

“智慧政务”中的文本挖掘应用

摘要：

近年来，随着互联网的广泛应用和网络留言的迅速发展，网络留言信息平台已成为群众反映社会问题的重要渠道。因此，运用网络文本分析和数据挖掘技术对网络问政平台的研究具有重大的意义。本文是利用智慧政务系统收集到的互联网公开来源的群众问政留言记录及相关部门对部分群众留言的答复意见信息，通过 R 软件建模解决网络问政留言平台中遇到的实际问题。

问题一：群众留言分类。利用文本分词和关键词提取的方法对留言主题进行文本挖掘，通过已知的数据信息，建立一套一级标签分类系统

问题二：热点问题挖掘。根据给出的一定时间内的留言内容，结合留言频次和群众对留言的关注度，通过综合筛选找出某段特定时间内特定地点或特定人群的留言热点问题，建立一套完整的热点问题评价体系。

问题三：答复意见的评价。根据给出的相关部门对留言的答复情况，通过按比例抽样的方式进行抽查，从答复的时间、字数、相关性、完整性和可解释性等多方面对答复意见质量进行综合评价，并通过结果给出相关部门今后工作的建议。

关键词： 中文分词 热点问题 综合评价 朴素贝叶斯算法 R 软件

目录

一、挖掘目标	3
二、分析方法与过程	3
2.1 问题 1 分析方法与过程	4
2.1.1 流程图.....	4
2.1.2 数据预处理.....	4
2.1.3 关键词的提取.....	8
2.1.4 F-Score 算法	10
2.2 问题 2 分析方法与过程.....	11
2.2.1 数据预处理.....	11
2.2.2 数据筛选	11
2.3 问题 3 分析方法与过程.....	11
2.3.1 抽样过程	11
三、结论与建议	14
3.1 结论.....	14
3.2 建议.....	14
四、 参考文献	14

一、挖掘目标

本次建模目标是利用智慧政务系统收集到的互联网公开来源的群众问政留言记录及相关部门对部分群众留言的答复意见信息，利用 jieba 中文分词工具对留言主题进行分词、通过 R 对关键词筛选及统计，达到以下三个目标：

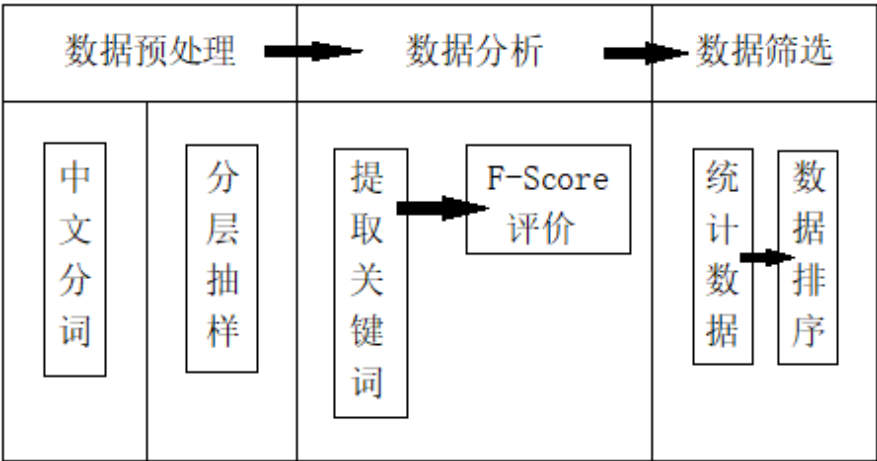
（1）利用文本分词和关键词提取的方法对留言主题进行文本挖掘，根据关键词提取的结果，结合题目给出的一级分类标题建立模型，使其形成一套完整且准确率较高的分类系统。

（2）根据给出的一定时间内的留言内容，结合留言频次和群众对留言的关注度，通过综合筛选找出某段特定时间内特定地点或特定人群的留言热点问题，建立一套完整的热点问题评价体系。

（3）根据给出的相关部门对留言的答复情况，通过按比例抽样的方式进行抽查，从答复的时间、字数、相关性、完整性和可解释性等多方面对答复意见质量进行综合评价，并通过结果给出相关部门今后工作的建议。

二、分析方法与过程

总体流程：



图一 总体流程

具体步骤：

本题主要包括如下步骤：

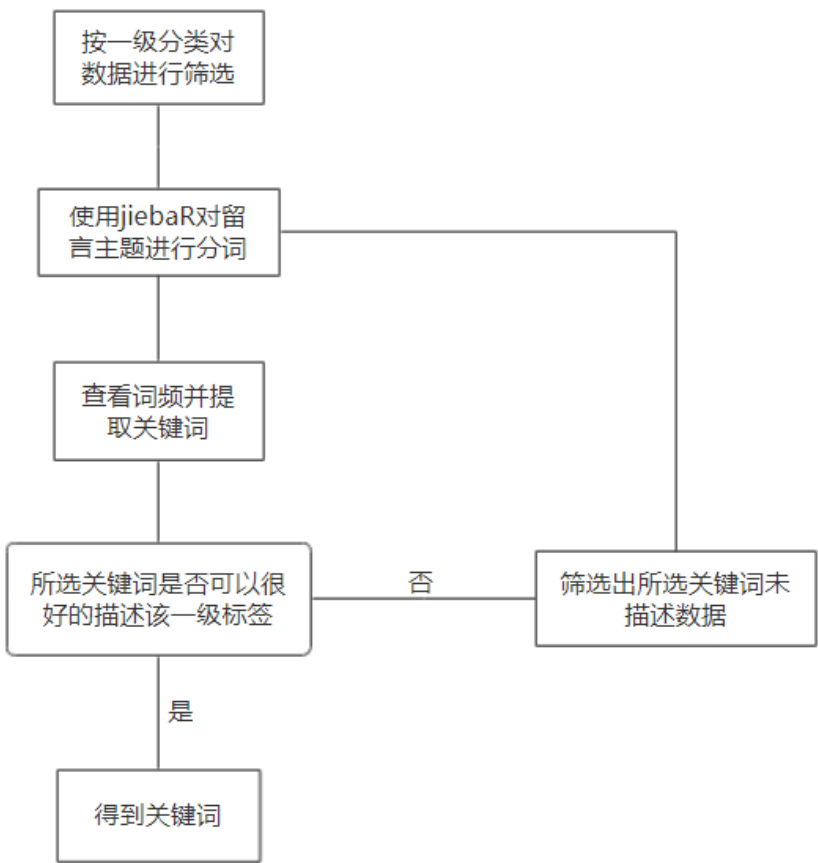
步骤一：数据预处理。将题目给出的数据进行中文分词，并在研究第三问时将数据进行分层、抽样。

步骤二：数据分析。在对数据进行分词后，找出每个一级标签描述的关键词并进行分类。采用 F-Score 算方对模型进行评价。对第三问抽样后的数据从相关性、完整性、可解释性等角度评价数据。

步骤三：数据筛选。统计相关数据，分类筛选汇总找出热点问题。

2.1 问题 1 分析方法与过程

2.1.1 流程图



图二 分类流程

2.1.2 数据预处理

2.1.2.1 对留言主题表进行中文分词

首先对附件 2 给出的内容按照一级标签的分类方式进行分类, 然后对分类后的内容分别进行文本挖掘。

在对留言内容进行挖掘分析之前, 先要把非结构化的文本信息转换为计算机能够识别的结构化信息。在附件 1 内容分类三级标签体系中, 以中文文本的方式给出了分类体系。为了便于分类, 先要对这些留言主题进行中文分词。这里采用 R 的中文分词包 jieba 进行分词。jieba 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的情况所构成的有向无环图 (DAG), 同时采用了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, 对于未登录词, 采用了基于汉字成词能力的 HMM 模型, 使得能更好的实现中文分词效果。

在分词的同时, 通过 R 提取关键词, 抽取留言主题中与内容相关的高频关键词, 这里采用 jieba 自带的语义词。

2.1.2.2 词频统计

在对留言主题进行分词后, 需要统计出在每个类别里关键词的频率, 以供建立合理的分类模型。通过 R 软件进行词频统计, 为了提高精确度对词性和数字进行限定, 统计程序如下:

```
f<-scan('E:/2.txt', sep='\\n', what='', encoding="UTF-8")
seg<-qseg[f]
seg<-seg[nchar(seg)>1]
seg<-table(seg)
seg<-seg[!grepl('[0-9]+', names(seg))]
length(seg)
seg<-sort(seg, decreasing=T)[1:100]
seg
```

对每个一级标签类别的词频统计, 结果如下图所示:

问题	小区	关于	房产证	公积金	严重	建议	解决
222	213	91	83	83	81	71	69
改造	规划	建设	质量	公园	居民	请求	住房
68	67	67	65	64	63	61	60
公司	何时	反映	违规	房屋	没有	为何	为什么
55	53	51	51	50	50	50	50
业主	什么	廉租房	咨询	项目	县县	办理	时候
49	48	46	46	45	43	42	42
投诉	扰民	社区	天然气	购房	公交车	违法	自来水
42	41	41	41	40	39	39	39
不能	工程	危房改造	市场	违建	补贴	存在	租房
38	38	38	37	37	36	36	36
广场	棚户区	建筑	希望	贷款	违章建筑	安置	工程款
35	35	34	34	33	33	32	32
管理	举报	可以	路灯	请问	申请	拖欠	县城
32	32	32	32	32	31	31	31
中心	花园	街道	垃圾	影响	政策	拆迁	国际
31	30	30	29	28	28	27	27
家园	开发商	施工	拆除	二期	房地产	物业	相关
27	27	27	26	26	26	26	25
不动产	附近	农村	有限公司	招租	收费	村民	房子
23	23	23	23	23	22	21	21
西地省	怎么	安全	何时能	建房	停水	无法	城管
21	21	20	20	20	20	20	19
道路	公路	开通	生活				
19	19	19	19				

图三 城乡建设词频

seg	医院	问题	西地省	咨询	政策	人民	生育
180	180	70	66	64	56	52	49
独生子女	独生子女	关于	请问	医生	二胎	卫生院	计划生育
47	47	43	43	43	41	34	32
医疗	医疗	办理	小孩	再婚	超生	患者	什么
29	29	28	28	26	25	23	23
行医	行医	何时	计生	家庭	准生证	请求	卫生局
22	22	22	22	22	22	21	21
是否	是否	为什么	上户	一个	奖励	时候	享受
20	20	20	19	19	18	18	18
药品	药品	疫苗	执业	独生子女证	二孩	非法	合法
18	18	18	18	17	17	17	17
可以	可以	卫生室	诊所	证明	病人	罚款	妇幼保健
17	17	17	17	17	16	16	16
孩子	孩子	户口	计生办	建议	市中心	为何	医师
16	16	16	16	16	16	16	16
计委	计委	社会	卫生	希望	相关	不能	反映
15	15	15	15	15	15	14	14
抚养费	抚养费	情况	投诉	出台	单独	第一	关注
14	14	14	14	13	13	13	13
结扎	结扎	领导	乱收费	人员	退休	服务	举报
13	13	13	13	13	13	12	12
没有	没有	人口	收费	中医院	工作	解决	落实
12	12	12	12	12	11	11	11
乡镇	乡镇	医保	百姓	发放	公共卫生	检查	农村
11	11	11	10	10	10	10	10
如何	如何	违规	新生儿	要求	医疗事故	补助	承包
10	10	10	10	10	9	9	9
楚雅	楚雅	待遇					
9	9	9					

图四 计生卫生词频

seg	出租车	快递	的士	交通	客运	问题	路面	收费
112	112	69	55	39	39	38	37	35
邮政	邮政	乱收费	严重	司机	建议	道路	公司	物流
35	35	33	29	27	24	23	23	22
公路	公路	何时	滴滴	运输	收取	关于	取件	出行
21	21	21	19	19	18	17	15	14
打表	打表	投诉	希望	涨价	管理	坑坑洼洼	邮政储蓄	超载
14	14	14	14	14	12	12	12	11
高速公路	高速公路	什么	为什么	西地省	反映	火车站	农村	为何
11	11	11	11	11	10	10	10	10
银行	银行	邮政局	重视	包裹	补贴	非法	何时能	请求
10	10	10	10	9	9	9	9	9
请问	请问	燃油	现象	EMS	办理	规划	货车	建设
9	9	9	9	8	8	8	8	8
没有	没有	市场	无人	咨询	资格证	城市	费用	规范
8	8	8	8	8	8	7	7	7
行业	行业	经营	路段	是否	县县	乡镇	硬化	政策
7	7	7	7	7	7	7	7	7
班车	班车	不能	车辆	存在	发展	服务	高速	高铁
6	6	6	6	6	6	6	6	6
关注	关注	还是	价格	解决	经常	拒载	路基	旅客
6	6	6	6	6	6	6	6	6
马路	马路	破烂不堪	汽车站	桥梁	通车	通行	网点	违规
6	6	6	6	6	6	6	6	6
县城	县城	小区	应该	政府				
6	6	6	6	6				

图五 交通运输词频

seg	问题	社保	西地省	职工	咨询	工资	关于	人员
277	277	198	197	172	163	136	134	134
退休	退休	医保	员工	解决	请求	公司	为何	工作
134	134	123	123	114	114	88	87	81
政策	政策	养老保险	企业	请问	报销	事业单位	有限公司	公务员
79	79	78	74	72	71	71	71	67
待遇	待遇	养老金	不能	办理	什么	新农	拖欠	工伤
66	66	64	63	57	57	56	53	52
何时	何时	反映	为什么	发放	考试	退休职工	农民工	是否
51	51	50	50	49	46	45	44	44
加班	加班	建议	劳动	就业	怎么	补贴	可以	缴纳
43	43	43	43	42	41	40	40	39
没有	没有	农村	时候	生育	享受	相关	如何	单位
39	39	38	38	37	36	35	34	33
医疗保险	医疗保险	保险	失业	乡镇	社局	投诉	缴费	津贴
33	33	32	32	32	31	31	30	30
劳动合同	劳动合同	要求	医院	西地	调整	报告	改制	居民
30	30	30	30	29	28	27	27	27
申请	申请	我们	机关	希望	参加	落实	社保局	下岗职工
27	27	27	26	26	25	25	25	25
工作人员	工作人员	退休金	异地	职称	标准	能否	市人	保障
24	24	23	23	23	22	22	22	21
合理	合理	何时能	劳动法	领导	社保卡	生活	才能	查询
21	21	21	21	21	21	21	20	20
费用	费用	工人	公平	购买				
20	20	20	20	20				

图六 劳动和社会保障词频

seg	电梯	传销	小区	问题	垄断	西地省	市场	建议
	163	151	89	88	86	71	55	54
	景区	乱收费	投诉	收费	旅游	公司	涉嫌	价格
	53	51	50	49	48	43	42	40
	举报	质量	严重	有限公司	故障	存在	组织	经营
	38	35	34	34	33	31	31	28
	违规	液化气	关于	是否	猪肉	屠宰场	为何	县县
	27	26	25	25	24	23	23	22
	诈骗	安全隐患	检疫	收取	销售	燃气	为什么	班车
	22	21	21	21	21	20	20	19
	打击	反映	非法	欺骗	请问	有人	产品	欺诈
	19	19	19	19	19	19	18	18
	窝点	西地	发展	广场	建设	请求	消费者	直销
	18	18	17	17	17	17	17	17
	安全	行为	活动	频发	涨价	部门	查处	购买
	16	16	16	16	16	15	15	15
	国际	屠宰	现象	月湖	管理	加强	检验	食品
	15	15	15	15	14	14	14	14
	虚假	业主	超市	公交车	广告	希望	严厉打击	一个
	14	14	13	13	13	13	13	13
	怎么	办理	合格	经常	老人	年检	中心	咨询
	13	12	12	12	12	12	12	12
	打造	解决	开发商	没有	煤气	无法	无证	物业
	11	11	11	11	11	11	11	11
	协会	宣传	风景区	华府				
	11	11	10	10				

图七 商贸旅游词频

seg	污染	严重	噪音	排放	居民	污染环境	问题	扰民
	190	152	74	70	61	56	53	52
	公司	环境	养猪场	污水	村民	有限公司	非法	生活
	44	39	39	38	36	35	34	33
	影响	排污	小区	环保局	附近	关于	环保	生产
	32	30	29	28	26	26	26	26
	砖厂	垃圾	违规	反映	环境污染	工厂	破坏	毒气
	26	24	24	23	23	22	22	21
	废气	请求	违法	周边	环评	采石场	基站	举报
	21	21	21	21	20	19	19	19
	水源	市镇	投诉	臭气熏天	工业园	健康	解决	气味
	18	17	17	16	16	16	16	15
	取缔	生态环境	造纸厂	刺鼻	化工厂	加工厂	开采	空气
	15	15	15	14	14	14	14	14
	企业	气体	危害	西地省	直接	直排	导致	关停
	14	14	14	14	14	14	13	13
	居民区	山镇	塘村	偷排	信号	猪场	无人管	无证
	13	13	13	13	13	13	12	12
	严重破坏	冶炼厂	油烟	发射塔	街道	每天	一个	长期
	12	12	12	11	11	11	11	11
	整治	重金属	拆除	楚江	处理	大量	废水	焚烧
	11	11	10	10	10	10	10	10
	社区	谁来管	晚上	百姓	报告	多年	化工	建设
	10	10	10	9	9	9	9	9
	老百姓	难闻	如何	山村				
	9	9	9	9				

图八 环境保护词频

seg	教师	问题	中学	学校	小学	学生	补课	教育局
	221	163	144	143	131	126	121	108
	教育	关于	老师	违规	招生	建议	解决	幼儿园
	103	74	67	64	59	58	52	51
	乱收费	西地省	请求	收费	培训	为什么	咨询	反映
	47	47	46	45	43	42	41	39
	孩子	为何	招聘	一中	机构	文化	希望	不能
	39	39	37	35	34	34	34	31
	小孩	严重	高中	政策	实验	是否	学院	考试
	31	28	27	27	26	26	26	25
	市二中	存在	农村	什么	收取	投诉	县一中	学籍
	25	24	24	24	24	24	24	24
	代课	读书	公平	民办	学费	请问	入学	中心小学
	22	22	22	22	22	21	21	21
	发展	如何	要求	重视	家长	举报	领导	强制
	20	20	20	20	19	19	19	19
	退休	一个	职业	初中	待遇	非法	工资	参加
	19	19	19	18	18	18	18	17
	高考	工作	管理	建设	没有	现象	职称	中心
	17	17	17	17	17	17	17	17
	二中	发放	乱象	落实	我们	相关	中考	中小学
	16	16	16	16	16	16	16	16
	班主任	行为	实验学校	暑假	有倦	转学	办学	城区
	15	15	15	15	15	15	14	14
	国家	何时	可以	录取				
	14	14	14	14				

图九 教育文体词频

2.1.3 关键词的提取

根据以上过程，我们对各类一级标签关键词的选取结果如下：

城乡建设：

工程、地铁、绿化、房产、人行道、洒水、基础设施、夜宵、业主、物业管理、用电、棚改、城铁、环境卫生、交房、摆摊、违建、危房房产证、公积金、改造、规划、公园、房屋、住房、廉租房、危房改造、租房、天然气、自来水、购房、建筑、路灯、贷款、花园、拆迁、建筑、公厕、供暖、停电、自来水、公交车、棚户区、家园、开发商、施工、二期、房地产、不动产、招标、房子、建房、停水、城管。

商贸旅游：

班车、广告、燃气、虚假、宣传、驾校、营业执照、中巴车、开户费、客车、欺诈、全票、物业费、车费、春运、房价、工商、敛财、涨价、票价、平台、质监局、质量技术监督局、保健品、超市、电梯、传销、垄断、景区、旅游、涉嫌、价格、故障、液化气、猪肉、屠宰、检疫、窝点、消费者、直销、频发、检验、严厉打击、年检、协会、风景区。

卫生计生：

行医、药品、准生证、执医、卫计委、村医、生育、计生、院长、药店、医务、结扎、中医、抚养费、医院、独生子女、医生、二胎、再婚、超生、患者、行医、家庭、准生证、卫生局、上户、奖励、疫苗、执业、二孩、卫生室、诊所、病人、罚款、妇幼保健、计生办、医师、抚养费、公共卫生、医疗事故。

劳动和社会保障：

报销、合作医疗、建造师、中级职称、人力资源、参保、社厅、低保、职业技能、工程师、社会保障、编制、人社、病退、比例、人才社保、人员、员工、待遇、拖欠、农民工、就业、单位、失业、津贴、机关、退休金、社保卡、职工、退休、养老保险、养老金、工伤、加班、缴纳、医疗保险、社局、劳动合同、社保局、异地、查询、工资、医保、企业、公务员、新农、退休职工、劳动、享受、保险、缴费、改制、下岗职工、劳动法、工人。

交通运输：

出租车、交通、路面、运输、打表、超载、包裹、货车、路基、快递、客运、司机、物流、取件、坑坑洼洼、高速公路、燃油、车辆、旅客、的士、邮政、滴滴、邮政储蓄、邮政局、EMS、拒载、破烂不堪、桥梁、网点、通车、通行。

环境保护：

污染、辐射、喷漆、孵化场、灌区、奖补、变电站、处理厂、高压线、恶臭、烟囱、水沟、新能源、噪音、浓烟、赤石、排放、纸厂、变电站、烟囱、环境、猪场、排污、环保、砖厂、工厂、破坏、毒气、废气、环评、采石场、基站、水源、臭气熏天、工业园、气味、刺鼻、化工厂、加工厂、开采、空气、气体、危害、直排、关停、偷排、信号、严重破坏、冶炼厂、发射塔、重金属、废水、焚烧、难闻。

教育文体：

教师、文物、上学、校车、足球、教育局、校园、篮球、普高、舞弊、作弊、游戏、上网、运动、学院、网吧、全运会、一中、体育、二中、文化、幼儿园、中学、学校、小学、学生、补课、教育、老师、招生、考试、学籍、培训、高中、代课、读书、民办、学费、入学、高考、家长、初中、班主任、中考、中小学、暑假、转学、办学、录取。

在已知词频的情况下，我们通过查准率和查全率两个条件来选取关键词，其中：

查准率：检出的相关文献与检出的全部文献的百分比。普遍表示为：

查准率=（检索出的相关信息量/检索出的信息总量）x100%。

查全率：是衡量某一检索系统从文献集合中检出相关文献成功度的一项指标，即检出的相关文献与全部相关文献的百分比。

初次选取：依次选取该一级标签中词频最高的关键词，并在 excel 中，将此关键词按照留言主题进行筛选，查看其查准率，若查准率>60%，则认为可以选取该词作为关键词，反之不能。当分词结果全部筛选完成后，查看所有关键词对该一级标签的查全率，若查全率达到 80%以上，则认为关键词选取较好，反之进行

二次选取。

二次选取：当初次选取，查全率小于 80%时，将未被选取的数据进行二次筛选。

重新分词、提取关键词、查看查准率以及查全率。

重复以上步骤直至查准率和查全率达到 70%以上，这是选取的关键词即为所求。

2.1.4 F-Score 算法

建立关于留言内容的一级标签分类模型后，我们使用 F-Score 对分类方法进行评价：

$$F = \frac{1}{n} \sum_{i=1}^n \frac{2P_i R_i}{P_i + R_i}$$

其中 P_i 为第 i 类的查准率， R_i 为第 i 类的查全率。通过计算得知：

第一类城乡建设查准率 $P_i=0.8102$ ，查全率 $R_i=0.6864$ ；

第二类环境保护查准率 $P_i=0.8557$ ，查全率 $R_i=0.822$ ；

第三类交通运输查准率 $P_i=0.8174$ ，查全率 $R_i=0.9396$ ；

第四类劳动和社会保障查准率 $P_i=0.8011$ ，查全率 $R_i=0.804$ ；

第五类卫生计生查准率 $P_i=0.7967$ ，查全率 $R_i=0.7777$ ；

第六类商贸旅游查准率 $P_i=0.7937$ ，查全率 $R_i=0.7029$ ；

第七类教育文体查准率 $P_i=0.8214$ ，查全率 $R_i=0.8798$ ；

$$F = \frac{1}{7} \left(\frac{2*0.8102*0.6864}{0.8102+0.6864} + \frac{2*0.8557*0.8220}{0.8557+0.8220} + \frac{2*0.8174*0.9396}{0.8174+0.9396} + \frac{2*0.8011*0.804}{0.8011+0.804} + \frac{2*0.7967*0.7777}{0.7967+0.7777} + \frac{2*0.7937*0.7029}{0.7937+0.7029} + \frac{2*0.8214*0.8798}{0.8214+0.8798} \right) = 0.8084$$

模型得分为 0.8084，由此可见，模型较好

2.2 问题 2 分析方法与过程

对于热度的分析，我们从留言条数以及点赞数和反对数两方面进行考虑。我们认为留言条数相对来说更能表现出群众对该问题的重视程度，所以认为留言条数占热度指数的 60%，点赞数和反对数占热度指数的 40%。

2.2.1 数据预处理

由于点赞数和反对数都代表该条留言收到关注，所以我们将点赞数和反对数两项指标相加命名为关注度，并将关注度按照降序的顺序排序。

2.2.2 数据筛选

选取关注度较高的十条留言，依次计算相关留言条数及关注度，以此计算相关留言热度指数。

从选取结果可知，排名第五的留言热度指数为 97.4，留言条数为 1，关注度为 242。由于本题热度指数中留言条数占比较大，而我们按照关注度选取，所以我们对选取结果进行检验。

从 2.2.1 对数据的处理可知，该二度指数排名第五的为排序后第 7 条，且前 50 条内除“月亮岛路”问题外并无重复相关问题。然 50 条后关注度最多为 21。所以我们可以假设关注度为 150，可以计算出这时需要留言条数达到 63 条，热度指数才可以大于 97.4，但通过筛选得知，并未出现 50 以上相关连的留言。

通过计算“月亮岛路”问题热度指数，留言条数为 13，关注度为 179，热度指数为 79.4，小于 97.4。

所以判断排名第五的留言即为热度第五的留言。

2.3 问题 3 分析方法与过程

根据附件 4 给出的留言内容与相应的答复情况，我们采取按比例抽样的方式对答复情况进行审核，通过对答复的时间、字数、相关性、完整性和可解释性等方面建立对答复的综合评价系统。

2.3.1 抽样过程

因为考虑到每年的留言数目不同，因此我们采取按比例抽样的方法，同时发

现 2011 年与 2020 年全年所给数据不足 10 条，因此在此次分析中暂不予考虑，针对 2012-2019 对各年的留言总数和占比情况统计如下：

表 1

2012	52	0.0185251
2013	109	0.0388315
2014	103	0.036694
2015	130	0.0463128
2016	227	0.0808693
2017	453	0.1613823
2018	847	0.3017456
2019	886	0.3156395
合计	2807	1

经过以上统计不难发现，2016-2019 数量较多，因此我们决定对 2012-2015 年的留言按年抽取，而对 2016-2019 年留言按月抽取，总体按照 1：20 的比例抽样，最终各样本抽取情况统计如下：

表 2

2012	3
2013	5
2014	5
2015	6
2016	11
2017	23
2018	42
2019	44
合计	140

表 3

	2016	2017	2018	2019
1	1	0	2	3
2	1	1	2	3
3	2	2	3	4
4	1	2	3	4
5	1	4	4	4
6	0	2	3	4
7	0	1	3	4
8	1	2	4	3
9	1	2	4	4
10	1	2	3	3
11	1	2	6	4

12	1	2	6	4
----	---	---	---	---

评价指标：对留言回复的时间进行计算：答复时间≤5 天：优；5 天<答复时间≤10 天：良；10 天<答复时间≤30 天：中；答复时间>30 天：差。对留言字数进行统计：留言字数≤100：差 100<留言字数≤300：中 300<留言字数≤500：良留言字数>500：优。通过对留言的回复时间以及留言字数进行统计：答复时间小于五天，留言字数大于 500，答复的相关性、完整性、可解释性至少两个达标，则说明答复的很好；答复时间大于 30 天，留言字数小于 100，答复的相关性、完整性、可解释性一个或者没有达标的，则认为答复敷衍；答复时间在 30 天以内，字数适中，答复的相关性、完整性、可解释性一个及以上达标的，则说明答复的一般。

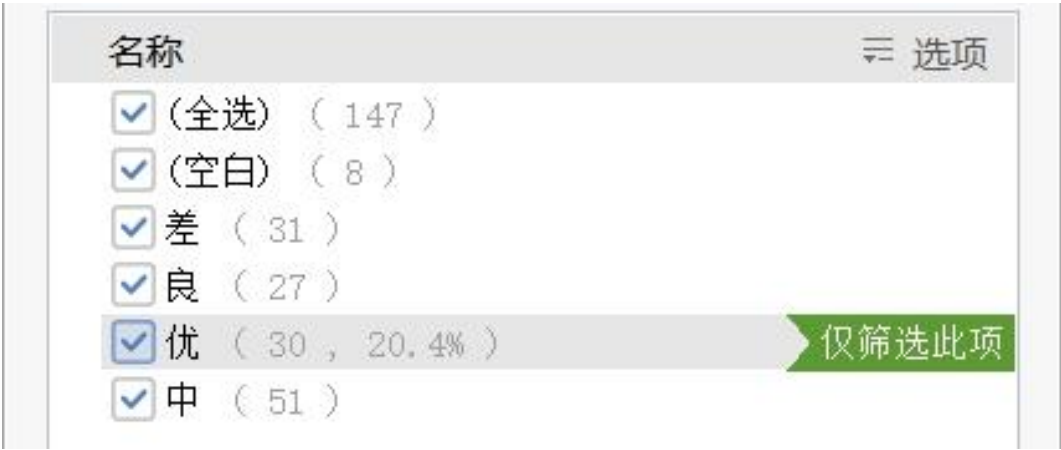


图 10 回复时间统计图

通过 excel 筛选，我们可知答复时间为优的占比 20.4%，说明政府相关人员的答复速度较慢，希望政府人员提高回复速度，或者加派人员进行相关问题回复。

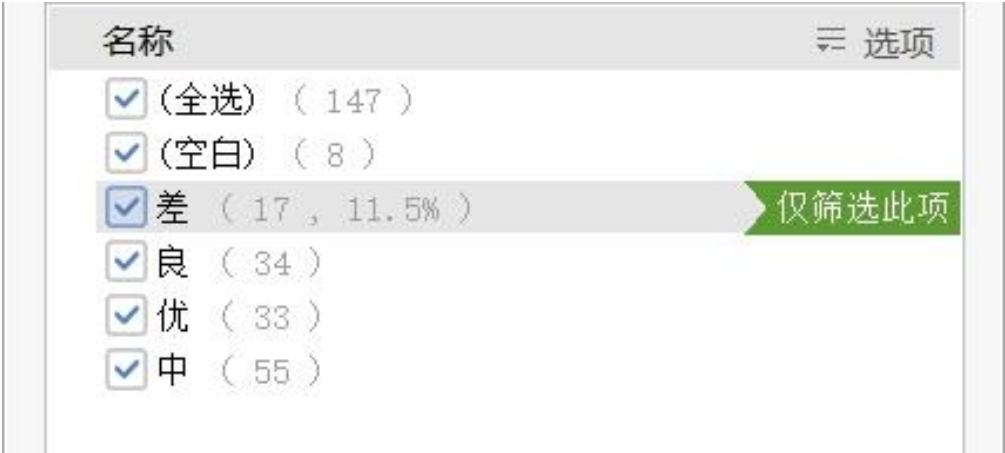


图 11 回复字数统计图

通过 excel 筛选，我们可知回复字数差的为 11.5%，说明政府人员回复时还是比较认真，对问题的回答也很尽力详细。

三、结论与建议

3.1 结论

通过建立多元化政务服务供给渠道，创新国家治理方式，加快新一代信息技术的应用推广，建设智慧政府，建立完善法律保障体系，改善促进国家治理现代化的网络环境，推动信息化社会发展，解决老百姓办事难的问题，使领导在线简单收集问题，第一时间看到市民反映情况，更快更高效的处理解决问题，缩短市民和领导直接沟通的时间，让市民跑更少的路来留言反映，避免更繁琐的人工手续。

3.2 建议

尽管通过电子政务和智慧城市建设，智慧政务从无到有取得了显著成绩。但我们也不难发现，智慧政务的建设是一个复杂的系统工程，它既要伴随着社会发展破解不均衡不充分的问题，也要顺应互联网+时代云计算、大数据和人工智能等技术迭代的完善。智慧政务建设是一个不断完善、不断提升、持续用力、久久为功的过程。通过本次实验，为网络问政平台提出一些改进建议：

1、希望政府提前将留言簿进行类别划分，使市民在相应的地方留言，而政府只需要在解答时检查市民的分类是否正确即可。

2、希望有关部门在进行解答时能够提高回复的解释性，尽量不要出现因为某些问题不属于自己的解答领域就推荐市民去某某部门咨询的情况，针对这个问题可以负责查看留言的人员将问题规划好后分配给各相应部门解答并将解答结果直接反馈给市民，或者通过留言系统自动分类给相应部门进行回答，而不要让市民自己再去找其他部门，做到有问必解。

四、参考文献

[1] 崔晓辉. 网络招聘信息的数据挖掘与综合分析. 北京林业大学. 2017

- [2] 王千, 王成, 冯振元, 叶金凤. K-means 聚类算法研究综述. 2012
- [3] 卜凡军. KNN 算法的改进及其在文本分类中的应用. 江南大学. 硕士学位论文. 2009
- [4] 曹卫峰. 中文分词关键技术研究. 南京理工大学. 硕士学位论文. 2009
- [5] 杨虎. 面向海量短文文本去重技术的研究与实现. 国防科技大学. 2007
- [6] 赵琳瑛. 基于隐马尔科夫模型的中文命名实体识别研究. 西安电子科技大学. 2007
- [7] 翟东海, 鱼江, 高飞, 于磊等. 最大距离法选取初始簇中心的 K-means 文本聚类算法的研究. 西安交通大学. 2014
- [8] 张晓辉, 李莹, 王华勇等. 应用特征聚合进行中文文本分类的改进 KNN 算法. 东北大学. 2003