对文本资料的数据挖掘及相关应用

摘要：

近年来，随着互联网技术的迅速发展，互联网技术逐步应用在了普通人日常生活的各个方面。以“智慧政务”为代表的互联网民政系统的日常使用日益增加，而自然语言处理等深度学习算法可以极大地提高政务系统的服务效率，给市民和工作人员带来更方便的体验。

本文研究了基于深度学习领域的自然语言处理对于大量文字文本的处理能力和应用能力。基于 python 编程语言，使用短文本分类工具 TextGrocery 实现“智慧政务”系统所需要的功能，并对 TextGrocery 所建立的模型的使用情况进行评估，并建立合理可行的热度评价模型。

关键词：自然语言处理，短文本分类，f1-score，热度评价

目录

第一章 绪论

1.1 选题背景

近年来，随着社会的迅速发展，互联网的深度应用、大数据、云计算、人工智能等技术得到迅猛的发展，新技术催生了大量的新鲜事物，并将其广泛地应用到了社会的方方面面。同时，新技术也下沉到与互联网技术关联不大的传统行业，大数据的广泛应用和其与人工智能的协同发展，让着两样新技术与传统产业融合成为可能，原有旧的产业与工作使用新技术以提高工作办事效率成为必然事件。

传统的互联网政务系统是政务系统与互联网技术的直接组合，市民可以直接在网上办理与线下相同的民政事务。但互联网政务的各项工作（包括留言处理，信息分类等）主要由政务系统内的工作人员由人工实现，工作繁琐，效率低下。政务系统工作日常工作所收集的文本资料，本身具有潜在的数据属性，但并不能直接使用，需要通过特定处理才能表露出来；而大量文本资料集合所形成的大量语料集在这样特定的数据转换手段下则可转化为能够表现其数据特征的大量数据集，从而满足人工智能的训练条件与大数据的分析工作，进而训练出符合工作要求的、可替代简单人工分类的分类模型和简单人工排名的排序模型。

分类模型与语料排序模型可以在相当程度上减轻政务系统审核人员的工作压力，提高工作效率，减少工作成本；同时也让市民感受到更舒适的体验，更便捷的服务。本文即将讨论实现的相关模型使用

python 编程语言实现，同时采用多种实现方法。

1.2 技术的发展状况

“智慧政务”与原有的互联网政务系统的区别，主要体现在“智慧”二字上。智慧政务系统的核心工作是依托大数据平台、数据处理组件、人工智能组件和数据安全组件等工具，对包括基础数据库与功能数据库等源数据进行收集和处理。为了让计算机实现上述的工作，需要相关的算法进行支持。机器学习作为一种使用算法进行数据解析的方法，对智慧政务系统的发展产生了积极的影响。在此基础上产生的深度学习则是在大数据处理与分析工作日趋成熟的今天逐步发展起来的一种实现机器学习的技术。深度学习技术需要海量的数据用于模型训练，最终实现各种原本难以实现的任务。

自上世纪 50 年代，人工智能概念提出算起，人类对人工智能的研究与实现付出了大量的精力，并产生了机器学习等分析方法。而机器学习等技术的进一步细分发展和大数据新兴技术的产生又催生了近年来深度学习等分支。目前机器学习与深度学习已经应用在诸多领域，并反哺了人工智能的进步。“智慧政务”系统正是这些技术与社会现有工作融合的一个缩影。而在本文探讨的“智慧政务”系统中，将使用深度学习的一个分支技术自然语言处理，以处理海量的日常使用语言，并将其转换、处理、分析、建模、训练，以满足现代工作的需要。

第二章　基于自然语言处理的标签分类模型

原有政务系统工作人员在处理网络问政平台的群众留言之前，需要先根据留言的内容对其进行标签分类，并根据标签把不同分类的留言递交到相应的政务部门处理。在本章当中，我们讨论如何使用编程语言和恰当的算法实现标签分类模型。

2.1 标签分类标准

首先是明确标签的内容。在实现模型工作之前，首先需要考虑的是什么样的源语料是可以用于处理的，什么样的素材是与模型需要达到的目标无关的。因而在模型构建工作开始前，先要构建出符合要求的分类标准。使用 python 编程语言对附件一给出的“城市建设和市政管理”表进行导入和处理后，我们得出以下内容：标签共分为三级，其中一级标签 15 个，二级标签 103 个，三级标签 390 个。内容如下：

（1）一级标签：

0	城乡建设
1	党务政务
2	国土资源
3	环境保护
4	纪检监察
5	交通运输
6	经济管理
7	科技与信息产业
8	民政
9	农村农业
10	商贸旅游
11	卫生计生
12	政法
13	教育文体
14	劳动和社会保障

（2）二级与三级标签：（左为二级标签，右为三级标签）

0	安全生产
1	城市建设和市政管理
2	城乡规划
3	村镇建设
4	工程质量
...	
98	就业培训
99	劳动保护
100	劳动关系
101	社保基金管理
102	退休政策及待遇

0	事故处理
1	安全生产管理
2	安全隐患
3	园林绿化环卫
4	城管执法
...	
385	退休人员待遇
386	内部退养人员待遇
387	退休政策
388	退休金发放
389	病退及提前退休人员待遇

其次是标签的简化与选择。我们注意到，二级和三级标签数目繁多，过于冗杂，在实际的初次分类工作中如果过分精细地划分标签分类并不利于政务工作的展开，反而会降低初分类的工作效率。因此我们选择着重对一级标签建立分类模型。

2.2 文本素材的预处理

中文文本的处理与英文文本不同。英文是一个一个词之间有着空格隔开。而中文不一样，中文没有空格隔开，我们需要将其变为与

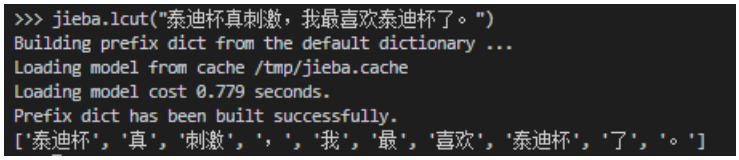
英文一样，一个词语接着一个词语的模式。用下面这句话作为示范：

泰迪杯真刺激，我最喜欢泰迪杯了。

而经过分词处理后，我们希望得到的是：

泰迪杯 真刺激 ，我 最喜欢 泰迪杯 了。

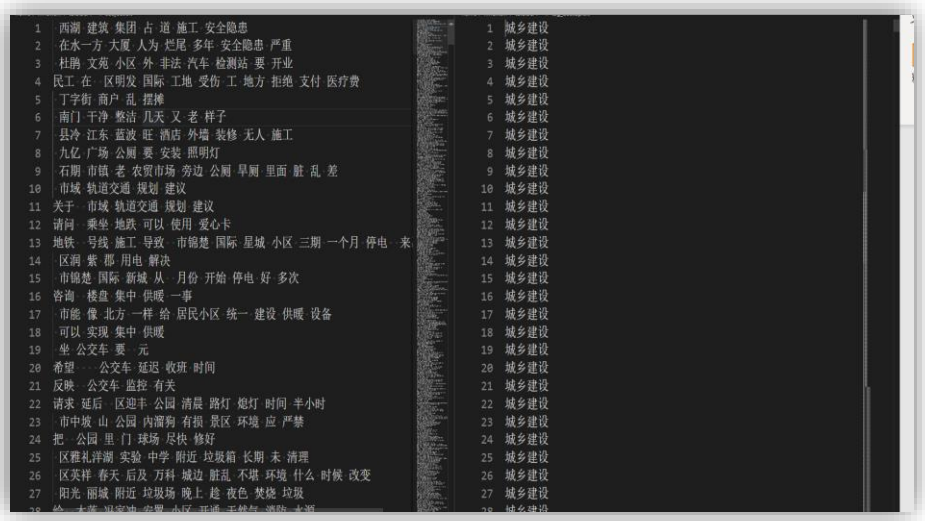
这样一来每个词语之间就有了明确的分界。为了实现这样的分词效果，我们决定采用 python 的 jieba 包来进行分词操作。Jieba 是一个较为好用的中文分词工具包。下面是 jieba 分词的结果。



可以发现，像“逗号，句号，真，最”这样的词语，对于我们进行标签分类是没有什么作用的，我们应该去掉这些词。再就是附件 2 中的文本包含大量的地点。在构建分类模型时，要是采用的信息大量包含了某地点，机器可能会误认为地点是该地点会是某个标签的特征词。为了避免上述的问题，我们还需要对文本进行一定的精简。构建了停用词表去除了如“逗号，句号，真，最”这样的词语。采用了正则表达式删除了文本中的英文与数字。达到了精简文本，提高模型学习数据的质量的目的。

在进行文本预处理时，我们发现附件二中只给出了 6 种标签的学习数据。我们决定到各大论坛去爬取他们的讨论贴标题，用来充当其余标签的学习资料。

最后将处理好的文本分为《subject》和《tag》。方便输入到模型中。



《subject》和《tag》的内容

2.3 分类模型的实现

模型我们采用的是 python 下的 TextGrocery 包，TextGrocery 包是一个高效易用的中文短文本分类工具。其效率较于 scikit-learn(朴素贝叶斯)要高出不少。

采用 TextGrocery 包中的 grocery.train()函数进行训练。grocery.train()的传入要求的是形如

train_src[i] = [tag[i],subject[i]]的二维列表。于是按这个格式导入文件。等待训练完成后，保存该模型，方便下次调用。

2.4 分类模型的评判

我们采用(F1 − score)评判方法来对我们的模型进行衡量。它是精确率和召回率的调和平均数。F1 − score的定义式：

$$F1 = \frac{1}{n}\sum_{i=1}^{n}\frac{2P_iR_i}{P_i + R_i}$$

其中：$P = precision$（查准率），$R = recall$（召回率），其定义为：

$$F1 - score = （precision * recall）/（precision + recall）$$

$$Precision = True\ Positive / (True\ Positive + False\ Positive)$$

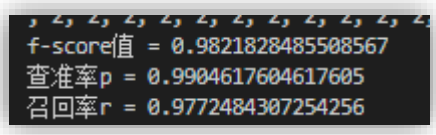
$$Recall = True\ Positive / (True\ Positive + False\ Negative)$$

在 python3 中，我们采用 f1_score 包进行计算。首先我们要将一级标签进行量化，即将其转化为数字，以方便用 f1_score 包进行评分。我采用的对应原则是根据标签在附件一内的排序，给予其对应的数字。考虑到样本之间数量的不平衡性，我们采用

$$f1\_score(y\_true, y\_pred, average = 'weighted')$$

函数来进行评分。

经过计算得到$f1 - score$的分数高的出乎意料，同时查准率P的值和召回率R的值同时逼近最大值 1，达到惊人的0.990和0.977，而一般的PR模型的典型特征是P与R的值呈现反比例相关。经过大量资料的比对，我们认为可能是由于测试数据过少，模型无法得到充足数据的训练，使得预测偏离实际。同时训练模型使用的测试资料的各标签文本差距较小，不利于机器学习这些标签的不同点。我们认为，如果输入更多的不同标签的测试资料用于训练阶段，可以提高机器对这些部分类似的不同标签的识别能力，从而让模型做到准确识别。

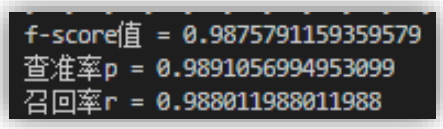


f1-score 的值

对于初始素材的不足，我们认为可以通过其它的手段补充其短板。我们在$f1-score$的计算过程中引入权重，即在

$$F1=\frac{1}{n}\sum_{i=1}^{n}\frac{2P_iR_i}{P_i+R_i}$$

的加权平均步骤中，按照输入的不同标签的训练资料数量占整体训练资料数量的权重，重新计算$F1$的值，结果如下：



改进加权后的 f1-score 的值

可以看到在略微改变加权步骤的计算后，$f1-score$值出现了微小的上升，虽然变化极小（数量级为10^{-3}），但是上升的趋势说明$f1-score$的计算改进的确有效；查准率P的值和召回率R的值同时向中间值0.98靠拢，虽然仍处于一个极高的数值，但其变化趋势说明P值和R值同时向R模型曲线的对称线$y=x$靠近，说明改进后的$f1-score$要优于原有的$f1-score$评判方法。

2.5 分类模型的总结

标签分类模型是一种对自然语言进行处理的模型，其最大的难度是要对原有的语料进行加工，使之能够成为机器可以处理的有效素材。在完成原始语料的分类标准后，其次是对符合标准的语料的预处理，产生可用于训练模型的材料。对模型的选取也很重要，适当的模型可以提高机器的效率，同时优化产出。最后是对模型的检验，选择出优秀的标签分类模型。

第三章　基于自然语言处理的热度评价体系

“智慧政务”作为一种智能化的信息处理系统，其工作除了为工作人员减少不必要的劳动负担外，还需要对市民反映的大量信息进行筛选，确定其中的热点信息，以相对的最小时间成本和最高效率解决最多人关心的问题。本章我们将讨论热度评价体系的实现与应用。

3.1 对文本素材的分析

素材附件三给出了内容与素材附件二类似的 30 条留言信息，其主体信息为语料较短的“留言主题”和具有大量语料内容的“留言详情”。

同时，每一条留言都有其对应的留言时间、点赞数和反对数。对于反对数，由于每一条评论的点赞数数目都是 0，因此反对数这一常量在此问题中并不具有引入的价值。而每一条留言的点赞数数目不尽相同，因此我们将其纳入热度评价体系的考虑范围内。同时留言素材的主体——语料具有可观的文本信息，其将会是我们建立此体系的重要因素。

另外，留言时间作为记录数据，本来不具有太多的信息。但在热度评价体系初步建立后，我们将探讨将留言时间的记录密度作为一个外生变量引入，用于检验我们已经建立起来的评价体系，并进行优化。

3.2 热度评价体系的实现

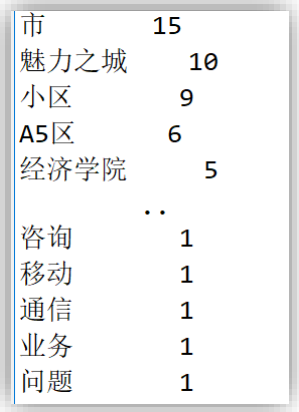
3.2.1 对文本素材的分词处理

素材附件三的“留言主题”和“留言详情”分别是概括性的简短语句和整段的、叙事性的段落文字，但二者的语句中所含有的名词、动词、固定地名、特定称谓却是其语句实际含义的承载体，这也是我们需要分离出来，用于模型训练和统计的实际使用材料。因此，我们对“留言主题”和“留言详情”单独摘取出来后，作相同的分词处理。我们使用 python 中的 jieba 工具包进行分词，区分开具有不同意义的实词，并对语料中的虚词、语气词、标点符号和其它无意义的词语进行剔除，得到预处理文件。

3.2.2 对分词的的排序统计与词云实现

对已经处理好的词语集进行数量统计，得到以下数据：

“留言主题”:



词量在 10 以上的“留言详情”:

’市’: 39, ’小区’: 21, ’月’: 19, ’年’: 17, ’居民’: 17, ’号’: 17, ’油烟’: 16, ’服务’: 16, ’系统’: 16, ’领导’: 15, ’严重’: 15, ’学生’: 14, ’说’: 14, ’窗口’: 14, ’学校’: 13, ’

烧烤’: 13, ’一个’: 13, ’社保’: 12, ’元’: 12, ’之前’: 11, ’记录’: 11, ’养老保险’: 11, ’套餐’: 11, ’实习’: 10, ’请’: 10, ’A5 区’: 10, ’影响’: 10, ’2019’: 10, ’标准’: 10

容易看出，使用 jieba 分词工具后，语句被拆分成了有实际意义的单个词语。其中“市”、“小区”、“A5 区”等词语同时出现在两个统计排序中，说明两部分反映出共同的问题。但同时，单个词语脱离了上下文语境后，理解单个词语变得困难，无法从词量统计中获得更多的意义。

在这样的情况下，词云则展示了同一区域内大量出现的各色热词。在词云中，字体越大的词语，表示其在次数统计中出现的次数越多，越小的词语表示其出现的次数越少。虽然这些词语仍然是脱离语境的单个词语，但在出现数量最多的数个热门词语汇集在一起，则更容易理解这些词语的逻辑联系与相关语境。

以下分别是“留言主题”的词云和“留言详细”的词云：

“留言主题“词云

“留言详细”词云

可以发现，词云相比于直接的词数统计，词云同样具有表现词数多少的功能；但在直观感受上，词云给予了阅图者更为丰富的信息，包括直接的词数权重、热词之间的有关联想。除此之外，出现在词云中的“显眼词汇”（也就是在直接词数统计中的大数量词语），也可以作为热度评价标准的重要依据。对于含义相近或者热度分类一级指标相近的不同留言，大数量词汇的是否出现与出现个数的多少将作为热度分类二级指标引入到评判过程当中。

3.2.3 点赞数对热度评价体系的主要影响

在讨论了词数排名和词云的参考价值后，我们指出其可作为热度评价体系的二级指标，但一级指标仍没有讨论，本小节我们讨论点赞数作为热度评价体系一级指标的重要性与合理性。

对 30 条留言进行汇总排名，可以发现：

（1）有 60%的留言的点赞数为 0，我们可以认为这些留言为无效留言，或称为未成为问题留言。由于每位留言者面对问题时可能具有主观角度、直接利益相关等因素，往往会把一些普通问题夸大，认为这是一个对自己而言的巨大、急迫的、难以解决的难题，并最终求助于社会协助系统。但这类问题往往难以得到社会其它成员的认可，这样的状况表现在政务系统留言上就是某些留言获得零点赞数。而在本素材中所有留言的反对数都为 0，因此没有把反对数作为点赞数的相反变量进行考量，而是直接排除。

（2）点赞数为 1 的留言和点赞数为 3 的留言在留言总体中各占 13%，我们把这类低点赞数的留言称为“一般问题留言”，其内容的确客观反映了存在问题，并得到了一些市民的认同。同时这些留言也契合了词数排名前列中出现的一些“热点词汇”，但这些留言仍然与接下来我们要讨论的另一些留言有着相当的差距。

```
('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),
('请求A市地铁2#线在梅溪湖CBD处增设一个站', 3),
('魅力之城小区临街门面油烟直排扰民', 3),
('A市经济学院强制学生外出实习', 3),
('J4县供销合作社在岗失业职工追缴社保', 2),
('A5区劳动东路魅力之城小区临街门面烧烤夜宵摊', 1),
('A5区魅力之城小区一楼被搞成商业门面，噪音扰民严重', 1),
('A市能不能提高医疗门诊报销范畴', 1),
('A市经济学院组织学生外出打工合理吗？', 1),
```

（一般问题留言：点赞数为 1～3 的留言）

（3）点赞数为 4 到点赞数为 9 的留言，我们将其称为“热点留言”。这类留言的点赞数从最小点赞数到最大点赞数，都是“一般问题留言”所收到点赞数的 3 倍或以上，说明其具有极为强烈的社会矛盾体现特性，可以作为热点问题。另一方面“热点留言”的内容也与词云中的“热点词汇”大部分重合，如“魅力之城”、“经济学院”“污染”等词语。

```
('A市经济学院体育学院变相强制实习', 9),
('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，急需处理！', 6),
('A5区劳动东路魅力之城小区油烟扰民', 4),
('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),
```

（“热点留言”）

综上，点赞数可作为本热度评价体系的一级指标，在出现有一级

指标相同或难以区分的情况下，可以使用词数统计和词云等的符合程度作为热度评价体系的二级指标。以上是热度评价体系的定性分析。

除定性分析外，我们还需要定义具体的可计算的热度指数以用于定量计算。我们可认为留言的热度与评价体系的一级指标X（点赞数）、二级指标Y（词云符合度）与热度指数η具有指数相关性，相关系数为K。我们给出以下定义式：

$$\eta = e^{K}(AX + BY) + C_0$$

其中A为点赞权重，表示一级指标X在此定义式中的权重，$0 \leq A \leq 1$；而B则是词数权重，表示二级指标Y在此定义式中的权重，$0 \leq B \leq 1$；C_0是一个实值常数，代表留言本身自带的热度属性，我们定义：

$$C_0 = \sum_{i=1}^{n} \frac{X}{n}$$

在素材附件三中，我们有$A = 0.75$，$B = 0.25$，$C_0 = 1.23$，$K = 1.5$。由此我们得到基于素材三的热度评价指标：

$$\eta = e^{1.5}(0.75X + 0.25Y) + 1.23$$

使用此评价体系所筛选出的热点问题 Top5：

热度排名	问题ID	热度指数	时间范围	地点/人群	问题描述
1	360114	34.842668	2017年6月	A市经济学院/学生	A市/经济学院/学院/变相/强制/实习
2	360108	25.87929	2019年8月	魅力之城/居民	A5区/魅力之城/小区/夜宵/污染/空气/处理
3	360101	18.036334	2019年7月	魅力之城/居民	A5区/魅力之城/小区/油烟/扰民
4	360107	13.554645	2019年7月	魅力之城/居民	A5区/魅力之城/小区/夜宵/污染/附近/空气
5	286572	11.3138	2018年10月	梅溪/乘客	A/市/地铁/2/梅溪湖/CBD/增设/站

3.3 引入时间密度对热度评价体系的分析

在上面的小节中我们实现了对热度评价指标的定义、计算以及对热度评价体系的初步构建。在已有的点赞数、词数统计、留言自身平

均热度外，我们引入时间密度的概念。

观察上表，我们可以发现，五个热点问题的时间跨度有 2 年之广，从 2017 年一直到 2019 年都有记录。而排名为 2-4 的热点问题的时间记录都集中在 2019 年的 7-8 月份，这部分的时间差只占总体的时间差的 10%，却占据了 60%的热点问题数量。在这极不均匀的时间分布中，我们认为：在均匀的时间轴中不均匀分布的时间记录，可以近似地看作其热点问题的时间密度。在单位时间内分布的热点问题记录越多，可认为此单位时间内的时间密度越大，换言之热点问题越突出，越重要，也就是越“热”。

引入时间密度概念后，我们定义时间密度指标及其计算：

T为总体热点问题的时间差，t为某几个需要讨论的热点问题的时间差；M为总体热点问题数量，m为需要讨论的热点问题个数。有：

$$S = \frac{m/M}{t/T}$$

例如在刚才的讨论中，排名为 2-4 的热点问题的S值为$S = 6$，而另两个热点问题因为在时间轴上相对独立，因而$S = 1$。具体的S值的引入因为缺乏大量的数据，因而不展开过多的讨论，但S值可以考虑作为为热度评价体系中的优化指标而引入。

3.4 热度评价体系的实现与总结

对“智慧政务”系统的实现在根本上是使用一系列的算法和模型，依托编程语言，实现我们需要的功能和工作。我们从自然语言处理中的具体语料意义选择出发，具体讨论了分词的必要性，分词的处理和

解析；分词的数据应用，包括分词的数据化和数据的可视化，并考虑分词数据移植到热度评价体系中的价值；讨论了最重要变量的点赞数的实际意义和应用价值；量化了热度评价体系的评价指标；最后提出时间密度的概念，并对已有的评价体系进行初步应用。

可以看出，基于自然语言处理的热度评价体系是一个多元化的评价体系，实际问题可以根据实际所得的有效数据进行提炼，从而得到有效展现语言特征的评价体系。同一语料集中还可以通过不一样的角度分析建立多样的优化指标，从而使单一评价体系的模型更为多样、能够适应更为复杂的情形。

第四章 基于自然语言处理的文本评价体系

4.1 对政务意见反馈的分析

作为“智慧政务”系统不可或缺的一环，相关部门的意见答复是必不可少的解决问题的一栏。但回复的意见并不一定是可以实行的解决方案，本章将对政务系统回复意见进行分析考量。

4.2 政务意见反馈系统的主要考量

评判一条意见是否是好的意见，我们将从以下几点考虑：

（1）相关性。如果答复意见和提出的问题的主要内容相关性不大，那这样的意见答复对提问人来说就没有参考价值。因此评判意见的优劣首先考虑其相关性。

（2）完整性。答复意见与原问题相关，但并不完整，也就是只有部分答复，无法完整地解决整个问题，只能部分解决。换言之，问题还在，只是换了一种形式。因此意见的优劣需要考虑其完整性。

（3）可解释性。如果一个意见提出来无法被理解，那这个意见和没有被提出来是等价的。当意见不存在时，讨论意见的价值与优劣是毫无意义的。

以上几点是对意见反馈的简单评判，在进一步的构建文本评价体系中，我们可以把这几个方面量化为一级指标，并对其进行建模。除此之外，我们还可以添加若干二级指标，使模型对各个因素的考虑更为完善，在此因能力问题不展开赘述。

第五章　总结

基于深度学习的自然语言处理是“智慧政务”系统的核心所在，其作用主要是支撑本文实现了对自然语言进行分类的标签分类模型，对多种因素进行汇总分析的热度评价模型，以及初步探讨了文本处理体系。

自然语言处理特指用机器来处理人类日常使用的自然语言，让机器能够“看懂”人类的语言，但实际上，自然语言处理是一个很宽泛的概念。在标签分类模型中，选择不同的模型工具会导致不同需要的语料集预处理，而不同的语料预处理工作又会导致产生不一样的模型训练材料。因此选择适当的实现模型，同时也是选择有不同偏向的处理方法和训练素材，并最终反映在模型的成品中。

一旦选定某个自然语言处理模型，并不代表着最终的模型和评价体系已经固定。对不同的变量进行选择，同时对变量在体系中的权重进行衡量，赋予不一样参数值，最终也会得到多样的评价结果。

同时，原始数据资料的素质与数量也是不可轻视的重要元素。原始数据素质过差，预处理得到的训练资料会难以支撑模型的训练与学习过程，最后成型模型的结果难以支撑正常的使用需要。或者可用的绝对训练数据过少时，用于预处理的训练资料过少，模型无法得到足够的训练，当面对部分“生僻”的条件时，模型也难以完成工作。

在以上章节的建模工作中我们也遇到了这两种问题，也是以后需要改进和学习的方面。

参考文献：

[1] 2shou　TextGrocery GitHub [DB/OL]

https://github.com/2shou/TextGrocery

[2] yucen 机器学习中的 F1-score 博客园 [DB/OL]

https://www.cnblogs.com/yucen/p/9912063.html

[3] ybdesire 详解多分类模型的 Macro-F1/Precision/Recall 计算过程 [DB/OL]

https://blog.csdn.net/ybdesire/article/details/96507733

[4] abel_cao 10 行 Python 代码的词云 [DB/OL]

https://www.jianshu.com/p/2c1748c69204,2017-03-06

[5] 一只 IT 小小鸟 Python 词云库 wordcloud 显示中文 [DB/OL]

https://blog.csdn.net/qq_34777600/article/details/77455674

[6] Kumata Python 排列函数：sort、sorted [DB/OL]

https://www.cnblogs.com/kumata/p/9058405.html

[7] adnb34g 智慧政务大数据局解决方案之 DKH 详解[DB/OL]

https://bbs.51cto.com/thread-1562958-1.html

附录：

数据来源：

用于模型训练的语料集来源：

上海市民政局：http://mzj.sh.gov.cn/gb/shmzj/node4/node10/n2756/index.html

民主与法制网：

http://www.mzyfz.com/cms/yifaxingzheng/fazhigongzuo/gongzuodongtai/html/1459/

农村新闻网：

http://www.nongcun5.com/news/4/

通信人家园：

http://www.txrjy.com/forum.php

第一财经：

https://www.yicai.com/search.html?keys=中国经济论坛

社旗县国土资源局：

http://www.sqgtj.gov.cn/a/xwdt/

国家监察委：

http://www.ccdi.gov.cn/yaowen/index.html

燕山大学党务公开：

https://xxgk.ysu.edu.cn/dwgk.htm