

“智慧政务”中的文本挖掘应用

摘要：

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。在处理网络问政平台的群众留言时，工作人员首先按照一定的划分体系对留言进行分类，以便后续将群众留言分派至相应的职能部门处理。目前，大部分电子政务系统还是依靠人工根据经验处理，存在工作量大、效率低，且差错率高等问题。

在本次数据挖掘过程中，我们首先用Rstudio进行数据预处理，分词以及停用词过滤操作，实现了对评价数据的优化，并提升了其可建模度，建立关于留言内容的一级标签分类模型以便更加快捷地对留言内容进行分类

并且用F-Score 对分类方法进行评价：

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

其中 P_i 为第 i 类的查准率， R_i 为第 i 类的查全率。

接着对某一时段内反映特定地点或特定人群问题的留言用python进行归类，定义合理的热度评价指标，并给出评价结果（某一时段内群众集中反映的某一问题可称为热点问题），及时发现热点问题，有助于相关部门进行针对性地处理，提升服务效率。

最后，相关部门对留言的答复意见，从答复的相关性、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，并尝试实现。

关键词：自然语言分类模型；文本挖掘；信息提取；模型预测；方案评价；

Text mining application in "smart government"

Abstract: In recent years, with wechat, microblog, mayor's mailbox, sunshine hotline and other online political platforms gradually becoming an important channel for the government to understand the public opinion, gather people's wisdom and gather people's spirit, the amount of text data related to various social situations and public opinions has been increasing, which has brought great challenges to the work of relevant departments that used to rely mainly on human to divide messages and sort out hot spots. At the same time, with the development of big data, cloud computing, artificial intelligence and other technologies, the establishment of smart government system based on natural language processing technology has become a new trend of social governance innovation and development, which has a great role in promoting the management level and efficiency of the government. When dealing with the public message on the network politics platform, the staff first classifies the message according to a certain division system, so as to assign the public message to the corresponding functional departments for processing. At present, most of the e-government systems still rely on manual processing according to experience, there are many problems, such as heavy workload, low efficiency, and high error rate.

In the process of data mining, we first used Rstudio to preprocess, segment and filter the stop words to optimize the evaluation data, improve its modeling degree, and establish a first-class label classification model for the message content so as to classify the message content more quickly

The classification method was evaluated by F-score

Then reflect the specific location or specific for a certain period of time
The messages of crowd problems are classified, reasonable heat evaluation indexes are defined, and evaluation results are given (a problem reflected by the crowd in a certain period of time can be called a hot issue). Timely finding hot issues is helpful for relevant departments to deal with them pertinently and improve service efficiency.

Finally, the relevant departments give a set of evaluation scheme for the quality of the reply from the perspective of relevance, integrity and interpretability of the reply, and try to achieve it.

Keywords: natural language classification model; text mining; information extraction; Model prediction; scheme evaluation;

目录

1. 挖掘目标	5
2. 分析方法与过程	5
2.1 一级留言分类模型建立	5
2.1.1 数据载入及缺失值处理	5
2.1.2 数据清洗整理	9
2.1.3 分词统计	10
2.2 热点问题排序	11
2.2.1 热点问题表	11
2.2.2 热点问题明细表	18
2.3 评价方案	36
3. 结论	36
4. 参考文献	36

1. 挖掘目标

挖掘数据收集自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。利用自然语言处理和文本挖掘的方法解决建立一级留言分类模型，热点问题排序以及形成一套评价方案，并尝试实现。

在当前社会环境下，网络问政平台已成为政府与公民之间直接沟通的重要渠道。然而，面对日益增长的留言数量和多样化的问题类型，政府部门在处理留言时往往面临工作效率低下、信息处理不及时等问题。因此，本研究旨在利用先进的自然语言处理和文本挖掘技术，从网络问政平台的群众留言中挖掘出有价值的信，以解决以下问题：

提高政府部门处理群众留言的效率和准确性：通过建立一级留言分类模型能够自动将群众留言按照一定的标准分类，从而实现快速、准确地将留言分派至相应的职能部门处理，减少人工干预的时间和成本。

发现热点问题和舆情趋势：通过对留言内容的文本挖掘和分析，可以识别出当前社会关注度较高的热点问题，及时采取措施应对，提升政府公共管理水平。

建立更加智能化的留言处理系统：基于自然语言处理技术，建立留言情感倾向性模型，可以更好地理解留言背后的情感倾向，有助于政府部门更加智能化地回应民意，提高公众满意度和信任度。

促进政府与公民之间的良好互动与沟通：通过对群众留言的及时响应和有效处理，能够增强政府的透明度和公信力，建立起政府与公民之间良好的互动机制促进社会稳定和发展。

因此，本研究不仅仅局限于简单的留言分类，更致力于构建一个智能化、高效率的留言处理系统，为政府提供更加科学、准确的决策支持，实现政府治理的现代化和智能化。

2. 分析方法与过程

2.1 一级留言分类模型建立

2.1.1 数据载入及缺失值处理

处理工具：Rstudio(R-3.6.3版本)

取到文本后，首先要进行文本评论数据的预处理。文本评论数据里面存在大量价值含量很低甚至没有价值含量的条目，如果将这些评论数据也引入进行分词、词频统计乃至情感分析等，则必然会对分析造成很大的影响，得到的结果的质量也必然是存在问题的。那么在利用到这些文本评论数据之前就必须要先进行文本预处理，把大量的这些无价值含量的评论去除。

在进行文本数据的预处理过程中，除了去除无价值含量的评论外，还需要考虑数据的质量和完整性，以确保后续分析的准确性和可靠性。以下是对文本数据预处理过程：

数据载入：

在数据载入阶段，首先需要从数据源中获取文本数据。这可能涉及从数据库、文本文件、API接口或网络爬虫等方式获取数据。在获取数据的过程中，要注意数据的格式、编码以及可能存在的特殊字符，确保数据能够被正确解析和处理。

文本去重：

文本去重是指在数据集中去除重复的留言，以避免重复信息对分析结果造成的干扰。常见的文本去重方法包括基于文本内容的去重和基于文本特征的去重。基于文本内容的去重可以使用哈希函数或者相似度匹配算法（如编辑距离、余弦相似度等）

来识别相似的文本，并将其合并或删除。基于文本特征的去重则是通过提取文本的特征向量，如TF-IDF向量或词向量，然后利用聚类算法（如K-means、DBSCAN等）或近似最近邻算法（如LSH）来识别重复文本。

机械语料压缩：

机械语料压缩是指删除一些常见的机械性文本，如无意义的符号、表情符号、链接、HTML标签等。这些内容通常对于情感分析或主题分类等任务并不具有实质性的帮助，因此需要在预处理阶段将其去除。常见的方法包括使用正则表达式进行模式匹配，识别并删除这些机械性文本。

短句删除：

短句删除是指删除过短的文本，因为这些文本通常包含的信息量较少，对于后续的分析任务没有实质性的帮助。短句删除的阈值可以根据具体任务和数据集的特点来确定，通常可以通过设置最小词数或字符数来进行过滤。

短句删除：

短句删除是指删除过短的文本，因为这些文本通常包含的信息量较少，对于后续的分析任务没有实质性的帮助。短句删除的阈值可以根据具体任务和数据集的特点来确定，通常可以通过设置最小词数或字符数来进行过滤。

以上步骤是数据预处理的基本流程，通过这些步骤可以清洗和准备好文本数据，为后续的分词、词频统计和情感分析等任务打下良好的基础。

留言编号	留言用户	留言主题	留言时间
1	24 A00074011	A市西湖建筑集团占道施工有安全隐患	2020/1/6 12:09:38
2	37 U0008473	A市在水一方大厦人为烂尾多年，安全隐患严重	2020/1/4 11:17:46
3	83 A00063999	投诉A市A1区苑物业违规收停车费	2019/12/30 17:06:14
4	303 U0007137	A1区禁路南路A2区华庭楼顶水箱长年不洗	2019/12/6 14:40:14
5	319 U0007137	A1区A2区华庭自来水好大一股霉味	2019/12/5 11:17:22
6	379 A00016773	投诉A市盛世耀凯小区物业无故停水	2019/11/28 9:08:38

```
> dim(Data)
[1] 9210    6
> class(Data)
[1] "data.frame"
> names(Data)
[1] "留言编号" "留言用户" "留言主题" "留言时间" "留言详情" "一级标签"
> sum(is.na(Data))
[1] 0
```

```
factor(0)
8905 Levels: ?F市南片区的危房拆迁就这么难吗? ?K4县白牙市镇洗衣厂污水直排 ... 做了《亲子鉴定》为何医院还不给办理出生证?
```

```
factor(0)
9052 Levels: \n \n -强烈要求M3县委、政府归还原工商局被违规侵占的行政编制\n\n 经机构改革，原M3县工商局、质监局、食药监局合并为M3县市场监督管理局，共有干部职工320余人，其中大部分为国家公务员。原M3县工商局在垂直管理期间，拥有行政编制的国家公务员196人，实行一人一编。但是，三局合一后，人员增加了，行政编制不但没有增加，反而被大幅缩减，县编办只核给了62个行政编制。新成立后的市场监督管理局按照国务院的“三定”要求，承担着行政法和市场监管的重要职责。原本机构改革的初衷就是要通过整合执法资源，加大执法力度，更好地维护市场经济秩序。如果光有公务员身份，没有行政执法编制，不具有合法的执法主体资格，又如何去履职执法、维护市场经济秩序呢？\n\n 为此，我们试问：\n\n 1、原M3县工商局的196个行政编制，是省里按照“三定”方案定的，并没有占据地方的行政编制，下放到地方管理后就被强行缩减了一百多个行政编制，那这么多行政编制去哪儿了？是被违规侵占还是做了其它用途？\n\n 2、对调整省级以下工商、质监体制的问题，早在去年6月8日，西地省委办公厅就下发[政府发文]25号红头文件，对编制、人员划转和工资福利待遇等进行了具体明确。省委办公厅专门出台文件，就是对下放在即的工商、质监这些关心和敏感的问题做出政策性的规定，相信同时也有对工商、质监由上收到下放的安抚之意，从而确保人员稳定，推进机构改革。但是，当我们还来不及放开双手热情投入地方怀抱的时候，地方政府却无情地给了我们当头两棒，一是强行缩减了一百多个行政编制，二是强行将住房公积金缴纳比例从原来的12%降到8%。这两棒不轻啊！让人心惊胆颤，不寒而栗。据了解，西地省各市、州及M市的其他几个县市编办，在接收工商、质监编制时，都是按照下放时的不变。28日上午，县编办的一位领导还打了M5市编办的电话，问了情况，M5市也是按照下放时的编制性质没变。为什么M3县委、政府、编办却要另搞一套，并且执意违[政府发文]25号文件精神？难道其中另有隐情？是政府在获取权力寻租吗？\n\n 对此，我们向M3县委、政府和编办反映了情况，并向M3县委全体常委递交了报告，我们提出如下诉求：\n\n 1、强烈要求M3县委、政府、编办严格执行并落[政府发文]25号文件精神，保持原M3县工商局在下放前的196个行政编制不变，归还被违规侵占的所有行政编制，确保原工商局广大公务人员的合法权益。\n\n 2、如果这一问题得不到有效解决，我们将向上级人民政府、纪委及新闻媒体如实反映情况，直到问题解决。\n\n 原M市M3县工商局全体公务人员\n\n 二〇一六年六月二十九日\n\n \n \n ...
```

情感倾向性模型

训练生成词向量

构建词表：

将经过预处理的文本数据构建成词表（**vocabulary**）。词表是由数据集中所有不同词语组成的集合，每个词语都被**赋予**一个**唯一**的**索引**编号。

One-hot Representation：

One-hot Representation是将每个词语表示为一个很长的二进制向量，向量的长度等于词表的大小，其中只有一个**元素**为1，其余**元素**为0。具体而言，对于词表中的每个词语，都会有一个对应的位置被设置为1，其余位置为0。这样就完成了将词语数字化表示的过程。

嵌入层（Embedding Layer）：

在深度学习模型中，通常会使用**嵌入层**来学习词向量。**嵌入层**将每个词语的**索引**编号作为输入，然后通过训练学习到每个词语的**稠密**向量表示（即词向量）。这些词向量是在模型的训练过程中学习到的，能够**捕捉**词语之间的语义信息。

训练模型：

一旦词向量被学习到，就可以将其应用于情感分析任务中。通常情况下，会使用深度学习模型，如**循环神经网络（RNN）**、**长短期记忆网络（LSTM）**、**卷积神经网络（CNN）**等来训练情感分类模型。在训练过程中，将词向量作为模型的输入，然后通过模型学习文本数据中的情感信息，从而实现情感分类的目标。

模型评估与优化：

训练完成后，需要对模型进行评估和优化。可以使用**交叉验证**等方法对模型进行评估，进而调整模型的超参数或优化算法，以提高模型的性能和**泛化**能力。

通过以上步骤，可以建立一个基于情感倾向性模型训练生成词向量的情感分析系统，用于分析文本数据中的情感倾向，并进行情感分类任务。这样的系统在文本情感分析、**舆情**监测、**用户**评论分析等方面都具有**广泛**的应用价值。

留言集子集的人工标注与映射

利用词向量构建的结果，再进行留言集子集的人工标注，城乡建设留言标为1，环境保护留言标记为2，交通运输留言标记为3，教育文体留言标记为4，劳动和社会保障留言标记为5，商贸旅游留言标记为6，卫生计生标记为7，然后将每条评论映射为一个向量，将分词后评论中的所有词语对应的词向量相加做平均，使得一条评论对应一个向量。

人工标注：

首先，需要人工对留言集子集进行标注，将每条留言按照其所**涉及**的主题或类别进行标记。根据给定的标签体系，例如城乡建设、环境保护、交通运输等，对每条留言进行分类标记，分别标记为1到7或其他相应的标签。这一过程需要具备**相关领域**知识的专业人员进行标注。

词向量构建：

利用已经训练好的词向量模型，将每个词语表示为一个**固定**长度的向量。这些词向量应该能够**捕捉**词语之间的语义相似性，以便后续对留言进行语义分析。

向量化表示：

对于每条留言，首先进行分词处理，将留言中的每个词语**转换**为其对应的词向量。然后将每个词向量相加，并将结果除以留言长度，以求得平均向量。这样，每条留言就能够表示为一个**固定**长度的向量。

留言分类模型训练:

将得到的向量化表示作为训练数据，同时将对应的人工标注作为标签，用于训练留言分类模型。可以采用各种机器学习算法，如支持向量机（SVM）、决策树、随机森林等，或者深度学习模型，如卷积神经网络（CNN）、循环神经网络（RNN）等进行模型训练。

模型评估与调优:

使用交叉验证等方法对训练好的模型进行评估，检验其在未见过的数据上的分类性能。根据评估结果进行模型参数调优，以提高模型的泛化能力和分类准确率。

通过以上步骤，可以建立一个基于词向量和人工标注的留言分类模型。该模型能够根据留言内容自动识别其所属的主题或类别，从而实现留言的自动分类和分派，提高政府部门处理留言的效率和准确性。

2.1.3 分词统计

```
# 采用默认的jieba分词器
engine<-worker(user='E://Udacity//R//R3.6.3//library//jiebaRD//dict//user.dict.utf8',encoding = 'UTF-8')
# 去除无关的词
word.lis <- lapply(CN.clean$description, function(x){
  v <- gsub('[\u4e00-\u9fa5|0-9|\\.|\\-|\\_]', '', segment(x,engine))
  v <- v[v!='']
  return(v)
})
# 所有的词语转化成大写,避免出现大小写的错误
segWord <- toupper(unlist(word.lis))
# 加载stop词语列表
stopWords<-toupper(readLines('E://Udacity//R//R3.4.3//library//jiebaRD//dict//stop_words.utf8',encoding = 'UTF-8'))
# 过滤分词
# 此处确保我要得到的前15个关键技能是正确的数据分析技能
removewords <- function(targetword,stopword){
  targetword = targetword[targetword%in%stopword == F]
  return(targetword)
}
segword<- sapply(X=segWord, FUN = removewords,stopWords)
```

```

word_freq <- top.freq(unlist(segword),15)
# 成有id和keyword构建的数据框
id <- NULL
keyword <- NULL
for(i in index(word.lis)){
  id <- c(id,rep(i,length(word.lis[[i]])))
  keyword <- c(keyword,word.lis[[i]])
}
keyword.df <- data.frame("id"=id,"keyword"=toupper(keyword))
keyword.df <- droplevels(keyword.df[keyword.df$keyword %in% word_freq$x,])
merge.df <- merge(CN.clean,keyword.df,by = 'id')
summary(merge.df)

```

2.2 热点问题排序

2.2.1 热点问题表

处理工具：python以及Rstudio及Excel筛选工具

数据预处理：

首先，我们需要加载附件3中的留言数据集，并对其进行预处理。这包括去除重复留言、清洗文本数据（如去除特殊符号、停用词等）、分词等操作。

热点问题分类：

我们将根据留言的分类信息，特别是第一分类（classification）中的数据种类，查看数据集中classification（第一分类）里面的数据种类，并查看每一种类意见的比例分布，来确定热点问题的大方向。通过查看每种分类下留言意见的比例分布，我们可以确定出现频率较高的几种类别，如劳动和社会保障、城乡建设、教育文体等。这些类别往往代表了当前社会关注的焦点问题，因此可以作为热点问题的参考指标。

借助代码如下：

```

import pandas as pd
f=open('D:\RandRstudio\示例数据泰迪\data.csv','r')
df=pd.read_csv(f)
df.head()
print(df['classification'].unique())
print(df.groupby('classification').size())

```

实现如下结果：

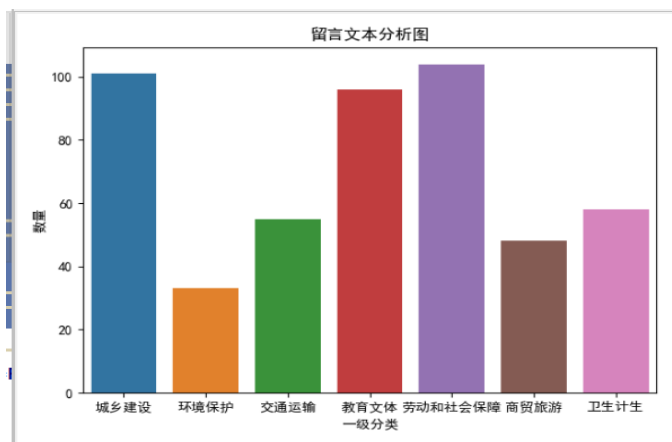

```
In[2]: import pandas as pd
...: f = open('D:\RandRstudio\示例数据泰迪\data.csv', 'r',)
...: df=pd.read_csv(f)
...: df.head()
...: print(df['classification'].unique())
...: print(df.groupby('classification').size())
...:
['城乡建设' '环境保护' '交通运输' '教育文体' '劳动和社会保障' '商贸旅游' '卫生计生']
classification
交通运输      55
劳动和社会保障    104
卫生计生      58
商贸旅游      48
城乡建设    101
教育文体      96
环境保护      33
dtype: int64
```

可知，在各种意见中，劳动和社会保障、城乡建设、教育文体意见所占比例最高，故热点问题极大方向是出自这三种留言意见中。

可用更加直观的条形图做演示，借助如下代码：

```
import seaborn as sns
import matplotlib.pyplot as plt
sns.countplot(df['classification'],label="Count")
plt.rcParams['font.sans-serif']=['SimHei']
plt.rcParams['axes.unicode_minus'] = False
plt.title("留言文本分析图")
plt.xlabel("一级分类")
plt.ylabel("数量")
plt.show()
```

得到以下直观的条形图演示



所得结论同上，故热点问题大方向确定为城乡建设，教育文体，劳动和社会保障。

接下来对数据集进行简单地词频统计并生成简单的词云
借助代码如下：

```
library(jiebaR)
library(wordcloud)
file <- readLines("E:/示例数据/附件2.txt", encoding = "gbk")
#使用qseg类型分词，并把结果保存到对象seg中
seg <- qseg[file]
#统计词频
seg <- table(seg)
#去除数字
seg <- seg[!grepl('[0-9]+', names(seg))]
#查看处理完后剩余的词数
length(seg)
#降序排序，并提取出现次数最多的前100个词语
seg <- sort(seg, decreasing = TRUE)[1:100]
#显示如下图1
seg
#设置图片形成效果
bmp("comment_cloud.bmp", width = 500, height = 500)
par(bg = "black")
#云图显示如下图2
wordcloud(names(seg), seg, colors = rainbow(100), random.order=F)
#关闭词云
```

```
dev.off()
```

得到结果如下：

seg	的	了	是	在	我	和	我们	市	有	不	年	县	也	都	没有	
4911	2136	1077	1052	1036	869	777	776	743	558	522	496	481	461	423	400	
344	344	323	289	286	282	282	280	267	257	247	247	241	229	229	220	
规定	工作	后	区	等	政府	相关	现在	但	这	并	学校	国家	被	教育	医院	
215	214	210	208	204	201	198	197	195	192	188	188	185	184	181	181	
说	上	又	劳动	将	进行	中	可以	去	日	能	职工	要求	让	而	小区	
179	176	176	174	171	171	168	164	164	163	162	162	161	158	157	155	
该	规划	严重	元	希望	不能	人员	单位	政策	来	情况	请	已	企业	他们	已经	
152	151	151	151	150	148	148	146	146	144	141	141	133	132	132	131	
办理	解决	有关	居民	下	业主	一	向	管理	工资	西地省	作为	却	时间	一直	想	
127	127	127	126	126	126	124	123	122	121	114	113	112	112	112	111	
本人	地	个	什么													
110	109	109	109													

可知热点词频最高的大部分意见是关于政府，学校，教育，医院，劳动，规划，小区，办理，居民，业主，工资，企业等热点问题。

生成热点问题表：

根据热度评价指标，我们可以按照降序排列将热点问题列入热点问题表中。通常情况下，我们会选择排名前几位的热点问题进行详细分析和处理。这些热点问题将被保存为一个Excel文件，命名为“热点问题表.xls”。

再借助Excel筛选排名前五的热点话题，可得如下热点问题表格

①关于“政府”热点意见表格：

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
4	3742	A00013884	A3区杜鹊文苑小区外的非法汽车检测站要开业了！	018/12/26 10:13:35	（省级）：水务局：指	城乡建设
5	5315	A0009647	民工在A6区明发国际工地受伤，工地方拒绝支付医疗费	2018/7/17 7:13:58	地现在拒绝支付医疗费	城乡建设
11	100825	A0006699	K市域轨道交通规划建议	2017/3/15 20:50:44	不利于促进城镇化的发	城乡建设
12	104720	A0006699	关于K市域轨道交通规划的建议	2017/3/15 20:53:39	不利于促进城镇化的发	城乡建设
19	4082	A00080958	A市可以实现集中供暖吗？	2018/11/20 9:56:55	家户户的大问题，由于	城乡建设
28	7257	A00012800	A3区阳光丽城附近垃圾场上趁夜色焚烧垃圾	017/11/28 16:01:15	至12345，转城管，反映	城乡建设
29	31807	A00014358	请给A5区木莲路冯家冲安置小区开通天然气和消防水源	019/10/12 19:44:00	几年来这里的问题已	城乡建设
30	107378	A00060328	K1区珠山街的下水设施需要完善	2019/4/18 9:14:48	街上的水有一米高，为什	城乡建设
33	379	A00016773	投诉A市盛世锦凯小区物业无故停水	2019/11/28 9:08:34	物业不是为业主服务的	城乡建设
35	22776	A000102162	楚江世纪城段沿江风光带园林区野草泛滥成灾	2014/8/25 11:41:11	（简称）原来是市水利	城乡建设
40	18801	A00075703	A7县松雅西地省片儿需要配置公立幼儿园	2015/7/15 9:56:06	带来担忧和恐慌。	城乡建设
41	18811	A00073803	A7县凉塘路东六线至东七线之间的三厂房会搬迁吗？	2015/7/9 0:07:09	的三厂房会搬迁，改	城乡建设
42	18819	A00014971	A7县世锦小学垃圾站规划在小区门口离最近住户不到5米	2015/7/7 9:57:31	球场边，哪里不是设置	城乡建设
43	70294	A00068088	E4县宋家塘和顺家园唐租房的规划设计合理吗	2017/6/14 21:00:24	现在蓉邵铁路运输专线	城乡建设
44	144304	A00012528	请责令梅苑开发区新世界建材城附近的违规建筑停工，重新规划设计	2013/7/14 10:47:44	四周与已有建筑只相隔	城乡建设
45	149985	A00083039	限高十二米竟然建了七层二十余米M7市新城的规划执法形同虚设	2014/4/3 0:52:25	投诉，相邻间距最宽处	城乡建设
47	131654	A00078525	L9县一幼建设教学楼不要与L9县十三五规划修维馆公馆相背离	2016/2/11 16:53:00	然，笔者在春节前，	城乡建设
48	150509	A00017418	关于对M7市首府印象增加容积率和提高建筑密度的质疑	2016/8/5 9:57:00	广大业主的一个最起码的	城乡建设
49	157560	A0009784	强烈要求M4市湾田学校不要盲目改址	014/12/23 13:43:26	解四中的压力，况且两	城乡建设
50	174873	A00063523	投诉L6市新时空小区不发放房产证	2019/5/14 11:56:30	业主的生命财产安全！	城乡建设
54	38275	A00057801	为加快B5县朱行特色小镇建设提出几点建议	2013/2/21 11:02:58	应，打名人“品牌”，	城乡建设
56	46322	A00070313	请求政府加强整顿C4市金数乡镇问题	2017/5/9 8:14:25	评估，我们金数镇居民	城乡建设
57	12050	A00097179	为何A6区白岩铺镇商户自己买地建的商铺连招牌都不能换？	2014/8/13 10:43:06	牌认可，获得加盟资格	城乡建设
58	12826	A00073498	A3区雨敞坪镇喜发塘为何至今不通公路	2013/7/6 10:53:52	的泥巴公路（现泥巴公	城乡建设
59	12330	A0007948	A市联美地产建豪宅渗水搭盘	2014/4/29 12:30:46	在开展群众路线教育	城乡建设

414	153828	A000100428	D6县市场监督管理局收费是合理的吗	019/10/10 23:47:20	000元，以此类推。营业	商贸旅游
416	38430	A00016710	请严查B3市三家饭店冒用找店招牌和商标的侵权行为	017/10/20 15:07:10	两字，他们也承认加	商贸旅游
419	73121	A00087180	E7县商品混凝土质量是否可靠？	013/11/27 14:04:04	百姓怎能不担忧？	商贸旅游
421	13793	A00031618	请A市重视物价过快上涨的问题	2019/4/8 12:51:48	全国垫底，百姓生活及	商贸旅游
425	84973	A00076795	G4县东门符家巷华东食品店直销店噪音污染	2019/1/16 8:50:02	睡，苦不堪言，四周、	商贸旅游
426	170840	A00049260	G8县吉美磁化水直销店伙伙虚宣传	2017/4/10 9:50:43	皮肤病。。。总之	商贸旅游
427	175437	A00031896	I市通联违反劳动法扣网络直销员基本工资	2014/5/16 15:19:26	听免费，全国无长途	商贸旅游
433	51251	A0007819	D10县莲花西路合力超市的猪肉是否经过检验检疫？	2019/3/7 14:50:43	看不出是否经过检验检	商贸旅游
439	165639	A000113117	C4市板托村村民没有依法享受公共卫生服务	2014/6/27 11:15:16	公共卫生服务。开展服	卫生计生
440	166008	A00029725	乡村医生公共卫生政府补贴钱不应被拖欠	2011/5/31 21:19:01	每个卫生室就领了五六	卫生计生
441	166015	A00048447	西地省公共卫生拨款的去哪了？	2011/5/22 20:21:31	贴，他们却跑路油钱都	卫生计生
444	14699	A00091797	A市的独生子女费何时发放？	2014/6/2 17:07:23	在A市A2区，请市长大	卫生计生
447	65824	A00015776	E10县水庙镇政协联工委主任罗生武作为计生干部，非法超生	2013/3/5 17:47:02	查了西女孩后超生了一	卫生计生
448	65851	A00014206	E7县鸭田镇计生办人员强行关押超生困难群众	2013/1/12 18:25:24	个月大的儿子正发高烧	卫生计生
449	6215	A3127561379	A市卫计委的回复文不对题	2018/4/23 14:45:30	将有关情况回复如下：	卫生计生
450	11352	A000107599	A2区计生办工作人员百般刁难恶意中伤群众	2015/6/11 17:48:18	别材料，她告警已经	卫生计生
455	164389	A00031481	请解决失独老人门诊医疗费用	2017/3/7 14:59:30	家庭扶助金340元还要	卫生计生
457	13184	A000104162	找超生要罚31万，找一家七口还吃饭吗？	2012/5/26 0:44:08	字，并要我签下放弃	卫生计生
458	46265	A00064327	C4市龙洞镇楠管村村委会当违法违法超生	2017/7/11 20:36:22	某村民代表的情况下，	卫生计生
460	28918	A000100848	A9市太平桥镇新菜市场保健品直销欺骗老年人	2017/7/19 10:29:55	药公司生产的松花粉能	卫生计生
463	122667	A000110005	L市中医院私自采购药品	2015/9/23 14:17:55	且没用完，我当时就心	卫生计生
467	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	卫生计生
470	84058	A00034825	使用过期药品，投诉G5县卫生局不受理	2014/8/11 13:15:15	室医师用过期药品我	卫生计生
472	13253	A00089558	请公平对待艾滋病患者	2012/4/4 0:34:52	吗？不是到处广告，免	卫生计生
480	73641	A00028837	E10县真山医院医护人员严重失职导致足月顺产新生儿脑损伤	2017/2/5 13:45:01	真小孩做康复，目前	卫生计生

②关于“学校”热点意见表格：

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
26	5869	A00015381	A5区雅礼洋湖实验中学附近垃圾箱长期未清理	2018/5/27 10:01:38	病，影响学生、教职工	城乡建设
41	18811	A00073803	A7县凉塘路东六线至东七线之间的三厂房会搬迁吗？	2015/7/9 0:07:09	的三厂房会搬迁，改	城乡建设
42	18819	A00014971	A7县世锦小学垃圾站规划在小区门口离最近住户不到5米	2015/7/7 9:57:31	球场边，哪里不是设置	城乡建设
49	157560	A0009784	强烈要求M4市湾田学校不要盲目改址	014/12/23 13:43:26	解四中的压力，况且两	城乡建设
116	10689	A00061748	关于取消在A3区杨美公园建设110KV变电站的请愿书	2016/1/8 16:48:00	尊重附近居民的意见	环境保护
118	138263	A00077882	M4市数千居民与学生跪求政府	2017/4/9 14:21:13	下水长期接触会自然影	环境保护
181	166998	A00030849	大学录取通知书，EMS邮政最后一公里这么黑	2017/7/17 9:51:41	款少40元）的形式敷衍，	交通运输
191	9955	A000111921	A市麓山国际实验学校教师工资低，平时每月1500元左右	2016/8/6 20:48:47	个月的税后工资扣完五	教育文体
194	5549	A00023078	A市未来之星教育学校部分培训老师师德需加强	2018/6/25 17:44:11	态度冷漠骂他蠢，并把	教育文体
195	37579	A00055859	B9市王仙镇二棚小学师资力量明显不足	2018/6/25 11:24:11	小学。同时二棚小学6	教育文体
196	50903	A00011653	D市慧南实验中学师资力量差	2019/6/28 11:43:24	一要分科到高二的学生	教育文体
197	56988	A00085407	D6县西渡镇咸水学校师资力量严重不足	2018/3/8 18:01:44	推介。但今天想与看	教育文体
200	162961	A00078750	建议我省出台原民办代课教师录用为非教师的国家正式职工的工龄认定相关政策	2015/2/28 9:51:10	间的工龄，理由是：	教育文体
201	106767	A00046530	K5县三小的教学管理太差了	017/10/25 11:15:57	。学生这么多，而老师	教育文体
202	152328	A0007834	M8县二中请加强教学管理	019/12/30 15:22:31	二中还真有一个英语	教育文体
204	47380	A00082617	C4市教育经费问题的疑惑	2014/1/21 0:54:27	为什么有的大校、困难	教育文体
205	96293	A00060024	J5县油麻乡中心学校强制要求各村小报账欺骗取国家教育经费	2015/11/16 9:07:40	元下拨款，为什么却	教育文体
206	16123	A0009621	2019年开始民办学校招生微机派位，A7县如何同步参加？	2019/1/18 18:49:03	招生。A7县内多所民办	教育文体
207	39443	A00026503	B市民办学校贷款，政府是否有贴息补助	2017/8/10 15:52:22	下B市民办学校贷款，	教育文体
208	51129	A00077591	办小学校撤并D1区政府变通为民办学校	2019/4/15 9:27:34	校，为何区政府批准准	教育文体
212	69522	A00084539	E10县藕塘小学离校区学习环境差	2019/9/11 15:20:50	！！每个孩子都是祖国	教育文体
214	177531	A00072711	H3县江坪镇中学关于请求封闭校园东南部围墙缺口的呐喊	2019/4/23 14:04:04	校园引诱、捉走本校	教育文体
216	26481	A00061612	小学生转学难，A市的学位谁家有人？	2015/4/24 23:19:34	给予回复为：直接回	教育文体
217	33042	A00010066	B市中科柏睿学校的学历毕业证书是国家认可的吗？	2019/8/14 7:28:11	市人工智能技术职业学	教育文体
218	42561	A00047070	C市大学实行一人一网政策，加大学生负担	2018/3/4 9:05:19	50元，加大学生负担，	教育文体

228	63276	A000106447	质疑E市E12市教育局关于“代课教师问题”的回复	2019/5/28 11:31:04	同意设立E12市王龙幼	教育文体
229	89367	A00073021	津市职教集团的教育体制有很多大的问题，必须要彻底改革	2015/3/19 22:07:44	资却很少，受到的批评	教育文体
230	111899	A00035554	K8县教育局应尽早建立督学责任区	2012/5/31 12:28:36	或专业水平的领导，不	教育文体
234	457	A00080495	咨询A市小学生转学问题	019/11/18 22:49:18	该的学习及生活在很	教育文体
235	167856	A00050547	又到一年高考时，I市监考规则不能“潜规则”	2012/5/9 16:56:26	的时间流逝，有些市	教育文体
236	176538	A000102231	G4县龙池学校高考作弊，监考老师未按规定用G2的老师	2015/6/15 9:50:24	矿泉水作接印号，龙	教育文体
237	176737	A00092673	I4县玉潭学校老师参与本校师范生监考，有失公平公正	2019/5/12 12:46:37	，关系到学生的前途	教育文体
239	2168	A00023791	反映长郡A9市实验学校特长生招生名不符实的问题	2019/5/31 0:29:11	，我认为该学校有导	教育文体
240	4710	A00040612	西地省成考招生报名系统一直不能预约！！	2018/9/12 16:56:11	等学校招生统一考试	教育文体
242	57664	A00079065	D10县官家嘴镇中心小学招生人数缩减不公平	2019/6/15 6:20:35	嘴镇中心小学招收一	教育文体
244	62297	A00041376	D11市蔡子池中学将上网上报名在家	2013/1/11 14:22:05	老师，我们也进行过	教育文体
246	76444	A00063804	F5县新河乡小孩被学校劝退辍学并注销了学籍！	2019/8/9 22:24:28	教育！现在14岁想去继	教育文体
247	2153	A00022064	请问非电力局的人能否在A市电力局的体育馆内打球？	2019/6/1 7:40:37	在那里，不刷脸，里面	教育文体
249	4722	A00041093	对A2区小学体育教育的一点看法	2018/9/12 10:29:51	起上一次体育课，可	教育文体
250	133667	A00072626	L6县三中运动会怎么回事？	2017/9/26 11:40:11	校举报，跑道又是返	教育文体
272	150475	A00057513	M7市旅游和文化广电新闻出版局多次违规审批网吧	2018/5/26 0:10:54	门由市教育局批准的	教育文体
321	85682	A000110357	关于C市教育局本次工资调整的几点看法	2015/12/31 9:05:21	的职称作为衡量工资	劳动和社会保障
325	62280	A00035367	D11市人社局网站应公布就业培训学校的情况	2014/6/14 22:04:11	应该在网上公开	劳动和社会保障
336	159144	A00070522	请问西地省哪里有意招人就业培训班	2017/5/10 9:08:33	等等，适合宜理就业	劳动和社会保障
338	163075	A00083968	A类1000元的技能培训补贴为何被扣	2014/6/8 14:24:03	“小利益”悄然扩张	劳动和社会保障
340	2187	A00087773	A市人社局是否有职业培训的项目？	2019/5/29 14:03:51	就业中心负责具体培	劳动和社会保障
342	169315	A00049384	G4县超越电脑职业培训学校虚假培训骗取国家补贴几十万元	2014/1/29 19:16:33	的人名和身份证号都	劳动和社会保障
406	155982	A00022577	请省府给予小微型创业企业公平竞争环境	2015/9/3 5:24:50	反同tracert西地省移	商贸旅游
466	132582	A00034624	关于L3县开展食品安全专项整治工作的建议	2016/8/18 9:39:06	，可是这些地方才真	卫生计生
473	130464	A0009708	依法治院依法行政，是化解医患纠纷的有效途径	2012/5/5 17:11:53	解。（1）首先死者	卫生计生

③关于“教育”热点意见表格：

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
40	18801	A00075703	A7县松雅西地省片区需要配置公立幼儿园	2015/7/15 9:56:08	带来担忧和恐慌。	城乡建设
42	18819	A00014971	A7县世禧小学垃圾站规划在小区门口离最近住户不到5米	2015/7/7 9:57:31	球场边，哪里不是设置	城乡建设
47	131654	A00078525	L9县一幼建设教学楼不要与L9县十二五规划修编熊公馆相背高	2016/2/11 16:53:00	然而，笔者在春节前，	城乡建设
49	157560	A0009784	强烈要求M4市湾田学校不要盲目改址	014/12/23 13:43:26	解四中的压力，况且	城乡建设
59	12330	A0007948	A市联美小学建粪渗水楼盖	2014/4/29 12:30:46	在开展群众路线教育	城乡建设
69	19386	A000108690	A市轴承厂厂棚改造项目问题的举报	2014/10/2 10:45:22	违纪，违法的问题，对	城乡建设
116	10689	A00061748	关于取消在A3区杨美公园建设110KV变电站的请愿书	2016/1/8 16:48:00	，尊重附近居民的意见	环境保护
192	12926	A00042779	请彻底解决A市属下区中学退休教师福利待遇不公的问题	2013/3/19 17:18:06	早已办好，而对我们	教育文体
193	47762	A0009849	C3县教育局、人社局克扣教师工资	017/11/21 17:10:00	前几人才在平台回复	教育文体
194	5549	A00023078	A市未来之星教育学校部分培训老师师德需加强	2018/6/25 17:44:11	态度冷漠骂他蠢，并	教育文体
197	56988	A00085407	D6县西渡镇咸水学校师资力量严重不足	2018/3/8 18:01:44	推介。但今天想与看	教育文体
198	29080	A00042795	道A9市原民办代课老师的工龄问题	2016/9/1 14:38:37	1、法律依据：根据《	教育文体
200	162961	A00078750	建议我省出台原民办代课教师录用为非教师的国家正式职工的工龄认定相关政策	2015/2/28 9:51:10	间的工龄，理由是：	教育文体
201	106767	A00046530	K5县二小的教学管理太乱了	017/10/25 11:15:58	，学生这么多，而老师	教育文体
202	152328	A0007834	M8县二中请加强教学管理	019/12/30 15:22:31	二中还真有一个英语	教育文体
203	39684	A00072228	我想了解教育经费在B9市是如何安排的？	2013/4/3 20:25:50	分自主经费，我很想知	教育文体
204	47380	A00082617	C4市教育经费问题的疑惑	2014/1/21 0:54:27	为什么有这么大的权、困难	教育文体
205	96293	A00060024	J5县油麻乡中心学校强制要求各村小报虚账骗取国家教育经费	2015/11/16 9:07:46	元下拨款，为什么却	教育文体
208	51129	A00077591	公办学校结果被D1区政府变通为民办小学校	2019/4/15 9:27:38	，为何区政府批准为	教育文体
211	2067	A00072941	A市民办培训机构乱象丛生，温斯翰英语培训强盗霸王条款	2019/6/11 0:43:37	日，本人到温斯翰英语	教育文体
212	69522	A00084539	E10县藕塘小学夷江校区学习环境差	2019/9/11 15:20:51	！！每个孩子都是祖国	教育文体
213	77221	A00026174	F1区为何校园暴力事件屡屡发生？	2014/6/17 13:37:11	又有一伙30多名来历不	教育文体
214	177531	A00072711	H3县江押镇中学关于请求封闭校园东南部围墙缺口的呐喊	2019/4/23 14:04:01	入校园引狼、拆走本校	教育文体
215	553	A00077598	西地省中医药大学的专业认证为何需要一个月？	2019/11/7 9:43:45	时间。现在是信息化	教育文体
216	26481	A00061612	小学生转学难，A市的学位谁家有的？	2015/4/24 23:19:31	给予回复为：直 接问	教育文体
216	26481	A00061612	小学生转学难，A市的学位谁家有的？	2015/4/24 23:19:31	给予回复为：直 接问	教育文体
219	86577	A00033881	G5县九潭私立学校作业太多，学生负担重！	2019/2/23 19:56:46	布置的作业都相当繁重	教育文体
220	107239	A00061072	K1区柳子中学的门票卡收费不便宜，增加学生负担	2019/9/5 18:42:07	钱到底是落入哪些人的	教育文体
223	63210	A000106551	E市袁山培英学校乱收费，一涨就一千	2019/6/27 16:09:11	盐碱、一锅乱炖倒是常	教育文体
226	57773	A0002381	D10县民办高中有统一的收费标准吗	2018/8/24 10:19:55	，“读不起就不要去	教育文体
227	3292	A00094030	希望A市能加快教育体制改革，让A8区的孩子享受市区同样的教育资源	2019/2/15 11:43:34	起将来自小孩的读书问	教育文体
228	63276	A000106447	质疑E市E12市教育局关于“代课教师问题”的回复	2019/5/28 11:31:04	同意设立E12市王龙幼	教育文体
229	89367	A00073021	津市职教集团的教育体制有很多大的问题，必须要彻底改革	2015/3/19 22:07:44	资却很少，受到的批评	教育文体
230	111899	A00035554	K8县教育局应尽早建立督学责任区	2012/5/31 12:28:36	或专业水平的领导，不	教育文体
234	457	A00080495	咨询A市小学生转学问题	019/11/18 22:49:18	该的学习及生活在很	教育文体
235	167856	A00050547	又到一年高考时，I市监考规则不能“潜规则”	2012/5/9 16:56:26	的时间流逝，有些市	教育文体
236	176538	A000102231	G4县龙池学校高考作弊，监考老师未按规定用G2的老师	2015/6/15 9:50:24	矿泉水作接印号，龙	教育文体
237	176737	A00092673	I4县玉潭学校老师参与本校师范生监考，有失公平公正	2019/5/12 12:46:37	，关系到学生的前途	教育文体
240	4710	A00040612	西地省成考招生报名系统一直不能预约！！	2018/9/12 16:56:11	等学校招生统一考试	教育文体
241	141813	A0006681	M3县思源小学2019年招生录取政策怎么变了	2019/8/26 19:08:44	育局。直接内定思源。开	教育文体
242	57664	A00079065	D10县官家嘴镇中心小学招生人数缩减不公平	2019/6/15 6:20:35	嘴镇中心小学招收一	教育文体
243	149166	A000109451	M2县的贫困幼儿补助哪儿去了？	2015/7/29 10:01:32	右下发！问题二：同样	教育文体
244	62297	A00041376	D11市蔡子池中学将上网上报名在家	2013/1/11 14:22:05	老师，我们也进行过	教育文体
245	74432	A00063499	F市经开区五年级女生转学被拒，被迫辍学在家	2019/4/14 10:58:07	！怎么办啊？都说教育	教育文体
246	76444	A00063804	F5县新河乡小孩被学校劝退辍学并注销了学籍！	2019/8/9 22:24:28	教育！现在14岁想去继	教育文体
249	4722	A00041093	对A2区小学体育教育的一点看法	2018/9/12 10:29:51	起上一次体育课，可	教育文体
254	168670	A00055926	C2区体育馆卫生状况堪忧	017/12/18 21:10:43	忍目睹！让人恶心至	教育文体
265	6752	A00074792	西地省卫视能不能多办一点文化类节目？	2018/2/13 15:44:22	多举办这类型的节目	教育文体
272	150475	A00057513	M7市旅游和文化广电新闻出版局多次违规审批网吧	2018/5/26 0:10:54	门由市教育局批准的	教育文体
273	156616	A00090072	请严查西地省际天传媒公司及其领导	2014/11/9 14:28:11	仅不支付。 4、无法	教育文体
273	156616	A00090072	请严查西地省际天传媒公司及其领导	2014/11/9 14:28:11	仅不支付。 4、无法	教育文体
284	160196	A00055169	关于批准抗日英烈任初襄墓为省、市级文物保护的报告	2015/4/15 15:08:21	计其数。“胜利后首个	教育文体
286	160188	A00073968	请求西地省文物局对认定文物发生争议依法予以裁定	2015/8/15 20:54:41	46》号令《文物认定	教育文体
321	85682	A000110357	关于C市教育局本次工资调整的几点看法	2015/12/31 9:05:21	的职称作为衡量工资	劳动和社会保障
339	148414	A00059441	建议要求西地省驾校培训一考取教练员等级证书才能上岗	2019/4/14 23:23:51	切的关系。 一是对	劳动和社会保障
341	123778	A00088194	提高L市民警素质，比如免费职业培训班	2011/5/2 15:00:11	我渴求学习……增加	劳动和社会保障
353	102782	A00013011	水市水库职工生存艰难，恳请领导体察民情	2014/2/25 0:48:56	库依然只有基本工资	劳动和社会保障
377	75713	A00020934	岳化教育系统剥离到地方提前退休老师退休工资低	014/11/21 21:10:11	党的有关政策的，也	劳动和社会保障
406	155982	A00022577	请省府给予小微型创业企业公平竞争环境	2015/9/3 5:24:50	反同tracert西地省移	商贸旅游
439	165639	A000113117	C4市板托村村民没有依法享受公共服务	2014/6/27 11:15:11	公共卫生服务。开展	卫生计生
488	28644	A0009360	再谈A9市人民医院医师职业素养不高、医疗服务差	2018/10/14 21:22:04	治疗，由于患者以前	卫生计生

④关于“医院”热点意见表格：

347	77082	A00010891	关于请求予以认定冯辉金为工伤的报告	2015/9/12 21:26:01	的规定，市人社局不予	劳动和社会保障
351	26255	A00039515	A市交警总队逼迫孕妇辞职	2014/8/11 23:12:33	徐岗位！交警总队不好	劳动和社会保障
352	175415	A00037367	为什么要强制女专业技术人员提前退休？	2014/6/14 18:43:43	重新划分了干部和工人	劳动和社会保障
353	102782	A00013011	水市水库职工生存艰难，恳请领导体察民情	2014/2/25 0:48:56	库依然只有基本工资。	劳动和社会保障
354	159501	A00099133	请保护L市劳动保护用品公司职工的合法权益	2012/5/23 13:16:19	主席替我们做主。	劳动和社会保障
355	180378	A00087073	还我尊严，给我们一个自食其力的机会！	012/10/28 20:15:33	。国家号召全社会都	劳动和社会保障
356	16291	A00097798	请解决A7县蓝思科技职难的问题	2018/9/22 20:58:06	不批，只有自离，自离不	劳动和社会保障
358	31441	A00044109	为何辞退员工还要员工打辞职报告？	2012/5/28 17:39:18	道如何写，家人问起不	劳动和社会保障
359	52404	A00068418	盼严查D6县电信公司非法用工	017/12/15 15:40:24	辞退我们，还我们五险	劳动和社会保障
360	102052	A00057142	X市达福鑫非法用工，劳动监察竟然无人敢管	2015/8/4 15:23:30	劳动合同非法解雇员工	劳动和社会保障
361	175271	A00084829	西省采丰商贸有限公司好润佳I6市店违反劳动法非法用工	2015/7/7 8:07:40	的他们也不管，领导们来	劳动和社会保障
362	8276	A00070937	A市艾依格家具商行不给员工签订正式劳动合同	2017/7/6 23:13:47	给我任何不存在劳动关系	劳动和社会保障
363	67040	A00036372	请求落实解除劳动合同后相关问题	2014/6/1 17:47:40	有限公司应向我支付双倍	劳动和社会保障
364	98174	A00032396	请恢复K2区农村信用合作联社职工的劳动合同，让我们老有所养	2019/1/20 16:29:36	一、（西地省人口与计	劳动和社会保障
366	138971	A00041489	M3县信用社劳务派遣工资只有正式工的百分之二十	2013/8/9 12:30:51	还置若罔闻，这样的行	劳动和社会保障
367	159494	A00020144	关于劳务派遣员工薪金同工不同酬问题	2012/6/29 18:22:33	请员工。现我们公司基	劳动和社会保障
369	5680	A00081901	A5区雅华物业有限公司侵犯农民工的合法权益	2018/6/13 14:07:18	工资怎么养家糊口怎么	劳动和社会保障
370	41544	A00094339	请求维护C市电力设备集团有限公司农民工权益	019/11/15 15:36:58	资。另前后添置设备花	劳动和社会保障
371	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 16:21:57	劳动者和用人单位不存	劳动和社会保障
372	149637	A00036352	关于M7市民申请办理恢复劳动关系的报告	2017/6/28 18:46:06	初字第599号民事判决	劳动和社会保障
373	162634	A00072950	关于请求为中风病人办理转移社保基金办理病退的报告	2017/6/22 18:52:26	个人。在1996年底，	劳动和社会保障
378	123619	A000108792	L市特困企业职工病退如何这么难？	2012/3/17 12:35:33	提出申请办理因病提前	劳动和社会保障
382	1083	A00032432	A8县畜牧系统退休工人何时才能领取退休金？	2019/9/11 9:28:14	一分钱退休金。公司在	劳动和社会保障
386	14137	A00068563	A4区临时环卫工退休工资为何一月才640元？	2016/7/27 20:03:04	所长吃喝玩乐吹空调，	劳动和社会保障
427	175437	A00031896	I市通联违反劳动法扣网络直销员基本工资	2014/5/16 15:19:26	听免费，全国无长途费	商贸旅游
473	130464	A0009708	依法治院依法行政，是化解医患纠纷的有效途径	2012/5/5 17:11:53	解。（1）首先死者	卫生计生

接下来即可由以上筛选表格进行热点问题表的汇总，如下：

表1-热点问题表

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 政府	2 0 1	2 0 1 1 / 1 1 / 1 至 2 0 1 9 / 9 / 8	A市小区 居民	居民利益得不到保障
2	2 学校	1 8 8	2 0 1 2 / 5 / 3 1 至 2 0 1 9 / 9 / 5	A市学校 教育部 门	学校教学管理问题
3	3 教育	1 8 4	2 0 1 1 / 5 / 2 至 2 0 1 9 / 9 / 5	A市教育局	教育局监管不力学生基本利益得不到保障
4	4 医院	1 8 1	2 0 1 1 / 9 / 2 8 至 2 0 1 9 / 8 / 6	各县市 医院	医护工作不到位及医疗工作人员利益得不到保障
5	5 劳动	1 7 4	2 0 1 1 / 9 / 2 8 至 2 0 1 9 / 9 / 1 1	A市员 工	争取员工劳动权益

2.2.2 热点问题明细表

表2—热点问题明显表

问题ID	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
1	31833	A00047470	A5区华银电力金电厂退休职工的房改权益为何被剥夺	2019/9/8 5:21:31	<p>谁也不会相信，我为电力企业奉献四十多年，房改的权益被剥夺。我是华银电力金电厂退休职工，孤寡老人，住A市厂属公房三十多年，申诉多年要求按房改政策买下，但领导们（企业属不属党管，企业属不属政府的）不但拖诿不办，反而要按市场价出卖，（不过欣喜的是，同一单位同一栋楼的领导在当时享受了房改的权益）具有讽刺意味的是，这个企业是全国扶贫济困的先进单位，本企业职工几十年的问题却推诿不办，在外可投送千金，对内仅仅五十平米都被刁难，这叫人何以为堪？青天在上，政府很多阳光政策，尤其是“老有所依，人有所居”，大获民心，我希望有一丝丝阳光温暖我心！</p>	1	0

					<p>作，只是政策不给力。 5、我们的工龄没法认定，有的到其他单位上班还达不到退休时的工作年限（15年），影响正常的退休，同时严重伤害我们的合法权益，也直接影响到我们的退休工资和现时工资水平。</p> <p>0000-00000000</p>		
--	--	--	--	--	---	--	--

2.3 评价方案

可根据如上操作步骤对留言内容进行分类，首先建立留言一级分类模型，对留言内容进行训练以保证对新的留言内容进行归类并转接到相关部门，其次对热点问题的讨论，从一级分类可看热点问题大方向确定为城乡建设，教育文体，劳动和社会保障。再细分词频统计，可知热点词频最高的大部分意见是关于政府，学校，教育，医院，劳动，规划，小区，办理，居民，业主，工资，企业等热点问题。根据这一套评价方案可对接下来收到新的留言内容以便更好地处理。

在评价答复意见的质量之前，我们首先需要明确评价的指标和评价方案。针对附件4中的相关部门对留言的答复意见，我们可以从答复的相关性、完整性、可解释性等角度来评价答复的质量。

相关性：答复意见的相关性指答复内容与留言内容的关联程度。一个好的答复应当能够直接回应留言所提出的问题或反映的需求，解决群众关注的**核心问题**。

完整性：答复意见的完整性指答复内容是**否全面、详尽**。一个完整的答复应当包含必要的信息和解决方案，不应**遗漏关键细节或重要环节**。

可解释性：答复意见的可解释性指答复内容是**否清晰明了、易于理解**。一个可解释性强的答复应当避免使用过于专业的术语或**复杂**的语言，以确保群众能够理解其含义和目的。

时效性：答复意见的时效性指答复是**否及时给出**。及时给出答复可以有效增强群众的**满意度**，提高政府的公信力。

模型搭建

在构建评价答复意见质量的模型时，我们需要考虑到评价指标的多样性和复杂性。下面是模型**搭建**的具体步骤：

特征选择：

首先，我们需要选择**合适**的特征来表示答复意见的质量。根据我们所关注的评价指标，例如相关性、完整性、可解释性和时效性，我们可以选择相应的特征。这些特征可以包括答复内容的长度、包含关键词的数量、回复时间等。

数据预处理：

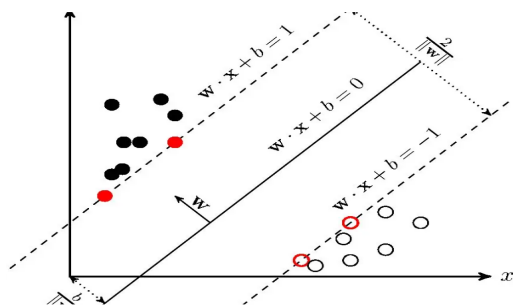
对于答复意见的数据，我们需要进行预处理以清洗和**准备**数据。这可能包括去除无关信息、处理缺失值、进行文本分词、去除停用词等操作。通过数据预处理，我们可以**准备**好用于模型训练的数据集。

模型选择：

根据我们的评价指标，我们可以选择**适合**的机器学习模型来训练和评估答复意见的质量。我们可以选择分类模型（支持向量机）来对答复进行分类，或者选择回归模型来预测答复的质量分数。

支持向量机

支持向量机（Support Vector Machine, SVM）是一种常用于分类和回归分析的机器学习算法，其主要思想是在**特征空间**中找到一个最优的超平面，将不同类别的样本**分隔**开来，并尽量使**分隔边界**的间隔最大化。



支持向量机算法的**核心思想**：支持向量机的**核心思想**是在**特征空间**中找到一个最优的超平面，使得不同类别的样本尽可能地被分开，并且使得**分隔超平面**与样本点之间的**间隔**最大化。

超平面：在二维**空间**中，超平面是一个一维的线性子空间，可以用一个直线来表示；在高维**空间**中，超平面是一个（d-1）维的线性子空间，可以用一个超平面来表示。对于二分类问题，超平面可以将两个类别的样本点**分隔**开来。

间隔最大化：SVM 的目标是找到一个超平面，使得样本点到超平面的**距离**（**间隔**）尽可能大。这个**间隔**可以用数学表示为样本点到超平面的**函数距离**的最小值。

核函数：在实际问题中，有些样本可能不是线性可分的，此时可以通过引入**核函数**来将样本映射到高维**空间**，使其在高维**空间**中线性可分。常用的**核函数**包括线性**核函数**、多项式**核函数**和高斯径向基**核函数**等。

支持向量：在 SVM 中，支持向量是指**距离**超平面最近的那些样本点，它们对确定超平面起着关键作用。这些支持向量决定了**最终**的分类**边界**。

正则化参数：在 SVM 中，通常会引入一个正则化参数（C），用于**平衡**间隔最大化和分类错误的**惩罚**。**C**越大，表示对分类错误的**惩罚**越大，间隔最大化的影响**越小**；反之，**C**越小，表示对间隔最大化的影响**越大**。

优化算法：SVM 的训练过程可以通过求解凸优化问题来实现，常用的优化算法包括序列最小最优化算法（SMO）、梯度下降等。

模型训练：

使用准备好的数据集，我们可以对选定的模型进行训练。在训练过程中，我们需要将数据集划分为训练集和测试集，并使用交叉验证等技术来评估模型的性能。

模型评价：

训练完成后，我们需要评估模型的性能以确保其准确性和可靠性。我们可以使用各种评价指标（如准确率、精确度、召回率、F1分数等）来评估模型在测试集上的表现。

模型调优：

如果模型性能不够理想，我们可以通过调整模型参数、优化特征选择、增加数据量等方式来改进模型性能。调优过程需要反复尝试和实验，直到达到满意的效果。

通过以上步骤，我们可以构建一个能够评价答复意见质量的模型。这个模型可以帮助政府部门更好地了解群众的需求和关切，提高答复的质量和效率。

代码如下：

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import classification_report

# 加载数据
data = pd.read_csv("附件4.csv")
# 数据预处理
# （假设答复意见的文本数据列名为'reply'）
reply_text = data['reply'].apply(lambda x: clean_text(x))

# 特征提取
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(reply_text)

# 定义标签（假设已有标记，标签列名为'label'）
y = data['label']

# 划分训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 模型构建与训练
model = SVC(kernel='linear')
model.fit(X_train, y_train)
# 模型评估
y_pred = model.predict(X_test)
print(classification_report(y_test, y_pred))
```

结果分析

性能评估：

首先，我们需要使用一系列评价指标来评估模型在测试数据上的性能。对于分类问题，常用的评价指标包括准确率、精确度、召回率、F1分数等。我们可以计算这些指标来衡量模型的分类准确性和效果。

混淆矩阵：

通过混淆矩阵可以更详细地了解模型的分类表现。混淆矩阵显示了模型在每个类别上的分类情况，包括真正例、假正例、真负例和假负例的数量。通过分析混淆矩阵，我们可以确定模型在每个类别上的分类准确性和误差情况。混淆矩阵是在分类问题中用于评估模型性能的一种表格形式，它将模型的预测结果与真实标签进行对比，以便更好地理解模型在不同类别上的分类情况。混淆矩阵的结构如下所示：

		预测类别	
真实类别		真正例	假正例
		假负例	真负例

在混淆矩阵中，行代表真实类别，列代表模型的预测类别。矩阵中的每个单元格表示模型将样本预测为某个类别的数量。

真正例（True Positive, TP）：模型正确地将正例样本预测为正例的数量。

假正例（False Positive, FP）：模型错误地将负例样本预测为正例的数量。

假负例（False Negative, FN）：模型错误地将正例样本预测为负例的数量。

真负例（True Negative, TN）：模型正确地将负例样本预测为负例的数量。

通过分析混淆矩阵，我们可以得到以下评价指标：

准确率（Accuracy）：指模型正确分类的样本数占总样本数的比例，计算公式为：

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

精确度（Precision）：指模型预测为正例的样本中实际为正例的比例，计算公式为：

$$Precision = \frac{TP}{TP+FP}$$

召回率（Recall）：指实际为正例的样本中被模型预测为正例的比例，也称为灵敏度或真正例率（True Positive Rate, TPR），计算公式为：

$$Recall = \frac{TP}{TP+FN}$$

F1分数：综合考虑精确度和召回率，是精确度和召回率的调和平均数，计算公式为：

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

通过分析混淆矩阵和计算上述指标，我们可以更全面地评估模型在不同类别上的分类表现，并确定模型的优劣。

特征重要性分析：

如果我们在模型中使用了特征选择或特征工程，可以进一步分析模型中各个特征的重要性。这可以通过查看特征权重、特征重要性得分等方式来实现。特征重要性分析可以帮助我们理解模型的决策过程，确定哪些特征对模型的分类结果影响最大。

3. 结论

一级分类模型的建立过程：该结论表明数据集中没有缺失值，意味着我们可以充分利用整个数据集进行建模和分析，而无需担心数据缺失的影响。另外，指出根据关键词的选择不同会导致不同的分类模型，这强调了在建立模型时需要**仔细选择特征**和关键词，并针对不同的问题领域进行定制化的处理，以提高模型的准确性和泛化能力。

热点问题的探讨：根据对热点问题的探讨，我们可以得知城乡建设、教育文体、劳动和社会保障是民众关注的**焦点问题**。进一步对词频进行统计后得出前五大热点为，政府、学校、教育、医院、劳动等词汇频率**较高**，这反映了民生问题在群众留言中占据重要地位。这些结论为政府部门**提供了**重要的参考，可以**帮助他们**更好地了解民众关切，优先处理与城市发展和民生福祉相关的问题，从而提升治理效能和民生水平。

综合考虑以上结论，我们可以得出一个更深层次的结论，即政府部门应该加强对城市**基础设施建设**、教育医疗保障和劳动就业等方面的关注和投入。通过解决这些热点问题，政府可以提高民生福祉，增强人民群众的获得感和**幸福感**，进而推动城市的可持续发展。同时，政府部门还应积极**倾听民意**，加强与民众的**沟通和互动**，建立更加**开放和透明**的治理机制，共同推动城市社会的进步和发展。

4. 参考文献

[1] https://blog.csdn.net/Eton2016/article/details/78414737?depth_1-utm_source=distribute.pc_relevant.nonetask&utm_source=distribute.pc_relevant.none-taskR语言文本分词以，词频统计，云图显示的简洁方法

[2] <https://blog.csdn.net/Eton2016/article/details/78419531>用R进行文本挖掘与分析--软件分词统计词频

[3] https://blog.csdn.net/Blackrosetian/article/details/78467629?depth_1-utm_source=distribute.pc_relevant.none-task&utm_source=distribute.pc_relevant.none-task