摘　要

本文旨在基于群众问政的留言记录和相关部门对群众留言的答复意见信息，通过自然语言处理和文本分析技术对留言记录进行分类，并筛选关键特征来挖掘其中热点问题。结合量化答复意见信息，从时效性、相关性、完整性、可解释性方面构建答复意见的质量评价体系。

针对任务一，基于“留言主题”和“留言详情”内容分别建立了多种分类模型，通过分类准确性和模型训练时间综合比较得出支持向量机（SVM) 在这两个指标下优于其他模型。由此采用支持向量机模型实现了对留言类别的分类预测。

针对任务二，构建了留言数量、支持数、留言长度、文本情感、留言时段、留言频率等多个热点问题评价指标，通过层次分析法对评价指标设定相应权重。随后对留言明细进行聚类，并根据聚类结果输出由热度排名、问题 ID、热度指数、时间范围、热点问题发生地点和人群、以及问题描述组成的热点问题表和热点问题留言明细表。

针对任务三，建立了从时效性、相关性、完整性、可解释性方面对答复意见内容的具体评价准则，并形成综合评价模型，给出了答复意见的综合评价得分。

关键词：自然语言处理；文本分类；支持向量机；层次分析法；聚类分析；评价模型

Abstract

The purpose of this paper is to classify the message records, filter the key features to uncover the hot issues in them by natural language processing and text analysis techniques, and construct a qualitative evaluation system of the responses in terms of timeliness, relevance, completeness and interpretability.

For Task 1, various classification models were built based on the content of ”Message Subject” and ”Message Details” respectively. Support vector machine (SVM) was superior to other models in terms of classification accuracy and model training time. The prediction of message categories was thus achieved by the support vector machine model.

For Task 2, the evaluation indexes of hot issues such as the number of messages, the number of supports, the length of messages, the text emotion, the period of messages, and the frequency of messages were constructed, and the corresponding weights were set to the indexes through the analytic hierarchy process. Then the messages are clustered, and the hot problem table and hot problem message details were output according to the clustering results.

For Task 3, specific evaluation criteria for the content of the responses in terms of timeliness, relevance, completeness and interpretability were established, and a comprehensive evaluation model was developed.

Key words: Natural language processing; Text classification; Support vector machine; Analytic hierarchy process; Cluster analysis; Evaluation model

1 问题分析

在当前这个电子技术高速发展的大数据时代，民众可以通过多种途径与相关政府部门进行沟通。在此情况下，以往的通过人工来对留言进行划分以及对热点问题进行整理等方式无法适应庞大的数据量。因此，我们可以借助自然语言处理技术及相关文本分析技术对相关文本数据进行系统性的划分和整理，从而提高平台处理效率。此外，我们也需要对相关部门的答复意见进行智能评价，从而提升民众与相关部门的沟通体验。

问题共给出四个附件数据。附件一提供了针对群众留言的内容分类三级标签体系。附件二的内容为群众留言，包括留言编号、留言用户、留言主题、留言时间、留言详情以及对应的一级标签。附件三的内容与附件二相似，但增加了对留言的反对数和点赞数信息。附件四的内容为群众留言信息及政府相关部门对应的答复意见。

本题共给出三个亟待解决的任务。任务一要求根据附件二提供的材料进行留言的一级分类，我们运用了多种模型进行分类，并选取了其中分类准确性最高的模型作为分类方法。任务二要求根据附件三提供的材料进行聚类，并定义合理的热度评价指标，根据热度评价指标进行热点问题挖掘与排序，最后根据聚类结果输出由热度排名、问题 ID、热度指数、时间范围、热点问题发生地点和人群、以及问题描述组成的热点问题表和热点问题留言明细表。我们采用 python 中的结巴分词对文本数据进行了处理并构造了词典以计算相关性从而进行聚类。任务三要求根据附件四提供的材料设计相关的评价指标并进行答复意见的评价。我们通过构建对答复意见的时效性、相关性、完整性、可解释性的具体评价标准，给出了答复意见的综合评价得分。

2 任务一：群众留言分类

2.1 数据处理

本任务是根据群众留言进行分类的问题，数据集中对分类有效的变量为“留言主题”，“留言详情”和“一级标签”，我们将这三个变量的数据提取出来进行分析。根据“留言主题”，“留言详情”内容，我们首先将其中对分类结果没有影响的地区信息进行剔除，以减少对分类的干扰。同时将“一级标签”转化为类别变量，以便于建模。然后，我们利用多种模型分别对“留言主题”和“留言详情”进行分类，并将数据集按 80% 和 20% 的比例随机分成训练集和测试集，通过分类指标 F-Score 来选择更为准确的模型。

由于本任务是文本分类问题，因此我们需要先将文本分词后去除没有实际意义或对分类无效的停用词，再利用词频-逆文本频率指数（TF-IDF）将分词转换为词向量，最后根据词向量进行分类。

2.2 建立模型

2.2.1 朴素贝叶斯（Naive Bayes）

朴素贝叶斯分类 [3] 是文本分类的经典方法，其假设每个特征之间是相互独立的，通过对每个类别在训练样本中出现的频率和每个特征划分对每个类别的条件概率，利用贝叶斯原理进行分类。

朴素贝叶斯的优缺点：

· 优点：

– 模型分类效率稳定，可以处理多分类问题。

– 对缺失数据不敏感，算法简单。

· 缺点：

– 模型假设特征之间相互独立，在实际中这个假设不一定成立。

– 先验概率的设定会影响分类结果准确性。

2.2.2 支持向量机（SVM）

支持向量机 [6] 基本模型是一种用于二分类的方法，基本原理是求解能够正确划分训练数据集并且几何间隔最大的分离超平面。根据基本原理，支持向量机模型可以扩展用于多分类的问题。

支持向量机的优缺点：

· 优点：

– 模型在小样本量的数据集上表现一般优于其他方法。

– 对非支持向量样本的删增不敏感，算法简单。

· 缺点：

– 模型在大样本量的数据集下需要大量的计算内存和时间消耗。

– 参数的调节会影响分类结果准确性。

2.2.3 逻辑回归（Logistic Regression）

逻辑回归 [7] 是一种广义的线性回归分析模型，适用于自变量是离散型的，是分类问题中常用的一种方法。

逻辑回归的优缺点：

· 优点：

– 模型参数表示每个特征对自变量的影响，可解释性强。

– 算法简单高效，可以在大样本量数据集上使用。

· 缺点：

– 模型在特征之间相关时分类效果会降低。

– 模型对极端值敏感，容易受其影响。

2.2.4 决策树（**Decision Tree**）

决策树 [5] 利用对特征的分割，将不同的类别样本进行分类，直至将所有样本分开，由此形成了以树结构形式表达的预测分析模型。

决策树的优缺点：

· 优点：

– 计算量较小, 且容易转化成分类规则，可解释性强。

– 样本特征可以是连续的也可以是离散的类型。

– 适合高维数据。

· 缺点：

– 模型容易过拟合。

– 模型忽略特征之间的相关性。

2.2.5 随机森林（**Random Forest**）

随机森林 [1] 是一个包含了多个决策树的分类方法，随机选取样本子集进行决策树的训练，并且根据内部决策树的分类结果选择其众数作为输出的类别。

随机森林的优缺点：

· 优点：

– 模型相对决策树更加稳定。

– 有良好的抗过拟合能力。

· 缺点：

– 模型相对于决策树计算成本更高。

– 对小样本数据和低维数据分类不一定准确。

2.2.6 XGBoost

XGBoost [2] 是一种基于树模型为基学习器的集成模型，利用最小化预测值和真实值之间的损失函数与树模型的复杂度之和，来避免对训练样本的过拟合，提高泛化能力。

XGBoost 的优缺点：

· 优点：

– 模型可以对特征进行并行处理，提高运算速度。

– 有良好的泛化能力。

– 模型对缺失值不敏感。

· 缺点：

– 数据量很大和特征较多时会占用大量内存，时间消耗大。

2.2.7 TextCNN

TextCNN [4] 是将卷积神经网络（CNN）应用到文本分类问题，利用多个不同大小的卷积核来提取句子中的关键信息，以捕捉句子中的局部相关性质。TextCNN 首先将文本句子矩阵通过卷积层和最大化池化层将不同长度的句子变成定长的表达，然后通过全连接的 softmax 层输出每个类别的概率，最后选取概率最高的类别作为预测值。

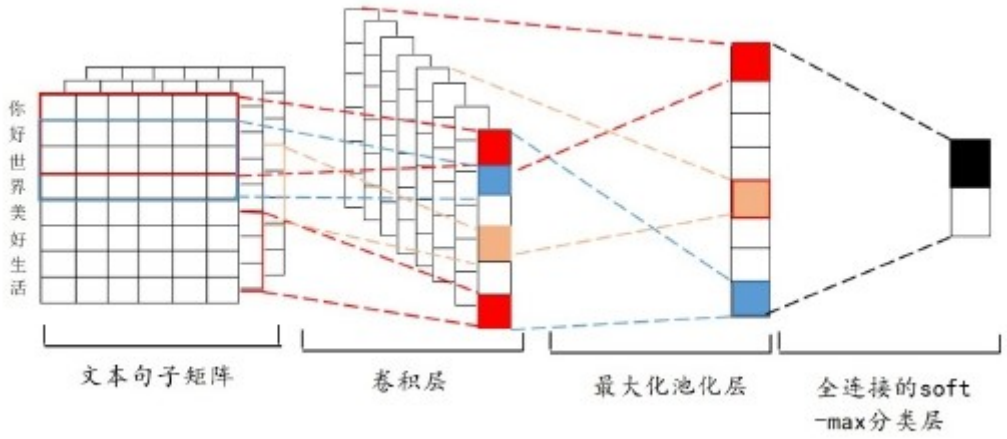

图 1: TextCNN 示例图

TextCNN 的优缺点：

· 优点：

– 模型通过卷积层可以自动进行特征提取。

– 可以处理高维数据。

· 缺点：

– 输入大样本量数据时会占用大量内存，训练时间一般较长。

– 池化层会丢失信息，忽略局部与整体的关联性。

2.3 分类结果比较

2.3.1 留言主题

我们先利用 TF-IDF 将“留言主题”中的文本数据转换为词向量后，使用训练集训练模型中提到的方法，然后根据测试集中不同方法得到的分类指标 F-Score 来比较其分类的准确性，以选择更为准确的分类方法。

表 1: 留言主题分类结果比较

	NB	SVM	LR	DT	RF	XGBoost	TextCNN
F-Score	0.72	0.84	0.85	0.75	0.76	0.74	0.81
运行时间（秒）	0.01	4.62	21.82	0.98	0.51	191.32	25,620

其中 NB：朴素贝叶斯，SVM：支持向量机，LR：逻辑回归，DT：决策树，RF：随机森林。

2.3.2 留言详情

我们同样先利用 TF-IDF 将“留言详情”中的文本数据转换为词向量后，使用训练集训练模型中提到的方法，然后根据测试集中不同方法得到的分类指标 F-Score 来比较其分类的准确性，以选择更为准确的分类方法。

表 2: 留言详情分类结果比较

	NB	SVM	LR	DT	RF	XGBoost	TextCNN
F-Score	0.78	0.89	0.89	0.73	0.78	0.85	0.76
运行时间（秒）	0.01	18.76	36.28	2.62	0.82	425.33	11,160

2.4 模型选择

从分类结果的比较中，我们可以得出利用留言详情的分类结果较留言主题更为准确，且支持向量机（SVM) 在留言详情中的各种方法中 F-Score 最高、运行时间较短。因此我们采用留言详情数据作为分类的输入数据，将支持向量机作为模型来对其进行分类。

3 任务二：热点问题挖掘

3.1 数据处理

3.1.1 文本处理

若文本中存在停用词，所有文本向量的相似度将会被提高, 从而影响后续的相似度计算阈值的确定，因此我们借助了一个包含 2749 个中英文停用词的文档进行了对于文本中停用词的去除。

另外，在对附件进行读取的时候，我们发现读取的文本中包含大量的转义字符，因此对其进行了去除。

3.1.2 日期处理

用 python 对附件三中的日期数据进行遍历后，我们发现日期数据共有两种不同格式:datetime 及 str。我们将所有日期格式统一转换为 datetime 格式，以便进行后续的统计操作。

3.2 问题聚类

我们利用结巴分词对去停用词和转义字符后的文本数据进行切词，然后用 gensim 包中的 corpora.Dictionary 生成了文本数据的词典，并将文档转化为向量形式。为了聚类，我们计算出每一条留言之间的相似度，并认为相似度大于 50% 的留言为同一类。在代码中，此 50% 阈值可根据实际认知进行相应的更改。

另外，我们同时也使用了切词后的文本数据进行词频统计，并将词频转化为权值（TF-IDF），最后使用 kmeans 算法进行聚类。我们将类别设置为 100 以进行对比。

3.3 问题聚类结果

```
np.unique(np.array(cluster_kmeans[:,0]))
array([ 1., 53.])
```

图 2: kmeans 聚类结果

使用 kmeans 聚类后，我们发现所有留言共被分为两类，这显然不符合我们的聚类要求，因此在后续分析中我们不再使用此算法的聚类结果。

使用向量余弦相似度进行计算后，留言共被分为 3621 类。

3.4 热点问题评价指标构造

基于任务二的需求，我们共构造了如下六个热点问题评价指标：

1）留言数量

问题的留言数量能够反映公众对其的关注程度，也能够在一定程度上反映该问题对社会造成的影响的深远程度，因此我们依据聚类结果统计了每一类问题的留言数量。

2）支持数

对每一条留言，赞成数和反对数能够体现公众对其的支持程度，因此我们计算出每一条留言的支持数为点赞数减去反对数，并根据聚类结果将一类问题中所有留言的支持数相加，最终得到这一类问题的支持数。

3）留言长度

留言长度可以在一定程度上反映留言者留言时的心态，我们认为留言长度越长代表留言者对该问题的关注度越高。因此，我们根据聚类结果将一类问题中的所有留言的长度取平均值，最终得到这一类问题的留言平均长度。

4）文本情感

一般而言，在一段留言中，留言者所表达的情感越消极，他/她所受到问题的影响就越大。因此我们使用 python 中的 SnowNLP 包对留言详情进行了中文文本情感分析。然后，我们根据聚类结果，将一类问题中所有留言的情感分析返回值的平均值作为该文本的情感值。

5）留言时段

心理学研究表明，人的心理情绪在一天中的波动变化是有明显模式的。一个人的情绪在早晨时最为低落，上午时较为煎熬，午后的心理程度慢慢减轻，黄昏时较为好转，晚上状态最好，睡前又起焦虑。而一个人在情绪较为低落时的留言将会带有更加消极的情绪，因此，为了对文本情感分析结果进行矫正，我们对留言时段进行了分类。我们将一天中的 24 小时共分成了 6 个区间：[00:00 - 5:59], [6:00 - 9:59], [10:00 - 13:59], [14:00 - 16:59], [17:00 - 18:59], [19:00 - 23:59]，每个区间的情绪程度值分别为 1，3，4，5，7，2，程度值越低表示情绪越低落。最后，我们根据聚类结果，返回一类问题中留言最多的时段的情绪程度值。

6）留言频率

在一段时间内，相同类型留言的出现次数可以反映该类问题的群众关注度和该问题的社会影响程度。因此，我们根据聚类结果将每一类问题中的留言按时间先后排序，计算出每条留言的间隔时间，最后对前三个最短的时间间隔取平均，作为该类问题的留言频率。

3.5 热点问题评价指标权重设定

我们随机抽取了二十名群众进行了对上述六个指标的重要性进行了调查，并采用了 1 10 的离散型评级标准。该问卷的详细内容在附录中给出。经调查，留言数量的重要性

平均得分为 7.50，留言支持数的重要性平均得分为 7.45，留言长度的重要性平均得分为 6.60，文本情感的重要性平均得分为 6.85，留言时段的重要性平均得分为 4.35，留言频率的重要性得分为 7.05。该结果能够较好地反映公众对上述六个指标重要性的普遍认知，也较为符合我队的初步设想。

3.6 热点问题热度指数计算

为进行合理的热度指数计算，我们采用了层次分析法进行决策。其中，目标层为确定热点问题，准则层为上述设计的六个指标，方案层为留言问题类别。由于在层次分析法中需要进行矩阵一致性的检验，且在一般情况下，矩阵阶数越大，出现一致性随机偏离的可能性也就越大，因此，我们根据调查结果，选择了留言数量位居前十的问题类别进行热度指数计算。

3.6.1 因素的权重设定与计算

对于指标因素层，我们依据调查结果及 Saaty 给出的 9 个重要性等级及其赋值，将留言数量、留言支持数、留言长度、文本情感、留言时段、留言频率的权重分别设定为 7，7，3，4，2，5。

对于留言类别因素，我们根据指标设定，计算出每个类别问题的对应指标值，将留言数量、留言支持数、留言长度、留言频率的权重设定设定为与返回值一致。对于文本情感，由于 SnowNLP 包的情感分析结果取值区间为 [0,1]，返回的情感值越接近 1 代表该文本整体感情越积极，情感值越接近 0 代表该文本整体感情越消极，所以我们返回的平均情感值小于 0.2 的问题的情感权重设定为 7，返回的平均情感值介于 0.2 至 0.4 之间的问题的情感权重设定为 6，返回的平均情感值介于 0.4 至 0.6 之间的问题的情感权重设定为 5，返回的平均情感值介于 0.6 至 0.8 之间的问题的情感权重设定为 4，返回的平均情感值介于 0.8 至 1 之间的问题的情感权重设定为 3。对于留言时段，我们根据心理研究结果，将 [00:00 - 5:59], [6:00 - 9:59], [10:00 - 13:59], [14:00 - 16:59], [17:00 - 18:59], [19:00 - 23:59] 这六个时间段的权重分别设定为 1，3，4，5，7，2。

3.6.2 层次分析法的实现

在层次分析法中，比较矩阵 A 需满足两个条件：$a_{ij} = 1/a_{ji}$, 其中 A 中元素 a_{ij} 表示要素 i 和要素 j 的重要性比较结果，以及通过一致性检验（检验系数 $CR < 0.1$)。在代码中，函数 comparison 会根据上述设定的权重构造比较矩阵，isConsist 函数会队矩阵进行一致性检验。当所有比较矩阵均通过一致性检验时，ComImp 函数依照层次分析法计算出留言数量前十的热点问题的综合重要性，即为我们设定的热点问题热度指数。

3.7 热点问题热度指数计算结果及排序

经计算后，我们得到这十类的热点问题热度指数，并按指数值大小排列如下：

table1

	热度排名	问题ID	热度指数	时间范围	地点/人群	问题描述
0	1	1	0.176024	2019/01/08至2019/08/12	A3区全晖优步花园	精装价格虚高，配套不落实
1	2	2	0.127348	2019/01/07至2019/07/17	A7县	恒基凯旋门万婴格林幼儿园办普惠园的咨询
2	3	3	0.127235	2019/07/07至2019/09/01	A市伊景园滨河苑	车位捆绑销售
3	4	4	0.115590	2019/06/28至2019/12/30	A市国王陵考古公园	公园建设刻不容缓
4	5	5	0.108485	2019/02/14至2019/09/09	A7县星沙凉塘路	旧城改造

图 3: 热点问题排序表

3.8 结果表格输出

3.8.1 时间范围

将数据中的日期格式统一后，我们提取出该每一类问题中最早的留言日期及最晚的留言日期，并以‘年/月/日’的格式输出，再在两个日期间添加文本‘至’，作为最终的时间范围输出文本。

3.8.2 地点提取与问题描述

经过对附件三数据的查看后，我们发现留言主题是留言内容的精简提要，并且包含了地点、人群信息和问题描述。因此，我们选择通过聚类后问题的留言主题进行地点/人群信息及问题描述的提取。

首先，我们运用结巴分词对一类问题中所有的留言主题进行切词，并返回切词结果和对应词性。接着，分别提取名词和动词中词频最高的四个词，检索出该类问题中包含高频词最多的一条留言主题，再次进行分词。由于留言主题没有统一格式，我们对这五条留言主题进行了手动分词

3.8.3 结果表格

最后，将上述提取的信息进行整合并用 Data.Frame 存储，保存为 xlsx 文档。输出的结果表格部分信息如下。

	问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	反对数	点赞数
4242	1	288728	A00080248	关于A市金晖优步花园相关问题的反映	2019/1/8 17:07:55	\n\t\t\t\t\t\n\t\t\t\t\t关于A市金晖优步花园(金晖惟楚花园）相关问题...	0	9
403	1	198111	A00053914	A3区金晖优步花园精装价格虚高，配套不落实	2019/2/25 11:29:26	\n\t\t\t\t\t\n\t\t\t\t\t金晖优步花园位于A市A3区A3区大道与杜鹃路...	0	0
1667	1	228091	A00072601	A3区金晖优步花园欺骗消费者	2019/3/22 18.29:28	\n\t\t\t\t\t\n\t\t\t\t\t金晖优步花园位于A市A3区A3区大道与杜鹃路...	0	2
4165	1	286663	A00011794	A6区时代年华小区毛坯房中的入户门计算进了全装修价格且价质不符	2019/8/7 10:20:44	\n\t\t\t\t\t\n\t\t\t\t\t时代年华小区毛坯中的入户门计算进了全装修价格...	2	3
373	1	197483	A00011794	再次反映A市时代年华项目装修价质不符问题	2019/8/12 16:24:48	\n\t\t\t\t\t\n\t\t\t\t\t本人是A市A6区时代年华楼盘，13栋业主，于...	0	3
...	...	...	...	...	...	...	...	...
2607	5	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	\n\t\t\t\t\t\n\t\t\t\t\t星沙的旧城改造今年就要全部结束了，而凉塘路的...	0	6
3411	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	\n\t\t\t\t\t\n\t\t\t\t\t在2017年6月份，星沙四区凉塘路的居民根据...	1	4
3638	5	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	\n\t\t\t\t\t\n\t\t\t\t\tA7县星沙街道的旧城改造今年就要全部完工了，...	0	6
4055	5	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	\n\t\t\t\t\t\n\t\t\t\t\t今年政府工作报告里说2019年要全部完成星沙...	0	6
924	5	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	\n\t\t\t\t\t\n\t\t\t\t\tA7县星沙街道的旧城改造今年就要全部完工了，...	0	1

图 4: 热点问题详情表

表格中留言详情显示的转义符在储存为 csv 文档时会自动隐去。

4 任务三：答复意见评价

4.1 数据处理

4.1.1 文本处理

文本中存在停用词会影响答复意见评价指标的准确性，因此我们对答复意见分词时剔除了停用词。

同时，对答复意见内容分析时，我们发现存在大量无意义的固定语句在答复意见的开始部分，这些语句特征一般具有“回复如下”、收悉”、已阅”、您好”、你好”、网“友”等词汇。因此我们将这些语句进行剔除。

4.1.2 日期处理

用 python 对附件四中的日期数据进行遍历后，我们发现日期数据共有两种不同格式:datetime 及 str。我们将所有日期格式统一转换为 datetime 格式，以便进行后续的统计操作。

4.2 评价指标设计

附件四提供了留言具体信息和相关部门对留言的答复意见，基于文本数据的特性及任务三的要求，我们构造了如下四种答复意见评价指标：

1）时效性

留言答复的时效性往往可以展现相关部门对群众留言反馈的效率和积极度。我们根据留言时间和答复时间计算出留言反馈的小时数，并正则化到 [0,1] 之间。由于答复时间与留言时间间隔越长，答复意见的时效性越低，我们将 1 减去正则化后的小时数作为答复意见的时效性得分。

2）相关性

由于附件四提供了留言详情和答复意见详情，我们可以计算出留言详情和答复意见详情的文本相似度，作为该答复意见的相关性得分。首先，我们对所有的留言详情和答复意见进行词典构造，并将文本转化成向量，然后使用 TF-IDF 技术进行加权，最后计算出每一条答复意见与其对应的留言详情的 LSI 相似度 (取值在 [0,1])，作为相关性得分。

3）完整性

答复意见的完整性可以很好地反映相关部门对留言进行答复时的认真程度，具有更高完整性的答复意见也能提高公众的留言体验。我们通过计算答复长度作为完整性的评价指标。答复长度越长代表相关部门对该问题的答复完整性越高。因此，我们统计了答复意见的长度，并将其正则化到 [0,1] 之间，将 90 分位数作为满分得分的阈值，超过阈值视为完整性满分，其余答复意见长度与阈值的比值作为完整性得分。

4）可解释性

可解释性可以等价于我们可以获取的可被理解的信息。在答复意见中，如果引用的条例、规范等文献越多，说明该答复越有理可依，且可以提供额外信息。另外，若答复意见的措辞越规范，则越容易被人理解。因此，我们对答复意见文本进行了是否引用的检测及是否有多点排序的检测，将它们的均值 (取值在 [0,1]) 作为可解释性得分。

根据四种答复意见评价指标，我们将每个答复意见在四种指标上的得分均值乘 10 后向上取整作为该答复意见的综合得分（取值为 1 到 10 的整数）。

4.3 指标分布展示

根据四种答复意见评价指标，我们计算了每一个答复意见在四种指标上的得分，并画出不同评价指标得分的频数分布图。从时效性得分频数分布图可以看出，绝大部分的答复意见回复速度很快，因此得分都较高。相关性得分频数分布图中出现大量得分几乎为 0 的答复意见，我们查看了相对应的答复意见，发现这些意见未包含具体的答复内容，例如“网友：您好！留言已收悉”或只有时间没有相应内容。这也符合相关性得分低的特征，说明我们相关性计算方法合理。完整性得分频数分布图中显示有 10% 的意见内容得到满分，符合对应评价标准。可解释性得分频数分布图中展示到有不到一半的答复意见使用引用条例规范或多点排序，通过查看对应的答复意见，我们发现这些内容理解上更容易被人接受，也更加具有可信度。

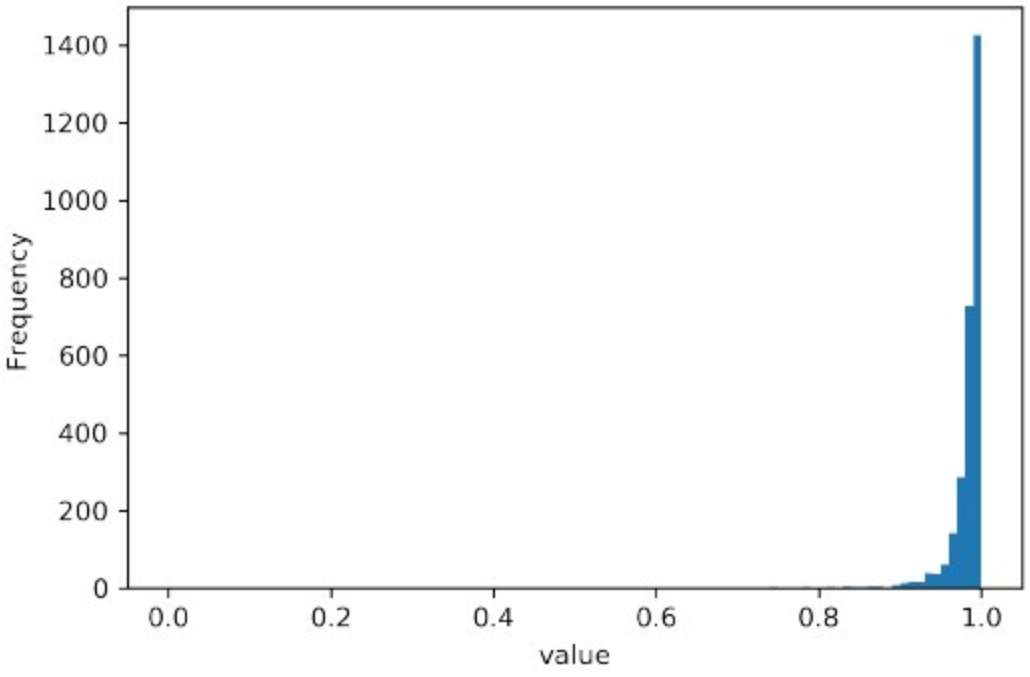

图 5: 时效性得分频数分布

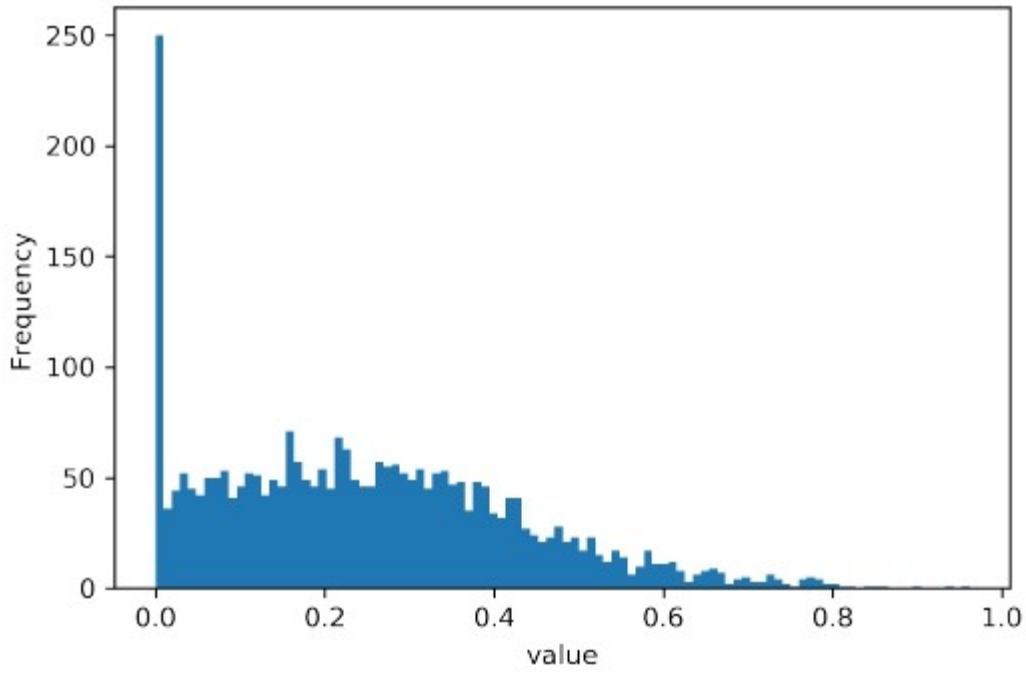

图 6: 相关性得分频数分布

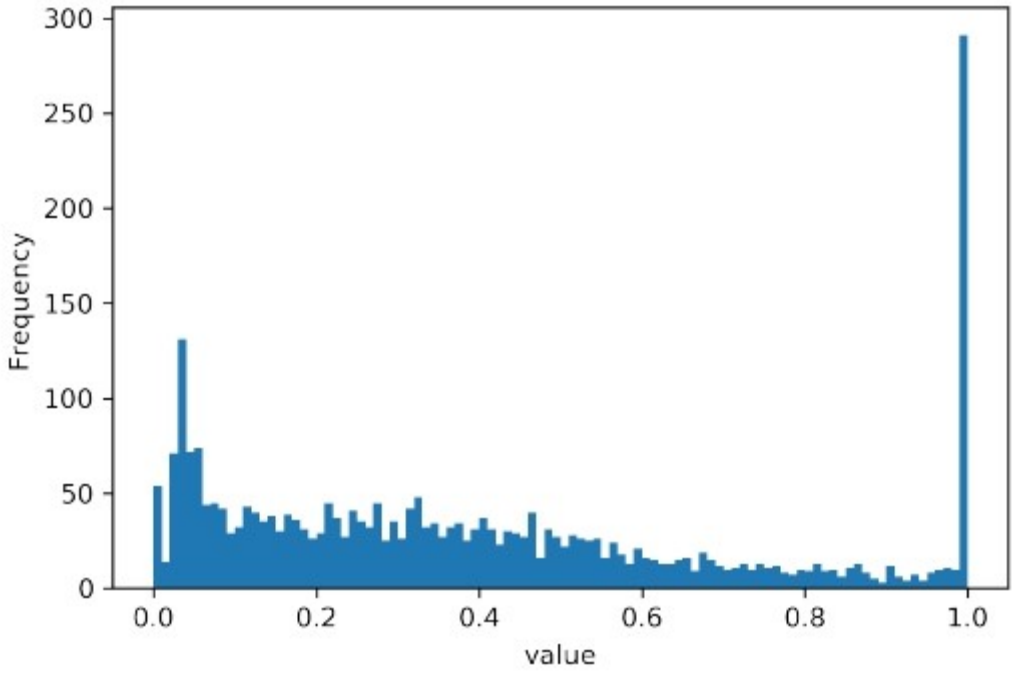

图 7: 完整性得分频数分布

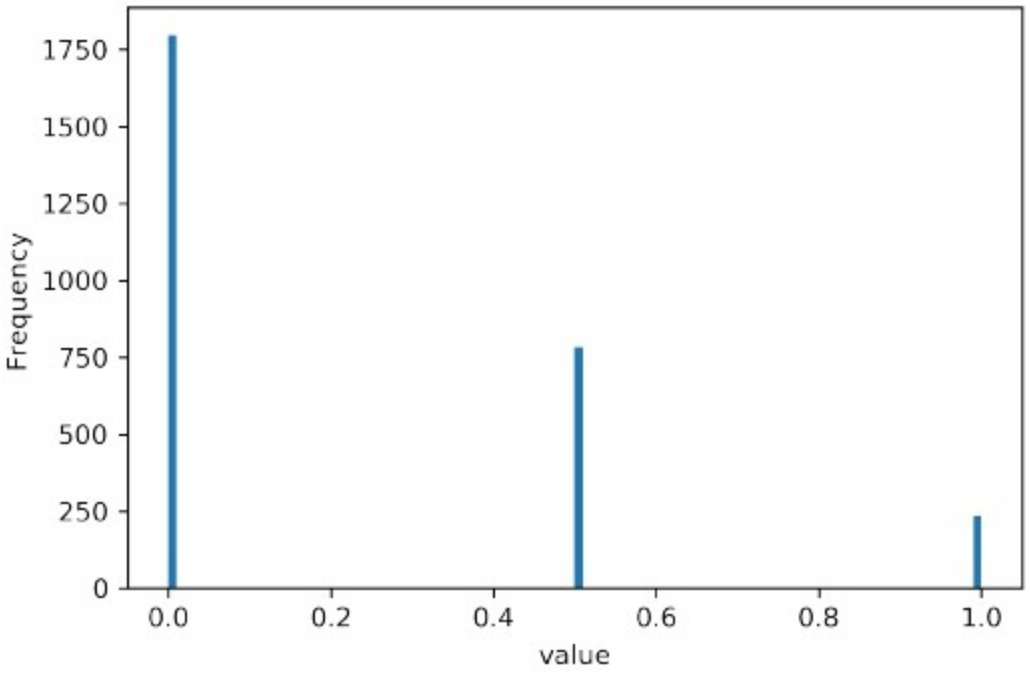

图 8: 可解释性得分频数分布

4.4 综合指标结果

通过对四种答复意见评价指标的综合，我们可以得到每个答复的综合得分。从综合指标结果频数分布图上，我们可以看出大部分的答复得分处在中间部分，少部分的得分处在高分部分，整体的分布图近似正态分布的单峰结构，因此我们设计的评价指标和综合得分计算方法合理，且计算方式易行，是答复意见评价的合适方案。

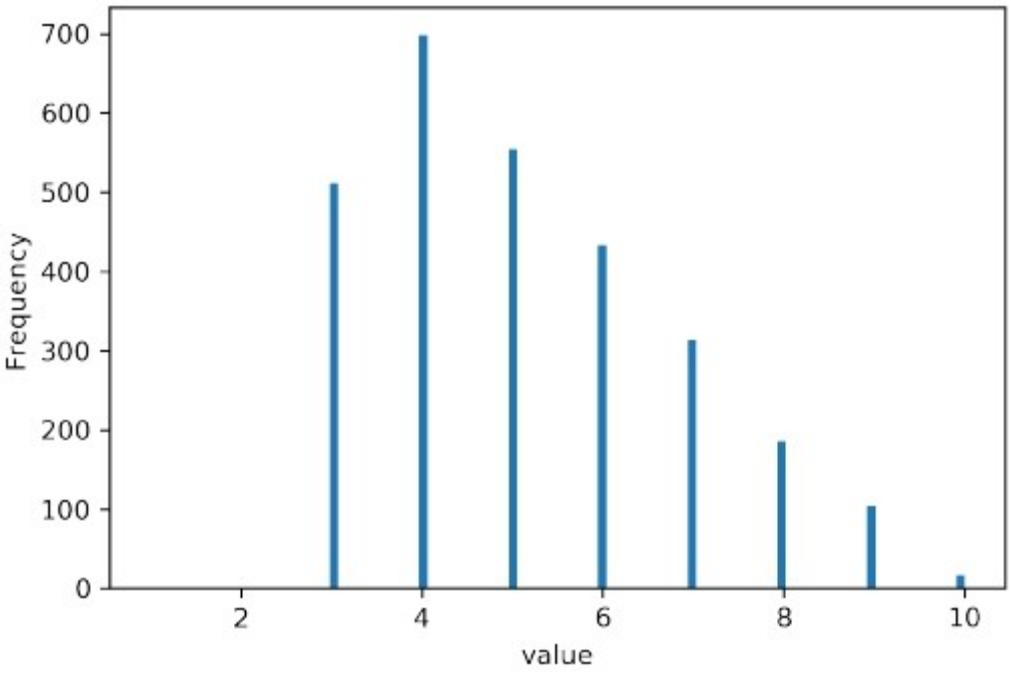

图 9: 综合指标结果频数分布

参考文献

[1] GÃŠrard Biau. Analysis of a random forests model. *Journal of Machine Learning Research*, 13(Apr):1063–1095, 2012.

[2] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[3] Murat Kantarcıoglu, Jaideep Vaidya, and C Clifton. Privacy preserving naive bayes classifier for horizontally partitioned data. In *IEEE ICDM workshop on privacy preserving data mining*, pages 3–9, 2003.

[4] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[5] S Rasoul Safavian and David Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674, 1991.

[6] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[7] Raymond E Wright. Logistic regression. 1995.

[8] 源初心理工作室. 人的心理情绪在一天中的波动变化[DB/OL]. http://blog.sina.com.cn/s/blog_142728cb00102v97i.html, 2015-02-15.

[9] lhxsir. 中文文本的关键字提取[DB/OL]. https://blog.csdn.net/lhxsir/article/details/83304173, 2018-10-23.

[10] 致林.python 进行中文文本聚类（切词以及 kmeans 聚类）[DB/OL]. https://www.cnblogs.com/bincoding/p/8878098.html, 2018-04-18.

[11] 王天泽. Python 之 gensim 自然语言处理库 [DB/OL]. https://www.cnblogs.com/wangqingyi/articles/5911647.html, 2016-09-27.

[12] 百度百科. 层次分析法（运筹学理论）[DB/OL]. https://baike.baidu.com/item/层次分析法/1672?fr=aladdin.