

# 第八届泰迪杯数据挖掘挑战赛

## ——C 题 “智慧政务” 中的文本挖掘应用

参赛队编号：2020107417

# 目 录

1 问题背景.....	2
2 群众留言分类.....	1
2.1 留言一级标签多分类模型.....	1
2.1.1 数据集划分-10 折交叉验证.....	1
2.1.2 文本预处理.....	2
2.1.3 文本表示模型.....	3
2.1.4 文本特征提取.....	4
2.1.5 KNN 文本多分类模型.....	4
2.2 分类模型评价.....	5
3 热点问题挖掘.....	6
3.1 热点问题分类模型.....	6
3.1.1 数据预处理.....	6
3.1.2 基于 word2vec 的词向量训练.....	6
3.2 热点问题评价指标体系.....	8
4 答复意见的评价.....	9
5 智慧政府的建议与展望.....	10
6 参考文献.....	11

# 1 问题背景

近年来，随着微信、留言、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。

根据来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。本文利用自然语言处理和文本挖掘的方法解决了群众留言分类问题、热点问题挖掘、答复意见评价等问题，为提高政务智能化信息化效率贡献了自己的一份力量。

## 2 群众留言分类

### 2.1 留言一级标签多分类模型

#### 2.1.1 数据集划分-10 折交叉验证

文本集合一般分为两个部分：训练样本集合和测试样本集合，训练样本集合用于训练分类模型，测试样本集合用于测试分类模型的性能。

10 折交叉验证先将附件 2 中的样本数据随机的分成 10 份，每次随机的选择 9 份作为训练集，剩下的 1 份做测试集。当这一轮完成后，重新随机选择 9 份来训练数据。若干轮（小于 10）之后，选择损失函数评估最优的模型和参数。通过反复的交叉验证，用损失函数来度量得到的模型的好坏，最终我们可以得到一个较好的模型。

其中利用重复分层交叉验证，保证各个一级标签类别中的数据均可被均匀的划分到训练集和测试集中。共计 9210 条数据，其中 6447 条为训练集数据，2763 条为测试集数据，具体划分结果如下图所示。

A	B	C	D	E	F
留言编号	留言用户	留言主题	留言时间	留言详情	一级标签
24	A00074011	建筑集团占道施工有违	20/1/6 12:09:	围墙内。每天尤其	城乡建设
37	U0008473	大厦人为烂尾多年，	20/1/4 11:17:	着，不但占用人行道	城乡建设
83	A00063999	市A1区苑物业违规收	9/12/30 17:06:	决。不知程明物业如	城乡建设
303	U0007137	南路A2区华庭楼顶水箱	19/12/6 14:40:	，霉是一种强致癌	城乡建设
319	U0007137	2区华庭自来水好大一	19/12/5 11:17:	，霉是一种强致癌	城乡建设
379	A00016773	市盛世耀凯小区物业无	19/11/28 9:08:	业不是为业主服务的	城乡建设
382	U0005806	询A市楼盘集中供暖一	9/11/27 17:14:	月亮岛片区近年规划	城乡建设
445	A00019209	西路可可小城长期停水	9/11/19 22:39:	帮助至今没有找到	城乡建设
476	U0003167	收取城市垃圾处理费不	9/11/15 11:44:	在的物业公司也未给	城乡建设
530	U0008488	A3区魏家坡小区脏乱差	9/11/10 18:59:	让人好好休息一下	城乡建设
532	U0008488	A市魏家坡小区脏乱差	9/11/10 12:30:	让人好好休息一下	城乡建设
673	A00080647	四届非法业委会涉嫌侵	9/10/24 11:29:	法业委会主任不敢	城乡建设
994	U0005196	梅溪湖壹号御湾业主用	19/9/18 22:43:	别的城市都已经一	城乡建设
1005	U0006509	翡翠湾强行对入住的业	19/9/18 13:36:	地产公司和金晖物	城乡建设
1110	A00099772	市锦楚国际星城小区三	19/9/9 11:07:	是无通知，突然断	城乡建设
1309	U0005083	和紫郡用电的问题能不	19/8/21 15:12:	之后，我们的用电	城乡建设
1440	A0003288	际新城从6月份开始停	19/8/6 10:28:	的生活，而且我们	城乡建设
1775	U0002150	成区南西片区城铁站设	19/7/4 18:52:	A市，并且规划有	城乡建设
1783	U0004763	政府加大对滨水新城的	19/7/4 14:25:	的或者几个半大小	城乡建设
1827	U000613	楚府线几个小区经常停	19/7/1 20:14:	已停电三次。说是	城乡建设
2603	A00099650	日及西地省辉东安建工	19/4/20 16:50:	(2015年~2016年)	城乡建设
3607	A00046529	水嘉园1栋三单元群租	19/1/8 10:08:	患，投诉给物业公	城乡建设
<div> <div>总数据集</div> <div>训练集</div> <div>测试集</div> <div>十</div> </div>					

## 2.1.2 文本预处理

文本预处理的主要作用是对输入的文本进行分词处理和结构化处理，通过预处理，原来完整的文本文档就转换成了词语序列。

### 分词处理

预处理的第一步是要进行分词，标记，如停用词，同义词等等。分词处理对于中文文本分类而言非常重要，因为汉字的书写习惯，在中文文档中没有一个很明显的词语分隔符，词语和词语之间的切分标志不明显，因此需要中文分词技术对中文文本进行分词，自上世纪年代文本分类在中国发展以来，针对政务文本分类相关技术中文语言的特性，提出了各种各样的分词算法，但是其算法主要可以归纳为三大类：基于词典的分词算法，基于语料统计的分词算法和基于语义分词算法。不同类型的分词算法所侧重的方面也不一样，我们可以按照文本集合的特点来选择适合于分词算法，以取得更大切词准确率和更高的切词速度。本文所使用的中科院汉语词法分析系统正是基于多层的隐马尔可夫模型隐马尔可夫模型正是基于语料统计的分词算法的两种主要模型的其中之一。

本文所使用的停用词词集来源于 GitHub，详见参考文献中的【1】

### 结构化处理

其次我们要对文本进行结构化处理，结构化处理主要是去掉文本中一些无用

的标记，如停用词，同义词等等。中文文本的结构化处理是要去掉在文本中频繁出现，但是本身对于文本而言没有多大作用的词语<sup>1</sup>。这一类主要包括了语气助词、副词、介词、连词等，通常自身并无明确意义，只有将其放入一个完整的句子中才有一定作用的词语，这类词语如果不过滤掉，因为其频率较高，

很难被特征选择算法去掉，将会大大增加文本空间的维数，增加计算量，从而降低文本分类的效率和速度，另外一方面也影响文本表示的准确度，分词以后的词语集合我们可以按照停用词表来去掉停用词，因为停用词的选取会影响分类的词集和性能，在制订停用词表的时候应该是基于不降低文本分类性能和减少文本分类特征项这两个原则。

### 2.1.3 文本表示模型

就中文文本而言，文本主要是由众多的中文文字组成，很难直接用于计算机中来进行训练和分类，因此我们需要将文字抽象成机器能够识别的形式，用计算机能够处理的方式将文本表示出来，在文本的预处理中，我们使用字或者词语切割了文本，将文本形成了词语的集合，但是字和词语仍然不能直接让计算机处理，必须将文本抽象成某种表示模型。本文主要应用的是世纪年代等人提出了向量空间模型该模型最早被应用到信息检索领域中。著名的系统就是该模型的成功应用。

模型的基本思路是：将文本以特征词的权重向量的方式表示为： $w_i$  为第个特征词的权值，可见基于的文本分类最重要的步骤就是从文本分词后的特征词集合中提出关键特征词。

在空间向量模型中，文本集合是由文本矩阵来表示的，矩阵表示如下

$$T = (W_{ij})_{M \times N}$$

公式中，表示词在第个文本中的特征权值，表示文本集合中词语的总数，表示文本集合中文本的总数。之所以得到广泛的认可，是因为它有以下两个优点第一，该模型仅仅提供了一个理论框架，其中的特征项权重计算方法没有固定，可以通过比较选出一种较好的算法，从而提高了文本分类的性能；第二，一部分匹配策略可以使过滤得到的结果文本与用户的需求更加接近。

### 2.1.4 文本特征提取

对于文本分类技术而言，首先经过文本预处理步骤中的中文分词处理后，原始文本转换为由众多词汇组成的集合，无论采取何种文本表示模型，由这些特征词组成的集合仍然是一个维度很高的空间，我们就需要对特征空间降维处理，降维的原因有两个：第一，维度越高，计算量越大，降维能够提高实际执行的效率，节省运行时间。第二，并不是所有的特征词都是关键特征词，有些特词对分类的贡献率比较低，应该去掉，依此来降低这些噪音对分类的影响。“特征提取就是从预处理完成后的原始文本特征集合里面找出一个子集，挑选的子集的标准就是不能降低分类的准确度，可见降维并不是改变原始文本集合的特性，而是选择有代表性的关键词，组成了一个新的文本集合。

对于 8 类一级标签，我们共提取了如下表所示的关键词词袋，以下为以环境保护和 为例的词袋表格（完整的一级分类特征词详见附件 1[一级分类特征词]）

一级分类	特征词			
环境保护	生态示范和模范城区创建	专项资金使用	大气污染	环评报告不实
	环境监测	生态破坏	其他污染	未批先建
	环境资质管理	危险化学品污染	电磁污染	污染损害纠纷调解
	环境执法	噪音污染	水污染	固体污染
	淘汰落后产能补偿	放射性污染	土壤污染	未验先用

一级分类	特征词			
党务政务	制度建设	台湾事务	工会工作	出版管理
	组织建设	外交外事	共青团工作	新闻管理
	作风建设	军事国防	妇联工作	统战工作
	廉政建设	宗教事务	互联网信息监管	政治体制改革
	侨务	民族政策	精神文明建设	政协工作
	港澳事务	民族事务	广播影视管理	人大工作

### 2.1.5 KNN 文本多分类模型

最近邻简称文本分类算法文本分类算法，是一种常用的文本分类算法，它是基于类比学习的分类方法，利用算法进行文本分类的基本思路是：、特征选择完

成后，用特征子集来表示训练样本，形成训练样本特征词空间。、用特征子集来表示待分类样本，计算训练样本特征词空间里面的每一个文本与待分类样本之间的距离。、从中选出个与待分类样本间距离最近的文本，并按降序排列。在这个文本中，如果某一类别的文本数最多，那这个类别就是待分类文本的类别。计算文本相似度最常用的方法是余弦距离，公式如下：

$$\text{Sim}(d_1, d_2) = \frac{\sum_{i=1}^n w_{1i} w_{2i}}{\sqrt{\sum_{i=1}^n w_{1i}^2 \sum_{i=1}^n w_{2i}^2}}$$

其中， $w_{1i}$ 和 $w_{2i}$ 分别表示文本 $d_1$ 和 $d_2$ 的文本向量中第 $i$ 个特征项的权重，公式中的值越大则说明文本 $d_1$ 和 $d_2$ 之间的差别越小， $d_1$ 和 $d_2$ 更有可能归属于同一个类别，反之，若 $w_{1i}$ 和 $w_{2i}$ 的乘积越小，说明两个向量属于某一个类别的可能性很小。文本分类算法是在训练样本中抽取出 $n$ 个作为分类依据，缺陷是很明显的，当样本空间分布不均匀，某一类的样本数目特别大，某一类很小，这可能会造成在分类的时候，样本数目大的训练样本类别在中占有很大的一部分，

肯定会造成分类的偏差。它的另一个不足之处都是算法对于权值的计算量比较大，因为对于待分类文本，需要与已知类别的样本空间里面的所有样本进行距离计算排序，才能得到前面 $n$ 个距离最近的文本。就目前而言，经常用到的做法是事先对已知类别的样本进行样本的分类重要性分析，去掉一些对分类而言用处不大的样本，文本分类算法比较适用于对样本容量比较大的文本空间进行分类，对那些样本容量比较小的文本空间，如果采取文本分类算法，非常容易产生误分，影响分类的效果。

## 2.2 分类模型评价

基于题目要求，本文按如下 F1 分类模型评价指标对分类模型与超参数不断优化调整

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

最终得到的分类模型训练集数据为 0.766，测试集数据测试为 0.677，具体分类结果详见附件 2[群众留言分类结果表]

## 3 热点问题挖掘

### 3.1 热点问题分类模型

#### 3.1.1 数据预处理

对留言文本进行分词和去停用词处理

在文本中使用频率很高，但不具有检索意义、不能表达文本中心思想的词语即为停用词。本文利用 GitHub 停用词表（详见附件 1），在利用结巴分词处理时，将得到的结果与停用词匹配，若匹配成功，就将这个词去掉，否则就保留下来。经去停用词后的部分评论文本分词后的结果如表所示。

表 3 经去停用词和分词后的部分评论文本

评论	分词结果
我在东北,过来感谢有绵阳警方这样的好警察,人民放心!	东北 过来 感谢 绵阳 警方 好 警察 人民 放心
绵阳警方的工作态度还是值得点赞的	绵阳 警方 工作 态度 值得 点赞
作为大连人看到这些微博和大连公安的微博对比 真心让大连公安弄寒了心	大连人 看到 微博 大连 公安 微博 对比 真心 大连 公安 弄 寒 心
中国警方在线官方微博,心系群众,时刻关注人民群众的重大事件,辛苦了!	中国 警方 在线 官方 微博 群众 时刻 关注 人民 群众 重大事件 辛苦
头部那几下重击好可怕	头部 几下 重击 好 可怕
不相信当事人仅仅情伤 感觉肯定很严重 不要放过一个坏人	不 相信 当事人 仅仅 情伤 感觉 肯定 很 严重 不要 放过 坏人

#### 3.1.2 基于 word2vec 的词向量训练

利用 word2vec 分布式的词向量训练方法，将词汇以向量形式展现出来，



从而达到将目标文本用于向量 运算的目的，然后更好地完成情感分析的任务。在进行 word2vec 的评论文本向量表达时，word2vec 模型的性能在判定词语间的语义关系时的表现，比 word2vec 模型更好，因此采用 word2vec 型对词库进行向量表达，并转化成特征向量的格式。步骤如下：

（1）对于上下文窗口的取值，实验中将窗口值设定为 5。

（2）基于精确度的考虑，词向量维度选择 300。例如：“可怕”的词向量如图 1 所示。

（3）在大规模语料库中训练出合适的词向量的基础上，评论文本本身的特征向量由对每条留言文本中所有词语的词向量累加后求平均得到。

```
In [21]: word_vec["可怕"]
Out[21]: array([-0.01923663,  0.04251761,  0.0007377,   -0.00613203,  0.00434406,
  0.09681892,  0.05613354,  0.02423764,   -0.02447239,  0.04273675,
  0.02985588,  0.06017527,   -0.03006638,   -0.06660105,  0.08285648,
  0.02231164,  0.02048188,  0.04652022,  0.06148026,  0.02726142,
  0.05668839,   -0.05619558,  0.02858041,   -0.05009842,   -0.02665614,
  0.03132824,   -0.03857632,  0.03423441,   -0.04878307,  0.01440702,
  -0.00704557,   -0.04852435,   -0.05323927,   -0.01095064,  0.00191462,
  -0.07492163,  0.02641462,  0.00255926,  0.07423229,   -0.02796243,
  0.00609034,  0.05468189,  0.14315921,  0.0022704,   -0.10236841,
  0.01395187,   -0.03428966,  0.05282713,   -0.03449227,   -0.05431299,
  0.0364626,   0.03222817,   -0.06181728,  0.01261381,  0.02497935,
  0.07051624,   -0.08532371,   -0.01995298,   -0.14631232,   -0.06203252,
  0.0123847,   0.02000811,   -0.01068336,  0.04804851,   -0.01893,
  -0.00811378,   -0.01815335,   -0.05918461,   -0.07282688,  0.06798108,
  0.02249823,   -0.03904066,   -0.00541221,  0.06283333,  0.04481772,
  -0.07386754,   -0.0140551,   -0.08728141,   -0.01943137,   -0.04115195,
  -0.02060916,  0.0484324,   -0.14185615,  0.05695656,  0.00175413,
  0.01837851,   -0.04511354,  0.08228002,  0.01037211,  0.0431748,
  0.0408273,   -0.02562444,   -0.14277475,   -0.02687152,  0.07314779,
  -0.06285559,   -0.10058171,   -0.17809546,   -0.03002307,  0.05693362,
  0.04480229,   -0.03831559,  0.07456064,   -0.11251874,  0.00447048,
  0.1239674,  0.08936244,   -0.14481111,  0.03233752,   -0.06975207,
  -0.02371453,  0.04546782,   -0.01798619,  0.00192425,  0.06846134,
  -0.07737852,   -0.00562542,   -0.04199971,  0.11728409,  0.07683342,
  -0.03293564,  0.02505319,  0.06991251,   -0.05145117,  0.06333594,
  -0.05134864,  0.00223926,  0.04939523,   -0.00574974,   -0.00317287,
  -0.00642216,   -0.05348864,  0.03624403,   -0.03243352,  0.06679387,
  0.0050741,   0.04882168,   -0.04612289,  0.07556996,   -0.01376744,
  -0.01785317,  0.07838932,   -0.12242651,  0.02944395,   -0.01142338,
  0.00159366,  0.04934741,  0.0347144,   -0.06260303,   -0.0432827,
  0.01438515,   -0.04012575,   -0.02622108,   -0.00545251,   -0.04754533,
  0.04841965,  0.01212817,  0.03981027,  0.01424227,   -0.04242923,
  -0.01863278,   -0.02859562,  0.00293603,   -0.01561744,   -0.0139687,
  -0.02352707,   -0.04014032,  0.01002742,   -0.06889735,  0.07192254,
  0.11831215,   -0.04403852,   -0.05625745,   -0.05614448,   -0.04920628,
  0.06810467,  0.00928134,  0.01221126,  0.10166306,  0.13306671,
  0.06632066,  0.02004768,   -0.07979675,   -0.02162714,   -0.04413408,
  -0.02499398,  0.01692105,   -0.06462193,   -0.13418321,   -0.00945978,
  -0.00216132,   -0.00253022,  0.00123339,   -0.00889911,  0.01881627,
  0.02729082,  0.02949569,  0.07127014,   -0.07488373,   -0.01779155,
  0.12050654,  0.01644542,   -0.01408007,   -0.10752223,  0.05796301,
  0.00155152,  0.00141139,  0.0015248,   -0.04524712,   -0.04956149,
  -0.00512413,  0.07287639,   -0.06399087,  0.08711482,   -0.00464064,
  -0.02900962,  0.02481902,  0.01942531,   -0.06211924,   -0.00774317,
  -0.01489269,  0.04414375,  0.08080114,  0.05634598,   -0.01022195,
  0.00724274,  0.05179468,  0.02839515,  0.02920082,  0.0362813,
  0.04199455,  0.00398712,   -0.05327526,  0.04091891,   -0.05304243,
  -0.05857945,   -0.04327781,   -0.13050438,   -0.13933928,   -0.04697328,
  -0.05013372,  0.00493107,   -0.01845442,   -0.0122899,  0.01883358,
  0.01256311,  0.02409304,  0.00696331,   -0.01418954,   -0.13294268,
  -0.10510134,  0.05316874,  0.03124507,   -0.08945451,  0.04596215,
  -0.00762733,  0.02376078,  0.0353476,   -0.08200548,  0.11311081,
  0.02568805,   -0.04061917,  0.00606511,   -0.17188834,   -0.05023827,
  0.05662892,   -0.04279474,   -0.04047481,  0.09976099,   -0.05851443,
  -0.00320903,  0.00223372,   -0.0329018,   -0.02012902,  0.02862214,
  -0.08242793,  0.01019183,   -0.05784925,  0.04411228,  0.02837087,
  -0.07306232,   -0.05548963,  0.02793604,  0.08898094,   -0.06748359,
  0.03824453,   -0.01708548,  0.00413033,  0.05148997,   -0.08207003,
  0.05803924,   -0.05692313,   -0.02097125,  0.01392251,  0.01084378,
  -0.00096467,  0.12782772,   -0.02597438,  0.0361142,   -0.05473543],
```

## 3.2 热点问题评价指标体系

为了对两种不同的分类方法进行比较,首先确定模型的评价指标。在本文中,微博评论情感极性分析为二分类问题,模型评价指标需要考虑正向和负向两个方面的模型效果,考虑训练模型不同预测结果的数量如表所示。

查准率、召回率、F1值和准确率四种评价指标是数据挖掘任务中常用的模型评价标准。下面对这四种评价标准进行详细介绍:准确率(Accuracy)表示训练模型对样本的分类性能,公式如

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

查准率(Precision)表示某一类别中训练模型预测数量占实际数量的比例,公式如

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

召回率(Recall)表示某一类别中训练模型预测数量占该类别实际记录数量的比例

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

## 4 答复意见的评价

表 4.9 各指标感知与期望差异

维度	指标	感知	期望	差距
有形性	政务微信页面布局简洁易懂	5.92	7.29	-1.37
	政务微信发布的信息形式多样化	6.09	7.41	-1.32
	政务微信的导航功能清楚直观	5.95	7.24	-1.29
	政务微信发布的信息很规范	5.30	6.97	-1.67
可靠性	政务微信提供的信息权威准确	6.54	7.72	-1.18
	政务微信提供的信息完整全面	6.15	7.40	-1.25
	政务微信提供的信息具有时效性	5.98	7.61	-1.63
	政务微信能为用户解决日常政务问题	5.60	7.24	-1.64
易用性	公众能快速查找并关注政务微信公众号	6.07	7.32	-1.25
	政务微信的基本功能使用方便、快捷	6.58	7.46	-0.88
	利用政务微信获取政务信息更便捷	6.30	7.51	-1.21
移情性	政务微信能随时为用户提供服务	6.01	7.54	-1.53
	政务微信服务态度友好	6.18	7.63	-1.45
	政务微信提供的服务符合用户需求	4.99	7.08	-2.09
	政务微信针对不同用户提供个性化服务	4.60	7.22	-2.62
响应性	政务微信对用户各项操作指令反应速度快	6.15	7.41	-1.26
	政务微信对用户的需求及意见及时予以答复	5.87	7.38	-1.51
	政务微信告知用户解决问题需要的时间	5.74	7.54	-1.80
	政务微信的回复与客户需求一致	5.82	7.46	-1.64
安全性	政务微信平台是安全可信赖的	6.02	7.73	-1.71
	政务微信能够保护用户个人账户等隐私信息	5.86	7.74	-1.88
	政务微信提供的信息不会有其他安全风险	6.18	7.67	-1.49

本文通过信度和效度的分析，证明该评价模型具有良好的信度和效度；相关性分析结果也表明各维度与政务微信公众服务质量总体评价均呈现不同程度的正相关关系，说明本文构建的政务微信公众服务质量评价模型具有良好的科学性和适用性。在服务质量分析结果上，通过对模型的应用，得知我国政务微信服务质量得分为在整体上还没有达到公众期望的标准。各维度服务质量与公众的预期也有一定差距。

在相关性方面，公众对政务微信的整体直观体验不能达到预期，页面布局不够简洁易懂，发布信息的形式不够多样化、规范化，导航功能也不够一目了然，说明政务微信在设计友好性上还存在一些问题，需要更专业的团队来进行打造。

在完整性方面，公众的感知总体水平较高，但可靠性是公众能准确有效地获取政务信息和政务服务的最重要的保障，因此公众对可靠性的期望也很高，导致感知仍然未能达到预期水平。政务微信信息内容的质量还需要进一步的提升。

在可解释性方面，虽未达到公众满意的程度，但在该维度上公众的感知与期望分值都较高，差距最小，表明公众认为政务微信来获取政务信息或政务服务的方便。

## 5 智慧政府的建议与展望

“智慧政府”不仅利用物联网、云计算、移动互联网、人工智能、数据挖掘、知识管理等技术，还强调以用户创新、大众创新、开放创新、共同创新为特征创新 2.0 方法论，提高政府办公、监管、服务、决策的智能化水平，形成高效、敏捷、便民的新型政府。“智慧政府”是电子政务发展的高级阶段，是提高党的执政能力的重要手段。

随着物联网、云计算、移动互联网、Web 2.0 等新一代信息技术飞速发展，电子政务正由电子政府到“智慧政府”转变。智慧政府不仅强调新一代信息技术的应用，也强调以用户创新、大众创新、开放创新、共同创新为特征的创新 2.0。现代政府事务日益复杂，传统政府的智能水平已经难以应付这种新的形势，必须建立“智慧政府”。

政府的四大职能是经济调节、市场监管、社会管理和公共服务。“智慧政府”就是要实现上述职能的数字化、网络化、智能化、精细化、社会化。与传统电子政务相比，“智慧政府”具有透彻感知、快速反应、主动服务、科学决策、以人为本等特征。

在智能监管方面，智能化的监管系统可以对监管对象的自动感知、自动识别、自动跟踪。例如，在主要路口安装具有人脸识别功能的监视器，就能够自动识别在逃犯等；在服刑人员、嫌疑犯等身上植入生物芯片，就可以对他们进行追踪。智能化的监管系统可以对突发性事件进行自动报警、自动处置等。例如，利用物联网技术对山体形变进行监测，可以对滑坡进行预警。当探测到火情，建筑立即自动切断电源。智能化的监管系统可以自动比对企业数据，发现企业偷逃税等行

为。智能化的移动执法系统可以根据执法人员需求自动调取有关材料,生成罚单,方便执法人员执行公务。

## 6 参考文献

停用词集 <https://github.com/6/stopwords-json/blob/master/dist/zh.json>

- [1] 魏姮清. 基于机器学习的政务微博评论情感分类研究[J]. 现代商贸工业, 2020, 41(11):161-162.
- [2] 刘哲思. 基于 SERVQUAL 模型的政务微信公众服务质量评价研究[D]. 湘潭大学, 2017.
- [3] 李国祥. “智慧政府”背景下公共服务供给改革研究[D]. 南京大学, 2016.
- [4] 于冠一, 陈卫东, 王倩. 电子政务演化模式与智慧政务结构分析[J]. 中国行政管理, 2016(02):22-26.
- [5] 康红霄. 智慧政务模型构建及其推广研究[D]. 燕山大学, 2015.
- [6] 李超民. 治理现代化视阈中的智慧政务建设[J]. 社会主义研究, 2014(04):81-88.
- [7] 赵玓, 陈贵梧. 从电子政务到智慧政务:范式转变、关键问题及政府应对策略[J]. 情报杂志, 2013, 32(01):204-207+197.
- [8] 徐光明. 省教育厅认真做好政务环境评议评价群众意见建议回复工作[J]. 江西教育, 2003(23):10.