

# “智慧政务”中的文本挖掘应用

## 摘要

随着时代的发展，建立基于自然语言处理技术的智慧政务系统对提升政府的管理水平和施政效率有极大的推动作用。本文运用了多种自然语言处理技术和文本挖掘方法解决问政平台中的群众留言分类问题、热点问题挖掘任务和答复意见评价问题。

在群众留言分类部分，为了选择恰当的文本分类方法，我们首先分析了问政平台留言的特点并撰写了文本分类模型的研究综述。在此基础上我们选择建立**融合注意力机制的卷积神经网络模型 (ACNN)**来解决问政平台中的留言分类问题，该模型可以关注更为重要的文本特征，从而提高文本分类的效果。在该模型中，我们先利用正则表达式、命名实体识别、Jieba 分词、去停用词等技术对文本数据进行预处理。接着，为了使文本数据能够被 ACNN 模型所使用，我们利用 **Word2vec 模型**将中文词语转化为词向量。然后，为了优化 ACNN 模型的文本分类效果，我们构建了 3 层卷积层来提高模型提取文本特征的能力，并设置了 **Dropout 操作与 Batch normalization 操作**以避免训练时出现过拟合问题同时加速模型训练。最后，我们使用 F-Score 对模型分类效果进行评价，并通过多次实验确定了模型的最优参数组合。结果表明：相对于传统的卷积神经网络模型，我们所建立的模型可以将  $F_1$  值提高 0.1576，将分类精度提升 11.84%；在利用改进的词向量训练模型后，ACNN 模型仅训练 3 分钟验证集分类准确率即可达到 90%左右，且训练结束后  $F_1$  值为 0.8940。

在热点问题挖掘部分，我们首先同样对文本数据进行新词识别、Jieba 分词、去停用词等预处理工作，同时利用**正则表达式**依据留言问题所属行政单位对留言进行划分。该操作不仅能提高热点问题聚类精度，还能够大大降低后续文本相似度的计算量。接着，我们利用词语的词性信息和 **TF-IDF 值**筛选关键词，并在利用 Word2vec 训练词向量与计算**词语余弦相似度**的基础上定义了文本相似度的计算方法。在该方法中还利用词语的 TF-IDF 值对文本相似度计算进行**加权处理**，充分将文本表层信息与文本语义信息相结合，使得文本相似度计算结果更加合理准确。在得到文本相似度矩阵后，我们利用**谱聚类**完成热点问题的聚类任务。最后我们选取问题的被关注度、涉及人群的广度和被认同度作为热度评价指标，并综合层次分析法和 TOPSIS 评价法建立 **AHP-TOPSIS 评价模型**选出热度前五的问题，具体结果请见附件。

在答复意见评价部分，我们从两个不同的出发点去评价答复意见。一是从答复意见自身质量出发，选取**相关性、完整性、可解释性、可读性和时效性**指标来评价答复意见。其中相关性表示答复意见围绕相应留言主题展开回复的紧密程度；完整性象征答复意见结构的规范程度；时效性表示答复的及时程度；可读性表示答复意见易于理解的程度；可解释性表示答复意见是否引用相关文件进

行解释和是否对留言涉及事件进行调查后做出回复。二是从群众的体验感出发，通过 NLPIR 大数据语义智能分析平台对每条留言进行**情感分析**得到情感得分。然后运用层次分析法确定评价答复自身质量的各个指标的权重并结合留言情感得分给出综合评价方案。最后，对于评价答复自身质量的各个指标，我们给出了其年平均值的变化趋势并加以简单的分析，以供相关部门参考。

此外，我们还剖析了各个模型的优缺点。

**关键字：**自然语言处理 ACNN Word2vec 文本相似度 TF-IDF AHP-TOPSIS

## Abstract

With the development of the times, the establishment of a smart government system based on natural language processing technology will greatly promote the management level and governance efficiency of the government. This article uses a variety of natural language processing techniques and text mining methods to solve the problem of mass message classification, hot topic mining tasks, and the evaluation of reply comment in the political inquiry platform.

In the part of mass message classification, in order to choose the appropriate text classification method, we first analyzed the characteristics of the online platform for political participation and wrote a research review of the text classification model. On this basis, we choose to build a **Convolutional Neural Network Model (ACNN)** that integrates the attention mechanism to solve the message classification problem in the online platform for political participation. This model can focus on more important text features, thereby improving the effect of text classification. In this model, we first use regular expressions, named entity recognition, Jieba word segmentation, stop word removal and other technologies to preprocess text data. Next, in order to make the text data be used by the ACNN model, we use the **Word2vec Model** to convert Chinese words into word vectors. Then, in order to optimize the text classification effect of the ACNN model, we constructed a 3-layer convolutional layer to improve the model's ability to extract text features, and set up **Dropout operations** and **Batch normalization** operations to avoid overfitting problems during training and speed up model training. Finally, we use F-Score to evaluate the model classification effect, and determine the optimal parameter combination of the model through multiple experiments. The results show that, compared with the traditional convolutional neural network model, the model we built can increase the F<sub>1</sub> value by 0.1576 and improve the classification accuracy by 11.84%; after training the model with the improved word vector, the ACNN model only trains the validation set for 3 minutes. The classification accuracy can reach about 90%, and the F<sub>1</sub> value after training is 0.8940.

In the mining of hot issues, we first perform new word recognition, Jieba word segmentation, and stop word preprocessing on the text data. At the same time, we use **Regular Expressions** to divide the messages according to the administrative unit to which the message question belongs. This operation can not only improve the clustering accuracy of hotspot problems, but also greatly reduce the calculation amount of subsequent text similarity. Next, we use the part-of-speech information of the words and **TF-IDF** value to filter the keywords, and define the calculation method of text similarity based on the use of Word2vec to

train the word vector and **calculate the cosine similarity** of the words. In this method, the TF-IDF value of words is also used to **weight** the text similarity calculation, which fully combines the text surface information and the text semantic information, so that the text similarity calculation result is more reasonable and accurate. After obtaining the text similarity matrix, we use **spectral clustering** to complete the clustering task of hot issues. Finally, we selected the attention degree, the breadth of the people involved and the degree of recognition as the heat evaluation indicators, and integrated the AHP and TOPSIS evaluation methods to establish the **AHP-TOPSIS** evaluation model to select the top five heat problems. For the specific results, please see the attachment .

In the comment evaluation section, we evaluate the comments from two different starting points. First, starting from the quality of the response comments, we select **relevance, completeness, interpretability, readability and timeliness** indicators to evaluate the response comments. Relevance means how closely the reply comments are centered on the subject of the corresponding message; completeness symbolizes the degree of standardization of the reply opinion structure; timeliness means the timeliness of the reply; readability means that the reply opinion is easy to understand; and interpretability means whether the opinion refers to the relevant documents for explanation and whether to respond after investigation. The second is from the masses' experience, through the NLPiR big data semantic intelligence analysis platform for **sentiment analysis** of each message to get sentiment score. Then use the analytic hierarchy process to determine the weight of each indicator to evaluate the quality of the response and combine with the sentiment score of the message to give a comprehensive evaluation plan. Finally, for each indicator that evaluates the quality of the response itself, we give a change in its annual average and make a simple analysis for reference by relevant departments.

In addition, we also analyzed the advantages and disadvantages of each model.

**Key words:** Natural language processing   ACNN   Word2vec   Text similarity   TF-IDF   AHP-TOPSIS

# 目录

摘要.....	1
一、介绍.....	7
1.1    问题背景.....	7
1.2    问题描述.....	7
1.3    论文的结构安排.....	7
二、文本数据预处理方法介绍.....	7
2.1    数据清洗.....	7
2.1.1    数据去重.....	8
2.1.2    数据去噪.....	8
2.2    中文分词.....	8
2.2.1    命名实体识别.....	8
2.2.2    新词识别.....	9
2.2.3    词性标注.....	9
2.2.4    Jieba 分词.....	9
2.2.5    去停用词.....	9
2.3    词向量训练.....	10
三、模型假设.....	10
四、群众留言分类.....	11
4.1    留言分类模型的选择.....	11
4.1.1    问政平台留言特点.....	11
4.1.2    文本分类模型研究综述.....	12
4.2    留言分类模型的建立.....	14
4.2.1    数据预处理.....	14
4.2.2    融合注意力机制的卷积神经网络的建立.....	15
4.2.1    模型评价方法的介绍.....	17
4.3    实验与结果.....	18
4.3.1    模型训练方法说明.....	18
4.3.2    网络参数的确定.....	18

4.3.3 结果分析.....	21
五、热点问题挖掘.....	23
5.1 相似问题归类模型的建立.....	24
5.1.1 数据预处理.....	24
5.1.2 留言相似度的计算.....	25
5.1.3 相似文本聚类.....	27
5.2 热度评价体系的建立.....	27
5.2.1 热度评价指标的选取与量化.....	28
5.2.2 热度指数的计算.....	28
5.3 结果.....	30
六、答复意见的评价.....	31
6.1 评价模型的建立.....	31
6.1.1 评价指标的选取与量化.....	31
6.1.2 基于群众情感分析的答复意见评价模型.....	39
6.2 结果.....	41
6.2.1 评价指标.....	41
6.2.2 基于群众情感分析的答复意见综合评价.....	44
七、模型评价.....	46
7.1 模型的优点.....	46
7.2 模型的缺点.....	47
八、参考文献.....	48

## 一、介绍

### 1.1 问题背景

十九届四中全会明确提出“保持党同人民群众的血肉联系，把尊重民意、汇集民智、凝聚民力、改善民生贯穿党治国理政全部工作之中”。

近年来，随着以互联网为代表的数字媒体快速发展，网络问政平台逐渐成为了政府了解民意、汇集民智、凝聚民力的重要渠道。以人民网《领导留言板》为例，据人民网 2020 年 1 月报道，2019 年全国各地网民通过《领导留言板》留言近 50 万件，同比增长 58%，创历史新高。网络问政平台的公众留言数量不断攀升，给以往主要依靠人工进行留言处理的相关工作部门带来了巨大的压力，如：在有大量文本信息需要处理的情况下，无法同时保证留言分类的效率与准确率；为了及时发现热点问题有针对性地解决问题，计算热点问题的热点指数也增加了相关部门的工作量。同时，大量的留言回复给相关政府部门对回复进行评价与改进带来了困难。因此，建立基于自然语言处理技术的智慧政务系统显得尤为迫切。

### 1.2 问题描述

在以往处理网络问政平台的留言时，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相应的职能部门进行处理，然后会对热点问题整理以便相应职能部门有针对性地进行处理，最后相关职能部门会给留言留下答复意见。因此本文需要解决的主要问题有三：一是根据已有的分类数据建立关于留言内容的一级标签分类模型；二是通过定义合理的热度评价指标来计算出热点问题的热点指数；三是从答复的不同角度对答复意见的质量给出一套评价方案。

### 1.3 论文的结构安排

本文首先统一介绍所用到的文本数据预处理方法与阐述模型假设，然后在论文的第四、五、六部分分别阐述三个问题具体的建模过程与结果分析，最后对模型的优缺点进行剖析。

## 二、文本数据预处理方法介绍

这部分将介绍本文会用到的几种文本数据预处理的方法，包括数据清洗、分词、命名实体识别、新词识别、词性标注、去停用词和词向量训练，后续在模型中具体应用时会再加以简单的说明。

### 2.1 数据清洗

### 2.1.1 数据去重

考虑到希望问题得到解决的迫切心情，某些用户可能会针对某一问题多次提交相同留言，而这样的数据在后续的问题中是有意义（尤其在热点问题挖掘部分），所以对于相同留言标题、留言详情的数据我们不予去除。我们仅去除【留言时间】相同的数据，这可以去掉由系统失误在同一时间产生的重复数据。

### 2.1.2 数据去噪

针对本问题而言，没有意义的符号会增加数据分析的负担。所以我们利用正则表达式去除文本中的空格、x 序列和标点符号。

正则表达式是对字符串（包括普通字符（例如，a 到 z 之间的字母）和特殊字符操作的一种逻辑公式<sup>[4]</sup>，可以完成以下两种工作：一是判断给定的字符串是否符合正则表达式的过滤逻辑；二是从字符串中获取我们想要的特定部分。因此，正则表达式具有非常强的灵活性、逻辑性和功能性，可以用于完成本文的数据去噪等任务。

## 2.2 中文分词

中文分词操作是文本预处理过程中必不可少的环节，是中文文本分类研究的基础。当下比较常见的分词方法有三种：

- 基于字符串匹配的分词算法：把待分析的文本与词典中的词进行匹配，如果匹配成功，即算词识别成功。依据识别扫描方向的不一致，可以进行正向匹配、反向匹配与双向匹配。这种方法较为简单，但却完全依靠词典进行匹配，不能进行歧义判断，而且词典本身就小不能完全概括所打调，不适用于大规模的文本分词操作。
- 基于统计的分词算法：判断上下文中相邻的字之间的关联程度，一起出现的次数越多，构成一个词的可能性越大，如果这种紧密程度高于某个数值，就可以认定它们组成了一个词。此方法可以识别未登录词，但像“他的”、“这一”这些高频组合不属于词的范畴，会影响分词结果。
- 基于理解的分词算法：基于理解的分词算法是一个推理过程，需要对整个语法进行语法和语义的分析，会用到大量的语言和信息知识来对词做界定。而中文博大精深，语义复杂，很难直接从语义语法角度将文字分析组织成计算机可以理解的形式，目前这个方法还处于研究当中。

### 2.2.1 命名实体识别



命名实体识别（Named Entity Recognition，简称 NER），是指识别文本中具有特定意义的实体，主要包括人名、地名、机构名、专有名词等。在本问题中，命名实体识别可以识别出小区名、机构名等名词。进一步，我们将这些名词加入到分词自定义词典中。这样可以将文本中小区名、机构名等名词更好地识别出来，有助于文本分类等进一步分析。

在这里我们对比了 Stanford CoreNLP<sup>[2]</sup>和 NLPIR 大数据语义智能分析平台<sup>[3]</sup>，发现两者的命名实体识别效果基本相同。但前者运算速度慢且操作复杂，所以我们使用后者进行本题的命名实体识别。识别后将词纳入自定义词典中。

### 2.2.2 新词识别

新词识别能从文本中挖掘出具有内涵新词、新概念，用户可以用于专业词典的编撰，还可以进一步编辑标注，导入分词词典中，提高分词系统的准确度，并适应新的语言变化。在这一步，我们通过 NLPIR 大数据语义智能分析平台来进行新词识别，并将结果纳入自定义词典。

### 2.2.3 词性标注

在本题中，词性标注主要用于我们对某些重要词语（如小区名）进行词性观察，了解其词性。这有助于我们在提取关键词时通过限定词性来提取出这些重要词语。这一需求我们通过 Jieba 库来实现。

### 2.2.4 Jieba 分词

目前，中文分词有很多可以直接使用的分词程序：NLPIR（由中科院的张华平研发），但准确率较低；Jieba 分词软件，是在工业界口碑很好的分词软件；THU Lexical Analyzer for Chinese 分词软件，同时具有分词与词性标注功能<sup>[4]</sup>。比较后，我们选择的是 Jieba 分词工具，该工具简单实用，具有快速处理的优势，分词的准确率也较高。同时加入由附件一中的标签词、命名实体识别出的词、新词发现和人工提取的关键词组成的自定义词典，让分词结果更为准确。

### 2.2.5 去停用词

停用词是指像“的”、“了”、“得”、“在”这样的自身无明确意义的词。在分词结果中存在大量的停用词，而停用词对于本问题的解决没有帮助。所以为了减少储存空间，提高运算效率，我们将停用词从分词结果中剔除。停用词主要来自于哈工大停用词表、百度停用词表<sup>[5]</sup>和人工添加。

## 2.3 词向量训练

为了便于文本分析，所以我们需要将词数字化。在此我们使用词向量来表示词的特征。目前中文词向量的训练主要有下面三种方法：Word2Vec、Glove 和 FastText。本题我们使用的是这三者中最常用的 Word2Vec 模型。

Word2Vec 中包含两个模型——CBOW 模型和 Skip-gram 模型。两个模型都包含三层：输入层、投影层和输出层。前者是在已知词 $w_t$ 的上下文 $w_{t-1}$ 、 $w_{t-2}$ 、 $w_{t+1}$ 、 $w_{t+2}$ 预测当前词 $w_t$ ；后者恰恰相反，在已知当前词 $w_t$ 的前提下，预测其上下文 $w_{t-1}$ 、 $w_{t-2}$ 、 $w_{t+1}$ 、 $w_{t+2}$ 。

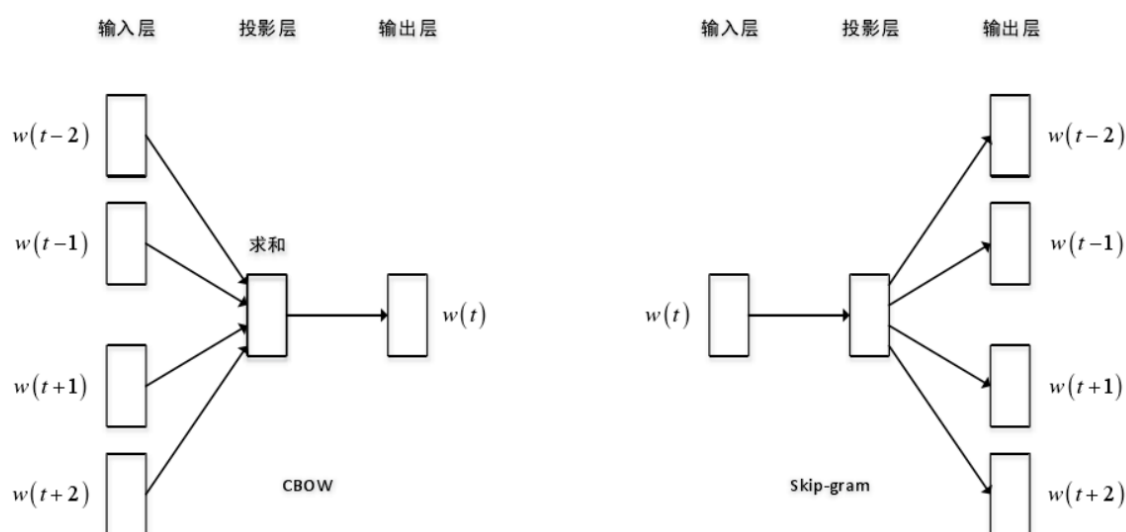


图 1 CBOW 模型和 Skip-gram 模型示意图

本题最终选择 Skip-gram 模型，因为 Skip-gram 模型对低频词效果更好，但是当语料充足后，两者的训练结果效果差距不是很大，但是 Skip-gram 模型的训练时间较 CBOW 模型的训练时间有较多的增加。另外，更丰富的语料和更高维的词向量能更好的拟合语料中的语义信息<sup>[6]</sup>。

经过试验，我们发现仅凭题目所给语料训练出来的词向量效果不好。所以我们采用以下办法进行改进：

- 1) 使用北京师范大学开源的预训练词向量模型<sup>[7]</sup>
- 2) 在此基础上利用题目所给语料扩充预训练词向量模型

经过测试，通过该方法得到的词向量模型效果良好。

## 三、模型假设

➤ 假设 1：假设不同人拥有不同的留言用户 ID 且同一人在不同时间留言其用户 ID 不变；

虽然一般留言用户 ID 是与网络 ID 或账号相对应的，但是通常每个人只会使用一个网络 ID

或注册一个账号，而且该假设可以使在热点问题挖掘部分定义问题涉及人群的广度时有合理的量化方法。因此该假设有一定的合理性与必要性。

➤ **假设 2：**群众留言时会说明问题发生地的省级或市级或区级行政单位；

通常群众在留言时都会有说明问题发生地的具体行政单位的习惯，同时在问政平台还可以考虑强制群众注明问题发生地的行政单位。而利用该假设我们既能够在热点问题挖掘部分大大降低文本相似度的计算量，又能够在一定程度上提高热点问题聚类的准确率。因此该假设也具有一定的合理性和必要性。对于本次题目所给数据(附件三)中没有注明行政单位的留言，我们首先找出与其所描述的问题相同的留言，将这些留言中所注明的行政单位当作该留言所在的行政单位；对于无法通过这种方法找到其所属行政单位的留言，我们将其视为面对西地省提出的问题，即其所属行政单位为西地省。

➤ **假设 3：**留言点赞数高于 50 的为人为恶意刷赞。

我们初步处理数据后发现，99%的留言点赞数在 0-50 这个区间内，但有个别留言的点赞数高达 1500+。因此，我们将点赞数高于 50 的留言视为群众为了引起政府关注而恶意刷赞的留言，并将其数据剔除，以保证后续热点问题挖掘的准确性。

## 四、群众留言分类

在这部分我们将解决问政平台群众留言分类问题，实际上这是一个文本多分类问题，因此我们建立了基于注意力机制的卷积神经网络模型来解决这一问题。在这部分我们将主要对以下三个内容进行详细的阐述：一是文本分类模型的选择理由；二是模型的构建过程；三是模型的实验结果。

### 4.1 留言分类模型的选择

我们选择留言分类模型的思路为：首先分析问政平台群众留言的特点，然后在深入了解多种文本分类方法的适用情形与优缺点后依据留言特点以及分类的要求选取恰当的分类模型。具体分析如下。

#### 4.1.1 问政平台留言特点

在初步浏览数据以及查阅相关文献后，我们认为群众在问政平台的留言主要有以下几个特点：

- ① 语境性强。存在一词多义的现象，不同留言使用同一词汇，需结合具体的语境分到不同类别。例如：在“交通局的亲属拖欠我们工资”与“交通局对某地交通的管制不到位”两条留言中虽然均存在“交通局”一词，但是这两条留言属于不同类别。因此适合采用考虑语境与目标词汇映射关

系的 Word2Vec 来训练词向量<sup>[8]</sup>。

- ② 留言正文文本冗余，标题蕴含信息较少。群众留言的正文中对于问题的描述和情绪表达占据较大篇幅，因此若仅采用正文进行分类会影响准确性。而留言的标题虽含有关键信息但信息量较少，若仅采用标题文本进行分类会有较大误差。因此我们在所建立的文本分类模型中将综合利用留言的标题和正文信息

### 4.1.2 文本分类模型研究综述

文本分类的流程一般为文本预处理、文本表示、特征提取、分类器分类和分类结果评价，在这部分我们主要介绍各个分类器的优缺点以及它们的适用情形。

20 世纪 60 年代，Maron 创新性的提出概率索引模型，利用贝叶斯公式解决文本分类问题<sup>[9]</sup>。到了 20 世纪 70 年代，Salton 等人提出了向量空间模型，往后的一段时间里，文本自动分类方法主要依靠传统的知识工程。直至 20 世纪 90 年代，基于机器学习和统计方法的文本分类模型逐渐发展起来，使得文本分类更加准确、高效并成为了目前文本分类的主流方法。

文本分类模型可分为基于传统机器学习的分类模型和基于深度学习的分类模型两大类，较为常用的传统分类模型有：朴素贝叶斯、支持向量机、神经网络算法、k 最近邻、决策树、Rocchio 算法，它们的优缺点如表 1 所示。

分类算法	优点	缺点
朴素贝叶斯	算法简单，在特征属性相关性较小时具有最优的性能；所需估算的参数少，对缺失数据不敏感，适合增量式训练	分类效果依赖于先验概率；在属性较多或属性间相关性较大时分类效果较差
支持向量机	使用的训练集少，可以处理高维稀疏文本数据和解决非线性问题，对特征相关性不敏感；可以避免神经网络结构选择和局部极小点问题	样本较多时训练速度慢，对缺失数据较为敏感；核函数的选择没有统一的标准
神经网络算法	能解决较为复杂的非线性问题，分类准确度高；对噪声数据不敏感	需要确定大量参数，且难以解释其输出结果；模型训练需要花较长时间

K 最近邻	不需要进行特征选取，易处理类别数目多或类别交叉的情况，且性能稳定	空间复杂度高，样本容量小时分类准确度不高；K 值的选取直接影响分类的性能
决策树	易于解释和理解；对数据预处理要求低；计算较为简单，且可通过静态测试对模型进行评价；可构造多属性决策树	难以处理时序型数据、连续型数据和缺失数据；易出现过拟合以及忽略属性之间的相关性
Rocchio 算法	算法极易实现和理解，分类效率高	受文本集分布的影响，原型向量可能落在相应的类别外

表 1 传统分类模型的优缺点

随着计算机性能的提升，深度学习技术在语音、文本、图象等信息的处理中有突破性的发展。相对于传统的文本分类模型，深度学习方法能够通过多层语义操作，获得更高层更抽象的语义表征，并将特征提取工作融合于模型的构建过程中，减少人为设计特征的不完备性与冗余<sup>[10]</sup>。因此我们将选取深度学习模型来解决本次的留言分类问题。

在文本分类领域较为常用的深度学习分类模型有：循环神经网络、卷积神经网络，以往也有研究者将卷积神经网络与循环神经网络结合使用或在深度学习模型中融入注意力机制、LSTM 等方法。循环神经网络(RNN)可以在时间序列上对文本分类进行建模，其优点是具有很强的捕获上下文信息的能力，理论上每个时刻的隐态可以包含之前所有输入的信息。循环神经网络的缺点是：在模型学习的过程中，随着网络层数的增加，使得模型面临梯度爆炸或梯度消失的问题，最后影响分类结果。而且其记忆偏重于最后阶段的学习内容<sup>[11]</sup>。但随着对循环神经网络研究的深入，有研究者在普通循环神经网络的基础上引入门控机制控制信息的累积，有效缓解了梯度爆炸和梯度消失<sup>[12]</sup>。如记忆单元更新的长短期记忆网络(long short-term memory, LSTM)，但是如此一来会导致计算量剧增<sup>[13]</sup>。因此本次我们将不考虑使用 RNN 模型。

2014 年 Facebook 的研究员 Kim 发表了基于卷积神经网络(CNN)做文本分类的经典论文<sup>[14]</sup>，相比与循环神经网络，卷积神经网络的优点在于：在计算特征向量的过程中保留了有用的局部词序信息，多尺寸卷积核具有强大的捕获局部特征的能力，并且可以高度并行化，模型训练速度较快<sup>[12]</sup>。卷积神经网络的缺点是模型中卷积核尺寸的设置没有统一的方法，而且卷积核的尺寸会直接影响最终分类效果。

深度学习中的注意力机制即仿照我们人类的关注机制，即根据需要将注意力放在某一方面，而不是平均地关注所有内容。这种方法加快了信息的处理速度，并且筛选出来的都是高价值的信息，

也提升了信息处理的准确性<sup>[15]</sup>。近年来，注意力机制在图象识别、语音识别和文本分类等领域均有广泛的使用。Bahdanau 等人提出了注意力机制使模型自动学习到区分句子中重点单词的能力，实验表明如此结合有较好的效果<sup>[16]</sup>。将注意力机制与 CNN 模型相结合可以使模型关注更为重要的特征并对文本上下文内容的关联性处理有更好的效果<sup>[17]</sup>，从而提高 CNN 文本分类的效果。

基于以上所有分析，我们选择建立融合注意力机制的卷积神经网络模型来进行群众留言分类工作。

## 4.2 留言分类模型的建立

这部分将具体阐述模型的建立过程和相关的优化操作。卷积神经网络的网络结构包含输入层 (Input Layer)、卷积层(Convolutional Layer)、池化层(Pooling Layer)、全连接层(Connection Layer)，其中输入层输入数据，卷积层与池化层对文本信息进行特征提取，全连接层对数据进行分类计算操作并输出分类结果<sup>[18]</sup>。在数据进入输入层之前，我们将对数据进行新词识别、命名实体识别、分词、去除停用词、计算每个词的 TF-IDF 值以及利用 Word2vec 训练词向量等数据预处理操作。而且在池化层之后我们将加入注意力层，提取高维特征信息，从而将注意力机制与卷积神经网络模型相融合。整个模型的结构如图 2 所示。

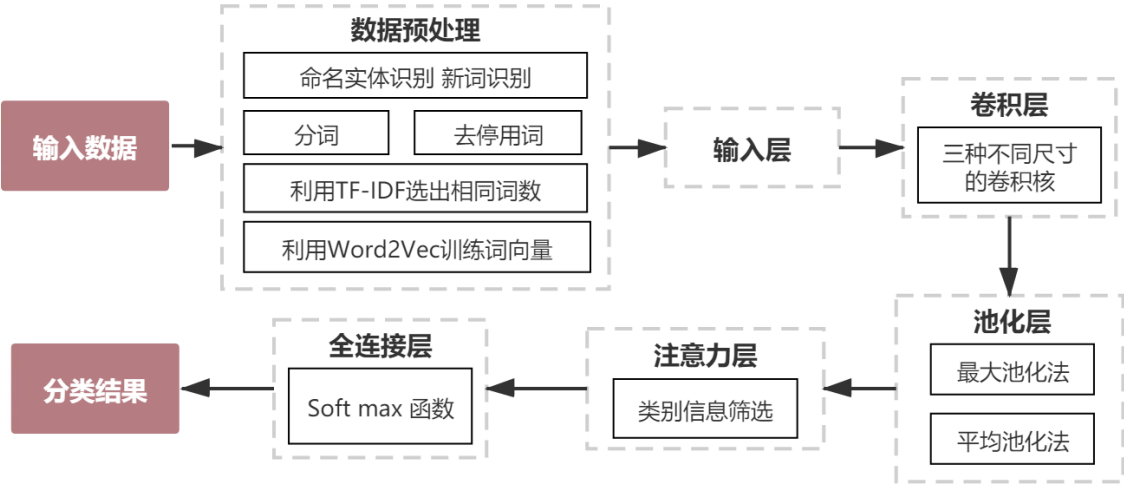


图 2 模型结构图

### 4.2.1 数据预处理

在问题一的数据预处理方面，我们首先利用新词识别与命名实体识别技术识别出留言中出现的新词、地点、人物与机构，如表 2 所示。然后将识别出来的词导入分词词库，以增加分词的准确性从而提高文本分类准确率。接着，我们计算每个词的 TF-IDF 值（该方法的具体介绍见后文），并依

据 TF-IDF 值选出每条留言的关键词。在这一步我们令每条留言选取的关键词数量相同，以方便后续数据处理，对于词数不够的留言在训练词向量后进行补 0 操作。最后我们利用 Word2vec 训练词向量，将文本转化成向量，方便后续将数据代入训练网络。

茅竹镇 n	竹山桥镇 n	圩镇 n	珠山 n	西路 n	黄兴镇 n
楚永路 n	涛圩镇 n	镇原 n	南洲镇 n	合心村 n	沙湾 n
丽北路 n	承德市 n	河北省 n	玉江村 n	溪湖 n	甘肃天水 n
河南鄆城 n	山东曹县 n	失陷村 n	浙江青田县 n	忠义乡 n	四川内江 n
公田镇 n	向家村 n	贵州普安 n	贵州省 n	博上村 n	月湖 n
江苏徐州 n	楠香村 n	岩乡 n	青山村 n	拖坪村 n	……

表 2 命名实体识别部分结果

#### 4.2.2 融合注意力机制的卷积神经网络的建立

##### 1、 输入层

将数据预处理后的数据输入卷积神经网络，即建立留言矩阵  $X \in R^{n \times m}$ ，其中  $n$  是留言分词后的词语数量， $m$  是词向量的维度。

$$X = [x_1, x_2, \dots, x_n]^T$$

##### 2、 卷积层

卷积层是整个卷积神经网络的重点，能够实现文本的特征提取并实现特征降维。卷积操作后输出的结果是多张特征图。简单的卷积神经网络中，一个卷积层中只含有一个卷积核。但因为留言文本较长且上下文具有一定的相关性，所以本此所建立的模型将利用多卷积核进行特征提取工作，这可以使提取到的特征更加精确<sup>[19]</sup>。当卷积核的滑动窗口为  $l$  时，卷积操作的公式如式(1)所示。

$$d_i = f(w \cdot c_{i:i+l-1} + b) \quad (1)$$

其中  $f$  为激活函数； $w$  为权重矩阵； $c_{i:i+l-1}$  代表第  $i$  到  $i + l - 1$  个词组成的词向量矩阵； $b$  为偏置量。本文使用的激活函数是 Relu 函数，它不仅能最大限度地加快模型收敛速度(远快于 sigmoid 函数和 tanh 函数)，而且还能避免训练过程中正区间的梯度消失问题，提高了模型的综合性能<sup>[15][20]</sup>。

##### 3、 池化层

池化层进行降维操作，可以有效地减少特征图的尺寸以及全连接层的参数数量，从而提升模型的运算速度并避免过拟合问题。为了考虑代表性的特征，本文采用最大池化法进行池化操作。

#### 4、注意力层

大多数注意力机制的研究都附着在 Encoder-Decoder 的框架下，结合注意力机制的 Encoder-Decoder 框架<sup>[21]</sup>如图 3 结合注意力机制的 Encoder-Decoder 框架所示。注意力机制的特点是动态地分配权重，以便将注意力资源分配到更有效的信息上。

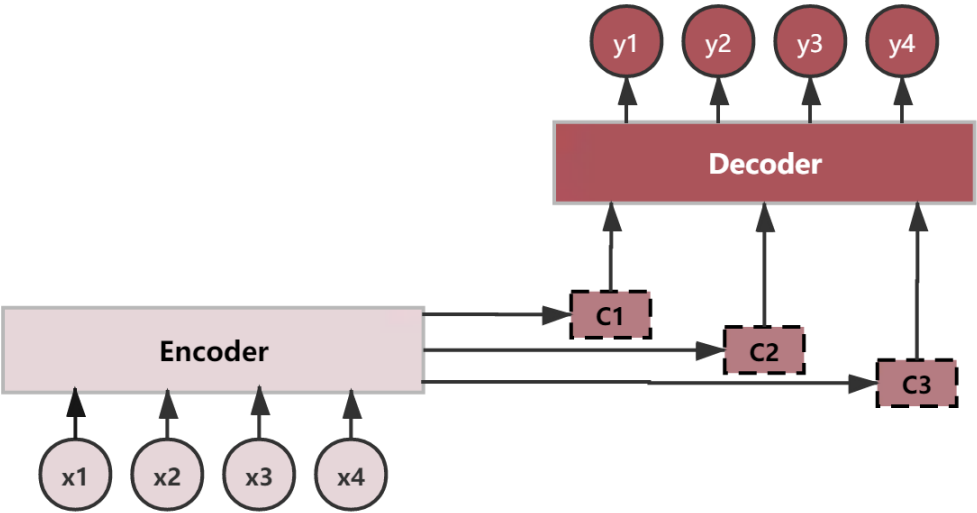


图 3 结合注意力机制的 Encoder-Decoder 框架

但是目前大多数研究均将注意力机制用于提取文本内容的低层特征，然后再将处理好的数据代入神经网络进行分类。这种方法忽略了文本高层特征的提取操作，且最终的分​​类准确率并不是很高。因此本文参考文献的做法，利用注意力机制对文本高层特征进行整合提取，获得利于分类的高层特征。主要内容即为以下两个方面：一是考虑特征词在类别信息的表达上面的贡献率大小，关注与选取出表达类别信息较多的特征词；二是将特征图做进一步的特征提炼，筛选出具有代表性的特征词同时减小特征图大小，减轻计算机的负担。本文注意力层的基本结构如图 4 所示

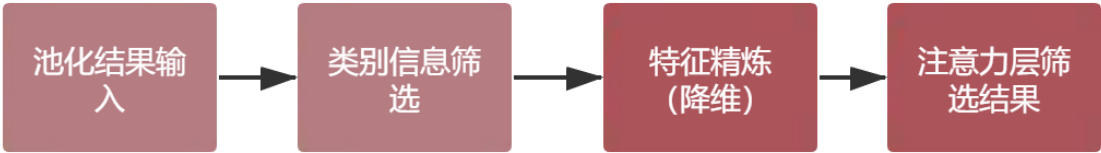


图 4 注意力层基本结构

在类别信息筛选部分将利用池化层的结果选取出表达类别信息较多的特征词。平均池化关注特征的整体有效性，最大池化则关注对于类别区分更加明显的特征，类别信息筛选则是综合考虑两个池化的结果，对特征做进一步筛选。具体的类别信息筛选结果 A 的计算



公式如式(2)。

$$A = \sigma(W_1(P_1) + W_2(P_2)) \quad (2)$$

其中 $\sigma$ 为 Relu 激活函数,  $W_1, W_2$ 为训练得到的权重矩阵,  $P_1$ 是平均池化结果,  $P_2$ 是最大池化结果。

在特征精炼部分利用卷积操作对类别信息筛选结果进行特征降维, 最终便得到注意力层的筛选结果。

## 5、全连接层

全连接层利用 SoftMax 函数整合前面所有层级处理后的特征信息完成留言文本分类任务, 输出最终的分类结果。

## 6、模型优化设置

### (1) Dropout 操作

Dropout 操作又名随机失活操作, 其作用是: 每次训练迭代时随机使一部分神经元处于失效的状态, 从而避免训练时出现过拟合问题。实验表明当 dropout 概率为 0.5 时, 分类准确率最高, 训练时间也可接受<sup>[19]</sup>。因此我们取 dropout 的概率为 0.5。

### (2) Batch normalization

Batch normalization 操作指的是在每一层输入前对数据进行标准化处理。这样即保证了输入可以回到标准的正态分布状态, 同时又消除数据的分布变动, 使梯度维持在较大的状态, 损失函数能以较快地速度向最优值靠近, 加速模型训练直至收敛<sup>[15]</sup>。

## 4.2.1 模型评价方法的介绍

对于分类方法的评价, 我们通常使用  $F - Score$  评价方法:

$$F_1 = \frac{1}{n} \sum_{i=1}^n \frac{2P_i R_i}{P_i + R_i} \quad (3)$$

其中 $P_i$ 为第 $i$ 类的查准率,  $R_i$ 为第 $i$ 类的查全率。

下面就式子中出现的查准率、查全率以及式子的含义进行说明:

### 1. 查准率 (precision)

分类器在一个类别中正确分类的文本数与分类器在该类上实际分类的文本数的百分比, 准确率越高表明分类器在该类上出错的概率越小。用公式表示如下:

$$P_j = \frac{TP_j}{TP_j + FP_j}$$

其中,  $TP_j$ 指的是被分类器正确分类到类别 $c_j$ 的文本数;  $FP_j$ 指的是实际不属于类别 $c_j$ 却被分

类器错误地分到类别 $c_j$ 的文本数。

## 2. 查全率 (recall)

分类器在一个类别中正确分类的文本数与该类应有分类正确的文本数的百分比，召回率越高表明分类器在该类上可能漏掉的分类越少。用公式表示如下：

$$R_j = \frac{TP_j}{TP_j + FN_j}$$

其中， $TP_j$ 指的是被分类器正确分类到类别 $c_j$ 的文本数； $FN_j$ 指的是实际属于类别 $c_j$ 却没有被分类器正确地分到类别 $c_j$ 的文本数。

## 2. $F - Score$

在深度学习中，精确率(Precision)和召回率(Recall)是常用的评价模型性能的指标，从公式上看两者并没有太大的关系，但是实际中两者是相互制约的。我们都希望模型的精确了和召回率都很高，但是当精确率高的时候，召回率往往较低；召回率较高的时候精确率往往较低。因此，很多时候，我们需要综合权衡这 2 个指标，这就引出了一个新的指标 $F - Score$ ，综合考虑 Precision 和 Recall 的调和值

$$F - Score = (1 + \beta^2) \frac{Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \quad (4)$$

当 $\beta = 1$ 时，成为 $F_1 - Score$ ，这时查准率和查全率都很重要，权重相同。

## 4.3 实验与结果

在这部分我们首先阐述模型的训练方法和相关网络参数的确定过程，然后呈现最终的模型结果。最后对比我们所建立的融合注意力机制的卷积神经网络模型的与其他文本分类模型的分类效果。注：问题一用到的所有方法的源代码请见附件，且由于题目所给数据（附件二）仅涉及 7 类一级标签，因此我们在训练模型时设置总类数为 7。

### 4.3.1 模型训练方法说明

在本次构建的卷积神经网络模型中，我们采用具有动量的随机梯度下降(SGDM)训练网络。传统的随机梯度下降(SGD)训练网络在遇到沟壑时容易陷入震荡，因此可以为其引入动量(Momentum)，加速 SGD 在正确方向的下降并抑制震荡<sup>[22]</sup>。而且该训练网络的训练速度更快，并有一定的摆脱局部最优的能力。我们将 SGDM 训练网络的最大训练轮数设置为 4。

### 4.3.2 网络参数的确定

我们将数据集的 90%用于训练模型，将其余 10%作为验证集。然后对不同的参数组合的网络进行训练，对比验证集分类准确度、 $F_1$ 值与训练时间。实验环境配置如表 3 所示

开发工具	配置情况
操作系统	Windows 10
开发语言	MATLAB, Python
CPU	Intel Core i5-7300HQ CPU @ 2.50GHz
内存	8G
硬盘	1T

表 3 实验环境配置

卷积神经网络不同于其他传统算法，网络参数的选择会直接影响整个网络的性能与最终文本分类的准确率。本次模型所涉及到的各个参数的最优取值如表 4 所示，下面我们将以卷积层大小、关键词数目和 Dropout 概率为例阐述参数确定的过程以及其合理性。

网络参数	取值
词向量维度	128
卷积层大小	$1 \times 128 / 1 \times 128 / 1 \times 128$
关键词数目	60
Dropout 概率	0.5
SGDM 初始学习率	0.001
SGDM 最大训练轮数	4

表 4 网络参数取值

### 1、卷积层大小

卷积层的大小影响着特征选取的准确率，若卷积层的窗口尺寸过小，则会引发特征图的维度灾难问题；若卷积层的窗口尺寸过大，则会影响计算级运行速度，且有可能在特征提取的过程中引入噪声。正如前文所说，为了使提取到的特征更加精确，我们设置了三层卷积层进行特征提取工作，并选取了多种尺寸进行多次实验，具体如表 5 所示。

卷积层窗口尺寸	验证集分类准确率	F <sub>1</sub> 值	训练时间
1×128/2×128/3×128	78.94	0.7697	76m2s
3×128/4×128/5×128	76	0.7392	69m57s
4×128/5×128/6×128	73.51	0.7113	90m59s
5×128/6×128/7×128	75.68	0.7275	76m36s
6×128/7×128/8×128	73.18	0.7043	78m23s
10×128/11×128/12×128	69.6	0.6536	92m22s
1×128/1×128/1×128	81.22	0.7935	52m7s
2×128/2×128/2×128	79.8	0.7729	60m6s
3×128/3×128/3×128	76.44	0.7374	66m55s

表 5 不同卷积层窗口尺寸实验结果

注：这些实验中其他参数均相同，词向量维度为 128；关键词数目为 40；Dropout 概率为 0.5；SGDM 初始学习率为 0.001；SGDM 最大训练轮数为 4。

由表可知，最优的窗口尺寸组合为：1×128/1×128/1×128，此时验证集分类准确率与F<sub>1</sub>值均最高，且所用的训练时间最短。因此我们选取该组合作为本次模型的卷积层窗口尺寸组合。

## 2、关键词数目

每条留言选取的关键词数目的多少会影响最终分类精度与网络训练时间，若关键词数目过少，则会因输入特征不足而导致分类进度下降；若关键词数目过多则会加大计算难度，增加网络训练时间。正如前文所说，我们将利用 TF-IDF 法选取关键词。因此，我们分别选取每条留言中 TF-IDF 值为前 40、前 60 和前 80 的词语进行实验，结果如表 6 所示。

关键词数量	验证集分类准确率	F <sub>1</sub> 值	训练时间
40	81.22	0.7935	52m7s
60	85.02	0.8383	63m52s
80	84.69	0.8332	85m11s

表 6 不同关键词数量实验结果

注：这些实验中其他参数均相同，词向量维度为 128；卷积层窗口尺寸为1×128/1×128/1×128；Dropout 概率为 0.5；SGDM 初始学习率为 0.001；SGDM 最大训练轮数为 4。

由表可知，当关键词数量为 60 是有最高的分类准确率和 $F_1$ 值，且训练时间可以接受。因此我们选取 60 作为本次模型的关键词数量。

### 3、Dropout 概率

随机失活操作 Dropout 的主要作用是避免网络训练的过程中出现过拟合现象。Dropout 操作会在每一层随机除去一些冗余的神经节点，保留具有代表性的信息。我们分别选取 Dropout 概率为 0.1-0.9 做了实验，结果如表 7 所示。

Dropout 概率	验证集分类准确率	$F_1$ 值	训练时间
0.1	81.43	0.8007	66m29s
0.2	83.17	0.8188	68m41s
0.3	84.69	0.8293	70m29s
0.4	83.39	0.825	68m19s
0.5	85.02	0.8383	63m52s
0.6	83.5	0.8266	66m53s
0.7	80.13	0.7838	67m32s
0.8	18.68	0.045	77m53s
0.9	18.68	0.045	66m45s

表 7 不同 Dropout 概率实验结果

注：这些实验中其他参数均相同，词向量维度为 128；卷积层窗口尺寸为 $1 \times 128 / 1 \times 128 / 1 \times 128$ ；关键词数目为 60；SGDM 初始学习率为 0.001；SGDM 最大训练轮数为 4。

由表可知，当 Dropout 概率为 0.5 时，验证集分类准确率最高，训练时间也最短。故本次模型的 Dropout 概率选为 0.5。

此外，词向量维度、SGDM 初始学习率与 SGDM 最大训练轮数也是经过多次实验后选定的，其过程类似以上几个参数，因此这里便不再赘述。

### 4.3.3 结果分析

最优参数组合的模型训练结果如图 6 所示，由此可知我们所建立的融合注意力机制的卷积神经网络模型分类准确率高达 85.02%， $F_1$  值为 0.8383，训练时间仅需 63 分钟 52 秒，效果出众。

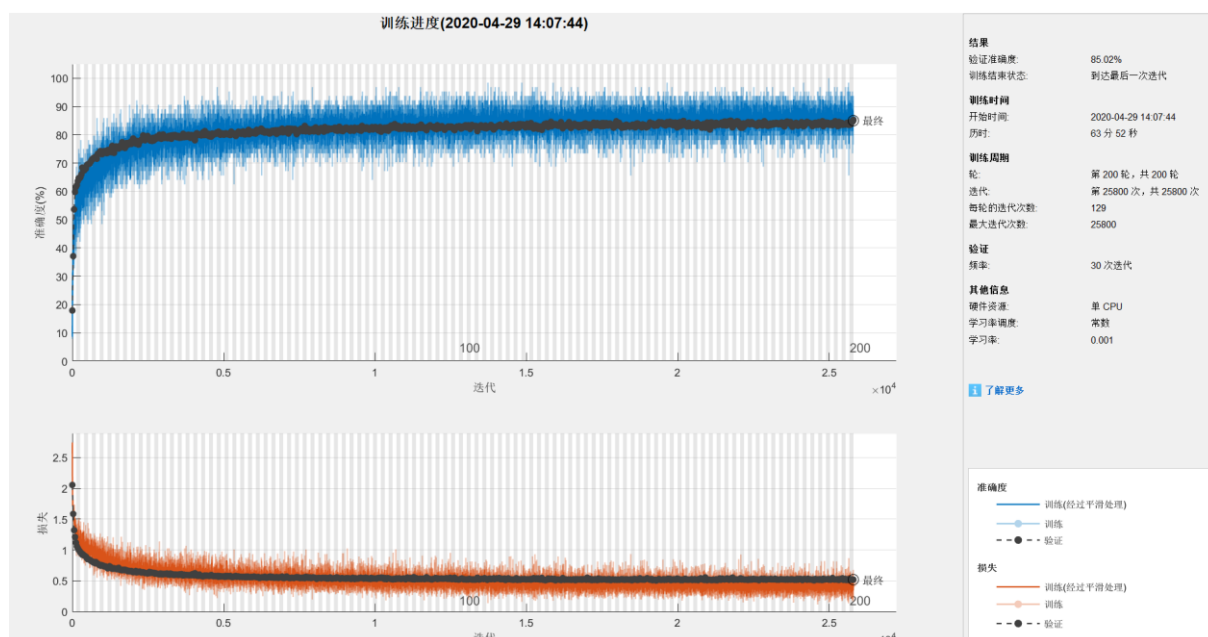


图 5 模型训练结果

由图 5 可知，当模型训练到第 30 轮时就具有较为不错的性能，这也正是我们所构造的模型的一大优点。因此我们认为，若时间紧迫或数据量增多，可以适当减少模型训练轮数，以节省模型训练时间。

此外，我们还设计了另外两个卷积神经网络模型，对比其训练结果与我们模型的差异，具体如图 6 所示。

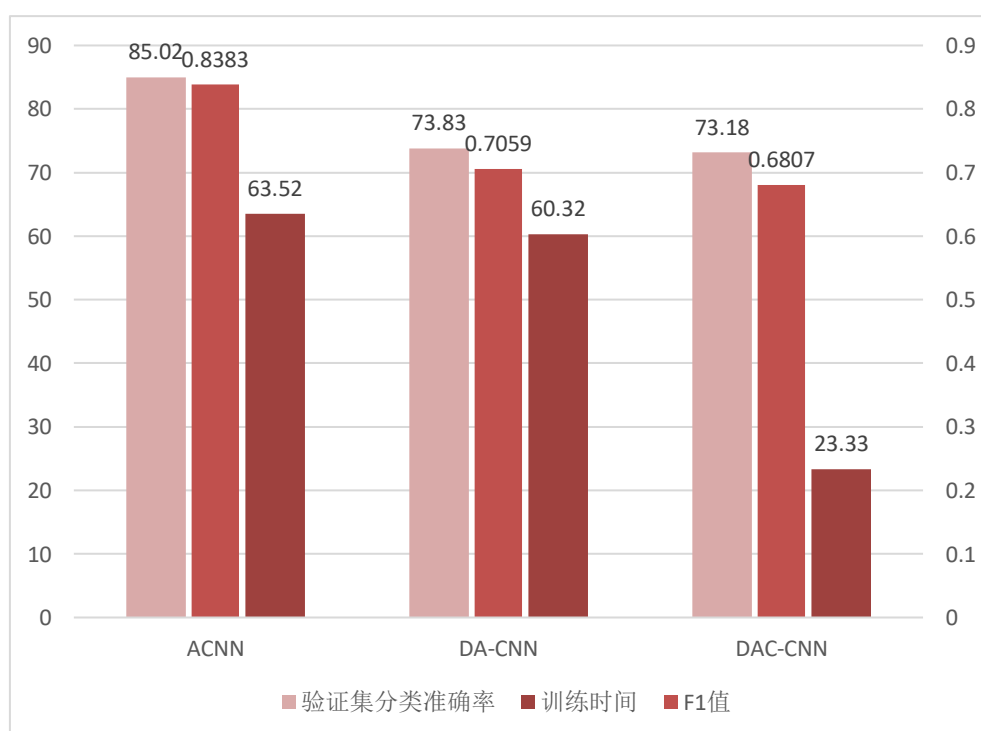


图 6 不同模型结果对比

注：除了卷积层个数差异外，三个模型的其他参数均相同。其中 ACNN 为我们所建立的融合注意力机

制的卷积神经网络模型；DA-CNN 为没有注意力机制的卷积神经网络模型；DAC-CNN 为没有注意力机制且仅有一个卷积层的卷积神经网络模型。

由图可知，我们所建立的 ACNN 模型分类精度与 $F_1$ 值均是最优的，且对于传统的卷积神经网络模型(DAC-CNN)，我们所建立的模型可以将分类精度提升 11.84%，将 $F_1$ 值提高 0.1576。这说明我们的改进是有效的，且模型性能出众。虽然我们的改进导致了模型训练时间的增加，但是最终的训练时间仍在可接受的范围。

以上实验所用的词向量训练模型均是用题目所给语料训练的，但正如本文 2.3 节所说，我们经过调试后发现该词向量模型训练的不好，并对其做了改进。利用改进的词向量训练模型，我们用最优参数组合(卷积层窗口尺寸改为 $1 \times 300/1 \times 300/1 \times 300$ )对融合注意力机制的卷积神经网络进行了训练，结果如图 7 所示。

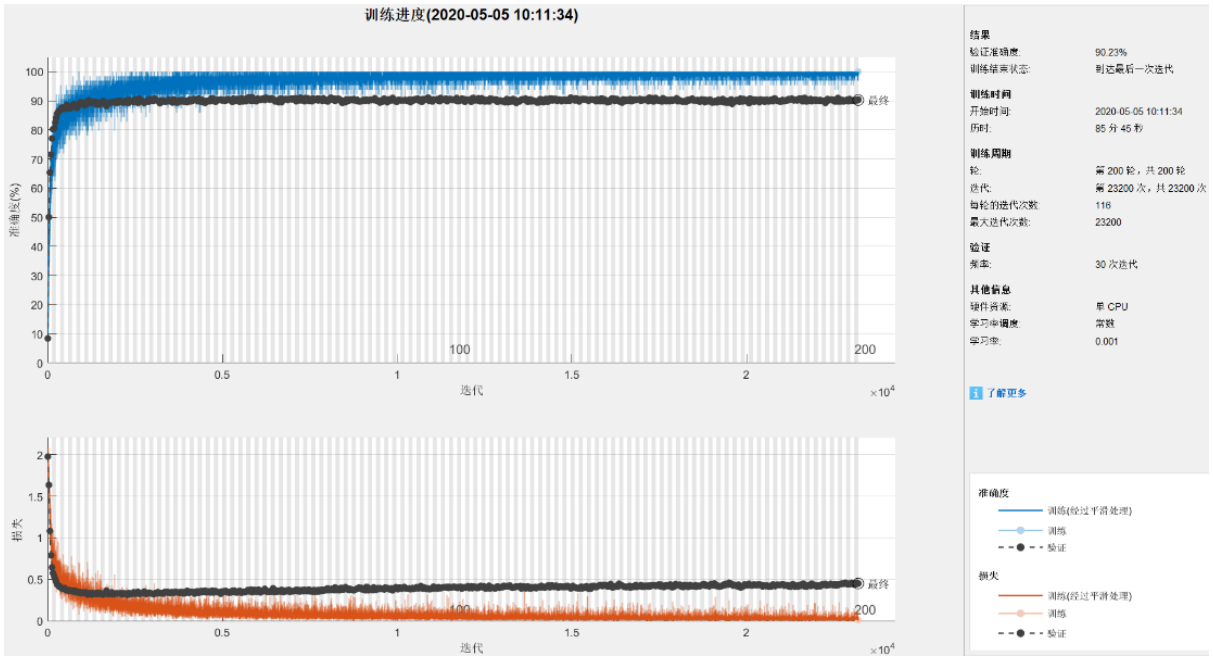


图 7 词向量训练模型改进后的结果

由图可知，该模型的具有非常优秀的分类能力，验证集分类准确率高达 90.23%，且训练时间也可接受，经计算该模型的 $F_1$ 值为 0.8940。此外我们认为，若有更加丰富的数据用于训练 ACNN 模型，将可以得到更加优秀的分类效果。

## 五、热点问题挖掘

在这部分我们需要完成以下两个任务：一是识别出反映同一问题的留言并进行归类；二是定义合理的热度评价指标，计算各个问题的热度指数，最终识别出某一时间段内的热点问题。因此我们将构建相似问题归类模型与热度评价体系来完成以上两个任务，下面具体阐述建模过程。

## 5.1 相似问题归类模型的建立

在这部分，我们首先利用正则表达式对各个区的留言进行筛选与分类，并利用词性标注以及计算 TF-IDF 值的方法选出每条留言的关键词，有效降低相似度计算模型的维度。其次，利用 Word2vec 训练词向量并利用余弦距离定义两个词之间的相似度。然后为了体现文本中不同重要程度的词语对文本相似度产生的不同影响，我们在定义文本相似度的计算方法时引入了加权因子，以更加准确地计算出文本间的相似度。最后，我们利用能在任意形状的样本空间上聚类且收敛于全局最优解的谱聚类<sup>[23]</sup>得出最终的归类结果。下面阐述具体的实现过程。

### 5.1.1 数据预处理

在问题二的数据预处理方面，我们同样进行新词识别、命名实体识别、分词、去停用词和词性标注等数据预处理工作，同时考虑到问政平台群众留言的特点，我们在停用词库中添加了如表 8 所示的停用词，以提高分词效果以及最终文本聚类效果。

问题	请问	小区	业主	投诉	居民
反映	建议	请求	希望	恳请	.....

表 8 停用词表

此外，我们还利用正则表达式（具体介绍见本文 2.1.2 节）筛选出各个地区的留言。正如模型假设部分所说，我们假设群众留言时会说明问题发生地的省级、市级和区级行政单位，而相同问题的留言一定会有相同的省级、市级和区级行政单位，因此我们利用行政单位留言进行分类是合理的，即能提高热点问题挖掘精度又能够大大降低后续文本相似度的计算量。具体分类过程如下：

- **Step1:** 利用正则表达式初步处理数据，发现所有留言中说明的省级行政单位仅有“西地省”，而市级、区级行政单位已被字母和数字替代。因此我们利用正则表达式判断留言中是否 26 个英文字母，将出现次数多的行政单位作为该留言的行政单位，而对于没有出现市级行政单位但出现“西地省”的留言，我们认为其描述的问题是面对整个西地省提出来的，并将其划分至西地省。例如：若某留言文本中含有三个“A”与一个“B”，则将该留言分到 A 市。
- **Step2:** 利用省级和市级行政单位对留言进行分类后，我们发现 A 市的留言有 4006 条，约为总数的 92.6%。因此我们再利用正则表达式对 A 市进行区级行政单位细分，对于没有注明具体区级行政单位的留言我们认为其描述的问题是面对整个 A 市提出来的。我们最后将所有留言分到了 12 个区域，具体如表 9 所示。



地区	留言数量	地区	留言数量
西地省	152	A4 区	289
A 市	1376	A5 区	245
其它市	131	A6 区	191
A1 区	253	A7 区	698
A2 区	321	A8 区	100
A3 区	505	A9 区	65

表 9 各区域留言数量明细

### 5.1.2 留言相似度的计算

#### 1、文本特征选取

我们利用每条留言中各个词的 TF-IDF 值与词性选出代表该留言进行文本相似度计算的关键词。对于一条留言来说，TF-IDF 方法是一种用于评价某一字词对于它们的重要程度的有效方法。其中 TF 是根据词频来增加词项权重的，而 IDF 则是一个处于特定条件下的词项概率分布的交叉熵。其计算公式为：

$$TF-IDF(w_{i,d}) = \frac{n_{w_i,d}}{\sum_k n_{w_k,d}} \times \log\left(\frac{|N|}{n_{w_i}}\right) \quad (5)$$

其中  $n_{w_i,d}$  表示某一特征词  $w_i$  在文本  $d$  中出现的次数， $|N|$  表示所有文本数， $n_{w_i}$  表示文本集合中包含特征词  $w_i$  的文本数量。

因此，可以利用 TF-IDF 从每一篇文本中挑选出若干个关键词，然后用这些关键词代表文本进行文本相似度计算。这样做的好处是能保证在不影响提取文本特征的前提下，尽一切可能来减少特征向量的维度<sup>[24]</sup>。此外，我们认为在一篇文本中，能够表示文本主干意思的一般是名词与动词，而形容词等往往起修饰作用。因此为了能更加准确地选出代表文本的关键词，我们还利用词性标注技术识别出每个词的词性，然后结合 TF-IDF 选出代表文本的关键词。具体步骤如下：

➤ **Step1:** 对每篇文本中的各个词进行词性标注，筛选出其中的词性为表 10 所示的词语：

词性	简写
名词	n
人名	nr
地名	ns
机构团体	nt
其他专名	nz
动词	v

表 10 选择的词性

- Step2: 计算所筛选出来的词语的 TF-IDF 值，并进行排序；
- Step3: 设定阈值 $P_1, P_2$ ( $P_1, P_2$ 为百分比)，在留言标题中选取 TF-IDF 值为前 $P_1$ 的词，在留言详情中选取 TF-IDF 值为前 $P_2$ 的词，而且令在留言标题和留言详情中分别选取的关键词均不超过 5 个，最后取并集即得到该留言的关键词。之所以要在留言标题和详情中分开取词，是因为我们认为留言标题是留言详情中关键信息的浓缩，具有较大的参考价值。此外，由于我们已经利用正则表达式对各个区域的留言进行了划分，因此为了避免误差，我们将剔除关键词中出现的区域词，如“西地省,A 市,A1,A2”等等。参考黄承慧等人<sup>[24]</sup>的研究，我们令 $P_1 = 70\%, P_2 = 30\%$ 。

如此一来，相较于传统的 TF-IDF 方法，一篇文本的关键词向量维度大大降低，从而较大地提高了后续算法的效率。

## 2、词相似度的计算

我们利用训练好的 Skip-Gram 模型输出得到所有词语的词向量表示（具体介绍见本文 2.3 节）。接着利用余弦距离定义词与词之间的相似度，从而即可计算出词语之间的相似度。

## 3、文本相似度的计算

正如前文所说，我们用选出来的关键词代表文本，从而文本间的相似度计算即可转化成词语间的相似度计算。设 $v_i, v_j$ 是两篇不同文本的关键词向量，其中 $v_i = (w_{i1}, w_{i2}, \dots, w_{im})$ ,  $v_j = (w_{j1}, w_{j2}, \dots, w_{jn})$ , 参考黄承慧等人的方法<sup>[24]</sup>，我们利用式(6)计算文本相似度：

$$Textsem(v_i, v_j) = wf_{ij} \times VectSim(v_i, v_j) \quad (6)$$

上式中 $wf_{ij}$ 为计算 $v_i$ 与 $v_j$ 之间相似度的加权因子， $VectSim(v_i, v_j)$ 为 $v_i$ 与 $v_j$ 之间的相似度。引入加权因子是为了体现文本中不同重要程度的词语对文本相似度产生的不同影响，我们认为两篇文本相似度较高的关键词的 TF-IDF 值所占比例越高，说明它们越能代表这两篇文

本进行相似度计算，因此我们利用相似度大于阈值的关键词的 **TF-IDF** 值所占的比例进行加权。此外，我们还消除每篇文本关键词向量维度不同的影响，使其满足基本的相似度量标准。具体定义如以下几式：

$$wf_{ij} = 1 + ave(i, j) \times \left( \sqrt{VectSim(v_i, v_j)} - VectSim(v_i, v_j) \right) \quad (7)$$

$$ave(i, j) = \frac{1}{2} \left[ \frac{\sum_{k \in \Lambda_i} TF \cdot IDF(w_{ik})}{\sum_{k=1}^m TF \cdot IDF(w_{ik})} + \frac{\sum_{l \in \Lambda_j} TF \cdot IDF(w_{jl})}{\sum_{l=1}^n TF \cdot IDF(w_{jl})} \right] \quad (8)$$

$$\Lambda_i = \left\{ k: 1 \leq k \leq m, \max_{1 \leq l \leq n} \{Sim(w_{ik}, w_{jl})\} \geq \mu \right\} \quad (9)$$

$$\Lambda_j = \left\{ l: 1 \leq l \leq n, \max_{1 \leq k \leq m} \{Sim(w_{jl}, w_{ik})\} \geq \mu \right\} \quad (10)$$

$$VectSim(v_i, v_j) = \frac{1}{2} \left[ \frac{1}{m} \sum_{k=1}^m \max_{1 \leq l \leq n} \{Sim(w_{ik}, w_{jl})\} + \frac{1}{n} \sum_{l=1}^n \max_{1 \leq k \leq m} \{Sim(w_{jl}, w_{ik})\} \right] \quad (11)$$

其中  $Sim(w_{jl}, w_{ik})$  为  $w_{jl}$  与  $w_{ik}$  的词相似度(余弦距离)，我们利用连续双射将词相似度取值范围从  $[-1, 1]$  映射至  $[0, 1]$ ； $\mu = 0.75^{[24]}$  为所设定的阈值。

完成以上三个步骤，我们便可以得到文本间的相似度矩阵，下面则利用相似度矩阵进行聚类操作。

### 5.1.3 相似文本聚类

在聚类方面，我们选择了谱聚类算法。谱聚类算法建立在谱图理论基础上，与 K 均值、EM 算法等传统的聚类算法相比，它具有能在任意形状的样本空间上聚类且收敛于全局最优解的优点<sup>[23]</sup>。

具体算法步骤如下：

- **Step1:** 输入待聚类的样本集、相似度矩阵 S 和聚类数目 K；
- **Step2:** 利用输入的相似矩阵 S 构建邻接矩阵 W 和度矩阵 D；
- **Step3:** 计算出拉普拉斯矩阵 L，并进行标准化操作： $D^{-\frac{1}{2}}LD^{\frac{1}{2}}$ ；
- **Step4:** 计算  $D^{-\frac{1}{2}}LD^{\frac{1}{2}}$  的特征值和特征向量构建特征向量空间；
- **Step5:** 将特征向量组成的矩阵按行标准化，得到特征矩阵；
- **Step6:** 用传统 k-均值聚类算法(k-Means)进行聚类；
- **Step7:** 输出聚类结果。

聚类结果中每一类即代表着同一个问题的留言，因此可以直接将聚类结果用于问题热度评价。

## 5.2 热度评价体系的建立

在这部分我们首先定义了以下三个指标从不同侧面来衡量问题的热度：问题的被关注度、问题

涉及人群的广度和问题的被认同度；接着给出了各个指标量化评价的方法；最后结合层次分析法和 TOPSIS 法计算出每个问题的热度指数，并由此选出热度指数前 5 的问题，得到“热点问题表.xls”和“热点问题留言明细表.xls”。

5.2.1 热度评价指标的选取与量化

我们认为若一个问题被提及的次数越多、被越多人关注与认同，则说明这个问题的热度越高。因此，我们选取了以下三个指标来衡量问题的热度：问题的被关注度、问题涉及人群的广度和问题的被认同度。下面介绍三个指标的量化方法。

- (1) 问题的被关注度 = 该问题的留言次数之
- (2) 问题涉及人群的广度 = 该问题的留言人数

由模型假设可知，不同留言用户编号代表不同用户，因此我们统计出该问题的留言中共有多少个不同的留言用户编号，并将该数值作为该问题的留言人数。

- (3) 问题的被认同度 = 该问题所有留言的点赞数 - 该问题所有留言的反对数

5.2.2 热度指数的计算

在这部分我们将层次分析法与 TOPSIS 评价法进行结合，建立 AHP-TOPSIS 评价模型，并以此来综合评价每类问题的热度，即计算每类问题的热度指数。AHP-TOPSIS 模型建立过程如下。

1、层次分析法确定指标权重

层次分析法完全依靠主观评价确定因素的权重，在决策过程中既充分吸收了定性分析的结果，又体现了定量分析的优势。且其所需数据量少，运算时间短，具有很强的科学性和条理性。我们运用层次分析法确定指标权重的步骤如下：

- Step1: 依据模糊数学理论方法得出的比较标准<sup>[25]</sup>（如表 11 所示）以及我们自己的主观理解建立重要程度比较矩阵。

标准值	定义	说明
1	同样重要	因素 $X_i$ 与 $X_j$ 的重要性相同
3	稍微重要	因素 $X_i$ 的重要性稍微高于 $X_j$
5	明显重要	因素 $X_i$ 的重要性明显高于 $X_j$
7	强烈重要	因素 $X_i$ 的重要性强烈高于 $X_j$
9	绝对重要	因素 $X_i$ 的重要性绝对高于 $X_j$

表 11 AHP 比较标准

注：2，4，6，8 分别表示上手两相邻标准判断的中值。

我们认为相比于点赞认同，一个人去撰写留言更能体现这个问题对他产生了严重影响；而相比于一个人多次留言同一问题来说，多人留言同一问题更能体现问题的热度；因此我们将重要程度比较矩阵定义如下：

	被关注度	人群广度	被赞同度
被关注度	1	1/3	3
人群广度	3	1	5
被赞同度	1/3	1/5	1

表 12 AHP 重要程度比较矩阵

- Step2: 进行一致性检验；利用 MATLAB 计算得  $C_R = 0.0332 < 0.1$ , 通过一致性检验。
- Step3: 计算各个指标权重，结果如表 13 所示。

	被关注度	人群广度	被赞同度
权重	0.3237	0.4507	0.2256

表 13 热度指标权重

## 2、综合利用 TOPSIS 法和权重计算热度指数

TOPSIS 法（Technique for Order Preference by Similarity to Ideal Solution）又称逼近理想解排序法，是一种常用的综合评价方法，其能充分利用原始数据的信息精确地反映各评价方案之间的差距。具体操作步骤如下：

- Step1: 构建决策矩阵  $A = (a_{ij})_{m \times n}$ ，其中  $a_{ij}$  为第  $i$  条留言第  $j$  个指标的数据， $m$  为总留言数， $n$  为指标数量。
- Step2: 构建标准化决策矩阵；

为了消除由不同指标具有不同的量纲产生的不可公度性，我们对指标进行无量纲化处理。在本问题中三个指标均为效益性指标，其无量纲化的计算公式如下：

$$b_{ij} = \frac{a_{ij} - \min_j a_{ij}}{\max_j a_{ij} - \min_j a_{ij}} \quad (12)$$

由此得到标准化决策矩阵  $B = (b_{ij})_{m \times n}$ 。

- Step3: 利用层次分析法得到的权重建立加权标准化决策矩阵

$$C = (c_{ij})_{m \times n} = \begin{bmatrix} w_1 b_{11} & w_2 b_{12} & w_3 b_{13} \\ \vdots & \vdots & \vdots \\ w_1 b_{n1} & w_2 b_{n2} & w_3 b_{n3} \end{bmatrix} \quad (13)$$

- Step4: 计算正理想解( $c^+$ )与负理想解( $c^-$ )以及各个问题与正、负理想解的距离( $d_i^+, d_i^-$ )，

计算公式如下:

$$c_j^+ = \{ \max_i c_{ij} \}, c_j^- = \{ \min_i c_{ij} \}, j \in J_1, J_1 \text{ 为效益性指标集}$$

$$d_i^+ = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^+)^2}, d_i^- = \sqrt{\sum_{j=1}^n (c_{ij} - c_j^-)^2}$$

➤ Step5: 各个热点问题贴近度计算, 即计算热度指数; 贴近度计算公式如下:

$$E_i^+ = \frac{d_i^-}{d_i^+ + d_i^-} \quad (14)$$

➤ Step6: 选出热力指数前 5 的问题, 即贴近度前五的问题。

## 5.3 结果

我们利用 python 和 MATLAB 实现问题二的各个步骤, 所有源代码请见附件。在进行相似问题聚类时, 我们发现主观设定的聚类数目在很大程度上会影响最终的聚类效果。因此我们进行了多次实验, 确定了每个地区聚类效果较优的聚类数目, 具体如表 14 所示。

地区	聚类数目	地区	聚类数目
西地省	平均 1.7 条留言一类	A4 区	平均 3 条留言一类
A 市	平均 10 条留言一类	A5 区	平均 2 条留言一类
其它市	平均 1.7 条留言一类	A6 区	平均 1.7 条留言一类
A1 区	平均 3 条留言一类	A7 区	平均 2 条留言一类
A2 区	平均 4 条留言一类	A8 区	平均 2.1 条留言一类
A3 区	平均 2 条留言一类	A9 区	平均 2.1 条留言一类

表 14 各地区较优聚类数目

由于无法确定最优的聚类数目, 所以在聚类结果中个别类别的聚类效果不好。因此我们综合考虑聚类效果后在每个地区选取热度指数排名前三的热点问题, 详情请见附件“问题二热点问题最终排名汇总.xls”。接着我们在选出来的 36 个热点问题里再运用一次 AHP-TOPSIS 模型进行热度评价, 选出热度指数前五的热点问题, 具体如表 15 所示。同时按照题目要求, 在附件中给出“热点问题表.xls”和“热点问题留言明细表.xls”。

热度排名	问题ID	热度指数	时间范围	地点/人群	问题描述
1	4	0.79	2019/04/10 至 2020/01/26	A 市 A2 区丽发新城小区	小区附近建搅拌站影响环境和扰民
2	28	0.49	2019/07/07 至 2019/09/01	A 市伊景园滨河苑小区	小区车位违规捆绑销售
3	19	0.48	2019/04/17 至 2019/09/06	A 市 A7 县地铁 3 号线松雅西地省站	反映地下通道建设问题
4	29	0.37	2019/01/25 至 2019/12/04	A 市居民	反映地铁手机扫码乘车与楚行一卡通的相关问题
5	7	0.33	2019/03/24 至 2019/03/29	A 市 A3 区郝家坪小学	咨询郝家坪小学何时扩建

表 15 热度排名前五的热点问题

## 六、答复意见的评价

在这部分我们将针对附件 4 相关部门的答复意见建立一些评价指标，并且对答复意见的质量给出一套评价方案。因此在这一部分我们将主要对以下两个内容进行阐述：一是评价指标的建立；二是综合评价方案的制订。

### 6.1 评价模型的建立

首先，我们为附件四中的中的答复意见建立一些评价指标，选取了以下 5 个指标来评价答复的不同侧面：相关性、完整性、可解释性、时效性以及可读性；其次，在这些指标的基础上，建立综合评价方案，并且给出每条答复意见的具体评价结果。

#### 6.1.1 评价指标的选取与量化

我们建立了 5 个评价指标从 5 个不同角度出发评价每条留言的答复意见：从答复意见与相应留言的关联程度出发，建立了相关性指标；从答复意见的结构完整程度出发，建立了完整性指标；从答复是否及时出发，建立时效性指标；从答复意见对于相应留言是否给予充分解释出发，建立可解释性指标；从答复意见的可理解性出发，建立了可读性指标。下面我们对这几个指标进行详细说明：

##### 1.相关性L

我们通过计算答复文本与留言文本的相似度来评价答复与留言之间相关性，因此我们首先选取答复与留言文本中的关键词，然后利用 Word2Vec 将词语转化成词向量，最后利用词与词之间的余弦相似度计算文本之间的相似度。具体步骤如下：

### （1）提取关键词

值得注意的是，选取关键词通常会运用 TF-IDF 法，但是 IDF 在本问题中并不适用，具体原因如下：由前文介绍可知，若在所有文本集中包含关键词  $W$  的文本集数量越少，则词  $W$  的 IDF 值越大；而在计算留言  $T_1$  与其回复  $T_2$  的相似度时，不需要考虑其他留言与回复，因此所有文本集应定义为  $\{T_1, T_2\}$ ；如此一来，若留言  $T_1$  中的某个词在其答复中也存在，则该词的 IDF 值会降低，这显然本问题的要求相矛盾。

因此我们通过利用词频统计与词性标注筛选出每条留言与答复词频前 10 的名词与动词，将这些词作为相应文本的关键词，并用于后续的文本相似度计算。

### （2）留言与回复相似度计算

参考黄承慧等人的方法<sup>[24]</sup>，留言与回复的相似度计算公式定义如下：

$$VectSim(T_i, T_j) = \frac{1}{2} \left[ \frac{1}{10} \sum_{k=1}^{10} \max_{1 \leq l \leq n} \{Sim(w_{ik}, w_{jl})\} + \frac{1}{10} \sum_{l=1}^{10} \max_{1 \leq k \leq m} \{Sim(w_{jl}, w_{ik})\} \right] \quad (15)$$

其中  $T_i$  为某条留言， $T_j$  为与之对应的回复，其他参数的具体解释见本文的 5.1.2 小节。

### （3）相关性指标的量化

由前文可知， $VectSim(T_i, T_j)$  的取值范围为  $[0, 1]$ ，因此为了将相关性指标定义为百分制，我们利用式(16)来量化留言与回复的相关性  $L$ 。

$$L = 100 \times VectSim(T_i, T_j) \quad (16)$$

## 2.完整性指标 $I$

完整性指标用于判断一条回复的完整性如何。下面是一般网络问政平台的回复格式：

<p style="text-align: center;">网络问政平台回复格式</p> <p>您好！</p> <p>您的来信咨询（投诉/建议）的信件已收悉，先将办理情况答复如下：</p> <p>（正文）</p> <p>以上是我单位对您来信的回复，如果您对我们的回复有什么不满意的地方或有新的意见建议，请致电 0000-00000000，感谢您对****事业的关心、理解和支持，祝您生活愉快。</p>
---



我们认为一个完整的答复应具有以下几个结构：表示问候、表示感谢、表示收到、提供进一步提问的渠道。

(1) 具体实施方法



图 8 完整性具体实施流程

(2) 选取结构词

其中我们选取了一系列结构词分别来代表各个结构：

问候：

你好	您好	.....
----	----	-------

感谢：

谢谢	感谢	感谢你	感谢您	.....
----	----	-----	-----	-------

收到：

已阅	知悉	收悉	收悉	.....
----	----	----	----	-------

提供进一步询问渠道：

致电	电话	投诉电话	咨询电话	.....
----	----	------	------	-------

(2) 结构词数量确定

在每条答复意见找代表各个结构的结构词，若有一个结构中的至少一个结构词，则代表该回复有相应结构。

具体方法如下：

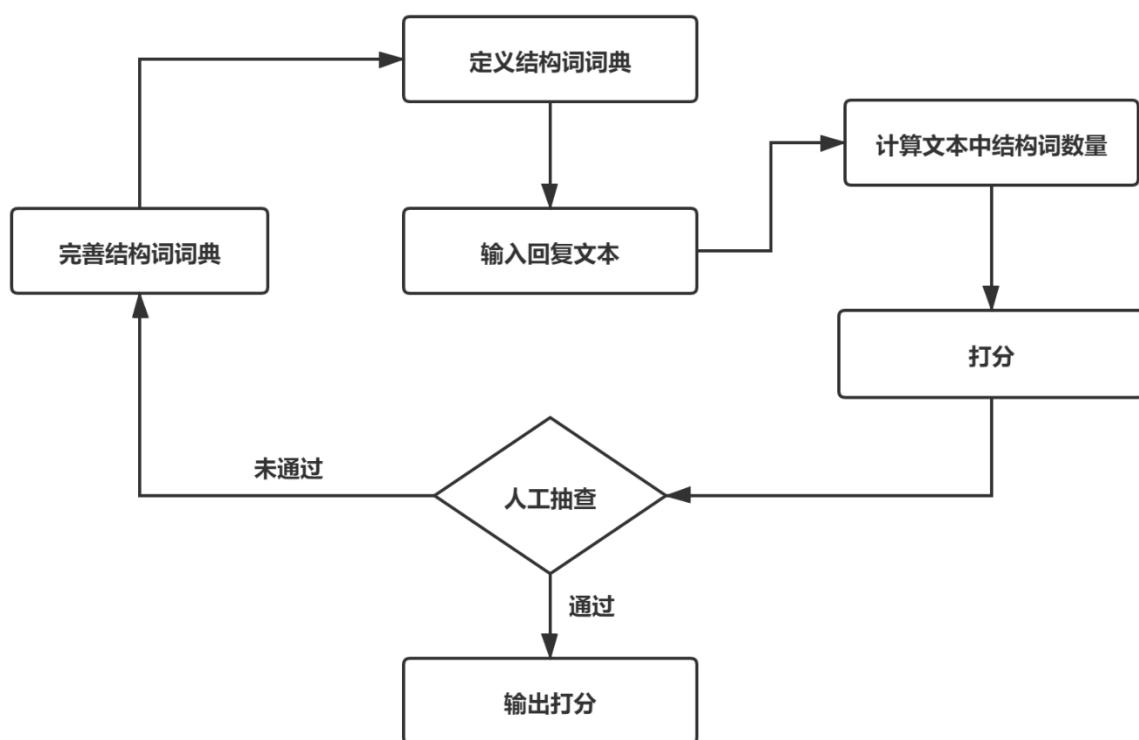


图 9 算法流程图

### （3）完整性指标的量化

我们对具有不同数量结构的答复意见进行评分：

具有结构个数	评分
4	100
3	75
2	50
1	25
0	0

表 16 完整性评分表

我们将具有这四个结构的答复意见的完整性评为 100 分；将具有这四个结构中的三个结构的答复意见评为 75 分；将具有这四个结构中的两个结构的答复意见评为 50 分；将具有这四个结构中的一个结构的答复意见评为 25 分；将不具有这四个结构的答复意见评为 0 分。我们认为具有两个及两个以上结构的答复意见具有较好的完整性。

### （4）具体评价方法

在每条回复中找结构词，若出现一个结构的结构词则说明该回复具有相应结构，看一共出现了几个结构。例如：

【4004】

网友“UU0081054”您好！您的留言已收悉。现将有关情况回复如下：A市火车站作为省会窗口地区以及省内最大的人、车流集散地，流动的电的、走廊上喊客人员、流浪乞讨人员、夜间违规的士等乱象等或多或少存在。对此，A1区公安、民政、残联、交通、交警、工商、城管、食药监、市政、园林、环卫等各部门，开展了多部门联合执法集中整治行动，打击查处各种乱象。一年来，共治安拘留喊客人员521人；查封出售假烟门店8家次、查处违停20582台次；查扣摩的16843台次；查处黑车、网约车、违规运营大客车718台次；反复劝阻、劝离、送返流浪乞讨人员约1320人次；查处流动摊贩413人次，查处乱吐乱扔580人次，查处违停抄牌锁车352台次。下一步，各部门会加强整治力度，确保火车站周边各种乱象得到进一步治理。感谢您对我们工作的支持、理解与监督！

2018年12月6日

回复中有结构词“您好”，则该回复具有问候结构；回复中具有结构词“收悉”，则该回复具有收到结构；回复中具有结构词“感谢”，则该回复具有感谢结构；而这条答复意见中不具有提供进一步询问渠道的结构，因此该条答复意见共有三个结构，它在完整性这个指标的评分为75。

3.时效性指标T

时效性指标用于评判答复意见回复的及时性。我们收集了多个省份对于政府答复群众网上留言办理时限的规定，具体如表15所示：

省市	文件出台时间	文件名称	办理时限（一般）	办理时限（复杂）
山西	2009年3月25日	《山西省人民政府办公厅网民留言办理工作实施方案》	一个月	两个月
安徽	2009年4月11日	《人民网网友给省委书记留言办理工作暂行规定》	一个月	
天津	2009年7月13日	《关于人民网等网上留言办理工作的规定》	十五天	
辽宁	2010年7月14日	《关于网友在人民网等网络媒体向省委主要领导留言办理工作的暂行规定》	一个月	一个半月

		《中共上海市委办公厅关于		
上海	2011 年 5 月 7 日	做好人民网网友给市委领导	两个星期	
		留言回复办理工作的通知》		
		《网民留言办理工作实施办		
甘肃	2012 年 3 月	法(试行)》	五天	二十天

表 17 留言办理机制一览表

根据以上 6 个省市的答复办理时限我们认为答复时间在距留言时间的一个月以内时较为及时。

#### (1) 时效性指标的量化

用 Excel 中的 DAYS 函数分别计算答复时间与留言时间之间相隔天数。

#### (2) 评价方法

下面我们给出具体评分方法：

答复时间区间	分数
[0,10]	100
(10,20]	90
(20,30]	80
(30,40]	70
(40,50]	60
(50,60]	50
(60,70]	40
(70,80]	30
(80,90]	20
(90,100]	10
else	0

表 18 时效性评分方法

### 4. 可读性指标R

可读性指标用于评判答复意见的可理解性，答复意见的可读性或读者的理解能力可以用自动化可读性指数ARI(Automated Readability Index)<sup>[26]</sup>来表示，在过去的信息科学研究中用可读性来定性地检验一些（英文）文本的特征，表示某文本内容吸引人的程度。

#### (1) 可读性指标量化方法

本文针对对象是中文文本，我们对公式进行一些改变使其使用于中文文本：

$$ARI = 4.71 \times \left( \frac{\text{总字数}}{\text{总词数}} \right) + 0.5 \times \left( \frac{\text{总词数}}{\text{总句数}} \right) \quad (17)$$

其数值近似等于我们可能理解一段文字的最低程度。

## (2) 具体实施方法

1. 用 Excel 中的 LEN 函数计算出每条答复意见中的字符数  $a_1$ ;
2. 用 Excel 中的 LEN 函数计算出删去表示一句话结束的标点 “，”、“。”、“？”、“！”、“；” 后的字符数  $a_2$ ，其中删去操作是用的 Excel 中的替代功能， $a_1 - a_2$  即得到每条答复意见的句数;
3. 用 Excel 中的 LEN 函数计算出删去所有标点的字符数  $a_3$ ——即为总字数;
4. 用 Python 进行分词并且用 MATLAB 将每条答复意见的词语进行计数，计数为  $b$ ;
5. 则(17)式变为:

$$ARI = 4.71 \times \left( \frac{a_3}{b} \right) + 0.5 \times \left( \frac{b}{a_1 - a_2} \right) \quad (18)$$

## (3) 评价方法

我们定义如下评分方法:

ARI 排名	分数
1	100
1-前 10%	90
前 10%-前 20%	80
前 20%-前 30%	70
前 30%-前 40%	60
前 40%-前 50%	50
前 50%-前 60%	40
前 60%-前 70%	30
前 70%-前 80%	20
前 80%-前 90%	10
前 90%-前 100%	0

表 19 时效性评分表

## 5. 可解释性指标 $E$

可解释性指标用于评判答复意见是否有据可依或者是否经过一定的调查研究做出的答复。我们通过判断每个答复意见是否出现文件型和调查研究型相关关键词来判断该答复意见的可解释性，文件型关键词用于判断答复意见是否有据可依，调查研究型关键词用于判断答复意见是否经过一定的

调查研究。

(1) 具体实施方法

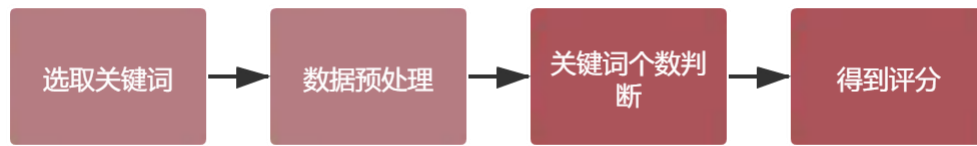


图 10 可解释性具体实施流程

(2) 关键词的选择

我们通过人工检索在附件四的答复意见中寻找关键词，关键词分为两部分：

文件型关键词：

实施意见	实施办法	实施方案	工作方案	实施条例
实施意见	管理办法	相关要求	文件	规划
通知	细则	规程	标准	.....

调查研究型关键词：

据查	核实	经审查	查勘	严查
查勘	考察	核查	核实	据调查
勘察	调研	勘测	经了解	.....

(3) 关键词数量确定

在每条答复意见中找关键词并且进行计数。具体算法如下：

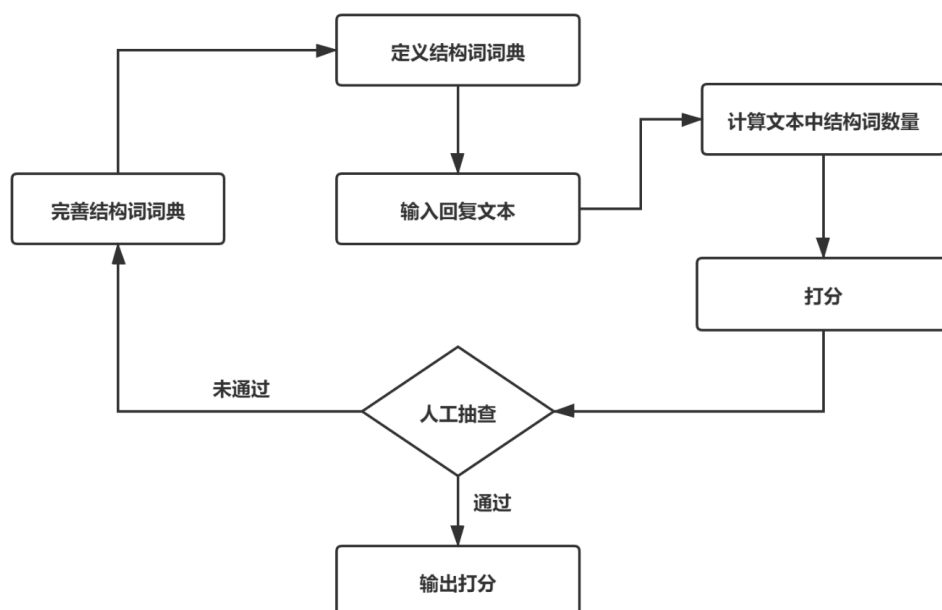


图 11 具体算法流程

（4）评价方法

根据经验，我们认为出现关键词的个数与答复意见的长度有着一定关系——答复意见的文本越长，越有可能出现更多的关键词。因此我们定义以下评分方法：

$$C_i = \frac{c_i}{a_i} \tag{19}$$

$c_i$ :第*i*条答复意见中出现的关键词的个数；

$a_i$ : 第*i*条答复意见中的总字数；

$C_i$ 排名	分数
1	100
1-前 10%	90
前 10%-前 20%	80
前 20%-前 30%	70
前 30%-前 40%	60
前 40%-前 50%	50
前 50%-前 60%	40
前 60%-前 70%	30
前 70%-前 80%	20
前 80%-前 90%	10
前 90%-前 100%	0

表 20 时效性评分表

6.1.2 基于群众情感分析的答复意见评价模型

由于答复意见的答复对象是留言的群众，因此我们综合群众情感和答复意见的质量做出综合评价。其中答复意见质量指的是前文用于评价答复意见的指标：相关性、完整性、时效性、可读性和可解释性。而留言群众情感则是我们用文本情感分析来得到的每条留言的情感值。

1.文本情感M分析

我们通过 NLPPIR 大数据语义智能分析平台对每条留言进行情感分析，得到每条文本的正向情感值和负向情感值，我们将二者直接进行相加，得到我们每条留言的情感值。

2.综合评价方法

（1）层次分析法确定权重

层次分析法详细介绍请见本文 5.2.2 小节，该问题各指标的重要程度比较矩阵图表 21 所示。

	相关性	完整性	时效性	可解释性	可读性
相关性	1	7	3	5	9
完整性	1/7	1	1/7	1/3	3
时效性	1/3	7	1	3	9
可解释性	1/5	3	1/3	1	7
可读性	1/9	1/3	1/9	1/7	1

表 21 AHP 重要程度比较矩阵

经计算， $C_R = 0.0684 < 0.1$ ，通过一致性检验，具体权重结果如表 22 所示。

权重指标	权重
相关性 $l$	0.50
完整性 $g$	0.06
时效性 $t$	0.28
可解释性 $e$	0.13
可读性 $r$	0.03

表 22 各项评价评价指标权重

### (2) 综合评价

根据经验，当留言群众在进行留言时情感倾向为负面，若此时相应的答复意见质量差，则对于留言群众而言，答复意见的质量会在原来质量的基础上降低；相反地，当留言群众进行留言时情感倾向为正面，若此时相应的答复意见质量好，则对于留言群众而言，答复意见的质量会在原来质量的基础上上升。

因此，我们将情感分析后得到的留言群众的情感值纳入综合评价中：若第*i*条基本评价分数 $s_i$ 低于所有答复意见基本评价分数平均值 $\bar{s}$ ，则考虑相应留言的情感值，若情感值也低于所有留言的情感值，则在原来分数的基础上打“折扣”；相反地，若第*i*条基本评价分数 $s_i$ 不低低于所有答复意见基本评价分数平均值 $\bar{s}$ ，则考虑相应留言的情感值，若情感值也不低于所有留言的情感值，则在原来分数的基础上给予“加分”

$$s_i = (lL_i + gG_i + tT_i + eE_i + rR_i)$$

若 $s_i \geq \bar{s}$ ,

$$S = \begin{cases} \left( \frac{m_i - \bar{m}}{M_1} + 1 \right) s_i, & m_i - \bar{m} \geq 0 \\ s_i, & m_i - \bar{m} < 0 \end{cases} \quad (20)$$

若 $s_i < \bar{s}$ ,



$$S = \begin{cases} s_i, & m_i - \bar{m} \geq 0 \\ \left( \frac{m_i - \bar{m}}{M_2} + 1 \right) s_i, & m_i - \bar{m} < 0 \end{cases} \quad (21)$$

$s_i$ 表示基本评价分数； $\bar{s}$ 表示所有 $s_i$ 的平均值； $S$ 表示综合评分； $m_i$ 表示第 $i$ 条留言的情感值； $\bar{m}$ 表示所有留言的情感值的平均值； $L_i$ 表示第 $i$ 条留言的相似度评分； $G_i$ 表示第 $i$ 条留言的完整性评分； $T_i$ 表示第 $i$ 条留言的时效性评分； $E_i$ 表示第 $i$ 条留言的可解释性评分； $R_i$ 表示第 $i$ 条留言的可读性评分；

$$M_1 = m_{\max} - \bar{m}$$

$$M_2 = \bar{m} - m_{\min}$$

## 6.2 结果

### 6.2.1 评价指标

#### 1.相关性

经过计算所有答复意见的相关性平均评分为 48.6，可以看出相关性处于中等水平，仍有很大提升空间。

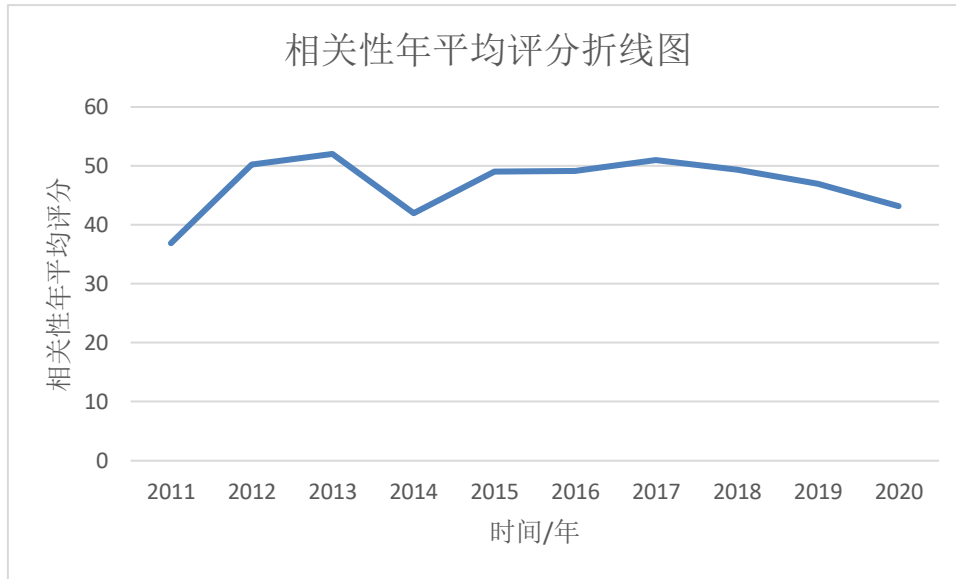


图 12 相关性年平均评分折线图

由图 12 可以看出，相关性年平均评分处于 40-50 左右，从 2011 年到 2019 年相关性年平均评分波动较小，且近几年有下降趋势，相关部门应注意提高答复的相关性。

#### 2.完整性

经过计算所有答复意见的完整性平均评分为 57.1，可以看出每条答复意见平均具有两个以上的结构，完整性处于较高水平，但仍可以在完整性方面做更多的改进，以此来树立良好的政府形象。

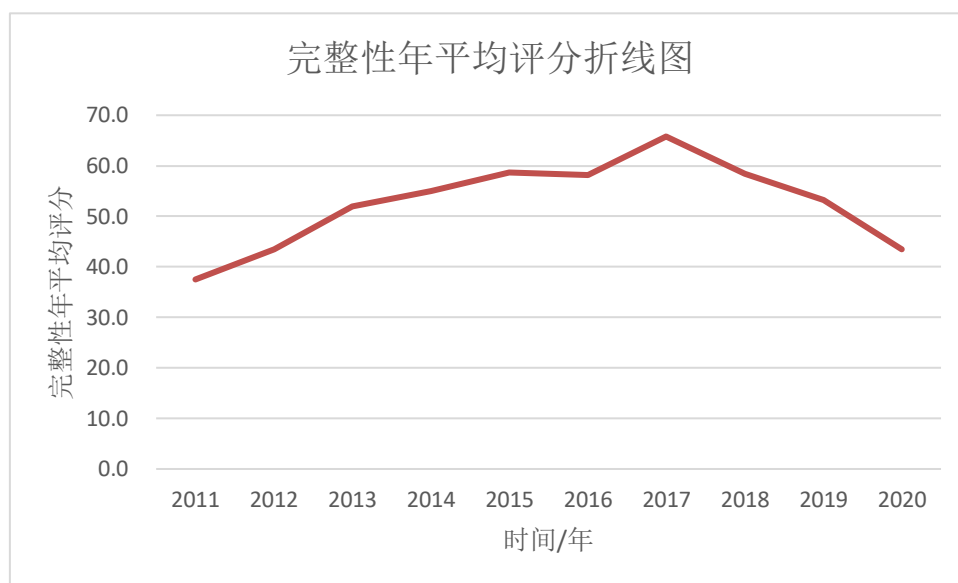


图 13 完整性年平均评分折线图

由图 13 可以看出，2011-2017 的完整性平均评分持续上升，在 2017 年做的最好。但是 2017 年-2020 年有明显的下降趋势，相关政府部门需多注意；此外，2011 年的完整性年平均评分虽然最低，但我们认为原因可能是 2011 年的数据较少，从而造成了误差。

结果分析：

得分	占比
100	8.4%
75	41.5%
50	26.5%
25	17.4%
0	6.2%

表 23 完整性结果

网络增大了信息的开放性和透明度，人人都可以通过网络了解到政府的一举一动，随时随地对政府行为进行监督<sup>[27]</sup>。基于这样的舆论监督下，党政机关应该注意话语表达方式，而答复意见的完整性也体现着政府的亲和度，因此我们需要注意答复意见的完整性。由上述表中结果，我们可以看出超过 75% 的答复意见的都具有两个及两个以上的结构，整体完整性较好，但仍还有少部分的答复意见不具有四个结构中的任意一项。

### 3.时效性

结果分析：

答复时间-留言时间 (天)	占比
3个工作日内回复	18.3%
10个工作日内回复	47.6%
20个工作日内回复	71.8%
30个工作日内回复	83.8%
超过60个工作日	5.5%

表 24 时效性部分结果

政府的迅速回应，不能只体现在对重大社会事件处理上，还应当表现在对事关群众切身利益的一些小事上。对于小事、贴近民众生活问题的回应必须做到及时迅速，这样才能获得民众的认可<sup>[28]</sup>。从以上表中结果可见，超过 80%的答复意见都能在一个月内做出回复；不仅如此，还有将近一半的答复意见在 10 个工作日内做出回复，并且还有相当一部分的答复意见能在留言后的 3 个工作日内做出回复。由此可以看出，当下政府的回应速度较快，但是仍存在小部分答复意见超过 2 个月的时间才对留言进行回复。

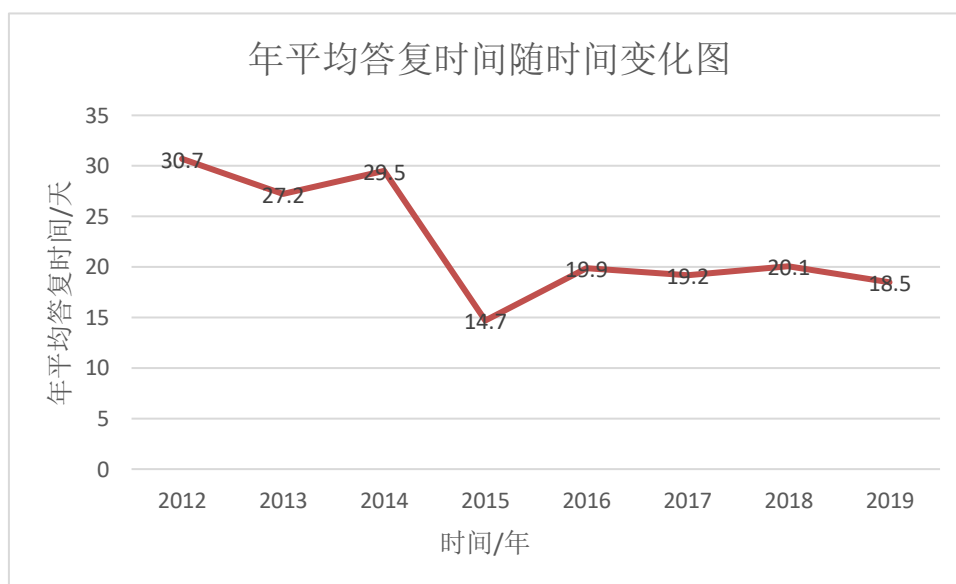


图 14 年平均答复时间随时间变化图

由图 14 可知，年平均答复实际总体呈下降趋势，这说明政府的答复速度越来越快，办事效率越来越高，值得肯定。

#### 4.可读性

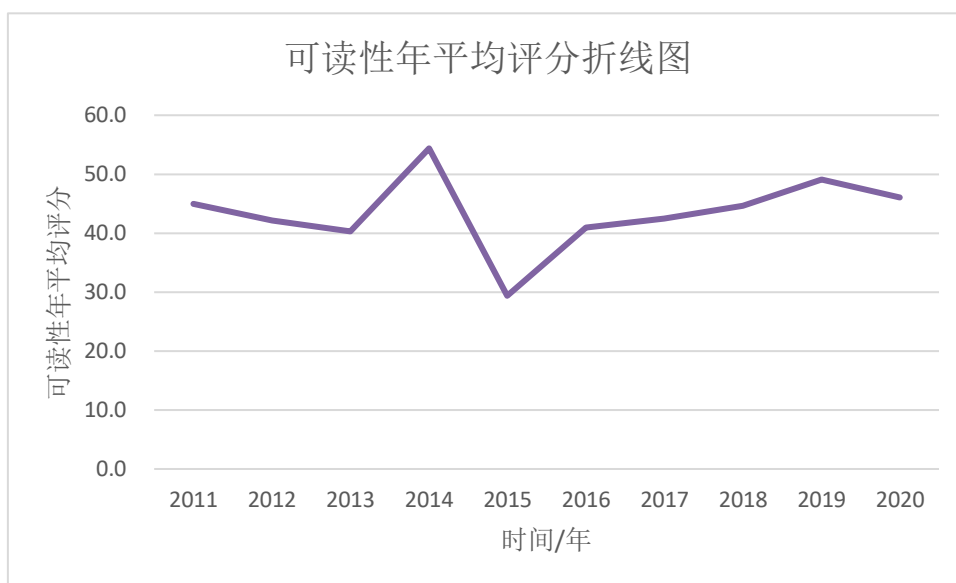


图 15 可读性年平均评分折线图

由图 15 可知，2011-2020 年的可读性年平均评分在小范围内波动，说明答复意见在语言的“通俗易懂”方面长年保持的较好。

#### 5.可解释性

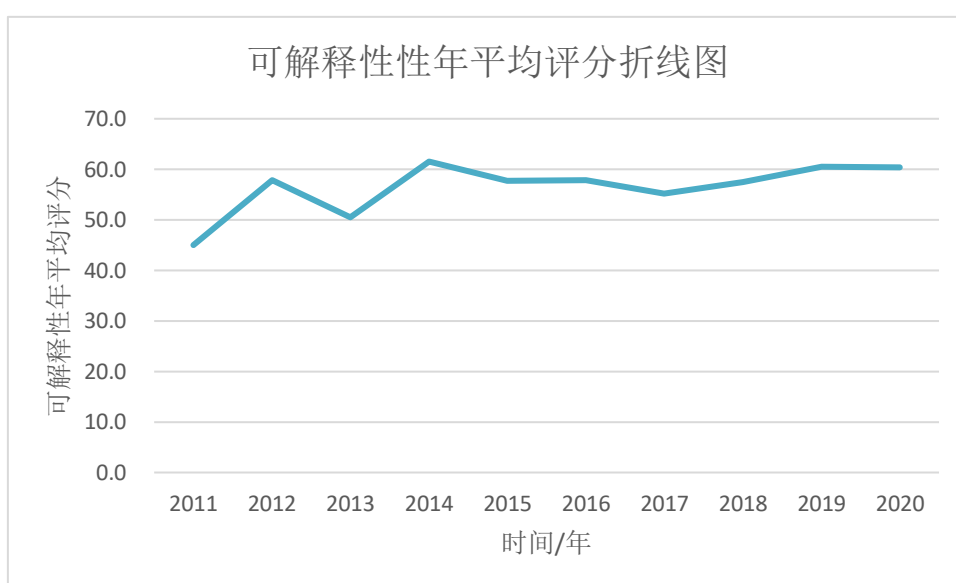


图 16 可解释性年平均评分折线图

由图 16 可见，2011-2020 年的可解释性年平均评分呈缓慢上升的趋势，说明政府在为群众进行答复时越来越注意答复意见是否有据可依以及在答复前是否有进行一定实际调查，但仍然还有上升空间。

## 6.2.2 基于群众情感分析的答复意见综合评价

下面我们选取两个例子来展示基于群众情感分析的答复意见综合评价结果：

【82186】综合评分：29.3

留言时间：2019/5/15

我是全体 G 市鸿正苑小区正义业主代表，我们已经维权快 2 个月了。之前 2015 年也到市政府上访几次都有记录的，一直没得到上级领导回复，而且我们的组织者被开发商请的人打伤住院，还警告我们谁牵头就打谁。现在开发商强制我们业主买他们的车位，已经变向换了几个花样的公告逼我们了，从事发至今这两个月地下车位一个都没有卖出去，全部是空车位，恳请上级有关部门能帮帮我们。

答复时间：2019/12/3

高山街鸿正苑小区首届业主代表大会于 9 月 6 日顺利召开，在住建局、街道和社区的细心指导下历时四个月成立了第一届业主委员会！在业委会和新物业的共同努力下，小区有了很多变化。业委会成员与新物业跟开发商针对鸿正苑地下停车场收费问题进行了协商，开发商现在将该停车场收费权交给新物业，收费标准为 260/月，可以月租也可以年租，但仅对鸿正苑业主内部，外来车辆只可以临时停车，临停 3 小时以内收费 5 元，这样的收费标准得到了大家的一致认可！

根据我们的评价指标，此答复意见的完整性差，没有出现四个结构中的任何一个结构；时效性差，政府部门在留言后 5 个多月才做出答复，没能给留言群众及时的答复；可解释性差，没有对留言中所提情况进行调查，也未用文件对留言情况进行解释；再加上我们对留言进行情感分析得到，留言的情感倾向为负向，因此综合评价评分低。

【11149】综合评分：82.9

留言时间：2017/3/22

高铁南站是 A 市对外窗口，但东广场公交站一直无遮雨棚，A 市春季多雨，夏季高温，希望有关部门尽快建好公交站遮雨棚，打通服务群众最后一公里问题

答复时间：2017/4/14

网友“UU0081719” 您好！您的留言已收悉。现将有关情况回复如下： 据查，您提到的楚江北路加油站位于鹅秀片控制性详细规划的范围内，控规编制的成果已依法进行了公示，并上报市政府批准同意。该加油站的选址及位置符合控制性详细规划要求；该加油站与周边建筑的安全防护距离满足国家规范 GB50156-2014《汽车加油·加气站设计与施工规范》中的规定，并且已征得消防、环保部门的同意。因此，该项目的建设符合法律法规及规划审批程序。此外，金侠经济开发区正在制定楚江北路沿江风光带的景观规划，楚江北路加油站建设单位已提供多个外立面方案给金侠经济开

发区审核，楚江北路加油站外立面将严格按照楚江北路沿江风光带的景观规划的要求实施。 感谢  
感谢您对我们工作的关心、监督与支持！A市城乡规划局 2015年9月1日

根据我们的评价指标，此答复意见的完整性强，具有四个结构中的其中三个结构；时效性好，政府部门在留言后一个月内做出答复，给留言群众及时的答复；可解释性好，对留言中所提情况进行调查，也用相关文件的规定对留言情况进行解释；再加上我们对留言进行情感分析得到，留言的情感倾向为正向，因此综合评价评分较高。

所有答复意见的具体结果见“问题三全部指标和综合评分结果.xlsx”。

## 七、模型评价

### 7.1 模型的优点

- **群众留言分类模型：**该模型将注意力机制与卷积神经网络相结合，同时还设置了多个卷积层提取留言的特征，从而使该模型在具有 90% 以上的分类准确率，且  $F_1$  值高达 0.89。此外，该模型学习速度极快，在题目所给的数据量下训练 3 分钟验证集即具有 90% 的分类准确率，如表 17 所示。因此我们所构建的模型可以轻松完成海量数据的训练任务，同时能保证验证集分类准确率。

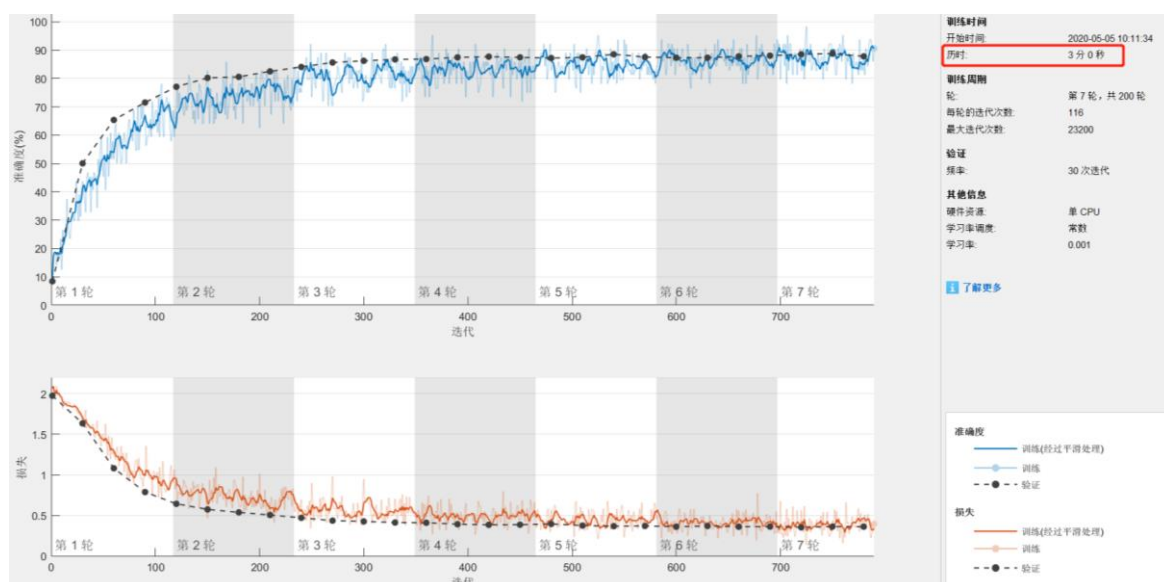


图 17 留言分类模型学习能力展示

- **热点问题挖掘模型：**在该模型中，我们利用了多种有效方法在不同方面进行降维，即能够大大降低最后文本相似度计算的复杂性，同时还能提高热点问题聚类的准确率。例如我们运用 TF-IDF 与词性对关键词进行筛选、降维；利用正则表达式对问题所属行政单位进行划分等等。此外，该模型中提出的文本相似度计算方法是在利用 Word2vec 训练词向量与计算词语余弦相似度的

基础上，利用词语的词性信息和 TF-IDF 值筛选关键词并进行加权处理，将文本表层信息与文本语义信息相结合，使得文本相似度计算结果更加合理准确。而且命名实体识别和新词发现的使用有利于使得同一地区的问题之间获得更高的相似度，有利于热点问题挖掘。

- **答复意见评价模型：**该模型的亮点是从两个不同的出发点去评价答复意见的质量，一是从答复意见自身质量出发，如该模型中所建立的相关性、完整性、可解释性、可读性和时效性指标；二是从群众的体验感出发，对群众留言进行情感分析然后结合群众的情绪对答复意见进行评价。

## 7.2 模型的缺点

- **群众留言分类模型：**该模型有多个参数需要调节，但最优参数的确定没有一套确切的可以参照的方法，因此在寻找模型最优参数组合时需要花费一定的时间。此外因为时间限制，我们没有寻找其他语料进行实验，因此不确定目前所选出的最优参数组合是否适用于其他语料。
- **热点问题挖掘模型：**该模型依据问题所属行政单位对留言进行划分，而此次题目所给数据对行政单位进行了处理，均用字母和数字以及“西地省”代替，这使得我们的划分存在些许误差。因为某些留言内容中出现表达其他意思的字母和数字也被我们识别为行政单位。但若在实际情形中，利用具体的行政单位名字进行划分则可以在很大程度上减小该误差。
- **答复意见评价模型：**该评价模型考虑的不够全面。一是在评价答复的时效性时，我们没有考虑其对应的留言问题的复杂性。即不管群众所留言的问题是复杂还是简单，我们都用同样的标准衡量答复的时效性。二是没有考虑答复意见中出现错别字的情形。三是对于留言错板块的留言，例如：

【18126】

【留言】我们是 I 市汽车北站进站口的周围居民。在这里的马良社区马园口组，有一栋六层近一万平方米的新大楼。5 月份，该楼业主和有人合作，拟将建设私营的 I 市平康肾病医院。目前，该医院的负责人说会出钱找“黑关系”，保证环保达标。对于周围居民的意见，他置之不理。一旦医院建成，它的设备辐射、废弃药物、尸体、病毒传染等将严重危害周围居民健康，造成巨大的环境污染。为此，我们强烈请求你们督促西地省、I 市、I1 区的环保部门，尽快阻止这医院的建成，为居民造福！

【答复】您的留言已收悉。关于您反映的问题，已转 I1 区委、区人民政府调查处理。

我们没有将此类答复意见单独挑出来评价，而是将此类的答复意见与其他答复意见一起进行评价。

## 八、参考文献

- [1] 胡军伟,秦奕青,张伟.正则表达式在 Web 信息抽取中的应用[J].北京信息科技大学学报(自然科学版),2011,26(06):86-89.
- [2] <https://stanfordnlp.github.io/CoreNLP/>
- [3] <http://www.nlpir.org/wordpress/>
- [4] 朱梦. 基于机器学习的中文文本分类算法的研究与实现[D].北京邮电大学,2019.
- [5] <https://github.com/goto456/stopwords>
- [6] 成昊. 基于 Word2Vec 的中文问句检索技术研究及系统实现[D].哈尔滨工业大学,2016.
- [7] <https://github.com/Embedding/Chinese-Word-Vectors>
- [8] 段尧清,姚兰.政媒融合问政平台非正式文本自动分类匹配研究[J/OL].情报理论与实践:1-9[2020-05-07].<http://kns.cnki.net/kcms/detail/11.1762.G3.20200109.1229.010.html>.
- [9] 陆旭. 文本挖掘中若干关键问题研究[M]. 中国科学技术大学出版社, 2008.
- [10] 汪岩,刘柏嵩.文本分类研究综述[J].数据通信,2019(03):37-47.
- [11] 梁思程. 基于深度学习的文本表示与分类研究[D].西安工程大学,2019.
- [12] 张晓敏. 深度学习在文本分类中的研究与应用[D].电子科技大学,2019.
- [13] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [14] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [15] 陶文静. 基于卷积神经网络的新闻文本分类研究[D].北京交通大学,2019.
- [16] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. Computer Science, 2014, arXiv: 1409. 0473 .
- [17] 孙嘉琪,王晓晔,周晓雯.基于神经网络模型的文本分类研究综述[J].天津理工大学学报,2019,35(05):29-33.
- [18] 侯宇昆.卷积神经网络概述[J].中国新通信,2017,19(09):45.
- [19] 朱梦. 基于机器学习的中文文本分类算法的研究与实现[D].北京邮电大学,2019.
- [20] ]闫琰. 基于深度学习的文本表示与分类方法研究[D].北京科技大学,2016.
- [21] Cho K, Courville A, Bengio Y. Describing multimedia content using attention-based encoder-decoder networks[J]. IEEE Transactions on Multimedia, 2015, 17(11): 1875-1886.
- [22] Qian N. On the momentum term in gradient descent learning algorithms[J]. Neural networks, 1999, 12(1): 145-



151.

[23] 张吉文. 基于谱聚类的文本聚类算法研究[D].贵州大学,2015.

[24] 黄承慧,印鉴,侯昉.一种结合词项语义信息和 TF-IDF 方法的文本相似度量方法[J].计算机学报,2011,34(05):856-864.

[25] 黄贯虹, 方刚. 系统工程方法与应用[M]. 广州: 暨南大学出版社, 2005: 45-48.

[26] Hu N, Bose I, Koh N S, et al. Manipulation of online reviews: An analysis of ratings, readability, and sentiments[J]. Decision support systems, 2012, 52(3): 674-684.

[27] 周丽. 我国政务微博的现状、问题与对策研究[D].江西财经大学,2012.

[28] 朱仁显,樊山峰. “互联网+” 背景下政府回应问题研究[J].长白学刊,2017(06):42-48.