

基于深度学习的智慧政务平台建设

摘要

在 5G 网络高速发展的背景下，随着云计算、AI、大数据、物联网等前沿技术的发展和运用，政务行业转型已进入一个全新的节点—智慧政务，即“互联网+政务”。同时，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。本文主要通过数据建模，从政务人员角度出发，构建合理的标签分类及问题归类模型，并提出留言热度评价数学模型及答复内容质量评价数学模型，以替代一部分人工实现智能化。

针对问题一，建立一个基于有监督学习的多分类模型。利用百度 API 通过**回译法**对部分数据实现**数据增强**，以解决数据不平衡问题。由于网络问政平台的特殊性，我们选择以能够学习长期的规律的 **LSTM 长短期记忆网络**来构建一级标签分类模型，并使用**混淆矩阵**和 **F_1 分数**评价其分类效果

针对问题二，问题归类实际为一个**文本聚类问题**。首先对需要进聚类的留言内容观察其特点，便于有针对地调整模型参数，以达到更优的聚类效果。在对数据进行基本清洗后，利用 Python 爬虫在二手房交易网站上爬取全国社区名**扩充词库**，使用 HanLP2.0 进行带**地名识别和词性标注**的分词，利用 TF-IDF 计算权重，根据数据特点对名词和地点名词进行一定程度**加权**。对处理后的留言内容，通过 PCA 法进行降维，通过 **DBSCAN 聚类方法**对同一主题进行聚类，**对比不同参数聚类效果**，选择最适合的数据结果。确定热度评价指标，设定准则即**人群广度准则**、**“羊群效应”准则**、**时间衰退准则**，依此构建热度评价模型，并给出热点问题排名表。

针对问题三，要求对网络问政平台答复意见建立合理质量评价指标，我们结合数据特点从群众角度建立属于本文的评价指标，问政平台答复从规范性上比在线评论要易于判断。相对的，答复文本的长短主要取决于留言内容，文本长度与答复质量上并无线性关系。我们借用前人“4C”概念，对答复进行评价模型建立属于本文的**“3C1T”答复内容质量评价模型**。

最后，给出每个模型的评价及改进方法。

关键词：LSTM 长短期记忆网络 DBSCAN 聚类方法 “3C1T”答复质量评价指标

一、 问题重述

1.1 问题背景

在 5G 网络高速发展的背景下，随着云计算、AI、大数据、物联网等前沿技术的发展和應用，政务行业转型已进入一个全新的节点—智慧政务，即“互联网+政务”。同时，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。目前各类电子政务系统的成熟度、智能化尚有欠缺，管理实效有待提升，服务供给能力、服务响应能力以及服务智慧能力不足。因此，建立完善的智慧政务系统急不可迫。

在处理网络问政平台的群众留言时，政府工作人员按照一定的划分体系对留言进行归类，再分派到各相应职能部门处理，但是人工工作有着工作量大、效率低、出错率高的弊端，故迫切需要建立留言内容标签分类模型，以提高政府工作人员效率。其次，对留言内容实现自动归类并提取主题内容，建立合理的热度评价模型，可有助于及时发现热点问题，使相关部门进行有针对性地处理，提升服务效率。对于服务型政府，致力于提升服务质量，对此，有必要建立一套答复意见评价方案，对答复内容做出客观评价，不断改进服务质量，杜绝个别人尸位素餐，网络问政平台沦为“空架子”。

1.2 问题提出

根据以上背景，以及给出的四个附件，需要解决以下问题：

1. 按照一定划分体系（参考附件 1），根据附件 2 给出的数据，建立关于留言内容的一级标签分类模型，并使用 F_1 分数评价其分类效果。
2. 根据附件 3 将某一时段内反映特定地点或特定人群问题的留言进行归类，定义合理的热度评价指标，并给出评价结果。
3. 针对附件 4，对网络问政平台答复意见建立合理质量评价指标，并尝试实现。

二、 问题分析

2.1 问题一分析

问题一实际上是建立一个基于有监督学习的多分类模型。难点在于群众留言表达口语化、表达多样性以及网络问政平台长期接受群众留言，随着时间推移，数据量的增大，后期数据可能会遇到的爆炸以及梯度消失。

我们需要根据附件 2 给出的数据，建立关于留言内容一级标签的分类模型。我们的目标是将留言分到其所属的一级标签中去，且每条留言只能对应一个一级标签。由于平台上留言表达的多样性，我们的模型需要具有记忆性，以便“理解”留言文本中语句关系。除此之外，网络问政平台长期接受群众留言，为防止后期数据可能遇到的爆炸以及梯度消失，模型只需要保留重要的信息，忽略不重要的信息。

标签分类模型构建完成后，我们对其使用混淆矩阵和 F_1 分数评价其分类效果。

2.2 问题二分析

问题二要求将某一时段内反映特定地点或特定人群问题的留言进行归类，定义合理的热度评价指标，并给出评价结果。此问题是一个文本聚类问题。此问题难点在于，群众留言表达口语化且具有多样性，如何实现优秀的聚类效果。其次，合理的热度评价指标需要满足以下条件：

1. 反映问题人数越多，该问题热度越高。
2. 降低“羊群效应”影响，点赞数，反对数对问题影响应次于符合规则留言条数。
3. 工作人员可及时发现近期热点问题，留言热度需随时间推移衰退。

2.3 问题三分析

问题三要求对网络问政平台答复意见建立合理质量评价指标，我们结合数据特点从群众角度建立属于本文的评价指标，问政平台答复从规范性上比在线评论要易于判断。相对的，答复文本的长短主要取决于留言内容，文本长度与答复质量上并无线性关系。我们借用前人“4C”概念，对答复进行评价模型建立属于本文的“3C1T”答复内容质量评价模型。

三、 基本假设

1. 网络问政平台有效答复为 20 个工作日。
2. 数据中所涉及信息皆为真实。

四、 符号说明

符号	说明	单位
h^*	留言基本热度	/
z	留言受赞同程度	/
θ	时间衰退因子	/
t	观测时间	天
t_0	留言时间	天
α	热度系数，控制有效留言数因素对热度影响，本文 $\alpha = 1$	/
β	冷却系数，本文 $\beta = 0.0345$	/
Complete	完整性衡量答复是否具有规范格式	/
Correlation	相关性检测答复主题与留言主题相关度	/
aCcountable	可解释性检测答复判断依据所占比	/
Timeliness	时效性检测是否在规定时间内进行答复	/
Q	答复内容质量评分	/
w_i	表示各指标对应的权重 $i = 1,2,3,4$	/

注：未列出符号及重复的符号以出现处为准

五、 问题一模型的建立及求解

5.1 问题的分析

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。我们需要根据附件 2 给出的数据，建立关于留言内容一级标签的分类模型。我们目标是将留言分到其所属的一级标签中去，且每条留言只能对应一个一级标签。由于平台上留言表达的多样性，我们的模型需要具有记忆性，以便“理解”留言文本中语句关系。除此之外，网络问政平台长期接受群众留言，为防止后期数据可能遇到的爆炸以及梯度消失，模型只需要保留重要的信息，忽略不重要的信息。综合考虑，我们选择能够学习长期的规律的 LSTM 长短期记忆网络来构建一级标签分类模型。

标签分类模型构建完成后，我们对其使用 F_1 分数评价分类效果。

5.2 模型的建立

5.2.1 数据预处理

附件 2 中数据各标签数据数量分布如图 5.1，由图我们可以观察到，城乡建设、劳动和社会保障数据量较大，均为 2000 左右。教育文体、商贸旅游、环境保护、卫生计生数据量依次减少。数据量最小的交通运输仅有 638 条，还未到数据量最大的城乡建设的三分之一。

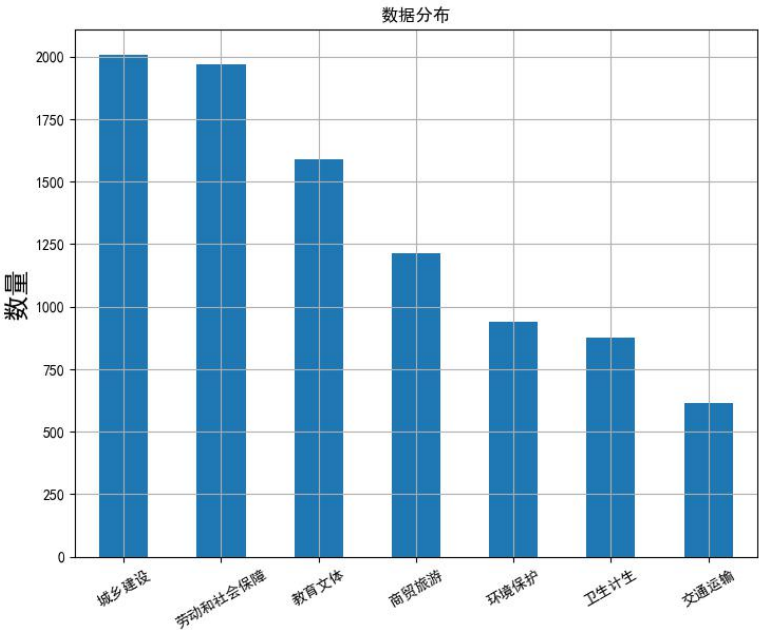


图 5.1 附件 2 数据分布直方图

数据不平衡会影响多标签分类模型效果。针对此的情况，我们需要对数据量小的标签作数据增强处理。然而，文本数据增强不同于数值数据或图像数据，前者是离散的。对于文本数据而言，通过线性插值法后的句子可能根本不存在或因一个较小的扰动影响语义，从而影响分类效果。为更好地分类效果，我们采用回译法（指将留言文本翻译为另一语言再翻译成中文）解决留言文本数据不平衡问题。此方法优点在于增强数据同时保证其留言文本句子语义不变。我们利用百度 API 来实现回译法数据增强。完成后，数据分布如图 5.2，各标签数据数量基本持平。

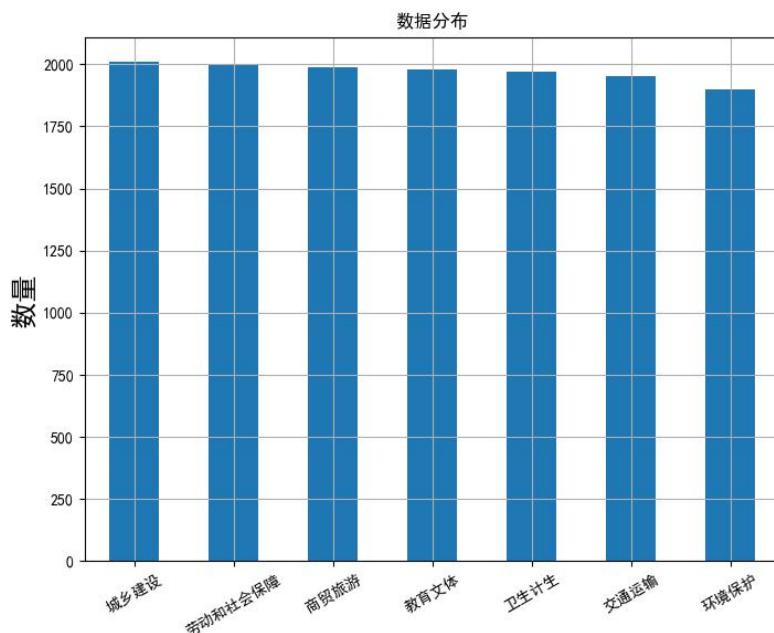


图 5.2 数据增强后数据分布

解决完数据不平衡问题后，我们还对一级标签进行 id 化处理，对文本数据进行去除标点符号及无意义停用词，分词等操作，以保留文本重要信息，便于接下来的分类模型训练。

5.2.2 基于 LSTM 的多标签分类模型构建

数据预处理完成后，接下来我们开始构建 LSTM 模型。我们首先将数据进行向量化处理，转化为一个整数序列向量，提取最频繁使用的 50000 个词，使每条留言最大的词语数为 250 个（超过的将会截去，不足的将会被补 0），拆分训练集和测试集（比例为 9:1）。我们需要构建一个可把留言文本分到对应的 7 个一级标签中，且每条数据只能对应一个标签。

由此定义一个 LSTM 的序列模型：

1. 模型的第一层为嵌入层，它使用长度为 100 的向量来表示每一个词语
2. SpatialDropout1D 层在训练中每次更新，将输入单元的按 0.2 比率随机设置为 0，以防止过拟合
3. LSTM 层包含 100 个记忆单元
4. 输出层为包含 7 个分类的全连接层
5. 多分类问题，将激活函数设置为 'softmax'，损失函数设为分类交叉熵 'categorical_crossentropy'。

定义的 LSTM 模型各项参数如表 5.3：

层（类型）	输出（Output Shape）	参数（Param）
Embedding	(None, 250, 100)	5000000
SpatialDropout1D	(None, 250, 100)	0
LSTM	(None, 100)	80400
Dense	(None, 7)	707
Total params:5,081,107		
Trainable params:5,081,107		
Non-trainable params:0		

表 5.3 LSTM 模型各项参数

经过 5 个周期的训练，损失函数趋势图和准确率趋势图如下图 5.4，5.5，

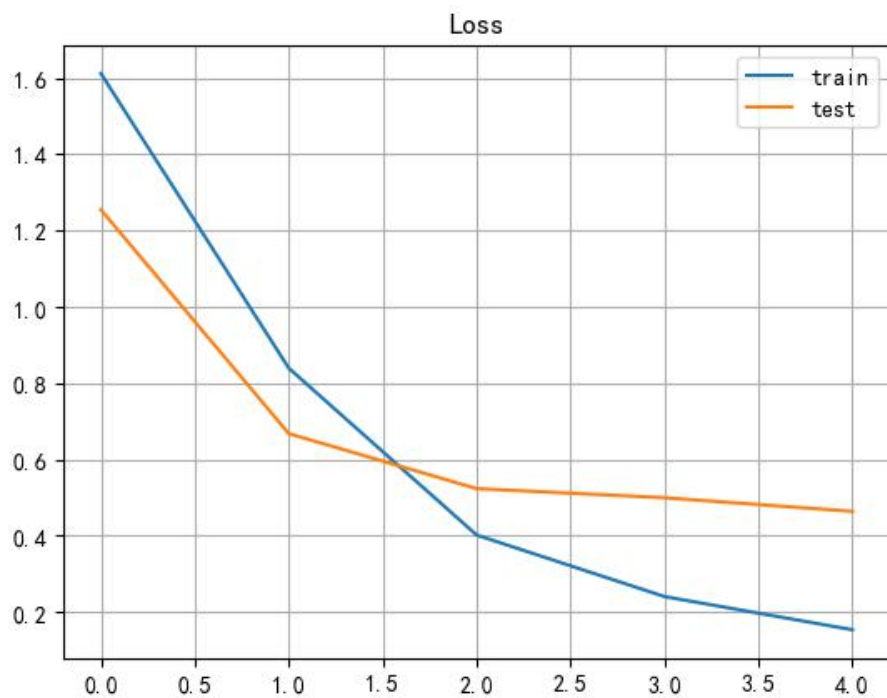


图 5.4 损失函数趋势

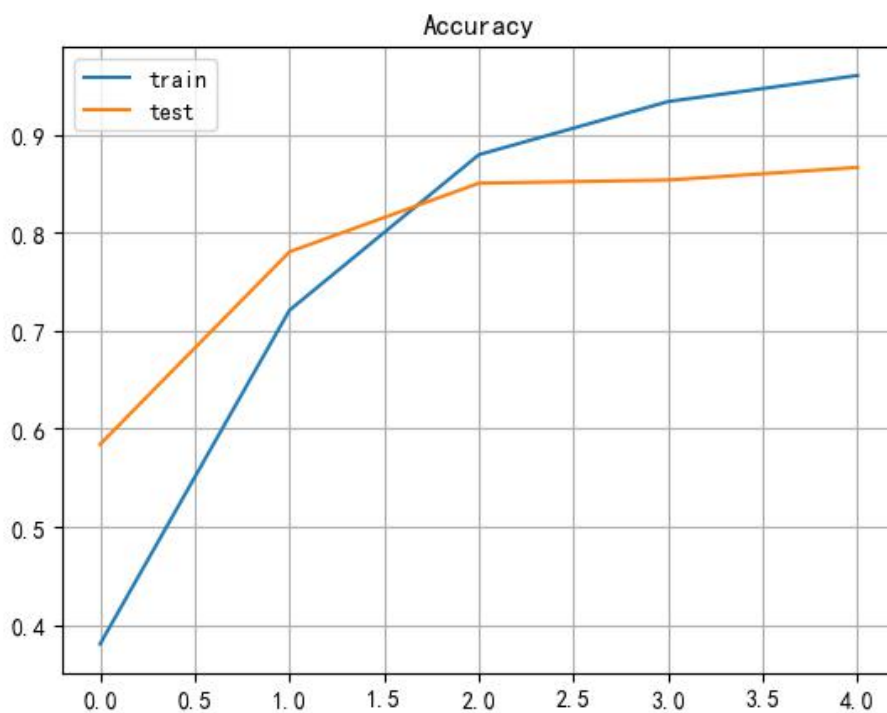


图 5.5 准确率趋势

从图上，随着训练周期的增加，模型在训练集中损失逐渐减小接近 0，这是典型的过拟合现象。而在测试集中，损失随着周期的增加，曲线逐渐平缓，不再下降。从图 中我们可以发现，随着训练周期的增加，模型在训练集中准确率越来越高逐渐接近 1，而在测试集中，准确率随着训练周期的增加，曲线逐渐平缓，不再上升。

5.3 模型的评价

我们通过混淆矩阵和 F_1 分数对其评价分类效果。混淆矩阵如图 5.6，混淆矩阵的主对角线表示预测正确的数量，除主对角线外其余都是预测错误数量。从上面的混淆矩阵可以看出劳动和社会保障标签正确率最高，商贸旅游分类正确率相对较低。

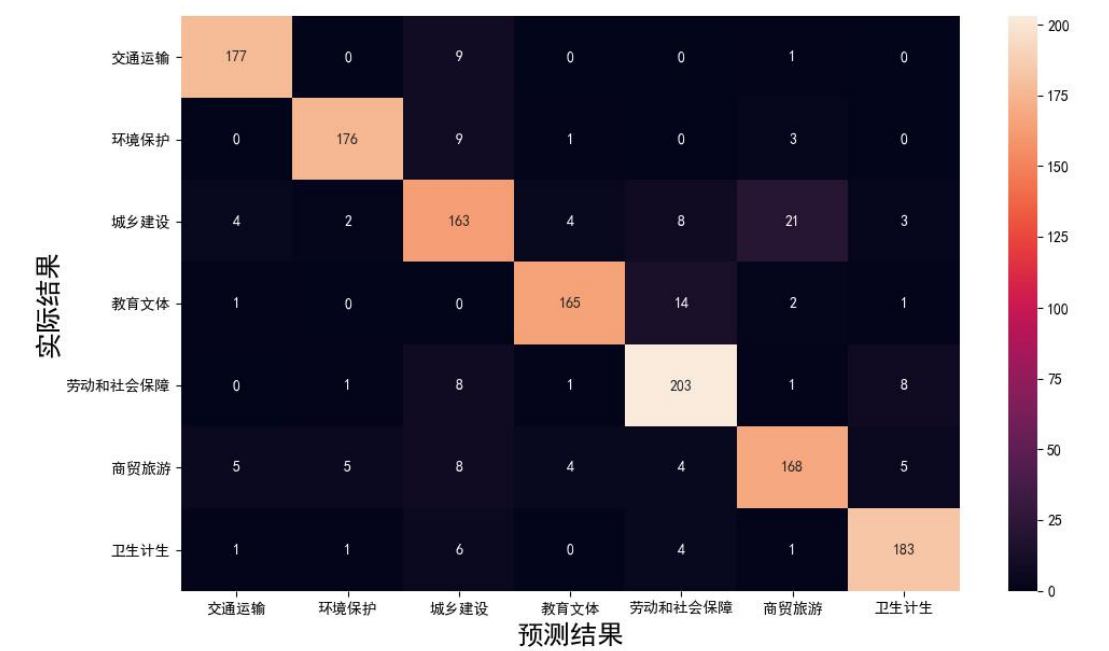


图 5.6 混淆矩阵

接下来，我们对模型使用 F_1 分数评价分类效果，多分类模型分类效果如表 5.6，各一级标签 F_1 分数均在 0.8 以上，最高的交通运输和环境保护达到了 0.94，此模型准确率达到 0.8。

	precision	recall	f1-score	support
交通运输	0.94	0.95	0.94	187
环境保护	0.95	0.93	0.94	189
城乡建设	0.80	0.80	0.80	205
教育文体	0.94	0.90	0.92	183
劳动和社会保障	0.87	0.91	0.89	222
商贸旅游	0.85	0.84	0.85	199
卫生计生	0.92	0.93	0.92	196
accuracy	/	/	0.89	1381
macro avg	0.90	0.90	0.90	1381
Weighted avg	0.89	0.89	0.89	1381

表 5.6 基于 LSTM 的多分类模型效果

综上所述，基于 LSTM 建立的多分类模型分类准确率达到 0.8，但考虑到通过百度 API 回译法进行数据增强后，会有一些的数据重复，对分类效果判断有一定负面影响。

六、 问题二模型的建立及求解

6.1 问题的分析

问题二要求我们根据附件 3 将某一时间段内反映特定地点或特定人群问题的留言进行归类。通过对附件 3 进行观察，发现数据具有以下特点：

- (I) 包含大量路名、社区名等专有名词。
- (II) 同一地点出现多个问题。
- (III) 出现“灌水”现象，相同用户 ID 同一问题重复留言。
- (IV) 点赞数、反对数出现极个别数量极大。

其中，(I) 影响文本分词效果，(II) 会对聚类效果造成影响，(III) (IV) 会影响热度指数排名效果。另外，为了建立更为合理的热度评价指标，我们提出以下三条准则：

1. 人群广度准则：涉及人群广度越大，留言条数越多，热度越高。
2. “羊群效应”准则：留言点赞数、反对数越大，用户盲从赞同或反对数占比越大。
3. 时间衰退准则：留言热度随时间流逝减小。

问题二流程如下，在对数据进行清洗后，使用 Python 爬虫在二手房交易网站上爬取全国各地社区名，同时通过网络搜索添加路名等数据作为自定义字典，使用 HanLP2.0 进行带地名识别和词性标注的分词，利用 TF-IDF 计算权重。对处理后的留言内容，使用 DBSCAN 方法对同一主题进行聚类，对比不同参数聚类效果。随后确定热度评价指标，构建模型，对反映问题进行排序，热度前五的热点问题整理成表格。

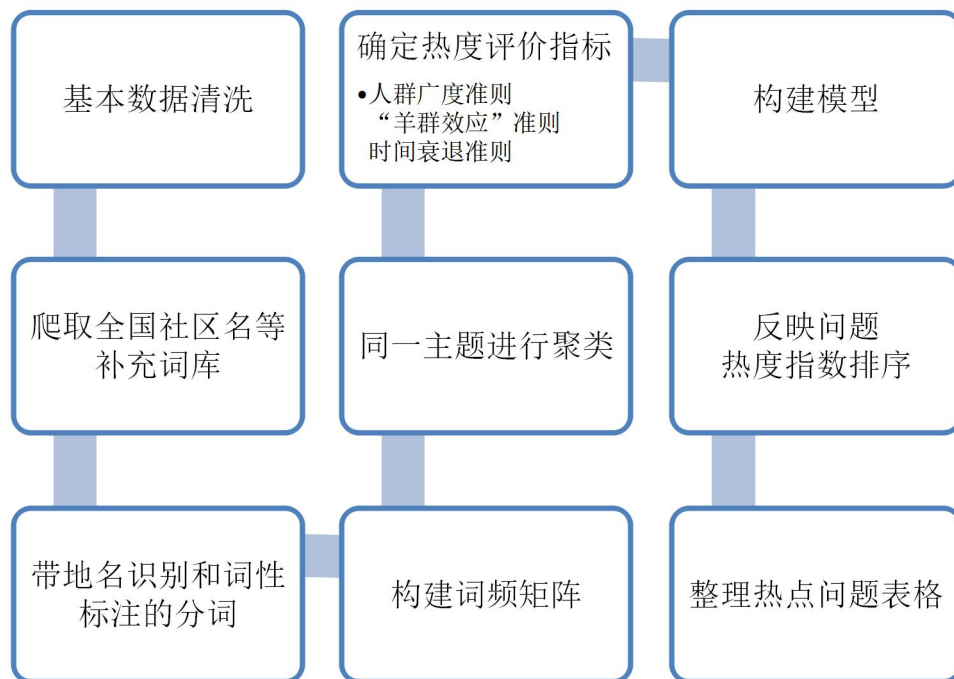


图 6.1 问题二流程图

6.2 模型的建立

6.2.1 数据预处理

在基本的数据清洗操作后，数据中包含大量路名，社区名等专有名词，给分

词造成难度。且从政务人员角度出发，热点问题是反映某一时间段内反映特定地点特定人群的问题，因此正确地识别留言中的地点名词至关重要。为了实现更好地文本分词效果，我们对分词词库扩充地点专有名词。

我们选择在某二手房买卖平台网站爬取全国主要城市小区名，该网站具有专门的小区板块，易于爬取数据。



图 6.2 房产买卖交易平台

扩充前：
A , 市 , 万科 , 魅力 , 之 , 城 , 自 , 打 , 交 房 , 入 住 , 后 , 底 层 , 商 铺 , 无 , 排 烟 , 管 道

扩充后：
A 市 万 科 魅 力 之 城 , 自 打 , 交 房 , 入 住 , 后 , 底 层 , 商 铺 , 无 , 排 , 烟 , 管 道

图 6.3 词库扩充前后效果对比

词库扩充前后 HanLP2.0 对留言分词效果对比如图 6.3，地点专有名词不再被拆分，使后续模型训练效果更优。

另外，附件 3 中数据留言时间范围是 2017/6/8 17:31:20 至 2020/1/26 19:47:11。

6.2.2 基于 DBSCAN 的文本聚类模型

在文本预处理过后，构建文本聚类模型之前，我们首先将文本转化为词频矩阵，通过 TF-IDF 统计每个词语的权值。在问政平台留言中，名词能反映文本内容，其中地点名词最具代表性，因此我们对名词和地点名词增加其权重。和传统的 K-Mans 算法相比，DBSCAN 优势在于可发现任意形状聚类簇，同时它在聚类的时候还可以找出异常点，有助于热点问题的提取，去除反映人数少的主题，减少计算量。综合衡量，我们选择基于 DBSCAN 构建文本聚类模型。由于文本数据为离散型，所构成向量维度较高，为更好地聚类效果，我们通过 PCA 法进行降维。下面是模型各项参数设置，不同效果对比。名词权重设置为 1.1，地名权重设为 1.25 及 1.5，聚类半径分别为 0.8、1.0、1.2，聚类效果如图 6.4, 6.5，

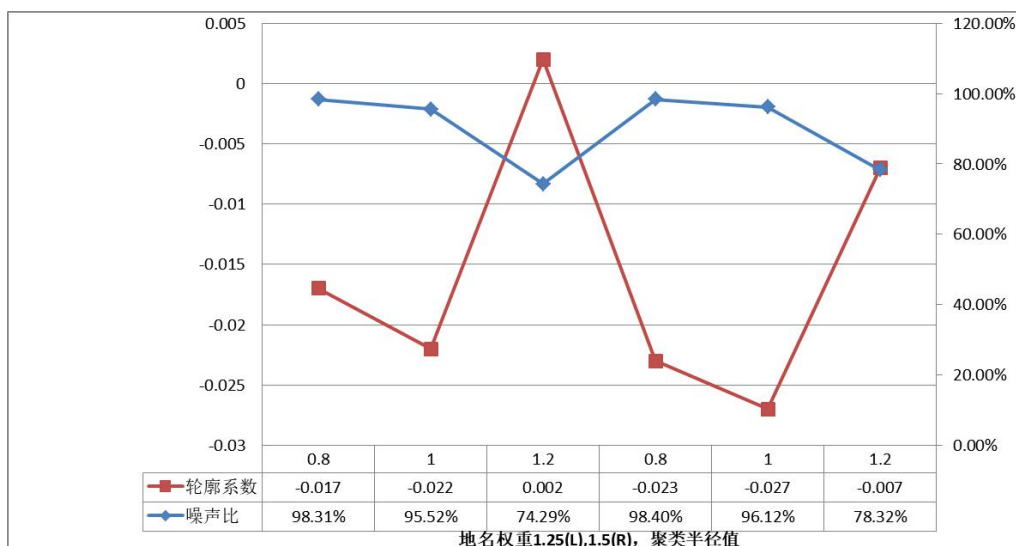


图 6.4 不同参数下 DBSAN 聚类效果

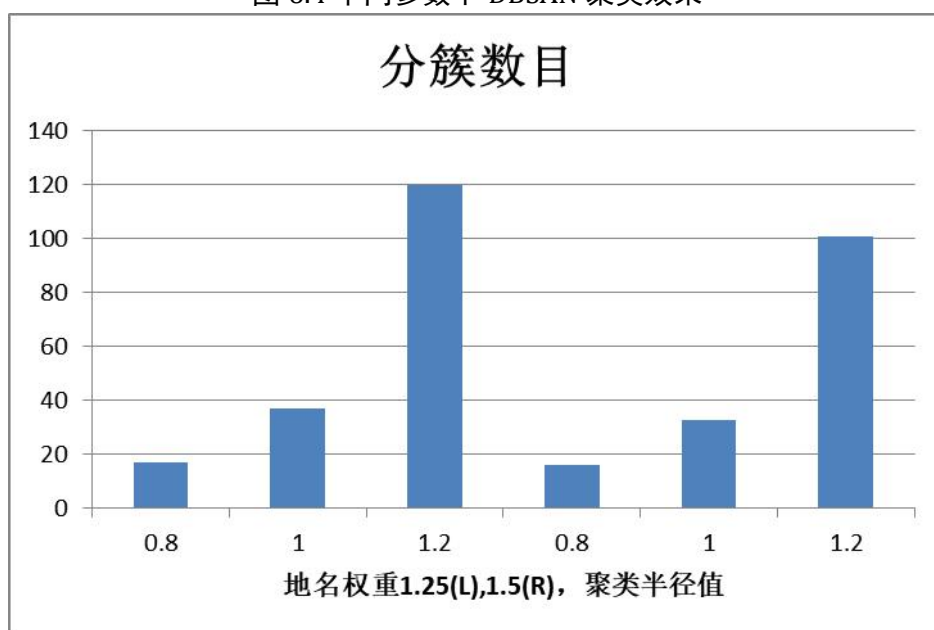


图 6.5 不同参数下 DBSAN 聚类分类簇目

由图可看出，随着聚类半径增大，分簇数目依次增大，半径为 1.2 分类簇目是半径 1 的三倍有余。文本聚类的特殊之处在于文本内容具有语义，以及我们目标为关注热点问题，所以聚类半径过大可能导致主题不同但具有一定相似性的问题归为一类，给后续热度评价带来困难。轮廓系数表示聚类效果的好坏，值越大越好。同时，我们还需对聚类后数据进行观察。综合考虑后，我们选择聚类半径为 1，地名权重为 1.25 的聚类结果作为下面热度评价指标数据。

6.2.3 建立热度评价模型

热点问题受留言数量、时间因素、赞同数及反对数的影响。为了定量分析这种影响，建立合理的热度评价指标。

首先，我们认为，留言数量为主要影响因素，在数据预处理中已对违规灌水行为的数据做出处理。其次，对于赞同及反对行为，有极个别数量极大，我们认为人存在羊群效应，即“跟风”行为。赞同数多的留言，有更大的几率被人们点赞，继而造成“赞多者愈多赞”现象。根据孙连等人提出一种用户社交网络排

序算法，利用“牛顿冷却定律”整合了时间衰退度因子。问政平台与社交网络相比，具有流量小，更新频率慢等特点，我们首要保证在一定时间内留言数量为主要影响因素，其次保证新留言较旧留言有更高热度。通过合理设置“牛顿冷却定律”中冷却系数可有效控制热度随时间衰退速度，以达到预想效果。

由此提出三条准则：

4. 人群广度准则：涉及人群广度越大，留言条数越多，热度越高。
5. “羊群效应”准则：留言点赞数、反对数越大，用户盲从赞同或反对数占比越大。
6. 时间衰退准则：留言热度随时间流逝减小。

综上所述，对单条留言建立以下模型以计算热度指数：

1. 基础热度 h^*

$$h^* = \text{热度系数}\alpha \times \text{留言数量}(\text{number})$$

热度系数 α 可根据实际情况决定， α 越大代表话题总热度受留言人数因素影响越大。为简化计算，本文 $\alpha = 1$ 。对单条留言数量即为 1，此处引进是为了热度系数服务。

2. 赞同数与反对数之差 x

$$x = \text{赞同数}(\text{ups}) - \text{反对数}(\text{downs})$$

留言受赞同程度 z

$$z = \max(x, 1)$$

z 表示留言收到的净赞同数，若反对数大于或等于赞同数，则 $z = 1$ 。

3. 时间衰退度因子 θ

$$\theta = e^{-(\text{冷却系数}(\beta) \times \text{间隔时间}(t-t_0))}$$

冷却系数(β)根据实际情况决定， β 决定时间因素的影响程度， β 越大，表示随着时间流逝热度下降速度越快。本文 β 的值为 0.0345，即留言赞同程度相等的情况下，在二十天前留言热度较当天留言下降百分之五十。间隔时间($t-t_0$)单位为天， t_0 为留言时间， t 为当前时间，本文 t 为 2020 年 1 月 27 日零点。

结合以上变量，热度 h 计算公式为，

$$h = \theta(h^* + \log_2 z)$$

将此公式中 $\log_2 z$ 为被赞同及反对数影响的部分，表示赞同数超出反对数部分越多，得分越高。基于“羊群效应”准则，人们对赞同数多的留言有更大几率跟风点赞。考虑到政务留言平台，人流量较小的因素，我们对 z 取以 2 为底的对数，也就是说前 8 人赞同与后 56 人赞同得分相同，留言受赞同程度越大，越到后面点赞对得分的影响越低，以减弱“羊群效应”影响。

6.3 热度指数排序结果

将反映同一类问题的各条留言热度求和，计算出所有反映问题的热度，对其进行排名，得出排名前五的热点问题。

我们选取最后一条留言日期后一天零点作为观测时间，即在 2020 年 1 月 27 日零点时，热点问题排名如下表 6.6，

热度排名	时间范围	地点/人群	问题描述
1	2019/11/2-2019/12/26	A 市丽发新城小区	丽发新城小区旁边建的搅拌站噪音灰尘等污染严重
2	2019/4/29-2019/12/16	A 市市公交站房	A 市常规公交运营不善
3	2019/9/11-2019/12/3	A3 区中海国际社区	中海国际社区三期与四期的空地违规施工
4	2019/10/31-2019/11/2	A 市南塘小区公交站	建议加建一个城轨公交站
5	2019/8/23-2019/9/6	A4 区绿地海外滩小区	渝长厦高铁最新的红线征地范围以及走向经过违规

表 6.6 热点问题排名

七、 问题三模型的建立及求解

7.1 模型的建立

7.1.1 3C1T 评价指标构建

关于文本质量评价方面，郭银灵提出通过改进前人经典评价指标以及研究发现的数据质量评价指标，形成与构建在线评论质量评价指标体系。同理，对于问政平台答复，我们也可结合数据特点从群众角度建立属于本文的评价指标。问政平台答复从规范性上比在线评论要易于判断。相对的，答复文本的长短主要取决于留言内容，文本长度与答复质量上并无线性关系。

从数据角度出发，来自《数据整理实践指南》，用“4C”概括数据质量分析，即完整性（Complete）、一致性（Coherent）、准确性（Correct）与可解释性（aCcountable）。准确性在留言答复中没有这一说法，遂不考虑此因素。问政平台需要及时答复留言，具有时效性（Timeliness）。在此，我们可以借用“4C”概念，对答复进行评价模型建立，具体如下：

1. 完整性（Complete）可以理解为答复具有一定格式的规范开头语及结束语。简单来说，完整的答复标记为 1，不完整则标记为 0。
2. 一致性（Coherent）本指各数据之间依照一定逻辑建立关系，一致性检查是针对关系的检查。此处我们简单替换成相关性（Correlation），即留言和答复主题的相关性，是否“答其所问”。
3. 可解释性（aCcountable）本意指可追踪数据来源。从群众角度出发，对留言答复而言，我们可理解为是否“出之有因”（答复内容有迹可循），是否有给出判断依据，有无引用相关条例做出解释。通过计算判断依据及引用部分在答复文本中所占比例作为可解释性数据。计算采用以下公式：

$$aCcountable = \frac{\text{判断依据总字词数}}{\text{文本总字词数}}$$

4. 时效性（Timeliness）是否在规定时间内进行答复。根据《贵港市网络问政平台运行管理办法》规定，根据内容、性质、缓急程度，对信息进行分类办理。根据事件复杂程度回应时间依次为 3 个工作日，10 个工作日，20 个工作日和两个月内。在此，我们做简化处理，以 20 个工作日为基准线，判断回复是否具有时效性。

那由此形成的问政平台答复质量评价 3C1T 指标如下表 7.1:

3C1T 指标	说明
完整性（Complete）	衡量答复是否具有一定规范格式
相关性（Correlation）	检测答复主题与留言主题相关度
可解释性（aCcountable）	检测答复判断依据所占比
时效性（Timeliness）	检测是否在规定时间内进行答复

表 7.1 问政平台答复质量评价指标

7.1.2 确定权重

确定权重前，我们需要将指标量化。对于完整性，我们通过对比答复内容截取，将附件 4 中具有一定规范格式的开头语，结束语提取入词库，通过对比识别，将具有一定的规范格式的答复内容标记为 1，否则记为 0。对于相关性，我们建立向量空间模型 VSM（vector space model），对留言内容及答复内容分别提取主题，选取特征项，通过 TF-IDF 为特征值赋权，计算两文本的相似度，以此为相关性。对于时效性，我们通过调用 chinese_calendar 模块，计算留言与答复时间中工作日的数量，若小于等于 20 则记为 1，否则记为 0。

根据提取的指标特征，将各指标量化后，构建答复质量评价模型，此处用 Quality 表示答复质量评论的质量，得分记为 Q，建立以下模型：

$$Q = w_1 \text{Complete} + w_2 \text{Correlation} + w_3 \text{aCcountable} + w_4 \text{Timeliness}$$

其中， $w = (w_1, w_2, w_3, w_4)$ 表示各指标对应的权重，Complete 为完整性，Correlation 为相关性，aCcountable 为可解释性，Timeliness 为时效性。

通过前人相关研究，确定各指标权重，由表可知，人们认为相关性对网络问政平台答复内容质量影响最大，完整性对质量影响最小。

3C1T 指标	权重（0,1）
完整性（Complete）	0.15
相关性（Correlation）	0.35
可解释性（aCcountable）	0.25
时效性（Timeliness）	0.25

表 7.2 相关研究确定权重

由此我们可确立问政平台质量评价模型，将留言答复内容质量量化表示。

7.2 模型的评价

由于时间关系及技术原因，本文并未继续进行深入研究，也是本次遗憾之处。答复内容质量量化对政务人员而言，可有效规范答复内容，提供相关改进方向。对政府而言，此指标可作为政务人员工作评价的参考指标，防止个别人尸位素餐。对群众而言，此指标的提出可使老百姓“有处可说理”，防止网络问政平台沦为“空架子”。

八、 模型评价、改进

8.1 问题一：

优点：

1. 通过回译法增加数据，使各标签数据量基本持平，基本完整保留数据信息，从而获得更好的模型模型训练效果。
2. 基于 LSTM 长短期记忆网络的多分类模型会保留重要信息，忽略无关信息，

计算量较小,可应对长时间序列的留言分类工作急速增长的数据量和可能出现的梯度消失。

缺点:

1. 通过回译法增强的数据会一定程度上与原数据重合,随着模型训练周期增加,易出现过拟合现象
2. 仅根据附件 2 七个标签训练模型,附件 1 中完整分类体系具有十五个一级标签

改进方向:

利用网络爬虫,爬取更多分类留言内容,训练更完整的一级标签分类模型。

8.2 问题二:

优点:

1. 对词库进行地点专有名词扩充,文本分词效果较优。
2. 根据数据特点,对具有代表性的地点名词和名词进行加权,优化聚类效果。
3. DBSCAN 聚类方法可去除噪音点,本问题中噪音点表示反映人数少的主题内容。
4. 根据问政平台特点,提出热度评价三准则,建立特有的热度评价模型。

缺点:

1. 由于文本是离散型数据,所在空间属于高维度空间,降维后维度依旧具有一定高度,DBSCAN 聚类方法对高维度空间点聚类并不理想。
2. 对于重复留言现象未设立相关准则,未能代码实现。

8.3 问题三:

优点:

1. 提出属于本文的网络问政平台答复质量“3C1T”评价指标。
2. 提出各指标的数值量化方法,并给出质量评分公式。

缺点:

1. 仅提出模型,未能对各指标量化方法代码实现。
2. 各指标权重未经过数据检验,仅参考前人研究设立。

改进方向:

代码实现后,可用灰色关联法等进行权重修正,以获得更合理的质量评分公式。

九、 参考文献

- [1]郭银灵. 基于文本分析的在线评论质量评价模型研究[D].内蒙古大学,2017.
- [2]相洋. 问答系统的答案优化方法研究[D].哈尔滨工业大学,2017.
- [3]陈涛,谢丽莎.在线评论文本信息质量等级的测量探析——基于模糊综合评价法[J].科技创业月刊,2012,25(07):50-52.