“智慧政务”中的文本挖掘应用

摘要

随着网络的飞速发展，网络问政逐渐成为一种主流的问政形式。各类社情民意相关的文本数据量不断攀升，传统人工已经渐渐不能够胜任大量数据的整理和归类筛选。为此，我们利用深度学习和自然语言处理等知识分别就留言分类、热点问题发掘和答复质量评价建立模型，并应用于提高该领域的工作质量和效率。

针对问题一：本文使用 TextCNN 文本分类模型对群众留言进行分类，经过随机重采样进行**数据增强**后，按照 7：1.5：1.5 的比例划分训练集、测试集、验证集。最终获得模型正确率平均为：95.17%，F1-Score 值平均为：95.3%，使用服务器训练花费时间平均时长为 51 秒。

针对第二题：本文通过热度指标 1 和热度指标 2 两个指标建立了热度评价指标。其中热度指标 1 为根据本问实际情况做出修改的 Reddit 热度指标算法所得，热度指标 2 为基于文本长度的热度指标算法所得，建立热度评价指标计算公式为:$\mathbf{Score} = \boldsymbol{w_1} \times \mathbf{热度指标1} + \boldsymbol{w_2} \times \mathbf{热度指标2}$。通过名实体识别获得具体地点并根据具体地点使用 Single-Pass 聚类算法进行话题提取。使用评价模型对所得话题进行热度评价。最后将热度前 5 的话题保存到热点问题表以及热点问题留言明细表文件。

针对第三题：本文通过文本相关性、可解释性、完整性和时效性四项指标建立答复意见评价模型算法，公式为：$\mathbf{Score} = \mathbf{w_1} \times \mathbf{相关性指标} + \mathbf{w_2} \times \mathbf{可解释性指标} + \mathbf{w_3} \times \mathbf{完整性指标} + \mathbf{w_4} \times \mathbf{时效性指标}$。通过评价模型对附件 4 答复进行评价，最后将结果保存到 C3 评价明细表。

关键词：TextCNN 分类模型；数据增强；热度评价指标；答复意见评价指标；文本分类；深度学习；

Abstract:

With the rapid development of the network, network politics has gradually become a mainstream form of politics. The amount of text data related to various social situations and public opinion is increasing, and traditional labor has gradually been unable to be competent to sort out and summarize a large number of data. Therefore, we use the knowledge of deep learning and natural language processing to build models for message classification, hot topic discovery and response quality evaluation to improve the work quality and efficiency in this field.

For question 1: In this paper, the TextCNN text classification model is used to classify the mass messages. After random resampling for data enhancement, the training set, test set, and verification set are divided according to the ratio of 7: 1.5: 1.5. The average accuracy rate obtained is 95.17%, the F1-Score value is 95.3%, and the average time spent using server training is 57 seconds.

For question 2: The calculation formula of the thermal evaluation model established in this paper is: Score=w1×score1+w2×score2, of which heat score1 is the index obtained by the Reddit heat index algorithm modified according to the actual situation of this question. Score2 is the heat index obtained by the heat index algorithm based on the text length. Use the evaluation model to evaluate the resulting topics. Finally, save the top 5 topics to the hotspot question list and hotspot question list file.

For question 3: The algorithm of the quality evaluation model of the reply opinion established in this paper is: Score = w1 × score1 + w2 × score2 + w3× score3 + w4 × score4. Evaluate the response to Annex 4 through the evaluation model, and finally save the result to the C3 evaluation schedule.

Keywords: TextCNN classification model; data enhancement; popularity evaluation index; reply opinion evaluation index; text classification; deep learning

目录

1 问题重述

1.1 问题背景

随着网络的飞速发展，网络问政平台逐步成为政府了解民意、汇聚民智、凝聚民气的重要渠道。由于省去了传统方式的繁琐的手续，因此在某种意义上提高了群众对问题反映的效率。同时，随着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。并且随着大数据、人工智能、云计算等技术的发展，建立基于自然语言处理技术的智慧政务系统已经是社会治理创新发展的新趋势，这对提升政府的管理水平和施政效率具有极大的推动作用。

1.2 要解决的问题

1) 根据附件 2 所给出的数据，建立关于留言内容的一级标签分类模型。

2) 根据附件 3 所给出的数据，将某一时段内反映特定地点或特定人群问题的留言进行归类，定义合理的热度评价指标，并给出评价结果。

3) 根据附件 4 相关部门对留言的答复意见，从答复的相关性、完整性、可解释性等角度对 答复意见的质量给出一套评价方案。

2 问题一：群众留言分类

2.1 问题分析

1. 挖掘任务

充分利用附件 2 的数据，通过数据探索、数据预处理和合理的数据清洗得到较为干净的数据进行数据增强等操作后，将文本向量化表示并划分为训练集和验证集。利用深度学习等知识训练一个准确有效的模型，并对后续未知留言进行分类预测。

2. 挖掘流程

如图 1 所示，挖掘任务主要由四部分组成，第一部分为爬取相关领域 3 万条留言数据以**扩充语料库**，在原有数据基础下训练预训练词向量 word2vec。第二部分为预处理部分，第三部分为**文本数据增强**，第四部分为**文本分类**部分。其中预处理主要包括去除空格符号、去除称呼部分、分词、去停用词、关键词获取、最后根据给定定长将文本进行序列填充。数据增强部分则通过**随机重采样**对数据进行增强。文本分类部分中，将准备好的数据通过 word2vec 词嵌入映射成一个二维的分布式文本向量，通过卷积神经网络的卷积层和池化层进行特征提取，再经过全连接层对特征进行整合并且表决投票，最后输出预测结果。挖掘流程图如图 1 所示：

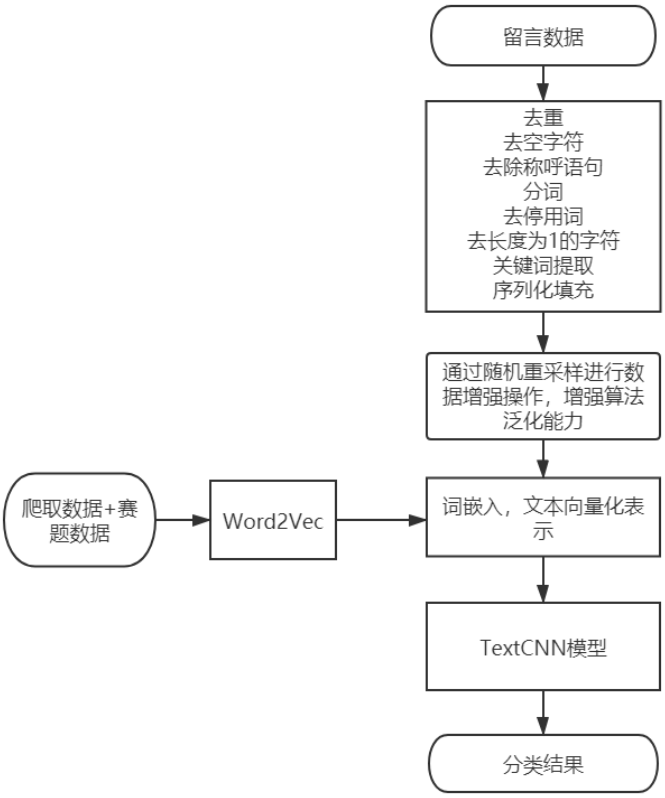


图 1 本问挖掘流程

2.2 模型建立

1. 模型简介

本问选用卷积神经网络文本分类模型（TextCNN）作为留言分类模型。CNN 不仅在计算机视觉领域中有着不凡的成效，而且在文本分类中也具有不错的表现。

由于本问分类任务主要是主题文本分类任务，主题文本分类的核心是获取关键词和关键语句，同时词作为文本的最小粒度，因此本质上便是获取关键词。通过借鉴 CNN 在图像处理领域的经验，CNN 经过卷积和池化操作在特征提取方面有着比较好的表现，进而转战到文本分类中。将文本通过词嵌入转换成 L*M（L 为最长文本长度，M 为词向量的特征长度）的矩阵，进而可以进行类似计算机视觉的识别操任务。

TextCNN 主要由词嵌入层、卷积层、池化层、全连接层和输出层组成。其中词嵌入层主要将文本数据通过 word2vec 转换成文本向量，卷积层主要通过卷积操作获取感受野内的特征，特别地，TextCNN 选取的卷积核宽度与词向量维度保持一致。池化层是通过卷次操作所得特征再进一步进行特征提取，能起到降维作用并且在理想情况下往往能保留原有的显著的特征，池化操作后需要对特征进行拼接以此获得最终的特征向量，特别地，TextCNN 的池化往往选用 1-max-pooling 以保证获取到的是特征向量最大值的特征。全连接层主要是对多个特征进行表决，如图 2 所示：

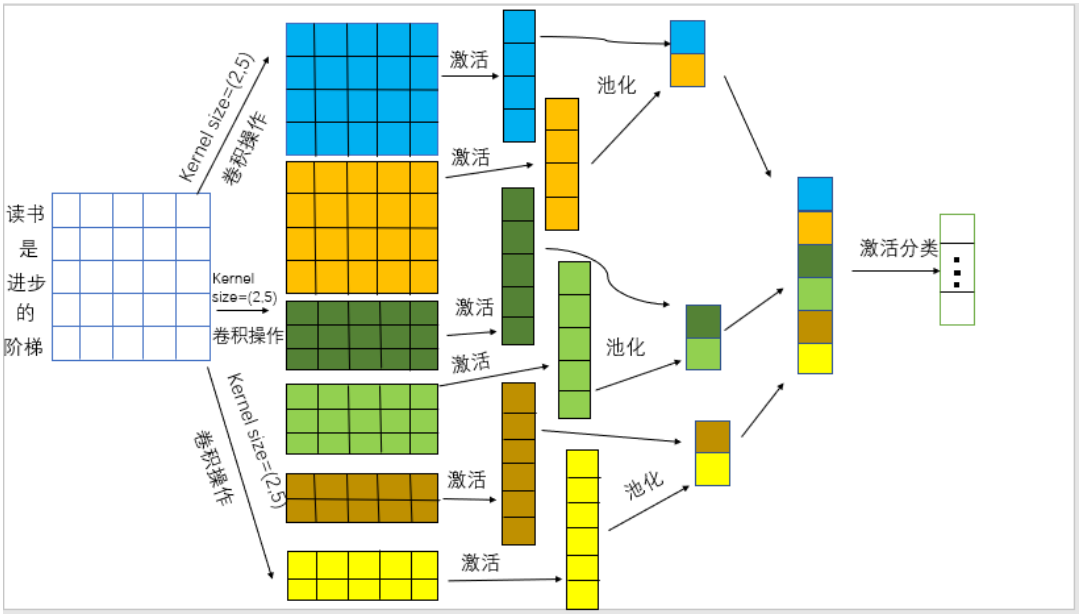


图 2　TextCNN 文本分类大致演示

2. 模型框架

本文搭建的 TextCNN 模型首先通过 Word2Vec 词嵌入将数据转换成文本向量，然后通过卷积操作进行特征提取，其中卷积核尺寸分别为 1*300，2*300、3*300、4*300，卷积核个数均为 150 个。通过 1-Max-polling 最大池化得到最强特征，

使用 tanh 激活进行激活，然后通过拼接级联得到最终特征表达。经过两层全连接隐藏层，均用 tanh 激活函数激活，并使用 0.5 的 dropout 防止过拟合。最后输出层通过 softmax 函数进行激活分类。模型框架如图 3 所示。

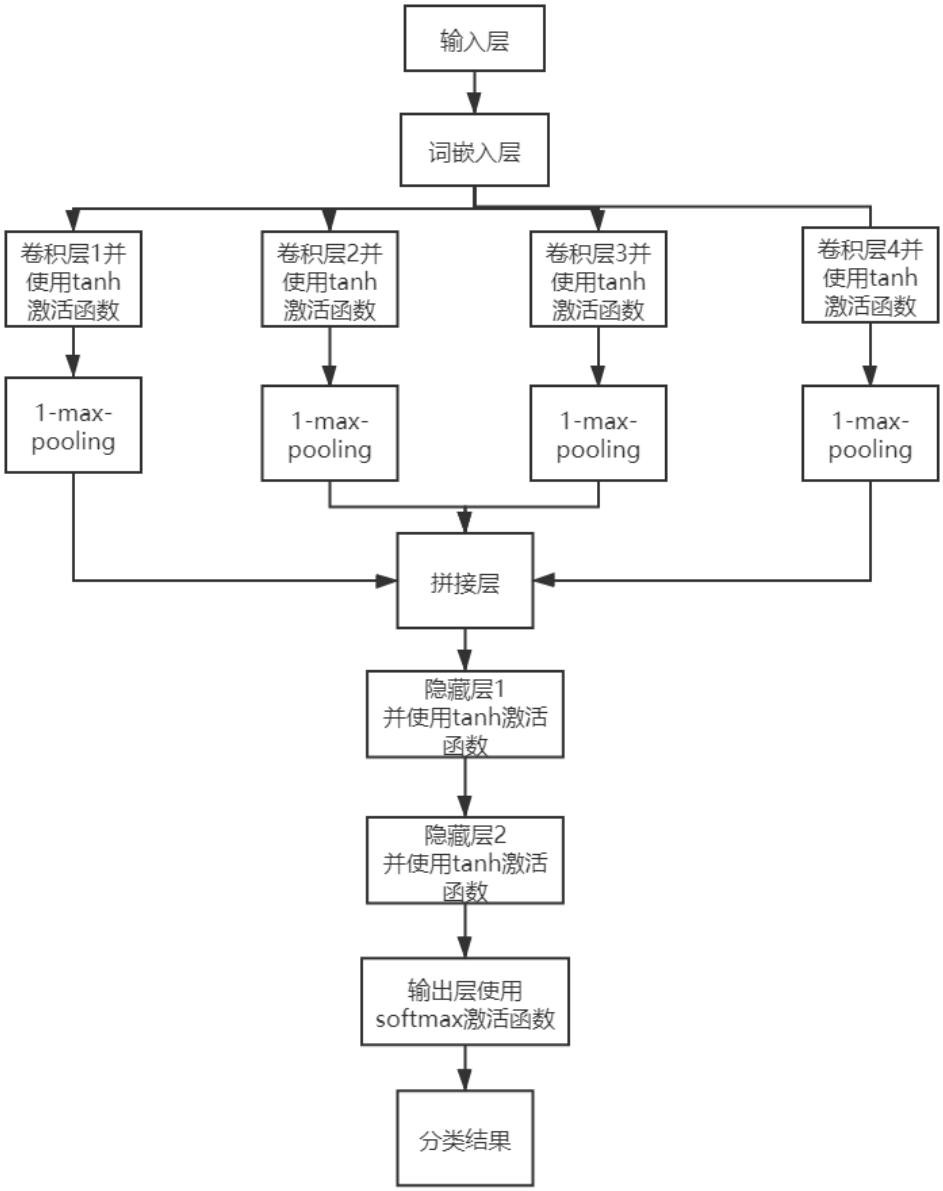


图 3 TextCNN 模型框架

2.3 模型求解

1. 实验条件

本次实验的硬件配置为深度学习服务器：CPU：Intel Xeon E5-2650 v4 ×29 mins，显卡：NVIDIA GeForce GTX Titan X ×8，内存：32G×7，SSD：1T。

实验软件环境为：Python3.6、keras2.3.1

2. 数据预处理

去重、去空字符：数据探索过程中发现数据中有158条留言有数据重复，以及几乎每条数据都存在空字符，为使算法效果不受影响，因此进行了去重和去空字符的处理。

分词：在中文自然语言处理中，词是最小的能够独立活动的有意义的语言成分，但是词语之间没有明显的区分标记。为此分词处理显得尤其重要。本文采用Python语言的jieba分词模块对文本进行分词处理。

去停用词：停用词通常指在文本挖掘过程中利用价值几乎可以忽略不计的词语文字。因此去停用词能够在一定程度上节省存储空间和提高运算效率，与此同时还能够较大程度上减少对高价值词语的干扰性。

关键词提取：本文采用将关键词集代替文本的方式作为数据清洗和特征提取的一种处理。原因有三：1.本文分类任务属于主题分类，主题分类任务有别于情感分类任务，对于上下文的语义的依赖较弱，主题分类更多的是根据部分重点词汇作为分类的依据。2.相较于完整文本，用关键词集代替完整文本更能够减少无关词的干扰，使得特征更加鲜明，更加有利于卷积神经网络的特征提取。3.由于选用关键词集代表文本，因此在一定程度上减少运算，提高运算效率。关键词提取主要是通过TF-IDF和TextRank算法进行提取，为防止信息流失，分别使用两种算法求得每则留言对应排名前30的关键词，然后综合两种算法的结果进行关键词集合并，一方面能保证关键词不遗漏，另一方面加大两个算法同时获得的关键词的权重。假设TF-IDF算法针对某留言获取的关键词集为[A,B,C]，TextRank算法针对某留言获取的关键词集为[A,C,D]，将两个词集合并则得到[A,A,B,C,C,D]。

序列化填充:根据以上步骤所获得数据，赋予每个词语一个独有的数值编号。再将数据用数值编号表示。设每条数据最大长度为60，假如文本长度不足60则用字符’PAD’填充。

Word2vec词嵌入训练：为了进一步扩充语料库进而加强数据的语义特征和提高算法鲁棒性，本文从国内一较为权威的网络问政平台——“问政四川”爬取

了 3 万多条数据（数据存储于四川.csv），并且结合附件 2、附件 3、附件 4 中的留言数据使用 Skip-gram 模式进行 word2vec 词向量的训练。

word2vec 是词向量的一种分布式表示，核心思想是邻近的词会存在某种语义相关的特性，因此可以通过神经元网络的方式进行训练使得隐含的特征被一定程度体现。如图 4 所示，Skip-Gram 训练模型是由三层网络——输入层、隐藏层、输出层组成，每层的激活函数均选用线性激活函数。Skip-Gram 模型核心思想是通过当前 B 词预测上下文的词 A 和 C（假设词窗大小为 1）。将 B 作为输入，输出则为预测 A 和 C 的概率，通过多次反向传导更新权重当损失函数足够低时便停止训练。最后将隐藏层的权重作为向量属性。相较于传统的独热编码和词袋模型，word2vec 很好的解决了高维稀疏的问题，同时赋予词语一定的语义信息。

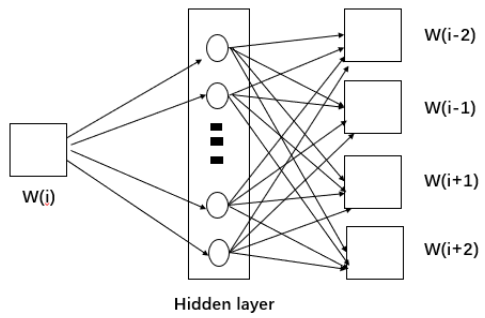


图 4 skip-gram 模型

Word2vec 效果展示

”红绿灯”语义最接近的前十个词语

[('斑马线', 0.919451117515564),
('信号灯', 0.9041105508804321),
('人行横道', 0.902644157409668),
('路口', 0.893820583820343),
('电子眼', 0.8856887221336365),
('指示牌', 0.8821775913238525),
('摄像头', 0.8796292543411255),
('天桥', 0.8674561977386475),
('十字路口', 0.8636099696159363),
('地下通道', 0.860063910484314)]

3. 数据增强以及数据选择

在附件 2 的数据中，总共有 9210 条数据 7 个类，其中经过去重后，交通运输类有 597 条，劳动和社会保障类有 1961 条，卫生计生类有 875 条，商贸旅游类有 1168 条，城乡建设类有 1992 条，教育文体类和环境保护类分别又 1570 条和 924 条，数据很不平衡。如图 5 所示。数据不平衡会导致训练过拟合，预测结果会偏向数据量多的类。因此本文采用**随机重采样**的方法进行**数据增强**。随机重采样核心思想是：对少数类的数据，通过随机抽取的方式抽取本类中的数据并进行填充，直到少数类数据条数等于多数类数据条数，使数据集变得均匀，如图 7 所示。最终每类数据条数均为 1992 条。进行数据增强后使得留言分类准确率和 F-Score 值均**增长了 5%-6%**，如图 6 所示。最后以 7：1.5：1.5 的比例划分训练集、验证集、测试集。

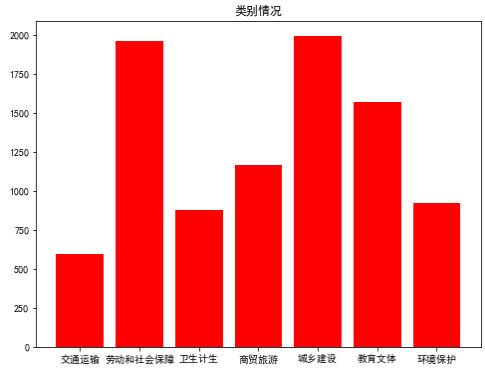


图 5 原始数据集各类数据条数

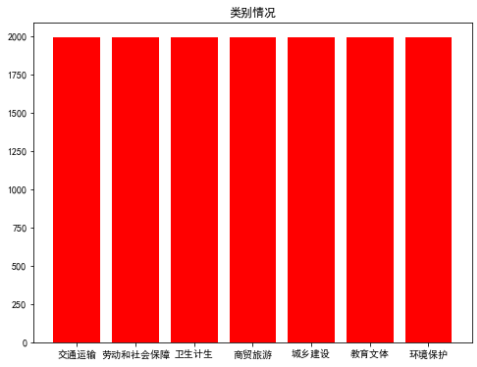


图 6 重采样后数据集各类数据条数

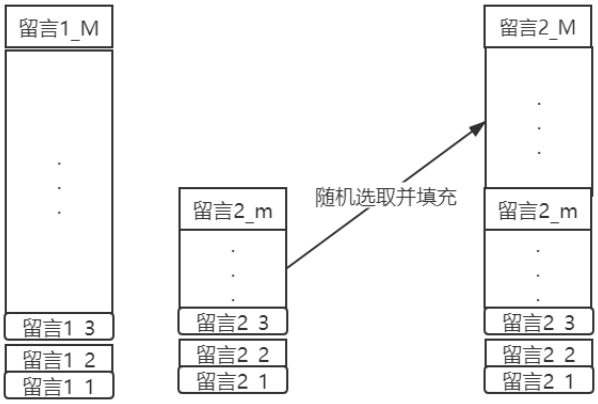


图 7 随机重采样

4. 输入数据

经过数据预处理、数据增强和数据选择后，将选择的数据的序列化向量输入到 TextCNN 模型中。

5. 模型的训练

该模型的代码全都整合在 Question_1.py 文件中，利用此程序对模型进行训练。模型主要参数如下表所示：

表 1 模型主要参数

参数	值
Input_dim（词袋中词的个数）	58182
Output_dim(词向量维度)	300
Input_length（文本长度）	60
Weights（词嵌入）	embedding_matrix（预训练词向量）
Trainable（是否对 Weights 进一步训练）	True
Kernel(卷积核长度)	[1,2,3,4]
Filter_size(卷积核个数)	[150,150,150,150]

Batch_size	1500
Epoch	30

6. 训练部分过程展示如表 2

表 2 部分训练过程

```
Train on 10160 samples, validate on 1793 samples
Epoch 1/100
10160/10160 [==============================] - 3s 274us/step - loss: 1.9169 -
accuracy: 0.2219 - F1_score: nan - val_loss: 1.5923 - val_accuracy: 0.5215 -
val_F1_score: 0.0011
Epoch 2/100
10160/10160 [==============================] - 2s 242us/step - loss: 1.4634 -
accuracy: 0.5703 - F1_score: 0.0790 - val_loss: 1.2039 - val_accuracy: 0.7546 -
val_F1_score: 0.3016
Epoch 3/100
10160/10160 [==============================] - 2s 242us/step - loss: 1.0954 -
accuracy: 0.6925 - F1_score: 0.4262 - val_loss: 0.8829 - val_accuracy: 0.8154 -
val_F1_score: 0.5603
Epoch 4/100
10160/10160 [==============================] - 2s 242us/step - loss: 0.8010 -
accuracy: 0.8091 - F1_score: 0.6545 - val_loss: 0.6450 - val_accuracy: 0.8516 -
val_F1_score: 0.7695
Epoch 5/100
10160/10160 [==============================] - 2s 242us/step - loss: 0.5758 -
accuracy: 0.8678 - F1_score: 0.8009 - val_loss: 0.4667 - val_accuracy: 0.8985 -
val_F1_score: 0.8730
Epoch 6/100
10160/10160 [==============================] - 2s 242us/step - loss: 0.4053 -
accuracy: 0.9167 - F1_score: 0.8923 - val_loss: 0.3529 - val_accuracy: 0.9141 -
val_F1_score: 0.9044
…
…
…
Epoch 18/100
10160/10160 [==============================] - 2s 241us/step - loss: 0.0394 -
accuracy: 0.9995 - F1_score: 0.9992 - val_loss: 0.1703 - val_accuracy: 0.9498 -
val_F1_score: 0.9501
Epoch 19/100
10160/10160 [==============================] - 2s 243us/step - loss: 0.0368 -
accuracy: 0.9995 - F1_score: 0.9992 - val_loss: 0.1696 - val_accuracy: 0.9509 -
val_F1_score: 0.9512
Epoch 20/100
10160/10160 [==============================] - 2s 241us/step - loss: 0.0338 -
accuracy: 0.9994 - F1_score: 0.9992 - val_loss: 0.1692 - val_accuracy: 0.9509 -
```

```
val_F1_score: 0.9524
总共花费 50.73589825630188 秒
Accuracy: 0.950237
F1_Score:0.951998
```

7. 模型的测试

混淆矩阵如下表所示：

表 3 混淆矩阵

预测值 / 真实标签	交通运输	劳动和社会保障	卫生计生	商贸旅游	城乡建设	教育文体	环境保护
交通运输	273	0	0	2	6	0	0
劳动和社会保障	0	281	8	0	1	5	1
卫生计生	0	4	314	2	1	1	1
商贸旅游	5	2	3	299	8	2	1
城乡建设	6	10	3	8	257	4	8
教育文体	0	5	0	3	3	298	0
环境保护	0	0	0	0	2	0	283

2.4 模型分析

1. F-Score

使用 F-Score 对分类方法进行评价，公式如下：

$$\mathrm{F-Score}=\frac{1}{\mathrm{n}}\sum_{i=1}^{n}\frac{2P_iR_i}{P_i+R_i}$$

其中P_i和R_i分别为查准率和查全率，公式如下：

$$P_i=\frac{TP}{TP+FP}$$

$$R_i=\frac{TP}{TP+FN}$$

本问中在 1899 条验证集中，F-Score 最大值时为 **0.9509**，在 2106 条测试集中，F-Score 为 **0.951998**。

2. 混淆矩阵

混淆矩阵如下表所示：

表 4 混淆矩阵

预测值 / 真实标签	交通运输	劳动和社会保障	卫生计生	商贸旅游	城乡建设	教育文体	环境保护
交通运输	273	0	0	2	6	0	0
劳动和社会保障	0	281	8	0	1	5	1
卫生计生	0	4	314	2	1	1	1
商贸旅游	5	2	3	299	8	2	1
城乡建设	6	10	3	8	257	4	8
教育文体	0	5	0	3	3	298	0
环境保护	0	0	0	0	2	0	283

3. 模型训练过程

如下表中的关于准确率、损失函数以及 F-Score 值的训练过程中得变化情况可知，三项验证集的指标大体趋势跟训练集的三项指标的同步，模型的准确率和 F-Score 值最后都能获得比较好的效果，损失函数也能够较好的收敛。

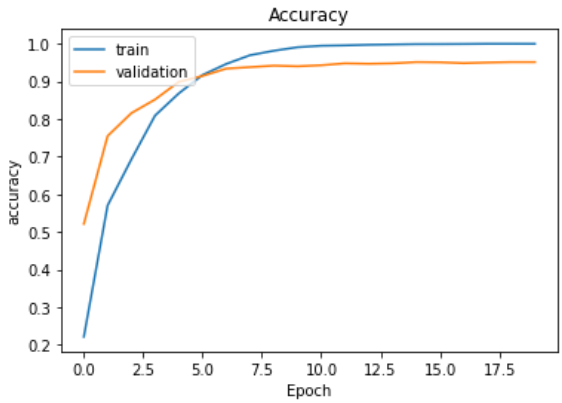


图 8 准确率变动过程

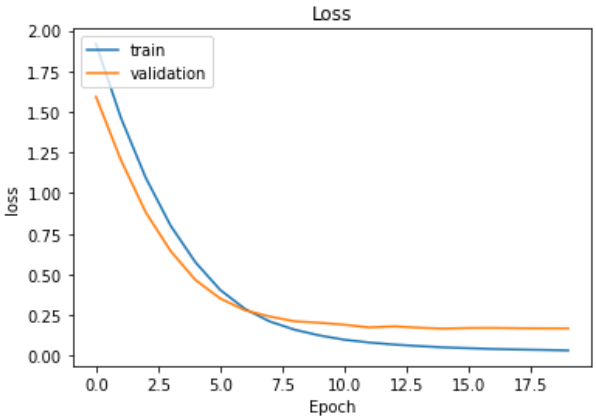


图 9 损失函数变动过程

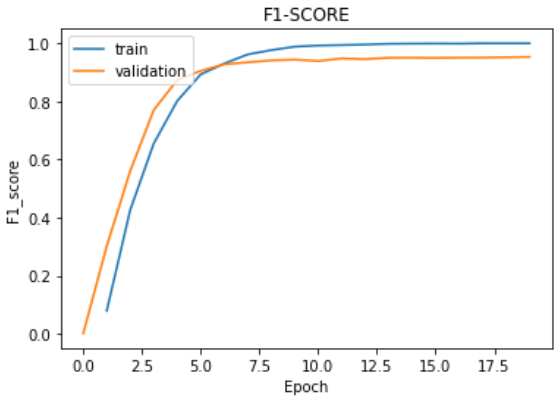


图 10 F-score 变动过程

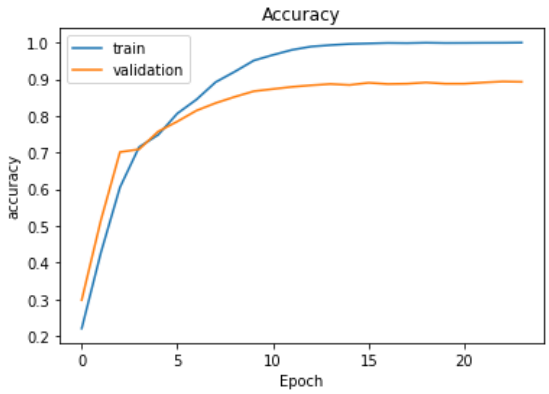
4. 数据增强前后比较

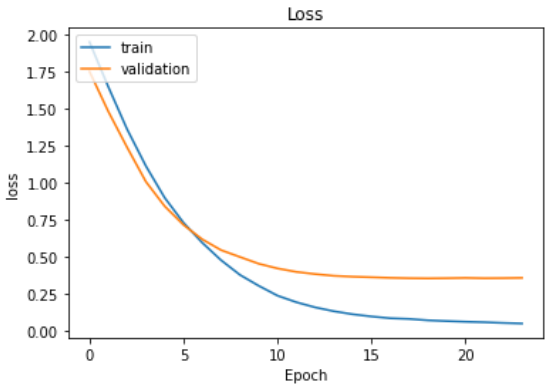
经过比较，数据增强对于模型的拟合能力和训练测试效果有着显著的作用。

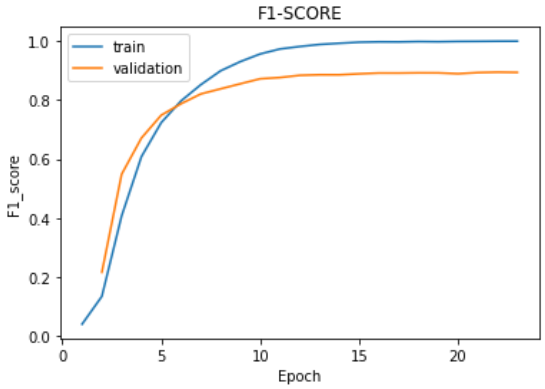
表 5 数据增强前后比较

	准确率	F-Score	训练时间
数据增强前	0.901592	0.902191	40.53372 秒
数据增强后	0.950237	0.951998	50.73589 秒

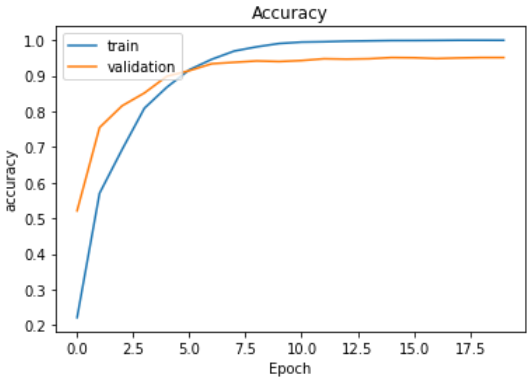
数据增强前

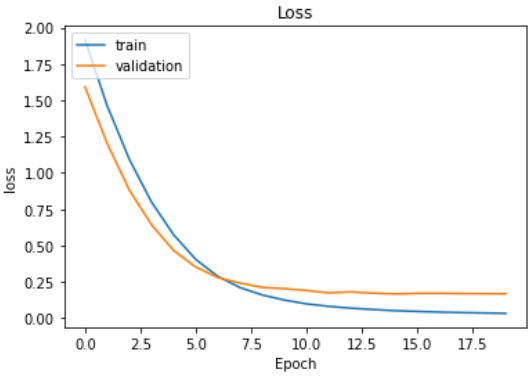


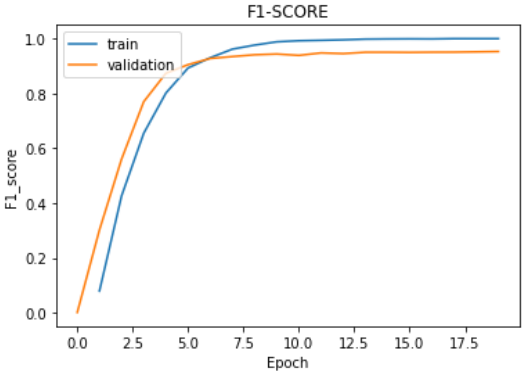




数据增强后







3 问题二：热点问题挖掘

3.1 问题分析

1. 挖掘任务

本问要求我们根据留言数据，将某段时间内反映的地点或者人群问题的留言进行归类，并且定义合理的热度评价指标，给归类后的留言话题进行打分，并且对每类话题进行问题描述和时间范围统计。最后通过热度值大小选择前五的话题作为结果。

2. 挖掘思路

本文关键点主要为聚类模型和地点识别，数据 3 共有 4326 条数据，涉及到的领域和地域都很广，因此直接使用聚类算法进行类别划分则会受到很多干扰因素影响从而影响最终结果，最明显的例子便是虽然文本内容相似但是会出现将不同地区的留言归为一类的情况。因此本文的思路是先进行地点人群的识别，然后以地点为对象对数据进行筛选匹配，之后再使用聚类算法进行归类得到话题，最后在以类别为讨论对象进行热度值评分、问题描述、事件范围归纳和热度排名。

3. 挖掘流程

如图 7 所示，挖掘流程大致可分为 3 部分，第一部分是进行命名实体识别，对小地点进行识别然后储存到地点集 A（形如 [地点，地点 2…地点 m]），同时通过数探索和统计得出市区县等大地点编号集合 B(形如[A5 区,L 县,A 市…])。第二部分主要是按照具体地点的留言事件使用 Single-Pass 聚类算法进行聚类得到话题。对 A、B 进行遍历获取每一个具体地点，再按照具体的地点的留言进行聚类。第三部分主要是对具体的话题进行问题描述、热度值打分（使用 Reddit 排名算法）、提取时间段、热度排名。

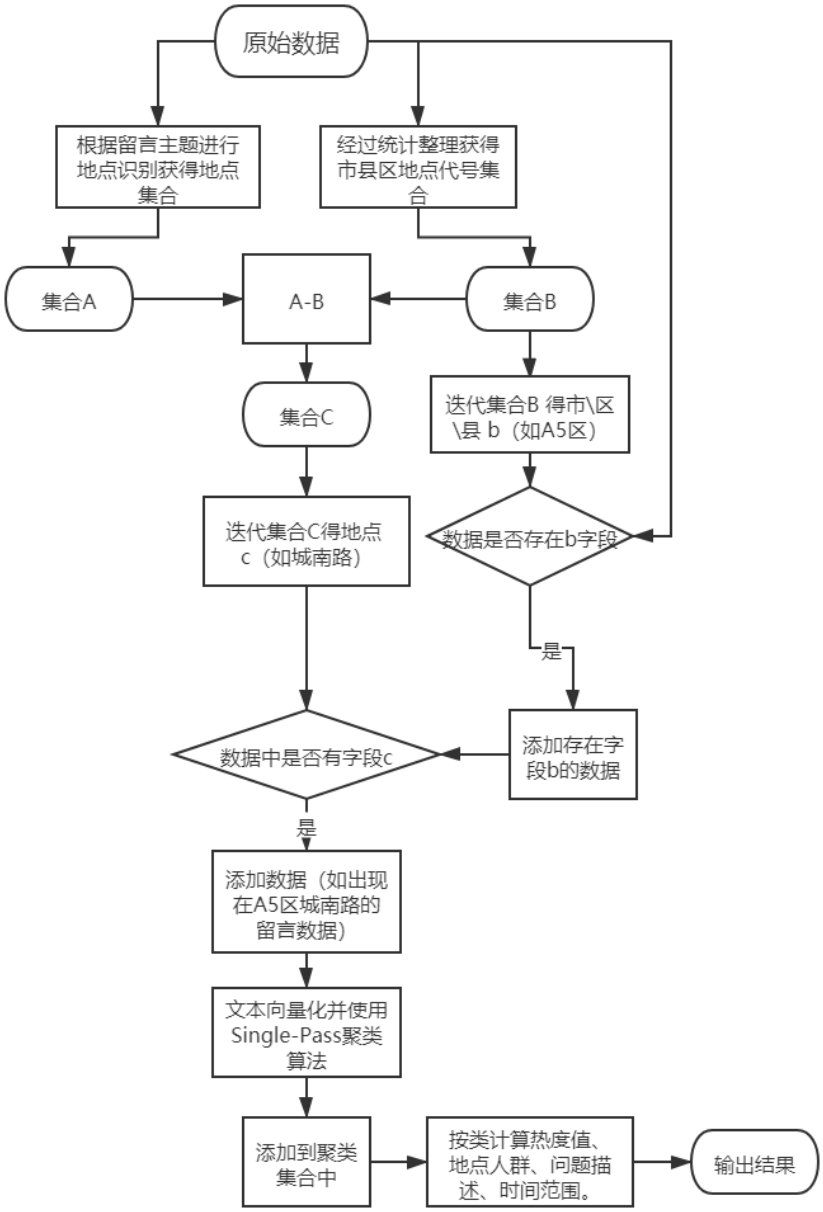


图 11 本问挖掘流程

3.2 热度评价指标建立

热度评价指标

热度评价模型由热度指标 1 和热度指标 2 共同构成，其中热度指标 1 和热度指标 2 如下：

热度指标 1：本文选用 Reddit 排名算法所得结果作为热度指标 1，Reddit 算法由美国的最大网络社区 Reddit 建立。该算法考虑到了留言的新旧程度、留言的赞成数反对数的差值大小、留言的受欢迎程度或者受反对程度同时还考虑到

了投票的方向（积极或者消极）计算公式如下：

$$t = 留言时间 - 2005年12月8日7:46:43$$

$$x = 赞成数 - 反对数$$

$$y = \begin{cases} 1\ (x > 0) \\ 0(x = 0) \\ -1(x < 0) \end{cases}$$

$$z = \begin{cases} |x|(x \neq 0) \\ 1(x = 0) \end{cases}$$

$$S1 = \log_{10} z + \frac{yt}{45000}$$

其中 t、x、y、z、和 S1 分别表示时间新旧程度、赞成数反对数之差、投票方向、受支持或受反对程度以及最终的热度分值。2005 年 12 月 8 日 7:46:43 表示 Reddit 社区创建时间。45000 表示的是 45000 秒，相当于 12.5 个消失。根据实际情景，本文对该算法就本问做了如下改进：

由于在数据中的最晚留言时间为 2017 年 6 月 8 日，并且由于问政平台一般实时性相对较低，一般回复周期为一周，因此对 t 进行改进

$$t = \frac{留言时间 - 2017年6月8日}{7}$$

由于 Reddit 算法得到的分值本质上是更倾向于受留言新旧影响，对于赞成数和反对数较为不敏感，原因是 S1 中的对数部分是以 10 为底的。同时由于问政平台更新不像日常生活中交流的论坛那般迅速，对于热度更新的速度理应较慢，因此做出如下更改：

$$S1 = \log_5 z + \frac{yt}{30}$$

需要说明的是，本文改进的 Reddit 算法在时间上是以天数作为单位，而原始的 Reddit 算法是以秒为单位。

热度指标 2：留言长度也能从侧面体现除群众对问题的关注程度，因此本文还建立了一个居于文本长度的热度打分模型，公式如下：

$$S2 = \left[\frac{\text{len(data)}}{100}\right]$$

其中符号[]表示取整。该模型以 100 为量级进行打分

结合热度指标 1 和热度指标 2，得出热度评价指标，公式如下：

$$\text{Score} = w_1 \times \text{热度指标 } 1 + w_2 \times \text{热度指标 } 2$$

其中根据主观经验，令$w_1 = 0.8$，$w_2 = 0.2$，并进行归一化处理：

$$\text{Score}' = \frac{\text{score} - \overline{score}}{\max(score) - \min(score)}$$。

3.3 模型求解

1. 实验条件

本问的实验的硬件配置为普通 PC：CPU：Intel(R) Core(TM)i5-5200U @2.20GHZ， SSD：120GB，RAM:12GB

实验软件环境为：Python3.6、keras2.3.1、Tensorflow2.0

2. 数据预处理

数据清洗：对数据进行去空格字符、去重复值、去称呼部分、分词、去停用词等操作，以此获得干净的数据。

关键词提取：使用 TF-IDF 和 TextRank 算法进行关键词提取，分别用两种算法求出每条数据中留言主题的前 5 个关键词和留言详情的 15 个关键词，然后将两种算法获得的关键词集求交集，以交集替代留言文本。

文本向量化：将每条数据的每个词通过 word2vec 转化成词向量，再求平均值作为文本向量。

地点识别：使用 Python 语言 hanlp 模块用于地点识别，通过 hanlp 模块的词性识别，选择词性为 ns 和 nt（地名和机构名）的词进行提取并添加到小地方**集合 A** 中。同时经过数据探索和统计，使用字段匹配得到**大地方集合 B**。如（“A 某区”、“A 某县”、“西地省”、“经开区”等编号的大地方）。

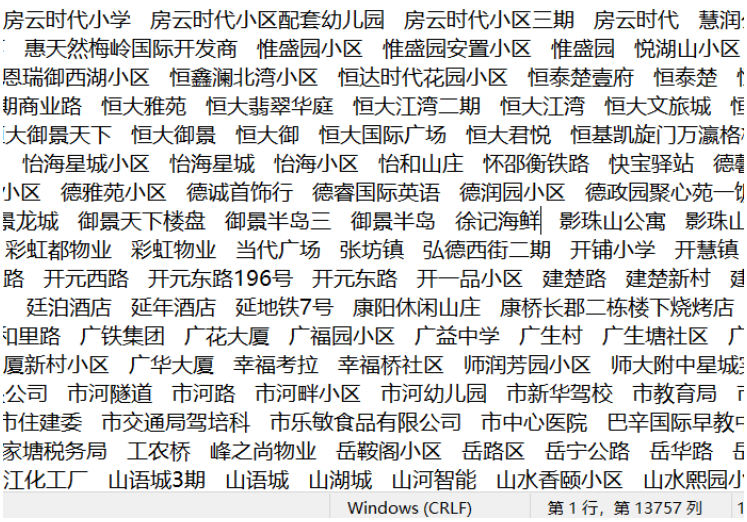


图 12 小地点展示识别

single-pass 聚类算法：single-pass 聚类算法法核心思想：设定阈值，将新的数据跟每条已经存在的话题进行相似度计算，假如相似度大于阈值，则将新数据归入相似度最大的话题中，假如每对相似度都低于阈值，则将当前数据归为新的话题中。

single-pass 算法流程如下：

第一步，输入文本向量集、输入阈值。

第二步对文本向量集进行遍历。

第三步，判断是否存在话题，假如不存在则将（当前）第一条文本向量加入到第一个类中，并且初始化一个新的类等待下一个话题出现。假如存在，则按以下步骤进行。

第四步，遍历已将存在的话题并且计算当前文本向量与每个话题向量的余弦相似度。其中话题的话题向量为话题中的文本向量的平均值。相似度每比前一次大便更新一次直到遍历完成。

第五步，遍历完成，比较更新后最大的相似度值，假如最大值小于阈值则将当前文本添加到新话题中，并且初始化一个新的类等待下一个话题出现，假如最大阈值大于阈值，则将当前文本添加到与之相似度最大的话题中。

第六步，判断是否存在未遍历的文本向量，假如有按照以上步骤进行，假如全部遍历完成则聚类过程完毕。其中余弦相似度计算公式如下：

$$similarity = \cos(\theta) = \frac{A \cdot B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}}$$

问题描述：本模算法于话题聚类之后，由于留言详情本身便有着很强的概括性，因此本文通过针对话题中文本对应的留言主题为描述对象，使用 TextRank 算法进行文本摘要。

3. 算法流程

第一步，根据附件 3 的数据，通过热度指标 1 和热度指标 2 对计算每条留言对应指标。

第二步，对大地点集合 B 进行遍历然后对小地点集合 A 遍历，**以具体的地点为对象利用 sing-pass 聚类算法得出话题**。

第三步：聚类完成并获得全部话题集，并分别对每个热度指标 1 和热度指标 2 进行求和归一化，再通过**热度评价模型得出热度值**。对话题内留言分别求得最新最旧日期进行日期范围统计，**使用问题描述算法对话题内留言进行问题描述**，将**第二步所得具体地点作为话题地点。**最后根据热度值做降序操作，筛选前五的话题。

3.4 模型分析

结果分析：本模型最终效果比较理想，首先在地点提取中精确度相当高，其次在话题提取效果中，能够很大程度上将留言归好类，问题描述借助这留言主图的高概括性的特点也能够比较突出的将问题进行描述。

结果展示

本题总共归纳出**2118 个话题**，涉及地点 1935 个（存在同地点不名称的情况）。本问模型更偏向于留言反映次数较多的话题，对于赞成数较大但是留言反映条数的话题较为不敏感。排名前五的热点如下表所示，前五的话题的详细信息如下表所示：

表 6 热点话题表

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	1	2019/07/07 至 2019/08/31	A 市伊景园	投诉 A 市伊景园滨河苑捆绑车位销售

2	2	0.665119	2019/04/30 至 2019/11/23	A7 县东六路	A7 县泉塘街道漓楚东路与东六路交汇处是否可以架设人行天桥
3	3	0.626674	2019/02/14 至 2019/09/09	A7 县凉塘路	A7 县星沙四区凉塘路旧城改造要拖到何年何月才能动工
4	4	0.609054	2019/01/03 至 2019/12/30	A 市地铁 3 号线	A 市地铁 3 号线松雅西地省站首开为什么没有直通松雅湖方向的出入口
5	5	0.585594	2019/03/26 至 2019/04/12	A6 区月亮岛路	关于 A6 区月亮岛路沿线架设 110KV 高压线杆的投诉

表 7 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	反对数	点赞数
1	191001	A909171	A 市伊景园滨河苑协商要求购房同时必须购买车位	2019-08-16 09:21:33	商品房伊景园滨河苑项目是由 A 市政府办牵头为广铁集团铁路职工定向销售的楼盘，作为集团的一名退休员工，我深深感觉到政府对铁路员工的关怀，在广铁辛苦几十年…	1	12
1	196264	A00095080	投诉 A 市伊景园滨河苑捆绑车位销售	2019/8/7 19:52:14	A 市伊景园·滨河苑现强制要求购房者捆绑购买 12 万车位一个，不买就取消购房资格，国家三令五申禁止捆绑车位销售…	0	0
…	…	..	…	…	…	…	…

1	289950	A00044759	投诉 A 市伊景园滨河院捆绑销售车位	2019/7/7 7:28:06	提问 A 市政府就广州局集团公司与 A 市政府及 A 市政建设有限公司协商达成对武广铁路职…	0	0
2	205332	A00028151	反映A7县泉塘街道漓楚东路与东六路区域白天夜晚飙车扰民问题	2019/11/23 17:13:43	尊敬的领导：您好！泉塘区域最近飙车党又开始猖狂起来了！从中午开始就飙车，凌晨一两点都在飙车！特别提醒：飙车党的停车区域应该在铁建国际城附近！重点位置：泉塘东六路和漓楚东路交界区…	0	4
…	…	…	…	…	…	…	…
2	234588	A00034367	A7 县开元路与东六路交叉口北半幅路面被围挡占道多年	2019/7/11 11:51:06	因地铁 3 号线施工，A7 县开元路与东六路交叉口北半幅路面被围挡占道多年，附近居民由东六路上开元路需要绕行十分不便，围挡上有挂牌公示占路时间…	0	3
…	…	…	…	…	…	…	…

表 8 地点提取效果表

岳鞍阁小区 A市第六都三期北区 A7县金科物业 A3区燕联小区3栋一KTV A1区马坡岭街道 A4区四方坪 A市越卓集团 西地省斯拉同舟法车业公司 A4区伍家岭街道 A4区市河幼儿园 省加家车位不动产管理咨询公司 A市杨山家立交桥 A2区金线街 A市富瑞星座小区 A5区五一新村 A市电力局 A市威尼斯城 A市高岭国际商贸城 A8县厚海玉龙国际 A市创世纪广场 中国工商银行西地省金融培训学校 A2区朝阳烧烤龙湾店 A市烈士公园 A市老街老地方美食广场 A市万芙路青园小学 三一重工 A7县龙塘安置小区 A3区楚雅路过江隧道匝 A7县合心村 A6区星月绿洲小区 A6区盛腾雅苑安置小区 A2区福绿新村 A市旭辉国际广场 A市金科时代中心物业 A3区学士街道 A市恒大江湾二期 A市北二环 A市双龙警苑 A市融城园 A5区莲湖村 A市梅溪湖壹号 A6区茜茜美甲店 A2区保利花园三期 A市九龙仓 A6区马龙口组 A4区恒鑫澜北湾小区 A市赤岭路 A3区云顶梅溪湖 西地省堂源华唯崇医药公司 A2区南门口地铁站 A市钰龙天 A市卓越浅水湾 A市西地省顺安房产 A市山水佳园小区 A6区顺舟旺城 A市梦想枫林湾 A市泰阳商城 A市九龙领仕汇小区 A3区联邦小区 A市金科天悦 A市魏家坡小区 A3区房云时代 A市经开区城北污水厂 A3区中海国际社区幼儿园 A市中联重科 A市金座雅居 A市明发国际 A2区老公寓 K9县金色梯田幼儿园 A市金地物业 A3区振业城二期车库 A3区学士街道主次干道 A市黑石铺街道 A3区华润橡树湾咔吉堡儿童乐园 西地省玖鼎能源环境科技公司 A市五一大道 A2区港子河公园 A4区宏金花园 A3区许家洲路 A3区东方红街道 A市金六福珠宝 A市嘉顺苑安置小区 A7县沙塘冲组 西地省长韶娄高速公路有限公司 A市西湖公园 A市盛庭岚 A市德雅苑生活小区 A1区华年世嘉酒店 A5区明昇壹城二期 A7县兆坤星悦荟 A3区润塘小区 A2区嘉苑偶寓 A5区井圭路农贸市场 A市业委会 A市星沙城 A市区配套幼儿园 A1区育才二小 A3区枫林美景小区 A3区新明园林 A市稻田中学 A4区清水塘溪泉湾小学 A市恒大文旅城 A6区渝长厦高铁 A市规划局 A6区紫郡鲲鹏物业 A市德睿国际英语 A市鑫源悦城 D7县畔塘村 A市汽车南站 A8县世纪花园 A1区顺发驾校 A7县星骑士 A市沙坪老街 A7县坤兆星悦荟 A市三一大道 A3区唐家洲 A7县星城首府 A市山庭苑 A3区梅溪湖壹号御湾 A3区高心麓城芭比罗幼儿园 A7县松雅安置小区 A3区学丰路 A市桃花源 A市广泰家园 A市旭辉御府 西地省堂源华唯崇医药连锁公司 A市枫雅名苑小区 A市恒泰楚 A7县车上组 A8县流沙河镇 西地省直公积金管理中心 A3区保利西海岸小区幼儿园 A1区紫来阁小区 A4区新河生鲜农贸市场 A4区第一湾小区 A市九峰小区 A5区明大佳园 A4区月湖公园 A市八一路 A市保利小区 A7县三一街区小区 A市惠天然梅岭国际开发商 A市市都昌能源科技有限公司 A7县武塘村 西地省建达鸿实业集团公司 A市山水湾 A6区观音岩片区 A1区锦城AB区 A6区兴邦花城小区 A市房云时代小区三期 A3区西雅博才小学 A1区苏家巷社 A5区井湾子 A6区区熙庭 A市澳海澜庭 A市保利麓谷 A2区狮子山社区 A市美联嘉园物业 A6区紫鑫御湖湾一号 A2区石碑中学 西地省科技职院 A市城市建设开发公司 A2区兴马州农家乐饭庄 A2区柠檬丽都小区 A1区东成大厦业主委员会 A8县夏铎铺亮之星企业 A市金邻公寓 A3区金科东方 A市交通银行 A5区红

表 9 话题提取效果表

A市伊景园滨河苑协商要求购房同时必须购买车位 投诉A市伊景园滨河苑捆绑车位销售 投诉A市伊景园滨河苑定向限价商品房违规涨价 A市伊景园滨河苑捆绑销售车位 A市伊景园·滨河苑欺诈消费者 A市伊景园滨河苑定向限价商品房项目违规捆绑销售车位 投诉A市伊景园滨河苑捆绑销售车位 投诉A市伊景园滨河苑捆绑车位销售 无视消费者权益的A市伊景园滨河苑车位捆绑销售行为 和谐社会背景下的A市伊景园滨河苑车位捆绑销售 投诉A市伊景园滨河苑定向限价商品房违规涨价 A市伊景园滨河苑诈骗钱财 违反自由买卖的A市伊景园滨河苑车位捆绑销售行为 A市伊景园滨河苑欺压百姓 投诉A市伊景园滨河苑开发商违法捆绑销售无产权车位 惊！！A市伊景园滨河苑商品房竟然捆绑销售车位 A市伊景园滨河苑坑害购房者 A市伊景园滨河苑捆绑销售车位是否合理 投诉A市伊景园滨河苑开发商违法捆绑销售无产权车位 A市伊景园滨河苑项目捆绑销售车位 无视职工意愿职工权益的A市伊景园滨河苑车位捆绑销售行为 投诉A市伊景园滨河苑开发商 投诉A市伊景园滨河院捆绑销售车位	反映A7县泉塘街道漓楚东路与东六路区域白天夜晚飙车扰民问题 A7县泉塘街道漓楚东路与东六路交汇处是否可以架设人行天桥 请求解决A7县漓楚路与东六路上下班时间交通拥堵的问题 A7县开元路与东六路交叉口北半幅路面被围挡占道多年 A7县东六路与漓楚路口照明光线太暗存在极大安全隐患 建议取消A7县盼盼路东六路至东四路段的三条人行道 建议A7县在漓楚路和东六路交汇处建地下通道或人行天桥 A7县漓楚路和东六路十字路口照明灯光太暗 希望A7县东六路泉塘中学段能增设人行横道 A7县凉塘路的旧城改造要拖到何时才能动工 请问A7县星沙凉塘路的旧城改造要拖到何时何月何时才能再次启动 A7县星沙凉塘路旧城改造究竟要拖到何年何月才能开始 A7县星沙街道凉塘路的旧城改造什么时候会启动 A7县星沙街道凉塘路旧城改造什么时候可以进行 A7县星沙四区凉塘路旧城改造要待何时 A7县星沙四区凉塘路旧城改造要拖到何年何月才能动工 A7县星沙镇四区凉塘路改造何时可以开始

4 问题三：答复意见的评价

4.1 问题分析

挖掘任务

针对附件 4 提供的留言数据和其对应的答复意见，针对答复的相关性、完整性和可解释等进行答复质量评价。

挖掘思路

由于是建立对答复质量的评价模型，我们本文联系现实生活，认为一个好的答复应该具有以下特质：第一，回复及时。第二，回复的内容应该是基于问题本身回答的，而非答非所问。第三，一个好的答复应该具备有充足的论点和论据，能够利用论据对论点进行展开。第四便是尽可能地详尽而非一笔带过。概括起来便是应该具备时效性、相关性、可解释性和完整性。为此本文将从这四个特点处入手对答复质量进行评价。

挖掘流程

本问的挖掘流程主要是建立**相关性指标、可解释性指标、完整性指标、时效性指标**四个指标。然后分别对每则答复进行四项指标评价并且分别**归一化**，最

后对四项指标**求和再归一化**。

4.2 模型建立

答复质量评价指标

留言答复意见评价指标由相关性、可解释性、完整性、时效性四项指标组成：

相关性指标

相关性指标主要是通过分别将留言文本和答复文本提取关键词，以关键词作为文本数据通过 word2vec 转化为文本向量，并且通过余弦相似度求得留言和答复文本向量之间的相似度。其中余弦相似度公式如下：

$$score1 = \cos(\theta) = \frac{A \cdot B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}}$$

可解释性指标

首先根据对数据的探索观察，发现留言答复均比较官方，因此使用正则表达式通过匹配关键字来获取每则回复引用过的法律、规定等文献，以此作为论据。同样使用正则表达式获取每条答复的分点作答的论点。再通过论点论据的个数对可解释性进行评估。

$$\text{score2} = \frac{3}{1+e^{-A}} + \frac{2}{1+e^{-P}}$$

其中 A、P 分别表示论据个数和论点个数

完整性指标

完整性指标是基于留言答复的文本长度来计算的，文本的有效文本长度越长，答复的完整性就相应的比较高。完整性指标公式如下：

$$\text{score3} = \left[\frac{\text{len(data)}}{100}\right] + 1$$

时效性指标

时效性指标主要是基于留言时间和答复时间之间的时间差建立的，时间越长表示的时效性越低。时效性指标公式如下：

$$\text{score4} = \frac{1}{\text{时差天数} + 1}$$

最终结合四项指标求得答复意见评价指标，公式如下：

$$\text{Score} = w_1 \times \text{相关性} + w_1 \times \text{可解释性} + w_3 \times \text{完整性} + w_4 * \text{时效性}$$

其中$w_1 = 0.4$，$w_2 = 0.3$，$w_3 = 0.2$，$w_4 = 0.1$。

4.3 模型求解

1. 实验条件

本问的实验的硬件配置为普通 PC：CPU：Intel(R) Core(TM)i5-5200U@2.20GHZ，SSD:120GB,RAM:12GB。实验软件环境为:Python3.6。

2. 数据预处理

数据清洗：对数据进行去空格字符、去重复值、去称呼部分、分词、去停用词等操作，以此获得干净的数据。

关键词提取：由于留言主题有很强的概括信息，因此在留言主题中存着大量的关键词汇，同样使用 TF-IDF，分别对留言和答复提取 10 个关键词。

有效信息提取：由于在留言数据和答复数据中存在着各式各样的礼貌格式和内容，为了能够进一步清洗出文本数据的有效内容，本文使用观察法和字段匹配法进行去冗余值处理，进而得到有效信息。

文本向量化：将每条数据的每个词通过 word2vec 转化成词向量，再通过词向量求和取平均值获得文本向量。

3. 数据选择

本问使用附件 4 全部数据进行答复质量评价模型的建立。数据总共有 2816 条数据。

4. 算法流程

本问的挖掘流程主要是建立相关性指标、可解释性指标、完整性指标、时效性指标四个指标。然后分别对每则答复进行四项指标评价并且分别归一化，最后对四项指标求和再归一化。如此便得到最终的质量评价分值。流程图如下所示：

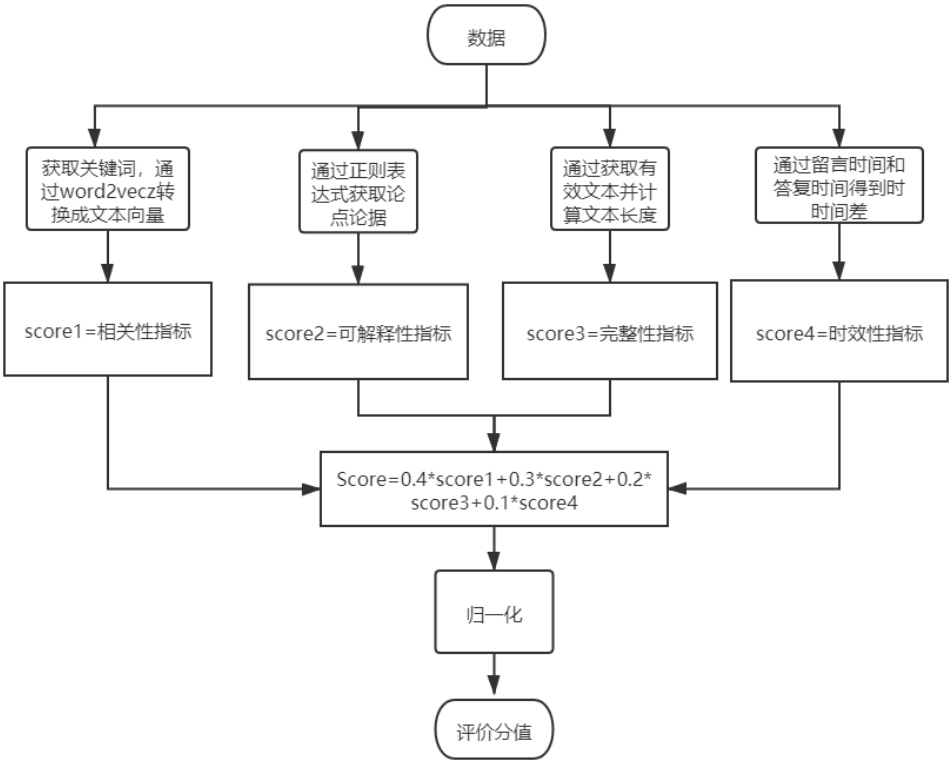


图 13 第三问算法流程图

5. 程序框架

本问程序均集成在 Question_3.PY 文件下，文件内容主要为预处理代码、4 大指标建立代码以及最终的质量评价分值代码。

4.4 模型分析

1. 模型评价

首先，本模型符合大部分人们对回复质量好坏的观点，具有较为充分的模型建立的依据。其次，本模型实现较为简单，运算成本低。

2. 成果展示

论点和论据的提取效果如图所示：

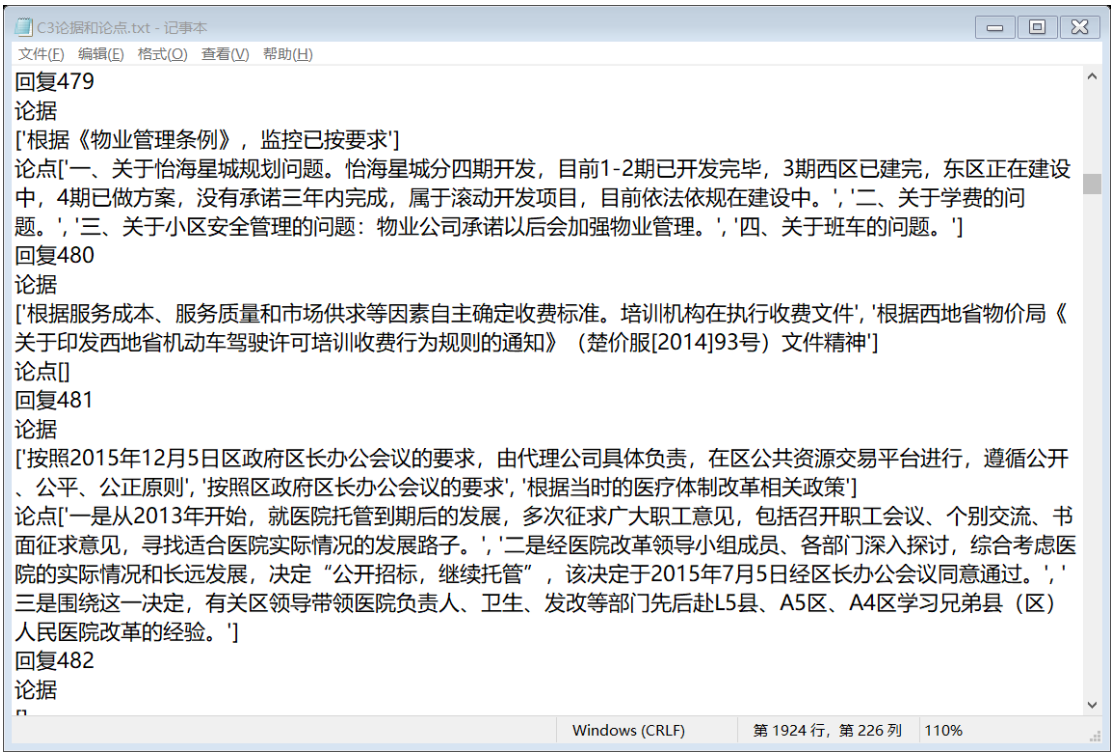


图 14 论点论据提取展示

通过对得分降序所得结果如下表所示：

表 6 第三问评价结果展示

F	G	H	I	J	K	L	M	N	O	P	Q	R	S
留言详情	答复意见	答复时间	复内容提	时间差	时效性得分	论据	论点	论据得分	论点评分	解释性得	完整性得分	相关性得分	总体评价分
咨询关于加快推进	首先感谢您对J9县	2019-01-07	为充分	2	0.333333333	['按照“统	['一、指导	0.9933	1	0.99197	1	0.8201101	1
咨询J9县林业局关	“UU0082390”首先	2019-01-03	为充分	19	0.05	['按照“统	['一、指导	0.9933	1	0.99197	0.99871959	0.83431358	0.973240825
咨询关于加快推进	“UU0082390”首先	2019-01-16	为充分	11	0.083333333	['按照“统	['一、指导	0.9933	1	0.99197	0.99871959	0.8201101	0.970498026
尊敬的各位领导：	“UU0081583”您好	2018-07-31	经调查	1	0.5	['《基本农	['一是根据	0.982	0.9997	0.97815	0.14340589	0.85078535	0.828838457
我于2017年购买	“UU0081201”您好	2019-08-06	1、延期	0	1	['《B市城区	['1、延期	0.982	0.9933	0.97306	0.040973111	0.76331675	0.820661895
据悉，因星	网友“UU0081241”	2017-10-30	轨道交	0	1	['根据地铁	['一是关键	0.8808	0.9991	0.85623	0.149807939	0.78753405	0.816460664
小区违规开	网友“UU0081323”	2016-11-24	业主您	30	0.032258065	['《医疗机	['一、关于	0.9933	1	0.99196	0.185659411	0.90316047	0.813382746
我想咨询下，我	“UU0081158”您好	2019-06-13	“UU0081	0	1	['《生育证	['一、生育	0.9526	0.9526	0.90515	0.075544174	0.76381386	0.805171304
再次对K10县委组	“UU0081304”关于	2018-06-26	“UU0081	19	0.05	['《从乡镇	['一、关于	0.9526	0.9999	0.94299	0.338028169	0.83581617	0.802426644
再次对K10县委组	“UU0081304”您好	2018-06-26	K10县逆	19	0.05	['《从乡镇	['一、关于	0.9526	0.9999	0.94299	0.33546735	0.83581617	0.801828379
融圣国际的厂	网友“UU008593”	2017-06-12	一、开	32	0.03030303	['按照地铁	['一、开发	0.9526	1	0.94309	0.241997439	0.88321885	0.799873499
尊敬的领导：	网友“UU008720”	2015-08-24	一、关	198	0.005025126	['《A7县货	['一、关于	1	1	1	0.409731114	0.7351264	0.786854936
投诉J7县安监局	我县根据人力资源和	2019-12-04	我县根据	15	0.0625	['按照县委	['一、发放	0.9526	0.9991	0.94236	0.119078105	0.90518441	0.784926252
投诉J7县安监局	现将有关情况答复如	2019-12-05	我县根据	16	0.058823529	['按照县委	['一、发放	0.9526	0.9991	0.94236	0.115236876	0.90518441	0.7835994

5 总结与展望

1. 总结

对于问题一：本文通过一系列对数据的清洗和增强，并通过 TextCNN 文本分类模型得到相当不错的效果，同时训练速度快。既保证了准确性还满足较高的效率。因此能够胜任对留言文本分类的任务。

对于问题二：本文通过精确的地点识别和适合本任务的文本聚类算法——Single-Pass 能够对话题精确有效的进行聚类提取，并且通过热度评价模型得到较为合理的结果。

对于第三问：本文通过比较合理的方法对相关性、可解释性、完整性和时效性四项指标进行计算，并且通过答复质量评价模型能够得到比较合理的结果。

2. 展望

对于第一问：使用了比较经典的模型进行文本分类，因此在数据预处理和数据增强方面需要花费大量的精力，同时模型参数也相对较多，调整参数的工作也比较繁琐，因此希望能在后期尝试更多的模型，尽可能让模型简单易用。

对于第二问：本文缺少对事件人群的提取功能，同时在地点识别期间虽然能够较为准确的提取出地名，但是任然面临着同一地名不同表达方式的识别，例如“A 市魅力之城小区”跟“A 市万科魅力之城”，因此在话题提取的时候会根据地名表达不同而将原本属于同一话题的留言分成不同话题。因此精度受到一定的影响。因此希望在后期能够针对这个问题进行完善。

对于第三问：使用了比较粗糙的论点论据提取方法，同时完整性指标的算法建立也相对简单，因此希望能够在后期加以完善得到更加有说服力的结果。

参考文献

[1] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.

[2] Rong X. word2vec parameter learning explained[J]. arXiv preprint arXiv:1411.2738, 2014

[3] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space.ICLR Workshop, 2013

[4] 李素建;基于语义计算的语句相关度研究[J];计算机工程与应用;2002 年 07 期

[5]毕华，梁洪力，王钰;重采样方法于机器学习[J].计算机学报;1016,2009,00862

附录

第一问数据清洗效果展示

初始数据：

'投诉 A 市盛世耀凯小区物业无故停水。\n\t\t\t\t\t\n\t\t\t\t\t 我在 2015 年购买了盛世耀凯小区 17 栋 3 楼，4 楼两层共计 2 千平方，一直以来我们按时足额缴纳物业费及其它费用，但由于小区的入住没有达到成立小区业委会要求，物业公司在小区为所欲为，以至于物业不是为业主服务的，而是管理业主的，想停那个的电就停那个电，想停那个水就停那个的水，小区的水电费都是我们交供电和供水公司，我就想问一问各级政府职能部门我们业主就没有一个投诉物业公司的部门了吗？\n\t\t\t\t\t\n\t\t\t\t\t\n\t\t\t\t\t'

去空字符、分词、去停用词

'投诉 盛世 耀凯 小区 物业 无故 停水 年 购买 盛世 耀凯 小区 栋 楼楼 两层 共计 千平方 以来 按时 足额 缴纳 物业费 费用 小区 入住 没有 达到 成立 小区 业委会 要求 物业公司 小区 为所欲为 物业 不是 业主 服务 管理 业主 想 停 电 停 电想 停 水 停水 小区 水电费 交 供电 供水 公司 想 问一问 各级 政府 职能部门 业主 没有 投诉 物业公司 部门'

TF-IDF 关键词提取并去单字词：

'小区 耀凯 盛世 业主 物业公司 电想 物业 楼楼 问一问 水电费 投诉 两层 为所欲为 供电 业委会 足额 停水 供水 无故 平方 物业费 共计 按时 入住 成立 各级 职能部门 达到 以来 缴纳'

TextRank 关键词提取并去单字词：

'小区 想 业主 物业公司 没有 停水 盛世 投诉 缴纳 物业 物业费 职能部门 供电 供水 交 费用 公司 水电费 停电 服务 管理 要求 耀凯 达到 业委会 成立 入住 无故 为所欲为 足额'

关键词合并:

'小区 耀凯 盛世 业主 物业公司 电想 物业 楼楼 问一问 水电费 投诉 两层 为所欲为 供电 业委会 足额 停水 供水 无故 平方 物业费 共计 按时 入住 成立 各级 职能部门 达到 以来 缴纳 小区 想 业主 物业公司 没有 停水 盛世 投诉 缴纳 物业 物业费 职能部门 供电 供水 交 费用 公司 水电费 停电 服务 管理 要求 耀凯 达到 业委会 成立 入

住 无故 为所欲为 足额'

第二问大地点获取

```
p=[i+j+k for k in list("区县") for j in list("123456789") for i in list("ABCDEFGHIJKLMNOPQRSTUVWXYZ")]
p1=np.unique([i for j in (data_1["Title"]+data_1["content"]) for i in p if i in j])
#出现过的区县名集合
p1=p1.tolist()
p1.extend(["西地省","高新区","经开区","A市"])
```

获得的大地点:

'A1区 A2区 A3区 A4区 A5区 A6区 A7县 A8县 B4区 B7县 C2区 C3区 C3县 D1区 D7县 D8县 E3区 E4区 E5区 E5县 E6县 E7县 E8县 F5县 F6县 F7县 G1区 G2区 G3县 G7县 G8县 H3县 I3县 I4县 I5县 J3县 J4县 J5县 J9县 K1区 K2区 K3县 K4县 K5县 K6县 K8县 K9县 L5县 L6县 L7县 M1区 M2县 M3县 M9县 西地省 高新区 经开区 A市'

第二问话题提取展示

话题提取和地点描述展示如表 8 和表 9 所示

A市伊景园滨河苑协商要求购房同时必须购买车位
投诉A市伊景园滨河苑捆绑车位销售
投诉A市伊景园滨河苑定向限价商品房违规涨价
A市伊景园滨河苑捆绑销售车位
A市伊景园·滨河苑欺诈消费者
A市伊景园滨河苑定向限价商品房项目违规捆绑销售车位
投诉A市伊景园滨河苑捆绑销售车位
投诉A市伊景园滨河苑捆绑车位销售
无视消费者权益的A市伊景园滨河苑车位捆绑销售行为
和谐社会背景下的A市伊景园滨河苑车位捆绑销售
投诉A市伊景园滨河苑定向限价商品房违规涨价

A 市伊景园滨河苑诈骗钱财

违反自由买卖的 A 市伊景园滨河苑车位捆绑销售行为

A 市伊景园滨河苑欺压百姓

投诉 A 市伊景园滨河苑开发商违法捆绑销售无产权车位

惊!! A 市伊景园滨河苑商品房竟然捆绑销售车位

A 市伊景园滨河苑坑害购房者

A 市伊景园滨河苑捆绑销售车位是否合理

投诉 A 市伊景园滨河苑开发商违法捆绑销售无产权车位

A 市伊景园滨河苑项目捆绑销售车位

无视职工意愿职工权益的 A 市伊景园滨河苑车位捆绑销售行为

投诉 A 市伊景园滨河苑开发商

投诉 A 市伊景园滨河院捆绑销售车位

反映 A7 县泉塘街道漓楚东路与东六路区域白天夜晚飙车扰民问题

A7 县泉塘街道漓楚东路与东六路交汇处是否可以架设人行天桥

请求解决 A7 县漓楚路与东六路上下班时间交通拥堵的问题

A7 县开元路与东六路交叉口北半幅路面被围挡占道多年

A7 县东六路与漓楚路口照明光线太暗存在极大安全隐患

建议取消 A7 县盼盼路东六路至东四路段的三条人行道

建议 A7 县在漓楚路和东六路交汇处建地下通道或人行天桥

A7 县漓楚路和东六路十字路口照明灯光太暗

希望 A7 县东六路泉塘中学段能增设人行横道

A7 县凉塘路的旧城改造要拖到何时才能动工

请问 A7 县星沙凉塘路的旧城改造要拖到何时何月何时才能再次启动

A7 县星沙凉塘路旧城改造究竟要拖到何年何月才能开始

A7 县星沙街道凉塘路的旧城改造什么时候会启动

A7 县星沙街道凉塘路旧城改造什么时候可以进行

A7 县星沙四区凉塘路旧城改造要待何时

A7 县星沙四区凉塘路旧城改造要拖到何年何月才能动工

A7 县星沙镇四区凉塘路改造何时可以开始

A 市地铁 3 号线星沙大道站地铁出入口设置极不合理!

反映 A 市地铁 3 号线松雅西地省站地下通道建设问题

希望 A 市地铁 3 号线松雅西地省站能靠近松雅湖正门的出入口

反映 A 市地铁 3 号线松雅西地省站西北方向 10 万民众安全问题

请问 A 市地铁 3 号线什么时候开通

A 市地铁 3 号线在深业睿城小区有出入口吗

A 市地铁 3 号线松雅西地省站首开为什么没有直通松雅湖方向的出入口

咨询 A 市地铁 3 号线星沙文体中心站出口设置问题

对 A 市地铁 3 号线的两点建议

A6 区月亮岛路架设高压电线环评造假谁为民众做主

反对 A6 区月亮岛路架设高压电线强烈要求重启环境评估

A6 区月亮岛路 11 万伏高压线没用地埋方式铺设

关于 A6 区月亮岛路沿线架设 110KV 高压电线杆的投诉

关于 A6 区月亮岛路沿线架设 110KV 高压线杆的投诉

关于 A6 区月亮岛路沿线架设 110KV 高压电线杆的投诉

关于 A6 区月亮岛路 110KV 高压线的建议