基于机器学习的“智慧政务”文本挖掘应用

摘要

随着互联网与移动互联网的快速发展，各网络问政平台每天产生的数据量也十分巨大。因此，“智慧政务”中的文本挖掘应用问题也成为了当下热门研究之一。

针对问题一，本文根据一级标签对附件 2 留言数据进行数据清洗、取出中文停用词以及 jieba 分词等数据预处理后，利用 Python 工具对多项式朴素贝叶斯、线性支持向量机、逻辑回归、随机森林四种分类模型进行训练对比，发现线性支持向量机的分类模型的分类效果更好且分类准确率达到 87.0%，随后建立 F-Score 评价体系对分类模型进行评价，最后对构建的留言分类模型进行验证和优化。

针对问题二，本文对附件 3 留言详情表进行相似问题归类，首先本文选取“留言标题”作为问题识别与问题聚类的依据，然后将留言信息按地方进行区域划分，保证了数据集的不同区域性。通过使用 jieba 库和自定义词典对数据集进行中文分词后，文本利用基于文本相似度的聚类模型将相似度较大的留言归为一类问题。本文还使用轮廓系数对聚类模型进行性能评估，发现聚类效果较好。然后定义热度评价指标 H，根据改进的 Reddit 热度排名算法对聚类问题进行热度评价，得出问题的热度排名并汇总成热点问题表格。

针对问题三，本文对附件 4 中数据构建模糊综合评价模型。结合实际背景，本文利用相关性、完整性、可解释性、答复效率四个指标对政府部门的答复意见质量进行量化评价，然后构造“答复意见总得分 S”和“答复得分等级 L”两个新指标，最终完成构建模糊综合评价模型，得到了良好的综合评价效果。附件 4 中的留言答复数据总体情况较好，有少数留言答复在相关性、完整性、可解释性、答复效率的某个方面上有待提高。

关键词： 机器学习； NLP； 线性 SVM； F-Score； 相似度计算； 模糊综合评价

Abstract

With the rapid development of the Internet and the mobile Internet, the amount of data generated by each network interrogation platform is also huge. Therefore, the application of text mining in "smart government affairs" has become one of the current hot research.

For problem one, this article performs data cleaning on the attachment 2 message data according to the first-level label, extracts Chinese stop words and jieba word segmentation and other data preprocessing, and uses Python tools to perform polynomial naive Bayes, linear support vector machine, logistic regression, Random forest four classification models were compared for training, and it was found that the classification effect of the linear support vector machine classification model was better and the classification accuracy rate reached 87.0%. Then, an F-Score evaluation system was established to evaluate the classification model, and finally the message classification was constructed. The model is verified and optimized.

For question two, this article classifies similar questions on the message detail table in Annex 3. First of all, this article selects the "message title" as the basis for question identification and question clustering, and then divides the message information by region to ensure the difference of the data set. regional. After using the jieba library and a custom dictionary to perform Chinese word segmentation on the data set, the text uses a clustering model based on text similarity to classify the messages with greater similarity as a type of problem. This paper also uses the contour coefficient to evaluate the performance of the clustering model and finds that the clustering effect is better. Then define the heat evaluation index H, evaluate the heat of the clustering problem according to the improved Reddit heat ranking algorithm, and obtain the heat ranking of the problem and summarize it into a hot question table.

In response to question three, this paper builds a fuzzy comprehensive evaluation model for the data in Annex 4. Combining with the actual background, this paper uses the four indicators of relevance, completeness, interpretability, and response efficiency to quantitatively evaluate the quality of the government's response opinions, and then constructs two total scores: "S" and "L". The new index finally completed the construction of the fuzzy comprehensive evaluation model, and a good comprehensive evaluation effect was obtained. The overall response data of the message reply in Annex 4 is relatively good, and a few message responses need to be improved in some aspects of relevance, completeness, interpretability, and response efficiency.

Keywords: machine learning; NLP; linear SVM; F-Score; similarity calculation; fuzzy comprehensive evaluation

目录

1 问题重述

1.1 问题背景

随着互联网与移动互联网的快速发展，各网络问政平台正逐步成为政府了解基层群众、帮助群众办事的重要工具，随之而来的是大量相关文本数据的生成，纵观这些数据而言，一个网络问政平台，在一个星期内产生的文本数据级别已经达到 TB 级别，所以如何能够快速的处理与挖掘出这些有用的数据是关键所在，在以往利用人工处理这些海量文本数据而导致效率低下的情况下，利用大数据技术、人工智能、深度学习等技术建立基于自然语言处理的智慧政务系统是社会发展的新趋势。

1.2 问题的提出

根据附件所给的群众问政留言记录数据以及相关部门对群众留言的答复意见数据，从以下几个方面地数据进行数据挖掘：

(1)针对附件 2 数据，建立基于群众留言内容数据一级标签的分类模型，并利用 F-Score 对分类模型进行评价。

(2)针对附件 3 数据，将群众留言内容进行相似归类，并定义热度评价指标，给出留言热度排名，列出排名前 5 的热点问题。

(3)针对附件 4 数据，对政府部门对留言的答复意见从答复相关性、完整性、可解释性等角度对答复意见的质量进行综合评价。

2 问题分析

2.1 问题 1 的分析

问题一旨在根据附件 2 给出的数据，建立关于留言内容的一级标签分类模型,并利用 F-Score 对该分类模型进行评价。

通过对附件 2 留言数据进行数据清洗、取出中文停用词以及 jieba 分词等数据预处理，利用多项式朴素贝叶斯、**线性支持向量机**、逻辑回归、随机森林四种分类模型进行分类训练，选择分类效果较好的分类模型，最后采用 **F-Score** 对分类模型进行评价，通过分析预测失误的例子详情和 F1-Score，对构建的留言分类模型进行验证和优化。

2.2 问题 2 的分析

问题二旨在根据附件 3 数据， 将某一时段内反映特定地点或特定人群的热度问题的留言进行归类，并定义合理的热度评价指标，并给出评价结果。

本问题中，可以利用 jieba 库和自定义词典以及常见的停用词列表相结合对留言标题以及留言详情进行中文分词，随后利用基于文本相似度的聚类模型将相似度较大的留言归位同一类，然后**定义热度评价指标 H，**最后可利用改进 reddit 算法得到热点问题排名，再汇总成热点问题表和热点问题留言明细表。

2.3 问题 3 的分析

问题 3 旨在根据附件 4 数据，对于群众留言答复意见从相关性、完整性、可解释性、答复效率等指标，来对留言答复质量进行综合评价并加以实现。

针对于相关性指标，利用 jieba 算法对基层群众留言详情数据列与政府相关部门答复数据列进行关键词统计与中文分词，随后利用 IF-IDF 算法计算基层群众留言详情数据列与政府相关部门答复数据列对应的 TF-IDF 值，最后将该值进行单位化处理，利用打分制得出文本的相关性。

针对于完整性指标，通过将问候语(如：您好，你好等)、处理问题告知语(如：解决、协商、上交等)、结束语(感谢、理解、关心等)等相关语料库加入到 jieba 自定义语料库对政府相关部门答复数据进行中文分词，利用 IF-IDF 算法判断政府相关部门答复数据是否与问候语、处理问题告知语、结束语相匹配，最后利用打分制得出文本的完整性。

针对于可解释性指标，利用法律法规解释、调查，调解用语匹配解释两个特征来衡量该指标，通过将法律法规，调查，调解用语等相关语料库加入到 jieba 自定义语料库对政府相关部门答复数据进行中文分词，利用 IF-IDF 算法判断政府相关部门答复数据是否与该特征相匹配，最后利用打分制得出文本的可解释性。

针对于答复效率指标，利用 MATLAB 工具对群众留言时间列与政府相关部门答复时间列以天为单位进行时间间隔计算，通过定义不同时间间隔区间判分模型，得出政府部门答复效率。

本文针对附件 4 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、答复效率四个角度对政府部门对留言的答复意见质量进行衡量，并进行量化评分，构建

起**模糊综合评价模型**并对附件 4 数据加以实现。

3 符号说明

符号	符号说明
H	热度评价指标
S	答复意见总得分
L	答复等级
a	样本与其所在相同聚类的平均距离
b	样本与其距离最近的下一个聚类里的点的平均距离

4 模型的建立与求解

4.1 问题一：群众留言分类

4.1.1 问题的分析

在处理网络问政平台的群众留言时，一般需要按照一定的划分体系对留言进行分类，本文根据附件 2 给出的数据，建立一个关于留言内容的一级标签分类模型。通过对附件 2 留言数据进行数据清洗、取出中文停用词以及 jieba 分词等数据预处理，利用多项式朴素贝叶斯、线性支持向量机、逻辑回归、随机森林四种分类模型进行分类训练，选择分类效果较好的分类模型，最后采用 F-Score 对分类模型进行评价，通过分析预测失误的例子详情和 F1-Score，提出了相关分类模型的改进措施。

4.1.2 数据预处理与模型准备

为了防止控制对模型的准确性造成的影响，本文对问题一的初始数据进行数据清洗，以便于清洗掉数据中的空值。经过对初始数据进行数据清洗，可以发现在一级标签数据列中空值为 0，留言详情数据列中空值为 1，可以说明此次初始数据的完整性较好。由于除字母、数字、汉字以外的所有符号都是没有意义的，因此也将相应数据进行清洗，以便后续的进一步分析。

为了能够更加直观地了解数据，本文对附件二的数据进行一级标签类别统计并作出以下可视化图像与相应表格。

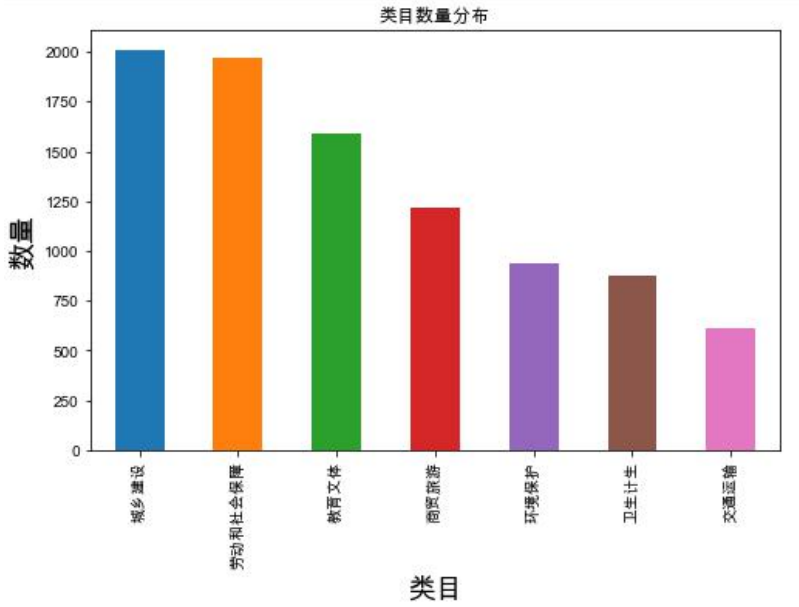


图 1 类别统计图

为了便于后期的分析和处理，本文对每一类一级标签添加类编号和统计出每个类别的数据，得到如下表所示：

表 1 类别统计表

类编号	一级标签	数据总量
0	城乡建设	2009
1	劳动和社会保障	1969
2	教育文体	1588
3	商务旅游	1215
4	环境保护	938
5	卫生计生	877
6	交通运输	613

通过上述图像，可以知道一级标签列含有 7 个类别，分别为：城乡建设、劳动和社会保障、教育文体、商贸旅游、环境保护、卫生计生、交通运输。同时，也可以观察到一级标签**各个类别的分布不一致**，其中，**市民反映较多的问题类别有：城乡建设、劳动和社会保障以及教育文体。**

由于本文使用的留言内容都是中文,所以为了加大模型的准确性对留言内容进行一些预处理工作是必不可少的。本文数据集文本中含有大量的标点符号、特殊符号、无意义词语等等，这些高频常用词无法反应出文本的主要意思,因此要被过滤掉。对于这些词语，本文选用了通用的**停用词文件**对数据集进行清洗。

本文采取的方法步骤如下图所示：

图 2 去中文停用词

本文过滤掉文本的标点符号和一些特殊符号、无意义词语后，接下来要在 clean_review 列的基础上进行 jieba 分词,把每个评论内容分成由空格隔开的单独的词语，得到的部分数据如下图所示。

cat_id	clean_review	cut_review
0	A3区大道西行便道未管所路口至加油站路段人行道包括路灯杆被圈西湖建筑集团燕子山安置房项目施工...	A3 区 大道 西行 便道 未管 所 路口 至 加油站 路段 人行道 包括 路灯 杆 被 圈...
0	位于书院路主干道的在水一方大厦一楼至四楼人为拆除水电等设施后烂尾多年用护栏围着不但占用人行道...	位于 书院 路 主干道 的 在水一方 大厦 一楼 至 四楼 人为 拆除 水电 等 设施 后 ...
0	尊敬的领导A1区苑小区位于A1区火炬路小区物业A市程明物业管理有限公司未经小区业主同意利用业...	尊敬 的 领导 A1 区苑 小区 位于 A1 区 火炬 路 小区 物业 A 市程明 物业管理...
0	A1区A2区华庭小区高层为二次供水楼顶水箱长年不洗现在自来水龙头的水严重霉味大家都知道水是我...	A1 区 A2 区华庭 小区 高层 为 二次 供水 楼顶 水箱 长年 不洗 现在 自来水 龙...
0	A1区A2区华庭小区高层为二次供水楼顶水箱长年不洗现在自来水龙头的水严重霉味大家都知道水是我...	A1 区 A2 区华庭 小区 高层 为 二次 供水 楼顶 水箱 长年 不洗 现在 自来水 龙...

图 3 jieba 分词部分图

通过 jieba 分词后可以获得每个类别大量文本词语数据，本文根据其中出现频率最高的**前 100 个关键词**作出每个类别的**词云图像**。得到的每个类别高频词的词云如下所示。

图 4 每一类高频词词云图

4.1.3 分类模型的建立

(1)模型的对比选择

在选取模型上，本文选取了多项式朴素贝叶斯、线性支持向量机、逻辑回归、随机森林四个模型，为了探索哪种分类器模型的分类效果更好，本文选取了附件 2 部分数据集对常用的 4 种分类器进行训练和测试，并评估它们的准确率。通过对常用的 4 种分类器进行训练和测试，得到了以下模型的准确率，如箱体图所示。

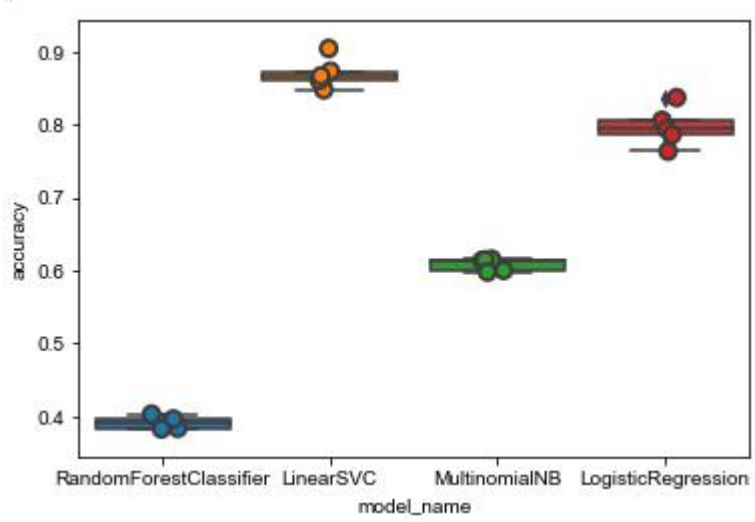


图 5 四种模型准确率箱体图

表 2 模型准确率表

模型名称	模型准确率
线性支持向量机	0.870464
逻辑回归	0.797273
多项式朴素贝叶斯	0.608423
随机森林	0.391790

从上述箱体图与模型准确率表可以看出有三个分类器的平均准确率都在 60%以上。其中**线性支持向量机的准确率最高**。随机森林分类器的准确率是最低的，因为随机森林属于集成分类器，一般来说集成分类器不适合处理文本数据,因为文本数据有太多的特征值,使得集成分类器难以应付。

通过上文的比较，发现线性支持向量机的平均准确率达到了 87.0%，**因此本文最终选择建立线性支持向量机分类模型作为问题一的留言分类模型。**

(2)模型原理

支持向量机模型包括：线性可分支持向量机、线性支持向量机、非线性支持向量机。当训练集线性可分，通过软间隔最大化学习的线性分类器为线性支持向量机。线性支持向量机是针对线性不可分的数据集的，这样的数据集可以通过近似可分的方法实现分类。对于这样的数据集，类似线性可分支持向量机，通过求解对应的凸二次规划问题，也同样求得分离超平面以及相应的分类决策函数。

1、求解原始问题的对偶问题

由上所述，我们得到线性支持向量机的原始问题：

$$\min_{w,b,\xi}\frac{1}{2}||\omega||^2+C\sum_{i=1}^{N}\zeta_i \tag{1}$$

$$s.t.\ y_i(\omega_i+b)>1-\zeta_i;i=1,2,...N \tag{2}$$

$$\zeta_i>=0;i=1,2,...,N \tag{3}$$

引入拉格朗日函数，利用拉格朗日函数的对偶性，将问题变成一个极大极小优化问题，得到了如下原始问题的对偶问题。

$$\min_{\partial}\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\partial_i\partial_j y_i y_j(x_i\cdot x_j)-\sum_{i=1}^{N}\partial_i \tag{4}$$

$$s.t.\sum_{i=1}^{N}\partial_i y_i=0 \tag{5}$$

$$0<=\partial_i<=C \tag{6}$$

然后设置惩罚参数C，并求解对偶问题，假设求得的最优解为∂_i^*；

2、计算原始问题的最优解

$$\omega^*=\sum_{i=1}^{N}\partial^* y_i x_i \tag{7}$$

选择∂^*中满足 0 的份量，计算：

$$b^*=y_j-\sum_{i-1}^{N}y_i\partial_j^*(x_i\cdot x_j) \tag{8}$$

3、计算原始问题的最优解：

分离超平面为：　$\omega^*.x+b^*=0$ (9)

分类决策函数为：　$f(x)=sign(\omega^* x+b^*)$ (10)

(3)模型的训练

通过上文的比较，发现线性支持向量机的平均准确率达到了 87.0%，因此本文就采用线性支持向量机分类模型对数据集进行划分训练集和测试集并训练。通过画出模型的学习曲线，可以直观地观察到模型的准确性与训练数据集大小的关系。最终实验表明当训练数据集占 67%时，算法的准确性较好。

	城乡建设	劳动和社会保障	教育文体	商贸旅游	环境保护	卫生计生	交通运输
总数据 100%	2009	1969	1588	1215	938	877	613
训练集 67%	1346	1319	1063	814	628	587	410
测试集 33%	663	650	525	401	310	290	203

图 6 训练集与测试集的划分

在模型训练方面，本文计算 cut_review 列的 TF-IDF 的特征值，TF 是词频，IDF 是逆文本频率指数。TF-IDF 是在词语计数的基础上，降低了常用高频词的权重,增加罕见词的权重，因为罕见词更能表达留言的主题思想。

本文使用 sklearn.feature_extraction.text.TfidfVectorizer 方法来抽取文本的 TF-IDF 的特征值。本文除了抽取评论中的每个词语外,还要抽取每个词相邻的词并组成一个“词语对”，如: 词 1，词 2，词 3，词 4，(词 1，词 2)，(词 2,词 3)，(词 3，词 4)。通过扩展本文特征集的数量,从而提高我们留言分类的准确度。

本文实验得出的 features 维度是(9209,817174),这里的 9209 表示我们总共有 9209 条评价数据,817174 表示我们的特征数量这包括全部评论中的所有词语数+词语对(相邻两个单词的组合)的总数。

下面本文通过**卡方检验**的方法来找出每个分类中关联度最大的两个词语和两个词语对。卡方检验是一种统计学的工具,用来检验数据的拟合度和关联度。在这里本文使用 sklearn 中的 chi2 方法。

图 7 每个分类中关联度最大的两个词语和两个词语对

通过实验结果可以看到经过卡方检验后，找出了每个分类中关联度最强的两个词和两个词语对。这些词和词语对能很好的反映出分类的主题。

模型训练完成后，使用划分的测试数据集对模型进行**测试**，测试结果使用混淆矩阵来展示，显示了预测标签和实际标签之间的差异。

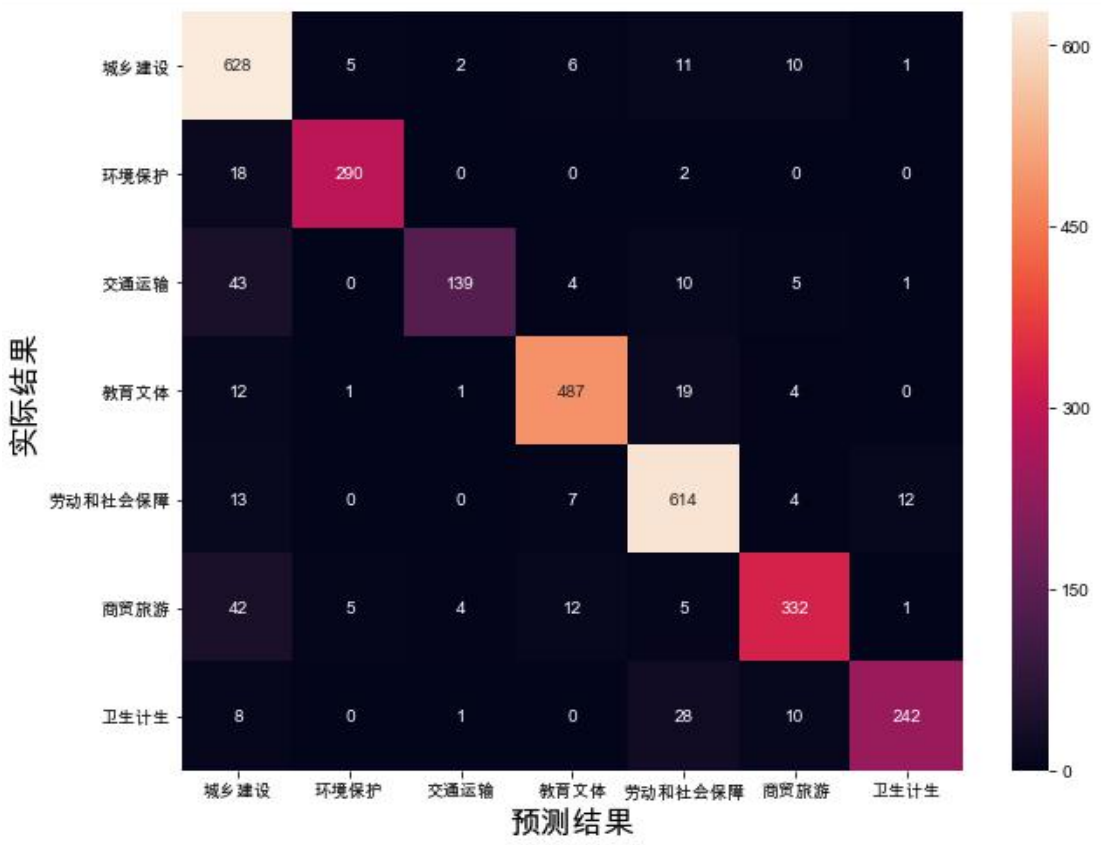


图 8 测试结果的混淆矩阵

通过上图描述，混淆矩阵的主对角线表示预测正确的数量,除主对角线外其余都是预测错误的数量。从混淆矩阵可以看出"环境保护"类预测最准确,只有 11 例预测错误。“城乡建设”和“劳动和社会保障”类别的预测错误数量较多。

4.1.4 模型的评价与优化

多分类模型一般不使用准确率(accuracy)来评估模型的质量,因为accuracy不能反应出每一个分类的准确性,因为当训练数据不平衡时，accuracy 不能反映出模型的实际预测精度。因此本文采用 F1 分数、ROC 等指标来评估模型。

通常使用 F-Score 对分类方法进行评价，F-Score 的定义如下：

$$F_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{2P_iR_i}{P_i+R_i} \tag{11}$$

其中 Pi 为第 i 类的查准率， Ri 为第 i 类的查全率。下面是查看各个类的 F1 分数，实验结果如下图所示。

accuracy 0.8989799276077657

	precision	recall	f1-score	support
城乡建设	0.82	0.95	0.88	663
环境保护	0.96	0.94	0.95	310
交通运输	0.95	0.69	0.80	202
教育文体	0.94	0.93	0.94	524
劳动和社会保障	0.89	0.94	0.92	650
商贸旅游	0.91	0.83	0.87	401
卫生计生	0.94	0.84	0.89	289
avg / total	0.90	0.90	0.90	3039

图 9 每个分类的 F1 分数

从上图 F1 分数上看,"环境保护"类的 F1 分数最大，“交通运输”类 F1 分数最差只有 80%，究其原因可能是因为“交通运输”分类的训练数据最少，只有 613 条,使得模型学习的不够充分,导致预测失误较多。

下面是预测失误的例子,希望能通过这些例子来改善本文的分类器，部分详细输出数据如下图所示。

环境保护 预测为 城乡建设 ： 18 例.

	cat	review
2160	环境保护	\n\t\t\t\t\t\n\t\t\t\t\t 我是一名在省会A市工作D市人之一，我和很...
2195	环境保护	\n\t\t\t\t\t\n\t\t\t\t\t尊敬的龚书记，你好！本人家住E市E2区宝庆西...
2099	环境保护	\n\t\t\t\t\t\n\t\t\t\t\t刘家巷，刘武昌按摩对面，搬来一冷库，太一市场...
2071	环境保护	\n\t\t\t\t\t\n\t\t\t\t\t2019年9月10日，B市生态环境局在《关于...
2041	环境保护	\n\t\t\t\t\t\n\t\t\t\t\t在含浦镇开心农场匝道附近，天气一放晴就大量烧...
2064	环境保护	\n\t\t\t\t\t\n\t\t\t\t\t尊敬的黎局长： 您好！ A3区施家港社区...
2337	环境保护	\n\t\t\t\t\t\n\t\t\t\t\t J市二医院家属区内，有人饲养家禽，臭气熏...
2067	环境保护	\n\t\t\t\t\t\n\t\t\t\t\tA市A4区雅居乐花园小区一楼商铺在环评报告明...

图 10 部分预测失误的例子详情

通过分析预测失误的例子详情和 F1-Score，本文提出了一些分类模型的改进措施。通过统计分析这些数据，本文将从以下几个方面对留言分类模型进行优化：优化分词算法、停用词列表，优化特征工程，增加有价值的特征和多项式特征，分类算法调优等等。然后再次训练本文的分类模型，从而本文的留言分类模型不断提高与完善。下面是每一类的数据的优化与改进措施。

表 3 模型优化改进的措施表

类别	错误率	F1_Score	改进措施
环境保护	4%	0.95	完善停用词列表
交通运输	5	0.80	增加该类训练样本
教育文体	6	0.94	完善停用词表
劳动和社会保障	11	0.92	完善停用词表
商务旅游	9	0.87	增加有价值的特征
卫生计生	6	0.89	增加该类训练样本
城乡建设	17	0.88	增加有价值的特征

4.2 问题二：热点问题挖掘

4.2.1 问题分析与识别

问题二旨在根据附件 3 数据， 将某一时段内反映特定地点或特定人群的热度问题的留言进行归类，并定义合理的热度评价指标，并汇总出热点问题。

根据文本内容的相似度，从而从众多的留言中找出那些相似度大的留言归为同一问

题。句子与句子之间的相似度计算，本文选取了“留言主题”来作为聚类的依据。随后利用基于文本相似度的聚类模型将相似度较大的留言归位同一类，然后**定义热度评价指标 H**，最后可利用改进 reddit 算法得到热点问题排名，再汇总成热点问题表和热点问题留言明细表。

4.2.2 聚类模型的建立

(1)数据预处理

1、 jieba 分词的自定义词典，本文添加的词典包含了常见的行业术语还有优化改进时新增的词语，格式是一行一个词，附件“dictionary.txt”。

2、 本文采用了常用 NLP 英文停用词文件“chineseStopWords.txt”，包含常见的数字、字符、基础语气词、代词、疑问词等等。由于停用词没有实际意义，因此本文将这些停用词过滤掉。

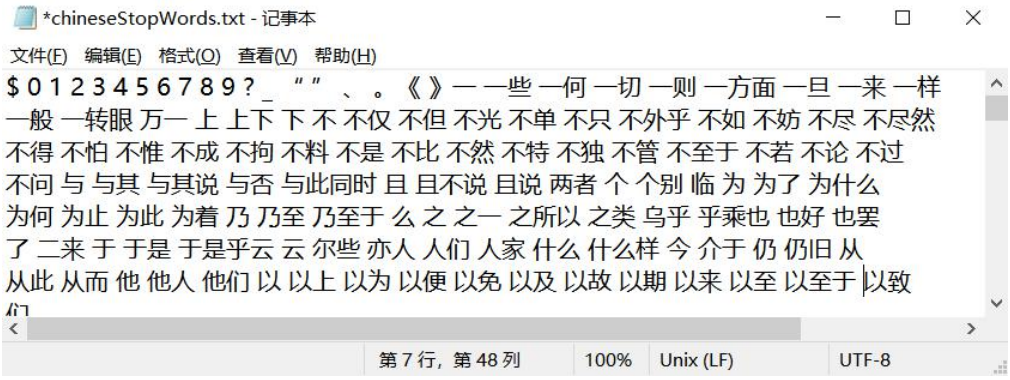


图 11 停用词列表

(2)相似度计算

本文增设了名称为“问题汇总”的 3 个 sheet 表头，分别为聚类问题编号、留言主题、留言次数、导航超链接”，其中“导航超链接”带超链接，可以导航到类的明细数据。

本文的相似度阈值设置为 50%。通过 jieba 分词进行处理，首先自定义词典以及排除信息才能 jieba 分词，然后形成一个二维数组。使用 gensim 中的 corpora 模块，将分词形成后的二维数组生成词典。将二维数组通过 doc2bow 稀疏向量，形成语料库。本文这里不使用 TF 模型算法，采用的是 LsiModel 模型算法，将语料库计算出 TF-IDF 值。然后获取词典 token2id 的特征数。计算稀疏矩阵相似度，建立一个索引。最后，通过 jieba 进行分词处理。通过 doc2bow 计算测试数据的稀疏向量。从而求得测试数据与样

本数据的相似度。

（3）聚类结果

根据照留言次数排序将结果写入目标 excel 表，整理出特定格式 result.xlsx 表格，部分数据如下图所示。

聚类问题编	留言主题	留言次数	导航超链接
1	A市地铁3号线和4号线出入口	179	A市地铁3号线和4号线出入口、站点选址的设置极不
2	A市多路公交车发车太不准时	165	A市多路公交车发车太不准时（例如：205、225、2
3	A2区暮云段道路改造期间，	122	A2区暮云段道路改造期间，造成交通出行不便、噪
4	建议加快A市国家中心城市建	98	建议加快A市国家中心城市建设进度且增强大气污染
5	A7县多地麻将馆扰民	87	A7县多地麻将馆扰民
6	A市温斯顿楚府英语机构恶意	72	A市温斯顿楚府英语机构恶意拖欠退款
7	A7县金科时代小区收取不合理	61	A7县金科时代小区收取不合理的物业费，而且小区
8	A7县楚龙街道红树湾小区的	51	A7县楚龙街道红树湾小区的消防设施瘫痪，安全隐
9	A3区枫林三路和中海国际社区	41	A3区枫林三路和中海国际社区夜间施工扰民
10	A7县松雅湖滨湖路和灰埠路_	32	A7县松雅湖滨湖路和灰埠路上飙车党扰民严重
11	A2区A1区南路夜间大货车飞	25	A2区A1区南路夜间大货车飞驰鸣笛
12	A7县星沙时代星城小区内摆摊	22	A7县星沙时代星城小区内摆摊设点严重
13	建议加快A市一圈二场三道步	19	建议加快A市一圈二场三道步伐力度
14	A7县黄兴镇卫生院剥夺职工休	12	A7县黄兴镇卫生院剥夺职工休息时间
15	A1区A2区华庭地下车库要变	10	A1区A2区华庭地下车库要变成垃圾场了

图 12 聚类结果图

4.2.3 聚类算法性能评估与优化

本文针对上述模型对聚类算法进行性能评估和优化，通过轮廓系数可以很好的衡量模型性能，轮廓系数可以在不需要已标记的数据集的前提下，对聚类算法的性能进行评估。轮廓系数由以下两个指标构成。针对本文数据，其轮廓系数 s 的值为:

$$s = \frac{b-a}{\max(a,b)} \tag{12}$$

针对本文数据集，其轮廓系数 s 为其所有样本的轮廓系数的平均值，轮廓系数的数值于介于[-1，1]之间。-1 表示完全错误的聚类，1 表示完美的聚类，0 表示聚类重叠。

通过计算本文聚类算法性能评估指标，输出得到本文的聚类算法的平均轮廓系数得分：

```
When cluster =589
The silhouette_score=0.7015
```

图 13 聚类算法的平均轮廓系数得分

由上图得知该聚类模型的聚类总体效果不错，也有需要优化的地方，例如：查阅附件 3 的原始数据，检查一下如果通过人工标记是否能够标记出某些留言的正确分类。也可以通过调整聚类的数量来让聚类效果达到更好。

4.2.4 根据热度排名算法，得出热点问题

(1)定义热度评价指标 H

通过对大量文献研究与分析，本文选取了留言问题 4 个有用的参数，分别为点赞数、反对数、留言次数、话题持续时间。通过改进 reddit 热度排名算法，定义出热度评价指标公式：

$$H = \mathrm{w1} * (x - y) + w2 * z + w3 * \log_{10}(t) \tag{13}$$

其中 x 为点赞数，y 为反对数，z 为留言次数，t 为问题持续时间。通过大量的查阅文献和对现有的数据分析统计，本文定义了合理的权重参数，分别为 w1=0.35，w2=0.45，w3=0.20 。定义热度评价指标 H 能够很好地反映了问题的热度，下图为聚类问题的几个参数统计情况，参见附件中“参数汇总”表。

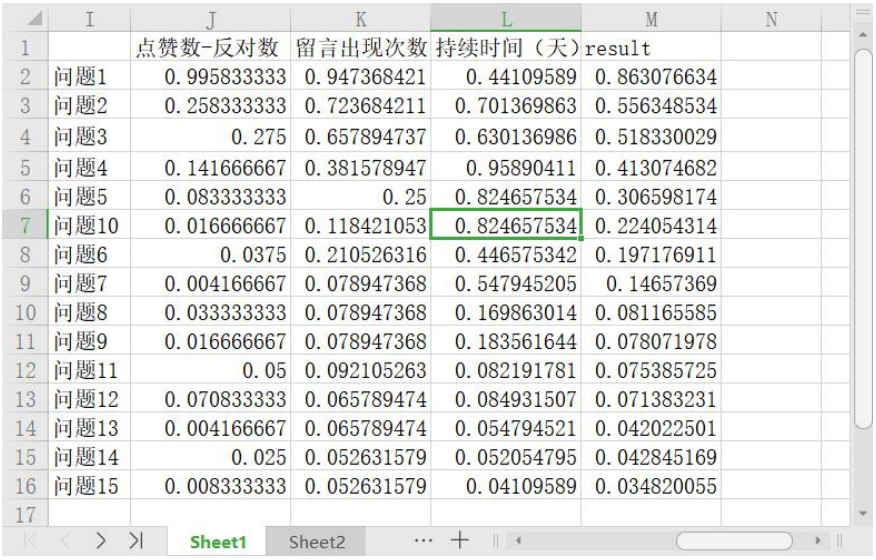
	A	B	C	D	E	F	G
1		点赞数	反对数	点赞数-反对数	留言出现次数	持续时间（天）	
2	问题1	244	5	239	72	161	
3	问题2	62	0	62	55	256	
4	问题3	73	7	66	50	230	
5	问题4	38	4	34	29	350	
6	问题5	21	1	20	19	301	
7	问题6	10	1	9	16	163	
8	问题7	1	0	1	6	200	
9	问题8	5	1	8	6	62	
10	问题9	5	1	4	6	67	
11	问题10	4	0	4	9	301	
12	问题11	12	0	12	7	30	
13	问题12	17	0	17	5	31	
14	问题13	1	0	1	5	20	
15	问题14	6	0	6	4	19	
16	问题15	2	0	2	4	15	
17							

Sheet1 Sheet2

图 14 聚类问题的参数统计情况

(2)计算并得出热点问题

为了规范数据，本文把相应的数据映射到 0～1 范围之内处理，归一化后得到以下排名结果，如图 15 所示。



	I	J	K	L	M	N
1		点赞数-反对数	留言出现次数	持续时间（天）	result	
2	问题1	0.995833333	0.947368421	0.44109589	0.863076634	
3	问题2	0.258333333	0.723684211	0.701369863	0.556348534	
4	问题3	0.275	0.657894737	0.630136986	0.518330029	
5	问题4	0.141666667	0.381578947	0.95890411	0.413074682	
6	问题5	0.083333333	0.25	0.824657534	0.306598174	
7	问题10	0.016666667	0.118421053	0.824657534	0.224054314	
8	问题6	0.0375	0.210526316	0.446575342	0.197176911	
9	问题7	0.004166667	0.078947368	0.547945205	0.14657369	
10	问题8	0.033333333	0.078947368	0.169863014	0.081165585	
11	问题9	0.016666667	0.078947368	0.183561644	0.078071978	
12	问题11	0.05	0.092105263	0.082191781	0.075385725	
13	问题12	0.070833333	0.065789474	0.084931507	0.071383231	
14	问题13	0.004166667	0.065789474	0.054794521	0.042022501	
15	问题14	0.025	0.052631579	0.052054795	0.042845169	
16	问题15	0.008333333	0.052631579	0.04109589	0.034820055	
17						

Sheet1 Sheet2

图 15 问题归一化处理

由于得到的排名结果不能很直观地反映问题的热度，因此本文根据归一化后得到的结果再划分热度指数星级，规则如下：

表 4 数据清洗

热度指数星级	result 得分
5 星级	0.51-1.00
4 星级	0.31-0.50
3 星级	0.11-0.30
2 星级	0.051-0.1
1 星级	0.020-0.050

通过上述星级定义，整理得到热点问题表、热点问题留言明细表，本文选取了热点问题表中热度指数前 5 名的热点问题，如下表所示。

表 5 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	5 星级	2019/3/25 - 2019/9/6	A 市地铁 3 号线和 4 号线各站点	A 市地铁 3 号线和 4 号线出入口、站点选址的设置极不合理，建议解决噪音扰民和增加扫码进站设备问题
2	2	5 星级	2019/2/12 - 2019/10/28	A 市多路公交车	A 市多路公交车发车太不准时（例如：205、225、208、808 路）
3	3	5 星级	2019/1/14 - 2019/8/4	A2 区暮云段道路	A2 区暮云段道路改造期间，造成交通出行不便、噪音扰民等问题

4	4	4 星级	2019/1/10 - 2019/12/30	A 市国家中心城市	建议加快 A 市国家中心城市建设进度且增强大气污染防治工作的力度
5	5	4 星级	2019/2/17 - 2020/1/7	A7 县多个镇麻将馆	A7 县多地麻将馆扰民

本文对所有留言聚类成不同的问题，根据聚类结果看可以把附件 3 数据聚成 590 个不同的问题，在此本文还列出了排名前 15 的热点问题，如下图所示。

	A	B
5	热点排名	热点问题
6	6	A市温斯顿楚府英语机构恶意拖欠退款
7	7	A7县金科时代小区收取不合理的物业费，而且小区停车难，强制收取停车费
8	8	A7县楚龙街道红树湾小区的消防设施瘫痪，安全隐患多
9	9	A3区枫林三路和中海国际社区夜间施工扰民
10	10	A7县松雅湖滨湖路和灰埠路上飙车党扰民严重
11	11	A2区A1区南路夜间大货车飞驰鸣笛
12	12	A7县星沙时代星城小区内摆摊设点严重
13	13	建议加快A市一圈二场三道步伐力度
14	14	A7县黄兴镇卫生院剥夺职工休息时间
15	15	A1区A2区华庭地下车库要变成垃圾场

图 16 排名前 15 的热点问题

4.3 问题三：答复意见的评价

在附件 4 中，给出了相关部门对留言的答复意见，在文本挖掘中，答复意见的质量决定了政府相关部门对基层群众反映问题的办事效率以及解决能力，为确定政府相关部门在解决基层群众反映的相关问题的解决能力，本文针对附件 4 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、答复效率四个角度对政府部门对留言的答复意见质量进行衡量，并进行量化评分，构建起模糊综合评价模型并对附件 4 数据加以实现。

4.3.1 模型准备

（1）相关性

在中文文本挖掘中，由于中文文化博大精深，一句话可以由多个句子进行表达，所以在处理文本间文本之间的相关性是一件比较复杂的事情，为判断政府部门答复与群众留言之间的相关性，本文采用文本相似度的方式来对相关性进行量化评分。并给出评分

方案，即基于文本相似度的计算情况下，文本相关性评分为 1，文本不相关性评分为 0.

在处理与计算文本之间的相似度上，本文采用基于 TF-IDF 算法的长文本相似度计算。TF-IDF 是一种基于统计的算法，以评估某字词在一个文本集合或一个语料库中重要程度。字词的重要性随着它在文件中出现的次数成正比增加，但同时会随着它在语料库中出现的频率成反比下降。对于该算法中，有两个相对重要的两个概念：词频(TF)、逆向文件频率(IDF)，其中词频(TF)为某一个给定的词在某文本中出现的次数，逆向文件频率(IDF)为文件夹中可能存在包含多个文档，如果包含某词语的文档越少，则 IDF 越大，则说明某词语有很好的代表作用。其中，词频(TF)和逆文档频率(IDF)的计算公式如下：

$$\begin{cases} TF = \dfrac{\text{某一文件中某词语}\omega\text{出现次数}}{\text{文本中所以词语数目}} \\ IDF = \log\dfrac{\text{语料库的文档总数}}{\text{包含某词}\omega\text{的文档数}+1} \end{cases} \tag{14}$$

通过上述算法的描述，可以得出 TF-IDF 的计算公式，其中该值越大，则代表某个词对文章的重要性就越高。

$$TF-IDF = \text{词频}(TF) * \text{反文档频率}(IDF) \tag{15}$$

文本相关性的计算

通过上述 IF-IDF 算法的描述，文本针对于基层群众留言详情数据列与政府相关部门答复数据列进行 IF-IDF 计算，为了方便该数据列的文本区分，本文首先利用 jieba 算法对给该数据列进行关键词统计和中文分词。关键词统计和分词效果如下图所示：

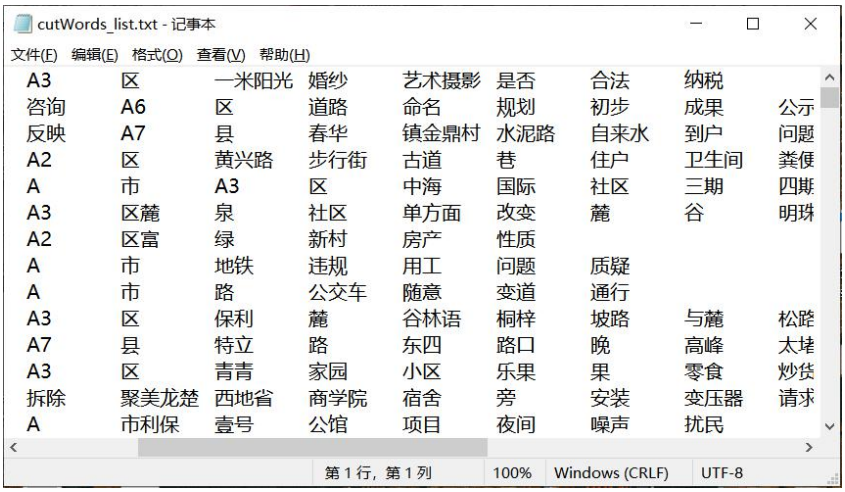


图 17 留言详情分词结果图

完成针对于基层群众留言详情数据列与政府相关部门答复数据列的中文分词后，本文通过 TF-IDF 算法计算出 TF 以及 IDF 的词频向量，进而确定基层群众留言详情数据列

与政府相关部门答复数据列对应的 TF-IDF 值，由于 IF-IDF 值经过向量化处理，为了更方便直观看出文本间的相似性，本文将 TF-IDF 值进行单位化处理，最终得到文本相关性的评价模型，即文本间相关，评分 1，文本间不相关，评分 0。

$$\begin{cases} 1, \text{文本间相关} \\ 0, \text{文本间不相关} \end{cases} \tag{16}$$

根据文档相关性数据，可以得出基层群众留言详情数据列与政府相关部门答复数据列的相关性，总数据为：2816，相关性总量：2761，不相关性总量：55，相关率为：98%；其各得分的数据量与占比如图所示：

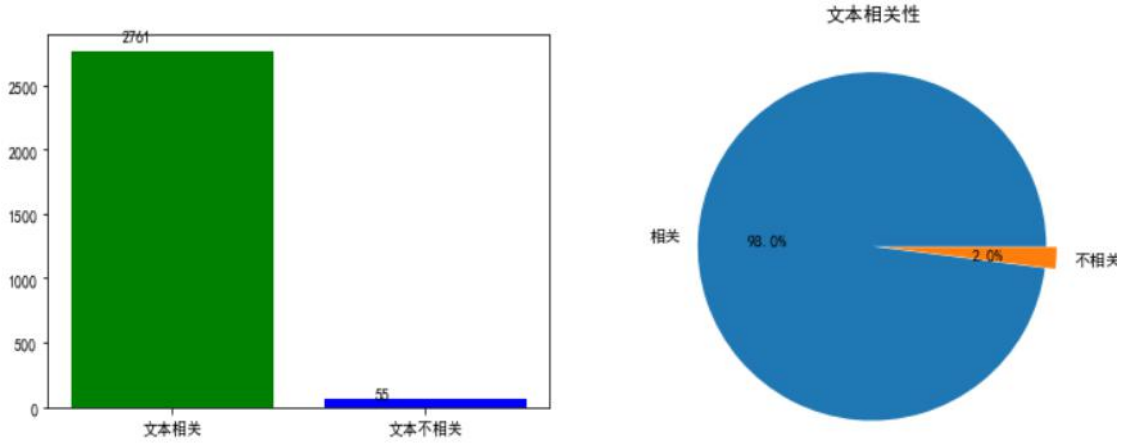


图 18 文本相关性绘制图

（2）完整性

在中文文本挖掘中，文本间的完整性也是衡量文本质量的一个好的衡量指标，为能反映出政府相关部门答复数据列的完整性，通过查询相关资料，本文从四个指标对府相关部门答复数据列的完整性进行量化衡量，即问候语匹配、相关性匹配、处理问题告知语匹配、结束语匹配。在处理政府相关部门答复数据列的完整性中，若能匹配到这四个指标其中一个，得 1 分，否则得 0 分。

通过上述描述，本文首先利用 jieba 对基层群众留言详情数据列与政府相关部门答复数据列进行中文分词，为了能够匹配和区分到上述四个指标的词语，本文在语料库中加入问候语(如：您好，你好等)、处理问题告知语(如：解决、协商、上交等)、结束语(感谢、理解、关心等)。在完成中文分词后，本文通过上述 TF-IDF 算法计算政府相关部门答复数据列针对于这四个指标的相关得分并计算总分，其中满分为 4 分。

根据文档完整性数据，可以得出政府相关部门答复数据列的完整性，总数据为：2816，完整性得分为 4：2499 份，占比为 88.7%，完整性得分为 3:295 份，占比为：10.5%，

完整性得分为 2:21 份，占比为：0.7%，完整性得分为 1:1 份，占比为：0.035%，其各得分的数据量与占比如图所示：

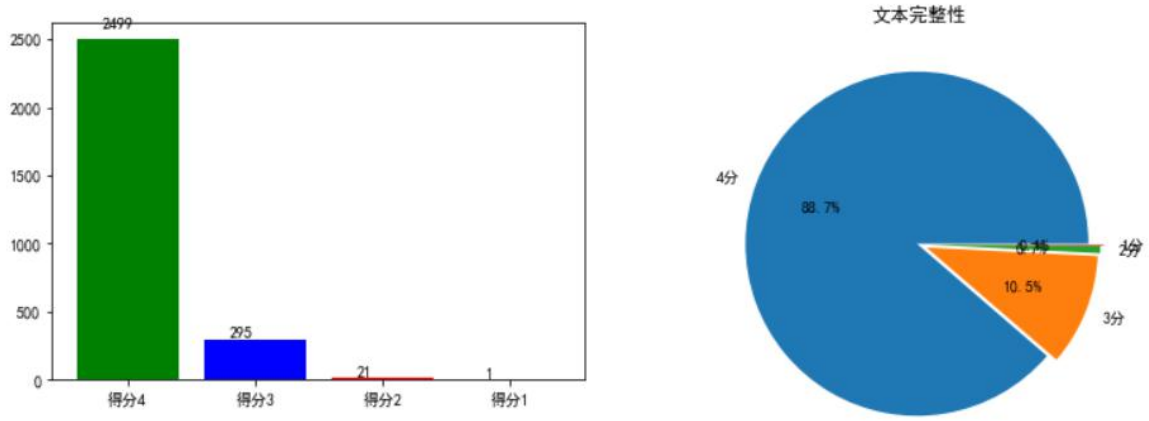


图 19 文本完整性得分汇总图

（3）可解释性

在中文文本挖掘，文本的可解释性也是衡量文本质量的一个好的衡量指标，可解释性是指回复的内容是否有依据可言。为能反映出政府相关部门答复数据列的可解释性，通过查询相关资料，可以得知政府回复群众留言中，都会涉及法律法规解释，除此之外还有不触犯法律法规的，即需要政府针对相关部门出门进行调查调节的问题。本文从两个指标对府相关部门答复数据列的完整性进行量化衡量，即法律法规匹配，调查、调解用语匹配。在处理政府相关部门答复数据列的可解释性中，若能匹配到这两个指标其中一个，得 1 分，否则得 0 分。

通过上述描述，本文首先利用 jieba 对基层群众留言详情数据列与政府相关部门答复数据列进行中文分词，为了能够匹配和区分到上面两个指标的词语，本文在语料库中加入法律法规，调查、调解用语。在完成中文分词后，本文通过上述 TF-IDF 算法计算政府相关部门答复数据列针对于这两个指标的相关得分并计算总分，其中满分为 2 分。

根据文档可解释性数据，可以得出政府相关部门答复数据列的可解释性，总数据为：2816，可解释性得分为 2：2557 份，占比为 91%，可解释性得分为 1:252 份，占比为：8.9%，可解释性得分为 0：7 份，占比为：0.1%，其各得分的数据量与占比如图所示：

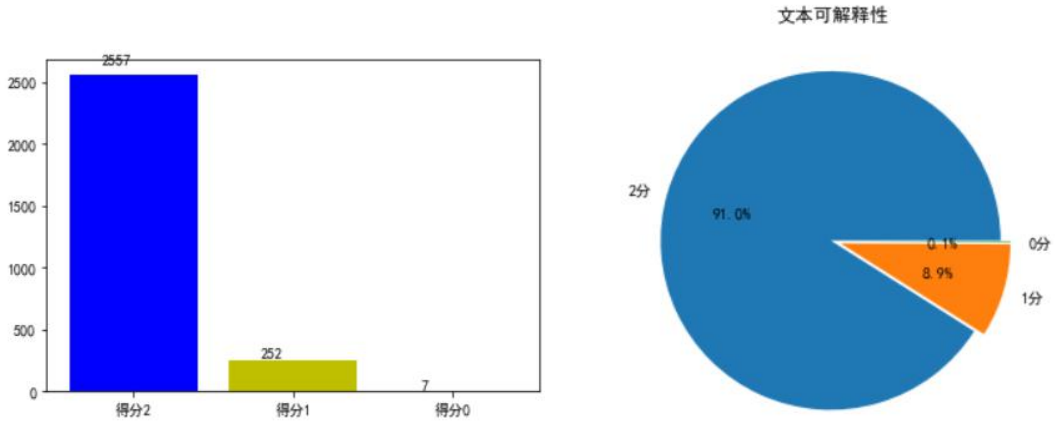


图 20 文本可解释性得分汇总图

（4）答复效率

政府相关部门是为人民服务的政府，政府的办事效率也可以直接影响人民的幸福指数，政府回复基础群众问题留言时，答复回复所需时间也决定了政府相关部门的办事效率。通过相关资料查询，可以得知在政府答复群众留言中，一般根据政府答复群众留言时间间隔(以天为单位)为基准来判断政府的办事效率。在判定政府的办事效率上，本文采用计分制来判断政府答复效率，不同时间间隔区间对应的分值如下，其中 x 表示答复时间间隔：

$$\begin{cases} 4, & 0 \le x \le 5 \\ 3, & 5 < x \le 15 \\ 2, & 15 < x \le 30 \\ 1, & 30 < x \end{cases} \tag{17}$$

通过上述描述，本文通过 MATLAB 工具对群众留言时间列与政府相关部门答复时间列进行时间间隔计算，通过上述模型对每条答复进行答复效率评分。最后将得分情况输出到 excel 表格上，其中满分为 4 分。

根据文档答复效率性数据，可以得出政府相关部门答复数据列的答复效率，总数据为：2816，答复效率得分为 4：740 份，占比为 26.3%，答复效率得分为 3:992 份，占比为：35.2%，答复效率得分为 2：623 份，占比为：22.1%，答复效率得分为 2：461 份，占比为：16.4%，其各得分的数据量与占比如图所示：

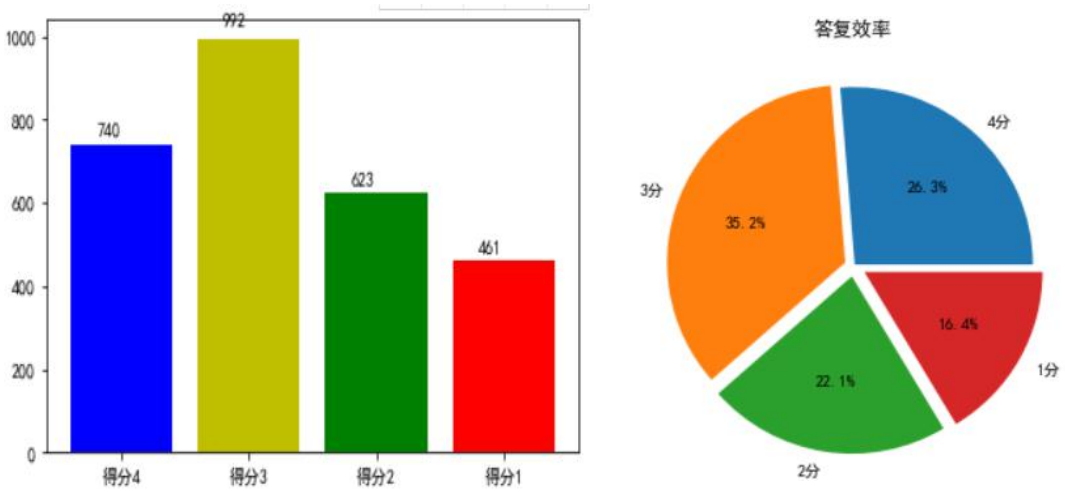


图 21 答复效率得分汇总图

4.3.2 模型的建立与实现

为了能更好地对答复意见的质量作出评价，在上文的分析统计基础上，建立起模糊综合评价模型，并对附件 4 的数据加以实现。

（1）建立综合评价的因素集

因素集是以影响评价对象的各种因素为元素所组成的一个普通集合，通常用 U 表示，$U=(u_1,u_2...,u_m)$，其中元素 u_i 代表影响评价对象的第 i 个因素。这些因素，通常都具有不同程度的模糊性。

本文的等级的指标集为 $U=(u_1,u_2,u_3,u_4)$，u_1 表示为相关性，u_2 表示为完整性，u_3 表示可解释性，u_4 表示为答复效率。

（2）建立综合评价的评价集

评价集是评价者对评价对象可能做出的各种结果所组成的集合，通常用 V 表示 $V=(v_1,v_2...,v_m)$，其中元素 v_j 代表第 j 种评价结果，可以根据实际情况的需要，用不同的等级、评语或数字来表示。

本文的等级的评价集为 $V=(v_1,v_2,v_3,v_4)$，v_1,v_2,v_3,v_4 分别表示很好、较好、一般、不好。

（3）进行单因素模糊评价，获得评价矩阵

若因素集 U 中第 i 个元素对评价集 V 中第 1 个元素的隶属度为 r_{i1}，则对第 i 个元素

单因素评价的结果用模糊集合表示为：$R_i=(r_{i1},r_{i2}...,r_{in})$，以 m 个单因素评价集 $R_1,R_2...R_m$ 为行组成矩阵 R_{m*n} 称为模糊综合评价矩阵。

通过以上算法得到单因素评判矩阵 R:

$$R=\begin{bmatrix}0.35 & 0.39 & 0.22 & 0.04\\ 0.17 & 0.35 & 0.39 & 0.09\\ 0 & 0.30 & 0.44 & 0.26\\ 0.09 & 0.22 & 0.30 & 0.39\end{bmatrix}$$

（4）确定因素权向量

评价工作中，各因素的重要程度有所不同，为此，给各因素 u_i 一个权重 a_i，各因素的权重集合的模糊集，用 A 表示 $A=(a_1,a_2,a_3,a_4)$。本文通过层次分析法 AHP 的成对比较阵来构造因素权向量，得到的权数向量为：$A=(0.28,0.25,0.27,020)$。

（5） 建立综合评价模型

确定单因素评判矩阵 R 和因素权向量 A 之后，通过模糊变化将 U 上的模糊向量 A 变为 V 上的模糊向量 B，即 $B=A_{1*m}$。 其中 $R_{m*n}=(b_1,...,b_n)$ 称为综合评价合成算子,本文取成一般的矩阵乘法即可。

本文从答复的相关性、完整性、可解释性等角度 4 个参数中组合了一个“答复意见总得分 E”指标，通过这个指标能很好地反映附件中每条答复意见的质量。

答复意见总得分 E 的计算公式：

$$E=u_1*a_1+u_2*a_2+u_3*a_3+u_4*a_4 \qquad (18)$$

其中 $A=(0.28,0.25,0.27,020)$，评定留言答复意见的等级的指标集为 $U=(u_1,u_2,u_3,u_4)$，u_1 表示为相关性，u_2 表示为完整性，u_3 表示我可解释性，u_4 表示为答复效率。

建立起模糊综合评价模型后，并对附件 4 的数据加以实现。最后根据答复意见总得分进行划分评价等级，本文选取的评价等级有：很好、较好、一般、不好，划分的规则如下：

答复意见总得分	答复等级
0.95-1.00	很好
0.80-0.95	较好
0.65-0.80	一般
0-0.65	不好

得到了附件 4 的所有留言答复的“答复意见总得分”和“答复等级”，部分数据如下图所示，完整数据见“答复意见总得分表”。

	A	B	C	D	E	F	G	H	I
1	留言详情	答复意见	相关性得	完整性得	可解释	答复效	答复意见总得分		答复等级
2	2019年4月以	现将网友在平	1	0.75	1	0.5	0.8375		较好
3	潇楚南路从2(	网友“A00023	1	1	1	0.75	0.95		很好
4	地处省会A市	市民同志：你	1	1	1	0.75	0.95		很好
5	尊敬的书记：	网友“A00011	1	1	1	0.75	0.95		很好
6	建议将“白竹	网友“A00092	1	1	1	0.5	0.9		较好
7	欢迎领导来A	网友“A00077	1	0.75	1	0.25	0.7875		一般
8	尊敬的胡书记	网友“A00010	1	1	1	0.25	0.85		较好
9	我做为一东澜	网友“UU0081	1	0.5	1	0.5	0.775		一般
10	我是美麓阳光	网友“UU0087	1	1	1	0.5	0.9		较好
11	胡书记好！根	网友“UU0086	1	0.75	0.5	0.5	0.7025		一般
12	我家住在A市	网友“UU0082	1	1	1	0.25	0.85		较好
13	胡书记：您好	网友“UU0088	0	0.75	1	0.25	0.5075		不好
14	尊敬的书记：	网友“UU0087	1	1	1	0.5	0.9		较好
15	尊敬的领导：	网友“UU0081	1	0.75	0.5	0.75	0.7525		一般
16	建议增开A市2	网友“UU0081	1	0.75	1	0.5	0.8375		较好
17	2016年下半年	网友“UU0084	1	1	1	0.25	0.85		较好
18	12月16日上午	网友“UU0081	1	1	1	0.75	0.95		很好
19	梅溪湖至今没	网友“UU0087	1	0.75	1	0.5	0.8375		较好
20	希望相关部门	网友“UU0082	1	1	1	0.75	0.95		很好
21	看病需要带社	网友“UU0081	1	0.75	0.5	0.75	0.7525		一般
22	希望潇楚一卡	网友“UU0081	1	1	1	0.75	0.95		很好
23	A9市北盛镇对	网友“UU0081	1	1	1	0.75	0.95		很好
24	尊敬的市委、	网友“UU0081	1	1	1	0.5	0.9		较好

答复意见总得分表 Sheet2 Sheet3

图 22 答复意见总得分和等级情况

本文通过构建模糊综合评价模型，对附件 4 的留言答复意见的质量构造了“答复意见总得分 S”和“答复等级 L”两个指标，能够很好地衡量留言意见的质量。通过下图的分析统计，可以了解到对于附件 4 的留言答复意见的等级分布。数据表明有 52%的留言答复等级为“很好”，有 42%的留言答复等级为“较好”，有 5.7%的留言答复等级为“一般”，有 0.3%的留言答复等级为“不好”。这也反映了该智慧政务系统的留言答复总体情况较好，有少数留言答复在相关性、完整性、可解释性、答复效率的某个方面有待提高。

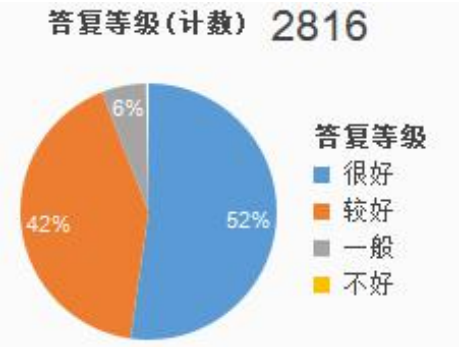


图 23 留言答复意见等级情况图

5 模型的评价

对于本文问题一构建的分类模型，充分比较了多项式朴素贝叶斯、线性支持向量机、逻辑回归、随机森林四种模型，选择了分类效果最佳的线性支持向量机作为留言分类模型。并通过 F-Score 评价体系对模型优化提出相关建议。

对于本文问题二构建的聚类模型，利用了文本相似度来构建聚类模型，聚类效果较好，同时使用了轮廓系数来衡量聚类模型的性能，最后能够根据实际背景来改进 Reddit 热度排名算法对聚类问题进行热度评价，得出较好的热度排名。

对于本文问题三构建的模糊综合评价模型，能够从相关性、完整性、可解释性和答复效率 4 个角度对政府部门的答复意见质量进行综合评价，并对附件 4 的数据加以实现，得到了良好的综合评价效果。

6 结论

随着网络问政平台的搭建，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。

本文建立了基于自然语言处理技术和文本挖掘技术的模型，有效解决了这些难题。

对于平台群众的留言信息，本文建立了基于机器学习的线性支持向量机分类模型，能有效地解决留言划分和热点整理的相关难题，同时也对模型进行了测试和评价，效果较好。

挖掘热点问题也是网络问政平台很重要的流程，本文建立了基于文本相似度的聚类模型，能较准确地挖掘出当下热点问题，并根据热度指数对挖掘的留言问题进行排名，能高效找出当下的热点问题，以便及时为热点问题提供相应的解决措施。

最后，本文构建了模糊综合评价模型，能够从相关性、完整性、可解释性和答复效率 4 个角度对政府部门的答复意见质量进行综合评价，并对附件 4 的数据加以实现，得到了良好的综合评价效果。

7 参考文献

[1] 高明霞,李经纬.基于 word2vec 词模型的中文短文本分类方法.山东大学学报. 2019(02).

[2] 王根生,黄学坚.基于 Word2vec 和改进型 TF-IDF 的卷积神经网络文本分类模型[J]. 2019(05).

[3] 张波,黄晓芳.基于 TF-IDF 的卷积神经网络新闻文本分类优化[J].西南科技大学计算机科学与技术学院.2016.

[4] 李海林,邹金串.基于分类词典的文本相似性度量方法.华侨大学信息管理系.2015.

[5] 毛郁欣,邱智学.基于 Word2Vec 模型和 K-Means 算法的信息技术文档聚类研究.浙江工商大学管理工程与电子商务学院.

[6] 范宁.基于文本挖掘在民宿满意度中的研究[J].广西师范大学.201906.

[7] 薛曌.基于文本相似度的主观题自动评分系统的设计与实现[J].北京邮电大学.20190603.

[8] 廉素洁.基于文本分类和情感评分的电信投诉文本挖掘研究[J].浙江工商大学.201812.

[9] 陈立孚,周宁,李丹.基于机器学习的自动文本分类模型研究[J].现代图书情报技术,2005(10):23-27.

[10] 黄晓斌,赵超.文本挖掘在网络舆情信息分析中的应用[J].情报科学,2009,27(01):94-99.

[11] 黄永昌.scikit-learn 机器学习-常用算法原理及编程实践[M].北京：机械工程出版社.