基于自然语言处理技术的
“智慧政务”中的文本挖掘研究

摘要

近年来，随着互联网的发展，网络问政成为国家与人民沟通的方式之一，也是人民监督政务的途径之一，但由于技术的限制，群众留言分类及热点问题挖掘等问题尚需人工处理，加大了政府工作量，基于等问题，使用自然语言处理技术以及文本挖掘，自动文摘，以及文本自动分类等技术建立了关于群众留言的一级标签分类模型，利用SPSS 及 EXCEL 等软件，对于热点问题挖掘更加精准化，也建立了留言评价方案，对于以及标签分类模型使用 F-score 的评价方法对模型进行评价。

关键词：网络问政 自然语言处理 文本挖掘

目录

1. 问题重述

1.1 问题背景

近年来，随着互联网应用的飞速发展，文本数据库得以迅速增长，人们迫切需要有效的数据挖掘工具从海量文本数据中提取有价值的知识；另一方面，随着电子政务的进一步发展，政府部门内部及政府部门之间产生了大量政务信息，并且微信、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。

但经过近些年来的电子政务基础资源的大规模建设，海量政务信息资源挖掘和电子政务知识管理等深层次应用正逐步进入电子政务舞台，对电子政务实施数据挖掘将成为政府信息化的一个新的研究方向。与此同时，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。

1.2 问题重述

根据附件所给出的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。需利用自然语言处理和文本挖掘的方法解决下面的问题：

（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。请根据附件 2 给出的数据，建立关于留言内容的一级标签分类模型。

（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。请根据附件 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，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”。按表 2 的格式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”。

表 1.1 热点问题

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	…	2019/08/18 至 2019/09/04	A 市 A5 区魅力之城小区	小区临街餐饮店油烟噪音扰民
2	2	…	2017/06/08 至 2019/11/22	A 市经济学院学生	学校强制学生去定点企业实习
…	…	…	…	…	…

表 1.2 热点问题留言明细

问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	点赞数	反对数
1	360104	A012417	A 市魅力之城商铺无排烟管道，小区内到处油烟味	2019/08/18 14:44:00	A 市魅力之城小区自打交房入住后，底层商铺无排烟管道，经营餐馆导致大量油烟排入小区内，每天到凌晨还在营业……	0	0
1	360105	A120356	A5 区魅力之城小区一楼被搞成商业门面，噪音扰民严重	2019/08/26 08:33:03	我们是魅力之城小区居民，小区朝北大门两侧的楼栋下面一楼，本来应是架空层，现搞成商业门面，嗓声严重扰民，有很大的油烟味往楼上窜，没办法居住……	1	0
1	360106	A235367	A 市魅力之城小区底层商铺营业到凌晨，各种噪音好痛苦	2019/08/26 01:50:38	2019 年 5 月起，小区楼下商铺越发嚣张，不仅营业到凌晨不休息，各种烧烤、喝酒的噪音严重影响了小区居民休息……	0	0
…	…	…	…	…	…	…	…
1	360109	A0080252	魅力之城小区底层门店深夜经营，各种噪音扰民	2019/09/04 21:00:18	您好：我是魅力之城小区的业主，小区临街的一楼是商铺，尤其是餐馆夜宵摊等，每到凌晨都还在营业，每到晚上睡觉耳边都充斥着吆喝……	0	0

续表 1.2 热点问题留言明细

2	360110	A110021	A市经济学院寒假过年期间组织学生去工厂工作	2019/11/22 14:42:14	西地省A市经济学院寒假过年期间组织学生去工厂工作，过年本该是家人团聚的时光，很多家长一年回来一次，也就过年和自己孩子见一次面，可是这样搞……	0	0
2	360111	A1204455	A市经济学院组织学生外出打工合理吗？	2019/11/5 10:31:38	学校组织我们学生在外边打工，在东莞做流水线工作，还要倒白夜班。本来都在学校好好上课，十月底突然说组织到外省打工……	1	0
…	…	…	…	…	…	…	…
2	360114	A0182491	A市经济学院变相强制实习	2017/06/08 17:31:20	系里要求我们在实习前分别去指定的不同公司实训，我这的公司的工作内容和老师之前介绍以及我们专业几乎不对口，不做满6个月不给实训分，不能毕业……	9	0

（3）答复意见的评价

针对附件 4 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。

2. 数据分析

2.1 数据预处理

(1)文本清洗

①去除链接地址

链接地址需要在进一步分析前被去掉，可以使用正则表达式实现去除。

②去除停用词

停用词是在每个句子中都很常见，比如英语中的“is”、“but”、“shall”、“by”，汉语中的“的”、“是”、“但是”等。从数据分析角度看，停用词没有意义。语料中的这些词可以通过匹配文本处理程序包中的停用词列表来实现去除。

③去除标点符号

标点符号显然对文本分析没有帮助，因此需要去除。

④去掉空白符

可以使用正则表达式去掉词前后的空白字符，只保留词本身。

⑤去除特殊字符

在进行了去除空白字符、数字和标点符号等操作后，一些形式特殊的链接地址等额外内容可能仍然未被去除，需要对处理后的语料再进行一次检查，并用正则表达式去除它们。

最后在除去数据中非文本部分时,少量的非文本内容的可以直接用 Python 的正则表达式(re)删除， 复杂的则可以用 beautiful soup 来去除。去除掉这些非文本的内容后，我们就可以进行接下来的文本预处理。

(2)分词

首先要做的预处理就是分词。英文单词天然有空格隔开容易按照空格分词，但是也有时候需要把多个单词作为一个分词，比如一些名词如“New York”，需要作为一个词看待。而中文由于没有空格，分词就是一个需要专门去解决的问题了。

通过标准语料库，可以近似的计算出所有的分词之间的二元条件概率，比如任意两个词 w_1， w_2，它们的条件概率分布可以近似的表示为：

$$P(w_2 \mid w_1)=\frac{P(w_1,w_2)}{P(w_1)} \approx \frac{freq(w_1,w_2)}{freq(w_1)}$$

$$P(w_1 \mid w_2)=\frac{P(w_2,w_1)}{P(w_2)} \approx \frac{freq(w_1,w_2)}{freq(w_2)}$$

其中 $freq(w_1,w_2)$ 表示 w_1，w_2 在语料库中相邻一起出现的次数，而其中 $freq(w_1)$，$freq(w_2)$ 分别表示 w_1，w_2 在语料库中出现的统计次数。

利用语料库建立的统计概率，对于一个新的句子，可以通过计算各种分词方法对应的联合分布概率，找到最大概率对应的分词方法，即为最优分词。

同样地，再次利用维特比算法中采用的动态规划来解决这个最优分词问题的，动态规划要求局部路径也是最优路径的一部分，很显然我们的问题是成立的。首先我们看一个简单的分词例子：″人生如梦境″。它的可能分词可以用图 2.1 表示。

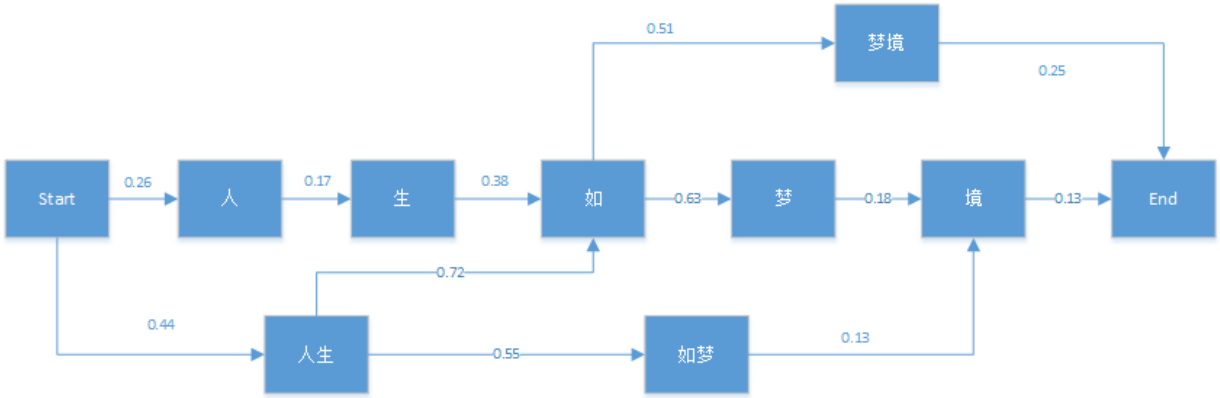

图 2.1 可能分词的表示

(3)文本转化

①生成词向量词典

可使用 word2vec 算法对分词后的语料库进行预训练生成词向量词典。其中的字词是不重复的，word2vec 会对文本中的字、词和标点基本元素的出现频率进行统计，通过无监督训练，获得作为基础语料基础构成元素的字词对应的指定维度的向量表征。

②文本数字化

将分词后语料库文本中对应的字词和①中词向量词典经行对比，获其索引，即在词向量词典对应的序号，这样文档中都以整数索引序号表示从而实现索引形式的数字化，有利于降低文本表示的数据维度。

③文本向量化

在数据进入模型训练前需按照词的索引序号从①中生成的词向量词典取出其对应的向量，这样附件中的文本被转化为向量的形式。

2.2 名词注释

为更明确所要解决的问题和问题分析的过程，对以下名词做一定的注释。

(1)网络问政

随着网络的日益普及，互联网在中国民众的政治、经济和社会生活中扮演着日益重要的角色，中国公民以网民的身份通过互联网行使知情权、参与权、表达权和监督权，这就是网络问政。

(2)自然语言处理

自然语言处理是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。因此，这一领域的研究将涉及自然语言，即人们日常使用的语言,所以它与语言学的研究有着密切的联系,但又有重要的区别。自然语言处理并不是一般地研究自然语言，而在于研制能有效地实现自然语言通信的计算机系统，特别是其中的软件系统。因而它是计算机科学的一部分。

(3)文本挖掘

文本挖掘是指从大量文本数据中抽取事先未知的、可理解的、最终可用的知识的过 程，同时运用这些知识更好地组织信息以便将来参考。直观的说，当数据挖掘的对象完全由文本这种数据类型组成时，这个过程就称为文本挖掘。文本挖掘也称为文本数据挖掘。

(4)查准率

Precision(精度)是衡量某一检索系统的信号噪声比的一种指标，即检出的相关文献与检出的全部文献的百分比。普遍表示为：查准率=（检索出的相关信息量/检索出的信息总量）x100%。

(5)查全率

查全率是衡量某一检索系统从文献集合中检出相关文献成功度的一项指标，即检出的相关文献与全部相关文献的百分比。

3. 问题分析与求解

3.1 问题一的分析与求解

近年来，随着科技的不断发展，互联网全面渗入到各个领域中。智慧政务应运而生，政府利用互联网将政务公开化处理，在阳光下接受人民的监督与建议，实现政府管理与公共服务的精细化、智能化、社会化，实现政府和公民的双向互动，但是对于智慧政务中的公众留言问题，依赖于人工进行留言分类，存在工作量大、效率低，且差错率高等问题，针对这一问题，预建立一个一级标签分类模型。

3.1.1 一级标签分类模型

(1)运用自动文摘技术形成相应标签

首先对于留言要进行主要内容提取，这就要运用到自动文摘技术，所谓自动文摘就是利用计算机程序自动地从文本中提取文摘，文摘是全面准确地反映某一文本中心内容的简单连贯的文字，就相当于新闻中的摘要；自动文摘技术主要有机械文摘和理解文摘两种，机械文摘能够适用于非受限域，符合当前自然语言处理技术面向真实语料、面向实用化的总趋势，理解文摘虽牺牲领域宽度，但换取了理解深度。

自动文摘分为基于统计的自动文摘、基于理解的自动文摘、基于信息提取的自动文摘、基于结构的自动文摘。对于群众留言分类，采取的是基于理解的自动文摘，对此分四步进行。

①语法分析

借助词典中的语言学知识对原文中的句子进行语法分析，获得语法结构树。

②语义分析

运用知识库中的语义知识将语法结构描述转换成逻辑和意义为基础的的语义表示。

③语用分析和信息提取

根据知识库中预先存放的领域知识在上下文中进行推理，并将提取出来的关键内容存入一张信息表。

④文本生成

将信息表中的内容转换成一段文本。

自动文摘是自然语言处理中较难的一部分，对此首先要利用词频模型进行辅助分析。TF-IDF(term frequency-inverse document frequency)词频-反转文件频率,是一

种用于情报检索与文本挖掘的常用加权技术，用以评估一个词对于一个文本或者一个语料库中的一个领域文件集的重复程度。基于词频模型的基础上，留言内容进行文本清洗之后，可利用 abstractive 式的 seq2seq 模型。seq2seq 的模型一般都是如下的结构：

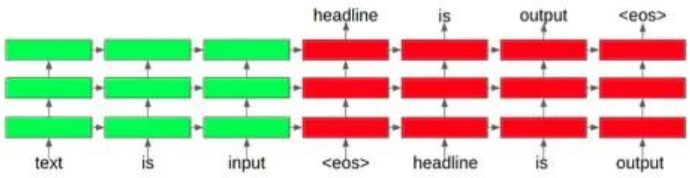

Figure 1: Encoder-decoder neural network architecture

图 3.1 seq2seq 的模型

encoder 部分用单层或者多层 rnn/lstm/gru 将输入进行编码，decoder 部分是一个语言模型，用来生成摘要。这种生成式的问题都可以归结为求解一个条件概率问题 p(word|context)，在 context 条件下，将词表中每一个词的概率值都算出来，用概率最大的那个词作为生成的词，依次生成摘要中的所有词，形成留言相应标签分类。

(2)运用多标签文本自动分类将结果精准化

由于留言内容多样化，导致在自动文摘时可能会形成多个标签，比如某个留言中即包含政治方面的内容又包含城乡建设方面的内容那么便会形成两个标签，接下来要解决的便是多标签文本自动分类问题。

文本自动分类即在给定的分类体系下，根据文本的内容用计算机程序确定文本所属类别的过程。一般采用机器学习的方法进行自动文本分类，即基于训练集的自动文本分类。文本分类也属于文本挖掘的基本范畴，从数学的角度而言，文本自动分类相当于映射，而精确化在于映射是多对一但不可以一对多，从而实现留言的精准分类，文本分类的映射规则是计算机系统根据现有的数据库对文本类别区分形成一定的分类规则，在此规则上对于新的文本进行分类。

对于多标签文本的分类，需构建基于算法适应的多标签分类模型，对现有的软模糊粗糙集模型进行改造用于多标签文本的分类。

假设文本数据库有 M 个标签集，样本集合 X 表示待分类样本，Y 表示某样本所属标签集合，通过以下步骤得到样本 X 的标签类别。

①输入多标签数据集，对其类标签的表示形式进行处理，“1”表示属于，“0”表示不属于。

②假定样本属于每个类，根据软模糊下近似隶属的计算方法，得出 M 个样本 X 对每个类的近似隶属度的值。

③给出一个界限，划分出对样本 X 有着较高贡献度的类别。

④输出这些类别的集合，就是样本 X 的标签。

3.1.2 使用 F-Score 对模型进行评价

对于一级标签分类模型采用 F-score 方法进行评价。

(1)计算公式

$$\text{F-score}=(1+\beta^2)\frac{Precision*Recall}{(\beta^2*Precision)+Recall}$$

(2)Precision(精确率)

$$\text{Precision}=\frac{TP}{TP+FP}$$

- TP(真阳性):正样本被正确预测为正样本
- FP(假阳性):负样本被错误预测为正样本
- TN(真阴性):负样本被正确预测为负样本
- FN(假阴性):正样本被错误预测为负样本

(3)Recall（召回率）

$$\text{Recall}=\frac{TP}{TP+FN}$$

(4)β值的介绍

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

(5)结果预测

结果分为准确预测与错误预测,按照下面的公式求预测准确率,用这个值来评估模型准确率。

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+TN+FN}$$

3.2 问题二的分析与求解

首先将附件 3 中的非结构化数据进行去重去空、中文分词及停用词过滤等数据预处理，然后基于 TFIDF 权重法提取 500 个候选特征词，形成词袋，构造词汇文本矩阵，由于这种方法具有高维度，高稀疏度以及同义词影响的缺点，因此，进一步利用基于潜在语义（LSA）分析的奇异值分解算法（SVD）对词汇-文本矩阵进行空间语义降维，语义压缩后的文本向量被认为投影在了同一空间里，再通过 k- means 聚类算法对职位的职业类型和专业领域进行划分。

3.2.1 挖掘热点问题

将附件 3 中结构化的数据数值化处理后，运用主成分分析法构建综合排名算法得出各个话题热门指标的排名进而制作出表 3.1、表 3.2 和表 3.3。

表 3.1 热点问题

问题排名	问题ID	热度指数	时间范围	地点/人数	问题描述
1	1	18	2017/6/8至2019/11/22	A市经济学院	学校强制学生去定点企业实习
2	3	13	2019/7/21至2019/9/25	A5区劳动东路魅力之城小区一楼的夜宵摊	严重污染附近的空气
3	2	11	2019/7/28至2019/9/10	A5区魅力之城小区临街门面	油烟扰民
4	4	4	2019/8/26至2019/9/4	A5区魅力之城小区一楼	噪音扰民严重污染附近的空气
5	5	3	2018/10/27	A市地铁2#线	在梅溪湖CBD处增设一个站

表 3.2 热点问题留言明细_1

问题ID	留言编号	留言用户	留言主题	留言时间	留言详情	点赞数	反对数
1	360114	A0182491	A市经济学院体育学院变相强制实习	2017/6/8 17:31	书记您好，我是来自西地省经济学院体育学院的一名即将大四的学生，系里要求我们在	9	0
1	360110	A110021	A市经济学院寒假过年期间组织学生去工厂工作	2019/11/22 14:42	关于西地省A市经济学院寒假过年期间组织学生去工厂工作，过年本该是家人团聚的时	0	0
1	360112	A220235	A市经济学院强制学生实习	2019/4/28 17:32	各位领导干部大家好，我是A市经济学院的一名学生，临近毕业，学校开始组织学生参加	0	0
1	360113	A3352352	A市经济学院强制学生外出实习	2018/5/17 8:32	A市经济学院强制1响电子商务跟正业物流专业实习。其中我们企业物流专业实习6个月，	3	0
1	360111	A1204455	A市经济学院组织学生外出打工合理吗？	2019/11/5 10:31	一名中职院校的学生，学校组织我们学生在外边打工，在外省做流水线工作，还要倒白夜班	1	0
2	360103	A0012425	A5区劳动东路魅力之城小区临街门面烧烤夜宵摊	2019/9/25 0:31	A5区劳动东路魅力之城小区临街夜宵摊、烧烤摊24小时经营，油烟扰民。违法经营。本	1	0
2	360107	A0283523	A5区劳动东路魅力之城小区一楼的夜宵摊严重污染	2019/7/21 10:29	局长：　　你好，A5区劳动东路魅力之城小区的一楼，开了几家夜宵快餐店，厨房内多	3	0
2	360108	A0283523	A5区劳动东路魅力之城小区一楼的夜宵摊严重污染	2019/8/1 16:20	局长：　　你好，A5区劳动东路魅力之城小区的一楼，开了几家夜宵快餐店，厨房内多	6	0

表 3.3 热点问题留言明细_2

3	360100	A324156	魅力之城小区临街门面油烟直排扰民	2019/9/5 12:29	魅力之城小区楼下烧烤摊、快餐店无证经营，长期油烟烧烤重死人。一天24小时都是烟。请政府关问处理这个地方没有烟道，不	3	0
3	360101	A324156	A5区劳动东路魅力之城小区油烟扰民	2019/7/28 12:49	尊敬的政府：A5区劳动东路魅力之城小区临街门面长期油烟直排。长期投诉无果。由于各部门互相扯皮，导致油烟机清洗也没有	4	0
3	360104	A012417	A市魅力之城商铺无排烟管道，小区内到处油烟味	2019/8/18 14:44	A市魅力之城小区自打交房入住后，底层商铺无排烟管道，经营餐馆导致大量油烟排入小区内，每天进出都搞得业主一身油烟味，而	0	0
3	360102	A1234140	A5区劳动东路魅力之城小区底层餐馆油烟扰民	2019/9/10 6:13	A5区劳动东路魅力之城小区，底层有几家餐管，油烟严重影响我们居民的生活，(衣服晾晒处无有油烟味，窗户打开通风就有油烟扑	0	0
4	360106	A235367	A市魅力之城小区底层商铺营业到凌晨，各种噪音好痛苦	2019/8/26 1:50	2019年5月起，小区楼下商铺越发嚣张，不仅营业到凌晨不休息，各种烧烤、喝酒的噪音严重影响了小区居民休息，并且烧烤炒菜的	0	0
4	360105	A120356	A5区魅力之城小区一楼被搞成商业门面，噪音扰民严重	2019/8/26 8:33	我们是魅力之城小区居民，小区朝北大门两侧的楼栋下面一楼，本来应是架空层，现搞成商业门面，噪声严重扰民，有很大的油烟味往楼上窜，没办法居住，影响我们的晚午	1	0
4	360109	A0080252	万科魅力之城小区底层门店深夜经营，各种噪音扰民	2019/9/4 21:00	您好： 我是万科魅力之城小区的业主，小区临街的一楼是商铺，尤其是餐馆夜宵摊等，每到凌晨都还在营业，每到晚上睡觉耳边都充斥着	0	0
5	286572	A23525	请求A市地铁2#线在梅溪湖CBD处增设一个站	2018/10/27 15:13	领导好：今天游了A5区山，往A5区山下湖入内堵了半个小时才停好车。联想到上周末在梅溪湖游玩的拥堵，心里突然有个想法：如果地铁2#线在商业综合体或在梅溪湖CBD商	3	0

3.2.2 评价挖掘指标

首先对附件 3 的数据进行预处理后得出相关热点问题，再根据关键词对热点问题进行分类归纳，同时每个热点问题的点赞数和反对数也是评价该问题热度的指标之一，一个点赞数即代表一个人同意该用户的留言，反之，反对数即代表不同意该用户的留言。最终所有“智慧政务”问题的热度取决于留言人数与点赞数。

3.3 问题三的分析与求解

针对附件 4 中各相关部门对留言的答复意见，从留言答复的评价目的、评价的现实意义、如何评价才对市民有益处等方面出发，从答复的时间和答复的相关性、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。

首先，将附件 4 中的留言时间与答复时间进行对比，观察出答复时间的间隔大小，再与问题二中归纳的热点问题做比较，看答复时间是否能够对应热点问题而有所缩短；其次，基于问题一中的一级标签分类模型，对答复意见也进行同样的分类，初步判断答复意见的相关性强弱；再者，将附件 4 中答复意见的具体内容进行文本检索，识别出“您好”、“针对该问题”、“现给予回复”、“您的留言已收悉”等字眼，检测答复意见内容的完整性，并给出一套完整的答复格式；最后，根据答复意见中内容中的相关解释，评价答复意见的优劣性。

依据各参与问政部门单位对群众网络留言的重视程度、回复办结情况、群众满意率情况和政务公开信息发布情况综合评价，以及以上的评价方案，将对此做一定数量的测试，并给予测试结果。

3.3.1 评价方案的建立

(1)评价目的

在网络已经深入千家万户的当今社会，网络也必将成为政府与群众沟通最通畅的渠道，网络问政必将成为一种常态，然而，网络问政能不能做到有始有终？此问题不仅有关网络问政的成效，更是有关政府公信，这值得网络问政组织者重视。百姓提出的问题不能“说了也白说”，领导的表态更不能直视几句口头禅，说过就完事，百姓提出的问题、长期未能解决的民生难题，有没有整改，整改效果如何，还得继续在网络上接受行业代表、网民的现场提问。因此，对网络问政中留言的答复意见进行评价是十分有必要的。

(2)评价现实意义

信息技术的发展促进了网络政的渠道的丰富化，网络问政作为政治参与主体通过互联网等信息平台，对公共事务公开表达其利益诉求，监督政府行为，直接或问接的影响政府政策和决定、政府的运行方式等的公共政治行为。网络问政的广度、深度和宽度不断扩展，网络问政的议题涉及公共事务的诸多方面，公众的参与行为影响着政府的政策制定和执政方式。

网络问政属于一种网络政治，是政治民主化在网络政治中的延伸。网络问政的本质是现实民意在互联网空间的映射。政府回应作为政府与公众关系的核心环节，政府积极地参与网络问政，及时有效的回应公众诉求，成为政府政治智慧和行政能力的体现，政府回应，即答复意见是政府和社会、政府和公众之间互动的过程，作为政府与公众和社会的一种互动方式，是公众参与网络问政有效性的基本前提和重要保障。

政府回应机制运行是政府回应的实践过程。是政府根据其与公众的互动关系原理，通过政策回应和公众参与，实现政府和社会公众之间的各个要素之间自觉、稳定、可持续的相互关系，同时进行有意识的制度设计，实现政府组织的内在机能和流程的良好运行的过程。政府回应机制的目的就是要实现公众利益表达的常规化，保障公众的基本诉求，实现社会的有效政治参与和管理。政府回应机制的构建的目的是实现公众利益表达的行为正常化和常规化的需要，是政府提高其信任度、增强其合法性、促使其治理创新、提高公共服务水平的需要。政府回应与网络问政的利益一致性、回应流程的单一化以及回应型政府建设的要求为政府回应机制的构建和优化提供了契机。

基于网络问政的政府回应机制运行是检验回应机制构建的有效性的一个过程。通过文献资料查阅、典型案例分析和网络问卷调查，通过统计数据，研究出针对一般性

政府回应机制运行过程中存在的回应主体理念不足、与公众互动回应较少、回应效力较低等现状，深层剖析出政府回应机制构建和运行过程中政府权力结构碎片化、回应方式和流程单一、回应非制度化和非理性化等原因。结合政府回应机制构建和运行的典型案例，运用协同学、系统学等相关理论，对政府回应机制运行过程进行优化，通过改革政府机构和转变政府职能，明确政府回应主体，转变政府和公众回应理念，同时建立制度化的政府回应机制，将反馈制度、绩效考核制度、问责制度等运用于政府回应机制运行中，再通过细化和改进政府回应流程，完善政府回应机制的运行过程，构建政府与公众的协同和融合回应模式，促使政府回应机制运行更具实践性和可操作性，对于实现从个案回应到制度性回应具有理论意义和现实意义。

(3)评价方案

本方案将根据答复意见的及时性、关于热点问题的答复效率、答复意见的相关性、答复意见的完整性、答复意见的可解释性等实行百分制评价。

① 答复意见的及时性（占比 18%）

在题目所给的附件 4 中，根据表格的留言时间与答复时间，另外给出一列计算留言时间与答复时间的时间差，表格的部分内容如图 3.2 所示。

	A	C	D	G	H
1	留言编号	留言主题	留言时间	答复时间	留言时间与答复时间的时间差
2	2549	A2区景蓉华苑物业管理有问题	2019/4/25 9:32:09	2019/5/10 14:56:53	15天5小时24分
3	2554	A3区潇楚南路洋湖段怎么还没修好？	2019/4/24 16:03:40	2019/5/9 9:49:10	15天17小时45分
4	2555	请加快提高A市民营幼儿园老师的待遇	2019/4/24 15:40:04	2019/5/9 9:49:14	15天18小时9分
5	2557	在A市买公寓能享受人才新政购房补贴吗？	2019/4/24 15:07:30	2019/5/9 9:49:42	15天18小时42分
6	2574	关于A市公交站点名称变更的建议	2019/4/23 17:03:19	2019/5/9 9:51:30	16天16小时48分
7	2759	A3区含浦镇马路卫生很差	2019/4/8 8:37	2019/5/9 10:02:08	31天1小时24分
8	2849	A3区教师村小区盼望早日安装电梯	2019/3/29 11:53:23	2019/5/9 10:18:58	41天22小时25分
9	33970	B7县二中违规乱补课	2018/8/12 10:56:10	2018/8/20 9:25:34	8天22小时29分
10	33978	市601小区钻石新村20栋楼下快乐休闲网吧扰民	2018/8/8 13:15:50	2018/8/17 9:49:43	9天20小时33分
11	33984	咨询在B市用临时身份证办理出生证明的问题	2018/8/3 21:26:53	2018/8/7 9:13:42	4天11小时46分
12	34239	B市泰民米粉厂存在严重安全隐患	2018/2/3 16:27:45	2018/4/16 11:33:45	72天19小时6分
13	34249	B市港口街乱象何时了	2018/1/25 12:01:13	2018/4/12 10:10:40	77天22小时9分
14	34252	栗雨东路和王家坪路的十字路口没有任何信号	2018/1/24 11:12:25	2018/4/12 11:09:34	78天23小时57分
15	35462	请问B市西环线两侧辅道何时能全线贯通	2019/12/2 19:34:28	2019/12/10 15:16:08	8天19小时41分
16	35467	请求农行B市分行公开年收入	2019/11/28 15:41:22	2019/12/2 16:37:17	4天0小时55分
17	35479	B2区泉中路车辆乱停乱放，占用人行横道	2019/11/19 9:32:36	2019/12/2 16:43:23	13天7小时10分
18	35492	三一歌雅君路口实在太危险了，需红绿灯和摄像	2019/11/5 21:51:41	2019/11/12 14:53:48	7天17小时2分
19	35798	咨询B市公摊精装修费用问题	2018/12/5 23:49:06	2018/12/21 9:25:42	16天9小时36分
20	35801	B4区金锦社区的路面硬化高低不平易积水	2018/12/4 10:09:51	2018/12/21 9:27:31	17天23小时17分
21	35812	能否在上下班高峰期增开B市63路公交车班次	2018/11/21 14:43:59	2018/11/23 10:30:56	2天19小时46分
22	35818	B市男职工生育险其配偶能报吗？	2018/11/9 14:58:05	2018/11/15 9:42:34	6天18小时44分
23	37459	请问B9市带小孩打疫苗要带什么证件	2019/1/13 1:56:01	2019/1/14 16:06:08	1天14小时10分
24	37467	反映B9市醴官公路超载问题	2019/1/4 16:10:28	2019/1/9 16:01:29	5天23小时51分
25	37474	9市农村户口村民可以一次性补缴养老保险吗？	2018/12/25 14:41:15	2018/12/30 17:41:59	5天3小时0分
26	37482	建议B9市规划一个校车接送计划	2018/12/7 18:48:01	2018/12/13 18:53:19	6天0小时5分
27	37483	关于B9市2019年度医保问题的反映	2018/12/6 8:34:31	2018/12/12 11:15:23	6天2小时40分

图 3.2 答复意见的及时性

可以看出留言时间与答复时间的时间间隔长则可达到 78 天，时间差较短的则需 1 天，在答复意见的及时性，其质量上参差不齐，本评价方案将会以时间差的平均数为基准，对留言时间与答复时间的时间差作层次分割，如表 3.4 所示。

表 3.4 答复意见评价表

时间差	分数	备注
8 天以内	100	分数*18% （该项的最后得分）
9-18 天	90	
19-28 天	80	
29-38 天	70	
39-48 天	60	
49-58 天	50	
59-68 天	40	
69-78 天	30	
79-88 天	20	
89-98 天	10	
99 天以上	0	

②关于热点问题的答复效率（占比 17%）

根据问题二中归纳的热点问题留言明细表，再做一份包含“答复时间”的热点问题留言明细表2,重点关注热点问题中的留言时间与答复时间,利用SPSS做出折线图，其中含“答复时间”的热点问题留言明细表 2（部分内容）如下图所示：

	A	C	D	G	H	I
1	留言编号	留言主题	留言时间	答复时间	留言时间与答复时间的时间差	时间差（单位：分）
2	2549	A2区景蓉华苑物业管理有问题	2019/4/25 9:32:09	2019/5/10 14:56:53	15天5小时24分	21924
3	2554	A3区潇楚南路洋湖段怎么还没修好？	2019/4/24 16:03:40	2019/5/9 9:49:10	15天17小时45分	22665
4	2555	请加快提高A市民营幼儿园老师的待遇	2019/4/24 15:40:04	2019/5/9 9:49:14	15天18小时9分	22689
5	2557	在A市买公寓能享受人才新政购房补贴吗？	2019/4/24 15:07:30	2019/5/9 9:49:42	15天18小时42分	22722
6	2574	关于A市公交站点名称变更的建议	2019/4/23 17:03:19	2019/5/9 9:51:30	16天16小时48分	24048
7	2759	A3区含浦镇马路卫生很差	2019/4/8 8:37	2019/5/9 10:02:08	31天1小时24分	44724
8	2849	A3区教师村小区盼望早日安装电梯	2019/3/29 11:53:23	2019/5/9 10:18:58	41天22小时25分	60385
9	33970	B7县二中违规乱补课	2018/8/12 10:56:10	2018/8/20 9:25:34	8天22小时29分	12869
10	33978	市601小区钻石新村20栋楼下快乐休闲网吧扰民	2018/8/8 13:15:50	2018/8/17 9:49:43	9天20小时33分	14193
11	33984	咨询在B市用临时身份证办理出生证明的问题	2018/8/3 21:26:53	2018/8/7 9:13:42	4天11小时46分	6466
12	34239	B市泰民米粉厂存在严重安全隐患	2018/2/3 16:27:45	2018/4/16 11:33:45	72天19小时6分	104826
13	34249	B市港口街乱象何时了	2018/1/25 12:01:13	2018/4/12 10:10:40	77天22小时9分	112209
14	34252	区栗雨东路和王家坪路的十字路口没有任何信号	2018/1/24 11:12:25	2018/4/12 11:09:34	78天23小时57分	113757
15	35462	请问B市西环线两侧辅道何时能全线贯通	2019/12/2 19:34:28	2019/12/10 15:16:08	8天19小时41分	12701
16	35467	请求农行B市分行公开年收入	2019/11/28 15:41:22	2019/12/2 16:37:17	4天0小时55分	5815
17	35479	B2区泉中路车辆乱停乱放，占用人行横道	2019/11/19 9:32:36	2019/12/2 16:43:23	13天7小时10分	19150
18	35492	三一歌雅君路口实在太危险了，需红绿灯和摄像	2019/11/5 21:51:41	2019/11/12 14:53:48	7天17小时2分	11102
19	35798	咨询B市公摊精装修费用问题	2018/12/5 23:49:06	2018/12/21 9:25:42	16天9小时36分	23616
20	35801	B4区金锦社区的路面硬化高低不平易积水	2018/12/4 10:09:51	2018/12/21 9:27:31	17天23小时17分	25877
21	35812	能否在上下班高峰期增开B市63路公交车班次	2018/11/21 14:43:59	2018/11/23 10:30:56	2天19小时46分	4066
22	35818	B市男职工生育险其配偶能报吗？	2018/11/9 14:58:05	2018/11/15 9:42:34	6天18小时44分	9764
23	37459	请问B9市带小孩打疫苗要带什么证件	2019/1/13 1:56:01	2019/1/14 16:06:08	1天14小时10分	2290
24	37467	反映B9市醴官公路超载问题	2019/1/4 16:10:28	2019/1/9 16:01:29	5天23小时51分	8631
25	37474	9市农村户口村民可以一次性补缴养老保险吗？	2018/12/25 14:41:15	2018/12/30 17:41:59	5天3小时0分	7380
26	37482	建议B9市规划一个校车接送计划	2018/12/7 18:48:01	2018/12/13 18:53:19	6天0小时5分	8645
27	37483	关于B9市2019年度医保问题的反映	2018/12/6 8:34:31	2018/12/12 11:15:23	6天2小时40分	8800

图 3.3 包含“答复时间”的热点问题留言明细表 2

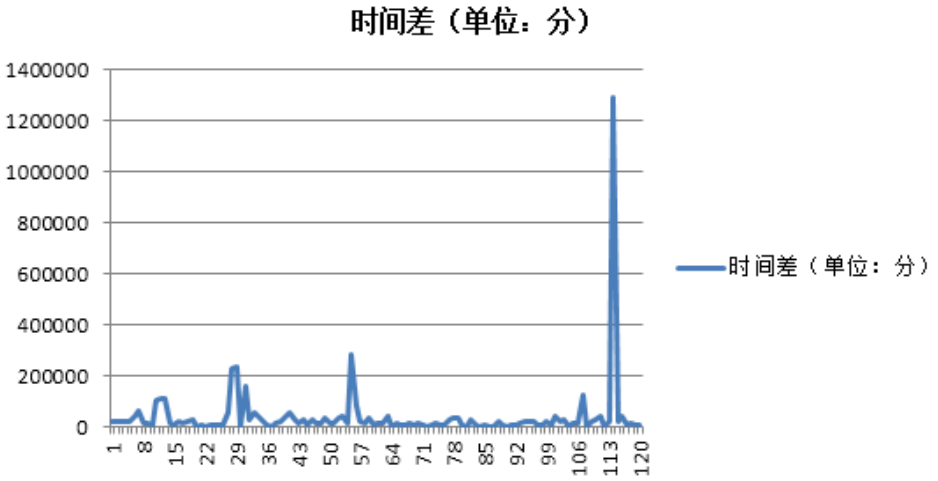

图 3.4 留言时间与答复时间的时间差折线图

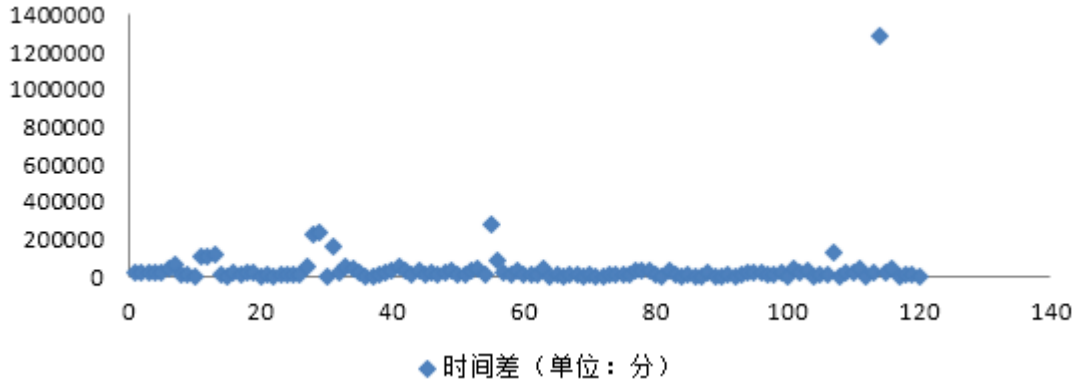

图 3.5 留言时间与答复时间的时间差散点图

相关性

		序号	时间差（单位：分）
序号	皮尔逊相关性	1	.060
	Sig.（双尾）		.512
	个案数	120	120
时间差（单位：分）	皮尔逊相关性	.060	1
	Sig.（双尾）	.512	
	个案数	120	120

图 3.6 时间差相关性

在散点图中，我们初步给出“关于热点问题的答复效率”的评价标准：

表 3.5 答复效率评价表

相关性	分数	备注
1.0	100	分数*17% （该项的最后得分）
0.8	80	
0.6	60	
0.4	40	
0.2	20	
0	0	

③答复意见的相关性（占比 20%）

首先针对附件 4 中的留言主题（留言详情）和答复意见，参考各类文献和问题一所归纳出的一级标签分类，做出如下领域本体部分对象属性表，如表 3.6 所示。

表 3.6 领域本体部分对象属性表

属性	说明	约束	
		定义域	值域
has part of	组成	政府机构	民政局， 人力资源和社会保障部
is part of	一部分	民政局，人力资源和社会保障部	政府机构
has function of	具有职能	民政局，人力资源和社会保障部	收入分配，社会保险等
is function of	执行职能	收入分配，社会保险等	民政局， 人力资源和社会保障部
has kind of	包括	社会保险	生育保险，养老保险，工伤保险， 失业保险，医疗保险
is kind of	属于	生育保险，养老保险，工伤保险， 失业保险，医疗保险	社会保险
is constrains of	受约束	生育保险管理，养老保险金管理，工伤保险管理，失业保险管理，医疗保险管理……	生育保险政策，养老保险金政策，工伤保险政策，失业保险政策，医疗保险政策……

has constrains of	指导	生育保险政策，养老保险金政 策，工伤保险政策，失业保险政策，医疗保险政策	生育保险管理，养老保险金管理，工伤保险管理，失业保险管理，医疗保险管理
is according as	依照	生育保险政策，养老保险金政 策，工伤保险政策，失业保险政策，医疗保险政策	生育保险政策，养老保险金政 策，工伤保险政策，失业保险政 策，医疗保险政策
has body of	实施主体	工伤认定、劳动能力鉴定、工伤 社保登记、待遇核准、工伤报销……	劳动局保险科、社保中心

属性可以指定定义域（Domain）和值域（Range），如对象属性 has function of 将定义域限定为机构类（民政局，人社部），值域限定为职能（收入分配，社会保险等），即某职能机构具有某固定职能。

其次对答复意见中的内容进行语义相似度检测。首先是语义相似度算法搭建。对于空间向量 P {（S1，W1），（S2，W2），„，（Sn，Wn）}，分别为关键词 S1，S2，…，Sn 在本体结构中匹配相应的概念，如直接匹配不成功，通过匹配各概念同义词组寻找，若同义词组也找不到，则使用中科院知网 Hownet 词库进行相似度匹配计算，并将相关数据返回。具体步骤如图 3.7 所示：

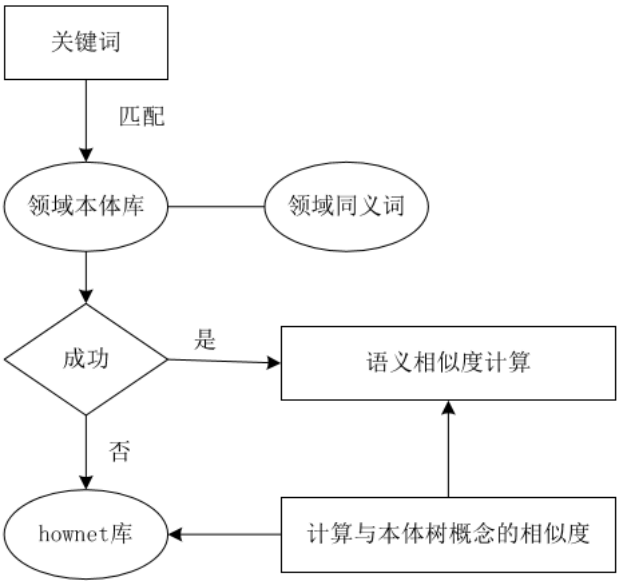

图 3.7 相似度匹配计算

对于调用hownet词库算法部分，本文采用java开源项目 chinesewordsimilarity 函数调用，此函数以 hownet 本体词库为基础，以中科院刘群等教授“基于《知网》的词汇语义相似度计算[69]”算法为依据编写，通过充分利用本体最小意义单位义原计

算两个词语的相似度，以语义距离为基础，计算两个词语义原相似度的最大值，每个义原又由第一独立义原、其他独立义原、关系义原、符号义原四部分组成，需分别计算相似度，给四部分分别定义权重，具体算法如以下公式所示，其中。本文通过使用 MyEclipse 直接调用此函数算法自动批量计算非本体概念相似度。

$$Sim(P,S)=\beta_1\times\frac{\sum_{i=1}^{n}\left(W_{Ai}\times\left(\frac{5}{Dis\ (Ai,S)+5}\right)\right)}{\sum_{i=1}^{n}W_{Ai}}+\beta_2\times\frac{\sum_{i=n+1}^{m}\left(SimHownet\ \ (Ai,B)\times W_{Ai}\times\left(\frac{5}{Dis\ (B,S)+5}\right)\right)}{\sum_{i=n+1}^{m}W_{Ai}}$$

其中β1 和β2 为权重值， β1 取 0.8 为可以直接映射到本体树上的概念相似度，β2 取 0.2 为不可直接映射到本体树的概念相似度，1 到 n 代表有 n 个可以直接映射到本体树的概念，n+1 到 m 代表有 m-n 个不可以直接映射的概念。通过计算空间向量 S 和部门概念 P 的相似度，如果相似度大于 0.5，认为此向量 S 与 P 相关。

通过反复验证计算，β1 取 0.8 为最佳值，因β1 取 0.7、0.6 准确率会明显下降，β1 取 0.9 时会过分加大某个概念的重要性，使非领域知识归入领域知识的机率加大，即查准率开始下降，表现结果如图 3.8 所示。

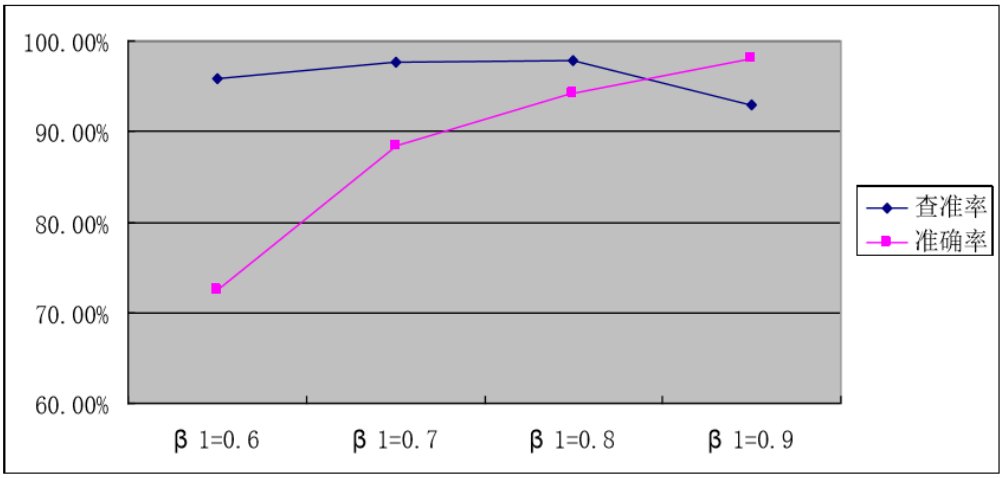

图 3.8 查准率随权重值变化的表现结果图

相关性匹配需要用到相似度算法和领域本体，以提取知识共享库中已存在的相关性知识，相关性匹配结束后，进行差异性的计算。

表 3.7 答复意见的相关性的评价表

差异性	得分	备注
0-20%	100	分数*20%

21-40%	80	(此项最后得分)
41-60%	60	
61-80%	40	
81-100%	20	

④答复意见的完整性（占比 15%）

伴随着网络信息的剧增，人们越来越依赖于信息检索技术来寻找信息，但目前基于关键字的传统搜索方法并不能很好地满足人们的信息需求。由于忽视了资源本身所含的语义信息，传统的基于关键字的信息检索，只能获得较低的查全率和查准率。本体理论和技术源于知识工程和人工智能领域，能够很好的处理基于语义的推理机制和自然语言理解问题，因而成为改进传统信息检索方式的良好途径。相比于传统的基于关键字检索方法，基于本体技术的语义信息检索能减少不相关的返回结果，提高检索结果的查全率和查准率，更加符合用户的需求。

对概念语义相似度计算方法进行分析和研究的基础上，提出了一种综合的语义相似度计算方法。在相似度计算时充分考虑数据层(Data Layer)、本体层(Ontology Layer)和上下文层(Context Layer)，并对本体层的语义相似度计算进行了细化，重点对基于语义距离的相似度计算方法进行了改进。实验验证上述方法的有效性。

提出了一种文本信息检索方法，把本体技术结合到传统的全文信息检索中。对初始查询进行查询扩展，利用概念语义相似度，对扩展词的规模和查询权重进行有效的控制，并结合向量空间模型(Vector Space Model，VSM)和本体技术对检索结果文档的分值进行计算,过滤与原始查询语义相关度较小的文档,最后实验验证该方法的效率。

对此，得益于覆盖率、新鲜度、准确性的提升，在内部评测中，腾讯 AI Lab 提供的中文词向量数据相比于现有的公开数据，在相似度和相关度指标上均达到了更高的分值。在腾讯公司内部的对话回复质量预测和医疗实体识别等业务场景中,腾讯 AI Lab 提供的中文词向量数据都带来了显著的性能提升。

由于采用了更大规模的训练数据和更好的训练算法，所生成的词向量能够更好地表达词之间的语义关系，如下列相似词检索结果所示：

输入	刘德华	兴高采烈	狂奔	自然语言处理
相似词	刘天王	兴高彩烈	飞奔	自然语言理解
	周润发	兴冲冲	一路狂奔	计算机视觉
	华仔	欢天喜地	奔跑	自然语言处理技术
	梁朝伟	兴致勃勃	狂跑	深度学习
	张学友	眉飞色舞	疾驰	机器学习
	古天乐	得意洋洋	飞驰	图像识别
	张家辉	喜笑颜开	疾奔	语义理解
	张国荣	欢呼雀跃	奔去	语音识别

图 3. 相似词检索结果图

因此，只需要在程序中输入“您好”、“留言”、“已收悉”、“针对此问题”、“经调查”、“规定”、“整治”、“感谢网友”等字词，就可以初步判断答复意见的完整性，并且如下：

表 3.8 完整性的评分表

完整度	分数	备注
100%	100	分数*15% (此项最后得分)
80%-99%	85	
60%-79%	65	
40%-59%	45	
20%-39%	25	
20%以下	5	

另外，将给出答复意见的基本格式：

格式 1：

尊敬的网友/市民：

您好！感谢您关注 XXX 问政，您于 XXXX 年 XX 月 XX 日给 XXX/关于 XXX 的留言已收悉，对您的留言我们高度重视，XXXX 主动认领查处。现将具体情况反馈给您：

经查：XXXX（说明调查情况）

下步，XXXX（做出解决方案/对目前无法解决的问题做出解释）

在此，感谢您对我们工作的关心和支持！

格式 2：

> 尊敬的网友/市民：
>
> 您好！感谢您关注 XXX 问政，您于 XXXX 年 XX 月 XX 日给 XX 的留言已收悉，对您的留言我们高度重视，XXXX 主动认领查处。现将具体情况反馈给您：
>
> 您反映的问题建议您拨打 XXXX 热线，可以就有关问题进咨询，谢谢！

⑤答复意见的可解释性（占比 30%）

许多政府网站的“网络问政”板块，互动及时、办事效果好，赢得了群众点赞。可是，有少数地方政府网站，面对群众疑虑或问题，回复互动倒是及时快捷、热情有加，但就是问题依然迟迟得不到解决。

网络问政如今已成民众与政府沟通的重要方式。老百姓有需要咨询或投诉的，敲敲键盘发个帖子，提交上去，得到及时回复，既公开透明，又快捷便民。但最近“僵尸网站”少了，线上政府热情回复，可是却难见下文，没了“回音”，千篇一律的回复被网友戏称“万能自动回复”，不仅浪费了公共资源，还有损政府形象。

针对“答复意见的可解释性”，首先对答复内容进行 Web 文本主题抽取。

Web 文本主题抽取是文本分类与知识发现的研究热点，既有的抽取方法一般存在主题粒度确定、主题语义解释、新网络词汇识别等难题，限制了其在开放应用领域的使用效果。论文借助百度百科词条背景，基于关系概念的概念分层以及主题连通的思想，面向中文文本构建了关系概念主题抽取模型(relational concept topic model，RCTM)，RCTM 模拟人的概念局部识别，上下文语境理解的并行阅读方式，由此实现中文文本的主题抽取。RCTM 中主题的表达相对独立、语义连通灵活，主题的描述具有更好的通用性与可解释性，为 Web 文本主题抽取提供了新的研究思路.实验表明，RCTM 具有良好的主题抽取准确率，文本抽取出的主题词，简洁直观、可解释性好.针对开放的 WEB 文本，具有更好的通用性、稳定性。

Web 文本提取过程如下：1）去除 HTML 标记及相关冗余，2）去除文档中残余的冗余信息，3）计算句子权重，提出主题文本。我们只是简单地利用文本长度和标点符号序列就很准确地将主题文本从文档中提取出来。

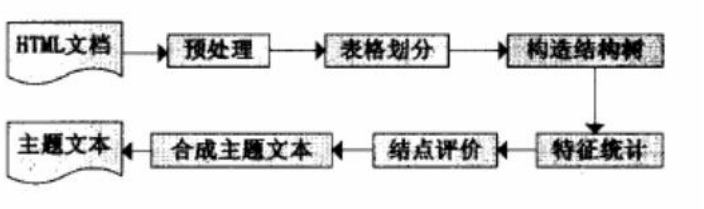

图 3.10 Web 文本提取过程

其次提取的主题文本与留言内容进行相似性匹配，其匹配方法类似于“答复意见的完整性”中的文本检索等方法，故不一一阐述。

其中，答复意见的可解释性将用答复与留言内容匹配度表示：

表 3.9 答复意见的可解释性程度表

答复与留言内容匹配度	分数	备注
100%	100	分数*30% (此项最后得分)
80%-99%	85	
60%-79%	65	
40%-59%	45	
20%-39%	25	
20%以下	5	

我们根据答复意见的及时性、关于热点问题的答复效率、答复意见的相关性、答复意见的完整性、答复意见的可解释性这五项的得分，最后做出一个简单的评价标准：

表 3.10 答复意见评价标准表

分数	评价	备注
85-100	优秀	暂无
70-84	良好	
55-69	合格	
54 及以下	差	

3.3.2　评价方案的测试及结果

以附件中的任意一例留言进行测试：

留言用户	留言主题	留言时间	留言详情	答复意见	答复时间
A00045581	A2区景蓉华苑物业管理有问题	2019/4/25 9:32:09	2019年4月以来，位于A市A2区桂花坪街道的A2区公安分局宿舍区（景蓉华苑）出现了一番乱象，该小区的物业公司美顺物业扬言要退出小区，因为小区水电改造造成物业公司的高昂水电费收取不了(原水电在小区买，水4.23一吨，电0.64一度）所以要通过征收小区停车费增加收入，小区业委会不知处于何种理由对该物业公司一再挽留，而对业主提出的新应聘的物业公司却以交20万保证金，不能提高收费的苛刻条件拒之门外，业委会在未召开全体业主大会的情况下，制定了一高昂收费方案要各业主投票，而投票不采用投票箱只制定表格要物业公司人员这一利害关系机构负责组织，对投票业主隐私权没有任何保护，还对投反对票的业主以领导做工作等方式要求改变为同意票，这种投票何来公平公正公开，面对公安干警采用这种方式投票合法性在哪？	现将网友在平台《问政西地省》栏目向胡华衡书记留言反映“A2区景蓉花苑物业管理有问题”的调查核实情况向该网友答复如下：您好，首先感谢您对我们工作的信任和支持，关于您在平台栏目给胡华衡书记留言，反映“A2区景蓉花苑物业管理有问题”的情况已收悉。现将我们调查处理情况答复如下：经调查了解，针对来信所反映的“小区停车收费问题”，景蓉华苑业委会于2019年4月10日至4月27日以“意见收集方式”召开了业主大会，经业委会统计，超过三分之二的业主同意收取停车管理费，在业主大会结束后业委会也对业主提出的意见和建议进行了认真疏理归纳并进行了反馈，业委会制定的停车收费标准不高于周边小区价格。针对来信所反映的“物业公司去留问题”，5月5日下午,辖区桂花坪街道牵头组织社区、物业公司、业主委员会、业主代表的会议，区住房和城乡建设局也参加了会议。在综合各方面的意见后，辖区桂花坪街道、区住房和城乡建设局已要求业委会依法依规召开业主大会，根据业主大会的表决结果再执行相应的程序。再次感谢您对我区工作的理解和关心。2019年5月9日	2019/5/10 14:56:53

其中：数据情况如下：

留言用户	留言主题	留言时间	答复时间	时间差
A00045581	A2区景蓉华苑物业管理有问题	2019/4/25 9:32:09	2019/5/10 14:56:53	15天5小时24分

留言与答复明细如下：

留言详情
2019年4月以来，位于A市A2区桂花坪街道的A2区公安分局宿舍区（景蓉华苑)出现了一番乱象，该小区的物业公司美顺物业扬言要退出小区，因为小区水电改造造成物业公司的高昂水电费收取不了(原水电在小区买，水4.23一吨，电0.64一度）所以要通过征收小区停车费增加收入，小区业委会不知处于何种理由对该物业公司一再挽留，而对业主提出的新应聘的物业公司却以交20万保证金，不能提高收费的苛刻条件拒之门外，业委会在未召开全体业主大会的情况下，制定了一高昂收费方案要各业主投票，而投票不采用投票箱只制定表格要物业公司人员这一利害关系机构负责组织，对投票业主隐私权没有任何保护，还对投反对票的业主以领导做工作等方式要求改变为同意票，这种投票何来公平公正公开，面对公安干警采用这种方式投票合法性在哪？

答复意见
现将网友在平台《问政西地省》栏目向胡华衡书记留言反映“A2区景蓉花苑物业管理有问题”的调查核实情况向该网友答复如下：您好，首先感谢您对我们工作的信任和支持，关于您在平台栏目给胡华衡书记留言，反映“A2区景蓉花苑物业管理有问题”的情况已收悉。现将我们调查处理情况答复如下：经调查了解，针对来信所反映的“小区停车收费问题”，景蓉华苑业委会于2019年4月10日至4月27日以“意见收集方式”召开了业主大会，经业委会统计，超过三分之二的业主同意收取停车管理费，在业主大会结束后业委会也对业主提出的意见和建议进行了认真疏理归纳并进行了反馈，业委会制定的停车收费标准不高于周边小区价格。针对来信所反映的“物业公司去留问题”，5月5日下午,辖区桂花坪街道牵头组织社区、物业公司、业主委员会、业主代表的会议，区住房和城乡建设局也参加了会议。在综合各方面的意见后，辖区桂花坪街道、区住房和城乡建设局已要求业委会依法依规召开业主大会，根据业主大会的表决结果再执行相应的程序。再次感谢您对我区工作的理解和关心。2019年5月9日

首先对答复时间进行测试，将其放入表格中计算得分为 90*18%=16.2 分；第二，由于该问题不属于热点问题，故取热点问题答复效率得分的合格值：60*17%=10.2 分；第三，关于答复意见相关性的判断，我们采取差异度的方法，得分为：87*20%=17.4

分；第四，在答复意见的完整性中，我们可以看到含有“您好”、“已收悉”、“经调查”等规定字词和相近字词，再放入程序中测试得分数为：98*15%=14.7 分；第五，根据留言主题与答复意见的匹配程度，得出分数：78*30%=23.4 分。

最后得出总分为：16.2+10.2+17.4+14.7+23.4=81.9，故被评为“良好”。

4. 总结

在电子政务基础资源的大规模建设的背景下，基于文本识别、文本检索、文本分类的现有技术，利用 Excel、SPSS 等程序软件，我们逐一解决了留言一级标签的重新分类、热点问题的挖掘、答复意见评价方案的建立等问题。

根据研究思路，从留言分类到挖掘热点问题，再到评价答复意见，建立了一套基于数据挖掘的智慧问政留言及答复管理模型，这套模型可以将大量的留言内容进行较为准确的一级标签分类，大大减少了只能依靠人工来进行留言划分的工作量；其次可以更为高效地挖掘和整理热点问题，对急需要解决的问题的答复提供了很好的帮助；再者，对政府及各有关部门关于市民留言的答复提出了一套评价方案，这套评价方案能够在一定程度上减缓政府及各有关部门敷衍大幅的情况；最后，我们分别对模型中的各个部分进行简单地测试，测试结果中出现的错误较少，以此说明该模型的建立方式基本正确。

5. 参考文献

[1] 吴旭刚. 基于网络问政的政府回应机制优化研究[D].湘潭大学硕士学位论文，2017

[2] 黄小慧. 基于本体的网络问政知识管理机制研究[D]. 华南理工大学硕士学位论文，2011

[3] 张成伟. 基于概念语义相似度的文本信息检索研究[D]. 安徽:安徽大学，2009. DOI:10.7666/d.d203878.

[4] 程春雷, 夏家莉, 曹重华,等. 关系概念的 Web 文本主题抽取模型研究%Research on Web Text Topic Extraction Model with Relational Concept[J]. 小型微型计算机系统, 2016, 37(5):972-977.

[5] 杨晖. 基于标签分类内容共享平台的网页自动文摘模型[D]. 庆大学计算机系统结构