

智慧政务中的文本挖掘应用

摘要

本文旨在基于互联网公开来源的群众问政留言记录以及相关部门对部分群众留言的答复意见，通过自然语言、数据分析和数据深入挖掘，解决群众留言分类、热点问题分类排序、答复意见评价等问题，为政务人员处理网络群众留言做探讨分析。

针对任务一，首先建立训练集测试集，采用贝叶斯算法对模型进行训练，再用训练好的模型对数据进行测试，并使用 F-score 分类方法对测试模型进行评价，从而实现群众留言一级分类。

针对任务二，本文构造了相似问题识别算法，通过多次识别，找到了某段时间内群众反映较多的问题。

针对任务三，本文使用 Python 建立了基于留言答复相关性、完整性、可解释性等相关建立了相应评价体系程序，通过计算表明，评价体系良好。

关键词：贝叶斯；预测；Python；识别；模型

Abstract

This article aims to solve the problems of the classification of mass messages, the classification of hot issues, the evaluation of answers, etc. through natural language, data analysis and in-depth data mining based on the public questioning message records of the public on the Internet and the response opinions of relevant departments on some mass messages , To discuss and analyze for the administrative staff to deal with the online mass messages.

For task one, first establish a training set test set, use Bayesian algorithm to train the model, and then use the trained model to test the data, and use the F-score classification method to evaluate the test model, so as to achieve the mass message 1 Level classification.

Aiming at task two, this paper constructs a similar problem recognition algorithm. Through multiple recognitions, it finds the problem that the masses reflect more in a certain period of time.

For task three, this article uses Python to establish the corresponding evaluation system procedures based on the relevance, completeness, and interpretability of message reply, and calculations show that the evaluation system is good.

Keywords: Bayesian; prediction; Python; recognition; model

目录

一、问题分析.....	4
二、任务一.....	4
2.1 基于 Python 的分词过滤	4
2.2 基于 Python 的预测模型	5
三、任务二.....	6
四、任务三.....	8
七、参考文献.....	9

一、问题分析

近年来，随着网络问政平台逐渐成为政府了解民意、汇聚民智、凝聚民气的重要渠道，因此，各类问政平台相关文本数据不断攀升。对以往采用人工进行留言分来和热点整理的工作人员带来极大困难。利用大数据、人工智能、云计算等技术实现网络留言分类和整理，有助于减少工作人员效率和提高问题解决的时效性，进一步提高网络留言优化管理。

问题给出了需要处理的表格数据，包括留言分类标准、留言编号、留言用户、留言主题、留言时间、留言详情、答复意见、答复时间、点赞数和反对数等指标。

任务一是要求我们建立模型去实现附件2中给出的一级分类标准，以使工作人员能够使用建立了的模型对网络留言进行分类。

任务二要求我们将附件3中某一时段群众集中反映的问题进行归类，并给出排名前5的问题（热点问题），以便相关部门能够及时解决。

任务三要求我们选取一些参数，对附件4相关部门的答复意见建立一套评价体系，以评判相关部门的工作。

二、任务一

2.1 基于 Python 的分词过滤

由于文本中含有大量无分类信息无关的文字和符号，像“嗯”这类词，它们对于程序编程没有实用价值。使用 Python 对留言主题和留言详情进行分类时，需要将这类无关此过滤掉。

首先，Python 中导入分词模块 jieba 库后，调用 jieba 库即可完成留言详情分词，之后再导出我们事先给出的停用词库 stoplist.txt，找出留言详情中与一级标签有关联的词。

留言编号	留言用户	留言主题	留言时间	留言详情	一级标签
24	A00074011	A市西湖建筑集团 古渔渔工有安全...	2020/1/6 12:09:38	...	城乡建设
37	U0005473	A市在水一方大厦 人为烂尾多年，...	2020/1/4 11:17:46	...	城乡建设
83	A00053995	投诉A市A1区药物 违法收停车费	2019/12/28 17:05:14	...	城乡建设
303	U0007137	A1区蔡岗南路A2 区违建房屋水箱...	2019/12/6 14:49:14	少的用品，要是 一种致病菌，我们住在这里连基本的健康保障都没有，请政府街道各领导	
319	U0007137	A1区A2区违建自 来水供水一报装	2019/12/5 11:17:22	...	城乡建设
379	A00016773	投诉A市盛世麒麟 小区物业管理...	2019/11/28 9:08:38	...	城乡建设
382	U0005806	咨询A市特盘集中 供暖一事	2019/11/27 17:14:11	...	城乡建设
445	A00019205	A2区桐梓板西路 市小城管中队	2019/11/19 22:34:36	...	城乡建设

图 2.1 留言详情

留言编号	留言用户	留言主题	留言时间	留言详情	一级标签
24	A00074011	A市西湖建筑集团 古渔渔工有安全...	2020/1/6 12:09:38	...	城乡建设
37	U0005473	A市在水一方大厦 人为烂尾多年，...	2020/1/4 11:17:46	...	城乡建设
83	A00053995	投诉A市A1区药物 违法收停车费	2019/12/28 17:05:14	...	城乡建设
303	U0007137	A1区蔡岗南路A2 区违建房屋水箱...	2019/12/6 14:49:14	少的生活必需品，是一种致病菌，健康保障，政府街道领导，请政府街道领导，健康生活环境	
319	U0007137	A1区A2区违建自 来水供水一报装	2019/12/5 11:17:22	...	城乡建设
379	A00016773	投诉A市盛世麒麟 小区物业管理...	2019/11/28 9:08:38	...	城乡建设
382	U0005806	咨询A市特盘集中 供暖一事	2019/11/27 17:14:11	...	城乡建设
445	A00019205	A2区桐梓板西路 市小城管中队	2019/11/19 22:34:36	...	城乡建设

图 2.2 留言详情分词过滤

Index	Type	Size	
0	str	1	】
1	str	1	【
2	str	1	,
3	str	1	!
4	str	1	会
5	str	1	的
6	str	1	日

图 2.3 分词停用词

2.2 基于 Python 的预测模型

对上述完成分词并过滤掉无关信息后的留言详情建立训练集测集 `train_label` 采用贝叶斯 `MultinomialNB` 对训练集测集进行训练，等待训练完成后对留言详情进行测试 `test_label` 最后对一级标签进行对比，得到测试集的准确度 详细模型参见 `untitled0.py` 。

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

其中 P_i 为第 i 类的查准率 R_i 为第 i 类的查全率 。

通过对附件 2 留言详情进行训练 对附件 2 中百分 80% 用来训练 20% 用来测试 得到训练精度如下：

```
# 贝叶斯算法
from sklearn.naive_bayes import MultinomialNB
model_nb = MultinomialNB()
model_nb.fit(train_cv, train_label)
# 训练集预测值
pre = model_nb.predict(train_cv) # 对训练样本预测
# 正确率
sum(pre == train_label) / len(train_label)
Out[9]: 0.9873737373737373
```

从训练集算法结果可知 训练集精确到高达 98% 这表明了建立的训练集模型能够对附件 2 留言详情内容进行训练 。

通过测试集对附件 2 进行测试 得到测试准确度和数据如下：

Index	一级标签
1664	城乡建设
2476	环境保护
5690	劳动和社会保障
8816	卫生计生
5517	劳动和社会保障
734	商贸旅游
7149	商贸旅游
6934	劳动和社会保障
2582	环境保护
8318	商贸旅游
1036	城乡建设
9066	卫生计生

train_label

	0
0	城乡建设
1	环境保护
2	劳动和社会保障
3	卫生计生
4	劳动和社会保障
5	商贸旅游
6	商贸旅游
7	劳动和社会保障
8	环境保护
9	商贸旅游
10	城乡建设
11	卫生计生

test_label

图 2.4 数据对比

F1_score(macro): 0.8345

F1_score(micro): 0.8528

F1_score(weighted): 0.8502

从测试集结果来看 本文建立的训练集能够实现留言详情分类 且测试精度在 85%左右 , 表明了本文建立预测模型有较好的预测度 。

三 、任务二

对数据附件 3 进行热点分类时 ‘识别热点问题可以依据两个因素 ‘留言文本相似程度 、点赞和反对数 ‘将 gensim 库导入 Python 中 ‘由于 gensim 库包含多种文本处理的模型 ‘像 corpora, models, similarities 等 ‘这些库能够解决上述热点问题分类 ‘首先对附件 3 数据进行 class DocumentSimilar 计算 ‘紧接着使用 chineseStopWords.txt 对上述数据进行分词过滤 ‘最后计算一些必要数据得到相似矩阵 。

计算热点问题时 ‘为了简化模型 ‘我们定义第一个留言详情问题相似度为 1 ‘别的留言详情则按照相似度原则进行数字化 ‘假设别的留言数据数字化后数据为 0.1 ‘那么我们认为这个留言数据和第一个留言有些类似 ‘别的数据也按照上述模型计算 ‘详细模型参见 f3.py 。。得到留言相似矩阵 aa 如下 ：

Index	Type	Size	Value
0	float32	(4326,)	[9.9999994e-01 1.6851961e-03 2.0801647e-04 ... 9.1393670e-04 6.7807414 ...
1	float32	(4326,)	[1.6851961e-03 1.0000001e+00 3.1563358e-03 ... 7.1213702e-03 6.2139996 ...
2	float32	(4326,)	[2.0801647e-04 3.1563358e-03 1.0000000e+00 ... 3.7290321e-03 1.5742084 ...
3	float32	(4326,)	[0.00107792 0.00098782 0.00761007 ... 0.00116571 0.00822193

图 3.1 留言相似矩阵

Index	Type	Size	
0	float32	(4326,)	[9.9999994e-01 1.6851961e-
1	float32	(4326,)	[1.6851961e-03 1.0000001e+
2	float32	(4326,)	[2.0801647e-04 3.1563358e-
3	float32	(4326,)	[0.00107792 0.00098782 0.0
4	float32	(4326,)	[[0.00686955] [0.00255567]
5	float32	(4326,)	[0.01090965 0.01470759 0.0
6	float32	(4326,)	[0.00197541 0.00277159 0.0
7	float32	(4326,)	[0.00246505 0.00450469 0.0
8	float32	(4326,)	[0.00413342 0.01498999 0.0
9	float32	(4326,)	[0.0092009 0.00043465 0.0
10	float32	(4326,)	[0.00036084 0.02362434 0.0
11	float32	(4326,)	[0.00398157 0.00305561 0.0
12	float32	(4326,)	[0.00719923 0.01432671 0.0
13	float32	(4326,)	[0.00304795 0.00596052 0.0

4 - NumPy array

	0
3	0.00680175
4	1
5	0.00837443
6	0.00840333
7	0.00213258
8	0.00627147
9	0.0767149
10	0.0120256
11	0.0189529
12	0.00903528
13	0.13377
14	0.00573145

图 3.2 留言数据详情

4	1-83024	2003080263	8-6-6 汇泰华新金昌里 水泥路, 自来水到户的问题	2019/7/19 18:59:11	5月都通知好了, 自来水没开就开了, 浪费资源	已	1
5	1-83039	2003081373	2区黄壳路步行街太古道管户二中间新外排	2019/3/19 12:48:23	和物业联系进行清理, 没有解决根本问题。三票	已	1
6	1-83039	2003087971	3区与旧社区三区和旧二中间空地做绿化工程	2013/11/25 16:51:42	8月23号法国国际社区三期四期中间, 即每天没封制, 几月没封制这块空地一直处于三不管状态, 物业公司不管也不管, 从去年开始做直道施工工程有争议的, 因为这种种在这块地上建, 那个时候土的时候再建呢, 而且这个状态一直在蔓延, 目前以上几个问题在天上四期四期中间	已	1
7	1-83073	6529134	黄泥土里方面路要修路时小区附近菜田管理	2019/3/11 11:49:42	修路, 要求施工单位修路期间, 政府应该派工作人员	已	1
8	1-83074	6529946	2区黄泥土里方面路的高压电杆	2019/1/12 20:51:12	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
9	1-83119	2003080262	2区黄泥土里方面路的高压电杆	2019/3/22 16:04:11	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
10	1-83170	200311323	2区黄泥土里方面路的高压电杆	2019/12/28 8:56:51	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
11	1-83219	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
12	1-83257	200313262	2区黄泥土里方面路的高压电杆	2013/10/18 11:12:45	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
13	1-83258	2003080262	2区黄泥土里方面路的高压电杆	2019/3/31 17:06:13	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
14	1-83258	2003080262	2区黄泥土里方面路的高压电杆	2019/5/15 16:28:10	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
15	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
16	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
17	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
18	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
19	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
20	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
21	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
22	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
23	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
24	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
25	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
26	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
27	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
28	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
29	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
30	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
31	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
32	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
33	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
34	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
35	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
36	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
37	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
38	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
39	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
40	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
41	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
42	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
43	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
44	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
45	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
46	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
47	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
48	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
49	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
50	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
51	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
52	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
53	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
54	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
55	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
56	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
57	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
58	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
59	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
60	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
61	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
62	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
63	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
64	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
65	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
66	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
67	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
68	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
69	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
70	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
71	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
72	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
73	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
74	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
75	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
76	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
77	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
78	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
79	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
80	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
81	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
82	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
83	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
84	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
85	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
86	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
87	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
88	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
89	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
90	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
91	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
92	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
93	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
94	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
95	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
96	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
97	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
98	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
99	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1
100	1-83259	2003080263	2区黄泥土里方面路的高压电杆	2019/9/17 4:38:06	电杆安全当问题修路时, 黄泥土里方面路的高压电杆安全当问题修路时	已	1

图 3.3 留言数据详情

图 3.1 我们可以看出 ‘相似矩阵数据都使用是用数值表示’通过计算可知计算的留言数据越接近 ‘则数值越大’。图 3.2 向我们展示了编号为 4 的留言数据 ‘从图中可以看到 ‘编号 13 数值大于 0.13’表明此条数据可能与第 4 编号数据接近’通过查看附件 3 数据 ‘编号 4 对应数据 6’编号 13 对应数据 15’得出两者数据如图 3.3 所示。从图 3.3 可知 ‘两者确实讲述同一个问题’皆为 A 市 A3 区工地施工问题’这更表明了本文建立模型的可靠性。

对图 3.1 中数据 对每条留言详情数值大于 0.1 的条数进行统计 得到留言相似条数 dd2 (热度) 和留言编号 dd3 如图 3.4 ' 并通过留言相似条数得到留言热度如图 3.3.

dd2	list	5	[252, 247, 246, 235, 226]
dd3	list	5	[1215, 24, 3543, 1668, 2290]

图 3.4 留言热度

E6						
	A	B	C	D	E	F
1	留言编号	留言用户	留言主题	留言时间	留言详情	反对数 点赞数
2	217139	AC0054222	A4区德雅路残疾人请求援助	2019/8/22 14:48:55	听说,你只有走法律途	C 0
3	188560	AC0075321	德鸿餐饮店拖欠工资,员工维	2019/10/6 15:01:36	达整改通知书是不是	C 0
4	271948	AC0094264	网站虚假宣传学位,房屋质量	2019/10/30 15:04:23	过调查,我们甚至还	C 0
5	228102	AC0063990	投诉A市A1区苑物业违规收停车	2019/12/30 17:06:14	已多次向物业和社区	C 0
6	242818	AC0018697	龙城二期小区幼儿园未按约定	2019/7/3 18:23:47	愿让我们这上千户家	C 0
7						

图 3.5 热点问题表

四、任务三

针对附件 4 相关部门对留言的答复意见，我们使用 gensim 库对留言答复意见进行相似识别。为简便运算，也使用数值进行相似度量。定义留言相似度大于 0.1 为相似，小于 0.1 则认为答复意见为未对留言进行相关回答。详细模型参见 f4.py。

建立模型后，对附件 4 处理数据如图 4.1 所示。数据详见 f4.py 运行程序得出结果 aa。

Index	Type	Size	Value
66	float32	(2,)	[0.07905694 0.]
67	float32	(2,)	[0.14660333 0.]
68	float32	(2,)	[0.28605482 0.]
69	float32	(2,)	[0.36633527 0.]
70	float32	(2,)	[0.0964396 0.]
71	float32	(2,)	[0.47894886 0.]
72	float32	(2,)	[0.327067 0.]
73	float32	(2,)	[0.5026118 0.]
74	float32	(2,)	[0.12423547 0.]

图 4.1 留言相似度

A	B	C	D	E	F	G	H
4.29F	110087139F	投诉B区小三儿发商和构造商。2018/11/21 10:00:00	投诉B区小三儿发商和构造商。2018/11/21 10:00:00	投诉B区小三儿发商和构造商。2018/11/21 10:00:00	投诉B区小三儿发商和构造商。2018/11/21 10:00:00	投诉B区小三儿发商和构造商。2018/11/21 10:00:00	投诉B区小三儿发商和构造商。2018/11/21 10:00:00

图 4.2 留言详情 1

[illegible]

从图 4.1 可知 所得结果第 69 行对应的数值为 0.366 在附件 4 中对应留言编号为 4236 ；第 70 行对应的数值为 0.09 ’此值小于 0.1 ’在附件 4 中对应留言编号为 4237, ’找到附件 4 中对应的留言与留言记录可知 ’并对两者进行分析可知 ’留言详情 1 中留言回复是经过相关调查得出结论后 ’对留言进行准确回复 ’而留言详情 2 回复只是按照官方规范进行回复 ’并未对提问者的问题进行详细回答 ’这与模型预测此问题回答为未相关接近 ’其余情况类似 ’这表明了本文建立的预测模型具有一定的可行性 。

七、参考文献