

一种基于自然语言处理的政务文本挖掘模型

摘要

近年来，各类社情民意相关的文本数据量不断飙升，给政府部门带来了极大的挑战，本文旨在构建基于自然语言处理（NLP）技术的文本挖掘模型，对文本数据进行分类以及聚类分析，解决政府工作部门现阶段的难题。

针对问题一：基于对“智慧政务”文本挖掘的理解和认识，本文将立足于以上背景和问题，构建基于人工神经网络库（Keras）的长短期记忆网络（LSTM）的文本分类模型，首先对数据进行清洗，再对数据进行分词、去除停用词处理后，对留言主题的特征进行提取，最后建立 LSTM 模型对留言主题完成一级分类与检验。在完成对题目所给问题集的数据分析以及预处理工作后，该模型与其他主流方案相比，在 F1-Score 上都表现出优越的效果。

针对问题二：本文构建了基于 TF-IDF 方法特征提取模型，并利用向量空间余弦相似度进行相似度分析。首先，将关键的问题进行筛选聚类，再利用熵权法排除了反对数作为热度衡量的指标，并确定将同类文本留言条数、同类文本的不同 ID 数、同类文本点赞数之和作为热度衡量指标，最后计算出热度排名前五的五个问题。

针对问题三：根据 Servqual 量度五维表，我们构建了政府答复质量评价层次结构模型，并利用隶属度转换算法和熵权法进行指标数量化以及所建立指标的权重计算，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了一套评价方案。

关键词：LSTM， F1-Score， 文本相似度分析， 熵权法， Servqual 量度五维表

A government text mining model based on natural language processing

Abstract

In recent years, the amount of text data related to various social conditions and public opinion has soared, which has brought great challenges to government departments. This article aims to build a text mining model based on natural language processing (NLP) technology to classify and aggregate text data analysis, to solve the current problems of government departments.

Aiming at Problem 1, based on the understanding and understanding of "smart government" text mining, this paper will build on the above background and problems to build a text classification model based on the artificial neural network library (Keras) long and short-term memory network (LSTM). After the data is cleaned, the data is segmented and the stop words are removed, and the characteristics of the message subject are extracted. Finally, an LSTM model is established to complete the first-level classification and test of the message subject. After completing the data analysis and preprocessing of the problem set given by the topic, the model shows superior results in F1-Score compared with other mainstream solutions.

Aiming at problem two, this paper builds a feature extraction model based on the TF-IDF method, and uses the vector space cosine similarity for similarity analysis. First, screen and cluster key issues, and then use the entropy weight method to exclude the antilogarithm as a measure of heat, and determine the number of similar text messages, the number of different IDs of similar texts, and the number of likes of similar texts The heat measurement index finally calculates the top five questions of the heat ranking.

In response to question three, according to the five-dimensional table of Servqual measures, we constructed a hierarchical structure model of government response

quality evaluation, and used membership conversion algorithm and entropy weight method to quantify the indicators and calculate the weight of the established indicators. The comments on the comments give a set of evaluation plans for the quality of the comments from the perspectives of relevance, completeness, and interpretability of the responses.

Key words: LSTM; F1-Score; Text similarity analysis; Entropy weight method; Servqual measure five-dimensional table

目录

摘要.....	1
Abstract.....	2
一、引言.....	5
二、问题一的分析与求解.....	5
2.1 模型框架.....	5
2.2 方案介绍.....	7
2.3 实验结果.....	10
三、问题二的分析与求解.....	13
3.1 模型框架.....	13
3.2 热点问题挖掘.....	13
3.3 相似度计算.....	14
3.4 熵值法衡量权重.....	16
3.5 衡量指标的确定.....	19
3.6 结果分析.....	25
四、问题三的分析与求解.....	26
4.1 答复意见质量评价体系的构建.....	26
4.2 指标数据化.....	28
4.3 运用熵权——隶属度转换算法模型评价.....	30
五、总结与展望.....	31
5.1 总结.....	31
5.2 展望.....	31
六、参考文献	32
七、附录.....	33

一、引言

近年来，随着微信、微博、市长信箱、阳光热线等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，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。同时，随着大数据、云计算、人工智能等技术的发展，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，对提升政府的管理水平和施政效率具有极大的推动作用。

在处理网络问政平台的群众留言时，工作人员首先按照一定的划分体系，对留言进行分类，以便后续将群众留言分派至相应的职能部门处理。而某一时段内群众集中反映的某一问题可称为热点问题，如“XXX 小区多位业主多次反映入夏以来小区楼下烧烤店深夜经营导致噪音和油烟扰民”。及时发现热点问题，有助于相关部门进行有针对性地处理，提升服务效率。但是目前，大部分电子政务系统还是依靠人工根据经验处理，存在工作量大、效率低，且差错率高等问题。

基于 TF-IDF 模型，本文将立足于以上背景和问题，通过数据清洗，关键词提取、文本分类、文本聚类等文本挖掘工作，实现对留言内容的一级标签分类，并通过熵权法建立合理的热度评价指标进行热点事件的确定，最终根据 Servqual 量度五维表，利用隶属度转换算法构建政府答复质量评价层次结构模型。

二、问题一的分析与求解

2.1 模型框架

针对第一问，为了使建立关于留言内容的一级标签分类模型，以便后续将群众留言分派至相应的职能部门处理，我们提出了一种基于 Keras（人工神经网络库）的长短期记忆网络（LSTM）的文本分类模型，该模型主要包括四个部分：预处理、特征提取、文本分类以及结果准确率检验。模型架构图如图 1 所示。

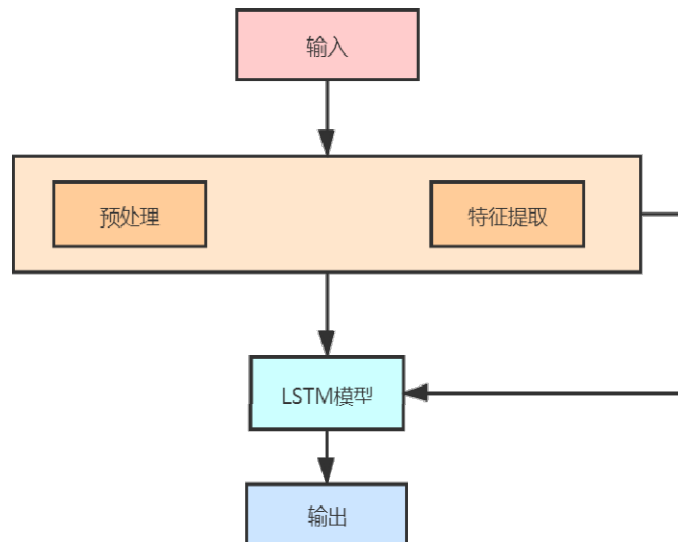


图 1 模型架构图

Step 1: 数据分析与预处理。我们对问题给出的数据集进行统计分析，并给出相应的预处理步骤；

Step 2: 特征提取：由于给出的留言主题数据量庞大，其经过数据预处理后的特征往往数量非常可观，如果不经过任何处理，直接进行文本分类，不仅对分类模型的准确率会产生一定的影响，而且其分类效率也不高。出于这些考虑，我们需要对 X' 进行特征提取，选取最有利于分类的特征向量，为后续的分类提高效率和准确率。

Step 3: 文本分类。我们采用使用 LSTM 的方法对数据进行文本分类。训练集为总数据的 65%（5894），验证集为 15%，测试集 20%首先引入我们需要的模块并预定义变量，这里我们将变量设置为留言主题，采用 Keras 给定的留言主题样本集合通过归纳分析进行训练，建立神经网络分类模型，得到分类规则，给出 LSTM 方法的算法实现。

Step 4: 模型评价。在基于 Keras 上采用 LSTM 对我们的留言主题进行文本分类后，再选取了 1474 个数据进行了检验，分别计算出了结果的准确率、宏平均和权重后，又建立了评价模型。

2.2 方案介绍

2.2.1 数据分析与预处理

在文本分类过程中，由于留言主题的多样化的特点，导致存储的大部分数据都为非结构化数据。面对这些复杂数据，计算机是无法直接处理的。这就需要先将文本进行预处理，并且将其转换成计算机能够识别出的形式。本文首先将留言主题引入数据库，把不同的评价数据分到不同的七个一级标签分类中去（分别为城乡建设、环境保护、交通运输、教育文体、劳动和社会保障、商贸旅游以及卫生计生），且每条数据只能对应七个类中的一个类，之后再对数据进行预处理。

数据预处理首先我们要对数据进行清洗，找到重复数据并去除，如表 1:

表 1 清洗掉的重复数据列表

153793	U0003678	A 市白沙晶城电梯经常发生各种掉楼层情况	2019/12/10	我是租户刚住两个月，电梯就	商贸旅游
			15:06:32	经常发生各种掉楼层...	
153876	U0008425	G1 区恒大华府的电梯三番五次出故障	2019/8/1	G 市恒大华府新交房的电梯	商贸旅游
			19:13:12	三番五次出故障...	

之后，预处理还包括删除文本中的标点符号,特殊符号，删除一些无意义的常用词，因为这些词和符号对系统分析预测文本的内容没有任何帮助，反而会增加计算的复杂度和增加系统开销,所有在使用这些文本数据之前必须要将它们清理干净，具体包括：

- （1）出去数据中的数字，保留文字留言主题
- （2）出去多余的左右空格和长度为 1 的字符
- （3）除去停用词并给数据打上标签

最后再进行分词预处理并使用向量空间模型（VSM）进行文本模式化。向量空间模型是一种使用向量表示数据的模型，通过向量空间的模式化，可以降低文本分类的难度。假设文档集合 Y 中某一文本 X，其中 Y 的文档数量为 N。对于文本 X，通过向量空间模型得到：

$$X'=\{(x_i,w_i)\}_{i=1}^n \tag{1}$$

其中 n 表示文本 X 中词的数量， x_i 表示文本 X 的第 i 个词， w_i 为 x_i 对应的特征权值，具体如下公式（2）所示：

$$w_i = f_i \log_2 \left(\frac{N}{m_{x_i}} + 0.01 \right) \quad (2)$$

其中 f_i 表示为 x_i 在文档 X 中的出现次数， m_{x_i} 表示为在集合 Y 中出现 x_i 的总文本数量。

对其进行归一化处理，则 w_i 由公式（3）所示：

$$w_i = \frac{f_i \log_2 \left(\frac{N}{m_{x_i}} + 0.01 \right)}{\sqrt{\sum_{i=1}^n \left(f_i \log_2 \left(\frac{N}{m_{x_i}} + 0.01 \right) \right)^2}} \quad (3)$$

2.2.2 特征提取

经过预处理模块后得到文本向量空间 $X' = \{(x_i, w_i)\}_{i=1}^n$ ，假设文档集合 X 的对应类别集合为， $C = \{(c_k)\}_{k=1}^l$ ，其中 l 为类别数量。

由于给出的留言主题数据量庞大，其经过数据预处理后的特征往往数量非常可观，如果不经任何处理，直接进行文本分类，不仅对分类模型的准确率会产生一定的影响，而且其分类效率也不高。出于这些考虑，我们需要对 X' 进行特征提取，选取最有利于分类的特征向量，为后续的分类提高效率和准确率。

在进行特征提取时，我们首先打散数据，形成数据分布更均匀的数据集，然后采用改进 Python 的 jieba 库对数据进行关键词提取。

jieba 是 Python 中一个重要的第三方中文分词函数库，分词原理是利用一个中文词库，将待分词的内容与分词词库进行比对，通过图结构和动态规划方法找到最大概率的词组。利用 jieba 库对数据进行词性的分词，如表 2。

表 2 利用 jieba 库对数据分词处理

输入	A3 区一米阳光婚纱摄影是否合法纳税了？
输出	[A3 区，一米，阳光，婚纱，艺术，设有，是否，合法，纳税，了]

分词后，再对留言主题进行词频统计，选出每个一级标签下的关键词。

2.2.3 文本分类

选取出了每个一级标签下的关键词后，我们要实现政务系统对关键词的自动分类，针对这个问题我们采用 LSTM 方法对数据进行文本分类[1]。

长短期记忆网络（LSTM）是一种专门用于处理时间序列数据的网络，传统的 RNN 神经网络的神经元是将输入运用函数进行计算后进行输出的单元，而 LSTM 将神经元变为记忆单元，每个记忆单元由输入门、遗忘门和输出门构成，其单元结构图如图 2 所示。

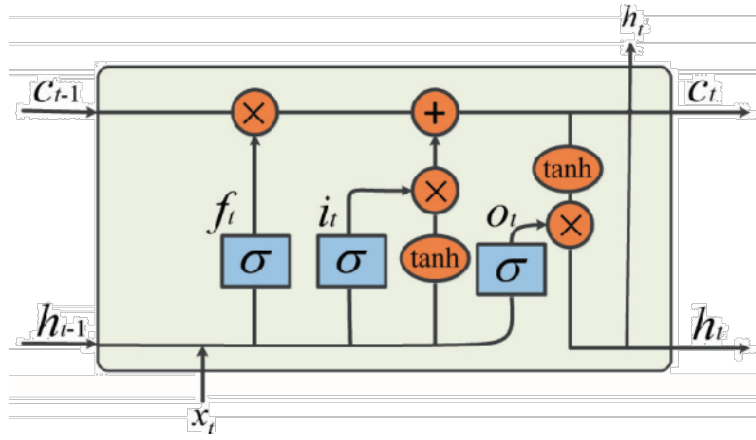


图 2 LSTM 单元结构图

其中长期状态 c 用于存储长期记忆信息，使得序列的长期状态可以保存下来，并传递到下一层，同时，遗忘门的设计又使得 c 得到更新，丢弃已经过时的信息。LSTM 的这一设计解决了 RNN 网络存在的梯度消失和梯度爆炸问题。 t 时刻的数据 x_t 到达网络后，与上一时刻 LSTM 的输出 h_{t-1} 一起作为输入，对 C_{t-1} 进行更新，得到新的长期状态 C_t ，计算公式如公式（4）所示。

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \end{aligned} \quad (4)$$

$$\bar{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$C_t = f_t \times C_{t-1} + i_t \times \bar{C}_t$$

之后，输入进行 sigmoid 计算后，与更新后的长期状态 C_t 进行计算，得到该时刻的输出 h_t ， h_t 的计算公式如公式（5）所示

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(C_t)$$

具体操作方法为:

首先引入我们需要的模块并预定义变量, 这里我们将变量设置为留言主题, 采用 Keras 给定的留言主题样本集合通过归纳分析进行训练, 建立神经网络分类模型, 得到分类规则, 给出 LSTM 方法的算法实现。

假设经过上述预处理和特征提取之后得到的文本 X 的对应特征向量为 $X'' = \{(x_i, w_i)\}_{i=1}^w$, 其中 $w \leq n$ 。通过已知对应类别标签的文本训练集对文本分类模型进行训练。本文采用 LSTM 神经网络作为文本分类模型进行分类训练。其算法伪代码如下所示: 定义输入为文本 Y, 其某个文本 X 经过预处理以及特征提取得到特征向量 $X'' = \{(x_i, w_i)\}_{i=1}^w$ 作为 LSTM 神经网络的输入节点, 输出为分类模型对所有文本集合 Y 做出的分类预测类别集合 C_Y 。

- 输入: 文本 Y
- 输出: 分类预测类别 C_Y
- 步骤 1: $Y' = \text{Pretreat}(Y)$; //其中 $X'' = \{(x_i, w_i)\}_{i=1}^w$
- 步骤 2: 以此计算文本集合 Y'中每个文本的互信息集 MI_Y^C
- 步骤 3: 根据控制阈值 b 获得模型输入特征集合 Y'';
- 步骤 4: $C_Y = \text{LSTM}(Y'')$;

对数据进行训练后, 我们开始下面我们来计算结果的准确率、宏平均和权重, 并通过求 F1 分数来评估我们模型的表现。

2.3 实验结果

本文第一问实验的数据来自题目给出的附件二中的所有数据数据。因为建立关于留言内容的一级标签分类模型, 我们选取了附件二数据中的留言主题一栏, 文本总量为 9211, 总共选取了 5894 个数据作为文本训练集用于训练模型, 剩余选取了 1474 个数据组成测试集测试分类模型的性能, 具体数据如表 3。

表 3 选取附件二的数据

数据类别	原始信息条数	比例	训练数据	验证数据	测试数据
总数	9210	1	5894	1842	1474
城乡建设	2009	0.22	1286	402	322
环境保护	938	0.10	600	188	150
交通运输	613	0.067	392	123	98
教育文体	1589	0.173	1017	318	254
劳动和社会保障	1969	0.213	1260	394	315
商贸旅游	1215	0.132	778	243	194
卫生计生	877	0.095	561	175	140

经过训练以及测试，模型的 acc、val_acc、loss 与 Epoch 的关系如图 3 所示：

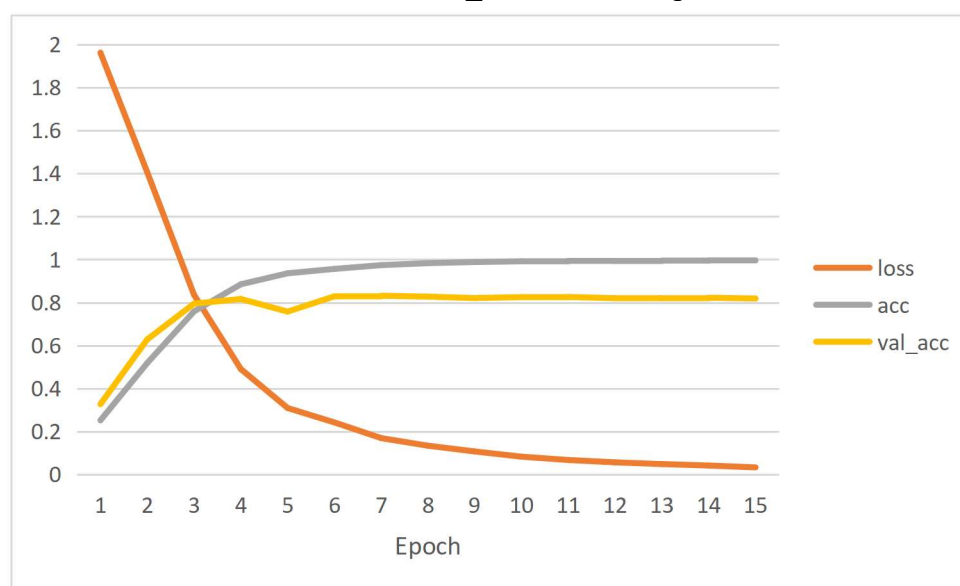


图 3 acc、val_acc、loss 与 Epoch 的关系

结果的准确率、宏平均、权重和 F₁-score 如下所示：

表 4 结果的 F₁-score 值

	Precision	Recall	F ₁ -score	Support
城乡建设	0.81	0.77	0.79	429
环境保护	0.84	0.81	0.83	187
交通运输	0.74	0.73	0.74	112
教育文体	0.88	0.86	0.87	325

劳动和社会保障	0.86	0.87	0.87	374
商贸旅游	0.72	0.85	0.78	227
卫生计生	0.84	0.83	0.83	188

表 5 结果的结果的准确率、宏平均、权重

精确度			0.82	1842
宏平均	0.81	0.82	0.81	1842
权重	0.82	0.82	0.82	1842

经过检验结果的 F_1 -score 值平均可以达到 81.5% 以上，测试结果较好，说明我们的文本分类成果很好。

三、问题二的分析与求解

3.1 模型框架

某一时段内群众集中反映的某一问题可成为热点问题，在大量的数据中，及时发现问题，有关部门再进行针对性处理，一方面可以化解舆论压力，解决百姓身边的困难，同样也可以减轻政府的工作压力，提高效率。为了找出附件中的热点问题，我们构建了基于 TF-IDF 方法特征提取模型将关键的问题进行筛选聚类，再利用熵权法排除了反对数作为热度衡量的指标对热点问题进行了排序并建立热点问题留言明细表。



图 4 问题二模型流程图

3.2 热点问题挖掘

我们将留言主题与留言内容提取出来后，制作成语料库，对其使用 TF-IDF 方法进行特征提取。TF-IDF 可以评估某一个字词对于一个语料库中的重要程度，字词的重要性随着它在语料库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。因此，TF-IDF 可以帮助我们过滤掉常见的词语，并且提取出重要的词语。

TF-IDF 的主要原理是[2]：如果某个词语在某一文档中出现的频率很高，但在其他文章中的频率很低，那么可以认为这个词语具有很好的分类能力。TF-IDF 实际上是 $TF(\text{词频}) * IDF(\text{逆向文件频率})$ 。TF 表示词条在文档 d 中出现的频率，IDF 的意义是若包含词条 t 的文档越少，即 n 越小，IDF 越大，则说明词条 t 具有较强的分类能力。若某类文档 C 中包含词条 t 的文档数为 m ，而其它类包含 t 的文档总数为 k ，则所有包含 t 的文档数 $n=m+k$ ， m 越大， n 越大，按照公式得到的 IDF 的值越小，则说明该词条 t 的分类能力较弱。

IF(词频)是对词数的归一化，以防止它偏向长的文件。对于在某一文件中的

词语 t_i ，它的重要性可表示为：

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (6)$$

其中 $n_{i,j}$ 是该词语在文件 d_j 中的出现次数，分母是在文件 d_j 中所有字词的出现次数之和。

IDF(逆向文件频率)用来衡量一个词语的普遍重要性。某一特定词语的 IDF，以由总文件数除以包含该词语的文件数目，再将得到的商取对数：

$$idf_i = \log \frac{|D|}{|\{j: t_i \in d_j\}|} \quad (7)$$

上式中 $|D|$ 是语料库中的文件总数， $|\{j: t_i \in d_j\}|$ 包含词语 t_i 的文件数目（即 $n_{i,j} \neq 0$ 的文件数目），若该词语不在语料库中，则被除数变为零，因此一般情况下使用 $1 + |\{j: t_i \in d_j\}|$ 。

3.3 相似度计算

我们使用 TF-IDF 提取出关键词后，再进行相似度计算。对于多个不同的文本，我们将这些文本中的词语映射到向量空间，形成文本中文字和向量数据的映射关系，通过计算几个或者多个不同的向量的差异的大小，来计算文本的相似度。

本文中，我们采用向量空间余弦相似度计算方法。余弦相似度用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异的大小。余弦值越接近 1，就表明夹角越接近 0 度，也就是两个向量越相似，这就叫"余弦相似性"。

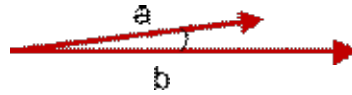


图 5 向量空间中两个向量夹角

图 5 中向量 a, b 的夹角很小，则说明向量 a, b 的相似性很高，即向量 a, b 代表的文本较为相似。

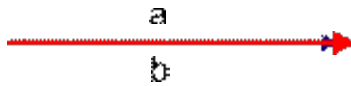


图 6 向量空间中两个向量夹角

图 6 中向量 a, b 重合，则可认为两向量相等，即 a, b 向量代表的文本是相等的。

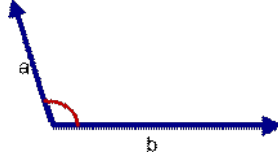


图 7 向量空间中两个向量夹角

图 7 中：向量 a, b 的夹角较大，则说明向量 a, b 相似度较小，即 a 和 b 向量代表的文本基本不相似。

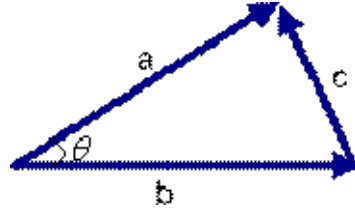


图 8 向量空间中的三角形

三角形中边 a 和 b 的夹角 的余弦计算公式为：

$$\cos(\theta) = \frac{a^2 + b^2 - c^2}{2ab} \quad (8)$$

在向量表示的三角形中，假设 a 向量是 (x_1, y_1) ， b 向量是 (x_2, y_2) ，则可以将余弦定理改写成下面的形式：

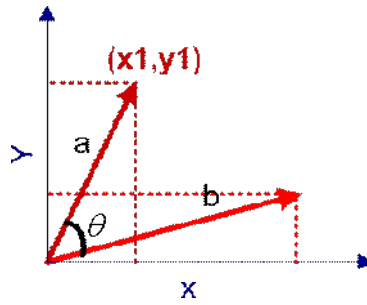


图 9 向量空间中的三角形中的 a, b

$$\begin{aligned} \cos(\theta) &= \frac{a \cdot b}{\|a\| \times \|b\|} \\ &= \frac{(x_1, y_1) \cdot (x_2, y_2)}{\sqrt{x_1^2 + y_1^2} \times \sqrt{x_2^2 + y_2^2}} \end{aligned} \quad (9)$$

$$= \frac{x_1 x_2 + y_1 y_2}{\sqrt{x_1^2 + y_1^2} \times \sqrt{x_2^2 + y_2^2}}$$

若向量 **a** 和 **b** 不是二维而是 **n** 维。假设 **a** 和 **b** 是两个 **n** 维向量，则 **a** 与 **b** 的夹角的余弦等于：

$$\begin{aligned} \cos(\theta) &= \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}} \\ &= \frac{a \cdot b}{\|a\| \times \|b\|} \end{aligned} \quad (10)$$

余弦值越接近 1，就表明夹角越接近 0 度，也就是两个向量越相似，夹角等于 0，即两个向量相等，即"余弦相似性"。

3.4 熵值法衡量权重

3.4.1 关键事件类别数据的预处理

经过相似度分析，我们得到出现频次较多的关键地点和事件，共 23 类，相似度分析能够在大数据中寻找出同一个事件出现过较多的次数，但是我们需要考虑到，某一事件仅被少数人提起，但是点赞量较多，也有可能是热点事件，因此我们通过 excel 进行对点赞量的排序，对此在 23 类的基础上，新增 3 类数据。

表 6 点赞数排序表

留言编号	留言用户	留言主题	留言时间	反对数	点赞数
253369	A00074795	穿 A 市城而过的京港澳高速（长楚高速）什么时候可以外迁至远郊？	2019/11/18 15:35:11	0	29
284571	A00074795	建议西地省尽快外迁京港澳高速城区段至远郊	2019/1/10 15:01:26	0	80
193091	A00097965	A 市富绿物业丽发新城强行断业主家水	2019/6/19 23:28:27	0	242
221996	A00080850	A 市博才长房云时代小学迟迟不开学，上万业主心急如焚	2019/1/4 11:36:26	0	28

26 类关键地点以及事件如下表所示如下表 7 所示（前后排序无特殊意义）：

表 7 经相似度分析所得关键地点和事件

地点、事件	
月亮岛 高压	三一大道改造
万润滨江装修	A7 华苑地下车位
温斯顿英语	房云时代房屋质量
凉塘路	丽发新城 搅拌
伊景园车位	郝家坪小学
车贷	辉煌国际饭店
魅力之城扰民	城际空间站
绿地海外滩扰民	人才
万境 房屋	A7 县社塘路白改黑
泉星公园优化	青青家园
和包支付	京港澳
丽发新城 断水	房云时代不开学
天池物业问题	A7 华苑两证合一

将 matlab 导出的每一类相似度较高的事件所包含的单位事件进行查找并进行整理，去除留言详情一列，以“万境 房屋”为例，如下表 8 所示（26 类详情见附录 1）：

表 8 “万境 房屋”类别所含具体数据

留言编号	留言用户	留言主题	留言时间	反对数	点赞数
208636	A00077171	A 市 A5 区汇金路五矿万境 K9 县存在一系列问题	2019/8/19 11:34:04	0	2097
234086	A00099869	A 市五矿万境 K9 县房子的墙壁又开裂了	2019/6/20 9:30:44	0	6
208069	A00094436	A5 区五矿万境 K9 县的开发商与施工方建房存在质量问题	2019/5/5 13:52:50	0	2
215507	A000103230	A 市五矿万境 K9 县存在严重的消防安全隐患	2019/9/12 14:48:07	0	1
252650	A00010531	A 市五矿万境 K9 县交房后仍存在诸多问题	2019/9/11 15:16:02	0	0
262599	A000100428	A 市五矿万境 K9 县房屋出现质量问题	2019/9/19 17:14:49	0	0

3.4.2 熵值法衡量权重介绍

熵是不确定性的度量，如果用 P_j 表示的 j 个信息不确定度（也即出现的概率）

则整个信息（设有 n 个）的不确定度量也可用下式表示：

$$S = -K \sum_{j=1}^n P_j \ln(P_j) \quad (11)$$

这就是熵。其中 K 为正常数，当各个信息发生的概率相等时，即 $P_j = 1/n$, S 取值最大，此时熵最大。

可利用熵信息的概念确定权重，假设多属性决策矩阵如下：

$$M = \begin{matrix} & \begin{matrix} A_1 & A_2 & \vdots & A_m \end{matrix} \\ \begin{matrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{matrix} \end{matrix} \quad (12)$$

则用 $P_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}}$ 表示第 j 个属性下第 i 个方案 A_i 的贡献度。

可以用 E_j 来表示所有方案对属性 X_j 的贡献总量：

$$E_j = -K \sum_{i=1}^m P_{ij} \ln(P_{ij}) \quad (13)$$

其中，常数 $K = 1/\ln(m)$ ，这样，就能保证 $0 \leq E_j \leq 1$ ，即 E_j 最大为 1。

由式中可以看出，当某个属性下各方案的贡献度趋于一致时， E_j 趋于 1；特别是当全相等时，也就可以不考虑该目标的属性在决策中的作用，也即此时属性的权重为零。

这样，可看出属性值由所有方案差异大小来决定权系数的大小。为此可定义 d_j 为第 j 属性下各方案贡献度的一致性程度。

$$d_j = 1 - E_j \quad (14)$$

则各属性权重 W_j 如下：

$$W_j = \frac{d_j}{\sum_{j=1}^n d_j} \quad (15)$$

当 $d_j = 0$ 时，第 j 属性可以剔除，其权重等于 0。

如果决策者事先已有一些经验的主观估计权重 λ_j ，则可借助上述的 W_j 来对 λ_j 进行修正。

$$W_0^j = \frac{\lambda_j W_j}{\sum_{j=1}^n \lambda_j W_j} \tag{16}$$

熵值法最大的特点是直接利用决策矩阵所给出的信息计算权重，而没有引入决策者的主观判断。

3.5 衡量指标的确定

3.5.1 留言条数、反对数、点赞数的确定

显而易见，根据题目以及所给数据，我们第一想法是将每一类的留言条数和、反对数和、点赞数和作为指标，我们将附录一所提 26 类关键事件进行对每一类的留言条数、反对数、点赞数求和，统计数据如下表 9：

表 9 26 类数据的留言条数、点赞数、反对数统计

事件	留言条数	反对数	点赞数
车贷	13	0	2383
万境	11	0	2112
绿地海外滩扰民	7	0	693
凉塘路	12	1	69
丽发新城 搅拌站	46	0	15
人才	28	4	42
月亮岛 高压	7	2	173
伊景园车位	28	1	18
泉星公园优化	10	0	51
和包支付	4	0	86
丽发新城 断水	1	0	242
京港澳	2	0	109
三一大道改造	3	0	74
魅力之城扰民	21	18	18
郝家坪小学	3	0	57
辉煌国际饭店	11	1	6
天池物业问题	7	2	39
万润滨江装修	4	0	28

城际空间站	6	0	18
A7 华苑地下车位	3	0	28
温斯顿英语	7	0	4
房云时代房屋质量	3	5	55
青青家园	8	1	0
A7 县社塘路白改黑	4	3	23
房云时代不开学	1	0	28
A7 华苑两证合一	4	0	2

Step 1:

将表 9 转化为矩阵:

$$M = \begin{bmatrix} 13 & 0 & 2383 \\ 11 & 0 & 2112 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 4 & 0 & 2 \end{bmatrix}$$

Step 2:

求得每一类留言条数、反对数、点赞数对关键事件的贡献度, 公式为:

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{26} x_{ij}} \quad (17)$$

将其转化为新的矩阵:

$$P = \begin{bmatrix} 0.02755905 & 5 & 0 & 0.33139808 & 6 \\ 0.18110236 & 2 & 0 & 0.00235368 & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \\ 0.02362207 & 4 & 0 & 0.00282441 & 6 \end{bmatrix}$$

Step 3:

$$\ln(P) = \begin{bmatrix} -3.591424118 & \infty & -1.10443495 \\ -1.708692871 & \infty & -6.051775394 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ -3.745574798 & \infty & -5.869453837 \end{bmatrix}$$

$$P \cdot \ln(P) = \begin{bmatrix} -0.09897625 & 0 & -0.36600762 \\ -0.30944831 & 0 & -0.01424394 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ -0.08847814 & 0 & -0.01657777 \end{bmatrix}$$

Step 4:

求出所有关键事件类别对留言条数、反对数、点赞数的贡献总量，其中：

$$E_j = -K \sum_{i=1}^{26} P_{ij} \ln(P_{ij}) \quad (18)$$

$$K = \frac{1}{\ln(26)} = 0.30692776430135 \quad (19)$$

得到 E 的矩阵 $E = [0.861413577 \quad 0.537452895 \quad 0.546728753]$

Step 5:

计算留言条数、反对数、点赞数下各关键事件类别贡献度的一致性程度，从而得到 d 的矩阵 $D = [0.138586423 \quad 0.462547105 \quad 0.453271247]$ 。最终得到：

$$\sum d_j = 1.05440477 \quad (20)$$

Step 6:

计算留言条数、反对数、点赞数的权重：

$$W_j = \frac{d_j}{\sum_{j=1}^3 d_j} = [0.131435694 \quad 0.438680776 \quad 0.42988353]$$

即留言条数所占的权重为 13.143%，反对数所占的权重为 43.868%，点赞数所占权重为 42.988%。

3.5.2 指标标准化+排除反对数作为指标

通过上述熵值法所得，留言条数所占的权重为 13.143%，反对数所占的权重为 43.868%，点赞数所占权重为 42.988%，在表 9 中，我们可以看出点赞数的数值是远远大于留言条数的数值，每个指标的数值与其权重的乘积之和为其热度综合评价价值，因此基于现在所建立的“留言条数+反对数+点赞数”的指标因素所得的最后结果，点赞数的评价得分是远远大于留言条数的评价得分，且反对数大多值为 0，但所占权重也要远远高于留言条数所占权重，最后结果会出现，点赞数最多的即是热点事件，这显然不符合常理，所以现在建立的指标因素以及数据处理是不合理的。

因此我们为了避免上述点赞数权重过大和反对数不合理的问题，我们将表 9 中每一类关键问题所涉及的不用用户 ID 个数作为指标，并将留言条数和、点赞数和、不同 ID 数进行由高到低排序，数值高的得分高，数值低的分数低，采用名次数据进行熵值法来衡量权重。

3.5.3 留言条数名次、点赞数名次、不同 ID 数名次作为指标

表 10 26 类数据的留言条数、点赞数、不同 ID 数统计与名次排序

事件	留言条数	点赞数	不同 ID	留言条数名次	点赞数名次	不同 ID 数名次
丽发新城 搅拌站	46	15	43	14	5	12
人才	28	42	26	13	10	11
伊景园车位	28	18	26	13	6	11
魅力之城扰民	21	18	17	12	6	10
车贷	13	2383	11	11	22	9
凉塘路	12	69	10	10	14	8
温斯顿英语	7	4	7	6	3	7
万境	6	2106	6	5	21	6
月亮岛 高压	7	173	5	6	18	5
泉星公园优化	10	51	5	8	11	5
天池物业问题	7	39	5	6	9	5
绿地海外滩扰民	7	693	4	6	20	4
城际空间站	6	18	4	5	6	4
青青家园	8	0	4	7	1	4
和包支付	4	86	3	4	16	3

三一大道改造	3	74	3	3	15	3
房云时代房屋质量	3	55	3	3	12	3
万润滨江装修	4	28	3	4	8	3
辉煌国际饭店	11	6	3	9	4	3
郝家坪小学	3	57	2	3	13	2
A7 县社塘路白改黑	4	23	2	4	7	2
A7 华苑两证合一	4	2	2	4	2	2
丽发新城 断水	1	242	1	1	19	1
京港澳	2	109	1	2	17	1
A7 华苑地下车位	3	28	1	3	8	1
房云时代不开学	1	28	1	1	8	1

Step 1:

将表 7 名次行列转化为矩阵:

$$M = \begin{bmatrix} 14 & 5 & 12 \\ 13 & 10 & 11 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & 8 & 1 \end{bmatrix}$$

Step 2:

求得每一类留言条数名次、点赞数名次、不同 ID 数名次对关键事件的贡献度，
公式为:

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^{26} x_{ij}} \quad (21)$$

将其转化为新的矩阵:

$$P = \begin{bmatrix} 0.08588957 & 0.01779359 & 0.09523809 \\ 0.07975460 & 0.03558718 & 0.08730158 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 0.00613496 & 0.02846975 & 0.00793650 \end{bmatrix}$$

Step 3:

$$\ln(P) = \begin{bmatrix} -2.45469287 & -4.02891675 & -2.35137525 \\ -2.52880084 & -3.33576957 & -2.43838663 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ -5.09375020 & -3.55891312 & -4.83628190 \end{bmatrix}$$

$$P\ln(P) = \begin{bmatrix} -0.21083251 & -0.0716889 & -0.22394050 \\ -0.20168350 & -0.11871066 & -0.21287502 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ -0.03125000 & -0.1013213 & -0.0383831 \end{bmatrix}$$

Step 4:

求出所有关键事件类别对留言条数名次、点赞数名次、不同 ID 数名次的贡献总量:

$$E_j = -K \sum_{i=1}^{26} P_{ij} \ln(P_{ij})$$

$$K = \frac{1}{\ln(26)} = 0.30692776430135$$

最终得到 E 的矩阵 $E = [-0.944405794 \quad -0.948161195 \quad -0.929492574]$

Step 5:

计算留言条数名次、点赞数名次、不同 ID 数名次下各关键事件类别贡献度的一致性程度, 得到 d 的矩阵 $D = [0.055594206 \quad 0.051838805 \quad 0.070507426]$

$$\sum d_j = 0.177940437$$

Step 6:

计算留言条数名次、点赞数名次、不同 ID 数名次的权重:

$$W_j = \frac{d_j}{\sum_{j=1}^3 d_j} = [0.312431547 \quad 0.291326726 \quad 0.396241728]$$

即留言条数名次所占的权重为 31.2431547%, 点赞数名次所占的权重为 29.1326726%, 不同 ID 数名次所占权重为 39.6241728%。

Step 7:

我们将每个关键事件的名次得分与其权重的乘积之和作为其综合热度评价
值：

$$M' = \sum_{j=1}^3 M_j * W_j = \begin{bmatrix} 10.58557601 \\ 11.33353637 \\ \cdot \\ \cdot \\ \cdot \\ 3.03928708 \end{bmatrix} \quad (22)$$

3.6 结果分析

将最后得分导入 excel 中进行排序，得到顺序如下表：

表 11 热点问题排序表

排名	事件	综合评价得分	排名	事件	综合评价得分
1	车贷	13.41211053	14	丽发新城 断水	6.243881064
2	人才	11.33353637	15	京港澳	5.973659159
3	丽发新城 搅拌站	10.58557601	16	房云时代房屋质量	5.621940532
4	凉塘路	10.37282345	17	温斯顿英语	5.52226155
5	伊景园车位	10.16822946	18	郝家坪小学	5.51702553
6	万境	10.05746934	19	辉煌国际饭店	5.165916005
7	魅力之城扰民	9.45955619	20	城际空间站	4.895084998
8	绿地海外滩扰民	9.286090706	21	万润滨江装修	4.769065175
9	月亮岛 高压	9.099678982	22	A7 县社塘路白改黑	4.081496722
10	泉星公园优化	7.685254994	23	青青家园	4.063314462
11	和包支付	7.099678982	24	A7 华苑地下车位	3.664150174
12	三一大道改造	6.495920709	25	房云时代不开学	3.03928708
13	天池物业问题	6.47773845	26	A7 华苑两证合一	2.624863093

我们发现排名前五位的热点问题的关键词是：车贷、人才、丽发新城 搅拌站、凉塘路、伊景园车位，排名前五的热点问题集见文件“热点问题.xls”，相应的热点问题对应的留言信息见“热点问题留言明细表.xls”

四、问题三的分析与求解

4.1 答复意见质量评价体系的构建

4.1.1 SERVQUAL 模型介绍

SERVQUAL 是 Service Quality 的缩写，即服务质量，模型的构成如图 10 所示：

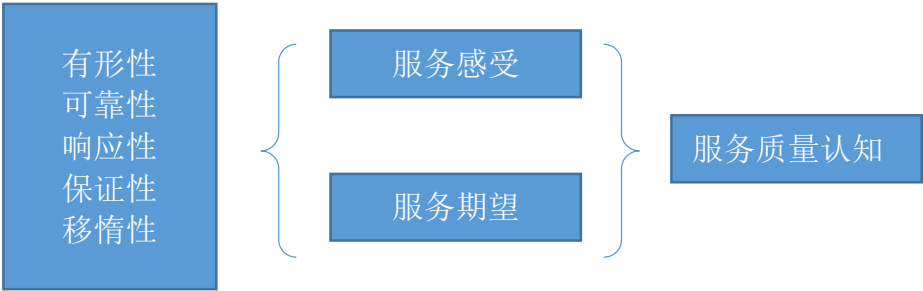


图 10 评价服务质量模型

为此，PZB 提出 5 个对服务质量的评价有关的服务质量维度，并根据 5 个维度设计了包括有 22 个问项的 SERVQUAL 量表[3]，其中 5 个维度如表 12 所示：

表 12 SERVQUAL 量表五维度

指标	意义
有形性	公司和公司内服务人员的形象等
可靠性	对履行服务承诺能力的可靠性
响应性	对提供便捷服务的时效性
保证性	顾客对公司的认可和信任程度
移情性	企业和顾客之间的友好关系

4.1.2 政府答复质量评价指标体系

根据 SERVQUAL 模型，结合政府答复质量评价的目的，运用层次分析法建立层次结构，其中政府答复质量评价为层次结构模型的目标层 A，影响政府答复质量的 4 个维度为层次结构模型的准则层 B，7 个细化指标为层析结构模型的指标层 C，提出新的政府答复质量评价指标体系，具体过程如下

（1）隶属度转换算法

隶属度转换算法的转换过程采用 $M(1, 2, 3)$ 模型，即“1 有效、2 可比、3 转换”的计算步骤。

(2) 确定指标区分权

已知影响目标 Q 状态的 n 个层次 m 个指标, 每个指标可以划分为 P 个等级, 表示第 C_k 个等级, 且 Q 级到 $C_k = 1$ 级划分有序。如果目标 Q 的底层 j 指标属于 C_k 等级的隶属度为 $\mu_{jk}(Q)$ ($k=1\cdots p, j=1\cdots m$)。已知, 且满足: $0 \leq \mu_{jk}(Q) \leq 1$, $\sum_{k=1}^p \mu_{jk}(Q) = 1$ 。若 $\mu_{jk}(Q)$ ($k=1\cdots p, j=1\cdots m$) 为目标 Q 的 k 指标属于 C_k 级的隶属度, 则可得到公式 x 和公式 x :

$$V_j(Q) = 1 - \frac{1}{\ln p} \left(- \sum_{k=1}^p \mu_{jk}(Q) \cdot \ln \mu_{jk}(Q) \right) \quad (23)$$

$$\alpha_j(Q) = \frac{v_j(Q)}{\sum_{i=1}^n v_i(Q)}$$

(3) 确定指标的类有效值和可比值

指标 j 属于等级 C_k 的隶属度 $\mu_{jk}(Q)$ 和指标 j 对目标 Q 的权重 $\beta_j(Q)$ 已知, 由此确定 j 指标的 k 类隶属度的有效区分值即 k 类有效值为 $\alpha_j(Q) \cdot \mu_{jk}(Q)$

($k=1, 2, \dots, p$)。再将有效值转化为可比值, 可比值矩阵 $N(Q)$ 如公式 6 所示

$$N(Q) = \beta_j(Q) \cdot \alpha_j(Q) \cdot \mu_{jk}(Q) \quad (k=1\cdots p) \quad (24)$$

(4) 确定可比和向量

若 $N(Q) = \beta_j(Q) \cdot \alpha_j(Q) \cdot \mu_{jk}(Q)$ ($k=1\cdots p$) 为 j 指标关于目标 Q 的 k 类可比值则称 $N_k(Q)$ 为目标 Q 的 k 类可比和, 可得公式

$$N_k(Q) = \sum_{j=1}^n \beta_j(Q) \cdot \alpha_j(Q) \cdot \mu_{jk}(Q) \quad (k=1\cdots p) \quad (25)$$

当目标 Q 的 k 类可比和越大, 则目标 Q 属于 C_k 级的程度就越大, 反之就越小。

(5) 确定上层指标的隶属度

经上述转换计算后的隶属度向量如公式 (26) 所示

$$\mu_k(Q) = \frac{N_k(Q)}{\sum_{i=1}^p N_i(Q)} \quad k = (1, 2, 3 \dots) \quad (26)$$

(6) 评价结果判别

评价等级的判别有最大隶属度准则法和置信度识别准则法。最大隶属度准则只在无序划分的情况下适用，若在有序划分(即 C_k 等级优于 C_{k+1})的情况下使用最大隶属度准则会出现不合理的结果。

设 λ ($0.6 < \lambda < 1$) 为置信度，则可得到公式 xx:

$$K = \min \left\{ k \left| \sum_{i=1}^k \mu(Q) \geq \lambda, 1 \leq k \leq 5 \right. \right\} \quad (27)$$

由此可以判定目标 Q 属于 C_k 等级。

4.2 指标数据化

运用熵权——隶属度转换算法[5]，可以将政府答复服务质量的评价分为“非常满意、比较满意、一般、比较不满意、不满意”5 个等级，评语等级为 C，则 $C = \{C1, C2, C3, C4, C5\} = (\text{非常满意}, \text{比较满意}, \text{一般}, \text{比较不满意}, \text{不满意})$

4.2.1 SERVQUAL 模型量度的规定

我们同样应用第一问的数据预处理来实现答复文本的中文分句、中文分词、词性标注和依存句法分析,本文利用 python 编写程序,首先进行了政府答复的分句处理,然后利用分句结果进一步展开了分词、词性标注和依存句法解析处理,最终得到了能够反映分句中各词语之间的依赖关系的结构化数据。

(1) 相关性：我们通过 python 编程得到答复与问题的相似度，并按照比例赋值如下表 13：

表 13 答复与问题的相似度

相关度	隶属度得分
0-0.2	1
0.3-0.4	2

0.5-0.6	3
0.7-0.8	4
0.9-1.0	5

（2）完整性：答复完整性主要包括：关键词涵盖面、答复态度。其中答复态度包括具体的礼貌性答复格式、是否给出后续的联系方式等。完整性我们采取关键点得分的方法，其中关键词涵盖面最高分为 3，礼貌性答复格式、是否给出后续联系方式各占 1 分。

表 14 完整性分数分配

关键词涵盖面 3	礼貌性答复格式 1	是否给出后续联系方式
1——2——3	1	1

（3）可解释性：可解释性同样采取赋值法，可解释性即为政府针对群众所提问问题的具体的解决意见：所反映问题的现状、如何解决、什么时候解决三方面，具体分配方式如下：

表 15 可解释性分数分配

问题现状	如何解决	是么时候解决
1	2	2

（4）对于群众来说，所提问题时效性显得更为重要，如果群众的问题很急，那么长时间给不了答复，那么政府的答复就失去了意义。根据所给附件进行答复时间与留言时间的时间差计算，得到数据如下图 11 所示。

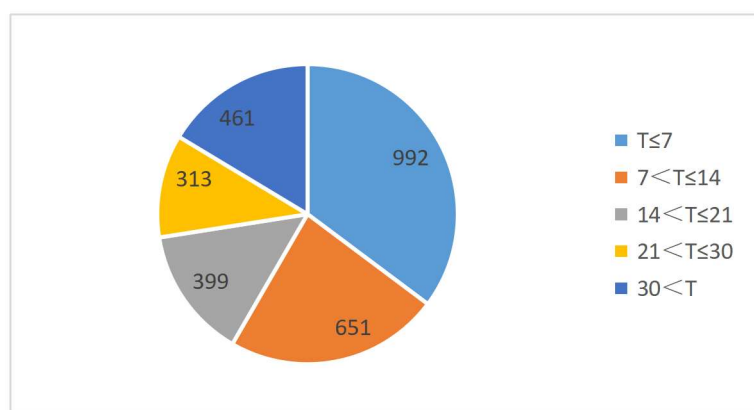


图 11 答复时间与留言时间的时间差

因此我们定义如下得分：

表 16 时效性分数分配

时间/天	$T \leq 7$	$7 < T \leq 14$	$14 < T \leq 21$	$21 < T \leq 30$	$30 < T$
得分	5	4	3	2	1

4.2.2 数据收集与整理

对通过第一问算法进行分类以及处理数据后，得到指标隶属度如下表所示：

表 17 指标隶属度

一级指标	二级指标	得 5 分	得 4 分	得 3 分	得 2 分	得 1 分
B1	B11	0.642	0.123	0.137	0.09	0.008
B2	B21	0.429	0.279	0.150	0.114	0.029
	B22	0.600	0.207	0.150	0.043	0.00
B3	B31	0.564	0.250	0.128	0.50	0.007
	B32	0.250	0.493	0.137	0.100	0.021
	B33	0.100	0.236	0.007	0.071	0.586
B4	B41	0.352	0.231	0.141	0.111	0.163

4.3 运用熵权——隶属度转换算法模型评价

4.3.1 熵权法计算指标权重

以表 17 中的指标隶属度为原始数据计算各指标权重如表 18 所示。

表 18 各指标权重

一级指标	权重	二级指标	权重
B1	0.158	B11	1
B2	0.164	B21	0.703
		B22	0.297
B3	0.509	B31	0.181
		B32	0.436
		B33	0.383
B4	0.299	B41	1

因此通过所建立的模型和熵权——隶属度转换算法[4]，可以得到，相关性所占权重为 15.8%，完整性所占权重为 16.4%，可解释性所占权重为 50.9%，时效性所占权重为 29.9%。

4.3.2 结语与建议

一、答复要明确

政府在回复群众的留言时，应该更注重具体的解决方式，或者需要政府出面解决的事情要说明解决的日期，给群众一个明确的答复。

二、答复要及时

通过附件四做日期差，我们发现有关部门做出的答复是在群众留言 1160 天，相如这样的答复，就失去了群众留言本身的意义。因此及时答复及时解决群众的问题，才能增加群众对政府的可信度。

五、总结与展望

5.1 总结

为了满足政府工作部门快速处理大刘文本数据的需求，本文基于人工智能大数据开发相关理论和实验，构建了一个智慧政务模型，让相关政府工作人员更高效、准确得实现人与机器的交流，模型的主要分为数据预处理、文本分类、文本聚类，并通过熵权法、相似度及算法解决，在对赛题的研究基础上，我们根据研究思路撰写论文，通过实验验证了本模型得可行性，较好实现了本赛题设立的目标。

5.2 展望

5.2.1 需进一步深化文本训练模型

在文本分类、聚类的模型上，应该考虑到群众在写留言时的心理情感，应从关键词折射出群众的情感动态，但是由于我们的知识有限，无法达到较深层次来建立模型。

5.2.2 需进一步热度评定模型

本文通过熵权法分析排除了反对数作为评判指标，并建立了留言条数、点赞数、不同 ID 数作为指标，但此种做法仅能在宏观上得出最终热点事件来，不能够较好地将时间作为一种变量一并考虑进去。

5.2.3 仍需进一步改进答复评价模型

本文以群众为主体，从群众期待如何解决问题与政府所给建议之间形成的落差入手，依据 SERVQUAL 模型建立了一套评判标准，但是在将文本数据化的过程中，应通过情感极性判断算法与情感强度计算规则更好的将文本赋值，但是我们由于受到知识限制，尝试过多次，未能实现此方法。

六、参考文献

- [1]杨帆.基于深度学习的法院信息文本分类[A].湖北工业大学,2019.
- [2]杨震.基于字符串相似性的网络短文本舆情热点发现技术[A].北京工业大学,2010.
- [3]邹沁含.交互文本质量评价模型的构建与实践[A].北京师范大学，2020.
- [4]李慧宗.基于熵权法的快递企业物流服务质量评价[A].安徽理工大学，2018.
- [5]徐奇钊.基于文本挖掘的文本情绪分类.云南财经大学，2016.

七、附录

附录一

第二问 26 类文本聚类详情

万境 房屋					
208636	A00077171	A 市 A5 区汇金路五矿万境 K9 县存在一系列问题	2019/8/19 11:34:04	0	2097
234086	A00099869	A 市五矿万境 K9 县房子的墙壁又开裂了	2019/6/20 9:30:44	0	6
208069	A00094436	A5 区五矿万境 K9 县的开发商与施工方建房存在质量问题	2019/5/5 13:52:50	0	2
215507	A000103230	A 市五矿万境 K9 县存在严重的消防安全隐患	2019/9/12 14:48:07	0	1
252650	A00010531	A 市五矿万境 K9 县交房后仍存在诸多问题	2019/9/11 15:16:02	0	0
262599	A000100428	A 市五矿万境 K9 县房屋出现质量问题	2019/9/19 17:14:49	0	0
丽发新城 搅拌					
188809	A909139	A 市万家丽南路丽发新城居民区附近搅拌站扰民	2019/11/19 18:07:54	0	1
189950	A909204	投诉 A2 区丽发新城附近建搅拌站噪音扰民	2019-11-13 11:20:21	0	0
190108	A909240	丽发新城小区旁边建搅拌站	2019-12-21 15:11:29	0	1
190523	A00072847	A 市丽发新城违建搅拌站，彻夜施工扰民污染环境	2019/12/26 13:55:15	0	0
203393	A00053065	A 市丽发新城小区侧面建设混凝土搅拌站，粉尘和噪音污染严重	2019/11/19 14:51:53	0	2
208714	A00042015	A2 区丽发新城附近修建搅拌站，污染环境，影响生活	2020-01-02 00:00:00	0	4
213464	A909233	投诉丽发新城小区附近违建搅拌站噪音扰民	2019-12-10 12:34:21	0	0
213930	A909218	A2 区丽发新城附近违规乱建混凝土搅拌站谁来监管？	2019-12-27 23:34:32	0	0
214282	A909209	A 市丽发新城小区附近搅拌站噪音扰民和污染环境	2020-01-25 09:07:21	0	0
216824	A909214	搅拌站大量加工砂石料噪音污水影响丽发新城小区环境	2019-12-25 12:15:57	0	0
217700	A909239	丽发新城小区旁的搅拌站严重影响生活	2019-12-21 02:33:21	0	1
231136	A909204	投诉 A2 区丽发新城附近建搅拌站噪音扰	2019-12-02 11:20:21	0	0

		民			
235362	A909215	暮云街道丽发新城小区附近水泥搅拌站 非法经营何时休	2020-01-06 20:45:34	0	0
238212	A909203	丽发新城小区附近建搅拌站合理吗？	2019-12-12 10:23:11	0	0
239336	A909213	A 市 A2 区丽发新城小区遭搅拌站严重污 染	2019-12-11 11:44:11	0	0
239648	A909211	A 市 A2 区丽发新城小区附近搅拌站明目 张胆污染环境	2020-01-06 22:41:31	0	0
243692	A909201	丽发新城小区附近的搅拌站噪音严重扰 民	2019-11-15 11:23:21	0	2
244335	A909135	A 市暮云街道丽发新城社区搅拌站灰尘， 噪音污染严重	2019/12/2 12:11:23	0	0
244512	A00094706	搅拌站丽发新城小区粉尘大的孩子生活 不了	2019-12-05 20:57:50	0	1
253040	A909202	投诉 A2 区丽发新城附近建搅拌站噪音扰 民	2019-12-04 12:10:21	0	0
258242	A909220	A 市暮云街道丽发新城社区搅拌站灰尘， 噪音污染严重	2019-12-02 12:23:11	0	0
258378	A00084226	丽发新城社区附近搅拌站修建严重影响 居民生活	2019-11-23 00:00:00	0	0
268109	A909230	我要举报 A 市 A2 区丽发新城小区开发商 违规建设搅拌站	2019-12-05 18:21:32	0	0
272224	A909224	丽发新城小区噪音大粉尘大,求撤走搅拌 站	2020-01-09 19:46:10	0	1
274004	A00026895	A 市暮云街道丽发新城社区附近搅拌站 噪音污染严重	2019-12-21 10:11:09	0	0
281546	A00051470	丽发新城小区附近搅拌站粉尘大,无法呼 吸	2019-11-29 14:19:27	0	1
281943	A909216	举报 A2 区丽发新城小区附近仍存在非法 搅拌站	2019-11-15 08:56:46	0	0
283482	A909232	丽发新城小区附近搅拌站的一些问题	2019-12-07 09:21:32	0	0
287331	A909217	A2 区李丽发新城附近无资质混凝土搅拌 站为何禁而不止？	2020-01-15 15:56:21	0	0
199379	A00092242	A2 区丽发新城附近修建搅拌厂，严重污 染环境	2019/11/25 10:17:56	0	0
215563	A909231	A2 区丽发新城小区旁边的搅拌厂是否合 法经营	2019-12-06 12:21:32	0	0
215842	A909210	A2 区丽发新城小区附近太吵了	2020-01-26 19:47:11	0	0
222831	A909228	噪音、灰尘污染的 A2 区丽发新城附近环 保部门不作为	2019-12-22 10:23:11	0	0
225217	A909223	A2 区丽发新城附近修建搅拌厂严重影响	2019-11-15 09:17:36	0	0

		睡眠			
233158	A909242	丽发新城小区旁建搅拌厂严重扰民！	2019-12-05 08:46:20	0	0
234327	A909212	昏天暗地噪声不断的丽发新城小区（A2区）	2019-12-26 21:44:13	0	0
255276	A909219	再次希望领导“拯救”丽发新城小区居民	2019-12-11 00:00:00	0	0
259788	A909221	A市暮云街道丽发新城社区搅拌厂危害居民健康	2019-12-07 00:00:00	0	0
260979	A909229	反应 A 市暮云街道丽发新城小区环境污染问题	2019-12-04 14:21:32	0	0
264944	A0004260	A2 区丽发新城附近修建搅拌厂噪音、灰尘污染	2019-11-02 14:23:11	0	0
267050	A909227	噪音、灰尘污染的 A2 区丽发新城附近已扰乱居民生活	2019-11-02 10:18:00	0	0
268300	A909225	A2 区丽发新城附近修建搅拌厂噪音污染导致生活不正常	2019-11-25 10:17:58	0	0
272361	A909242	丽发新城小区旁建搅拌厂严重扰民	2019-12-04 08:46:20	0	1
273282	A909226	A2 区丽发新城附近修建搅拌厂烟尘滚滚，声音刺耳	2019-12-25 10:17:59	0	0
281348	A909219	希望领导“拯救”丽发新城小区居民	2019-11-24 00:00:00	0	0
284485	A909222	黑心搅拌厂建在 A 市丽发新城小区附近	2019-12-18 00:00:00	0	0
丽发新城断水					
193091	A00097965	A 市富绿物业丽发新城强行断业主家水	2019/6/19 23:28:27	0	242
伊景园车位					
190337	A00090519	关于伊景园滨河苑捆绑销售车位的维权投诉	2019-08-23 12:22:00	0	0
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	1	12
196264	A00095080	投诉 A 市伊景园滨河苑捆绑车位销售	2019/8/7 19:52:14	0	0
205277	A909234	伊景园滨河苑捆绑车位销售合法吗？！	2019-08-14 09:28:31	0	1
205982	A909168	坚决反对伊景园滨河苑强制捆绑销售车位	2019-08-03 10:03:10	0	2
207243	A909175	伊景园滨河苑强行捆绑车位销售给业主	2019-08-23 12:16:03	0	0
209571	A909200	伊景园滨河苑项目绑定车位出售是否合法合规	2019-08-28 19:32:11	0	0
218709	A00010669 2	A 市伊景园滨河苑捆绑销售车位	2019/8/1 22:42:21	0	1
222209	A00017171	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	2019-08-28 10:06:03	0	0
223247	A00044759	投诉 A 市伊景园滨河苑捆绑销售车位	2019/7/23 17:06:03	0	0

224767	A909176	伊景园滨河苑车位捆绑销售！广铁集团做个人吧！	2019-07-30 14:20:08	0	0
230554	A909174	投诉 A 市伊景园滨河苑捆绑车位销售	2019-08-19 10:22:44	0	0
234633	A909194	无视消费者权益的 A 市伊景园滨河苑车位捆绑销售行为	2019-08-20 12:34:20	0	0
236301	A909197	和谐社会背景下的 A 市伊景园滨河苑车位捆绑销售	2019-08-30 16:32:12	0	0
239032	A909169	请维护铁路职工权益取消伊景园滨河苑捆绑销售车位的要求	2019-09-01 10:03:10	0	1
244243	A909198	关于伊景园滨河苑捆绑销售车位的投诉	2019-08-24 18:23:12	0	0
246407	A00099597	举报广铁集团在伊景园滨河苑项目非法绑定车位出售	2019-09-01 14:20:22	0	0
251844	A909167	投诉伊景园滨河苑项目违法捆绑车位销售	2019-08-20 13:34:12	0	1
255507	A909195	违反自由买卖的 A 市伊景园滨河苑车位捆绑销售行为	2019-08-20 12:34:21	0	0
258037	A909190	投诉伊景园滨河苑捆绑销售车位问题	2019-08-23 11:46:03	0	0
260254	A909173	投诉 A 市伊景园滨河苑开发商违法捆绑销售无产权车位	2019-08-30 18:10:23	0	0
268299	A909193	惊！！A 市伊景园滨河苑商品房竟然捆绑销售车位	2019-08-21 15:32:33	0	0
276460	A909170	A 市伊景园滨河苑捆绑销售车位是否合理？	2019-08-24 17:23:11	0	0
279070	A00095080	投诉 A 市伊景园滨河苑开发商违法捆绑销售无产权车位	2019/8/31 6:33:25	0	0
283879	A00044759	A 市伊景园滨河苑项目捆绑销售车位	2019/7/18 20:27:40	0	0
285897	A909191	武广新城伊景园滨河苑违法捆绑销售车位,求解决	2019-08-01 20:06:52	0	0
286304	A909196	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	2019-08-23 10:23:23	0	0
289950	A00044759	投诉 A 市伊景园滨河苑捆绑销售车位	2019/7/7 7:28:06	0	0
辉煌国际 饭店					
192338	A00010079 2	A 市辉煌国际城居民楼下开饭店，到底是哪些部门失职？	2019/1/30 13:24:52	0	0
192652	A00010079 2	A1 区辉煌国际城二期居民楼下商铺没有配套专用烟道，非法开饭店	2019/1/13 9:41:05	0	0
214307	A00010079 2	A1 区远大一路辉煌国际城二期物业公司提供虚假场地证明	2019/1/26 13:14:51	0	0
227607	A00010079 3	A 市辉煌国际城二期商铺非法营业	2019/1/7 21:20:20	0	1

241827	A00010079 1	A1 区辉煌国际城二期商铺被开发商违法出租给别人做餐饮	2019/1/6 11:02:20	0	0
246224	A00010079 3	A1 区辉煌国际城居民楼下商铺违法开饭店，老百姓从何谈安居乐业？	2019/4/9 18:07:10	0	0
270989	A00010079 1	老百姓不满意 A 市相关部门对“辉煌国际城二期问题”的回复	2019/3/14 19:08:37	0	1
276854	A00010079 3	A 市辉煌国际城二期居民楼下违法开饭店，维权半年无果	2019/5/22 8:01:00	1	2
279702	A00010079 2	A1 区远大一路辉煌国际城二期物业提供虚假场地证明，居民楼下商铺非法取证	2019/1/7 21:35:39	0	1
286629	A00010079 3	请取缔 A 市辉煌国际城二期居民楼下的违规饭店	2019/5/22 10:05:15	0	1
289930	A00010079 2	A1 区辉煌国际城二期居民楼下商铺违法开饭店，维权近三月没有用	2019/1/27 12:18:43	0	0
魅力之城 油烟、噪声扰民					
189381	A00010981 5	A 市万科魅力之城商铺无排烟管道，小区内到处油烟味	2019/12/4 16:25:06	0	0
195095	A00039089	魅力之城小区临街门面油烟直排扰民	2019/09/05 12:29:01	0	3
236798	A00039089	A5 区劳动东路魅力之城小区油烟扰民	2019/07/28 12:49:18	0	4
246598	A00054842	A5 区劳动东路魅力之城小区临街门面烧烤夜宵摊	2019/09/25 00:31:33	0	1
268914	A0006238	A5 区劳动东路魅力之城小区底层餐馆油烟扰民	2019/09/10 06:13:27	0	0
272122	A909113	A5 区劳动东路魅力之城小区一楼的夜宵摊严重污染附近的空气，急需处理！	2019/08/01 16:20:02	0	6
284147	A909113	A5 区劳动东路魅力之城小区一楼的夜宵摊严重污染附近的空气	2019/07/21 10:29:36	0	3
287386	A909116	A 市万科魅力之城商铺无排烟管道，小区内到处油烟味	2019/08/18 14:44:00	0	0
360100	A324156	魅力之城小区临街门面油烟直排扰民	2019-09-05 12:29:01	3	0
360101	A324156	A5 区劳动东路魅力之城小区油烟扰民	2019-07-28 12:49:18	4	0
360102	A1234140	A5 区劳动东路魅力之城小区底层餐馆油烟扰民	2019-09-10 06:13:27	0	0
360103	A0012425	A5 区劳动东路魅力之城小区临街门面烧烤夜宵摊	2019-09-25 00:31:33	1	0
360104	A012417	A 市魅力之城商铺无排烟管道，小区内到处油烟味	2019-08-18 14:44:00	0	0
360107	A0283523	A5 区劳动东路魅力之城小区一楼的夜宵摊严重污染附近的空气	2019-07-21 10:29:36	3	0
360108	A0283523	A5 区劳动东路魅力之城小区一楼的夜宵摊严重污染附近的空气，急需处理！	2019-08-01 16:20:02	6	0

242792	A909115	A5 区魅力之城小区一楼被搞成商业门面，噪音扰民严重	2019/08/26 08:33:03	0	1
245136	A909117	万科魅力之城小区底层门店深夜经营，各种噪音扰民	2019/09/04 21:00:18	0	0
246362	A909114	A 市魅力之城小区底层商铺营业到凌晨，各种噪音好痛苦	2019/08/26 01:50:38	0	0
360105	A120356	A5 区魅力之城小区一楼被搞成商业门面，噪音扰民严重	2019-08-26 08:33:03	1	0
360106	A235367	A 市魅力之城小区底层商铺营业到凌晨，各种噪音好痛苦	2019-08-26 01:50:38	0	0
360109	A0080252	万科魅力之城小区底层门店深夜经营，各种噪音扰民	2019-09-04 21:00:18	0	0
温斯顿英语					
188467	A00050188	投诉 A 市温斯顿英语培训学校拖延退费！	2019/3/28 19:57:19	0	1
190969	A0006925	A 市温斯顿英语梅溪新天地校区何时退费	2019/11/8 16:51:55	0	2
223497	A00072941	A 市民办培训机构乱象丛生，温斯顿英语培训强设霸王条款	2019/6/11 0:43:37	0	1
244103	A00082891	A 市温斯顿英语世纪金源校区还没退费给我	2019/9/27 17:28:01	0	0
254068	A00050454	投诉 A 市温斯顿英语培训机构拖延退费	2019/9/18 18:19:41	0	0
260766	A00042274	A 市温斯顿英语（梅溪新天地校区）店大服务差，退款难	2019/7/11 10:24:11	0	0
282638	A0005856	A 市温斯顿英语培训机构退费怎么这么难？	2019/11/29 9:42:59	0	0
青青家园					
188260	A00053484	A3 区青青家园小区乐果果零食炒货公共通道摆放空调扰民	2019/5/31 17:06:13	0	0
206379	A000103490	A3 区青青家园小区违规住改商存隐患	2019/6/6 10:50:52	0	0
212842	A00018295	A3 区青青家园小区公共通道破墙架烟道，人为制造消防火灾隐患	2019/5/31 10:04:12	0	0
236867	A00032346	A3 区天顶街道青山新村社区青青家园小吃店违规住改商	2019/8/5 13:56:42	0	0
250928	A00032346	反映 A3 区青青家园住改商审批的问题	2019/5/24 12:24:13	0	0
261103	A00032346	A3 区青青家园小区公共通道被架了烟道	2019/6/5 15:06:33	1	0
270086	A00032346	A3 区青青家园公共消防生命通道架设油烟管道，持续半年扰民	2019/5/8 9:45:25	0	0
284066	A00032346	A3 区天顶街道青青家园违规住改商问题到底谁能解决	2019/7/24 16:17:59	0	0
凉塘路					

210366	A00035629	A7 县凉塘路的旧城改造要拖到何时才能动工？	2019/7/31 16:42:51	0	1
217680	A00074409	A7 县东六到东八路之间的凉塘路还没拉通	2019/10/11 18:47:15	0	14
228139	A00098215	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？	2019/9/9 8:20:47	0	1
234952	A00035628	A7 县星沙四区凉塘路的旧城改造要拖到何时？	2019/7/31 11:34:33	0	0
245969	A00078335	A7 县星沙凉塘路东七线到东十线段今年会启动建设吗	2019/3/18 12:46:39	0	24
250512	A00035626	A7 县星沙凉塘路旧城改造究竟要拖到何年何月才能开始？	2019/9/2 14:32:27	0	6
259574	A00072486	A7 县星沙街道凉塘路的旧城改造什么时候会启动？	2019/7/23 7:39:44	0	1
260149	A00035631	A7 县星沙街道凉塘路旧城改造什么时候可以进行	2019/6/10 10:24:59	0	5
268757	A00072477	A7 县星沙四区凉塘路旧城改造要待何时	2019/7/4 14:10:30	1	4
274285	A00035628	A7 县星沙四区凉塘路旧城改造要拖到何年何月才能动工	2019/8/2 10:05:02	0	6
284120	A00035630	A7 县星沙镇四区凉塘路改造何时可以开始？	2019/2/14 10:07:59	0	6
车贷					
194343	A000106161	承办 A 市 58 车贷案警官应跟进关注留言	2019/3/1 22:12:30	0	733
217032	A00056543	严惩 A 市 58 车贷特大集资诈骗案保护伞	2019/2/25 9:58:37	0	790
218132	A000106090	再次请求过问 A 市 58 车贷案件进展情况	2019/1/29 19:15:49	0	0
220711	A00031682	请书记关注 A 市 A4 区 58 车贷案	2019/2/21 18:45:14	0	821
223787	A00034861	西地省 58 车贷案件创造全国典型诈骗案，立案至今无公告	2019/1/11 21:12:34	0	0
226265	A000106448	恳请 A 市经侦公正办理 58 车贷案件，还我们受害人一个公道	2019/5/28 15:08:51	0	3
234320	A000106592	不要让 A 市因为 58 车贷案件而臭名远扬	2019/7/8 17:16:57	0	0
240554	A00029163	A 市 58 车贷老板跑路美国，经侦拖延办案	2019/2/10 20:58:40	0	6
254532	A000106062	A 市 58 车贷恶性退出立案近半年没有发过一次案情通报	2019/1/14 22:08:20	0	3
264119	A00084445	58 车贷立案五个月过去，A4 区公安分局未公布过任何案情	2019/1/19 9:47:23	0	0

268251	A000106090	西地省 58 车贷立案近半年毫无进展，单位回复让人心寒	2019/2/2 15:03:05	0	25
272413	A000106062	西地省 A 市 58 车贷恶性退出，A4 区立案已近半年毫无进展	2019/1/14 20:23:57	0	2
272858	A00061787	A 市 58 车贷恶性退出案件为什么不发布案情进展通报？	2019/1/16 23:21:21	0	0
绿地海外滩 噪声扰民					
191951	A00041448	A4 区绿地海外滩小区距渝长厦高铁太近了	2019/8/23 14:21:38	0	1
202575	A00092007	咨询 A 市绿地海外滩二期与长赣高铁问题	2019/9/4 18:32:42	0	17
216316	A00097196	A4 区绿地海外滩二期业主被噪音扰得快烦死了	2019/9/6 10:16:27	0	2
243551	A00041448	A 市至赣州高铁对绿地海外滩二期小区影响太大了	2019/9/1 10:18:48	0	1
258708	A00092008	A 市绿地海外滩二期没交房就有楼房沉降开裂的问题	2019/1/30 10:00:04	0	2
263672	A00041448	A4 区绿地海外滩小区距长赣高铁最近只有 30 米不到，合理吗？	2019/9/5 13:06:55	0	669
266931	A00092007	按照当前的高铁规划，A 市绿地海外滩小区会饱受噪音困扰	2019/9/6 18:36:16	0	1
京港澳					
253369	A00074795	穿 A 市城而过的京港澳高速（长楚高速）什么时候可以外迁至远郊？	2019/11/18 15:35:11	0	29
284571	A00074795	建议西地省尽快外迁京港澳高速城区段至远郊	2019/1/10 15:01:26	0	80
和包支付					
200667	A00079480	请问 A 市为什么要把和包支付作为任务而不让市场正当竞争？	2019/1/16 17:01:25	0	78
275782	A00074795	建议和包支付整合楚行一卡通和 A 市地铁 app	2019/6/15 13:19:46	0	0
278393	A00015060	“和包支付”成了唯一方式支付党费的手段，合理吗？	2019/9/10 21:31:32	0	2
285619	A00042436	我们基层工作者推广和包支付已经毫无尊严	2019/5/21 17:51:54	0	6
月亮岛 高压					
218442	A00099016	A6 区月亮岛路架设高压电线环评造假，谁为民众做主	2019/4/8 21:19:40	0	22
231773	A00010141	反对 A6 区月亮岛路架设高压电线，强烈要求重启环境评估	2019/4/12 14:59:14	0	1

234885	A00060375	A6区月亮岛路11万伏高压线没用地埋方式铺设	2019/4/5 13:01:17	0	2
254865	A00099016	关于 A6 区月亮岛路沿线架设 110kv 高压电线杆的投诉	2019/4/3 17:36:58	0	5
262052	A00072424	关于 A6 区月亮岛路沿线架设 110kv 高压电线杆的投诉	2019/3/26 14:33:47	0	78
268250	A00072424	关于 A6 区月亮岛路沿线架设 110KV 高压电线杆的投诉	2019/3/26 10:17:31	0	10
272089	A00061602	关于 A6 区月亮岛路 110kv 高压线的建议	2019/4/9 17:10:01	2	55
房云时代 房屋质量					
220857	A00020702	A 市长房云时代 13 栋房屋有质量问题	2019/9/11 22:30:46	0	1
247713	A00045805	长房云时代二期房子出现很多严重的质量问题	2019/11/16 13:46:34	0	0
281898	A00096623	A 市长房云时代多栋房子现裂缝,质量堪忧	2019/2/25 15:17:38	5	55
房云时代不开学					
221996	A00080850	A 市博才长房云时代小学迟迟不开学,上万业主心急如焚	2019/1/4 11:36:26	0	28
三一大道 改造					
226723	A00040222	A 市三一大道全线快速化改造何时启动?	2019/9/15 15:31:19	0	66
252940	A0005950	A 市 A9 市河路至三一大道段一直没有拉通	2019/5/21 16:45:15	0	0
278028	A00053372	A 市德雅路(三一大道口---车站路口)什么时候改造?	2019/2/27 10:55:00	0	8
郝家坪小学					
205960	A00096376	A3 区郝家坪小学何时扩建?	2019/3/29 9:53:56	0	13
209742	A00012969	A3 区郝家坪小学什么时候能改扩建?	2019/3/24 21:07:12	0	41
233743	A00012969	A3 区观沙岭郝家坪小学何时改扩建?	2019/3/24 20:11:24	0	3
A7 华苑 地下车位					
211978	A0009303	询问 A7 县圣力华苑车位的问题	2019/5/7 14:16:11	0	4
253492	A0009303	请问 A7 县星沙圣力华苑出售地下车位是否合法?	2019/3/11 14:52:12	0	20
264693	A0009303	询问 A7 县星沙圣力华苑出售地下车位是否合法?	2019/5/6 16:38:01	0	4
A7 华苑 两证合一					
190346	A00058734	请求解决 A7 县圣力华苑小区房屋两证合一的问题	2019/7/10 22:46:03	0	1
193680	A00092446	A7 县圣力华苑房屋两证合一找什么部门可以解决	2019/7/8 15:17:54	0	0

251228	A00092446	请问 A7 县圣力华苑一房两证何时可以合并	2019/8/1 16:30:53	0	0
259350	A00092446	A7 县泉塘街道圣力华苑小区业主想办理“两证合一”该走什么程序？	2019/7/12 10:57:16	0	1
天池物业问题					
191493	A00018927	A 市南山十里天池虚假宣传及违规收取物业费	2019/2/15 16:52:20	0	6
232468	A0005801	A 市南山十里天池小区交房在即，但是小区正门却没有路	2019/2/15 15:55:29	0	5
238219	A00050125	A 市南山十里天池玉佩路出行不便，开发商在踢皮球	2019/2/27 10:12:23	0	1
245354	A00018298	A3 区南山十里天池（南麓苑）楼盘交付存在问题	2019/3/5 8:19:16	0	3
247851	A00018255	投诉 A 市南山十里天池开发商	2019/2/15 15:56:31	2	18
251901	A00021433	A 市南山十里天池存在严重质量问题，开发商物业不闻不问	2019/3/28 17:32:04	0	1
270683	A00021433	A 市南山十里天池门前有土堆，业主出行难	2019/2/13 10:22:45	0	5
万润滨江 装修					
202909	A00052058	A 市无良万润滨江天著牟取精装暴利	2019/1/3 19:37:36	0	28
217403	A00010385 7	A 市万润滨江天著忽悠业主延迟网签！	2019/2/26 18:09:45	0	0
241657	A00010385 7	请问 A 市万润滨江天著的全装修核价什么时候出来？	2019/3/6 21:07:31	0	0
250441	A00034164	A 市万润滨江天著精装修明细何时公示	2019/9/16 11:17:21	0	0
A7 县社塘路 白改黑					
192366	A00010350 6	A7 县社塘路白改黑及铁建一期社塘路人行道星沙供水公司井盖高出人行道	2019/4/30 17:17:00	0	2
223913	A00074795	A7 县金贸路什么时候拓宽并白改黑？	2019/10/9 11:12:38	2	5
225764	A00074795	A7 县金茂路什么时候拓改白改黑？	2019/9/27 17:11:19	0	16
274577	A00080323	给 A7 县社塘路白改黑的建议	2019/5/27 11:40:42	1	0
泉星公园 优化					
193337	A00080343	请问 A7 县泉星公园何时开工，工期多长？	2019/10/16 13:03:13	0	12
226408	A00080342	A 市经开区泉星公园项目规划需优化	2019/8/9 16:47:36	0	4
238692	A00080342	建议 A 市经开区泉星公园项目规划进一步优化	2019/8/12 13:15:05	0	16
245801	A00036276	A7 县泉星公园建设进度和力都大厦长期占用小塘路堆放材料影响百姓出行	2019/4/30 16:47:03	0	4

253331	A00080343	建议公示 A7 县经开区泉星公园项目景观效果图	2019/9/29 13:25:08	0	2
273741	A00080342	建议进一步优化 A 市经开区泉星公园项目规划	2019/8/13 16:02:24	0	0
278099	A00080323	A7 县泉星公园建设进度缓慢, 周边几万居民无处休闲散步	2019/6/24 22:22:11	0	0
278281	A00032698	建议 A 市经开区泉星公园项目规划进一步优化	2019/8/22 13:23:46	0	0
278545	A00036841	给 A 市经开区泉星公园项目规划进一步优化的建议	2019/8/26 13:00:06	0	13
289574	A00080342	对 A 市经开区泉星公园项目规划再进一步优化的建议	2019/8/16 13:33:59	0	0
人才					
189180	A000106515	A 市人才购房补贴申请是否与单位注册地有关?	2019/6/18 9:51:36	0	3
199435	A00024070	咨询 A 市人才购房补助发放问题	2019/7/2 13:11:31	0	0
200353	A0009832	社保补缴后, A 市人才新政补贴会补发吗	2019/7/30 11:56:47	0	0
203760	A000111954	咨询 A 市人才购房补贴政策	2019/6/25 15:43:23	0	0
205771	A00020115	夫妻共同买的房为何申请 A 市人才购房补贴不通过?	2019/5/29 17:12:29	0	0
206983	A00049301	A 市人才新政补贴最近两个月的怎么还没发?	2019/5/28 16:18:02	0	8
212128	A000110735	在 A 市买公寓能享受人才新政购房补贴吗?	2019/4/24 15:07:30	0	0
215591	A00073581	咨询 A 市人才新政落户等相关问题	2019/3/18 17:35:47	0	0
219026	A00083974	在 A 市购买二手房能享受人才新政购房补贴吗?	2019/9/27 9:17:14	0	0
224042	A00014225	咨询 A 市人才购房及购房补贴实施办法等相关问题	2019/1/16 11:58:48	0	0
225657	A00051791	关于 A 市人才购房补贴的疑问	2019/2/25 14:43:15	2	6
228559	A00082143	A 市人才新政落户后屡次申请购房补贴不成功	2019/2/15 10:09:13	0	2
229963	A00010465	咨询 A 市人才新政住房补贴申请的问题	2019/4/24 23:19:34	0	0
244951	A00026039	反映 A 市人才租房购房补贴问题	2019/7/7 19:12:32	0	0
247736	A00085481	A 市人才新政的补贴已经快两个月没发了	2019/5/28 11:54:08	2	7
250956	A00028389	A 市人才认证租房补贴通过名单在哪里查看?	2019/5/30 17:57:07	0	0
253740	A000106679	请问 A 市人才新政落户人员的档案如何管理?	2019/4/26 9:41:29	0	0

262512	A00043323	A 市 26 路公汽站台标有人才市场站，车却从来不停	2019/11/21 21:08:33	0	0
265577	A0004787	咨询 A 市人才购房补贴通知问题	2019/12/2 11:57:49	0	1
268023	A00030760	反映 A 市人才补贴问题	2019/10/29 19:32:19	0	0
268952	A00090921	咨询 A 市新引进高级技师人才奖励政策相关问题	2019/7/31 12:02:35	0	0
270015	A00020115	为何我的 A 市人才购房补贴申请不通过？	2019/5/29 16:51:26	0	0
272214	A00048770	A 市人才服务中心人事代理费补缴为何不能网上缴费	2019/10/16 12:07:49	0	0
282104	A00090921	关于高级技师申报 A 市人才新政购房补贴的相关问题咨询	2019/7/29 10:42:38	0	1
282248	A000106091	咨询 A 市人才购房补贴事宜	2019/1/30 15:42:19	0	0
283494	A00085185	关于《A 市人才购房及购房补贴实施办法（试行）》相关问题的咨询	2019/7/30 19:06:40	0	0
289408	A0012413	在 A 市人才 app 上申请购房补贴为什么通不过	2018-11-15 16:07:12	0	0
204245	A00024579	关于 A 市高级技师购房补贴的疑问	2019/3/1 10:58:43	0	14
城际空间站					
204531	A00082895	地绿 A 市城际空间站的建筑质量太差了	2019/9/24 8:02:57	0	1
208299	A00082895	A 市绿地城际空间站建筑质量堪忧	2019/9/23 11:20:33	0	0
228541	A00094264	举报 A 市地绿城际空间站的违法违规行为	2019/10/30 16:47:30	0	3
245738	A00090620	投诉 A 市绿地城际空间站虚假宣传	2019/10/30 13:10:18	0	1
271948	A00094264	A 市地绿城际空间站虚假宣传学位，房屋质量存在安全隐患	2019/10/30 15:04:23	0	0
282511	A00072388	A 市绿地城际空间站建筑质量太差了	2019/11/9 8:44:29	0	13

附录 2:

（第一问代码）

```
# -*- encoding:utf-8 -*-
import random
import jieba
import pandas as pd
import numpy as np
```

```

import chardet
stopwords = pd.read_csv(r"D:\\LSTM\\stopwords.txt", index_col=False, quoting=3, sep="\t",
names=["stopword"], encoding="UTF-8")
stopwords = stopwords["stopword"].values
jianshe_df = pd.read_csv(r"D:\\LSTM\\jianshe.txt", encoding="utf-8", sep=",")
huanjing_df = pd.read_csv(r"D:\\LSTM\\huanjing.txt", encoding="utf-8", sep=",")
jiaotong_df = pd.read_csv(r"D:\\LSTM\\jiaotong.txt", encoding="utf-8", sep=",")
jiaoyu_df = pd.read_csv(r"D:\\LSTM\\jiaoyu.txt", encoding="utf-8", sep=",")
laodong_df = pd.read_csv(r"D:\\LSTM\\laodong.txt", encoding="utf-8", sep=",")
shangmao_df = pd.read_csv(r"D:\\LSTM\\shangmao.txt", encoding="utf-8", sep=",")
weisheng_df = pd.read_csv(r"D:\\LSTM\\weisheng.txt", encoding="utf-8", sep=",")
jianshe_df.dropna(inplace=True)
huanjing_df.dropna(inplace=True)
jiaotong_df.dropna(inplace=True)
jiaoyu_df.dropna(inplace=True)
laodong_df.dropna(inplace=True)
shangmao_df.dropna(inplace=True)
weisheng_df.dropna(inplace=True)
jianshe = jianshe_df.segment.values.tolist()
huanjing = huanjing_df.segment.values.tolist()
jiaotong = jiaotong_df.segment.values.tolist()
jiaoyu = jiaoyu_df.segment.values.tolist()
laodong = laodong_df.segment.values.tolist()
shangmao = shangmao_df.segment.values.tolist()
weisheng = weisheng_df.segment.values.tolist()
def preprocess_text(content_lines, sentences, category):

    for line in content_lines:
        try:
            segs = jieba.lcut(line)
            segs = [v for v in segs if not str(v).isdigit()] # 除去数字
            segs = list(filter(lambda x: x.strip(), segs)) # 除去左右空格
            segs = list(filter(lambda x: len(x) > 1, segs)) # 除去长度为 1 的字符
            segs = list(filter(lambda x: x not in stopwords, segs)) # 除去停用词
            sentences.append((" ".join(segs), category)) # 打标签
        except Exception:
            print(line)
            continue

# 调用函数，生成训练数据
sentences = []
preprocess_text(jianshe, sentences, 0)
preprocess_text(huanjing, sentences, 1)
preprocess_text(jiaotong, sentences, 2)

```

```

preprocess_text(jiaoyu , sentences, 3)
preprocess_text(laodong, sentences, 4)
preprocess_text(shangmao, sentences, 5)
preprocess_text(weisheng, sentences, 6)
random.shuffle(sentences)
print ("选取随机 20 条数据检测")
for sentence in sentences[:20]:
    print (sentence[0])
    if(sentence[1]==0):
        print ("城乡建设")
    if (sentence[1] == 1):
        print ("环境保护")
    if (sentence[1] == 2):
        print ("交通运输")
    if (sentence[1] == 3):
        print ("教育文体")
    if (sentence[1] == 4):
        print ("劳动和社会保障")
    if (sentence[1] == 5):
        print ("商贸旅游")
    if (sentence[1] == 6):
        print ("卫生计生")

# 所有特征和对应标签
all_texts = [sentence[0] for sentence in sentences]
all_labels = [sentence[1] for sentence in sentences]

from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.utils import to_categorical
from keras.layers import Dense, Input, Flatten, Dropout
from keras.layers import LSTM, Embedding, GRU
from keras.models import Sequential
from keras import regularizers

MAX_SEQUENCE_LENGTH = 50
EMBEDDING_DIM = 200
VALIDATION_SPLIT = 0.15 # 验证集
TEST_SPLIT = 0.2 # 测试集

tokenizer = Tokenizer()
tokenizer.fit_on_texts(all_texts)
sequences = tokenizer.texts_to_sequences(all_texts)
word_index = tokenizer.word_index

```

```

print("Found %s unique tokens." % len(word_index))

data = pad_sequences(sequences, maxlen=MAX_SEQUENCE_LENGTH)
labels = to_categorical(np.asarray(all_labels))
print("data shape:", data.shape)
print("labels shape:", labels.shape)

p1 = int(len(data) * (1 - VALIDATION_SPLIT - TEST_SPLIT))
p2 = int(len(data) * (1 - TEST_SPLIT))

# 训练集
x_train = data[:p1]
y_train = labels[:p1]

# 验证集
x_val = data[p1:p2]
y_val = labels[p1:p2]

# 测试集
x_test = data[p2:]
y_test = labels[p2:]

model = Sequential()
model.add(Embedding(len(word_index) + 1, EMBEDDING_DIM,
input_length=MAX_SEQUENCE_LENGTH))
model.add(LSTM(200, recurrent_dropout=0.2))
model.add(Dropout(0.2))
model.add(Dense(64, activation="relu", kernel_regularizer=regularizers.l2(0.001)))
model.add(Dropout(0.3))
model.add(Dense(labels.shape[1], activation="sigmoid"))
model.summary()

model.compile(loss="categorical_crossentropy", optimizer="rmsprop", metrics=["acc"])
print(model.metrics_names)

model.fit(x_train, y_train, validation_data=(x_val, y_val), epochs=15, batch_size=128)
model.save("lstm.h5")

# 模型评估
print(model.evaluate(x_test, y_test))

#计算 F1-score
import seaborn as sns
from sklearn.metrics import accuracy_score, confusion_matrix
import matplotlib.pyplot as plt

```

```

y_pred = model.predict(x_test)
y_pred = y_pred.argmax(axis=1)
y_test = y_test.argmax(axis=1)

from sklearn.metrics import classification_report

print('accuracy %s' % accuracy_score(y_pred, y_test))
print(classification_report(y_test, y_pred))

```

运行结果：

##每例的第一行为关键词，第二行为其对应的分类。

```

选取随机 20 条数据检测
请问 县里 新农 合到 医院 报销
劳动和社会保障
K2 汽车 年检费
商贸旅游
港澳 高速 不宜 变为 城市道路
交通运输
城管 执法局 不顾 住户 行人 生命安全
城乡建设
I1 迎风 桥镇 生活 环境
环境保护
A4 代课 教师工资 学校 私自 克扣
教育文体
住房问题 咨询
城乡建设
请问 外省 公务员 工作 四年 辞职 后能 参加 西地 省考
劳动和社会保障
I2 区全 丰村 村民 铝壳 搬走
环境保护
办理 生育 独生子女证
劳动和社会保障
L8 县县 电力公司 更换 电表
城乡建设
请求 教育局 领导 孩子 学习 机会
教育文体
A5 教育
教育文体
我省 退休职工 养老金 调整 政策 出台
劳动和社会保障
投诉 A6 区中岭 小学 班主任 体罚 学生

```


教育文体
 请问 西地省 养老金 增资 往年
 劳动和社会保障
 乡镇 医院 生财有道 管管
 卫生计生
 建议 M5 公共交通 发展 扶持
 城乡建设
 H2 区中湖 乡镇 快递 不上 门派 额外 收取 费用
 交通运输
 L3 自建房 申请 临时 用电 开户费 何在
 城乡建设

Found 13400 unique tokens.
 data shape: (9210, 50)
 labels shape: (9210, 7)

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 50, 200)	2680200
lstm_1 (LSTM)	(None, 200)	320800
dropout_1 (Dropout)	(None, 200)	0
dense_1 (Dense)	(None, 64)	12864
dropout_2 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 7)	455

Total params: 3,014,319
 Trainable params: 3,014,319
 Non-trainable params: 0

['loss', 'acc']

Train on 5986 samples, validate on 1382 samples
 Epoch 1/15
 5986/5986 [=====] - 20s 3ms/step - loss: 1.9167 -
 acc: 0.2519 - val_loss: 1.7575 - val_acc: 0.3271

Epoch 2/15
5986/5986 [=====] - 18s 3ms/step - loss: 1.4056 -
acc: 0.5184 - val_loss: 1.1496 - val_acc: 0.6281

Epoch 3/15
5986/5986 [=====] - 18s 3ms/step - loss: 0.8356 -
acc: 0.7583 - val_loss: 0.8045 - val_acc: 0.7945

Epoch 4/15
5986/5986 [=====] - 18s 3ms/step - loss: 0.4905 -
acc: 0.8846 - val_loss: 0.7013 - val_acc: 0.8162

Epoch 5/15
5986/5986 [=====] - 18s 3ms/step - loss: 0.3092 -
acc: 0.9357 - val_loss: 1.0447 - val_acc: 0.7576

Epoch 6/15
5986/5986 [=====] - 18s 3ms/step - loss: 0.2422 -
acc: 0.9561 - val_loss: 0.7477 - val_acc: 0.8285

Epoch 7/15
5986/5986 [=====] - 16s 3ms/step - loss: 0.1692 -
acc: 0.9736 - val_loss: 0.8031 - val_acc: 0.8314

Epoch 8/15
5986/5986 [=====] - 16s 3ms/step - loss: 0.1338 -
acc: 0.9828 - val_loss: 0.8523 - val_acc: 0.8271

Epoch 9/15
5986/5986 [=====] - 17s 3ms/step - loss: 0.1072 -
acc: 0.9881 - val_loss: 0.9527 - val_acc: 0.8205

Epoch 10/15
5986/5986 [=====] - 18s 3ms/step - loss: 0.0828 -
acc: 0.9915 - val_loss: 1.0136 - val_acc: 0.8249

Epoch 11/15
5986/5986 [=====] - 17s 3ms/step - loss: 0.0674 -
acc: 0.9933 - val_loss: 1.0314 - val_acc: 0.8256

Epoch 12/15
5986/5986 [=====] - 17s 3ms/step - loss: 0.0565 -
acc: 0.9930 - val_loss: 1.0226 - val_acc: 0.8198

Epoch 13/15
5986/5986 [=====] - 16s 3ms/step - loss: 0.0484 -
acc: 0.9948 - val_loss: 1.0624 - val_acc: 0.8198

Epoch 14/15
5986/5986 [=====] - 17s 3ms/step - loss: 0.0417 -
acc: 0.9957 - val_loss: 1.0600 - val_acc: 0.8220

Epoch 15/15
5986/5986 [=====] - 18s 3ms/step - loss: 0.0330 -
acc: 0.9973 - val_loss: 1.1451 - val_acc: 0.8184

[0.9945666655396534, 0.8219326734542847]

accuracy	0.8219326818675353				
	precision	recall	f1-score	support	
0	0.81	0.77	0.79	429	
1	0.84	0.81	0.83	187	
2	0.74	0.73	0.74	112	
3	0.88	0.86	0.87	325	
4	0.86	0.87	0.87	374	
5	0.72	0.85	0.78	227	
6	0.84	0.83	0.83	188	
accuracy			0.82	1842	
macro avg	0.81	0.82	0.81	1842	
weighted avg	0.82	0.82	0.82	1842	

附录 3:

第二问文本相似度代码

```
import os
import numpy as np
import docx
import jieba
import re
import xlwt
from gensim import corpora, models, similarities
path = "D:\\泰迪杯\\C\\数据\\1"
files= os.listdir(path)
s=[]
doc_test="
```

```

for file in files:
    f = docx.Document(path+"/"+file)
    text=""
    for para in f.paragraphs:
        text+=para.text
    text_final = re.sub("[\s+\.!\\V_,$%^*(+\\'\\')]+|[+———()?【】“”! , 。 ? 、 : ~@#
    ¥%.....&* ( ) ; ]+", "",text)
    s.append(text_final)
test=docx.Document('D:\\泰迪杯\\C\\数据\\第二问.docx')
for test_para in test.paragraphs:
    doc_test+=test_para.text
doc_test_final=re.sub("[\s+\.!\\V_,$%^*(+\\'\\')]+|[+———()?【】“”! , 。 ? 、 : ~@#
    ¥%.....&* ( ) ; ]+", "",doc_test)
all_doc_list=[]
for doc in s:
    doc_list=[word for word in jieba.cut(doc)]
    all_doc_list.append(doc_list)
doc_test_list = [word for word in jieba.cut(doc_test_final)]

dictionary = corpora.Dictionary(all_doc_list)
corpus = [dictionary.doc2bow(doc) for doc in all_doc_list]
doc_test_vec = dictionary.doc2bow(doc_test_list)
tfidf = models.TfidfModel(corpus)
tfidf[doc_test_vec]
index = similarities.SparseMatrixSimilarity(tfidf[corpus],
num_features=len(dictionary.keys()))
sim = index[tfidf[doc_test_vec]]
#sorted(enumerate(sim), key=lambda item: -item[1])
xls=xlwt.Workbook()
sht1=xls.add_sheet('sheet1')
file_studentname=[]
for studentname in files:
    studentnames=re.sub(".docx","",studentname)
    file_studentname.append(studentnames)
for i,filename,_similarities in zip(np.arange(len(files)),file_studentname,sim):
    sht1.write(int(i),0,filename)
    sht1.write(int(i),1,str(_similarities))
xls.save('D:\\泰迪杯\\C\\数据\\相似度.xls')

```