

“智慧政务”中的文本挖掘应用

摘要

本文旨在近年来，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。

针对任务一，处理网络问政平台的群众留言时，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。根据给出的数据，建立关于留言内容的一级标签分类模型。并使用 F-Score 对分类方法进行评价：

$$F_1 = \frac{1}{n} \sum_{i=1}^n \frac{2P_i R_i}{P_i + R_i},$$

其中 P_i 为第 i 类的查准率， R_i 为第 i 类的全查率。

针对任务二，首先用附件给的 EXCEL 的找出排名前五的热点问题，。再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。再按照格式，对相应的热点问题给出对应的留言信息。

针对任务三，根据任务二的总结，对相关的部门的答复意见。从答复的相关性、完整性、可解释性等角度对答复意见的质量给出一套评价方案。

关键词：热点问题；F-Score；分类，归纳；评价结果；解决方案。

目录

摘要.....	1
一.问题重述.....	3
1.1 问题背景.....	3
1.2 要解决的问题.....	3
二. 问题一的分析与求解.....	4
三. 问题二的分析与求解.....	5
四. 问题三的分析与求解.....	6 与 7
五. 总结.....	8

一．问题重述

1.1 问题背景

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。

附件给出了收集自互联网公开来源的群众问政留言记录，及相关部门对部分群众留言的答复意见。请利用自然语言处理和文本挖掘的方法解决下面的问题。

1.2 要解决的问题

（1）群众留言分类

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（参考附件 1 提供的内容分类三级标签体系）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。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。建立关于留言内容的一级标签分类模型。

（2）热点问题挖掘

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。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，按表 1 的格式给出排名前 5 的热点问题，并保存为文件“热点问题表.xls”。给出相应热点问题对应的留言信息，并保存为“热点问题留言明细表.xls”。

（3）答复意见的评价

针对相关部门对留言的答复意见，从答复的相关性、完整性、可解释性等角度对答复意见的质量给出一套评价方案，并尝试实现。

二. 问题一的分析与求解

群众留言分类

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。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。根据给出的数据，建立关于留言内容的一级标签分类模型。

通常使用 F-Score 对分类方法进行评价：

β 是用来平衡 Precision, Recall 在 F-score 计算中的权重,取值情况有以下三种:

如果取 1,表示 Precision 与 Recall 一样重要

如果取小于 1,表示 Precision 比 Recall 重要

如果取大于 1,表示 Recall 比 Precision 重要

一般情况下, β 取 1，认为两个指标一样重要.此时 F-score 的计算公式为：

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

则根据附件 2 中的数据重新进行一级标签分类，然后和附件 2 的分类记过对比，计算 F1。

我们定义 卫生计生与教育文体为一级标签。

有如下过程：

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Accuracy=98.8%

Precision=49.32%

F1-score=65.46%

标记完成后的一级分类标签表格见附录中“一级标签分类”表。

三.问题二的分析与求解

热点问题挖掘

热点问题即在一段时间之内讨论次数最多，引起争议最大的问题。热度评价指标可以定义为一段时间内讨论人群的数目。在本体中可以认为是点赞数和反对数的总和。因为点赞和反对的人数越多，代表这个留言引起人们的争论越大，越能提现其为当下讨论的热点问题。

热度评价指标定义： 点赞数和反对数目之和。

表 1-热点问题表

热度排名	问题 ID	热度指数	时间范围	地点/人群	问题描述
1	1	2097	2019/8/12 至 2019/10/9	A 市 A5 区汇 金路五矿万 境 K9 县	小区群租房 泛滥成灾
2	2	1767	2019/1/18 至 2019/11/8	A 市金毛湾	A4 区 p2p 公 司 58 车贷， 非法经营近 四年
3	3	669	2017/6/8 至 2019/6/19	A4 区绿地海 外滩小区	A4 区绿地海 外滩小区距 长赣高铁不 合理高度
4	4	242	2017/6/8 至 2019/6/19	A 市富绿物 业丽发新城	A 市富绿物 业丽发新城 强行断业主 家水
5	5	80	2019/5/22 至 2020/1/8	京港澳高速 城区	建议西地省 尽快外迁京 港澳高速城 区段至远郊

另一个表格以附件（表 2-热点问题留言明细表）上传

四.问题三的分析与求解

答复意见的评价

从答复的相关性、完整性、可解释性等角度对 答复意见的质量给出一套评价方案。

答复的相关性：答复是否从 1，2，3 级相对应的分类标签中选出。

答复是否从问题的关键词入手。

答复的完整性：答复内容是否包含问题的所有关键词。

答复内容是否涵盖具有相同关键词的问题。

答复的可解释性：具有部分相同关键词的问题的答复之间逻辑是否合理不冲突。

问题的答复的逻辑在现实社会中是否可以实现。

根据相关部门对留言用户的答复意见，为了构建一个更加美好的社会环境，提高相关部门对留言用户的意见更加用心，负责的答复态度，规范化水平，为了实现社会主义核心价值观，结合相关部门的答复意见，特对答复质量实施细规。

一、评价原则

1、要实地考察，了解问题的相关性！

相关部门在接到留言用户的留言主题时，在了解了大致的情况下，在问题比较严重的情况下，应该主动联系留言用户或派遣人员进行实地考察，应及时发现问题，了解问题，积极探讨如何解决问题。结合实际情况给附近居民或有困难人员的及时解决问题。要坚持过程和结果相结合的原则，即注重过程，更加要关注结果。要督促问题的源，迅速解决对民众的困扰。

脚踏实地是一种风格，也是做事需要坚持的原则，求真务实肯干则是所需要展现出的姿态。当前的中国经历着最为广泛而深刻的社会变革和实践创新，在时代的急流浪潮中，广大干部要不忘初心，脚踏实地做事，以实际行动彰显使命担当。

“政之所兴在顺民心，政之所废在逆民心。”要牢固树立全心全意为人民服务宗旨，坚守初心，不忘人民，脚踏实地为人民干实事。将人民的切身利益作为一切工作的出发点，不为了追求政绩搞形式主义，虚假主义，大力抓落实，为民尽职尽责谋利益，带领人民走向更美好的未来。

2、要秉承一个实事求是的态度，可解释性就要求我们做事实求是！

相关部门在接到用户留言时，要秉承一个实事求是的态度。要寻求大多数人的意见，不能因为产生的问题解决过于困难或产生问题的人跟你有部分关系，而延缓处理，更有甚者是视而不见。中华人民共和国是由人民组成的社会主义国家，一切都要以人民的利益为先！随着国家政策的不断调整，工作内容也在不断发生变化，为人民服务为基本国策，其服务群众及提高人口质量的主要任务不会变。

党章中明确规定：“党的思想路线是一切从实际出发，理论联系实际，实事求是，在实践中检验真理和发展真理。”这阐述了思想路线的科学内涵。一切从实际出发，是实事求是思想路线的前提和基础；理论联系实际，是实事求是思想路线的根本途径和方法；在实践中检验真理，是实事求是思想路线的验证条件和目的。实事求是对于政党和国家之重要性不言而喻。在新中国成立前后均有一些典型事例可以证明，如革命期间从实际出发采取了农村包围城市、武装夺取政权的正确战略政策，也有因为未能坚持实事求是的原则而对国家造成了巨大损失，如“大跃进”和“文化大革命”事实上，对于个人而言，实事求是也是极为重要的处世原则，它包括尊重客观事实，不捏造事实，不讲谎言，不欺人不自欺，正视个人的缺点并积极改正等。这是我所理解的实事求是对于个人的要求。对于相关部门而言，实事求是

是也极为重要，在解决问题的过程中，它可以使我们不怕任何质疑的目光，从容应对任何人，任何地方的任何问题，拥有可解释的从容。如此，相信会受益匪浅。

3、要了解问题的全面，追求问题的完整性！

在用户举报一些问题时，不要过早的下结论，应该先制定相关计划，做出一套完整的行动计划表，相关部门再派遣人员，进行实地考察，问卷调查等方式，去了解这个问题的起因，如果反应所属正确，应迅速做出反应，给居民一个安静舒适的生活环境，如若反应有误，则应该严重批评造谣者，情节严重者，可以交给相关部门处理！

具体问题具体分析是马克思主义的一个重要原则。之所以要坚持具体问题具体分析首先在于它是人们正确认识事物的基础。世界上的事物之所以千差万别，就在于它们各有其特殊的矛盾，只有具体问题具体分析矛盾的特殊性，才能把不同质的事物区别开来。其次，具体问题具体分析是正确解决矛盾的关键。正确认识事物的目的是为了正确解决矛盾，事物的矛盾各不相同，事物的矛盾各不相同，解决矛盾的方法也不可能千篇一律。只有具体问题具体分析，才能找到解决矛盾的正确方法。

要做到具体问题具体分析，一是要承认矛盾的普遍性和客观性。事事有矛，时时有盾，这就是要求我们再认识事物和解决问题是，只有承认矛盾的普遍存在，才能进一步对矛盾作具体分析。二是要注意研究事情的特点、本质以及该事务存在的具体条件。不同的事情具有不同地特点，要做到具体问题具体化。坚持一切从实际出发，是做到具体问题具体化分析的前提。

二、评价方式

1、分级评价

给问题的严重情况，分为几个等级，优先处理评价等级高的问题。

2、分片评价

在同一块地区的问题，可以同时进行解决，可以最快的处理问题。

3、分组评价

将有处理不同问题经验的人员分在同一解决问题小组，这样在遇到不同问题或复杂问题可以在各自擅长的领域处理问题，互补！

三、评价内容

1、民事纠纷

小区环境，噪音，家庭问题等

2、刑事案件

建筑偷工减料，财务失窃等

3、环境问题

退草还林，野生动物保护等

4、商贸旅游

景区乱收费，导游强制买卖等

5、交通

超载，乱停乱放等

四、评价实施

1、留言板

设立不同问题，不同留言的板块或不同地区分板块

2、设不同的小组

在每个区设一个小组，在能处理的问题上，小组出面解决问题，比较严重的问题则上报交给上一级的小组解决！小组的成员则分共解决不同问题！

3、快速解决问题

在用户提出问题时，如若问题的情况较为紧急，应该快速组织应急小组，快速讨论并解决问题，不紧急的情况下，最短也要在 30 个工作日内解决问题！

4、每月一总

每个月总结一次，找出问题严重点，重点督促和解决此地方的问题。

总结

本文的主要目的是对收集的互联网公开来源的群众问政留言记录，及相关部门对部分群众留言的答复意见，按照一定的划分体系对留言进行分类，并用 **F-Score** 对分类方法进行评价。再根据一段时间内讨论人群的数目的多少找出来热点问题，并根据归纳分类，给相应热点问题给出对应留言信息。最后，我们根据相关部门对留言的答复意见，从答复的相关性、完整性、可解释性等角度给出了一套评价方案，并给出执行方案和分类总结！