

报名序号：2020106225

论文题目：基于“智慧政务”下的数据挖掘评价

目录

一、 问题重述.....	2
1.1 问题背景.....	2
1.2 提出问题.....	2
二、 问题分析.....	2
1. 问题一的分析.....	2
2. 问题二的分析.....	2
3. 问题三的分析.....	2
三、 模型假设.....	3
四、 符号说明.....	3
五、 模型的建立与求解.....	3
5.1.1 数据挖掘处理.....	3
5.1.2 应用模型.....	5
a 分层多粒度的标签体系.....	7
b 潜在语义分析.....	7
5.1.3 数据便准化处理校验.....	8
5.2 问题二模型的建立与求解.....	8
5.2.1 建立模型.....	8
5.2.2 模型求解.....	12
5.3 问题三的求解.....	15
5.3.1 问题分析与模型建立.....	15
5.3.2 模型的求解.....	16
6.1 模型的优点.....	21
6.2 模型的缺点.....	21
综合评价.....	21
参考文献.....	23
附录.....	24

摘 要

在大数据背景下，市民通过网络平台反馈到政府部门有关意见，对政府部门解决社会民生、法制建设、教育文化等提出了新的机遇与挑战，如何建立高效、准确的政府评价机制为本文研究的重点内容。

针对问题一群体留言分类，将文本、时间转换为数字进行量化处理，采用 Python 数据编程按照一定流程将数据剔除杂值，利用推荐机制模型深入探究数据背后的问题类别，最终得出目标数据，进而建立起一级分类标签评价模型。

针对问题二热点问题挖掘，首先建立其相关的评价指标，依据指标将数据利用 python 软件挖掘，利用模糊综合评价与层次分析法评价出热度相关性，运用 MATLAB 编程软件算出矩阵结果，最终得出热度问题。

针对问题三答复意见评价，建立评分机制，根据指标进行评分，采用综合评价模型分析得出政府答复满意度，效率性，完整性等综合评分，利用 MATLAB 软件测算出矩阵结果，利用灰色多层次分析出留言质量评价体制。

关键词：python 软件 MATLAB spss 模糊综合评价 层次分析法 答复评价模型 灰色多层次分析

Summary

In the context of big data, citizens feedback to relevant government departments through online platforms, presenting new opportunities and challenges to government departments in solving social livelihood issues, legal construction, education culture, etc. , how to establish an efficient and accurate government evaluation mechanism is the focus of this paper.

For the problem a group of message classification, text, time conversion to numbers for quantitative processing, the use of Python data programming in accordance with a certain process to eliminate the data out of miscellaneous value, the use of the recommendation-destroying mechanism model in-depth exploration of the problem types behind the data, finally, the target data is obtained, and then the evaluation model of the first-level classification label is established.

Aiming at problem two hot issues mining, firstly, the relevant evaluation index is established. According to the index, the data is mined by python software, and the correlation of heat is evaluated by using mould and AHP, using Matlab programming software to calculate the matrix results, the final heat problem.

In view of the evaluation of the responses to question 3, a scoring mechanism is set up, the indexes are graded, and a comprehensive evaluation model is adopted to analyze and obtain a comprehensive score such as the satisfaction degree, efficiency and completeness of government responses. The Matrix results are calculated by using Matlab software, using grey multi-level analysis to find out the evaluation system of message quality.

Keywords: Python Software Matlab SPSS module and comprehensive evaluation, analytic hierarchy process, reply evaluation model, grey multi-level analysis.

一、问题重述

1.1 问题背景

随着互联网大数据时代的日益进步发展，无论是青年群体还是中老年人群体，热衷于网络留言对政府工作的支持，利用网络留言成为了政府了解市民的重要渠道。一些市民通过公众号、微博等渠道及时给政府有关部门反馈社会、法制、教育、生产生活等信息，极大地提高了市民与政府部门之间的交流工作，同时也间接地推动了社会主义现代化建设的进程。在信息爆炸时代，利用“大数据”政府部门可以有效提取民众意见，民众有效的留言为政府工作以及区域建设做出积极的反馈，也体现出社会主义制度的优越性，关注民生，关注市民发展。

1.2 提出问题

问题一在市民较为多的数据中，如何利用模型划分市民留言的种类，以及如何将热点问题规划处理，并如何采用模型评价。

问题二如何利用模型将市民热点问题评价。

问题三如何建立高效性评价处理智慧平台让民众满意，如何建立数学模型将民众的留言保证科学化、效率化、公平化处理，提高效率并建立正确算法评价政府有关部门的回复评价。

二、问题分析

1. 问题一的分析

- 1: 采用 python 软件将数据归一化处理深入挖掘原始数据。
- 2: 制定处理流程图，将处理过程简练有效输出。
- 3: 采用科学体系描述出评价结果。
- 4: 利用题意 F-Score 公式将模型进行效度评价。

2. 问题二的分析

- 1: 建立合理、有效的影响热度问题的指标评价体系。
- 2: 利用 z-score 标准化方法将数据进一步处理。
- 3: 利用模糊综合评价模型和层次分析评价出热点问题。

3. 问题三的分析

- 1: 采用答复综合评价模型评分。

2: 建立评分机制。

三、模型假设

- 1: 假设文字转换为数字合理化。
- 2: 假设各指标对结果分析具有合理性。
- 3: 假设咨询专家对各个指标判断合情、合理。

四、符号说明

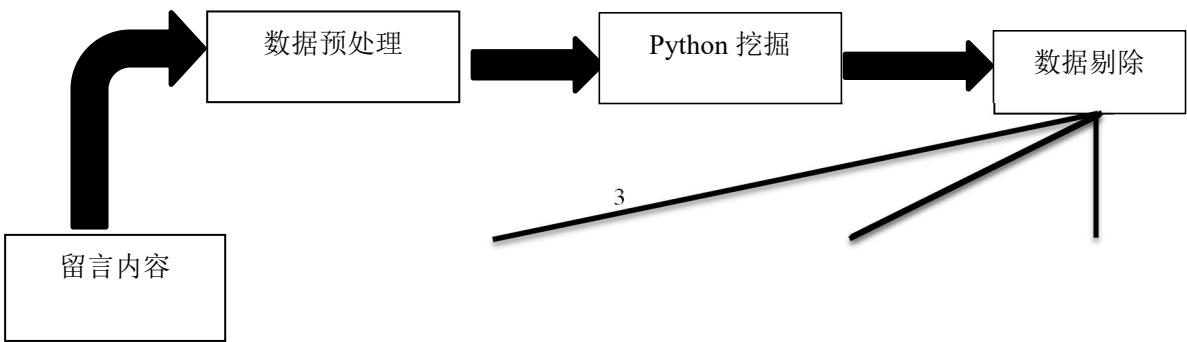
如 4-1 所示

符号	意义
λ_{\max}	最大特征值
ω	特征向量
CR	判断矩阵的随机一致性比率
RI	平均随机一致性指标
W_i	各个分向量的权重系数
C_{ijk}	各级指标

图 4-1

五、模型的建立与求解

5.1.1 数据挖掘处理



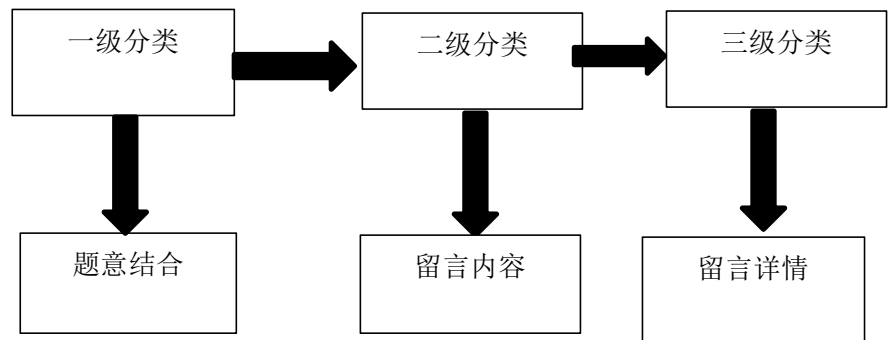


图 5-1

由于各文字、字符表述含义不同，因此采用数字转换形式，将每位用户留言转换为数字形式，例如城乡规划设置为一级分类为 1，安全事故处理设置为二级分类为 12 而 A3 道路安全问题设置为三级分类 123；环境保护设置为一级分类为 2，噪音问题设置为二级分类 23，而 A8 县公路噪音问题设置为 231，依次设置为具有一定长度的数字维度变量。

根据指标进行初步剔除。将数据导入 Python 软件根据以下步骤指标进行剔除得到有效数据。

如图 5-2 所示



图 5-2

在这里我们初步确定完初步有效数据后，在无忧地图上选取青岛不同区来代表附表中 ABC 市县区地区，如图 5-3 所示



图 5-3

通过挖掘地图找到公众提出问题的地点，从杂乱数据中精确定位到政府需要解决的公共地点，有效剔除重复值，利用大数据的准确性和有效性。

5.1.2 应用模型

利用经 python 处理后的数据进行研究阐述，得出如图 5-4 所示关系图

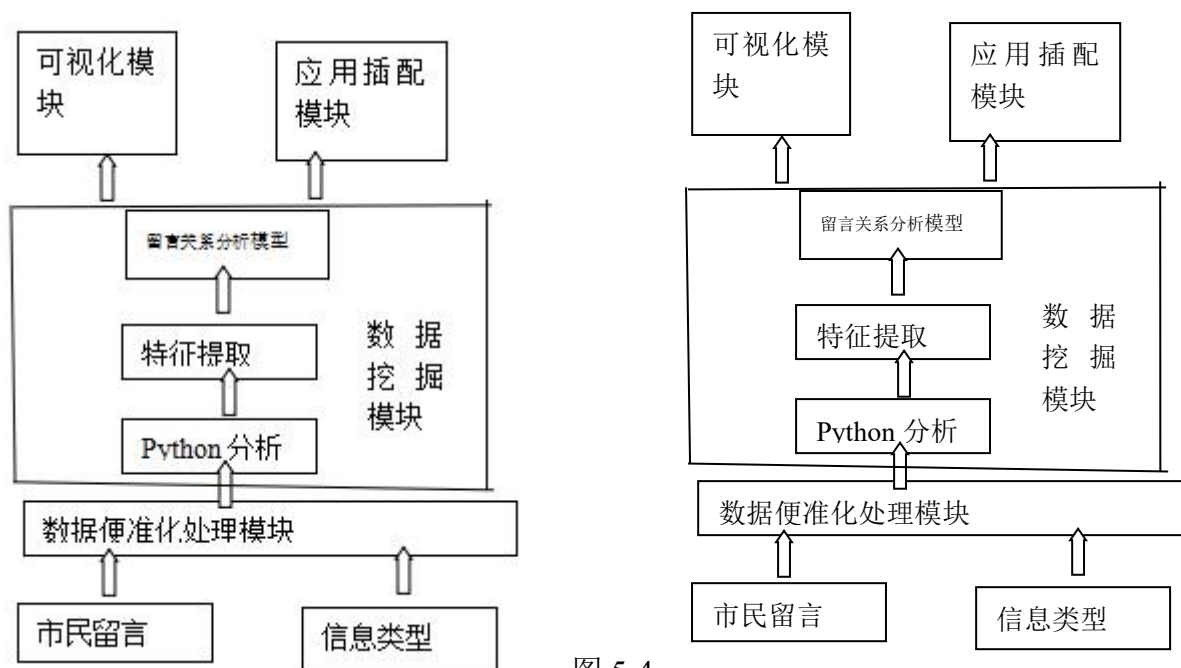


图 5-4

直观测绘出留言群体的留言类别，动态展示出留言群体的留言类别，有利于政府营销参考，并且支持扩展行为。

对留言群体的信息自动进行分类→生成了留言标签。

通过不同类别的标签通过 python 筛选对使用行为做出内容的相关分析，给政府查找工作提供便捷性。如图 5-5 所示

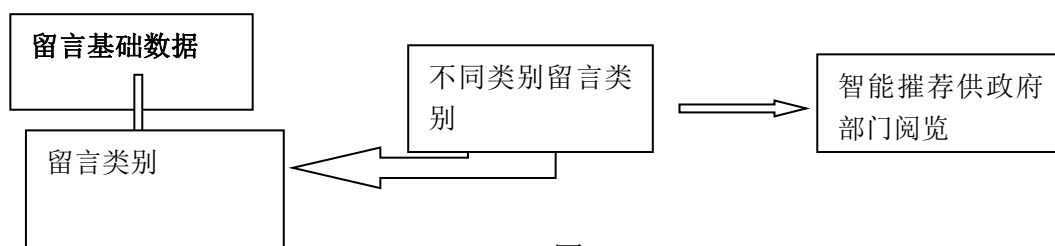


图 5-5

留言基础数据如表 5-1

基础数据库	行为留言信息				
	Python 剔除	反应问题地点	问题反应类型	问题所展现现象	解决途径

表 5-1

行为数据分析如表 5-2

10432	U0002971	住房问题	2016/4/6 20:00:48	楚 江新区新房 质量差	城乡建设
4445	U0004752	噪音问题	2018/10/11 10:49:50	坪塘大 道施工噪音 扰民	环境保护

表 5-2

留言信息数据分析类型：

数据变换、归一化处理：解决特征计量单位、大小范围差异大的解决问题，Kmeans 聚类：

使用基础数据聚类出类型的留言群体：如图 5-6 所示

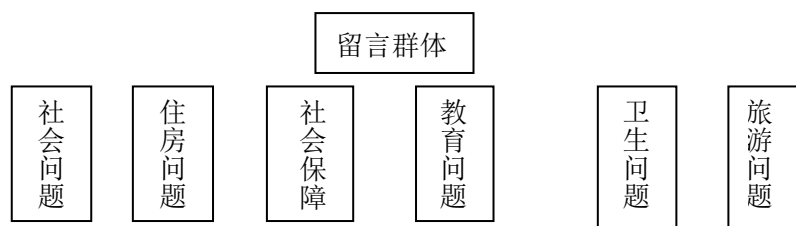


图 5-6

基于文本内容过滤推荐模型，如表 5-3 所示

功能	通过留言用户访问的数据特征，分类留言类型。
使用范围	政府部门
模型输入	在数访问的行为目标留言用户 u 、该数据分析过的类别，集合 I_u 以及集合中元素的关键词描述，留言数据未访问过的内容项集合 $R I_u$ 以及集合中元素关键词描述、对留言数据未访问过的内容项集合 $R I_u$ 以及集合中元素的关键词描述。
模型输出	留言用户 u 喜欢的内容项 Top-N 列表

表 5-3

a 分层多粒度的标签体系

留言特征按照大类-子类-关键词方式进行组织，能够适应政府查阅需求。

b 潜在语义分析

完成不同语境下多义词。
基于文本内容过滤推荐模型

idf_i = log (|D| / |{j : t_i ∈ d_j}|)

一个词语的权重为：

tfidf_{i,j} = tf_{i,j} × idf_i

根据留言类型，用向量 W_c = (w_{c_1} + w_{c_2} + ... w_{c_n}) 表示用户问题

政府部门对其关注兴趣度：r_{cs} = cos(w_c, w_s) = (w_c w_s) / (|w_c| |w_s|)

根据推荐机制关联规则，如表 5-4、表 5-5 所示

向量1	向量2	向量3	向量4	向量4	向量5	向量5	向量6	向量6	向量7	向量8
1	0	1	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0
0	1	1	0	1	1	0	0	0	0	0
1	0	0	1	0	0	1	0	0	1	0
0	0	1	0	0	0	0	1	1	0	0
0	0	1	1	1	0	1	0	0	0	1
0	0	1	0	0	0	0	0	1	0	0
0	0	1	1	0	0	1	0	0	0	0
0	0	1	0	1	0	0	0	0	0	1
0	0	0	1	0	0	0	0	0	1	0
0	0	1	1	0	0	1	1	1	0	0
0	1	1	1	1	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0
0	0	1	1	0	0	0	0	0	1	0
0	0	1	0	0	0	0	1	1	0	0
0	0	1	1	0	0	1	0	1	0	0
...

表 5-4

10	A00040612	高等教育	18/9/12 16:56:	教育市场化	教育文体
22	A00041093	招生考试	18/9/12 10:29:	考场纪律	教育文体
83	U0002289	招生考试	11/5/25 14:41:	跨省高考	教育文体
11	U0006346	招生考试	11/5/27 13:49:	信息公开制度	教育文体
31	A00072387	招生考试	18/7/15 10:30:	考试公平	教育文体
49	A00023078	招生考试	18/6/25 17:44:	烈士子女加分问题	教育文体

表 5-5

匹配出一级标签下相同的留言情况：

得出推荐查准率为 85.62%，查全率为 83.41%。

5.1.3 数据便准化处理校验

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

利用题意中给出公式测算出
最终测算出约为 83.82%。

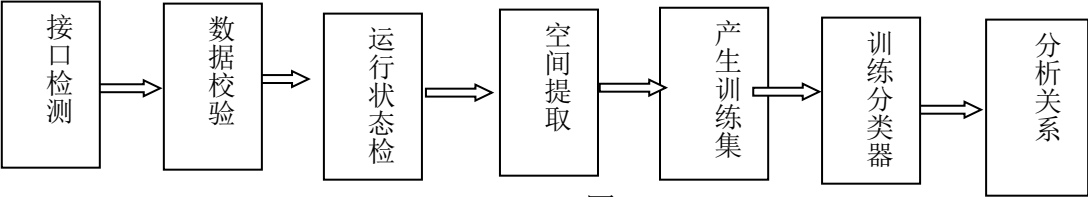


图 5-7

如图 5-7 所示，首先将目标文件检测配置，然后检测表的结构，清洗出数据，检查出实体性的完整性与参照完整性。检测模型运行状态，最后识别数据地点，得出出现频率较高地区，并使用 python 有效提选出市民所提出的问题所在，如下表 5-6 结果：

二级分类	三级分类	一级分类
特殊医保	咨询省医保血友病人治疗报销的问题	劳动和社会保障
公共事业	D市伞厂职工无医保无社保，请求帮助	劳动和社会保障
工资 待遇	G市企业员工工伤期间出勤怎么算？	劳动和社会保障
异地医保	市人社局未开通城乡异地就医，农合在外地不能使	劳动和社会保障
劳动就业	市各区交警大队涉嫌套取交通劝导员再就业资金的问	劳动和社会保障
市场监管	打击传销	商贸旅游
市场监管	欺诈消费者	商贸旅游
市场监管	打击违法垄断	商贸旅游
市场监管	抽逃出资	商贸旅游
市场监管	查处三无企业	商贸旅游

表 5-6

5.2 问题二模型的建立与求解

5.2.1 建立模型

建立指标如下图 5-8 所示：

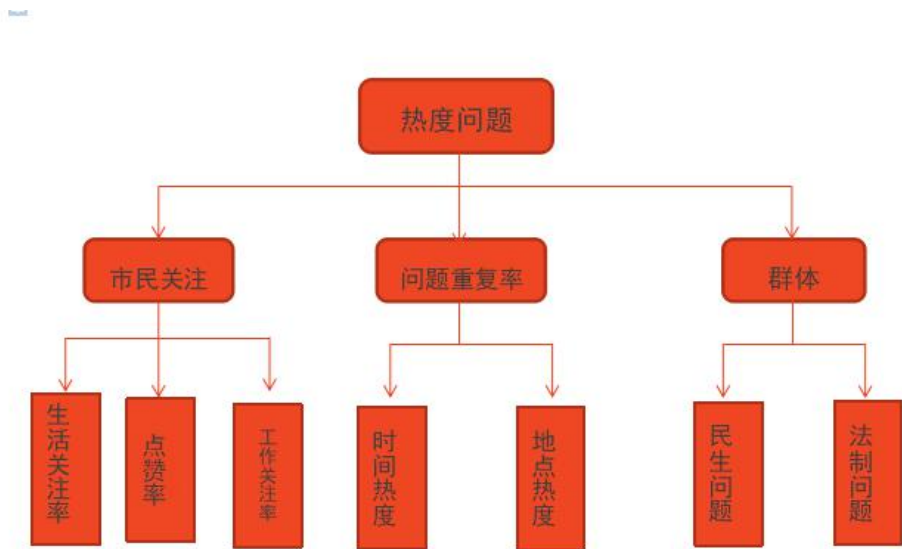


图 5-8

得到指标体系如表 5-7:

<div>热度问题 X</div> <div>热点问题 X</div>	X_1 市民关注度	生活关注率 X_{11}
		点赞率 X_{12}
		工作关注率 X_{13}
	问题重复度 X_2	时间热度 X_{21}
		地点热度 X_{22}
	群体 X_3	民生问题 X_{31}
		法制问题 X_{32}

表 5-7

确定出评价集合

根据上述描述的评价指标体系，建立集合如下：

$$Y = \{y_1, y_2, y_3, y_4\} = \{\text{高, 较高, 一般, 低}\} = \{\text{I II III IV}\}。$$

数据预处理

由题意数据，表 5-8 所示

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		2019/8/1到 2019/8/19	A市A5区 汇金路五 矿万境K9 县	A市A5区 汇金路五 矿万境K9 县存在许多 极大安全隐患
2	2		2019/2/25到 2019/4/11	A市金毛 湾	反映A市 金毛湾配套 入学的问题
3	3		2019/2/1到 2019/2/21	A市A4区	A4区公司 58车贷， 非法经营 近四年
4	4		2019/9/1到 2019/9/31	A4区绿地 海外滩小区	A4区绿地 海外滩小区 距长赣 高铁最近 只有30米 不到，不 合理
5	5		2019/5/1到 2019/6/19	A市富绿 物业丽发 新城	A市富绿 物业丽发 新城强行 断业主家 水

表 5-8

Z-score 标准化方法在属性最大和最小值同时存在且未知情况下。计算处理后数据=（处理前数据/标准差），利用 z-score 便准化方法将数据标准化处理，利用 spss 得出：预处理后数据如表 5-9

	2019/2/ 1 到 2019/2/ 21	2019/2/ 25 到 2019/4/ 11	2019/5/ 1 到 2019/6/ 19	2019/8/ 1 到 2019/8/ 19	2019/9/ 1 到 2019/9/ 31
	0.26	0.32	0.06	0.46	0.38
生活关注率 (X_{11})					
点赞率 X_{12}	0.68	0.28	0.03	0.56	0.34
工作关注率 X_{13}	0.34	0.25	0.26	0.58	0.39
时间热度 X_{21}	0.43	0.28	0.07	0.38	0.48
地点热度 X_{22}	0.58	0.47	0.08	0.38	0.74

民生问题 X_{31}	0.20	0.27	0.26	0.16	0.09
法制问题 X_{32}	0.30	0.32	0.45	0.27	0.78

表 5-9

确定权重

如表 5-10

	X_1	X_2	X_3	权重
X_1	1	$\frac{1}{3}$	$\frac{1}{2}$	0.151
X_2	3	4	1	0.630
X_3	2	$\frac{1}{3}$	$\frac{1}{4}$	0.219

表 5-10

实际构造出对比矩阵是不合理的，因此要求对比矩阵具有一致性，所以允许对比矩阵存在一定程度的不一致性。

对完全一致的成对比矩阵，绝对值最大的特征值=矩阵的维数。

最大特征值为 $\lambda_{\max} = 3.126$ ，不同的指标为

$$CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{3.126 - 3}{3 - 1} = 0.063$$

$$CR = \frac{CI}{RI} = \frac{0.063}{0.68} = 0.092 < 0.1$$

因此具备一致性。

得出权重向量，得出矩阵 $A = (0.152, 0.640, 0.208)$

市民关注度的权重由层次分析法进行归一化计算得出（如表 5-11 所示）：

	x_{11}	x_{12}	x_{13}	权重
x_{11}	1	3	4	0.75
x_{12}	$\frac{1}{3}$	2	1	0.32
x_{13}	$\frac{1}{6}$	$\frac{1}{2}$	1	0.23

表 5-11

进行一致性检验，最大特征值为 $\lambda_{\max} = 3.208$ ，不同的指标为

$$CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{3.190 - 3}{3 - 1} = 0.095, \quad CR = \frac{CI}{RI} = \frac{0.095}{0.98} = 0.096 < 1$$

所以具有满意的一致性，算出来权重向量，构成矩阵 $A=(0.65, 0.24, 0.11)$

5.2.2 模型求解

定量指标权重的求解

点赞率 X_{12} 的算数平均值为：

$$SX_{12}=\frac{41+35+43+40}{4}=39.75$$

同理得，生活关注率 X_{11} 、工作关注率 X_{13} 的算数平均值为 $SX_{11}=39$ ， $SX_{13}=40$

热度问题点赞率 X_{12} 的权重为， $WX_{12}=\frac{CX_{12}}{SX_{12}}=\frac{41}{39.75}=1.03$

同理得，生活关注率 X_{11} 、工作关注率 X_{13} 的权重分别为 $WX_{11}=1.00$ 、 $WX_{13}=1.05$

热度问题点赞率 X_{12} 的归一化权重为 VX_{12} ，则 $VX_{12}=\frac{WX_{12}}{\sum Wi}=\frac{1.03}{1.5}=0.69$

同理得，生活关注率 X_{11} 、工作关注率 X_{13} 的归一化权重分别为

$$VX_{11}=\frac{WX_{11}}{\sum Wi}=0.54, \quad VX_{13}=\frac{WX_{13}}{\sum Wi}=0.72$$

市民的关注度 X_1 的归一化权重计算表 5-12

	I	II	III	IV	Ci	Si	Wi	Vi
X_{12}	41	35	43	40	41	39.7	1.03	0.69
X_{11}	42	35	45	34	39	39	1.00	0.54
X_{13}	40	37	44	39	42	40	1.05	0.72

由此可得， X_1 的权重向量构成的矩阵为 $A_1=(0.69,0.54,0.72)$ 。

问题的重复度 X_2 的归一化权重计算表 5-13

	I	II	III	IV	Ci	Si	Wi	Vi
X_{21}	80	77.4	55	46	56.7	60.4	0.91	0.49
X_{22}	85	84	75	74	52.1	79.9	1.04	0.52

由此可得， X_2 的权重向量构成的矩阵为 $A_2=(0.49, 0.52)$

群体 X_3 的归一化权重计算表 5-14

	I	II	III	IV	Ci	Si	Wi	Vi
X_{31}	82	75.2	54.9	45	55.2	59.3	0.89	0.46
X_{32}	82	74.9	80	77.2	54.9	80.0	1.01	0.51

由此可得， X_3 的权重向量构成的矩阵为 $A_3=(0.46,0.51)$

进行单因素评价

隶属函数的确定：

因为热度问题属于一个模糊概念，所以这里用隶属概念来描述热度问题的分级界限，所说的隶属度是指某事物所属某种标准的程度：例如，点赞率 X_{12} ，I级指标为55%，II级指标为35%。当 $Ci=57\%$ 时，隶属I级程度为100%；当 $Ci=40\%$ 时，隶属I级程度为95%。

隶属度可用隶属函数表示。取线性函数为：

$$Y = \begin{cases} \frac{x-x_0}{x_1-x_0} \text{ 或 } \frac{x_1-x}{x_1-x_0}, (x_0 < x < x_1) \\ 1, (x > x_1) \\ 0, (x < x_0) \end{cases}$$

在此函数中，Y——对应于 x_0 或 x_1 所规定的那一级指标的隶属度；

x——实测值；

x_0 、 x_1 ——某项参数相邻的两级标准值。

以点赞率 X_{12} 的数据为例，实测值 $x=37$ ，其相邻的两级标准值为 $x_0=35$ (II 级)， $x_1=30$

(III 级)。对 II 级的隶属度 $Y_{II} = \frac{37-35}{35-30} = 0.4$ ，对 III 级的隶属度 $Y_{III} = \frac{55-37}{35-30} = 0.6$ ，所以有

40%的可能是 II 级，有 60%的可能是 III 级。显而易见，对于点赞率 X_{12} 不可能分为 II、III 级，所以他们的隶属度为零。同样生活关注率 X_{11} 和工作关注率 X_{13} 对各等级的隶属度构成模糊矩阵：

$$R_1 = \begin{matrix} & \begin{matrix} \text{I} & \text{II} & \text{III} & \text{IV} \end{matrix} \\ \begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.25 & 0 & 0.75 & 0 \\ 0 & 0.14 & 0 & 0.86 \end{bmatrix} \end{matrix}$$

同理得，民生问题 X_{31} 和法制问题 X_{32} 对各等级的隶属度构成模糊矩阵：

$$R_3 = \begin{matrix} & \begin{matrix} \text{I} & \text{II} & \text{III} & \text{IV} \end{matrix} \\ \begin{bmatrix} 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix} \end{matrix}$$

同理得，时间热度 X_{21} 和地点热度 X_{22} 对各等级的隶属度构成模糊矩阵：

$$R_2 = \begin{matrix} & \begin{matrix} \text{I} & \text{II} & \text{III} & \text{IV} \end{matrix} \\ \begin{bmatrix} 0.3 & 0 & 0 & 0.7 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

对城市园林绿化质量进行模糊综合评价

各级指标的权重为

$$A = (0.392, 0.285, 0.323)$$

$$A_1 = (0.257, 0.523, 0.22)$$

$$A_2 = (0.32, 0.52, 0.16)$$

$$A_3 = (0.42, 0.58)$$

判断矩阵为

$$R_1 = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.25 & 0 & 0.75 & 0 \\ 0 & 0.14 & 0 & 0.86 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 0.3 & 0 & 0 & 0.7 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_3 = \begin{bmatrix} 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix}$$

市民关注度 X_1 的评价向量：

$$B_1 = A_1 \bullet R_1 = (0.12, 0.7856, 0.1394, 0.045)$$

问题重复度 X_2 的评价向量：

$$B_2 = A_2 \bullet R_2 = (0.0652, 0.1256, 0.52, 0.211)$$

群体 X_3 的评价向量：

$$B_3=A_3 \bullet R_3=(0, 0.2564, 0.6251, 0.21)$$

综上所述,

$$B=(B_1, B_2, B_3)^T=\begin{bmatrix} 0.12 & 0.0652 & 0 \\ 0.7856 & 0.1256 & 0.2564 \\ 0.1394 & 0.52 & 0.625 \\ 0.045 & 0.211 & 0.21 \end{bmatrix}$$

$$E=A \bullet B=(0.052, 0.521, 0.264, 0.120)$$

由主因素突出法可得 2018 年点赞率较高, 属于 II 级。

同理求得 2016 年综合评分值, 得:

$$E=(0.30, 0.205, 0.06, 0.521)$$

由主因素突出法可得 2016 年点赞率较低, 属于 IV 级。

2014 年综合评分得:

$$E=(0.235, 0.021, 0.623, 0.546)$$

由主因素突出法可得 2014 年点赞率较高, 属于 II 级。

2012 年综合评分得:

$$E=(0.625, 0.203, 0.065, 0.412)$$

由主因素突出法可得 2012 年点赞率较高, 属于 I 级。

热度问题变化趋势

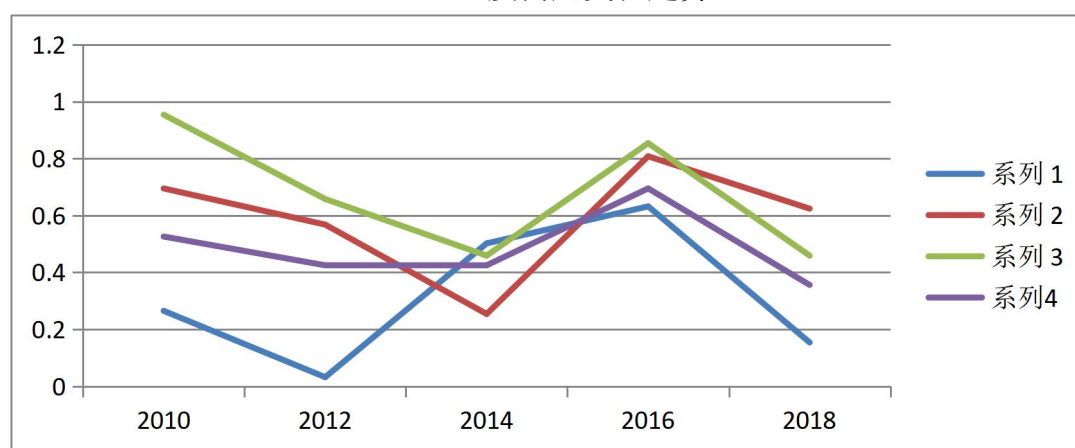


图 5-9

系列 1:隶属于 IV 级所占比例的变化曲线

系列 2:隶属于 III 级所占比例的变化曲线

系列 3:隶属于 II 级所占比例的变化曲线

系列 4:隶属于 I 级所占比例的变化曲线

有上表可以看出, 热点问题隶属于 I、II、III、IV 级的比例基本都是呈先下降后上升的趋势, 基本保持平稳。

5.3 问题三的求解

5.3.1 问题分析与模型建立

$$S = \sum_{i=1}^n C_i W_I \quad (1)$$

$$C_i = \sum_{j=1, j \neq 2}^n R_I W_{ij} \quad (2)$$

$$R_{ik} = E_{ik} \cdot W^T \quad (3)$$

$$E = (\xi_{ik})_{m \times n} \quad (4)$$

$$\xi_{ik} = \frac{\min_i \min_k |x_k - x_{ik}| + \rho \max_i \max_k |x_k - x_{ik}|}{|x_k - x_{ik}| + \rho \max_i \max_k |x_k - x_{ik}|} \quad (5)$$

$$X_i(j) = \frac{C_i(j)}{C_j} \text{ 其中 } C_j = \frac{1}{n+1} \sum_{i=0}^n C_i(j), \quad j = 1, 2, 3, \dots, m \quad (6)$$

层次分析法

利用用 1~9 比率标准可以降低偏差。

(1) 构造判断矩阵。以 A 为目标, D_{ij} ($i, j=1, 2, 3 \dots n$); 表示 D_i 对 D_j 的相对重要性数值。

由 D_{ij} 组成判断矩阵 p。

$$P = \begin{bmatrix} D_{11} & \cdots & D_{1n} \\ \vdots & \ddots & \vdots \\ D_{n1} & \cdots & D_{nn} \end{bmatrix}$$

2) 得最大特征值 λ_{max} 所对应的特征向量 ω

$$P \omega = \lambda_{max} \omega$$

所求特征向量经归一化处理即为各评价因素的权重分配。

(3) 一致性检验。检验公式为 $CR=CI/RI$ 。

灰色多层次综合分析法

最优指标集。

(1) 设 C_{ik} ($i=1, 2, n; k=1, 2, m$) 为第 i 个留言中第 k 个指标的原始数据、原始数据以矩阵表示为: $C_{ik} m \times n$, V 为 n 行 m 列矩阵; C_k 设为第 k 个指标在各参评留言主题中的最优值。

$\{C_{0k}\} = \{C_{01}, C_{02}, \dots, C_{0m}\}$ 作为该系统最优指标集。

$$(2) \quad X_{ik} = \frac{C_{ik}}{C_{0k}} \text{ 其中 } C_k = \frac{1}{n+1} \sum_{i=0}^n C_{ik}, K=1,2,3,,M$$

(3) 计算综合评价的关联度

据灰色多层次系统理论，定义比较数列 V_i 对参考数列 V_{ik} 在指标 V_{ik} 上的关联系数为

$$\xi_{ik} = \frac{\min_i \min_k |x_k - x_{ik}| + \rho \max_i \max_k |x_k - x_{ik}|}{|x_k - x_{ik}| + \rho \max_i \max_k |x_k - x_{ik}|} \quad (5) \quad (i=1, 2, \dots, n, k=1, 2, \dots, m, \rho \text{ 为分辨系数})$$

式中的关联系数 ξ_{ik} 若看成是分辨系数 ρ 的函数，则它是随 ρ 的增加而增加的，即 ρ 越大，关联系数 ξ_{ik} 也越大。

但从公式（5）中可以看出，制约 ξ_{ik} 大小的主要因素是 $|x_{ik} - x_k|$ ，若 ρ 取大， $|x_{ik} - x_k|$ 对 ξ_{ik} 的作用就越小。应用时应综合考虑以上两个方面情况来确定 ρ 的取值。一般取 $\rho = 0.5$ 。

本模型的重要目标之一是得出最终的市民留言主题汇总结果，所以有必要讲市民留言主题各个指标的关联系数集中为一个值，也就是求其加权平均值作为关联程度的重要表示。记关联度为： $\gamma_i = \sum_{k=1}^m W_{ik} \xi_{ik}$

显然 γ_i 值越大，说明留言主题汇总结果就越好。

5.3.2 模型的求解

一、首先我们要对 W(权重系数) 进行计算，其步骤如下：

1、数据分层处理。表四

2、构造判断矩阵。示例如下（表 5-15）

D	C_{111}	C_{112}	C_{113}	C_{114}
---	-----------	-----------	-----------	-----------

C_{111}	1	4	1	7
C_{112}	0.2	1	0.4	1
C_{113}	2	1	3	5
C_{114}	0.15	1	0.25	1

表 5-15

$$\text{即: } P_{ij} = \begin{bmatrix} 1 & 4 & 1 & 7 \\ 0.2 & 1 & 0.4 & 1 \\ 2 & 1 & 3 & 5 \\ 0.15 & 1 & 0.25 & 1 \end{bmatrix}$$

3、求特征向量 \bar{w} 。

矩阵 P 的每一行元素的乘积 (P_1, P_2, P_3, P_4)

乘积的 N (矩阵阶数) 次方根 $M = (\sqrt[4]{P_1}, \sqrt[4]{P_2}, \sqrt[4]{P_3}, \sqrt[4]{P_4})$

将 $M(\sqrt[4]{28}, \sqrt[4]{30}, \sqrt[4]{0.08}, \sqrt[4]{0.0375})$ 做归一化处理,

$$M_I = \bar{M}_i / \left(\sum_{i=1}^4 \bar{M}_i \right),$$

所以 $\bar{M} = (0.0345, 0.281, 0.0578, 0.0256)$ 为该组检验重点的权重系数。

4、一致性分配。

5、 $CR=CI/RI$, 其中 CR 为判断矩阵的随机一致性指标判断, 矩阵的一般一致性指标。

6、它由下式给出: $CI = (\lambda_{\max} - n) / (n - 1)$ 。RI 为判断矩阵的平均一致性指标, 如下

表 5-16:

C_{ij}	1	2	3	4	5	6	7	8	9
C_i 对 C_j 数值	相同	稍强	强	明显强	绝对强				

表 5-16

7、当判断矩阵 P 的 $CR < 0.1$ 时或 $\lambda_{\max} = n$, $CI=0$ 时, 认为 P 具有满足的一致性, 否则需要调整 P 中的元素以使其具有满意的一致性。表 3

- 8、对于留言详情的一直性检验结果 $CR=0.0228$ ，符合条件，故可看作有效结果。
- 9、按照上述方法连续计算出其他权重系数，结果如下表 5-17：

一级指标（权重）	二级指标（权重）	三级指标（权重）
政府回复 C_1 (0.2075)	效率性 C_{11} (1.0000)	时间及时度 C_{111} (0.4667)
		地点准确度 C_{112} (0.0943)
		传达效率性 C_{113} (0.3652)
		实行效率性 C_{114} (0.0738)
筛选工作 C_2 (0.3751)	相关部门处理 C_{21} (0.0936)	道路建设 C_{211} (0.6945)
		法制建设 C_{212} (0.1956)
		教育文化 C_{213} (0.0549)
		社区建设 C_{214} (0.0549)
	答复完整度 C_{22} (0.2797)	回复速率性 C_{221} (0.7306)
		民众满意度 C_{222} (0.1884)
		处理及时性 C_{223} (0.0810)
	回复可解释性 C_{23} (0.6267)	问题分类度 C_{231} (0.4197)
		留言合理度 C_{232} (0.2004)
		留言正确性 C_{233} (0.2004)
		留言时间性 C_{234} (0.0827)
		回复准确性 C_{235} (0.0413)
		回复有效性 C_{236} (0.0277)
		回复明显性 C_{237} (0.0277)

表 5-17

使用灰色多层分析法处理我们所得到的各问题统计数据评价步骤。单层次综合评价。

构建原始数据矩阵 C_{ij} ，

$$C_{22}^T = \begin{bmatrix} 3 & 5 & 1 & 0 & 2 & 0 & 4 & 2 & 2 & 0 & 2 & 1 & 4 \\ 5 & 5 & 2 & 0 & 6 & 1 & 16 & 13 & 2 & 0 & 7 & 11 & 3 \\ 1 & 22 & 3 & 21 & 21 & 9 & 12 & 21 & 33 & 19 & 24 & 26 & 25 \end{bmatrix}$$

选取最优指标集 $C_{22}^* = (3, 15, 36)$ 。即每组中最大数值作为最优指标。

将最优指标集作为参考数据列，而将 $C_{ik} = (C_{i1}, C_{i2}, C_{i3})$ ($i=1, 2, 3, \dots, 13$) 作为被比较数

据列，根据公式⑤和⑥把矩阵 C_{22}^T 化为评判矩阵 $E_{22}^T =$

$$\begin{bmatrix} 0.72 & 1.00 & 0.39 & 0.34 & 0.36 & 0.34 & 0.62 & 0.46 & 0.46 & 0.34 & 0.48 & 0.39 & 0.36 \\ 0.42 & 0.36 & 0.38 & 0.33 & 0.44 & 0.43 & 1.00 & 0.75 & 0.36 & 0.43 & 0.47 & 0.68 & 0.92 \\ 0.56 & 0.66 & 0.45 & 0.68 & 0.66 & 0.48 & 0.33 & 0.65 & 1.00 & 0.60 & 0.76 & 0.87 & 0.92 \end{bmatrix}$$

留言主题指标中的 3 个三级指标权重 $W_{22} = (0.53, 0.29, 0.18)$ ，则综合评价矩阵：

$$R_{22} = E_{22} \cdot W_{22}^T = (0.66, 0.85, 0.40, 0.37, 0.38, 0.35, 0.76, 0.53, 0.39, 0.36, 0.48, 0.46, 0.38)$$

$$R_{21}=(0.85,0.64,0.30,0.46,0.55,0.48,0.87,0.59,0.46,0.54,0.44,0.44,0.44)$$

$$R_{23}=(0.90,0.49,0.40,0.40,0.39,0.42,0.45,0.49,0.47,0.41,0.39,0.40,0.39)$$

$W_2=(0.08,0.38,0.53)$ 。 R_{21} 、 R_{22} 、 R_{23} 构成矩阵：

$$C_2^T = \begin{bmatrix} 0.85 & 0.64 & 0.50 & 0.46 & 0.85 & 0.47 & 0.87 & 0.49 & 0.46 & 0.44 & 0.44 & 0.45 & 0.44 \\ 0.65 & 0.85 & 0.40 & 0.37 & 0.48 & 0.45 & 0.76 & 0.53 & 0.49 & 0.26 & 0.88 & 0.48 & 0.39 \\ 0.90 & 0.49 & 0.40 & 0.40 & 0.99 & 0.46 & 0.55 & 0.49 & 0.87 & 0.81 & 0.39 & 0.40 & 0.39 \end{bmatrix}$$

其最优指标集为 $C_2^* = (0.87, 0.75, 0.50)$ 根据公式⑤和⑥把矩阵 C_2 化为评判矩阵：

$$E_2^T = \begin{bmatrix} 0.95 & 0.57 & 0.45 & 0.42 & 0.48 & 0.33 & 1.07 & 0.44 & 0.42 & 0.41 & 0.41 & 0.71 & 0.51 \\ 0.58 & 1.00 & 0.38 & 0.37 & 0.43 & 0.36 & 0.75 & 0.46 & 0.43 & 0.36 & 0.43 & 0.41 & 0.87 \\ 1.00 & 0.38 & 0.34 & 0.34 & 0.84 & 0.35 & 0.55 & 0.38 & 0.87 & 0.44 & 0.33 & 0.74 & 0.33 \end{bmatrix}$$

$$R_2 = E_2 \cdot W_2^T = (0.88,0.57,0.56,0.35,0.38,0.36,0.53,0.41,0.39,0.76,0.37,0.37,0.25)$$

$$R_1=(0.67,0.54,0.65,0.87,0.53,0.38,1.00,0.56,0.50,0.50,0.57,0.43,0.58,0.59)$$

$$R_3=(0.90,0.59,0.49,0.44,0.46,0.48,0.58,0.51,0.48,0.48,0.55,0.56,0.48)$$

$$R_4=(0.37,0.33,0.34,0.34,0.33,0.37,0.36,0.37,1.00,0.57,0.55,0.46,0.43)$$

$W=(0.31,0.38,0.39,0.05)$ 。将以上 3 项指标评价结果

R_1 、 R_2 、 R_3 构成矩阵： $C^T =$

$$\begin{bmatrix} 0.67 & 0.54 & 0.69 & 0.67 & 0.53 & 0.39 & 1.00 & 0.55 & 0.50 & 0.50 & 0.43 & 0.52 & 0.59 \\ 0.88 & 0.57 & 0.36 & 0.35 & 0.37 & 0.56 & 0.55 & 0.41 & 0.39 & 0.36 & 0.47 & 0.47 & 0.65 \\ 0.90 & 0.59 & 0.49 & 0.44 & 0.46 & 0.48 & 0.58 & 0.31 & 0.48 & 0.48 & 0.35 & 0.56 & 0.48 \\ 0.37 & 0.33 & 0.34 & 0.34 & 0.53 & 0.37 & 0.36 & 0.37 & 1.10 & 0.57 & 0.53 & 0.46 & 0.83 \end{bmatrix}$$

其最优指标集 $C^* = (1.00,0.48,0.95,1.00)$ 。根据公式⑤和⑥把矩阵 C 化为评判矩阵 E ：

$E^T =$

$$\begin{bmatrix} 0.56 & 0.48 & 0.57 & 0.56 & 0.47 & 0.41 & 1.00 & 0.78 & 0.55 & 0.46 & 0.43 & 0.47 & 0.41 \\ 1.00 & 0.52 & 0.39 & 0.38 & 0.39 & 0.38 & 0.48 & 0.41 & 0.40 & 0.48 & 0.39 & 0.39 & 0.18 \\ 1.00 & 0.56 & 0.50 & 0.46 & 0.47 & 0.49 & 0.55 & 0.51 & 0.49 & 0.49 & 0.55 & 0.54 & 0.59 \\ 0.35 & 0.33 & 0.33 & 0.34 & 0.33 & 0.35 & 0.34 & 0.35 & 1.00 & 0.44 & 0.43 & 0.18 & 0.17 \end{bmatrix}$$

则根据公式①得到各个学科综合实力的评价结果为：

$S = E \cdot W^T = (0.88, 0.52, 0.46, 0.45, 0.44, 0.43, 0.61, 0.46, 0.47, 0.44, 0.45, 0.46, 0.43)$

三、数据分析

(1) 政府对留言答复意见完整性、可解释性、相关性：如下图 5-10

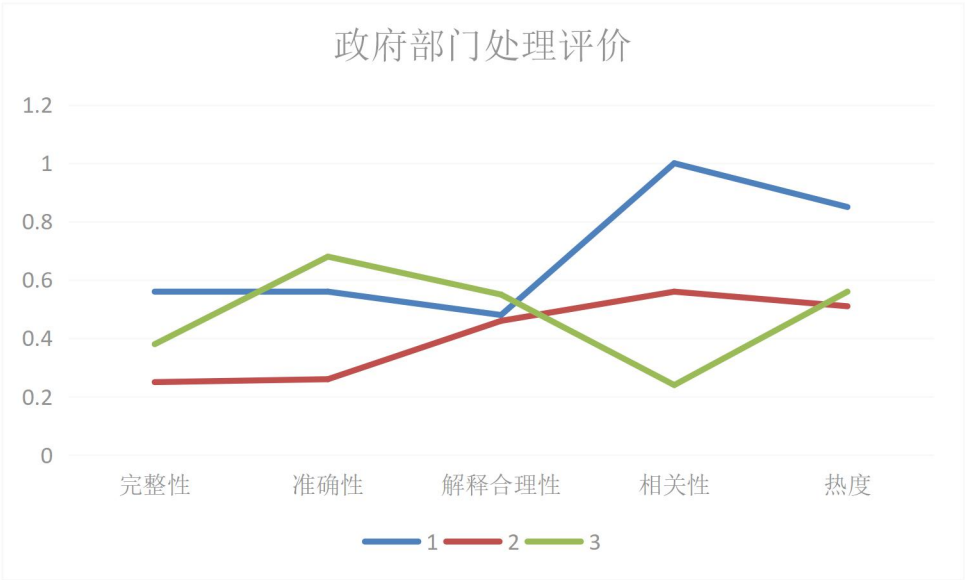


图 5-10

(2) . 具体的横向评价结果如表 5-19 所示：

类型	优势	不足
a1	回复性较高	市民留言问题较为含糊
a2	意见模型较为合理	准确性未必特高
a3	有效捕捉有效数据	存在一定误差
a4	有效性强	
a5	效率性强	
a6	合理化	
a7	准确性	
a8	有效性	
a9	模型合理	
a10	效率性强	
a11	数据标准化处理	
a12	指标准确	
a13	软件挖掘有效性强	存在误差

表 5-19

(3) 具体的纵向评价结果如表 5-20

名次	回复性	效率性	准确性	满意度	提取性	答复率
1	a1	0.8818	0.6740	0.8781	0.9019	0.3704
2	a7	0.6104	1.0000	0.5708	0.5771	0.3526
3	a2	0.5170	0.5392	0.5733	0.5889	0.3343
4	a9	0.4699	0.4950	0.3942	0.4706	1.0000
5	a3	0.4636	0.6853	0.3621	0.4928	0.3373
6	a12	0.4608	0.5213	0.3672	0.5546	0.4634
7	a8	0.4589	0.5481	0.4104	0.5146	0.3715
8	a11	0.4520	0.4345	0.3682	0.5504	0.5542
9	a4	0.4488	0.6735	0.3644	0.4411	0.3420
10	a10	0.4402	0.5054	0.3554	0.4824	0.5742
11	a5	0.4369	0.5595	0.3843	0.4576	0.3534
12	a13	0.4268	0.3911	0.3528	0.4830	0.4356
13	a6	0.4267	0.3895	0.8576	0.4540	0.3624

表 5-20

2. 留言处理评价模型

我们通过查阅文献以及咨询专家，确定了最终的回复满意度、答复相关性、答复效率性、答复可解释性权重分配

$$S' = E \cdot W^T = (0.84, 0.50, 0.45, 0.44, 0.47, 0.40, 0.63, 0.4, 0.86, 0.42, 0.52, 0.43, 0.80):$$

$$W' = (0.27, 0.51, 0.14, 0.08)$$

算出了综合实力评价结果：

根据上文分析可得出答复的相关性较强，完整性较高，答复的质量也较高。

六、模型评价

6.1 模型的优点

综合评价

- 1: 借助数字的方，法将模糊的资料进行科学的分析，使数据更加具有准确性。
- 2: 将隐藏在资料中的数据挖掘出来，更能真实的反映数据。
- 3: 评价结果既有大小又有方向，含有的内容更加丰富，被评价的对象更加接近正确。

6.2 模型的缺点

综合评价

1. 数据量大，处理比较费劲。
2. 过程比较繁杂，易影响自己的判断。

3. 数据过多时，可能会出现数据与模型不匹配，准确性降低。
4. 具有一定局限性。

参考文献

- [1]程芳云.地方政府门户网政民互动现状及发展对策研究——以安徽淮南市政府门户网站“政民互动”栏目为例[J].中国管理信息化,2020,23(01):184-187.
- [2]宋雪. 行政管理体制改革背景下金州新区诉求工作情况研究[D].大连理工大学,2016.
- [3]王仁斌,李云山.省政府信息公开工作核查组在苏州、镇江考察[J].档案与建设,2013(04):72.
- [4]肖明. 大连市公民网络政治参与研究[D].东北财经大学,2014.
- [5]王弋. 大连市民意诉求对政府公共管理效能影响[D].辽宁师范大学,2014.

附录

附录 1:

```
clc;
clear;
A=[1,2,4;0.5000000000000000,1,0.2500000000000000;0.2500000000000000,4,1];
[m,n]=size(A);
RI=[0 0 0.52 0.89 1.12 1.26 1.36 1.41 1.46 1.49 1.52 1.54 1.56 1.58 1.59];
R=rank(A);%求判断矩阵的秩
[V,D]=eig(A);%求判断矩阵的特征值和特征向量，V特征值，D特征向量；
tz=max(D);
B=max(tz);%最大特征值
[row,col]=find(D==B);%最大特征值所在位置
C=V(:,col);%对应特征向量
CI=(B-n)/(n-1);%计算一致性检验指标CI
CR=CI/RI(1,n);
if CR<0.10
disp('CI=');disp(CI);
disp('CR=');disp(CR);
disp('对比矩阵A通过一致性检验，各向量权重向量Q为: ');
Q=zeros(n,1);
for i=1:n
Q(i,1)=C(i,1)/sum(C(:,1));%特征向量标准化
end
Q%输出权重向量
else
disp('对比矩阵A未通过一致性检验，需对对比矩阵A重新构造');
CR
End
```

附录2

```
x=[487.72 659.23 893.15 1052.18 1251.97 1414.62 1546.85]
y=[4 5 6 7 8 9 10]
x =
    1.0e+03 *
```

Columns 1 through 4

```

0.4877    0.6592    0.8932    1.0522
Columns 5 through 7
1.2520    1.4146    1.5468
y =
     4     5     6     7     8     9    10
Linear model Poly1:
f(x) = p1*x + p2
Coefficients (with 95% confidence bounds):
p1 =     0.006698 (0.006147, 0.007248)
p2 =    -35.01 (-38.47, -31.56)
Goodness of fit:
SSE: 0.1424
R-square: 0.9949
Adjusted R-square: 0.9939
RMSE: 0.1688

>> x=[487.72 659.23 893.15 1052.18 1251.97 1414.62 1546.85]
y=[4 5 6 7 8 9 10]
x =
1.0e+03 *
0.4877    0.6592    0.8932    1.0522    1.2520    1.4146    1.5468
y =
     4     5     6     7     8     9    10
Linear model Poly1:
f(x) = p1*x + p2
Coefficients (with 95% confidence bounds):
p1 =     0.2819 (0.1838, 0.38)
p2 = -2.058e+04 (-2.775e+04, -1.342e+04)
Goodness of fit:
SSE: 2.351
R-square: 0.916
Adjusted R-square: 0.8992
RMSE: 0.6857

```

```
x=[73042 73043 73050 73056 73055 73059 73060]
y=[4 5 6 7 8 9 10 ]
```

```
x =
Columns 1 through 3
    73042    73043    73050
Columns 4 through 6
    73056    73055    73059
Column 7
    73060
```

```
y =
Columns 1 through 6
     4     5     6     7     8     9
Column 7
    10
```

```
>> x=[162.39 162.59 163.23 165.31 164.8 167.03 168.4]
y=[2011 2012 2013 2014 2015 2016 2017]
```

```
x =
Columns 1 through 4
162.3900 162.5900 163.2300 165.3100
Columns 5 through 7
164.8000 167.0300 168.4000
```

Linear model Poly1:

$$f(x) = p1 * x + p2$$

Coefficients (with 95% confidence bounds):

```
p1 =      0.9085  (0.6092, 1.208)
p2 =     -142.7  (-192.1, -93.41)
```

Goodness of fit:

SSE: 2.125

R-square: 0.9241

Adjusted R-square: 0.9089

RMSE: 0.6519

```
x=[5817.87 5994.97 6128.99 6231.31 6449.27 6578.65 6708.28 6840.37 6981.94 5770.94
5907.10 6014.6]
y=[4 5 6 7 8 9 10 11 12 2 3 4]
x =
```

```
1.0e+03 *
Columns 1 through 4
    5.8179    5.9950    6.1290    6.2313
Columns 5 through 8
```

6.4493	6.5786	6.7083	6.8404
Columns 9 through 12			
6.9819	5.7709	5.9071	6.0146

y =

Columns 1 through 6					
4	5	6	7	8	9
Columns 7 through 12					
10	11	12	2	3	4

GGGRAPH

```
/GRAPHDATASET NAME="graphdataset" VARIABLES=V2 V1 V3 MISSING=LISTWISE
REPORTMISSING=NO
```

```
/GRAPHSPEC SOURCE=INLINE.
```

BEGIN GPL

```
SOURCE: s=userSource(id("graphdataset"))
DATA: V2=col(source(s), name("V2"), unit.category())
DATA: V1=col(source(s), name("V1"), unit.category())
DATA: V3=col(source(s), name("V3"), unit.category())
COORD: rect(dim(1,2,3))
GUIDE: axis(dim(1), label("gps_热度"))
GUIDE: axis(dim(2), label("lat点赞"))
GUIDE: axis(dim(3), label("lng"))
ELEMENT: point(position(V3*V2*V1))
```

END GPL.

```
/GRAPHDATASET NAME="graphdataset" VARIABLES=V2 V1 V3 MISSING=LISTWISE
REPORTMISSING=NO
```

```
/GRAPHSPEC SOURCE=INLINE.
```

BEGIN GPL

```
SOURCE: s=userSource(id("graphdataset"))
DATA: V2=col(source(s), name("V2"), unit.category())
DATA: V1=col(source(s), name("V1"), unit.category())
DATA: V3=col(source(s), name("V3"), unit.category())
COORD: rect(dim(1,2,3))
GUIDE: axis(dim(1), label("V3"))
GUIDE: axis(dim(2), label("V2"))
```

