 6 词频词云图

5.2 聚类热点问题挖掘

基于已有分类类别的附件 3 数据，对预处理好的“留言主题”属性数据进行聚类分析，由于有些分类类别数据量比较大，采用 K-Means 聚类算法进行热点问题分析。K-Means 是常用的聚类算法，与其他聚类算法相比，其时间复杂度低，聚类的效果也不错。具体思想是将数据集划分为 k 个簇的算法，簇的个数 k 是未知的，每个簇通过其质心（簇中所有点的中心）来描述。K-Means 算法的工作流程是：

- (1) 随机确定 k 个初始点作为质心；
- (2) 将数据集中的每个点找到距离最近的质心，并将其分配到该质心对应的簇中；
- (3) 将每个簇的质心更新为该簇中所有点的平均值；
- (4) 重复第 (2)、(3) 步骤，直到簇的分配结果不再变化。

K-Means 算法能够保证收敛，但不能保证收敛于全局最优点，当初始中心点选取不好时，只能达到局部最优点，整个聚类的效果也会比较差，且分类的个数是未知的。因此，选择彼此距离尽可能远的那些点作为中心点；先采用层次聚类算法进行初步聚类输出 k 个簇，以簇的中心点的作为 K-Means 的中心点的输入；多次随机选择中心点训练 K-Means 来确定一个合适 k 值，选择效果最好的聚类结果。

层次聚类（Hierarchical Clustering）是一种聚类算法，通过计算不同类别数据点间的相似度来创建一棵有层次的嵌套聚类树。在聚类树中，不同类别的原始数据点是树的最低层，树的顶层是一个聚类的根节点，即采用“自底向上”的聚类策略。具体思路是先将数据集中的每个样本看作一个初始聚类簇，然后找出两个聚类最近的两个簇进行合并，不断重复该步骤，直到达到预设的聚类个数或某种条件。其中利用平均距离来计算两个簇之间的距离。

由前面表 6 中可知 9、11、14、15 类别的数量较少，进行层次聚类判定聚类个数并不适合，根据文本内容把这 4 类都聚为一类。对前 7 个分类类别每一类进行层次聚类，确定 K-Means 聚类的个数（ k 值）如表 9 所示。如表中可以看出，由于前 6 个分类类别的数量较大，则聚类的个数也相应增加。

表 9 一级分类类别对应的聚类个数

一级分类类别	1	2	3	4	5	6	7	9	11	14	15
聚类个数(k 值)	845	98	55	117	102	145	19	1	1	1	1

按照一级分类类别对聚类后的每一分类类别给出排名前 5 的热点问题，如图

7 所示，并保存为文件“热点问题表.xls”。

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	16	2019/07/07至2019/8/31	A市伊景园滨河苑	伊景园滨河苑项目捆绑销...
2	2	14	2019/01/4至2020/01/06	A市	加快现代服务建设力度
3	3	14	2018/10/27至2020/01/07	A市	地铁施工扰民
4	4	13	2019/1/18至2019/12/30	A市	反映A市公交车问题
5	1	11	2019/02/24至2019/12/23	A市	反映A市多所学校问题

图 7 热点问题

由图 7 中，可以给出相应热点问题对应的留言信息，部分分别如图 8 中的 1 类, 2 类, 3 类, 4 类, 5 类, 6 类, 7 类, 9 类, 11 类, 14 类, 15 类所示，全部内容保存在“热点问题留言明细表.xls”里。

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
1	1	289950	A00044759	投诉A市伊景园滨河苑捆绑销售	2019-07-07 07:28:06	提问A市政府就广州局集团公司与A市政府及A市	0	0
1	1	283879	A00044759	A市伊景园滨河苑项目捆绑销售	2019-07-18 20:27:40	关于铁广集团铁路职工定向商品房伊景园滨河苑	0	0
1	1	223247	A00044759	投诉A市伊景园滨河苑捆绑销售	2019-07-23 17:06:03	关于铁广集团铁路职工定向商品房伊景园滨河苑	0	0
1	1	285897	A909191	武广新城伊景园滨河苑违法捆绑	2019-08-01 20:06:52	我们是广铁集团铁路职工,武广新城片区的伊景园	0	0
1	1	199190	A00095080	关于A市武广新城违法捆绑销售	2019-08-01 22:32:26	武广新城为铁广集团的定向商品房,在未取得预	0	0
1	1	191001	A909171	A市伊景园滨河苑协商要求购房	2019-08-16 09:21:33	商品房伊景园滨河苑项目是由A市政府为牵头为	1	12
1	1	204960	A909192	家里本来就困难,还要捆绑买卖	2019-08-21 18:12:20	我是广铁集团铁路职工,因家人身体欠佳,常年就	0	0
1	1	244528	A909235	伊景园滨河苑开发商强买强卖	2019-08-21 19:05:34	A市广铁集团伊景园滨河苑商品房本来是政府为	0	2
1	1	289473	A000103438	反对滨河苑房子和车位捆绑销售	2019-08-22 00:00:00	现有伊景园滨河苑在对广铁集团员工定向销售的	0	0
1	1	214975	A909182	关于房伊景园滨河苑销售若干	2019-08-22 00:00:00	尊敬的领导,您好!感谢您在百忙之中抽出时间	0	3
1	1	258037	A909190	投诉伊景园滨河苑捆绑销售车	2019-08-23 11:46:03	尊敬的领导:我是广铁集团铁路职工,本来武广新	0	0
1	1	244243	A909198	关于伊景园滨河苑捆绑销售车	2019-08-24 18:23:12	广铁集团铁路职工定向商品房伊景园滨河苑项	0	0
1	1	222209	A00017171	A市伊景园滨河苑定向限价商品	2019-08-28 10:06:03	广铁集团与伊景园滨河苑开发商沆瀣一气,严重	0	0
1	1	236301	A909197	和谐社会背景下的A市伊景园	2019-08-30 16:32:12	广铁集团与A市政府及A市市政工程有限公司要求	0	0
1	1	260254	A909173	投诉A市伊景园滨河苑开发商违	2019-08-30 18:10:23	投诉A市伊景园滨河苑开发商违法捆绑销售无产	0	0
1	1	279070	A00095080	投诉A市伊景园滨河苑开发商违	2019-08-31 06:33:25	投诉A市伊景园滨河苑开发商违法捆绑销售无产	0	0
1	2	196908	A00031618	请加快A市汉江陵考古遗址公园	2019-01-04 15:00:57	地处银杉路旁边的国家级汉江陵考古遗址公园已	0	9
1	2	229359	A00031618	请加快A市现代服务业转型升级	2019-01-04 15:08:27	在国家大力倡导现代服务业转型升级的大背景下	0	0
1	2	240473	A00031618	请A市加快地铁的建设	2019-01-15 13:27:13	地处中部省会A市在地铁建设方面相比隔壁武汉	0	1
1	2	190213	A00031618	请A市加快自来水深度净化改造	2019-01-16 12:53:26	地处时代倾城小区的自来水烧过开水后电水壶就	0	0
1	2	225574	A00031618	请加快A6区桃花井社区配套功	2019-01-25 16:09:34	地处月亮岛街道桃花井社区配套的阅览室,警务	0	2
1	2	200611	A00031618	请A市加快高铁西站的建设力度	2019-02-15 15:47:46	A市在加快建设国家交通物流中心的大背景下何	0	3
1	2	206587	A00031618	请加快省图书馆的建设力度	2019-02-15 15:49:59	在建设文化强省的大背景下加快省图书馆建设刻	0	0
1	2	215805	A00031618	请加快体育强省的建设力度	2019-02-15 15:57:56	规划多年的省体育中心迟迟没有建设,在创建体	0	0

图 8 1 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
2	1	243692	A909201	丽发新城小区附近的搅拌站噪	2019-11-15 11:23:21	领导您好!我是暮云街道丽发新城小区的	0	2
2	1	281348	A909219	希望领导“拯救”丽发新城小区	2019-11-24 00:00:00	尊敬的张市长,您好!感谢您在百忙之中抽	0	0
2	1	283482	A909232	丽发新城小区附近搅拌站的一	2019-12-07 09:21:32	我是A市A2区丽发新城小区的居民,我	0	0
2	1	238212	A909203	丽发新城小区附近建搅拌站合	2019-12-12 10:23:11	请问在居民区附近百米的地方建搅拌站	0	0
2	1	287458	A909241	环境污染	2019-12-20 14:06:19	丽发新城小区旁建了一个大型搅拌站,造	0	0
2	1	217700	A909239	丽发新城小区旁的搅拌站严重	2019-12-21 02:33:21	开发商把特大型搅拌站,水泥厂从绿心	0	1
2	1	216824	A909214	搅拌站大量加工砂石料噪音污	2019-12-25 12:15:57	最近一段时间以来,A2区丽发新城小区	0	0
2	1	190523	A00072847	A市丽发新城违法搅拌站,彻夜	2019-12-26 13:55:15	领导您好!我是A市暮云街道丽发新城	0	0
2	1	235362	A909215	暮云街道丽发新城小区附近水	2020-01-06 20:45:34	暮云街道丽发新城小区附近搅拌站的存	0	0
2	1	214282	A909209	A市丽发新城小区附近搅拌站	2020-01-25 09:07:21	你们管不管A2区丽发新城小区啊!这个	0	0
2	2	232037	A00032452	A6区景城御苑相约本味饭店噪	2019-01-10 08:29:17	尊敬的领导,位于A6区高塘岭街道景城	0	5
2	2	257078	A000109087	A6区通景城空气质量太差	2019-03-12 11:47:17	只要不下雨,空气中一定是弥漫着焦臭	0	0
2	2	270549	A00089107	反映关于A6区景通城有难闻气	2019-04-19 00:50:22	昨天晚上10点半左右景通城小区及学校	0	0
2	2	201252	A00055550	A6区星月绿洲小区门面餐饮商	2019-07-09 15:40:58	我是A6区月亮岛街道杨凤社区星月绿洲	0	0
2	2	282172	A00040139	A6区高塘岭街道望福佳园附近	2019-08-13 11:17:45	位于A市A6区高塘岭街道郭亮路望福佳	5	1
2	2	238230	A00040581	A6区新城国际花都南1门至南2	2019-08-13 19:44:20	新城国际花都南1门至南2门马路对面(0	0
2	2	279313	A00096209	A6区莲湖路18号居民房屋被污	2019-08-26 13:52:43	本人是A市A6区莲湖路18号居民,前期	0	0
2	2	286565	A00063993	A6区乌山徐家桥碎石场噪音震	2019-11-22 12:04:38	乌山徐家桥碎石场噪音震天,灰尘满天	0	0
2	3	207467	A00014150	A4区金鹰小学附近一工地夜间	2019-05-06 09:26:39	位于金鹰小学、恒大雅苑居民楼、圣爵	0	0
2	3	228901	A0007228	A4区周楠中学旁铁路火车刹车	2019-05-08 06:27:03	A市A4区周楠中学(高中)旁铁路火车	0	1
2	3	240791	A00082221	A4区天健城夜间施工现场没有	2019-05-10 00:23:49	施工现场没有任何噪音降噪措施,其排	0	0
2	3	219748	A00053364	A市医学院后三街清吧夜间音	2019-06-06 00:26:21	A市医学院后三街清吧夜间音响放很大	0	0
2	3	226339	A00029956	A7县江背镇朱家桥社区工厂夜	2019-08-30 17:14:03	A7县江背镇朱家桥社区工厂夜间焚烧工	0	0
2	3	268755	A00082817	再次反映A4区珠江花园酒店夜	2019-10-07 15:09:54	本人在市长信箱上多次投诉珠江花园酒	0	0

图 8 2 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
3	1	263761	A00081409	A市雨敞坪镇雨敞坪社区妈妈驿站	2019-02-10 10:57:35	西地省A市雨敞坪镇雨敞坪社区类	0	0
3	1	207161	A00026895	我觉得A市橘子洲的烟火蛮好的	2019-02-19 14:36:03	又看到有人再吐槽橘子洲的烟火	4	0
3	1	202604	A00051608	A市白云路成断头路两年了	2019-02-21 11:59:14	前几年搞西二环拓改修的白云路	0	0
3	1	240074	A00032182	新买的纯电动封闭式面包车不到	2019-03-03 22:25:51	为响应国家节能减排及绿色出行	0	1
3	1	278070	A00060228	A市楚行一卡通APP为什么不能	2019-05-16 11:02:49	楚行一卡通APP怎么开发票,打	0	3
3	1	241859	A000106412	A市中海国际社区能考虑建地铁	2019-05-17 14:54:48	中海国际社区靠近高速口、环线	0	8
3	1	214348	A00050627	A市金桥互通的ETC通道为何经	2019-06-13 10:55:21	因为经常通过绕城高速在金桥收	0	1
3	1	188895	A00063289	票牛A市分公司不肯退我草莓盲	2019-06-28 09:31:25	22日A市草莓音乐节的票因为主	0	0
3	1	229036	A000107184	A市望月湖小区麓山雅苑消防通	2019-09-16 11:27:41	麓山雅苑有几百户住户,小区内	0	0
3	1	284145	A000101938	A市地铁三号线为什么还不通	2019-09-25 12:07:40	你好,A市地铁三号线真是几波几	1	0
3	1	275911	A0007322	A市能否恢复八一路等市区主道	2019-09-27 14:12:08	近日,A市对八一路等主要干线路	0	0
3	1	281367	A0007322	A市能否恢复八一路等市区主道	2019-09-27 14:20:04	近日,A市对八一路等主要干线路	0	0
3	1	271956	A00013072	A市全心全速搜投部邮政员涉嫌	2019-12-24 19:42:44	投诉A市火车站车站中路488号全	0	0
3	2	219675	A00088746	A市梅溪湖长郡中学附近交通乱	2019-04-18 10:03:52	我是居住在梅溪湖长郡中学附近	0	0
3	2	257881	A00013092	建议合理配置A市望仙路交通信	2019-04-18 17:22:22	近期经常上下班开车从望仙路、	0	0
3	2	228436	A00082452	A市人民中路主干道旁的废品回	2019-05-14 12:31:27	人民中路人民路立交桥西南角长	0	0
3	2	223915	A00028519	对A3区银双路、楚岳路交汇路口	2019-05-17 11:13:47	近期因为四号线建设结束,银盆	0	0
3	2	270717	A00017271	希望加强对A市楚府大桥匝道道	2019-05-22 17:26:09	最近一段时间,交警每到周一早	0	0
3	2	271407	A00099522	A市楚江路辅道交通信号灯设置	2019-05-24 10:03:41	A2区楚江路一桥下由北往南字	0	0
3	2	225850	A000106641	反映A1区星城世家全体业主急盼	2019-07-24 11:01:25	我们是A市A1区人民东路398号星	0	3
3	2	280150	A00031618	西地省师大附中星城实验学校第	2019-09-03 08:14:12	地处西地省师大附中星城实验学	0	0
3	2	250908	A00016827	A2区新韶西路与新联路交叉路口	2019-09-17 18:55:36	A2区新韶西路与新联路交叉的十	0	0
3	2	218882	A0007819	建议优化A市万丽家高架早晚高	2019-10-14 09:54:01	10月国庆黄金周过后,万丽家高	0	0
3	3	242501	A00018651	A3区雨敞坪镇潘家山村蓝塘红	2019-02-12 19:01:05	您好,希在距离A市二十公里不	0	0

图 8 3 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
4	1	224239	A00029915	A市外国语学校周日收费补课	2019-02-24 19:35:26	A市外国语学校初二近日打着家委	0	0
4	1	270174	A00080995	请解决A4区青竹湖片区的学校	2019-03-03 19:36:39	1青竹湖楚一外国语学校是否会公	0	0
4	1	244099	A00073143	A6区金悦花园后面建学校堵塞	2019-04-14 10:20:03	我们是原60号令安置在金悦花园	0	0
4	1	247999	A00055417	A市中雅培粹学校五一只放2天	2019-04-26 18:53:57	2019年,国务院通过了五一假调至	0	0
4	1	233281	A000107878	A市科技工程学校涉嫌欺骗学生	2019-06-03 11:48:10	尊敬的领导:我是一名9年级学生	0	2
4	1	230531	A00015289	A市外国语学校不分科的吗?	2019-06-17 21:04:03	我是A市外国语学校一个现高一年	0	0
4	1	245119	A00031618	请A市加快严查学校违规上课	2019-06-26 16:15:13	地处西地省师大附中星城实验学校	0	0
4	1	246662	A00093104	反映北京师范大学A市附属学校	2019-08-19 10:30:48	尊敬的县领导:您好!我是北京师	0	0
4	1	202989	A00061635	A3区中加学校的大喇叭太可怕	2019-11-05 10:17:54	1.A市A3区中加学校,自今年秋季	0	0
4	1	191968	A00032980	A3区实验小学东西校区明明属	2019-11-22 10:38:01	您好!我们看好A3区的发展,12年	0	1
4	1	257447	A00047653	A市中小学校机械执行制度,过	2019-12-23 08:57:32	现在已经进入冬天有一段时间了,	0	0
4	2	228393	A00026518	咨询A7县山水湾小区幼儿园性	2019-01-22 15:20:20	尊敬的各位领导:您好!我此次	0	4
4	2	266830	A000108342	咨询A市小学教师薪资问题	2019-04-09 19:39:00	我是博才梅溪湖小学的临聘老师,	1	2
4	2	238072	A00035965	咨询A7县县改区进展情况	2019-05-29 15:39:32	尊敬的领导:打扰了!关于A7县	2	18
4	2	217938	A000106629	关于A市初中生寒暑假时间缩短	2019-07-18 17:40:52	您好!请问为何今年初中生寒假不	1	0
4	2	261181	A000106632	咨询A市易地转业相关政策等问	2019-07-20 10:43:34	领导好:我想咨询下关于A市易	0	0
4	2	245162	A00054908	咨询A市小孩转学的问题	2019-09-03 00:20:44	配偶在市区有房的话,孩子可以从	0	0
4	2	206733	A00080495	咨询A市小学生转学问题	2019-11-18 22:49:10	尊敬的卢局长,您好!本人是19年	0	1
4	2	276992	A00080495	咨询小孩转入A市就读小学等问	2019-11-27 11:48:40	尊敬的市委领导,您好!本人是19	0	0
4	3	248501	A000110358	西地省专升本学位英语身份现	2019-03-16 15:15:17	我是一名参加成人高考专升本的英	0	0
4	3	229014	A00057791	反映西地省农业大学的孩子入	2019-07-05 19:40:00	敬爱的领导,您好!感谢您百忙之	0	0
4	3	241624	A00061091	西地省电子科技职业技术学院	2019-08-23 09:46:03	学校的学费竟高达17800是高出其	0	0
4	3	276513	A000106778	西地省一幼给孩子家长提出无	2019-08-29 00:14:49	您好!今天给您留言主要是反映新	0	0
4	3	218299	A00034254	西地省农大教授存在看学术不	2019-09-23 17:22:30	您好!在2019年西地省现代农业	0	2

图 8 4 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
5	1	228650	A88011323	对A8县市直田眼科医院非法套取医保	2019-01-05 12:07:37	A8县市直田眼科医院2018年套	0	0
5	1	233897	A00072330	A市医保局对肾衰竭患者的医保政策课	2019-01-07 11:50:28	众所周知,国家对大病医疗保险	0	0
5	1	255326	A000106082	A市住院无法使用医保,请问是否可以	2019-01-26 05:03:51	我父母均为残疾人,父亲左眼失	0	0
5	1	278560	A00011652	请问强直性脊柱炎的生物制剂进入A市	2019-02-28 03:09:52	我是一名强直性脊柱炎患者,生	0	0
5	1	238342	A00098426	咨询关于A市居民交职工医保的相关问	2019-03-18 15:48:38	本人在现在在A7县有职工医保,	0	0
5	1	247550	A00050022	投诉A7县医保局推卸责任	2019-04-01 21:12:01	我是一名孕妇,在2018年11月	0	4
5	1	248275	A000106377	外地父母在A市给新生儿买医保难	2019-05-09 11:10:49	衡花书记,您好.我与爱人都是	0	3
5	1	214658	A00051640	A市尿毒症病人反映医保问题	2019-06-02 00:40:12	我是一个尿毒症病人,从2012	0	0
5	1	209959	A00047610	先天性房间隔缺损能不能报医保?	2019-08-14 10:04:15	我儿肖*,19年3月4日至3月13	0	0
5	1	275314	A00092745	在A市补缴医保为何必须本地户口?	2019-11-12 10:36:15	女儿从2018年10月在A市禾丽医	0	0
5	2	275071	A00090192	请给A4区捞刀河镇高岭村村民补偿安置	2019-01-06 11:09:20	张坤树一家的情况说明一	0	6
5	2	220711	A00031682	请书记关注A市A4区58车货案	2019-02-21 18:45:14	尊敬的胡书记,您好!A4区p2p	0	821
5	2	244872	A00039516	A4区邮政公司劳务工特困补助还没到	2019-07-18 13:01:56	本人是A市A4区邮政公司的一名	0	0
5	2	278670	A00050197	A4区捞刀河街道高源村农民的土地被	2019-07-26 06:53:27	农民土地被非法侵占——请问	0	0
5	2	241551	A00095753	请A4区监察委员会重视解决拆迁户认	2019-08-07 16:15:26	本人于今年 2月12号在A市监察	0	5
5	2	236881	A00095158	在A4区兰亭风景小区路段施工,农民	2019-08-19 18:13:20	尊敬的领导,我在走头无路下	0	1
5	2	199229	A00025442	A4区怡然翠园退房事件怎么处理?	2019-09-02 10:30:31	2017年至今马上两年了,国有	0	0
5	2	260596	A00036572	能不能修一下A4区政府机关大院的篮	2019-11-02 17:08:27	政府机关大院里的篮球场多年	0	0
5	3	196664	A000100326	A市医学会泄露考生个人信息如此之	2019-02-22 14:38:33	我是一名医生,也是一名考生,	0	3
5	3	268495	A00094908	A市福田厂给工人频繁调岗、调岗降薪	2019-02-26 23:04:25	您好,很感谢您在百忙之中收	0	0
5	3	223054	A00081404	A市社保局12333电话为什么一直没	2019-04-03 16:36:09	我老婆快生育了,原本想去电	0	16
5	3	224537	A00027493	为什么我每次自主约考A市青竹湖考	2019-05-17 16:03:17	本人第一次在青竹湖考场考科	0	0
5	3	247052	A00086338	A市火车南站公安值班人员态度差	2019-06-30 15:30:49	6月30日上午8点,本人进站后	0	0
5	3	263836	A00047470	A市银华电力金电厂退休职工的房改	2019-09-08 02:07:17	我是银华电力金电厂退休职工,	0	0

图 8 5 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
6	1	235763	A00040573	A市新华铜联官窑公司销售石	2019-01-04 11:04:25	尊敬的书记：您好！我是A市居民，	1	6
6	1	285888	A000106041	举报B市新华驾校、B市交通局	2019-01-07 08:27:07	我想投诉无良驾校、退我血汗钱。五	0	0
6	1	285126	A00092781	A5区明昇壹城二期坑骗业主，	2019-01-07 12:55:44	尊敬的胡书记：我是明昇壹城二期业	0	0
6	1	232892	A00015335	A市万科魅力之城开发商未通	2019-01-08 09:54:00	您好！我是57栋业主，2018年12月2	0	1
6	1	281654	A00031618	举报A市时代倾城小区商业门	2019-01-10 09:58:08	地处时代倾城小区二期门面商铺常年	0	0
6	1	234343	A00039642	A市龙湖新壹城无故拖延客户	2019-02-12 15:37:32	2018年11月24日我和老公在A市龙湖	0	1
6	1	222196	A00028226	C3县石潭镇涟水河电鱼船不断	2019-02-25 16:36:10	石潭镇的涟水河也是的，电鱼船是鱼	0	0
6	1	217403	A000103857	A市万润滨江天著您悠业主延	2019-02-26 18:09:45	本人2018年7月12日在A市万润滨江	0	0
6	1	190462	A00027173	A市安顺驾驶员培训有限公司	2019-03-05 00:31:47	您好！关于驾校教练和学员的纠纷	0	1
6	1	192493	A000103957	A7县震天欢乐水世界广告活	2019-03-12 10:07:21	昨天路过A7县黄兴镇佳辉丽景鑫城	0	0
6	1	240648	A00082198	关于A6区美的翰城精装修的投	2019-03-20 10:12:56	投诉A市A6区美的翰城的精装修收取	0	0
6	1	268940	A00034825	A市好百年家具广场利亚斯家	2019-03-20 16:07:55	胡书记好，2017年中间，我在A市A	0	1
6	1	223530	A00087678	A市橘子洲景区服务不体恤百	2019-03-20 23:26:52	3月16日晚上，我带妻子（怀有身孕	2	0
6	1	188665	A000106234	A市松雅湖东方航标2栋2楼有	2019-03-26 00:18:26	您好 谢谢您能在百忙之中看到这份	0	1
6	1	269088	A0001918	A市香江商贸物流城欺诈骗售	2019-03-26 18:06:52	领导好！我系2017年7月8日全额付	0	0
6	1	268169	A000108008	A3区应加强景秀路流动活禽销	2019-04-06 17:59:09	每到周六，周日上午A3区景秀路工	0	0
6	1	268619	A00042895	A3区梅溪湖创新中心违规建	2019-04-16 15:02:16	A市A3区梅溪湖创新中心自2018年以	0	0
6	1	212540	A00084162	举报A市金科时代中心物业乱	2019-04-24 11:02:13	我是A市金科时代中心写字楼业主，	0	0
6	1	239627	A00063575	A市这边仅凭邮件签收确认企	2019-05-08 16:39:30	仅凭邮件签收确认企业经营地点异	0	0
6	1	208645	A00075393	A7县星湖湾二期房价的疑问	2019-05-11 21:15:55	尊敬的领导：您好！我是星沙星湖湾	0	0
6	1	250666	A000106382	A3区市场监管人员的工作应	2019-05-14 19:48:48	本人今天下午三点，正在会议室开	0	0
6	1	267582	A0004202	A7县行政执法局榔梨分局一主	2019-05-15 14:59:07	根据《关于报送p2p平台借款人逃废	0	0
6	1	258786	A00095795	A1区马坡岭街道新安小区门	2019-05-16 11:14:23	尊敬的领导：您好！ 在2018年10月	0	0
6	1	227746	A00063630	A6区金地自在城，请退回我的	2019-05-28 11:03:31	尊敬的A市住建委领导：您好，A6区	0	0

图 8 6 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
7	1	231442	A00050030	请求协调A6区康乃馨医院解	2019-01-04 22:58:13	我妈妈在20多天前在A6区银星路康乃	0	0
7	1	197160	A00046505	望能够解决去A市各大医院看	2019-03-20 16:20:17	望能够解决去医院看病大部分医院收	0	0
7	1	208655	A00049557	A市中医药研究院附属医院疑	2019-03-22 10:41:10	本人于2011年在西地省中医药研究院	0	0
7	1	219924	A00044049	怀疑A市南湖医院涉嫌销售假	2019-03-22 18:22:37	2019年2月28日本入委托彭美莹在A	0	0
7	1	261702	A000108637	在省中医附一医院治疗皮肤	2019-04-23 16:03:43	因“面部皮肤起红色丘疹”于2019年	0	0
7	1	196249	A000102310	A市楚雅二医院正式员工无购	2019-06-12 20:45:49	楚雅二医院有社保没养老保险竟然不	0	0
7	1	287544	A00086192	楚雅二医院看个病被踢来踢	2019-06-29 20:27:10	本人黄友秀，女，现龄56岁，M5市，	0	0
7	1	224178	A00030989	M市的救护车为何长期盘踞于	2019-08-01 17:46:22	M市救护车为何长期盘踞于A市多家医	0	0
7	1	269409	A00088766	A市星沙八医院不负责	2019-12-04 06:58:24	你好，我是2019年，4月29号，在星	0	0
7	1	195179	A00077887	A市阳光医院无故打 电话骚扰	2019-12-12 15:55:57	A市阳光医院2017年6月至7月期间无	0	0
7	2	281614	A00043022	关于A市医护人员健康证检查	2019-01-03 20:32:06	根据《食品安全法》、《公共场所卫	0	1
7	2	232887	A00088606	A7县妇幼医院连儿童常见病	2019-01-23 15:30:02	书记您好！ 经济强县，民生立县。一	7	7
7	2	288187	A00028389	举报A市黑救护车	2019-04-15 09:26:00	一黑救护车楚A89F93楚A620JV楚A1	0	0
7	2	271076	A00063592	请问现在推行居民健康卡可	2019-05-13 11:37:42	请问13就在推行居民健康卡现在可	0	0
7	2	198973	A00025783	A3区鑫庆医药公司拖欠员工	2019-05-28 16:20:42	西地省鑫庆医药有限公司（原西地省	0	0
7	2	203581	A0007153	A7县卫生健康局侵犯个人隐	2019-06-27 15:44:07	2019年6月，本人与妻子在A7县婚姻	0	0
7	2	229073	A000108884	A1区中联重科退休员工独生	2019-07-03 12:26:20	本人是中联重科混凝土分公司二〇一	0	0
7	2	254254	A00020146	反映A市省妇幼医院住院部重	2019-07-17 09:25:20	住院就必须完善各项检查，似乎成为	0	1
7	2	195089	A00072618	举报A市博长山水香颐小区违	2019-09-04 12:56:14	尊敬的A市政府领导，我们系A市博	0	0
7	3	282987	A00085459	A市计生委怎么不办理独生子	2019-01-06 09:52:02	我与丈夫于2014年8月1日生下一子，	1	1
7	3	261691	A00037654	A市为什么不能分级诊疗？	2019-02-09 12:27:29	1.分级诊疗请问是让患者分吗？（患	0	0
7	3	282959	A00032851	对A市献血积极分子亲属优先	2019-04-15 13:28:08	本人四次献全血、一次献成分血，早	0	0
7	3	247654	A00021549	A市的尿毒症患者太苦了！	2019-07-16 10:14:02	A市的尿毒症患者太苦了！ 尽管尿毒	0	0
7	3	276907	A0008611	投诉A市五洲泌尿专科医院恶	2019-07-18 17:34:24	2019年7月13日，因阴囊感染疼痛，	0	0

图 8 7 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
9	1	287113	A00076743	反映A7县农民土地的征收问题	2019/2/8 10:45:36	最近本人在黄兴，榔梨听说农民土地征	0	2
9	1	243134	A00075927	A7县松雅湖壹号土地使用期限	2019/3/9 22:38:00	我想购买星沙的松雅湖壹号房子，想	0	0
9	1	248372	A00030345	反映A7县楚龙街道高沙村土地	2019/5/23 17:44:47	看到很多偏远地区农村土地都已确权，	0	1

图 8 9 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
11	1	236233	A00077799	投诉A3区公安局经侦大队、银盆岭派出所	2019/3/2 19:47:08	尊敬的胡书记：您好！我是A	0	3

图 8 11 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
14	1	197717	A000106660	关于A市土地确权 and 粮田直补金问题的咨询	2019/7/30 17:53:33	我有两个问题想资讯一下，一	0	0

图 8 14 类热点问题留言明细

类别	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
15	1	189247	A000107866	建议在“我的A市”app中尽快接入法律	2019/12/21 9:45:46	互联网时代，“我的A市”app对于推	0	0

图 8 15 类热点问题留言明细

5.3 聚类性能指标分析

为了评价聚类的质量，定义一种用于衡量聚类效果的指标 SSE（Sum of Squared Error，误差平方和），误差是指样本到其质心的距离。SSE 值越小，表示数据点越接近质心。同时使用轮廓系数进行评估，轮廓系数是一个衡量一个结点与它属聚类相较于其它聚类的相似程度。取值范围-1 到 1，值越大表明这个结点更匹配其属聚类而不与相邻的聚类匹配。如果大多数结点都有很高的轮廓系数值，那么聚类适当。若许多点都有低或者负的值，说明分类过多或者过少。如图 9 所示。

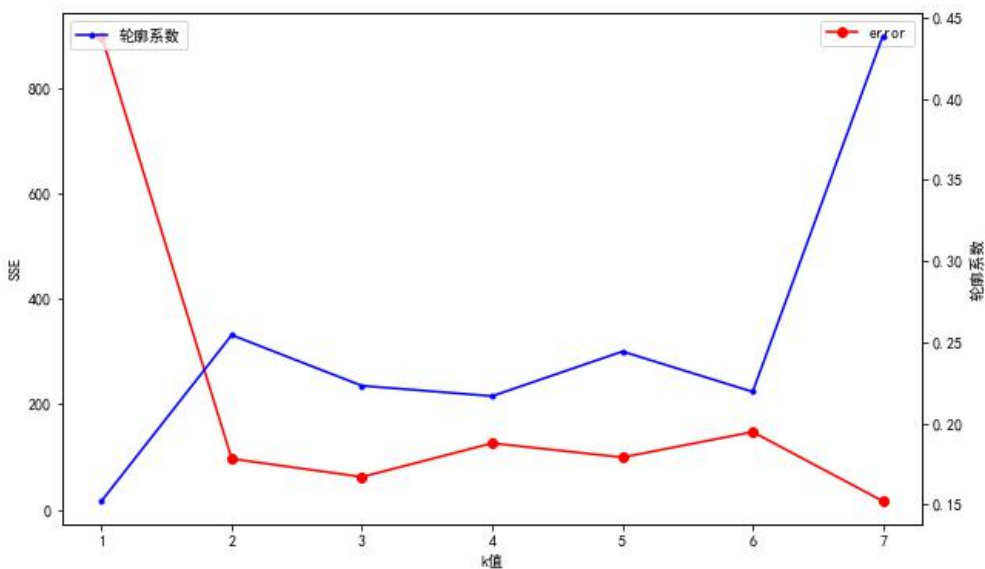


图 9 SSE 和轮廓系数性能指标分析

6. 针对答复意见建立评价方案

6.1 文本数据分析

对类似于网络问政的政民平台，国内外相关研究除了关注公众网络参与及其行为，还关注政府回应与政民互动的展开。在政府回应方面，越来越多的学者开始关注政府回应的方式、特征和条件等，认为网络问政中体现出的政府回应性是民主社会政府合法性的来源之一。网络问政也是一个从输入、输出到反馈的政府回应框架。

根据附件 4，其文本数据的属性为留言编号、留言用户、留言主题、留言详情、留言时间、答复意见和答复时间属性，根据问题三需要对附件 4 进行数据预

处理，从答复时间、数据分类、热点问题等方面进行数据分析，然后从答复意见的及时性、相关性、完整性、可解释性进行研究，最后提出一套评价方案。具体的流程如图 10 所示。

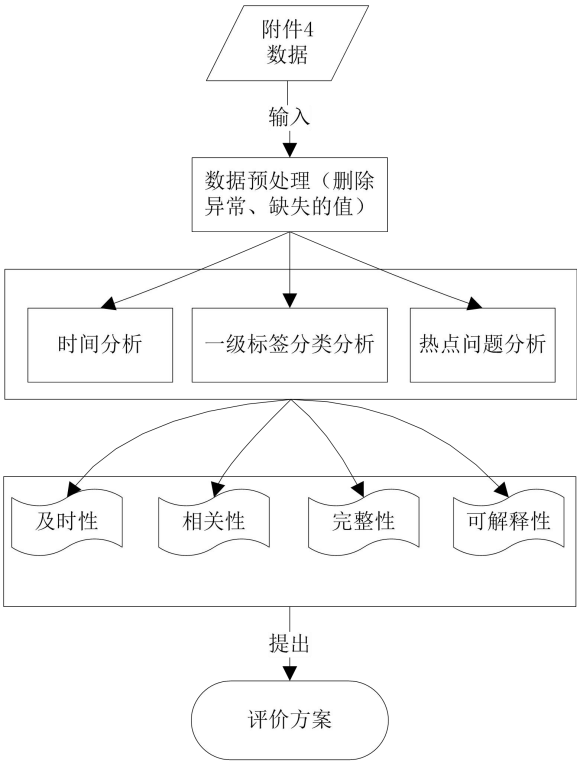


图 10 问题三的流程圖

6.1.1 时间的处理

对附件 4 进行预处理，留言时间和答复时间转换成%Y-%m-%d 格式，根据留言时间和答复时间增加留言年份(year)、留言月份(month)、回复时限(differ) 3 个属性。回复时限指的是答复时间和留言时间之间相差的天数，如式（10）。

回复时限=回复时间-留言时间

(10)

将回复时限采用五段时限法分为五个区间：1-5 天、6-10 天、 11-15 天、16-31 天以及 32 天以上，分别设置标签为{0, 1, 2, 3, 4}，其属性为 labeldays，此时附件 4 的数据如图 11 所示。

Index	留言编号	留言用户	留言主题	留言时间	留言详情	答复意见	答复时间	year	month	differ	labeldays
0	2549	A00045581	A2区景蓉华...	2019-04-25	...	现将网友在...	2019-05-10	2019	4	15	2
1	2554	A00023583	A3区潘楚南...	2019-04-24	...	网友“A00023...	2019-05-09	2019	4	15	2
2	2555	A00031618	请加快速高A...	2019-04-24	...	市民同志：...	2019-05-09	2019	4	15	2
3	2557	A000110735	在A市买公寓...	2019-04-24	...	网友“A00011...	2019-05-09	2019	4	15	2
4	2574	A0009233	关于A市公交...	2019-04-23	...	网友“A00092...	2019-05-09	2019	4	16	3
5	2759	A00077538	A3区含浦镇...	2019-04-08	...	网友“A00077...	2019-05-09	2019	4	31	3
6	2849	A000100804	A3区教师村...	2019-03-29	...	网友“A00010...	2019-05-09	2019	3	41	4
7	3681	UU00812	反映A5区东...	2018-12-31	...	网友“UU0081...	2019-01-29	2018	12	29	3
8	3683	UU008792	反映A市美麓...	2018-12-31	...	网友“UU0087...	2019-01-16	2018	12	16	3
9	3684	UU008687	反映A市洋湖...	2018-12-31	...	网友“UU0086...	2019-01-16	2018	12	16	3
10	3685	UU0082204	反映A2区大...	2018-12-30	...	网友“UU0082...	2019-03-11	2018	12	71	4
11	3692	UU008829	A5区鄱阳村0...	2018-12-29	...	网友“UU0088...	2019-01-29	2018	12	31	3
12	3700	UU00877	A4区万国城...	2018-12-29	...	网友“UU0087...	2019-01-14	2018	12	16	3
13	3704	UU0081480	举报A市芒果...	2018-12-28	...	网友“UU0081...	2019-01-03	2018	12	6	1
14	3713	UU0081227	建议增开A市...	2018-12-28	...	网友“UU0081...	2019-01-14	2018	12	17	3
15	3720	UU008444	关于A市新开...	2018-12-27	...	网友“UU0084...	2019-03-06	2018	12	69	4
16	3727	UU0081194	投诉A3区桐...	2018-12-27	...	网友“UU0081...	2019-01-03	2018	12	7	1
17	3733	UU008706	建议在A市梅...	2018-12-26	...	网友“UU0087...	2019-01-14	2018	12	19	3
18	3747	UU008201	希望相关部...	2018-12-25	...	网友“UU0082...	2019-01-08	2018	12	14	2

图 11 新数据部分展示

以附件 4 为研究对象，该文本数据是 2011 年至 2020 年 1 月之间为时间线，共计 2816 条有效数据，通过分析群众每年留言频数如图 12 所示，10 年间群众每月留言频数如图 13 所示。

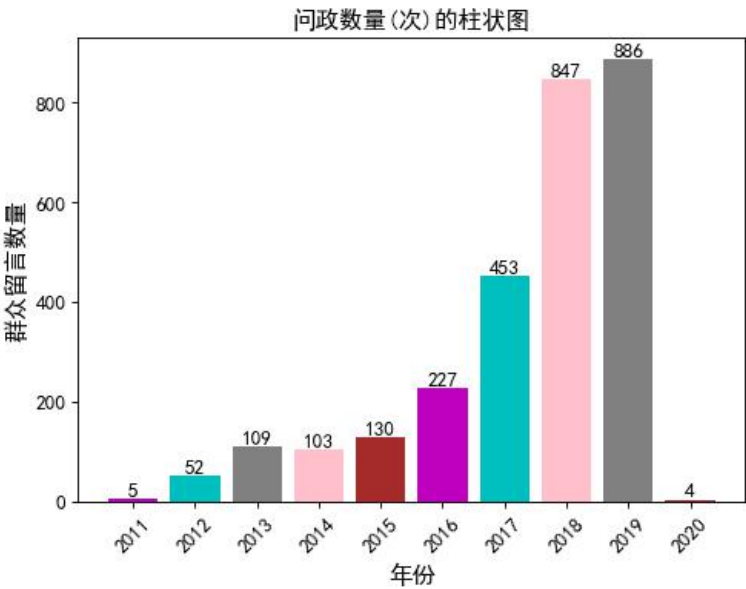


图 12 群众网络问政年柱张图

随着科技的发展，网络问政的逐渐普及其群总网上留言逐年增加，从 2016 年开始群总留言成倍增加，在 2018 年达到稳定状态，说明该网络问政平台在群众中渐渐普及。

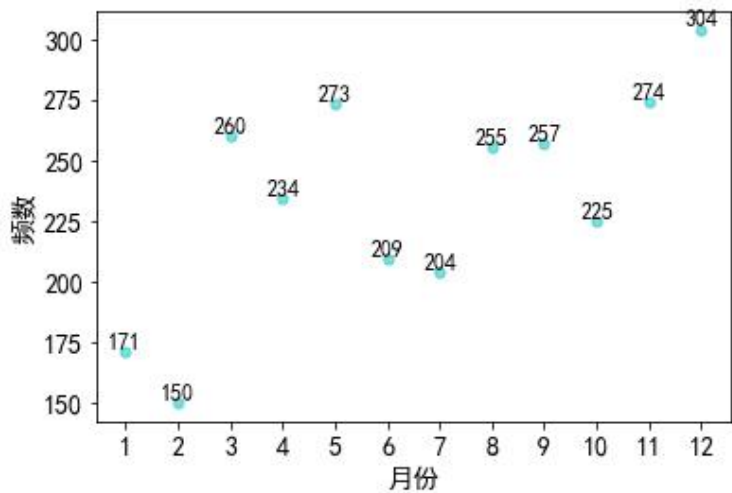


图 13 群众网络问政月散点图

从散点图看出群众在年初和年中的留言频数较少，而在其他月份留言频数相对较多，其中这 10 年间 12 月份群众留言最多为 304 条。

问政数量按年月分布如表 10 所示。总体上来看，网络问政的数量逐年波动增长，这种增长性反映了该网络问政平台具备良好的内在发展动力和外部发展环境。

表 10 网络问政量的年月分布

	2011 年	2012 年	2013 年	2014 年	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	总计
1 月	0	0	0	0	0	25	4	48	0	4	171
2 月	0	6	5	7	0	17	26	31	53	/	150
3 月	0	8	13	11	5	32	44	65	0	/	260
4 月	1	0	0	15	10	0	0	62	0	/	234
5 月	0	0	0	13	4	22	70	72	81	/	273
6 月	0	0	6	0	6	1	0	66	82	/	209
7 月	0	0	10	10	14	9	22	58	79	/	204
8 月	1	5	16	0	8	21	43	85	70	/	255
9 月	0	10	9	9	7	16	0	73	84	/	257
10 月	2	2	15	8	16	23	37	61	61	/	225
11 月	0	3	8	6	15	20	36	114	72	/	274
12 月	1	4	7	2	34	19	49	112	76	/	304
总计	5	52	109	103	130	227	453	847	886	4	2816

6.1.2 文本数据分类

将附件 4 的数据使用问题 1 的算法进行分类，对群众留言进行一级标签分类，以附件 2 的数据作为训练集，附件 4 数据作为测试集，并计算每一大类标签群众留言频数和占比。

问政诉求都是围绕一个主要话题展开的，话题标示着问政者的核心意见关切，是问政主体间交流和互动的基础。问政话题在不同的平台环境中有不同的区分，为利于研究的开展，附件 4 参考第一问的文本数据挖掘方法，将网络问政中的群总留言划分为城乡建设、环境保护、交通运输、教育文体、劳动和社会保障、商贸旅游、卫生计生 7 类（如表 11），通过验证分析，此类别划分具备较好的稳定性和合理性。其中城乡建设、劳动和社会保障、教育文体等话题是网络问政中最主要的话题。

表 11 群众留言一级标签分布

类别	城乡建设	环境保护	交通运输	教育文体	劳动和社会保障	商贸旅游	卫生计生
问政量	1425	195	124	365	440	166	101
比例	50.58%	6.92%	4.40%	12.95%	15.62%	5.89%	3.58%

6.1.3 热点问题

同理将附件 4 的数据根据问题 2 的思想，挖掘出每一级标签分类的热点群众留言，并计算热点数据频数和占比，如表 12、表 13、表 14、表 15、表 16、表 17、表 18 所示。

表 12 城乡建设热点留言频数分布

	Top1	Top2	Top3	Top4	Top5	Top6	other
频数	418	62	30	23	18	16	858
比例	29.33%	4.35%	2.10%	1.64%	1.26%	1.12%	60.21%

表 13 环境保护热点留言频数分布

	Top1	Top2	Top3	Top4	Top5	Top6	other
频数	14	10	10	6	6	5	144
比例	7.18%	5.13%	5.13%	3.08%	3.08%	2.56%	7.39%

表 14 交通运输热点留言频数分布

	Top1	Top2	Top3	Top4	Top5	Top6	other
频数	9	7	6	5	5	5	87
比例	7.25%	5.65%	4.84%	4.03%	4.03%	4.03%	70.16%

表 15 教育文体热点留言频数分布

	Top1	Top2	Top3	Top4	Top5	Top6	other
频数	86	15	11	10	7	6	230
比例	23.56%	4.11%	3.01%	2.74%	1.92%	1.64%	63.01%

表 16 劳动和社会保障热点留言频数分布

	Top1	Top2	Top3	Top4	Top5	Top6	other
频数	14	9	8	8	7	7	387
比例	3.18%	2.05%	1.81%	1.81%	1.59%	1.59%	87.95%

表 17 商贸旅游保障热点留言频数分布

	Top1	Top2	Top3	Top4	Top5	Top6	other
频数	49	9	4	4	4	4	92
比例	29.51%	5.42%	2.41%	2.41%	2.41%	2.41%	55.42%

表 18 卫生计生保障热点留言频数分布

	Top1	Top2	Top3	Top4	Top5	Top6	other
频数	11	9	6	4	4	4	63
比例	10.89%	8.91%	5.94%	3.96%	3.96%	3.96%	62.38%

6.2 数据研究

要根据答复意见的内容从时效性、相关性、完整性、可解释性的角度对相关部门的答复质量提出一套评价方案，步骤如下。

6.2.1 时效性

从表 19 看出，2011 年到 2020 年之间 1-5 天相关部门随着群众留言频数的额增加其答复时限也在提升，说明政民互动提高了社会生活质量，节约了时间成本。其中 a/b , a 表示频数, b 表示占比。

表 19 网络问政答复时效性分布

	2011 年	2012 年	2013 年	2014 年	2015 年	2016 年	2017 年	2018 年	2019 年	2020 年	总计
1-5 天	1/20%	20/38%	39/36%	24/23.5%	38/29%	76/33%	110/24%	176/20.5%	308/35%	4/100%	796/28%
6—10 天	0/0	10/19%	24/22%	23/22%	41/32%	32/14%	63/14%	177/21.5%	175/20%	0	545/19%
11-15 天	1/20%	4/8%	8/7%	15/15%	23/18%	37/16%	40/9%	161/19%	125/14%	0	414/15%
16-31 天	0/0	8/16%	12/11%	18/17%	17/13%	57/25%	155/34%	199/23%	159/18%	0	625/22%
32 天以上	3/60%	10/19%	26/24%	23/22.5%	11/8%	25/12%	85/19%	134/16%	119/13%	0	436/16%
总计	5/100%	52/100%	109/100%	103/100%	130/100%	227/100%	453/100%	847/100%	886/100%	4/100%	2816/100%

从及时性角度对每一级标签大类进行分析，如表 20 所示，可以看出交通运输、教育文体、劳动社会保障相关部门答复时效 1-5 天的占比都在 30%左右，说明这三大类是社会发展、经济增长的重要环节，其它一级标签大类每个阶段答复效率分布均匀。

表 20 群众留言时效性一级标签分布

	城乡建设	环境保护	交通运输	教育文体	劳动和社会保障	商贸旅游	卫生计生
1-5 天	393/28%	50/26%	36/29%	111/30%	141/32%	42/25%	23/22. 5%
6—10 天	270/19%	40/21%	19/15. 3%	81/22%	85/19%	32/19%	18/18%
11-15 天	193/13%	26/13%	18/14. 5%	67/18. 5%	54/12%	36/22%	20/20%
16-31 天	342/24%	46/24%	27/21. 7%	67/18. 5%	90/21%	30/18%	23/22. 5%
32 天以上	227/16%	33/16%	24/19. 5%	39/11%	70/16%	26/16%	17/17%
总计	1425	195	124	365	440	166	101

从及时性角度对热门话题的答复意见进行研究，如表 21 对热点排名前五的数据进行分析，可以看出群众留言的热点问题相关部门答复效率在一个月之内回复的概率都为 80%以上，说明热点问题答复效率较高。

表 21top5 留言热点时效性分布

	Top1	Top2	Top3	Top4	Top5
1-5 天	107/26%	23/27%	12/19. 2%	12/24. 5%	9/30%
6—10 天	76/18%	18/21%	12/19. 2%	10/20. 4%	5/17%
11-15 天	48/11. 5%	14/16%	10/16. 4%	8/16. 3%	6/20%
16-31 天	117/28%	16/18. 5%	16/26%	10/20. 4%	3/10%
32 天以上	70/16. 5%	15/17. 5%	12/19. 2%	9/18. 4%	7/23%
总计	418	86	62	49	30

6.2.2 相关性（杰卡德系数）

从相关性角度对相关部门答复意见进行研究，所谓相关性就是求解答复意见和留言主题或留言详情是否相似，因此通过 jieba 分词及去除停用词算法对每条群总留言和答复意见数据建立留言关键词、答复关键词两个向量，然后通过杰卡德系数对这两个向量计算文本相似度，数值越大，两个向量相似度越高。

Jaccard inde 又称为 Jaccard 相似系数(Jaccard similarity coefficient) 用于比较有限样本集之间的相似性与差异性。Jaccard 系数值越大，样本相似度越高。

给定两个向量 A、B，Jaccard 系数定义为 A 与 B 交集的大小与 A 与 B 并集的大小的比值，定义如式（11）：

$$J(A,B)=\frac{|A\cap B|}{|A\cup B|}=\frac{|A\cap B|}{|A|+|B|-|A\cap B|}$$

(11)

当集合 A，B 都为空时，J(A,B) 定义为 1。

与 Jaccard 系数相关的指标叫做 Jaccard 距离, 用于描述集合之间的不相似度。Jaccard 距离越大, 样本相似度越低。公式定义如式 (12) :

$$d_j(A, B) = 1 - J(A, B) \quad (12)$$

由于文本数据与很多干扰的词语, 所以这里当 $0.1 \leq \text{similar}$ 的情况下认为留言关键词和答复关键词这两个向量是相关。当 $0.05 \leq \text{similar} < 0.1$ 认为这两个向量半相关, 当 $0 < \text{similar} < 0.05$ 认为这两个向量弱相关, 最后当 $\text{similar} = 0$ 说明这两个向量完全不相关。并设置标签为 {3, 2, 1, 0}, 其属性为 labelsimilar。

从相关性角度对一级标签的相关部门答复意见进行相似度研究, 通过表 22 看出, 不相关的数据较少只有 301 条, 占总数据的 10.69%, 说明接近 90% 的留言关键词和答复关键词向量都是相关的。

表 22 群众留言相关性一级标签分布

相关度	城乡建设	环境保护	交通运输	教育文体	劳动和社会保障	商贸旅游	卫生计生	总计
相关	514	249	36	151	133	43	35	961
半相关	479	80	43	98	137	54	34	925
弱相关	289	45	29	75	127	41	23	629
不相关	143	21	16	41	43	28	9	301
总计	1425	195	124	365	440	166	101	2816

6.2.3 完整性

相关部门答复意见的完整性指的是相关部门答复了群总留言, 不关注回复时限, 只要答复意见不是网络自动答复或空值, 这条答复意见就是完整的。图 13 是经过以上步骤的数据处理及分析, 附件 4 数据处理之后完整的数据, 保存为 D:/附件/问题三/data2.csv

Index	留言编号	留言用户	留言主题	留言时间	留言详情	答复意见	答复时间	year	month	differ	留言关键词	答复关键词	一级标签	similiar	labeldays	labelsimiliar
3	2557	A00011...	在A市...	2019-...	...	网友“A00...	2019-05-09	2019	4	15	公寓 人才 ...	网友 A000...	5	0.126761	2	3
4	2574	A00092...	关于A...	2019-...	...	网友“A00...	2019-05-09	2019	4	16	公交站点 名...	网友 A000...	1	0.322581	3	3
5	2759	A00077...	A3区...	2019-...	...	网友“A00...	2019-05-09	2019	4	31	含浦镇 马路...	网友 A000...	1	0.0923077	3	2
6	2849	A00010...	A3区...	2019-...	...	网友“A00...	2019-05-09	2019	3	41	盼望 早日 ...	网友 A000...	1	0.0761905	4	2
7	3681	UU00812	反映A...	2018-...	...	网友“UU0...	2019-01-29	2018	12	29	东洲湾 社区...	网友 UU00...	1	0.0677083	3	2
8	3683	UU0087...	反映A...	2018-...	...	网友“UU0...	2019-01-16	2018	12	16	美麓 阳光 ...	网友 UU00...	1	0.083871	3	2
9	3684	UU0086...	反映A...	2018-...	...	网友“UU0...	2019-01-16	2018	12	16	洋湖 新城 ...	网友 UU00...	1	0.0992908	3	2
10	3685	UU0082...	反映A...	2018-...	...	网友“UU0...	2019-03-11	2018	12	71	大托 街道 ...	网友 UU00...	1	0.0700637	4	2
11	3692	UU0088...	A5区...	2018-...	...	网友“UU0...	2019-01-29	2018	12	31	鄱阳 安置房...	网友 UU00...	1	0.0978261	3	2
12	3700	UU00877	A4区...	2018-...	...	网友“UU0...	2019-01-14	2018	12	16	万国 段 修...	网友 UU00...	1	0.0277778	3	1
13	3704	UU0081...	举报A...	2018-...	...	网友“UU0...	2019-01-03	2018	12	6	举报 芒果 ...	网友 UU00...	6	0.0133333	1	1
14	3713	UU0081...	建议...	2018-...	...	网友“UU0...	2019-01-14	2018	12	17	增开 261 公...	网友 UU00...	1	0.0615305	3	2
15	3720	UU0084...	关于A...	2018-...	...	网友“UU0...	2019-03-06	2018	12	69	新开铺 披塘...	网友 UU00...	1	0.144796	4	3
16	3727	UU0081...	投诉A...	2018-...	...	网友“UU0...	2019-01-03	2018	12	7	投诉 桐梓坡...	网友 UU00...	6	0.0121951	1	1
17	3733	UU0087...	建议...	2018-...	...	网友“UU0...	2019-01-14	2018	12	19	梅溪湖 开办...	网友 UU00...	4	0.15	3	3
18	3747	UU0082...	希望...	2018-...	...	网友“UU0...	2019-01-08	2018	12	14	治理 中海 ...	网友 UU00...	1	0.0701754	2	2
19	3755	UU0081...	希望A...	2018-...	...	网友“UU0...	2019-01-04	2018	12	10	社保卡 医保...	网友 UU00...	5	0.0434783	1	1
20	3756	UU0081...	希望A...	2018-...	...	网友“UU0...	2019-01-04	2018	12	10	一卡通 手机...	网友 UU00...	1	0.0983607	1	2

图 13 最终结果数据部分展示

6.2.4 可解释性

所谓可解释性就是相关部门对群总留言及时回复，并且其答复意见与群总留言的信息是相关的，所以相关部门答复意见的可解释性必须同时满足上述两条条件。如表 23 所示，这些数据群众答复意见是不可行的。

表 23 不可解释性部分数据展示

留言主题	留言时间	答复意见	答复时间	differ	similiar	labeldays	labelsimiliar
咨询低保、残疾人补助的相关问题	2014/6/14	你好。请问当地民政部门。2017年8月18日	2017/8/18	1161	0	4	0
反映 A8 县金洲新区高新安置房分配问题	2014/4/10	网友：您好！留言已收悉	2014/5/16	36	0	4	0
建议 B9 市规划一个校车接送计划	2018/12/7	2018/12/12	2018/12/13	6	0	1	0

6.3 评价方案

通过时效性、相关性、完整性、可解释性角度对相关部门答复意见进行分析，得出以下结论。

在及时性方面，政府超过 31 天答复的所占比列为 16%；在相关性方面，相关部门答复意见与留言数据的不相关的数据占 10.69%；在完整性方面与相关性一样；在可解释性方面，必须同时满足相关部门及时回复群总留言，并且其答复意见与群总留言的信息是相关的，不满足的数据为政府超过 31 天答复的数和答复意见与留言不相关的数据的并集，通过计算不满足可解释性的数据拥有 690 条，占总数据的 24.5%，通过以上分析和推理，相关部门的答复质量不管从什么角度分析其有效数据都在 75%以上，其质量评价效果为良好。

评价方案通过 6.1 文本数据分析和 6.2 数据研究可得。

7. 总结

在群众留言分类问题方面。结合实际数据需要，对数据进行预处理的到关键词汇，根据关键特征对政务文本进行 TF-IDF 方法计算，分别利用贝叶斯算法、基于 TF-IDF 的贝叶斯模型和基于 TF-IDF 的支持向量机（SVM）模型来构建留言内容的一级标签分类模型。总的来说基于 TF-IDF 的支持向量机（SVM）模型的分类效果较好，能有效解决政府职工人员的服务效率问题。

在热点问题挖掘方面。基于群众留言分类问题建立的一级标签分类模型，进行附件 3 政务文本的一级标签分类，再利用层次聚类算法确定 K-means 聚类算法的 k 值文本进行聚类，根据 SSE 和轮廓系数进行聚类算法的评估，其效果还是不错的。

在答复意见评价问题方面。基于群众留言分类和热点问题挖掘问题，以附件 4 为研究对象，通过群众留言和政府回复的内容在及时性、相关性、完整性、可解释性角度进行分析，实现了一套评价方案。

参考文献

- [1] 杨萌萌, 黄浩, 程露红等. 基于 LDA 主题模型的短文本分类[J]. 计算机工程与设计, 2016, 37(12): 3372-3377.
- [2] 烧云飞. 融合主题模型与词向量的短文本分类方法研究[D]. 西安: 西安电子科技大学,

2019.

- [3] 程春惠. 公安犯罪案件文本挖掘关键技术研究[D].浙江: 浙江大学, 2010.
- [4] 尚迪. 社交媒体文本数据挖掘算法及应用[D].南京: 南京大学, 2017.
- [5] 姜大鹏. 基于词向量的短文本分类方法研究[D].浙江: 浙江大学, 2015.
- [6] 袁军鹏, 朱东华, 李毅等. 文本挖掘技术研究进展[J].计算机应用研究, 2006, 23(2): 1-4.
- [7] ZHANG Zhifei, MIAO Duoqian, GAO Can. Short text classification using latent Dirichlet allocation [J].Journal of Computer Applications, 2013, 33 (6): 1587-1590.
- [8] 王振振, 何明, 杜永萍. 基于 LDA 主题模型的文本相似度计算[J].计算机科学, 2013, 40(12): 229-232.
- [9] 贾璇. 基于文本挖掘的求职软件顾客评论情感分析[J].科技与创新, 2019(17):1-4.
- [10] 郭立秀. 基于文本挖掘的生鲜电商顾客满意度研究[D].成都: 西南交通大学, 2018.
- [11] 于君博, 李慧龙, 于书鹄. “网络问政”中的回应性——对 K 市领导信箱的一个探索性研究[J].长白学, 2018(02): 65-74.
- [12] 李清镇. 基于文本挖掘的笔记本电脑网评分析[D].兰州财经大学, 2019.
- [13] 李少温. 基于网络问政平台大数据挖掘的公众参与和政府回应问题研究[D].华中科技大学, 2019.
- [14] 李锋. 条块关系视野下的网络诉求与政府回应模式研究——基于中部某市网络问政平台的大数据分析[J].电子政务, 2019(05): 27-37
- [15] 沙勇忠, 王峥嵘, 詹建. 政民互动行为如何影响网络问政效果?——基于“问政泸州”的大数据探索与推论[J].公共管理学报, 2019, 16(02): 15-27+16
- [16] 李艳红, 程翔. 基于网络论坛文本挖掘的笔记本电脑满意度研究[J].微型机与应用, 2014, 33(18): 61-65.

附录

问题 1 的代码、数据、结果文件, 在 D:/附件/问题一/~
问题 2 的代码、数据、结果文件, 在 D:/附件/问题二/~
问题 3 的代码、数据、结果文件, 在 D:/附件/问题三/~