

C 题

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。

第一题

#出自 GitHub 里面的 py 开源项目 BERT 文本分类

调用过程，设置成自己的环境

```
export BERT_BASE_DIR=/home/chenz/BERT/chinese_L-12_H-768_A-12
```

```
export GLUE_DIR=/home/chenz/BERT-train2deploy/dat
```

```
export TRAINED_CLASSIFIER=/home/chenz/BERT-train2deploy/output
```

训练

```
python3 run_mobile.py \
```

```
--task_name=setiment \
```

```
--do_train=true \
```

```
--do_eval=true \
```

```
--data_dir=$GLUE_DIR \
```

```
--vocab_file=$BERT_BASE_DIR/vocab.txt \
```

```
--bert_config_file=$BERT_BASE_DIR/bert_config.json \
```

```
--init_checkpoint=$BERT_BASE_DIR/bert_model.ckpt \
```

```
--max_seq_length=32 \
```

```
--train_batch_size=16 \
```

```
--learning_rate=2e-5 \
```

```
--num_train_epochs=5.0 \
```

```
--output_dir=$TRAINED_CLASSIFIER
```

测试

```
python3 run_mobile.py \
```

```
--task_name=setiment \
```

```
--do_predict=true \
```

```
--data_dir=$GLUE_DIR \
```

```
--vocab_file=$BERT_BASE_DIR/vocab.txt \
```

```
--bert_config_file=$BERT_BASE_DIR/bert_config.json \
```

```
--init_checkpoint=$TRAINED_CLASSIFIER \
```

```
--max_seq_length=32 \
```

```
--output_dir=$TRAINED_CLASSIFIER
```

model_dir 就是训练好的.ckpt 文件所在的目录

max_seq_len 要与原来一致：

num_labels 是分类标签的个数,本例中是 3 个

```

# 模型文件转化
python3 freeze_graph.py \
    -bert_model_dir $BERT_BASE_DIR \
    -model_dir $TRAINED_CLASSIFIER/$EXP_NAME \
    -max_seq_len 32 \
    -num_labels 18
# 启动 bert
bert-base-serving-start \
    -model_dir $TRAINED_CLASSIFIER \
    -bert_model_dir $BERT_BASE_DIR \
    -model_pb_dir $TRAINED_CLASSIFIER \
    -mode CLASS \
    -max_seq_len 32 \
    -http_port 8091 \
    -port 65331 \
    -port_out 65332 \
    -device_map 1
# 调用
from bert_base.client import BertClient
classifier = BertClient("192.168.1.158", 65331, 65332, show_server_config=False,
check_version=False, check_length=False, mode='CLASS')classifier.encode(["农村人的歌，我家在
农村，非常喜欢的一首歌，这次我把一整首都唱完了，你也听听看"])

```

第二题

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	2019/8/19 11:34:04	A 市 A5 区汇金路五矿万境 K9 县	小区群租房泛滥成灾，擅自修改毛坯房，穿内裤随意走动，搬煤气做饭，宠物随便走动，车辆随便进入，逃生楼梯安排不合理
2	2	1762	2019/4/11 21:02:44	A 市金毛湾	金毛湾配套入学的问题
3	3	821	2019/2/21 18:45:14	A 市 A4 区 58	A4 区 p2p 公司 58 车贷，非法经营近四年，公安局未发一字立案公告和案件进展财产处置通报
4	4	790	2019/2/25 9:58:37	A 市 58	车贷特大集资诈骗案保护伞

					<p>这样安排合理？电摩充电不是都要遭殃？24栋25栋楼，在负一楼和一楼中间，然后一旦发生火灾，请回复，还正政府出面帮有让我们安居了我们才能</p>
2	223297	A00087522	反映 A 市金毛湾配套入学的问题	2019/4/11 21:02:44	<p>书记先生：您好！我是梅溪湖金毛湾业主，因为当初金毛湾没有交房，待今年上半年早一点安排配套和西雅中学两所学校之内近。得到缪副局长的解释，期盼着今年出具配套入学单位回复网友业主（hn项目交付时间为2018年12月，将金毛湾楼盘纳入配套入学汇报，明确于2018年7月梅溪湖中学或西雅中学配套入学，日仍然没有相关配套入学教育局咨询得到的结果是明年或以后不能保证，配套的配套是开发商给你们的样没有交房的建发央筑同区，金毛在梅溪湖的诸多还没有着落，现在又到开学的学位问题很是担心。恳请金毛湾2800户业主担心的业主</p>
3	220711	A00031682	请书记关注 A 市 A4 区 58 车贷案	2019/2/21 18:45:14	<p>尊敬的胡书记：您好！本人经营近四年。在受害人要求下，至今已6个月整。未发一处置通报，全国仅此一家许多人一生积蓄出借在平案情消息总是失望，四处</p>

					<p>违背两高一部《指导意见》</p> <p>处置，维护受害人权利的</p> <p>百忙中给予关注，督促 A</p> <p>展通报，追赃挽损成果，</p> <p>以安民心！谢谢！</p>
4	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	<p>胡市长：您好！西地</p> <p>车贷 https://baidu.com/</p> <p>贷爆雷，其法定代表人、</p> <p>（化名邢 ze）（夫妻关系</p> <p>58 车贷受害者向 A 市 A4 区</p> <p>侦毛钧回复没有办案经费</p> <p>8 月 20 日，全国各地 100</p> <p>门口，A4 区公安分局才被</p> <p>立案，但并不给报案人立</p> <p>承诺：办案会通报，重大</p> <p>其承诺无一实现。A4 区公</p> <p>除了冻结部分银行账户，</p> <p>车子、证券等其他资产，</p> <p>9 月份，查封了 58 车贷的</p> <p>现，该办公室已经解封，</p> <p>9 月份就说要通缉外逃的</p> <p>续，未通缉，对留在国内</p> <p>办案民警毛骏还为他们一</p> <p>年才进来的，进来后对 58</p> <p>管系统，对其发假标的事</p> <p>罪，还有功；说大股东苏</p> <p>是挂名；说担保公司法定</p> <p>的，但受害人查证，苏吕</p> <p>雄在 58 有高额工资，两</p> <p>的资金往来；羊毛头子孙</p> <p>毛骏说他是出借人。毛骏</p> <p>多的诈骗案都是出逃美国</p> <p>所有中国人，不管是外逃</p> <p>的苏吕、刘顺、陈杰雄、</p> <p>未控制人员，未查封资产</p> <p>修改网站，利用网站进行</p> <p>号，多次组织、操控伪代</p> <p>小、小事化了。涉案人员</p> <p>威胁说不投票视为投票，</p> <p>向毛骏反映，毛骏不仅不</p> <p>受害人发现，毛骏跟羊毛</p>

					多次去毛骏办公室，出来 利于涉案人员行动的消息 西地省体育局子弟，对 58 者怀疑西地省官场有苏纳 此举报，希望调查 58 车贷 败、违法行为，打掉保护
5	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	胡书记：您好！58 车贷案 也引起市领导的重视，公 人深受感动，也看到了希 没有跟进市领导的留言， 计报告还是没有公布。鉴 案，督促 A4 区经侦，领会 留言。

第三题

一、定目标

1、展现现有留言答复状况

附件四中有 121 例相关部门给出的留言答复意见，在这 121 例留言意见中，大部分的留言回复都是切合问题的实际的，只有少部分的回答模糊不清。

2、寻找当前留言答复的问题

在当前的解决了的 121 条留言答复中，据不完全统计相关部门给的大概的回复时间为 15 天，在这 121 条留言答复中相关部门给出的留言答复大部分都是紧紧扣住了群众的提问，只有少数几个问题给出的答复含糊不清。

3、预测留言答复发展趋势

随着互联网的推进，我国的网民数量也越来越多，已经实现了全省近 40%的群众都已经上网的实际，网络已经成为群众反映诉求的重要渠道。习近平总书记在网络安全和信息化工作座谈会中提出：“网民来自老百姓，老百姓上了网，民意也就上了网。群众在哪儿，我们的领导干部就要到哪儿去。各级党政机关和领导干部要学会通过网络走群众路线，经常上网看看，了解群众所思所愿，收集好想法好建议，积极回应网民关切、解释疑惑”。随着互联网的普及，网络已经成为广大人民群众了解政策、反映问题、建言献策的重要渠道，同时也是政府机构听民声，解民意，更好服务于人民群众的的新型理政方式。现如今，通过网上留言、网络问政平台、市长信箱投诉等方式，已经成为老百姓行使知情权、参与权、表达权和监督权的重要渠道，网络留言是反映千千万万百姓的心声，各级领导应高度重视网络舆论场，经常上网浏览网民留言，走好“群众路线”。回复并解决好每一条网络留言，在新形势下做到

为人民服务的宗旨。通过网络问政,及时了解实事,掌握多方面信息,积极回应,提高党政机关的工作效率,提高政府信息畅通透明度。同时也将一些懒政怠领导干部“显露原形”,人习惯于打官腔、官僚主义的干部不再有容身之地,在网络面前,有无数只眼睛在看着、监督着,人人都能成为一名新闻记者。

4、找出下一步工作的方向

网民的留言工作中,我们主要方向分为三点:

- ①、强化组织,工作运转高效有序。办好网上留言,需要一个健全的工作网络,有一个好的处理机制作为基本保证。目前,网上留言办理工作制度健全,程序规范,运行高效。
- ②、充实人员力量,健全工作网络。使得工作人员的经费得到保证,这些都是为了办好网民留言创造了良好条件。
- ③、制定和形成了规范的工作运行程序。这样使得每一个环节都有明确的工作要求,都有专人负责,形成了规范的处理机制,保证了工作的高效有序开展。

二、使用方法

关于留言答复意见的评价涵盖了各类不同学科的相关理论知识和分析方法,比如多元统计分析方法(其中又包括判别分析、聚类分析、因子分析、相应分析和典型相关分析等各种方法)、模糊综合评价法、灰色关联综合评价法、数据包络分析法和神经网络分析法等多种常用研究方法。下面对常用的评价方法进行梳理介绍:

1、多元统计学方法主要是利用数理统计方法解决多指标问题,通过计算指标间的关联性,抽象出隐藏在指标内的潜在信息,即从大量的评价指标中抽取少而精的概括性指标,利用这些指标完成综合评价。通过多元统计不仅能够对指标进行遴选并指标权重进行赋值,还能够按照应用需求进行相应的综合评价。不过,多元统计方法的应用条件较为苛刻,统计量要服从典型的概率分布,目标数据与各指标数据彼此为线性关系,各指标数据之间还要相互无关,而且进行元统计分析也需要较大的样本数据量,因此在分析实际情况中常常难以达到其要求。

2、模糊综合评价主要是将评价矩阵与权重进行模糊合成,得到结果向量从而完成综合评价的过程。步骤主要如下:首先,构建模糊综合评价指标。其次,根据评价因素的具体情况确定评价等级;再次,构建权重向量,建立模糊关系矩阵;第四,计算模糊合成值;最后,对评价目标执行模糊综合评价。模糊综合评价法的显著特点主要是以模糊数学的隶属度作为评价依据,以最优的评价因素值为基准。

3、数据包络分析法(简称 DEA)又叫做效率评价方法,属于线性规划的研究范围。若想将 DEA 运用在综合评价中,评价体系应满足如下条件:首先,评价体系中应该包含投入单元和产出单元,因为 DEA 是对多投入多产出的多个决策单元进行效率评价的主要方法,通过分析产出与投入的比率,清晰地说明各种投入产出组合,也因此常用于经济效益和运营效率的分析。其次,DEA 分析同样要求有足够的样本量,这是 DEA 分析可行性的基本条件。

4、BP 神经网络方法是一种非线性、自适应、自组织的网络,由大量简单的处理单元组成。BP 神经网络的结构主要是一个输入层、一个或多个中间隐含层和一个输出层。每一层都有若干个网络节点即神经元,处于同一层的神经元互相不存在连接关系。神经元可以通过传递函数和加权系数改变,上一个神经元对下一个神经元产生的刺激。BP 神经网络通过不断学习,可以从未知模式的大量复杂数据中发现规律,最终确定网络权值。权值确定之后,进行相应的数据处理,可得到输出

结果。

5、灰色系统理论主要研究少数据和贫信息的不确定性问题,希望能够通过科学手段,找出子因素间隐含的关系。灰色系统理论中的灰色层次分析主要是将评价体系内的不同影响因素按顺序分成多个层次,不同层次之间存在相互联系,按层次计算出各因素指标的关联度,最后对不同层次的关联度进行比较。

通过上述介绍我们可以发现,不同方法所使用的评价原理各不相同,适合的评价环境以及对原始数据的要求也各有不同。综合比较之后,灰色关联评价主要有如下六个特点: (1)对不同因素间的发展走向进行分析; (2)本质、上是定量分析,但是要以前期的定性分析为指导; (3)不要求样本数据量的大小; (4)不需要经验数据辅助分析; (5)不要求样本数据服从某一分布类型; (6) 处理过程简单易懂。本文建立的留言答复意见评价指标体系属于结构相同的多层指标体系, 样本量较大,因此我们可以采用灰色关联分析方法从系统科学中研究两个系统之间的关系发展。

三、评价标准

我们对于给出来的数据可以了解到,留言的回复是对过去、现在或者未来的情况做出评价或预测的,这样能否获得我们所需要的足够的可以理解的信息,比如反映出来的数据能否理解。我们先把反映出来的问题进行分类,然后每个分类群众反映的意见有多少,下面就是对相关部门给出的答复意见进行处理和分析。