“智慧政务”中的文本挖掘应用

摘要：随着信息时代的来临，电子政务应运而生，加速了社会公共服务和体系的建立与完善。本文通过对问政平台上群众的留言文本数据进行处理和分析，拟实现以机器代替人工分类工作，以期提高问政平台的政务服务效率。首先对群众留言文本数据进行分词、去停用词以及文本向量化等预处理，利用高斯贝叶斯方法训练分类模型，实现了群众留言的机器分类。其次，我们通过 DBSCAN 聚类算法将群众留言进行聚类，成功定义并计算热度评价指标，衡量出每一留言问题类簇的热度值，成功挖掘出了群众留言的热点问题。最后，利用以 AHP 层次分析法、FCE 模糊综合评价法为理论研究方法，对群众留言回复的相关性、完整性、可解释性的三个评价指标分层排序，获取合适权重，构建 AHP 调查表和模糊综合评价问卷，定性、定量地分析以期构建合适的评价方案，希望为政府“智慧政务”工作提供恰当参考。

关键词：文本分类；机器学习；聚类分析；层析分析法；AHP-FCE 综合分析

Text Analysis Application in “Smart Government”

Abstract: With the advent of the information era, E-government emerges, which accelerates improvement of social public services and system establishment. By processing and analyzing the message text data by masses on the online platforms, this paper aims at replacing the manual classification work with machine, and improves the efficiency of the government service of the platform. First of all, the text data is preprocessed by word segmentation, stop words depletion and text vectorization. The classification model is trained by using the Gaussian Bayesian method to realize the machine classification of the crowd message. Secondly, we use DBSCAN clustering algorithm to cluster the public comments, we successfully define and calculate the heat evaluation index, which measures the heat value of each message problem cluster. We successfully reveal the hot issues of the public comments. Finally, by using AHP and the fuzzy comprehensive evaluation method as theoretical research methods, we rank the three evaluation indexes of relevance, integrity and interpretability of the response of the masses’ message, obtain the appropriate weight, construct AHP questionnaire and fuzzy comprehensive evaluation questionnaire, analyze them quantitatively and quantitatively in order to build the appropriate evaluation scheme, and hope to provide references for the establishment of “smart government”.

Key words: Text classification; machine learning; Cluster analysis; Analytic Hierarchy Process; AHP-FCE

目录

一、研究目标

党的十七大报告提出“健全政府职责体系，完善公共服务体系，推行电子政务，强化社会管理和公共服务”。“智慧政务”是电子政务的提升，它是利用物联网、云计算、移动互联网、人工智能、数据挖掘、知识管理等技术，提高政府办公、监管、服务、决策的智能化水平，形成高效、敏捷、便民的新型政府服务模式。随着信息时代的来临，电子政务应运而生，加速了社会公共服务和体系的建立与完善。然而，随着越来越多群众通过微博、市长信箱、阳光热线等网络渠道反馈民情民意，相关的文本数据数量不断攀升，以往单纯依靠人工处理的电子政务模式亟待更新，基于自然语言处理的智慧政务成为大势所趋。本文拟通过自然语言处理的方式，实现以机器代替人工，对收集到的群众反馈文本数据进行分类，挖掘其中的热点问题，并构建一套有效的评价指标体系，对有关部门给出的答复的相关性、完整性和可解释性进行评价。

二、分析方法与过程

2.1 总体流程

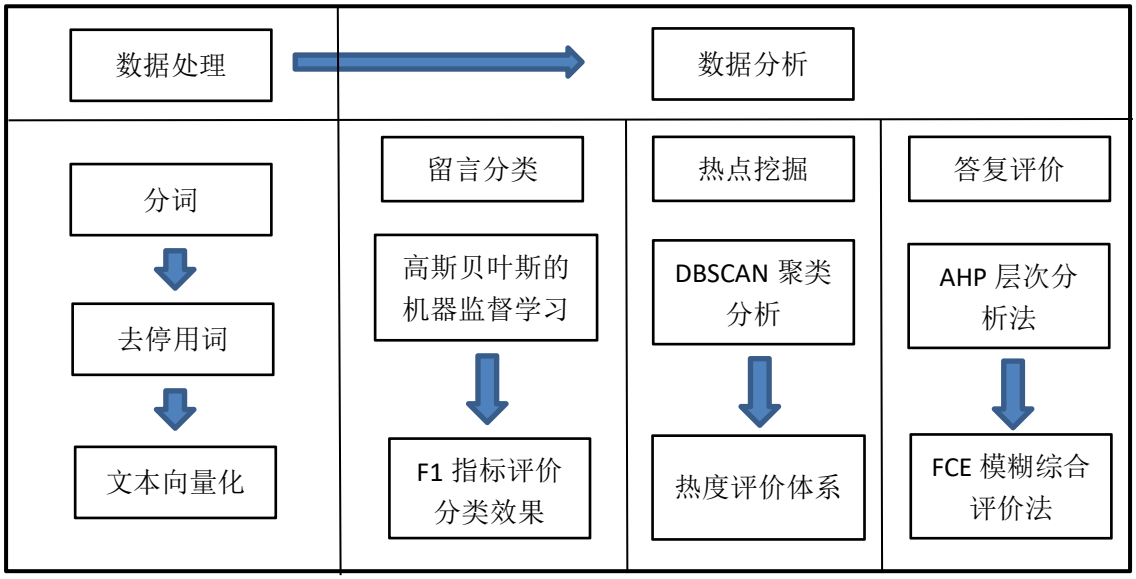


图 1 总体流程图

2.2 群众留言分类

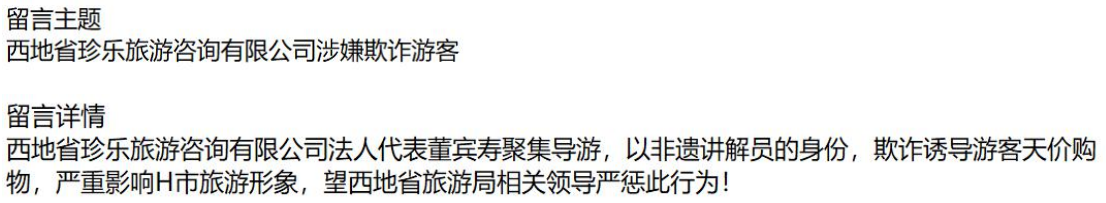
我们对本文中附件 2 的数据进行初步观察后，与预期一致，发现群众留言内容大多为有效文本信息，对于政务服务留言，政府部门不会也没有必要雇佣水军进行大量注水评论，群众也鲜少会在此类平台上留下无意义的言论或是“恶作

剧”。因此，我们可以省略对文本数据的去重、删除无意义留言等步骤，直接对文本数据进行分词操作。我们要解决的第一个问题是让计算机代替人工按照附件1给出的留言分类体系并对留言进行分类。附件2给出了留言编号、留言用户、留言主题、留言时间、留言详情以及该条留言所属的一级标签。因此，可以将此问题看作监督机器学习问题，通过建立训练模型，可以实现机器留言分类。

2.2.1 数据处理

在获取官网发布的文本数据后，第一步进行数据的预处理。因为群众的留言内容包含了很多无意义的虚词以及标点，并且由于语言表达的特点，同样语义的句子可能会以不同词语的组合形式重复出现。为了提高分类的准确度和质量，必须对文本数据进行分词、去停用词、提取特征词等数据处理操作，并且由于计算机无法读懂自然语言，我们还应当将处理后的文本转换成计算机能够理解和计算的语言，即文本向量化。

通过对题目以及文本数据的理解分析，我们认为仅保留留言主题和一级标签即可完成本题的目标。由于留言主题是群众想要反映的问题或建议的精炼和压缩，通常在20-30字以内。而留言详情是留言主题的扩展和细节描述，可以想象，人们在进行投诉建议时由于情绪上的不满，往往会加入很多带有情感色彩的形容词以及与事情本身无关的评价，比如下例所示：


留言主题
西地省珍乐旅游咨询有限公司涉嫌欺诈游客

留言详情
西地省珍乐旅游咨询有限公司法人代表董宾寿聚集导游，以非遗讲解员的身份，欺诈诱导游客天价购物，严重影响H市旅游形象，望西地省旅游局相关领导严惩此行为！

图2 原始数据举例

根据留言主题，我们可以很容易提取出该条留言的关键词“西地省”、“珍乐旅游咨询有限公司”、“欺诈游客”，整条文本中的无意义词非常少，该留言属于一级标签下的“商贸旅游”。而再看其对应的留言详情我们可以发现，留言详情的内容较长，并且包含了许多带情感色彩的细节描述和评价，比如“欺诈诱导游客天价购物，严重影响H市旅游形象，望西地省旅游局相关领导严惩此行为！”等等，这些文本内容包含的信息量非常少，如果将它们引入训练模型中，最后的效果会大打折扣，往往不理想。故我们直接保留留言主题，并对留言主题进行分词处理。

2.2.1.1 分词

目前的自然语言处理是建立在基于统计的语言模型，而这些模型是建立在词的基础之上，因为词是表达语义的最小单位。与西方语言不同，中文的词与词之间没有明显的间隔，我们进能够依靠标点符号来划分句子和段落。因此，为了得到词，首先需要对文本进行分词操作，才能进行后续的自然语言处理操作。

我们采用 python 中文分词组件“结巴”分词包对附件 2 中的留言主题文本数据进行分词。结巴分词器可以通过词典对中文文本进行分词、词性标注、关键词抽取等功能，并且支持用户自定义词典。因而分词效果取决于词典的质量，结巴分词采用 HMM（隐含马尔科夫模型）进行新词发现，以处理文本数据中出现的但在词典中未出现的词。

此外，我们也可以根据特定场景，使用用户自定义词典这一功能，将文本数据中出现频率很高，但词典中可能没有的词手动添加到词典中，以提高分词的质量。具体分词过程中，结巴分词首先通过对照词典生成句子的有向无环图，再根据词典寻找最短路径后对句子进行截取，对于不在词典中的词则使用 HMM 进行新词发现。结巴分词的总体工作流程如下图所示：

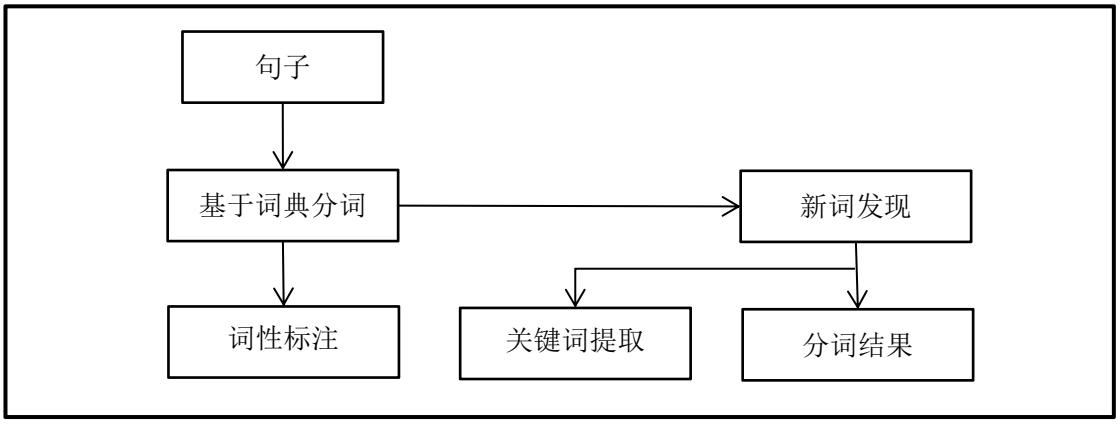


图 3 结巴分词工作流程

由于我们的目标仅在于分词，而不需要对文本进行词性标注，因此本步骤输出基于结巴分词的分词结果。

2.2.1.2 去停用词

经过分词操作后，我们得以将一个句子转换成若干词的合集。但从之前的示例可以看出，虽然留言主题已经相当精简和高度概括，但其中仍有部分文本数据存在，？！“”等标点符号，以及“吗”、“的”、“了”、“是”、“有”等无意义的虚词，这些词被称为停用词。这些词的存在不仅不能增加信息含量，反

而可能削弱其他实词的作用，导致结果精度不高。因此应当将这些停用词予以剔除。此外，从留言主题可以看到很多地名以字母和数字的组合替代，比如“A1市”、“G9 县”等等，这些词隐含了关键地名的信息，但在本题的分类目标中，地名对于分类是无贡献的，因此也应当剔除。

文本采用基于停用词表的文本停用词过滤方式，将分词结果于停用词表中的词语进行匹配，若匹配成功，则进行删除处理。结果示例如下：

去停用词前如下图所示：

E5县高滋村违法直销欺骗老人
G4县东门符家巷华东食品猪肉直销店噪音污染
G8县吉美磁化水直销团伙虚假宣传
I市通联违反劳动法扣网格直销员基本工资
推销中介以虚假广告的形式在A7县肆意诈骗
G市成邦地产中介拖欠购房尾款
投诉G7县二手房中介扰乱二手房市场
在西地省如何注销检验检测机构资质认定证书?
C1区农产品市场有无检验检疫证明啊?

去停用词后如下图所示：

高滋村 违法 直销 欺骗 老人
东门 符家巷 华东 食品 猪肉 直销店 噪音 污染
吉美 磁化水 直销 团伙 虚假 宣传
通联 违反 劳动法 扣 网格 直销员 基本工资
推销 中介 虚假 广告 形式 肆意 诈骗
成邦 地产 中介 拖欠 购房 尾款
投诉 二手房 中介 扰乱 二手房 市场
西地省 注销 检验 检测 机构 资质 认定 证书
农产品 市场 有无 检验 检疫 证明

2.2.1.3 文本向量化

计算机并不具有人类的智能，人在看到一个文本时，只要他懂得表达文本所使用的那门自然语言，那么他就可以根据自身的理解能力产生对文本内容所要传递信息的认识和理解（对于计算机而言相当于解码的过程）。但计算机并不能轻易地“读懂”自然语言，本质上说，计算机只具有高超的计算能力，它只认识 0 和 1，只能对文本进行“计算”，所以我们必须将文本转换为计算机可以识别的格式。

下面以附件 2 中的一个文本数据为例说明如何实现文本向量化。留言主题为“E7 县人民医院医务人员职称晋升好难”的留言被分类到一级标签“卫生计生”下。在经过分词和去停用词处理后，留言主题变成“医院 医务人员 职称 晋升 好难”。我们知道，这些词出现较多的留言主题应该比它们出现较少的留言主题与“卫生计生”的相关性更高。但如果一条留言主题的长度较长，某些关键词重复

出现次数更多的话更占优势，因为一般来说长留言比短留言包含的关键词更多。

因此，我们采用根据留言的长度对关键词的次数进行标准化，也就是用关键词的次数除以留言的总字数，称之为“单词文本词频”（Term Frequency，TF）。在上例中，该留言主题共含有 16 个字，“医院”、“医务人员”、“职称”、“晋升”、“好难”分别出现一次，则它们的词频分别都为 0.0625。

抽象来讲，对于在某一留言主题 D_i 里的词语 w 来说，w 的词频可表示为：

$TF_w=N_w/N$

其中，Nw 是关键词 w 在留言主题 D_i 中出现的次数，N 为留言的总字数。

那么，度量留言主题与一级分类的相关度的一个简单做法就是直接将各个关键词出现的词频加总：

$TF_1+TF_2+...+TF_N$

至此，我们就会发现如果简单将改一级标签下所有出现的词排成一列，然后将该留言主题按照“出现的词所在的位置取 1，未出现的词所在位置取 0”的规则将其向量化，会存在一个严重的缺陷。很明显，在该留言主题中，“医院”和“医务人员”对将该留言划分为“卫生计生”类问题起着更为重要的作用，其贡献更大，在“卫生计生”这一类的留言主题中，这两个词的出现频率更高、概率更大。

因此，我们会给每一个词赋予权重，以突出关键词的相对重要程度。而该权重必须满足这样的条件，即关键词预测主题的能力越强，权重越大。比如上例中的“医院”，看到这个词我们便能够或多或少想到“卫生计生”，或我们会认为将该留言分类为“卫生计生”具有合理性，但如果只给定关键词“好难”，我们无法从中获取任何有关其可能分类的信息，因为这个词可能出现在任何分类的留言主题中。即在上述公式中，两个关键词的 TF 相同，但一个是特定留言分类下才会出现的词，而另一个是各类留言中都会出现的通用词，那么显然第一个词应当被赋予更高的权重，因此，好的权重公式应当反映出关键词的分辨率或者说贡献率。

概括来说，如果一个词 w 在 D_w 个留言主题中出现过，那么 D_w 越大，w 的权重越小，反之亦然。我们使用逆文档频率（Inverse Document Frequency，IDF）这一权重方式：

$IDF_w=\log(D/D_w)$

其中，D 是全部留言数，D_w 是关键词 w 在全部留言中出现的频次。

最后，TF-IDF=TF*IDF

我们使用 python3.7 执行此步骤的代码操作。首先将前两步处理过后的数据导入 python，然后对数据进行切分处理，提取全部文本数据的 80%作为训练集，剩下 20%作为测试集，并分别将训练集和测试集转换为 TF-IDF 权值。

下面展示部分处理过程：

标签处理前：

```
1813        城乡建设
4204        教育文体
5588    劳动和社会保障
8607        卫生计生
168         城乡建设
          ...
2859        环境保护
2416        环境保护
8295        商贸旅游
2256        环境保护
8416        卫生计生
Name: 一级标签, Length: 1842, dtype: object
```

标签处理后：

```
In[22]: valid_y
Out[22]: array([4, 5, 1, ..., 3, 6, 2])
In[23]: type(valid_y)
Out[23]: numpy.ndarray
In[24]: valid_y.shape
Out[24]: (1842,)
```

文本向量后部分结果：

```
In[27]: tfidf_te
Out[27]:
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

词对应部分结果：

```
In[29]: print(vectorizer.vocabulary_)
{'请问': 10360, '东风': 698, '新区': 6665, '公园': 1820, '建设': 5587, '指挥部': 6249, '拆迁': 6179,
 '补偿': 10071, '协议': 2692, '补偿款': 10072, '没拿到': 7778, '西地省': 10127, '计划生育': 10225,
'条例': 7193, '修订': 1618, '出台': 2088, '大汉': 4168, '巨龙': 4880, '家园': 4571, '黑诊所':
11807, '市竹埠港': 5219, '工矿': 4852, '棚户区': 7399, '改造': 6478, '项目': 11511, '绩效': 9424, '
工资': 4866, '未发': 7072, '一分钱': 260, '县凯迪': 2915, '绿色': 9447, '能源开发': 9714, '有限公司
': 7027, '拖欠': 6202, '百来': 8700, '员工工资': 3577, '学校': 4402, '补课': 10089, '乱收费': 1049,
 '何时休': 1491, '新天地': 6679, '悍豹': 5872, '武装': 7516, '押运': 6164, '违法': 10798, '收取':
6444, '押金': 6165, '咨询': 3610, '教育局': 6551, '2015': 78, '免除': 1733, '托幼': 6031, '事宜':
1070, '南站': 2733, '端午节': 9161, '违规': 10808, '售票': 3640, '且度': 644, '恶劣': 5865, '市冷':
 4975, '农村': 1992, '旧房': 6814, '办个': 2247, '建房': 5564, '证好': 10283, '城区': 3915, '道路':
 10930, '红绿灯': 9330, '绿化带': 9441, '后移': 3528, '建议': 5586, '青兰': 11460, '乡江': 1008, '
家村': 4585, '矿业': 8912, '洗选': 7890, '隐身': 11379, '农家': 1986, '滴滴': 8132, '公司': 1816, '
地方': 3845, '运营权': 10744, '县古阳镇': 2934, '危房改造': 2793, '规避': 10190, '招标': 6212, '分包
': 2117, '电影': 8584, '退休': 10834, '职工': 9617, '近几年': 10750, '养老金': 1932, '增加': 4010,
'厕所': 2826, 'wc': 229, '本来面目': 7097, '黄茶': 11782, '89': 178, '华泰': 2678, '职业': 9607, '
学校食堂': 4403, '水箱': 7665, '噪音': 3698, '扰民': 6067, '揭露': 6381, '十四': 2612, '教育':
```

2.2.2 分类以及效果评价

2.2.2.1 训练方法与分类算法

训练方法和分类算法是分类系统的核心部分，目前存在多种基于向量空间模型的训练算法和分类算法。其中，传统算法包括支持向量机、决策树、多层感知器、朴素贝叶斯、最近 K 邻居方法和逻辑回归；集成学习算法包括随机森林、AdaBoost、lightGBM 和 xgBoost；深度学习算法包括前馈神经网络和 LSTM。这些方法各有利弊，下面对其中三种较为常用的传统方法进行概述：

（1）KNN

KNN（K 最近邻，K-NearestNeighbor）是数据挖掘分类技术中最简单的方法之一。所谓 K 最近邻，即每个样本都可以用它最接近的 K 个邻居来代表。KNN 没有训练过程，分类时直接将待分类文本与训练集合中的每个文本进行比较，然后根据最相似的 K 篇文本得到新文本的类别。在文本分类中，KNN 是最简单有效的分类算法，简单且容易实现，因此常常能够取得较好的结果。但当训练数据集很大时，需要大量的存储空间，而且需要计算待测样本和训练数据集中所有样本的距离，所以非常耗时。

（2）朴素贝叶斯

贝叶斯分类是一类分类算法的总称，这类算法均以贝叶斯定理为基础。贝叶斯方法是通过计算文本 D 属于每个类别 C_i（i= 1，2，… ，M）的概率 P（C_i，D），并将它们排序取其最大值来得到 D 所属的类别。根据贝叶斯公式，最后归结于求每个类别的概率 P（C_i）和从类别 C_i 生成文本 D 的概率 P（D，C_i）。这两个概率都可以通过训练语料得到。朴素贝叶斯（Naïve Bayes，NB）是贝叶斯

方法中使用最简单，使用最广泛的一种。在这种方法中,假设 D 由互相独立的多个特征 w_j（j= 1，2，… ，N，其中 N 是 D 中不同特征数）生成,于是 P（D，C_i）由可以归结成求 P（w_j，C_i）朴素贝叶斯方法被广泛用于文本分类中。朴素贝叶斯的公式如下：

$$P(y_i|x_1,x_2,\cdots,x_d)=\frac{P(y_i)\prod_{j=1}^{d}P(x_j|y_i)}{\prod_{j=1}^{d}P(x_j)}$$

朴素贝叶斯方法又可以分为伯努利贝叶斯（BernoulliNB）、高斯贝叶斯（GaussianNB）和多项式贝叶斯（MultinomialNB），分别对应数据满足高斯分布（正态分布）、多项式分布和伯努利分布的训练集数据。

（3）SVM

支持向量机（Support Vector Machine，SVM）是一种二类分类模型，其基本模型定义为特征空间上的间隔最大的线性分类器，其学习策略便是间隔最大化，最终可转化为一个凸二次规划问题的求解。SVM 可以直接用于线性可分问题，而对于线性不可分的情形，可以构造一个变换，将问题转换到一个新的空间，在这个新空间中线性可分。在文本分类中，SVM 是公认的较好的方法之一。

2.2.2.2 分类结果及评价

KNN 方法和朴素贝叶斯分类方法简单且效率较高，经过比较，我们采用高斯贝叶斯方法训练模型。关于结果的评价，根据不同的文本分类应用背景，有多种评估分类系统性能的标准可供选择，其中最常用的评估标准包括查全率（Recall）、准确率（Precision）、F1 评测值、微平均（Micro-average）和宏平均（Macro-average）。另外，还有一些较少使用的评估方法，包括平衡点（break-even point）、11 点平均正确率（11-point average precision）等。

按照本题目标要求，我们采用最通用的性能评价方法——查全率、准确率和 F1 评价值——来评价分类的效果。假设一个文本分类系统针对类别 C_i 的分类标注结果统计如表所示：

表 1 性能评价

文本与类别的实际关系 / 分类判断	属于	不属于
标记为“是”	a	b
标记为“否”	c	d

上表中，a 表示正确地标注测试集文本为类别 C_i 的文本数量；b 表示错误地标注测试集文本为类别 C_i 的文本数量；c 表示错误地排除测试集文本在类别 C_i 之外的文本数量；d 表示正确地排除测试集文本在类别 C_i 之外的文本数量。

对于某一特定的类别，准确率是所有判断的文本中与人工分类结果吻合的文本所占的比率，其数学公式表示如下：

准确率（Precision）=分类的正确文本数/实际分类的文本=a/（a+b）

查全率是指被正确分类的文档数和被测试文档总数的比率，即该类样本被分类器正确识别的概率。其公式表示如下：

查全率（Recall）=分类的正确文本数/应有文本数=a/（a+c）

查全率和准确率分别从两个方面考察分类器的分类性能。召回率过高可能导致准确率过低，反之亦然。所以综合考虑分类结果召回率和准确率的平衡，采用 F1 评价值比较合理。公式如下：

F1 评价值=（准确率 • 查全率 • 2）/（准确率+查全率）

另外，我们在查相关文献可知有微平均和宏平均两种计算准确率、查全率和 F1 值的方法。微平均从分类器的整体角度考虑，不考虑分类体系的小类别上的分类精度，利用被正确分类标注的文本总数、被错误分类标注的文本总数以及应当被正确分类标注而实际上却被错误地排除的文本总数分别替换前式中的 a、b、c 从而得到微平均召回率、微平均准确率和微平均 F1 值。而宏平均是从分类器小类别的整体考虑，首先计算出每一类别的召回率与准确率，然后对召回率与准确率分别取算术平均得到的宏平均召回率与宏平均准确率，最后根据宏平均召回率与宏平均准确率计算宏平均 F1 值。微平均从文本分类标注正确总数角度衡量分类精度，宏平均则从每一类别文本标注正确的角度衡量分类精度。

由于 F1 指标只能用于评价分类结果为二值的情形，我们将目标转换成多个二分类问题。运用 for 循环，每轮只学习并预测一个分类指标，这个分类指标为 1，其余为 0，共七个分类，七轮循环，得到的 F1 结果算术平均求得最终结果。

第一轮部分预测结果展示：

```
....
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0,
```

求得 7 轮循环的总的精确度和召回率：其中列表第一个数字是对表示为 0 的指标的预测评分，第二个数字是指标为 1 的预测评分。

```
In[31]: pre
Out[31]: array([6.58207295, 4.22079292])
In[32]: rec
Out[32]: array([6.47202966, 4.4698712 ])
```

平均后求得每一轮平均的召回率和精确度：

```
In[34]: rect
Out[34]: array([0.92457567, 0.63855303])
In[35]: pret
Out[35]: array([0.94029614, 0.60297042])
```

从而得出 F1 系数结果：

```
In[37]: F1
Out[37]: array([0.93236964, 0.62025182])
```

2.3 热点问题挖掘

本题第二小问的目标是根据附件 3 给出的文本数据进行热点问题的挖掘，即根据留言内容的相似性进行聚类分析，将文本相似性高的留言归为一类，再通过自定义热度评价体系，对每一类问题的热度进行评价和排序，找出群众在某一时间段内反映较多的问题。与之前的操作类似，本部分目标的实现也必须首先对文本数据进行分词、去停用词等操作，然后再进行聚类分析并构建热度评价指标，确定热点问题。

2.3.1 聚类算法概述及选择

聚类分析属于常见的无监督分类算法，较常用的聚类算法有两类：

（1）基于分区的算法，其中以 k-means 算法为典型代表。这类算法易于理解，但也存在着较为明显的缺陷。首先，需要人为事先确定聚类的个数，当数据集很大时，事先给出一个较为合适的值比较困难。其次，这类算法无法适用于具有任意形状的簇，而只适用于具有凸形状的簇。最后，由于其对内存的占用资源比较大，很难推广到大规模数据集；

（2）基于层次划分的算法，如层次聚类法。这类算法虽然不需要事先确定聚类的个数，但需要确定停止分裂的条件，并且其计算速度较慢。基于这两类算法缺点的考虑，本文中我们采用基于密度的聚类方法——DBSCAN 聚类算法，该算法能够有效解决基于分区和基于层次划分算法的缺陷。

DBSCAN 是一种基于密度的聚类算法，其基本假设是一个集群的密度要显著高于噪声点的密度，因此，该方法的基本思想是对于集群中的每一个点，在给定的半径范围内，其相邻点的数量必须超过预先设定的某一个阈值。在进一步介绍 DBSCAN 聚类算法的核心思想之前，有必要对一些关键的基本概念进行解释：

（1）核心对象：若某个点的密度达到算法设定的阈值，即邻域内点的个数不少于 MinPts，则其为核心点。

（2）Eps 邻域（或半径）：对于一个点 p，记其 Eps 邻域为 NEps(p)，定义如下：

$$\mathrm{N_{Eps}(p)} = \{\mathrm{q} \in D | \operatorname{dist}(\mathrm{p},\mathrm{q}) \leq \mathrm{Eps}\}$$

其中，D 表示整个数据集集合，dist(p,q)表示点 p 和 q 的距离。

$$\begin{gathered} \mathrm{p} \in \mathrm{N_{Eps}(q)} \\ |\mathrm{N_{Eps}(q)}| \geq \mathrm{MinPts} \end{gathered}$$

（3）直接密度可达：若点 p 和点 q 满足下式关系，则称点 p 直接密度可达点 q：

其中，MinPts 表示事先确定的某一核心对象的 Eps 邻域必须包含的最小数量。当点 p 和点 q 都是一个集群的核心对象时，此时直接密度可达对两个点来说都是对称的。但当点 p 是边界点时，直接密度可达不是对称的。

（4）密度可达：若存在一个点的序列 p1，p2，…，pn，p1=q，pn=p，有

pi+1 从 pi 是直接密度可达的，则称 p 从 q 密度可达，这实际上是直接密度可达的“传播”。

（5）密度相连：若从某核心点出发，点 q 和点 k 都是密度可达的，则称点 q 和点 k 是密度相连的。

（6）边界点：属于某一个类簇的非核心点，且不能发展下线。

（7）噪声点：不属于任何一个类簇的点，从任何一个核心点出发都是密度不可达的。

为更加直观地理解各类点的关系，以下图为例，点 A 表示核心对象，点 B 和点 C 表示边界点，点 N 表示噪声点：

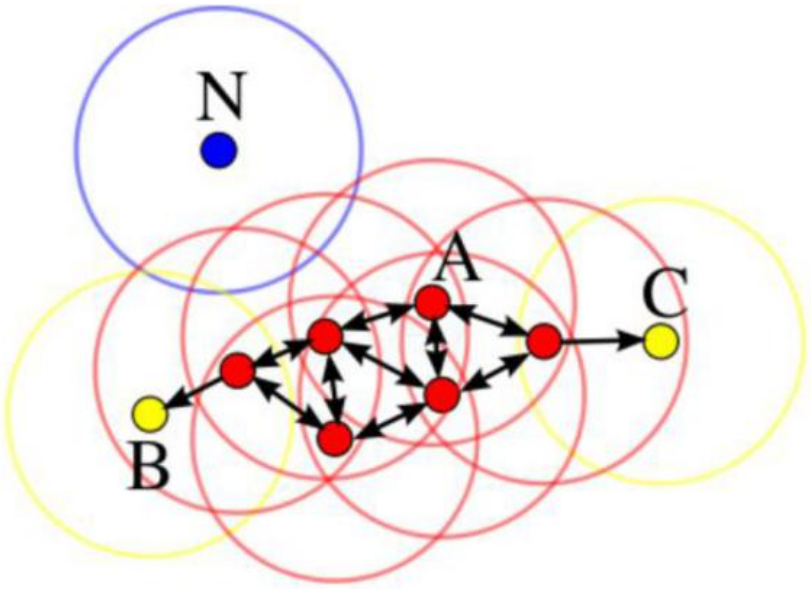


图 4 核心对象、边界点、噪音点

DBSCAN 聚类算法的定义为由密度可达关系导出的最大密度相连的样本集合，即为我们最终聚类的一个类别，或者说一个簇。这个 DBSCAN 的簇里面可以有一个或者多个核心对象。如果只有一个核心对象，则簇里其他的非核心对象样本都在这个核心对象的 Eps 邻域里；如果有多个核心对象，则簇里的任意一个核心对象的 Eps 邻域中一定有一个其他的核心对象，否则这两个核心对象无法密度可达。这些核心对象的 Eps 邻域里所有的样本的集合组成一个 DBSCAN 类簇。

那么怎么才能找到这样的簇样本集合呢？DBSCAN 使用的方法很简单，它任意选择一个没有类别的核心对象作为种子，然后找到所有这个核心对象能够密度可达的样本集合，即为一个聚类簇。接着继续选择另一个没有类别的核心对象去寻找密度可达的样本集合，这样就得到另一个聚类簇。一直运行到所有核心对

象都有类别为止。DBSCAN 聚类算法的总体工作流程如下图所示：

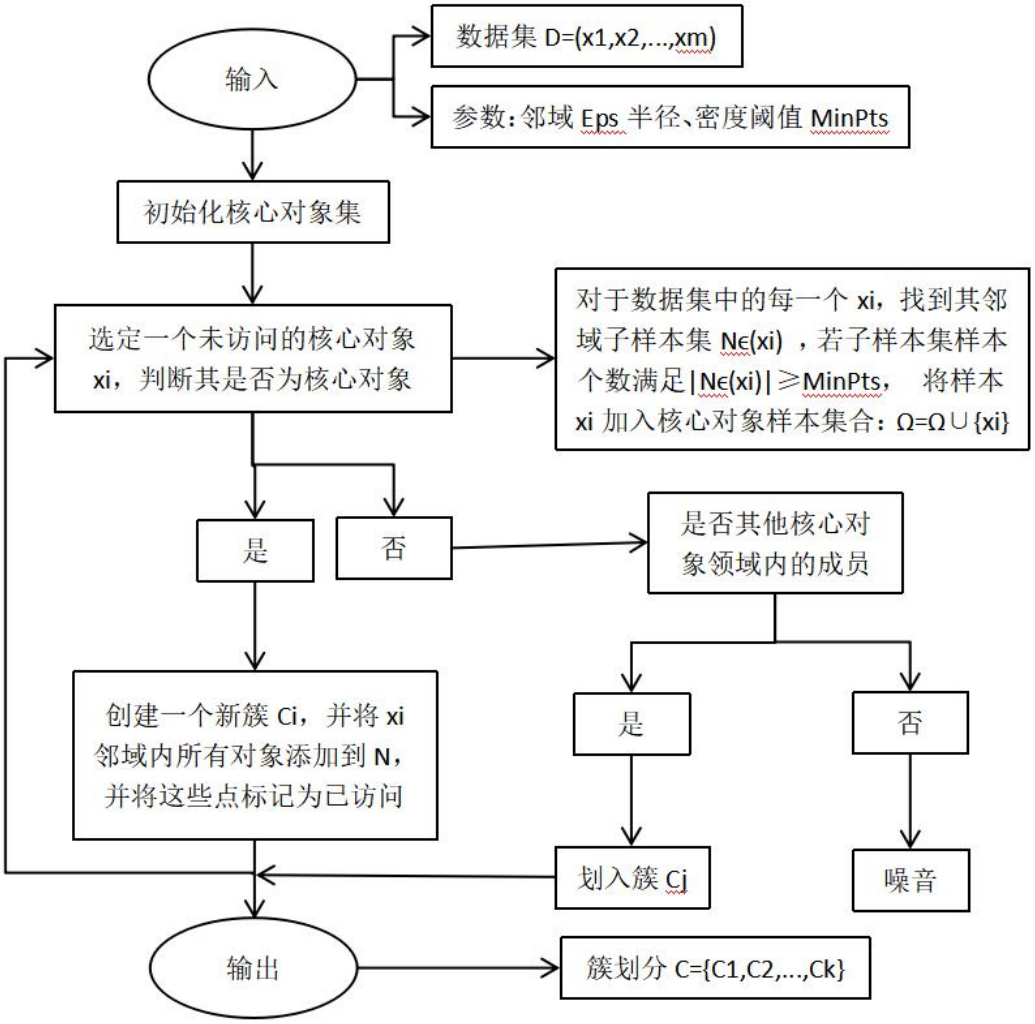


图 5 DBSCAN 聚类算法的总体工作流程

2.3.2 数据处理

2.3.2.1 分词和词性标注

要进行聚类分析，首先还是要进行文本数据的处理。在 DBSCAN 聚类算法下，我们需要对原始文本数据进行分词操做并记录词性。我们仍然使用结巴分词，关于结巴分词的工作原理和流程已经在前面留言分类的部分介绍过，此处不再赘述。与前述步骤不同的是，本步骤中在分词的同时，还要进一步标注每一个词的词性。部分分词结果示例参见下图：

```
In[4]: word_pos_dict
Out[4]:
{'留言': 'v',
 '主题': 'n',
 'A3': 'eng',
 '区': 'n',
 '一米阳光': 'nz',
 '婚纱': 'n',
 '艺术摄影': 'n',
 '是否': 'v',
 '合法': 'n',
```

2.3.2.2 文本向量化——TF-IDF 权重

使用 TF-IDF 对文本进行向量化，得到文本的 TF-IDF 权重。部分结果示例参加下图：

```
In[5]: weight
Out[5]:
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

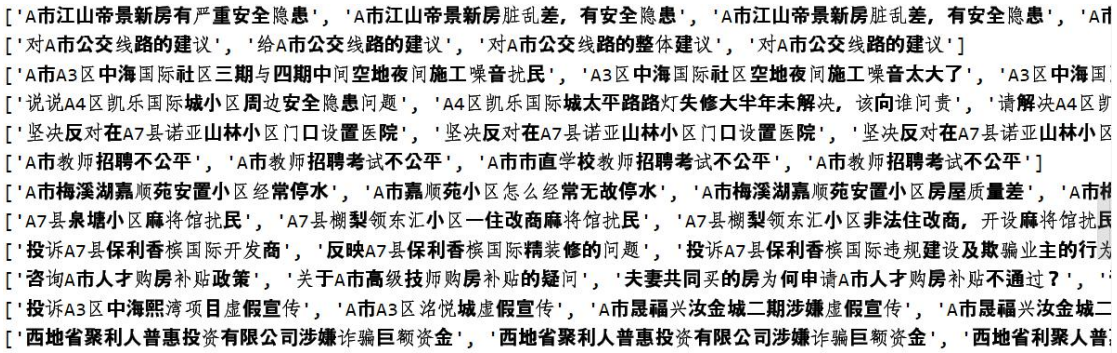
前面得到了分词的结果，并对词性进行了记录，接下来可以针对不同词汇的词性，给予其 TF-IDF 权重以不同的乘数，这样可以突出某些类型的词汇的重要性，在一定程度上有助于聚类的效果。

```
In[6]: new_weight
Out[6]:
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

2.3.3 聚类分析

在将文本数据进行上述初步处理之后，接下来就可以进入到正式的聚类分析

过程。由第一部分对 DBSCAN 聚类算法基本原理的介绍，我们可以知道这种算法的困难之处在于要确定两个参数，即 Eps 领域的半径，以及阈值 MinPts。我们设置 Eps 半径为 1，阈值 MinPts 为 4。通过 DBSCAN 聚类算法，我们共得到 35 个不同类别的留言簇，部分结果展示如下：


['A市江山帝景新房有严重安全隐患', 'A市江山帝景新房脏乱差，有安全隐患', 'A市江山帝景新房脏乱差，有安全隐患', 'A市
['对A市公交线路的建议', '给A市公交线路的建议', '对A市公交线路的整体建议', '对A市公交线路的建议']
['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', 'A3区中海国际社区空地夜间施工噪音太大了', 'A3区中海国
['说说A4区凯乐国际城小区周边安全隐患问题', 'A4区凯乐国际城太平路路灯失修大半年未解决，该向谁问责', '请解决A4区凯
['坚决反对在A7县诺亚山林小区门口设置医院', '坚决反对在A7县诺亚山林小区门口设置医院', '坚决反对在A7县诺亚山林小区
['A市教师招聘不公平', 'A市教师招聘考试不公平', 'A市市直学校教师招聘考试不公平', 'A市教师招聘考试不公平']
['A市梅溪湖嘉顺苑安置小区经常停水', 'A市嘉顺苑小区怎么经常无故停水', 'A市梅溪湖嘉顺苑安置小区房屋质量差', 'A市梅
['A7县泉塘小区麻将馆扰民', 'A7县棚梨领东汇小区一住改商麻将馆扰民', 'A7县棚梨领东汇小区非法住改商，开设麻将馆扰民
['投诉A7县保利香槟国际开发商', '反映A7县保利香槟国际精装修的问题', '投诉A7县保利香槟国际违规建设及欺骗业主的行为
['咨询A市人才购房补贴政策', '关于A市高级技师购房补贴的疑问', '夫妻共同买的房为何申请A市人才购房补贴不通过？', '
['投诉A3区中海熙湾项目虚假宣传', 'A市A3区洺悦城虚假宣传', 'A市晟福兴汝金城二期涉嫌虚假宣传', 'A市晟福兴汝金城二
['西地省聚利人普惠投资有限公司涉嫌诈骗巨额资金', '西地省聚利人普惠投资有限公司涉嫌诈骗巨额资金', '西地省利聚人普

2.3.4 构建热度评价指标

关于热度评价指标的构建，我们取聚成同一类的留言问题类簇中所包含的留言总数加上该类目下对应的原文本的点赞和反对的总数，来衡量该事件的热度。

聚类结果及热度指标示例如下图所示。文字部分表示聚成一类的原文本，末尾的数字则是热度。本题目标为输出前五类热点问题，通过观察，前两类聚类效果不好，原因可能是句子特征不明显，故删去，取第 3-7 类为最终结果。

('问问 A 市经开区东六线以西泉塘昌和商业中心以南的有关规划,问问 A 市经开区东六线以西泉塘昌和商业中心以南的有关规划,问问 A 市经开区东六线以西泉塘昌和商业中心以南的有关规划,问问 A 市经开区东四线以西、新安路以南的规划,问问 A 市经开区东四线以西、新安路以南的规划',
136),
('A 市长房云时代小区三期后面要建垃圾站？,A 市长房云时代 13 栋房屋有质量问题,A 市长房云时代附近建垃圾楼是真的吗？,A 市长房云时代多栋房子现裂缝，质量堪忧',
65),
('A 市经开区泉星公园项目规划需优化,建议 A 市经开区泉星公园项目规划进一步优化,建议进一步优化 A 市经开区泉星公园项目规划,建议 A 市经开区泉星公园项目规划进一步优化,给 A 市经开区泉星公园项目规划进一步优化的建议,对 A 市经开区泉星公园项目规划再进一步优化的建议',
39),

图 6 聚类结果展示

2.3.5 结果

经过以上步骤，我们最终得到 35 个留言问题类簇，通过定义热度评价指标为一类问题总数量与其点赞数和反对数之和，我们能够得到所有留言问题类簇的

热度排序，由于前两个类簇热度值存在异常，故我们将其剔除，取第三至第七留言问题类簇作为本文挖掘的热点问题。我们进一步按照要求将所得结果整理成如下表格。详情请参加附件。

表 2 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	136	2019/1/2 至 2019/2/11	A 市经开区东四线/东六线	A 市经开区东四线/东六线周边规划
2	2	65	2019/2/25 至 2019/9/11	A 市长房云时代小区	A 市长房云时代小区房屋质量反馈与询问周边规划
3	3	39	2019/8/9 至 2019/8/26	A 市经开区泉星公园项目	A 市经开区泉星公园项目需要优化
4	4	33	2019/2/14 至 2019/9/9	A7 县凉塘路	A7 县凉塘路的旧城改造进度缓慢
5	5	24	2019/7/8 至 2019/8/16	A7 县诺亚山林小区	反对在 A7 县诺亚山林小区门口设立医院

表 3 热点问题留言明细表

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	256358	A00080329	问问 A 市...	2019/1/2 20:27:07	A 市经开区东六线以...	0	29
1	233542	A00080329	问问 A 市...	2019/1/2 20:27:26	A 市经开区东六线以...	0	24
1	239670	A00080329	问问 A 市...	2019/1/11 15:46:04	A 市经开区东六线以...	0	41
1	275990	A00036259	问问 A 市...	2019/2/2 16:47:53	请问 A 市经开区东四...	0	16
1	261625	A000101635	问问 A 市...	2019/2/11 13:36:18	请问 A 市经开区东四...	0	21
2	188592	A00039456	A 市长房...	2019/6/18 10:38:44	长房云时代小区于今...	0	0
2	238038	A00012750	A 市长房...	2019/7/11 11:24:14	A 市长房云时代，其北...	0	0
2	220857	A00020702	A 市长房...	2019/9/11 22:30:46	领导您好：我是长房云...	0	1
2	281898	A00096623	A 市长房...	2019/2/25 15:17:38	领导：您好！我是长房...	5	55
3	226408	A00080342	A 市经开...	2019/8/9 16:47:36	目前 A 市经济技术开...	0	4

3	238692	A0008 0342	建议 A 市...	2019/8/12 13:15:05	目前 A 市经济技术开...	0	16
3	273741	A0008 0342	建议进一...	2019/8/13 16:02:24	目前 A 市经济技术开...	0	0
3	289574	A0008 0342	对 A 市经...	2019/8/16 13:33:59	目前 A 市经济技术开...	0	0
3	278281	A0003 2698	建议 A 市...	2019/8/22 13:23:46	目前 A 市经济技术开...	0	0
3	278545	A0003 6841	给 A 市经...	2019/8/26 13:00:06	目前 A 市经济技术开...	0	13
4	210366	A0003 5629	A7 县凉塘...	2019/7/31 16:42:51	A7 县星沙街道的旧...	0	1
4	228139	A0009 8215	请问 A7...	2019/9/9 8:20:47	A7 县星沙的旧城改...	0	1
4	234952	A0003 5628	A7 县星沙...	2019/7/31 11:34:33	A7 县星沙街道的旧...	0	0
4	250512	A0003 5626	A7 县星沙...	2019/9/2 14:32:27	星沙的旧城改造今年...	0	6
4	259574	A0007 2486	A7 县星沙...	2019/7/23 7:39:44	A7 县星沙街道四区...	0	1
4	260149	A0003 5631	A7 县星沙...	2019/6/10 10:24:59	2017 年 6 月 17 日星沙...	0	5
4	268757	A0007 2477	A7 县星沙...	2019/7/4 14:10:30	在 2017 年 6 月份，星...	1	4
4	274285	A0003 5628	A7 县星沙...	2019/8/2 10:05:02	A7 县星沙街道的旧...	0	6
4	284120	A0003 5630	A7 县星沙...	2019/2/14 10:07:59	今年政府工作报告里...	0	6
5	252413	A0007 26	坚决反对...	2019/8/16 8:36:38	爱尔眼科想在我们小...	0	0
5	226871	A0007 26	坚决反对...	2019/8/16 8:37:43	反对理由如下：一、在...	0	1
5	210107	A0004 2107	坚决反对...	2019/7/8 10:38:38	我们是诺亚山林小区...	0	14
5	194022	A0004 2107	坚决反对...	2019/7/8 10:39:54	我们是诺亚山林小区...	0	1
5	258186	A0004 2107	反对在 A7...	2019/7/8 10:48:31	我们是诺亚山林小区...	0	3

2.4 答复意见的评价

此部分我们组将结合层次分析法（AHP）和模糊综合评价法(FCE)综合分析

（简称 AHP-FCE 综合法）定性和定量尝试实现答复意见的质量评价方案的理论分析。

2.4.1 理论依据

我们在评价某一事物时，往往要涉及多个因素或者多个指标，同时根据这多个因素对事物作综合评价。因此，模糊综合评价法(FCE)是一种根据“模糊数学隶属度理论”把定性评价转化为定量评价的方法. 它结果清晰，系统性强， 能较好解决模糊的、难以量化的问题，适合各种例如评价方案这种非确定性问题的解决。

我们在使用 FCE 法计算需要确定各个评价指标的权重，即为权向量，一般由决策者直接指定. 但对于复杂的问题， 通常评价指标很多且相互影响，如果直接给出各个评价指标的权重比较困难，此时就需要 AHP 法进行分析。

在 AHP 中，通过分解问题，将复杂问题分解为多个子问题， 通过两两比较的形式给出决策数据，最终给出备选方案的排序权重，通常称为逐级排序法。因此本文我们会将评价指标作为 AHP 法的备选方案， 使用 AHP 法对问题分层建模，并根据专家对此模型的决策数据进行计算，得到备选方案计算出各个评价指标的排序权重，这样就能解决 FCE 中复杂评价指标权重的问题。

层次分析法在事实上是将一个复杂的多目标决策问题作为一个系统，将目标分解为多个目标或准则，进而分解为多指标（或准则、约束）的若干层次，通过定性指标模糊量化方法算出层次单排序（权数）和总排序，以作为目标（多指标）、多方案优化决策的系统方法。简而言之就是通过构建一套多层次的评价指标体系，完成对定性指标的定量化分析。因此，本文我们将用 yaaph 软件借助层次分析法的原理用于解决模糊综合评判中的权重问题。

2.4.2 建立层次结构模型

我们在使用 AHP-FCE 时，希望通过分解第三小问构造层次模型来完成评价指标的重要性排序，确定评价指标的合适权重，为构建评价方案提供理论依据。其基本原理是，首先利用 AHP 分层的思想对问题进行逐层分解， 把分层后的最下一层中间层要素(准则)作为评价指标，并将评价指标改为备选方案。

对于本文中的问题，我们对问题进行分层建模过程整理成表格如下：

表 4 AHP 分层建模过程

决策目标	答复意见质量评价方案
第一层中间层要素	答复相关性、答复完整性、答复有效性
第二层中间层要素	解释答复相关性用答复意见和留言详情之间的文本相似度衡量答复的相关性
	解释答复完整性用关键词识别(数据处理方法)，是否具有如下关键词：您好，感谢您，针对 XXX 问题，回复如下，感谢您对 XXX 的支持。
	解释答复有效性用答复时效性(通过留言时间和答复时间的时间差作为衡量答复时效性，我用 excel 中函数计算了答复时间和留言时间之间的时间差，以天为基数并取整，发现最大值为 1160 天，最小值为 0 天，根据实际情况综合考虑将时效性分阶段考虑：0-30 天，30—60 天，60-90 天，90 天及以上)，另外答复意见和留言详情之间的文本相似度也可作为衡量答复的有效性。

在解释答复相关性中，本文会采用 TF-IDF 相关性排序法同时做定量研究，利用余弦相似度计算文本之间的相关性，基本原理是在向量空间模型中，将两个表格中词条矢量，即用特征项集表示每一词条，随后以词频或者词的信息作为权重分析，分析关键点在于需要对数字映射后的每一条特征做一个余弦相似度的匹配。而这一问题在第一问的代码已经有所体现，这里主要解释下分析过程：

第一步: 对数据使用 DataFrame 化，并进行数组化

第二步：对数据进行分词，并去除停用词，使用' '.join 连接列表

第三步：np.vectorizer 向量化函数，调用函数进行分词和停用词的去除

第四步：使用 TF-IDF 词袋模型，对特征进行向量化数字映射

第五步：使用 from sklearn.metrics.pairwise import cosine_similarity， 对两两样本之间做相关性矩阵，使用的是余弦相似度计算公式

这一部分可通过机器学习衡量答复之间的相关性，由于人为的专家评判方法具有一定的偶发性，这种定性方法误差较大，且衡量标准较为模糊。因此通过机器学习可从定量的角度解释留言与答复之间的相关程度，弥补定性评分的缺陷。

随后经阅读相关文献，我们制定了一套评价方案评分标准，进一步整理成表格格式如下所示：

表 5 评价方案具体评判标准

衡量对象	衡量工具	专家评价细分准则（定性方法）	数据处理方法（定量方法）
答复相关性	答复意见与留言详情间的文本相关性程度	相关性程度较低　30 分	TF-IDF 相关性排序法
		相关性程度一般　50 分	
		相关性程度良好　70 分	
		相关性程度很高　90 分	
答复完整性	包含相关关键词个数	1-3 个　不太完整　40 分	关键词提取及其计数
		3-5 个　较为完整　60 分	
		5 个以上　十分完整　80 分	
答复有效性	(1)答复时效性，计算留言时间和答复时间的时间差	0-30 天　90 分	
		30—60 天　70 分	
		60-90 天　50 分	
		90 天及以上 30 分	
	(2)答复意见与留言详情间的文本相关性程度	同答复相关性衡量方法一致	同答复相关性衡量方法一致

2.4.3 模糊综合评判

2.4.3.1 建立模糊集

a) 建立因素集

根据影响答复意见的质量的三个因素，我们建立了如下的因素集：

U={答复相关性 U1，答复完整性 U2，答复有效性 U3}

b) 建立评价集

建立答复意见质量的评价级（四层）:V={优 V1，良 V2，中 V3，差 V4}

2.4.3.2 建立单因素评价矩阵

构建单因素评价矩阵我们会借助层次分析法解决评价矩阵中的评价指标的权重问题，利用评价集和评价指标的隶属集一一对应关系形成一级评判矩阵和二级评判矩阵，并进行对应的一致性检验，若均通过一致性检验，则可形成 AHP 调查表用于专家评分，这里由于专家评分具有一定的倾向性，定性分析具有一定

的误差，也可形成模糊综合评价问卷，用于更为精确地定量分析，我们主要注重其理论分析的可行性， 结合 TF-IDF 定量分析，形成 AHP-FCE 综合分析用于解释本题的评价方案制定。

2.4.4 层次分析法分层建模过程

2.4.4.1 分层建模模型

我们随后利用 yaahp 对所构建的层次结构进行分层建模，模型如下所示：

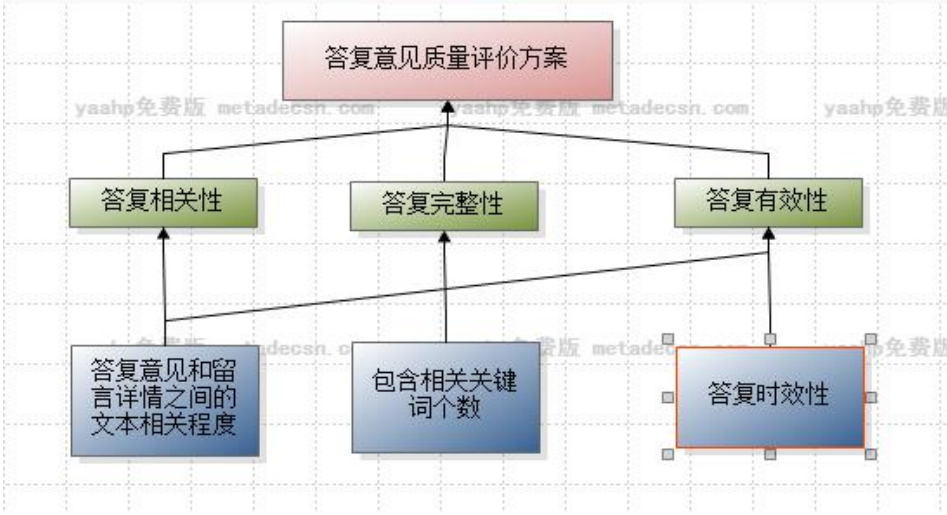


图 7 评价方案分层模型

2.4.4.2 构造判断（成对比较）矩阵

a)构造判断矩阵

我们通过对三个评价指标因素两两比较，给出判断矩阵如下所示：

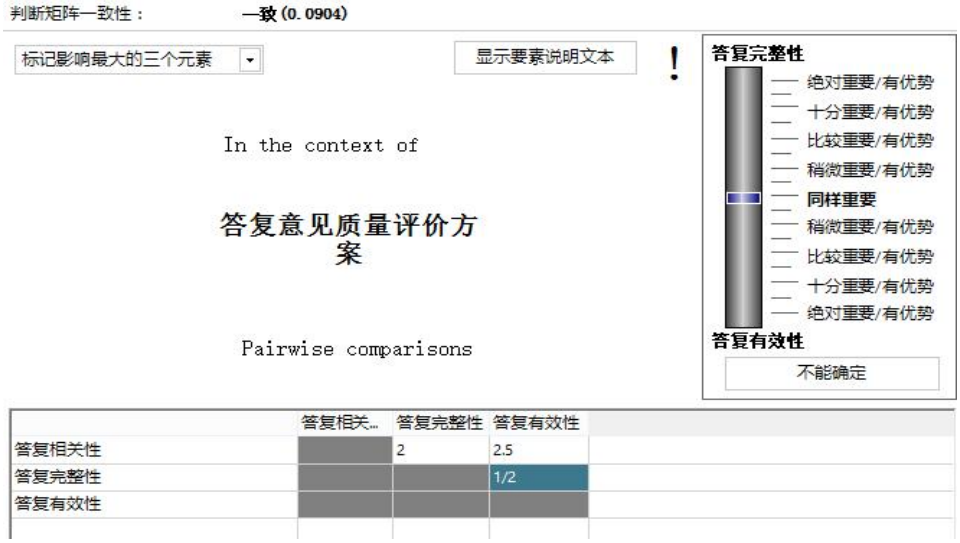


图 8 判断矩阵

这一步主要目的在于计算第二层中间矩阵的最大特征值；然后评价指标的权重 W，经不断尝试，这里我们设置答复时效性和答复意见和留言详情之间的文

本相似度之间的关系权重为 50%，答复完整性和答复相关性的权重为 2，答复相关性和答复有效性的权重为 2.5，上图显示判断矩阵的一致性为 0.094，第一层评价要素通过 RI 检验，即为通过一致性检验，满足了本文要求，随后继续构建第二层中间评价要素的矩阵检验，以确定评价要素的合适权重。

b) 计算第二层评价指标权重

中间矩阵经 yaahp 运行出来结果如下所示：

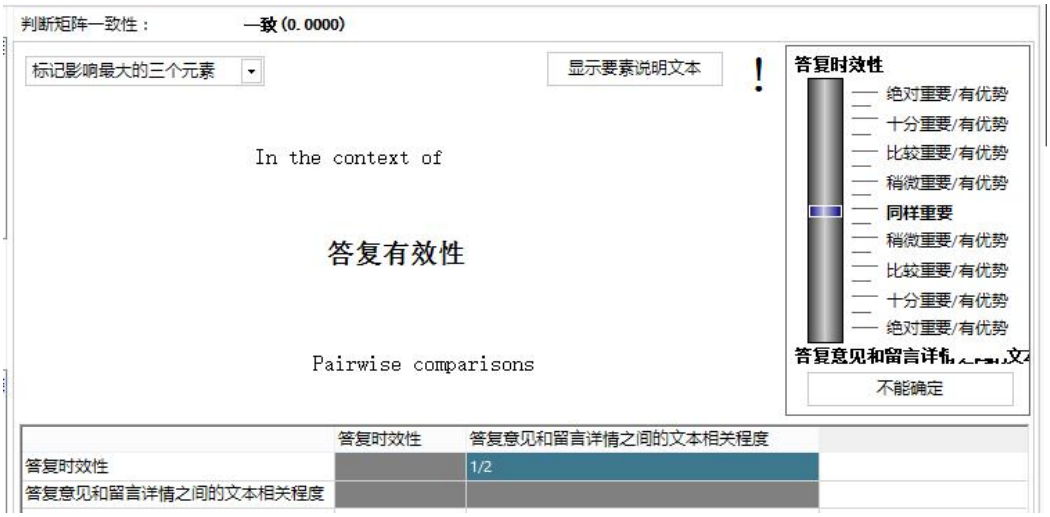


图 9 中间矩阵运行结果

这里我们设置答复时效性和留言、答复相关程度两个评价要素之间的权重为 1/2，结果显示通过一致性检验，可进一步进行层次单排序。

2.4.4.3 层次单排序及其一致性检验

进一步操作，设置备选方案通过一致性检验后，计算结果如下：

a)条状图如下：

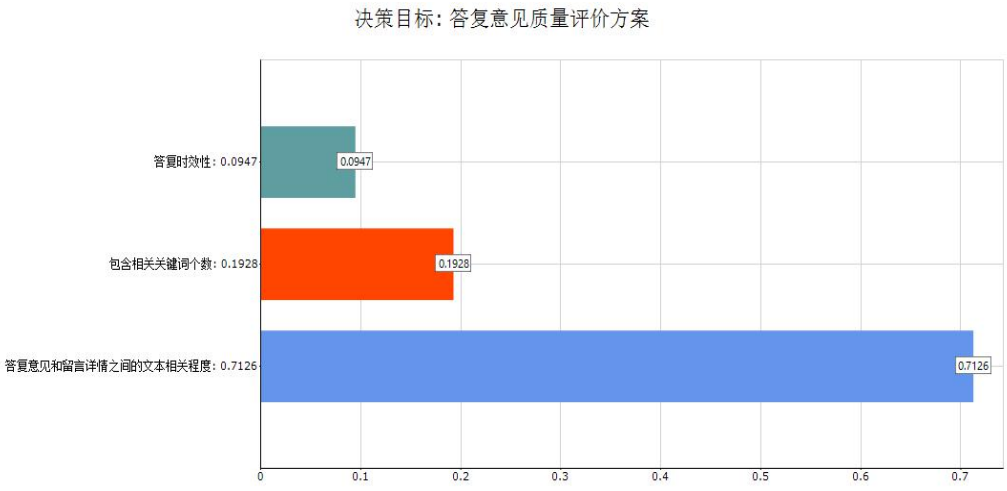


图 10 层次单排序的条状图结果

b)柱状图如下：

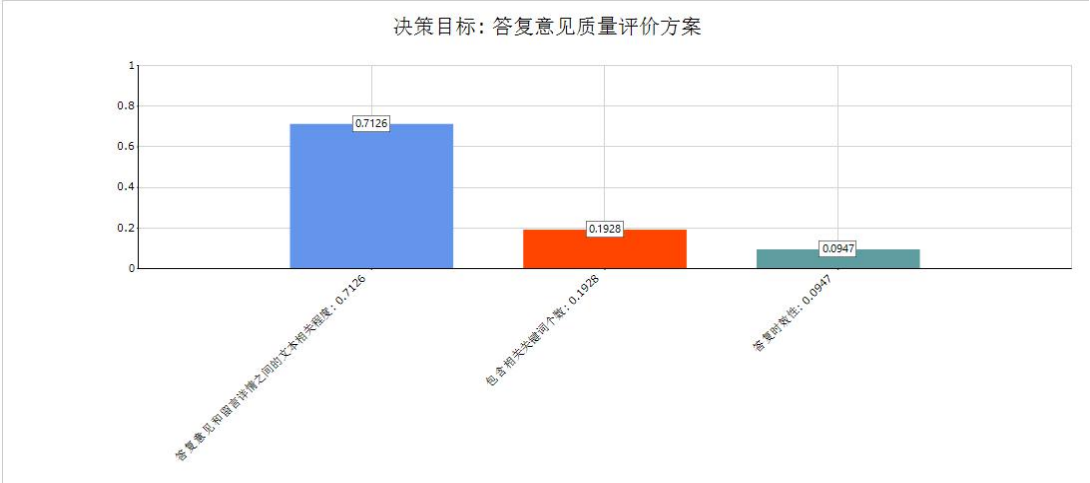


图 11 层次单排序的柱状图结果

c)饼图如下：

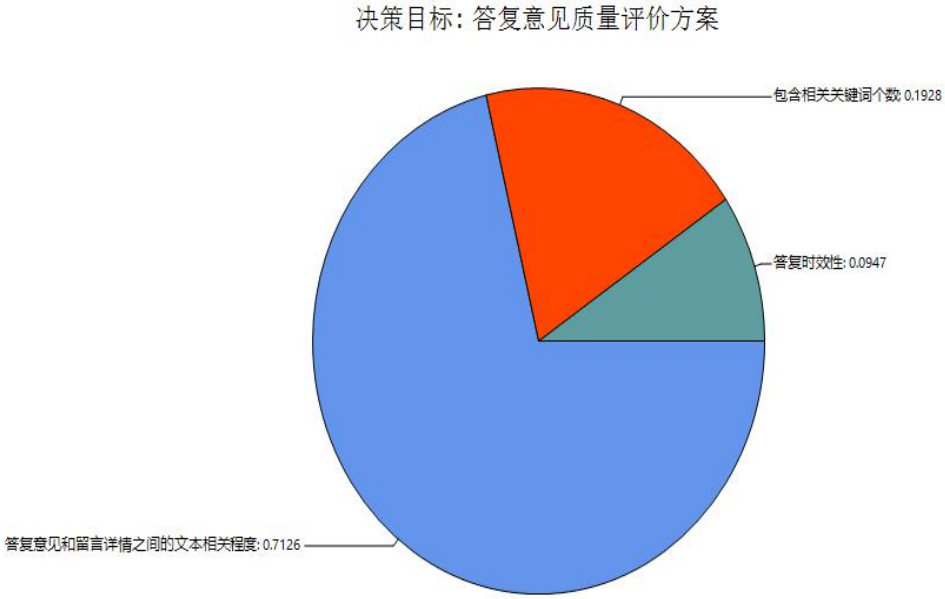


图 12 层次单排序的饼图结果

如上图所示，在第二层评价指标集中，答复意见和留言详情之间的文本相关程度这一要素在评价方案中重要性最大,占比为 17.26%；其次为包含关键词个数，占比 19.28%。最后考虑答复时效性，占比为 9.47%。

2.4.4.4 层次分析法计算结果的综合评价

我们经过 Yaahp 软件分析后，显示结果的详细数据如下：

a)权重分布图：

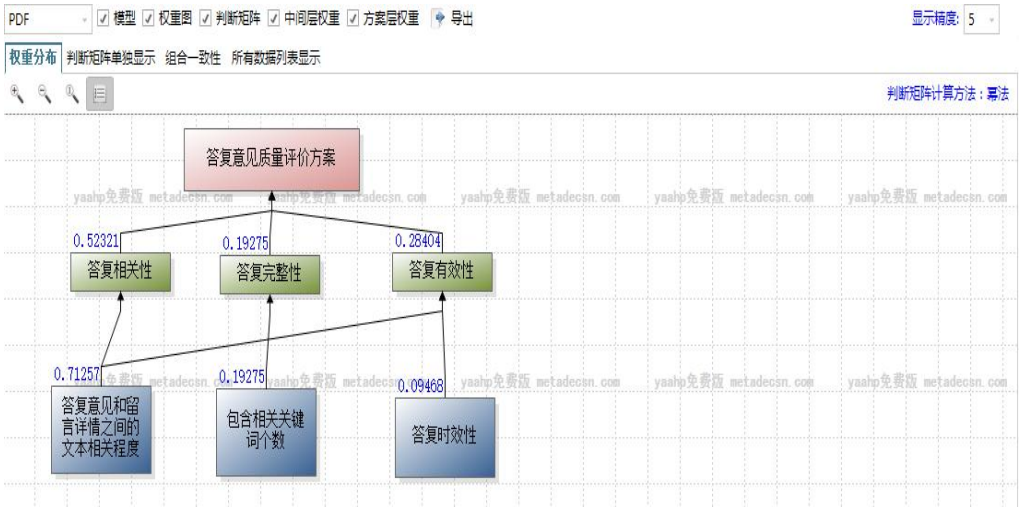


图 13 层次分析法计算结果的权重分布图

上图我们可看出，答复相关性的重要性最大，占比为 53.32%；其次答复有效性，占比为 28.40%；最后是答复完整性。说明我们在设置 AHP 专家评分表时，设置的权重排序应为答复相关性>答复有效性>答复完整性，就现实情况而言，分析结果也符合真实情况，具有可操作性。

b)判断矩阵单独显示如下：

图 14 层次分析法计算结果的判断矩阵

根据判断矩阵的结果显示，整个答复意见质量评价方案的一致性比例为 0.0904，且 λmax 为 3.094，通过了一致性检验。

c)组合一致性如下所示：

要素	权重	CI	RI(阶数)
方案层			
答复意见和留言详情之间的文本相关程度	0.71257		
包含相关关键词个数	0.19275		
答复时效性	0.09468		
第 1 准则层　组合一致性比例: 0.00000			
答复相关性	0.52321	0.00000	0.00000 (1)
答复有效性	0.28404	0.00000	0.00000 (2)
答复完整性	0.19275	0.00000	0.00000 (1)

图 15 层次分析法计算结果的组合一致性

结果显示组合一致性的分析结果同权重分布图的分析结果一致，在这里不再赘述。

d)所有数据列表显示：

方案层中要素对决策目标的排序权重

备选方案	权重
答复意见和留...	0.71257
包含相关关键...	0.19275
答复时效性	0.09468

第1个准则层中要素对决策目标的排序权重　组合一致性比例：0.00000

准则层要素	权重
答复相关性	0.52321
答复有效性	0.28404
答复完整性	0.19275

1. 答复意见质量评价方案　一致性比例：0.09040; 对"答复意见质量评价方案"的权重：1.00000; λmax：3.09402

答复意见质量...	答复相关性	答复完整性	答复有效性	Wi
答复相关性	1	2	5/2	0.52321
答复完整性	1/2	1	1/2	0.19275
答复有效性	2/5	2	1	0.28404

2. 答复相关性　一致性比例：0.00000; 对"答复意见质量评价方案"的权重：0.52321; λmax：1.00000

答复相关性	答复意见和留言详情之间的文本相关程度	Wi
答复意见和留...	1	1.00000

3. 答复完整性　一致性比例：0.00000; 对"答复意见质量评价方案"的权重：0.19275; λmax：1.00000

答复完整性	包含相关关键词个数	Wi
包含相关关键...	1	1.00000

4. 答复有效性　一致性比例：0.00000; 对"答复意见质量评价方案"的权重：0.28404; λmax：2.00000

答复有效性	答复时效性	答复意见和留言详情之间的文本相关程...	Wi
答复时效性	1	1/2	0.33333

图 16 层次分析法计算结果所有数据

综上所述，层次分析法结合模糊综合矩阵分析经过 yaahp 通过了一致性检验后，三个评价指标有了合适权重，便于进一步构造 AHP 调查表，用于专家评分，

这样模糊综合评判的权重问题就通过层次分析法的分层建模得以解决，这样本文对于评价方案的理论依据就站得住脚，由于缺少对应的专家对 AHP 调查表进行评分，本文在此就通过 AHP-FCE 层次分析结合模糊综合评判对构建答复质量评价方案进行了初步的理论估计，希望方便相关部门在经过专家评分和机器学习后的参考。

三、结论

本文对问政平台上群众的留言文本数据进行处理，首先采用高斯贝叶斯方法训练分类模型，实现了群众留言的机器分类，分类结果初步达到了预计设想，关键在于文本特征提取需要进一步改善精炼，更理想的文本分类方法是我们在数据处理工作中有待学习和拓展的；

其次，我们进一步通过 DBSCAN 聚类算法将群众留言进行聚类，定义并计算热度评价指标衡量每一留言问题类簇的热度值，成功挖掘出了群众留言的热点问题并进行较为精确的归类，并导出热点问题表和热点问题明细表；

最后，我们利用 AHP 层次分析法结合 FCE 模糊综合评判方法（AHP-FCE 综合分析）对相关部门对群众留言的答复进行了评价。结果显示，答复相关性、完整性和可解释性三个评价指标的重要性排序基本确定，且这三份评价指标均通过一致性检验，符合现实情况。AHP-FCE 综合分析法在分析评价方案领域属于比较前沿的方法，创新性较强但应用体系还不够完善，目前采用这种方法的学者还不多。我们此次在参考相关文献后使用该前沿方法，希望能够为本题评价方案的制定提供较为新颖的解决思路，为政府相关部门建设“智慧政务”提供借鉴和灵感，以期早日搭建起完善的政务服务体系，实现更高的群众满意度和服务效率。

四、参考文献

[1] 谢丽星, 周明与孙茂松, 基于层次结构的多策略中文微博情感分析和特征抽取. 中文信息学报, 2012(01): 73-83.

[2] 张立芳. 基于 AHP－模糊综合评判法的财政支出绩效评价研究[J]. 财经界:学术版, 2011(09):7-8.

[3] 陈炀,苗通,马欣.基于 AHP 模糊综合评价法的养老机构绩效评价研究[J].

中国物价,2020(05):90-93.

[4] 张东明,李亚东,黄宏伟.面向一流人才培养的研究生教育质量评价方法初探——基于层次分析与模糊综合评判的指标体系研究[J].研究生教育研究,2020(02):60-67.

[5] 田万宾.基于 AHP 的公共空间的评价方法研究——以百万庄“子丑寅卯”区为例[J].建筑与文化,2019(07):210-211.

[6] 苏金树,张博锋,徐昕.基于机器学习的文本分类技术研究进展[J].软件学报,2006(09):1848-1859.

[7] 庞剑锋,卜东波,白硕.基于向量空间模型的文本自动分类系统的研究与实现[J].计算机应用研究,2001(09):23-26.

[8] Martin Ester, Hans-Peter Kriegel, Joerg Sander, Xiaowei Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD. 1996: 226-231.

[9] 王宇. 基于 TFIDF 的文本分类算法研究[D]. 郑州大学, 2006.

[10] 谭松波. 高性能文本分类算法研究[D]. 中国科学院研究生院（计算技术研究所）.

[11] 巩知乐, 张德贤, 胡明明. 一种改进的支持向量机的文本分类算法[J]. 计算机仿真, 2009, 26(7):164-167.

[12] 郭金玉, 张忠彬, 孙庆云. 层次分析法的研究与应用[J]. 中国安全科学学报, 2008, 018(005):148-153.

[13] 吴殿廷, 李东方. 层次分析法的不足及其改进的途径[J]. 北京师范大学学报:自然科学版, 2004(02):125-129.

[14] 韩利, 梅强, 陆玉梅,等. AHP-模糊综合评价方法的分析与研究[J]. 中国安全科学学报, 2004(07):89-92.