C 题:“智慧政务”中的文本挖掘应用

摘要

随着现代化科技的发达,"互联网+政务服务"是深化放管服改革的关键之举，新型智慧城市的建设对于提升政府治理能力、推动数字经济发展具有重要意义,但数据安全问题的凸显目前正成为新型智慧城市建设的瓶颈。分析了新型智慧城市建设中政务数据共享开放存在的数据安全问题,梳理了各地政府在数据安全上采取的各类安全管理措施,以期全方位保障政务数据共享开放安全,促进新型智慧城市建设。微信，微博，抖音等一些网络平台，为政府了解民意、凝聚民气的重要渠道，通过网络平台人民群众所反映数据，可解决民众问题。

问题一处理：由附件 1 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。分别对一二三级分类由 matplotlib 数据可视化提取制图，聚类数据分析方法进行数据归类分总。由附件 2 对一级分类进行评价，给出通常使用 F-Score 对分类方法进行评价：且给出公式由模型处理附件 2 数据。分出三级之间是从属关系。

问题二处理：某一时间段人民群众集中反映的问题，针对留言情况，进行归类，定义合理的热度指标，运用 pythons 数据分析，并给出按前 5 热点问题“热点问题表”和“ 热点问题留言明细表”的表格。

问题三处理：从群众回复的问题里，由附件 4 数据分析，其相关性， 完整性,可解释性多角度给出一套完整的合理评价方案。

关键字:热点问题、智慧政务、归类分析、聚类数据分析、python 数据分析、合理方案。

1 挖掘目标

1.1 挖掘背景：

随着电子技术、信息技术和网络技术的发展,物联网、移动互联网和云计算等技术日益发展,并推动了大数据的发展,这为以提供公共服务为核心的智慧政务带来了发展的技术基础和推动力量。当前,学术界和理论界并未对智慧政务形成一个统一的概念,但从当前的研究现状来看,一般认为,智慧政务是在信息化时代背景下,综合运用互联网和信息网络技术,以大数据为核心,通过信息化手段为公众提供高效服务的政务模式,以此可以推动政府廉洁高效运转、政府政策制定更加精准、服务大众更为便捷、信息化透明化程度更高。智慧政务是信息通讯技术发展的结果,是政府治理发展的重要方向之一。

1.2 挖掘目标：

在信息化时代下，通过运用云计算、大数据、物联网等通讯技术，通过监测、整合、分析、智能等的响应，实现各职能部门的各种资源的高度整合，提高政府的业务办理和管理效率；加强职能监管，使政府更加廉洁、勤政、务实，提高政府的透明度；形成高效、敏捷、便民的新型政府，保证城市可持续发展，为企业和公众建立一个良好的城市生活环境。并联审批——政府各联网部门实现数据整合和信息资源共享，对政府工作流程进行优化和改造，以标准化服务的方式实现各类跨部门的联动业务，提高政府办事效率。

如今网络沟通发达，根据人们所浏览点赞，回复的热点事件，提取有效信息，并经过一系列处理和分析，方便政府及相关部门了解民意，互动沟通又增加创新的沟通渠道，提供市民与领导，企业与政府之间互动交流的平台机制，加强了与各界代表人士的协商，而且树立一个公平、公正、公开，并且响应快速高效的政府形象。政务建设为了智能化提取政务业务数据，并且指导业务决策和政策推行。

2 问题分析

2.1 问题一 群众留言分类

由于现网络信息众多，仅仅依靠人工处理，工作量及大而且效率非常低，所以需要精准找出人民群众所反馈的热点问题，而且差错率高等问题的解决，所以对所以相关数据进行优化和改造，以标准化服务的方式实现。

问题一中提供了附件 1 分类三级标签体系数据，附件 2 所给出的数据是对附件 1 标签中的一级分类的留言主题，留言时间，留言详情和留言用户和编号。

问题一所给的任务要求是之间的从属关系，解决三级标签分别由 excel 表格分出一级分类的总量及其对应的数量，绘制图表，附件 2 的留言主题和留言详情文本转换，关键字的提取。

2.2 问题二 热点问题挖掘

问题二中给出了附件 3 反映特定地点或特定人群问题的留言，且提供了点赞数和反对数。

问题二的任务要求是对留言主题和留言详情做归类整理，定义合理的热度评价指标。完成表 1“热点问题表.xls”和表 2”热点问题留言细明表.xls”

2.3 问题三 答复意见的评价

问题三的任务要求是如何将相关性和完整性，可解释性描述量化。此问为一个开放性问题，分析以上数据找出合理方案。

3 模型假设与符号说明

3.1 模型假设

a. 问题留言中，留言时间均在近两年，可忽略不考虑

b．使用一级标签分类模型，$F_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{2P_iR_i}{P_i+R_i}$，（Pi 为查准率，Ri 为查全率）

.c. 对热点主题 0-1 个数忽略不计，表中“其他“问题不考虑。

3.2 符号说明

符号	符号解释
X	留言主题
Y	主题条数

4 问题求解

4.1 问题一 群众留言分类

4.1.1 三级标签体系的分类

分类

样本	已观测	已预测 C	城乡建设	党务政务	国土资源	环境保护	纪检监察
训练	C	0	0	0	0	0	0
	城乡建设	0	46	0	0	0	0
	党务政务	0	0	21	0	0	0
	国土资源	0	0	0	11	0	0
	环境保护	0	0	0	0	14	0
	纪检监察	0	0	0	0	0	16
	交通运输	0	0	0	0	0	0
	教育文体	0	0	0	0	0	0
	经济管理	0	1	0	0	0	0
	科技与信息产业	0	0	1	0	0	0
	劳动和社会保障	0	0	0	0	0	0
	民政	0	0	0	0	0	0

		分类					
	农村农业	0	0	0	0	0	0
	商贸旅游	0	0	0	0	0	0
	卫生计生	0	1	0	0	0	0
	政法	0	0	0	0	0	0
	总计百分比	.0%	14.0%	6.4%	3.2%	4.1%	4.7%
测试	C	0	0	0	0	0	0
	城乡建设	0	4	0	0	0	0
	党务政务	0	0	1	0	0	0
	国土资源	0	0	0	1	0	0
	环境保护	0	0	0	0	1	0
	纪检监察	0	0	0	0	0	2
	交通运输	0	0	0	0	0	0
	教育文体	0	0	0	0	0	0
	经济管理	0	0	0	0	0	0
	科技与信息产业	0	0	0	0	0	0
	劳动和社会保障	0	0	0	0	0	0
	民政	0	0	0	0	0	0
	农村农业	0	1	0	0	0	0
	商贸旅游	0	0	0	0	0	0
	卫生计生	0	0	0	0	0	0
	政法	0	0	0	0	0	0
	总计百分比	.0%	15.6%	3.1%	3.1%	3.1%	6.3%

因变量: YIJIFL

(1)一级分类与二级分类的关系图表

a 安全生产
b 城市建设和市政管理
c 城乡规划
d 村镇建设
j 其他

1、 城乡建设

f 国有土地房屋征收与补偿
g 集体土地房屋拆迁与 补偿
h 建筑市场
I 住房保障与房地产
e 工程质量

a 党的建设
b 港澳台侨
c 国防外交
d 民族宗教

2、党务政务

e 群众团体
f 宣传舆论
g 政治体制
e 其他

a.海洋气象地震
b 破产资源管理
c 土地征收

3、国土资源

d 土地资源管理
e 其他

4、环境保护

a 环保管理
b 环境污染
c 建设项目审批
d 其他

a 党政处分
b 干部作风
c 滥用职权

5、 纪检监察

d 失职渎职
e 贪污贿赂
f 其他

a 出租车管理
b 建设管理
c 交通运输

6、交通运输

d 客货运输
e 其他
f 邮政管理

a 教师队伍和待遇
b 教育行政管理
c 教育收费
d 考试招生
e 教育体制

7、教育文体

f 其他
g 失学辍学
h 体育
I 文化
j 文物管理

a 安全生产
b 保险证券期货
c 国资监管
d 宏观调控

8、经济管理

e 金融财税
f 能源管理
g 其他
h 企业破产

9、科技与信息产业

a 电信
b 科学技术
c 其他
d 信息化建设
e 知识产权

a 城镇居民社会保险
b 工资福利
c 就业培训

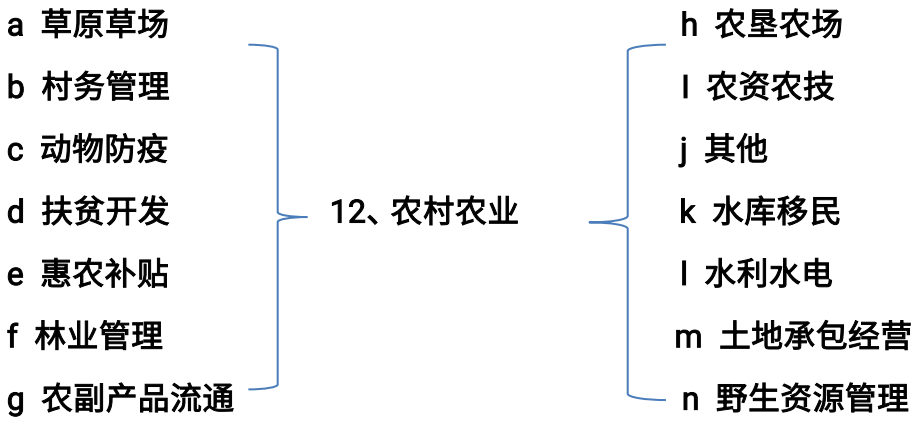
10、劳动和社会保障

d 劳动保护
e 其他
f 退休政策及待遇

a 福利慈善
b 复退安置
c 基层选举和社区建设
d 救灾募捐
e 民间组织

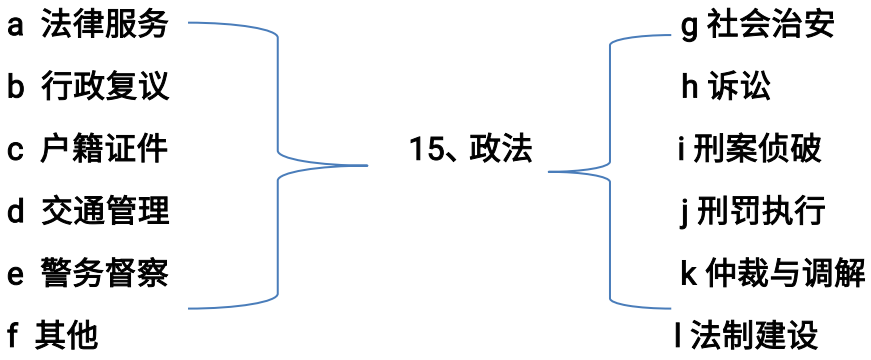
11、民政

f 其他
g 区划地名
h 社会服务
I 社会救助
j 优抚


a 草原草场
b 村务管理
c 动物防疫
d 扶贫开发
e 惠农补贴
f 林业管理
g 农副产品流通
12、农村农业
h 农垦农场
I 农资农技
j 其他
k 水库移民
l 水利水电
m 土地承包经营
n 野生资源管理


a 旅游管理
b 商业贸易
c 其他
13、商贸旅游
d 市场监管
e 质监检验检疫

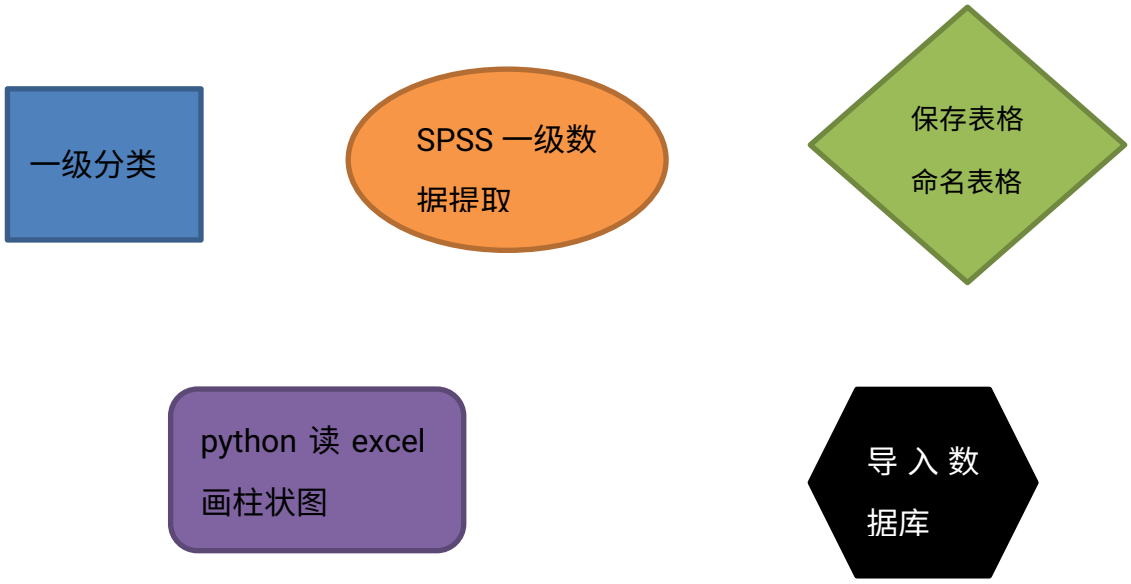

a 公共卫生
b 人口计生
c 其他
14、卫生计生
e 食品药品监管
f 医患纠纷
g 医政监管


a 法律服务
b 行政复议
c 户籍证件
d 交通管理
e 警务督察
f 其他
15、政法
g 社会治安
h 诉讼
i 刑案侦破
j 刑罚执行
k 仲裁与调解
l 法制建设

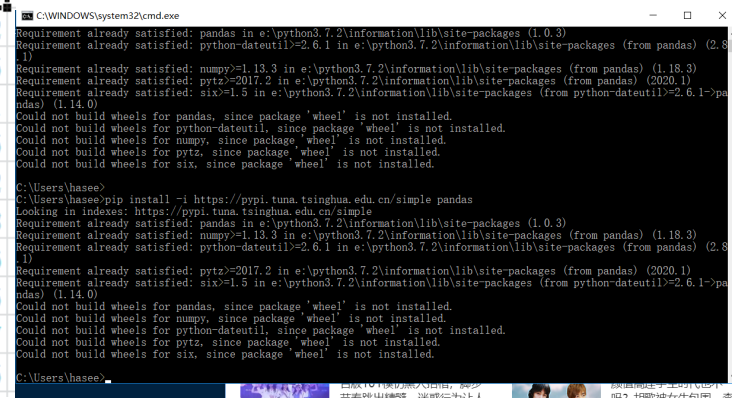
(2)一级分类与三级分类数量关系图表

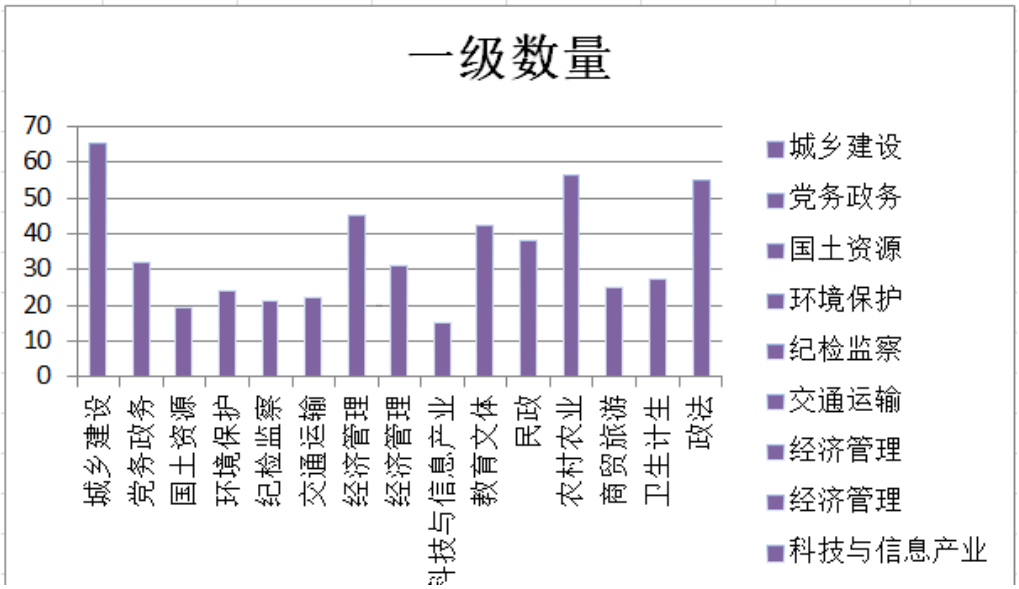
序号	一级标签名称	三级数量	序号	一级标签名称	三级数量
1	城乡建设	65	9	科技与信息产业	15
2	党务政务	32	10	劳动和社会保障	42
3	国土资源	19	11	民政	38
4	环境保护	24	12	农村农业	56
5	纪检监察	21	13	商贸旅游	25
6	交通运输	22	14	卫生计生	27
7	教育文体	45	15	政法	55
8	经济管理	31			

(2)三级分类流程图：



一级名称	一级数量
城乡建设	65
党务政务	32
国土资源	19
环境保护	24
纪检监察	21
交通运输	22
经济管理	45
经济管理	31
科技与信	15
教育文体	42
民政	38
农村农业	56
商贸旅游	25
卫生计生	27
政法	55





4.1.2 三级标签体系的关系

案例处理汇总

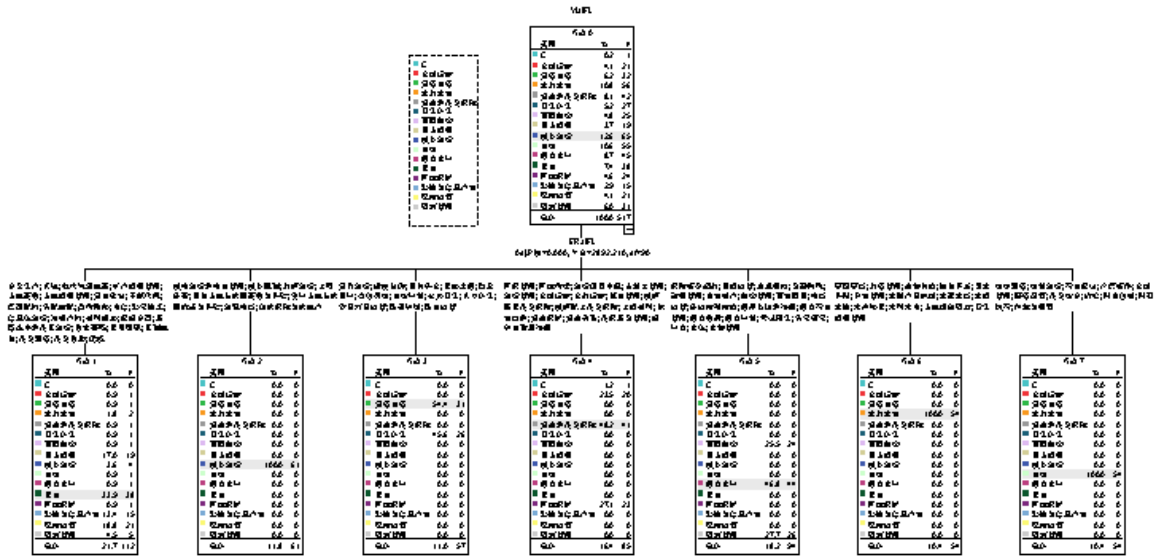
		N	百分比
样本	训练	344	91.5%
	测试	32	8.5%
有效		376	100.0%
已排除		141	
总计		517	

模型汇总

训练	交叉熵错误	85.649
	百分比错误预测	4.7%
	中止使用的规则	错误未减少的 1 连续步骤[a]
	培训时间	0:00:02.297
测试	交叉熵错误	25.829
	百分比错误预测	25.0%

因变量: YIJIFL

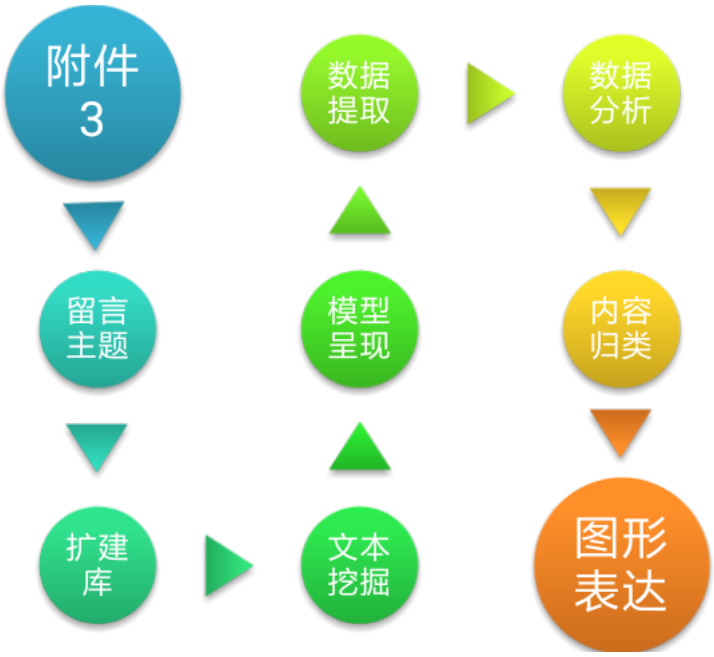
a. 基于检验样本的错误计算。



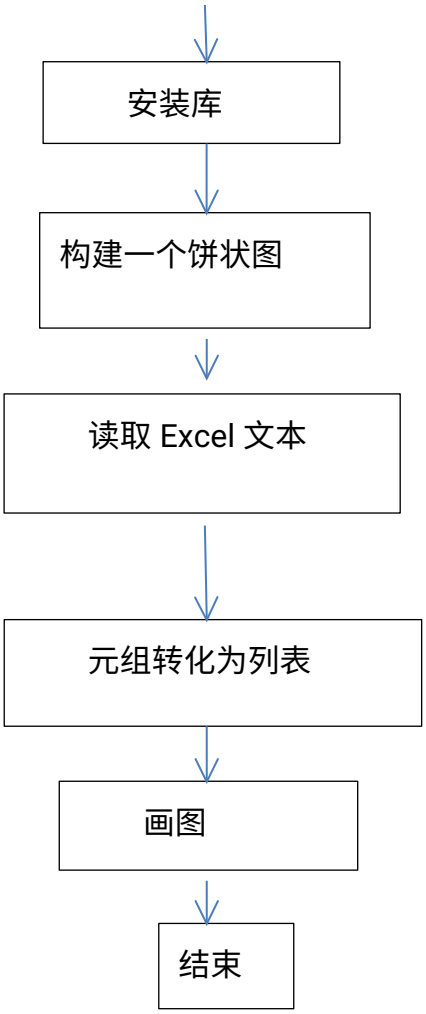
4.2 问题二 热点问题挖掘

4.2.1 问题识别

(1) 附件 3 留言详情流程图



开始

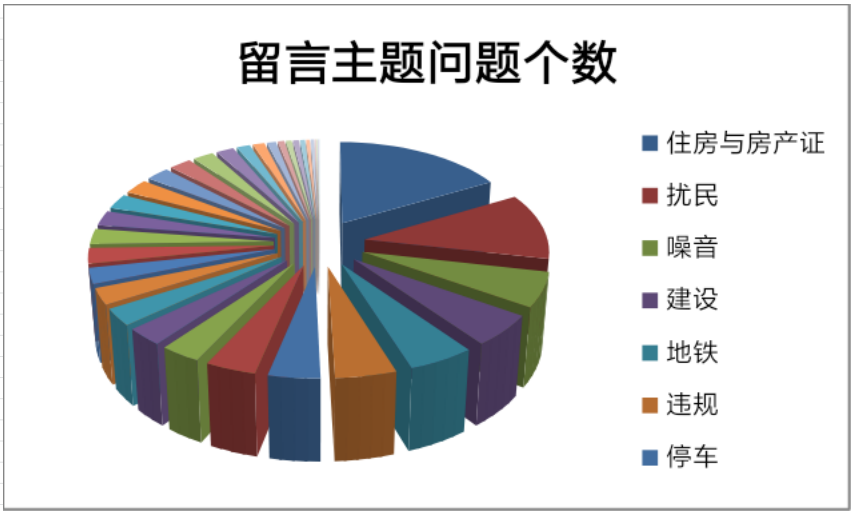


4.2.2 问题归类

⑴ 表 1:“热点问题表.xls” 及表 2:“热点问题留言表详情见附件文件。

⑵ 由附件 3 所提取的数据进行了汇总，如下结果。

	A	B
1	问题	问题个数
2	住房与房	419
3	扰民	278
4	噪音	147
5	建设	133
6	地铁	128
7	违规	117
8	停车	99
9	安全	97
10	幼儿园	78
11	拖欠	74
12	污染	71
13	公交车	68
14	拆迁	66
15	学校	65
16	垃圾	65
17	涉嫌	63
18	购房	63
19	非法	63
20	工资	58
21	环境	54
22	学院	51
23	诈骗	43
24	停水	33
25	质量问题	29
26	补课	22
27	违停	16



4.2.3 热度评价

热度排名	问题ID	热度指数	地点/人群	问题描述
1	1	1767	A市金毛湾/学生	A市金毛湾配套入学的问题
2	2	60	A市长房云/居民	多栋房子现裂缝，质量堪忧
3	3	57	A6区月亮岛/群众	A6110kv高压线的建议
4	4	43	A7县东六线/人民	建议加大拆迁力度
5	5	40	A7县/A7县人	出让星沙滨湖路以南，特立路以北的土地

上图为热度排名前五名，对入学，以及房屋质量，路段高压线，拆迁和土地问题，入学是每个家庭里都要面临的问题，对相关性问题的留言也提出的具体化。房屋质量的好坏上人们生活水平的提现，如今越来越多的人生活得到改善，对住房的要求也是有了更多更好的建议，路段安全问题时时刻刻都在提醒我们，生命只有一次，安全路段就是安全我们。拆迁也是现如今我国脱贫攻坚的这个重点问题。

提取留言主题数据中，居民住房问题，噪音扰民，人们出行交通等问题为首，住房问题一直是群众所关心的话题，结婚买房，住房条件和改善等都收关注。随着人们生活水平的提高，人们越来越注重生活环境的改善，其中噪音扰民问题就是一项严峻问题。

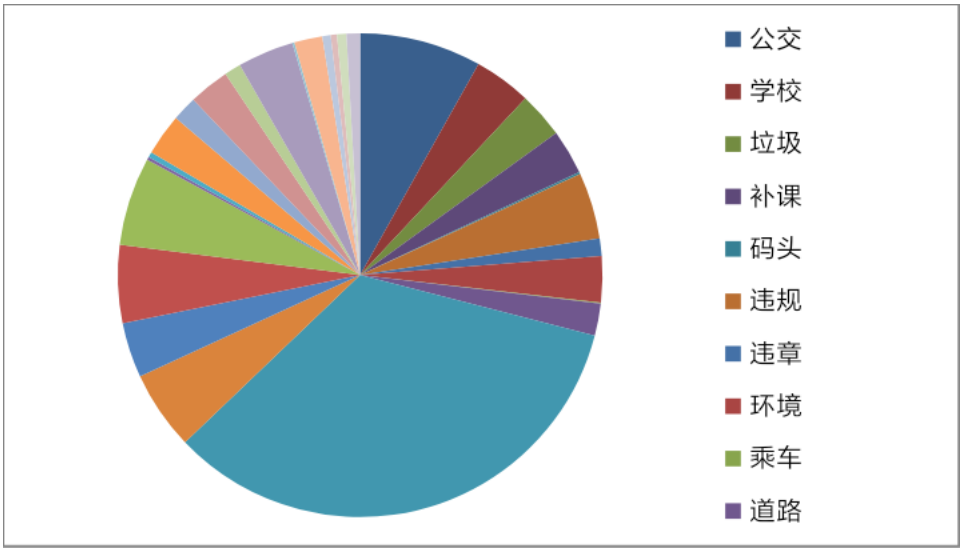
4.3 答复意见的评价

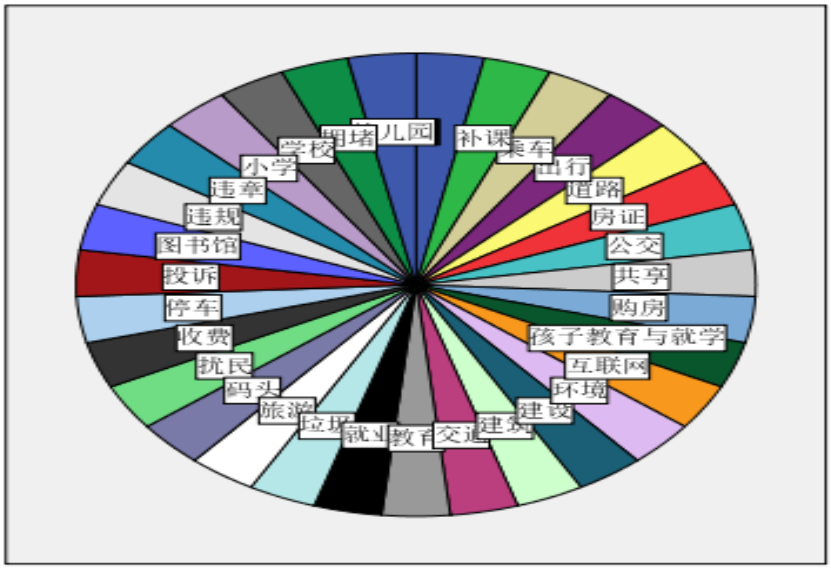
4.3.1 相关性

相关性

		Y	X
Y	Pearson 相关性	1	.[a]
	显著性（双侧）		.
	N	31	0
X	Pearson 相关性	.[a]	.[a]
	显著性（双侧）	.	
	N	0	0

附件 4 热点问题留言主题





热点问题与我们生活息息相关，学校，住房，教育问题，是人们时刻关心的话题，这些问题它不是独立存在的，总体分布均匀，学校问题中也会包括教育，建设，环境维护，交通方面，有安全，违章，乱停车等等，每一个问题的存在它不是独立存在，而是有所关联，问题载体的根源其实来源群体，人们活动相互交错，就会有一系列相关性问题产生，互联网问题也有所提高的趋势，我们对互联网的安全，隐私等问题，意识上都在开始逐步提高。

4．3.2 完整性

现部分政务服务中心审批部门以及工作人员在一定程度上存在重视前置审批,后续跟踪不足的问题. 问题堆积等，但这些热点所提出的问题都有一一的答复，例如一些住房问题和一些交通安全问题都有提出一些解决的方法和措施。而一些对于互联网和住房问题等等所答复的结果也不完整，但基本上所有应当回答的问题都已完整的回复了，所以整体来说是具备完整性的。

4.3.3 可解释性

方案通过对智能化设备和应用系统的集成应用，让办事企业和个人在服务大厅时刻体验科技元素、贴心服务，为公众提供了全方位、高智能、更加方便快捷的服务体验。包括：综合导视、排队叫号、电子样表、数字化窗口工位、交互评价、自助办理、自助取件等应用系统。信息技术在住房，交通安全，噪音扰民，学校，交通工具，线上实行整理，也加快了智慧城市建设，从实际出发，运用大数据，人工智能等新一代的技术推动城市化治理现代化，让生活变得更加美好。

5 总结

推动城市治理体系和治理能力现代化的必由之路。要通过夯实智慧城市建设基础、提升智慧政务服务水平、大力发展数字经济、构建智慧民生服务体系、打造数字孪生城市等途径，加快推进智慧城市建设，真正实现城市让生活更美好

智慧城市是数字经济的重要载体，数字经济是智慧城市的基础支撑。发展数字经济，一是 打造基于城市开放数据的创新应用平台。鼓励第三方机构开展基于开放数据的增值开发和创新应用，促进数据资产的交易流通，释放数据红利，提升数字经济活力。二是加快 5G 网络建设，深化融合应用。高速率、大容量、低时延的特性，使 5G 网络成为实现万物互联、各种技术创新应用最关键的基础设施。三是积极推进数字产业化、产业数字化，引导数字经济和实体经济深度融台。

智慧城市的目标是满足人民群众日益增长的美好生活需要，为群众提供高质量的就业、教育、医疗、养老等公共服务，让群众真正感受到智慧城市的成效。构建智慧民生服务体系，一是整合各类信息资源及应用，为群众提供本地化、统入口的公共信息服务。是建立一 站式线上线下办事大厅，让数据多跑路，让群众少跑腿。三是按照政府主导、统一规划、市场运作、资源共享、一卡多用的原则，推广应用集政务服务、社会保障、生活服务于一体的市民一 卡通，并围绕群众需求不断创新应用场景，提升民生服务水平。

6 参考文献

[1]周君,王显强.新型智慧城市下政务数据安全管理的研究[J].信息通信技术与政策,2020(03):29-33.

[2]李艳.智慧政务背景下优化营商环境路径分析[J].中国市场,2019(27):189-190.

【3】让城市更聪明更智慧。[N】 陈显中 河北日报 2020-04-22

附件

1. **建模所使用软件：**Microsoft Word 2010. Microsoft Excel 2010. IBM SPSS Statistics 22. Python 3.7 64-bit. SPSS Statistics 17.0 (代码如下)

```
import pandas as pd

df - pd. read excel."一级数据表. x1sx")
df
#读取文件中的数据
from pyecharts import options as optsfrom pyechar ts import Bar

bar= (
    Bar ()
    .add_ xaxis(df[”一级名称“]. to 1ist ())
    .add_ Yaxis("- 级数量”，df[”-级数量”]。to_ 1ist())
```

```
        set_ g1obal_ opts (title_ opto=opto.
    Tit1e0pto(title= "一级数量对比图"))

    bar. render_ netbook  ()
```

```
from itertools import chain, combinations
from openpyxl import load_ wor kbook

def loadDataSet () :
'''加载数据，返回包含若干集合的列表'''
result = []
# x1sx 文件中有 3 列，分别为一级分类、二级分类、三级分类
ws = load workbook('附件 1.x1sx') .worksheets[0]
for index, row in enumerate(ws. rows) :
#跳过第一行表头
        if  index==0:
                continue
        result .append (set (row[2] .value.split(', ')))return result

def createC1 (dataSet) :
'''dataSet 为包含集合的列表，每个集合表示一个项集
返回包含若于元组的列表，
每个元组为只包含一个物品的项集，所有项集不重复'''
return sorted (map (lambda i: (i,), set (chain(*dataSet))))

def scanD(dataSet, Ck, Lk, minSupport) :
```

```
'''dataSet 为包含集合的列表，每个集合表示一个项集
  ck 为候选项集列表，每个元素为元组
  minsuppp 为最小支持度阈值
返回 Ck 中支持度大于等于  minSupport  的那些项集”..
#数据集总数量
    total = len (dataset)
    supportData = {}
    for candidate in Ck:
#加速，k-频繁项集的所有 K-1 子集都应该是颇繁项集
    if LK and    (not all (map(lambda item: item in Lk,
    combinations (candidate,
    len (candidate)-1)))) :


    continue
#遍历每个候选项集，统计该项集在所有数据集中出现的次数
#这里隐含了一个技巧: True 在内部存储为 1
    set_ candidate = set (candidate)
    frequencies = sum (map (lambda item: set_ candidate<=item,dataSet) )

#计算支持度
    t = frequencies/total
#大于等于最小支持度，保留该项集及其支持度
    if t >= minSupport:
    supportData [candidate] = t
return supportData
def aprioriGen(Lk, k):
          '''根据 k-项集生成((k+1)-项集'''
result = []
     for index, iteml infor    enumerate (Lk) :
```

```
        for index, teml in Lk[index+1:]:
          if sorted(item1 [:k-2]) == sorted(item2[ :k- 2]) :
          result.append (tuple(set (item1) Iset (item2)))
    return result

def apriori (dataSet, minSupport=0.5):
  '''根据给定数据集 dataSet,
    返回所有支持度>=minSupport 的频繁项集'''
    C1 = createC1 (dataSet)
    supporttData = scanD (dataSet, C1, None, minSupport)
    k =2
    while True;
      LK = [key for key in supportData if len(key)==k-1]
      Ck =    aprioriGen(Lk, k)
#筛选满足最小支持度的(k+1)-项集
      supK = scanD (dataSet, Ck，  Lk, minSupport)
#无法再生成包含更多项的项集，算法结束
      if not supK:
                                        break
          supportData. update (supK)
          k =k+1
    return supportData

def findRules (supportData,
minConfidence=0.5) :
"""查找满足最小置信度的关联规则"""
      #对频繁项集按长度降序排列
          supportDataL = sorted (supportData.items()，
         key=lambda item: len (item[0]),reverse=True)
```

```
        rules = []
    for index, pre in enumerate (supportDataL) :
                for aft in supportDataL[index+1:] :
#只查找(k-1)-项集到 k-项集的关联规则
if len(aft[0]) < len(pre[0])-1:
       break
#当前项集 aft[0]是 pre[0]的子集
#且 aft[0]==>pre[0]的置信度大于等  于最小置信度阈值
       if set(aft[0]) <set (pre[0]) and\
pre[1]/aft[1]>=minConfidence:
        rules . append([pre[0],aft[0]])
  return rules
                                dataSet = loadDataSet ()

supportData = apriori (dataSet, 0.2)

#在所有频繁项集中查找并输出关系二级分类

bestPair = [item for item in supportData    if len (item)==2]
print (bestPair)
for item in findRules (supportData,    0.6) :
          pre,    aft = map (set,    i tem)
          print (aft, pre-aft, sep='==>')



from pyecharts import options as opts
from pyecharts.charts import  pie
def create_pie(datas,title) -> pie:
```

```
    pie =pie()
    pie.add("",datas)
    pie.set_global_opts(
        title_opts=opts.TitleOpts(title=title),
        legend_opts=opts.LegendOpts(pos_right="right")
    )
    pie.set_series_opts(label_opts=opts.LabelOpts(formatter="{b}:{c}:{d}%"))
    return pie
import pandas as pd
df = pd.read_excel(r&apos;文件所在位置.xlsx&apos;)
df_wenti = df.groupdy("wenti").size().sort_values(ascending=False)
df_wenti
datas = list(zip(df_wenti.index.to_list(),df_wenti.to_list()))
datas
pie = create_pie(datas,"饼图-问题对比")
pie.render_notebook
```