“智慧政务”中的文本挖掘应用的分析与实现

摘要

本文以政府提供微信、微博、市长信箱、阳光热线等网络问政平台所得到的留言数据为研究对象。结合目前主流的中文自然语言处理方法和文本挖掘技术，建立了对留言以及回复内容进行分类、评价等的模型。首先，建立了高准确率的留言分类模型并得到可用的分类器，提供 F-Score 进行评价，并依此实现对于留言进行快速一级分类。其次，针对对于数据中的热点问题建立搜索模型，定义了多个指标综合的热点评价标准，在留言中得到反映热度最高的五个热点问题，建立了热点问题留言明细表。最后，结合分类、关联预测等等方式，对留言以及回复内容进行文本挖掘，综合留言时间和内容等信息，提供了对留言回复评价的多因子评价模型。综上所述，本文模型针对问政平台所提供的服务方式提供了一些可以实际运用的优化，分类与热点提取可以提高平台服务人员的行政办公效率，问政群众可以尽快得到相应回复，提高对平台满意度。

关键词：文本分类、多因子评价模型、文本挖掘、自然语言处理、K-means、ERNIE、Bert

目录

0. 引言

0.1 问题重述

(一) 根据附件 2 给出的数据以及附件 1 的分类标签，针对网上问政平台得到的留言，建立关于留言内容的一级标签分类模型。并且使用 F-Score 对分类方法进行评价。

(二) 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，给出排名前 5 的热点问题，并分别得出热点问题表和热点问题留言明细表。

(三) 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。

0.2 本文主要工作和创新点

本文基于题目所给的数据以及信息，结合相关领域已有的研究工作的基础，如分词、关键词提取，机器学习以及分类、类聚的思想，建立了一套对于政务平台留言分类，识别热点问题并生成热点信息表，对留言及回复进行初步评价的评价系统。

（一） 机器学习模块效果的角度。本文选用多种分类模型进行训练与测试，并且结合题目数据特点进行微调，观察多个分类模型的表现，通过准确度与 F-Score 的表现进行评价，最后选择表现最优异的模型。并且在该模型基础上，实现了对应的实用的分类器，可对为分类数据进行分类，模型后续可继续拓展，新增数据后获得更强大的分类能力，具有实用性和拓展性。

（二） 文本聚类与命名实体识别。本文使用 Jieba 对语料进行分词，采用 k-means 算法完成对事件文本的聚类；选用 BiLSTM + CRF 的框架加入 BERT 模型作为 embedding 的特征获取层的方法完成命名实体的模型的构建，使用改模型对文本中的地址/人群信息进行抽取。

（三） 通过多指标建立综合评价模型。使用文本相似度的算法检验答复意见的评价质量。通过文本内含有的关键词的重合度来判别相关性。而留言子集小于答复子集时即可以认为该答复可以解释留言所需，即解释性得到满足。此外，通过格式匹配，判断留言答复是否完整。

1. 问题分析方法与过程

1.1 问题 1 分析方法与过程

1.1.1 流程图

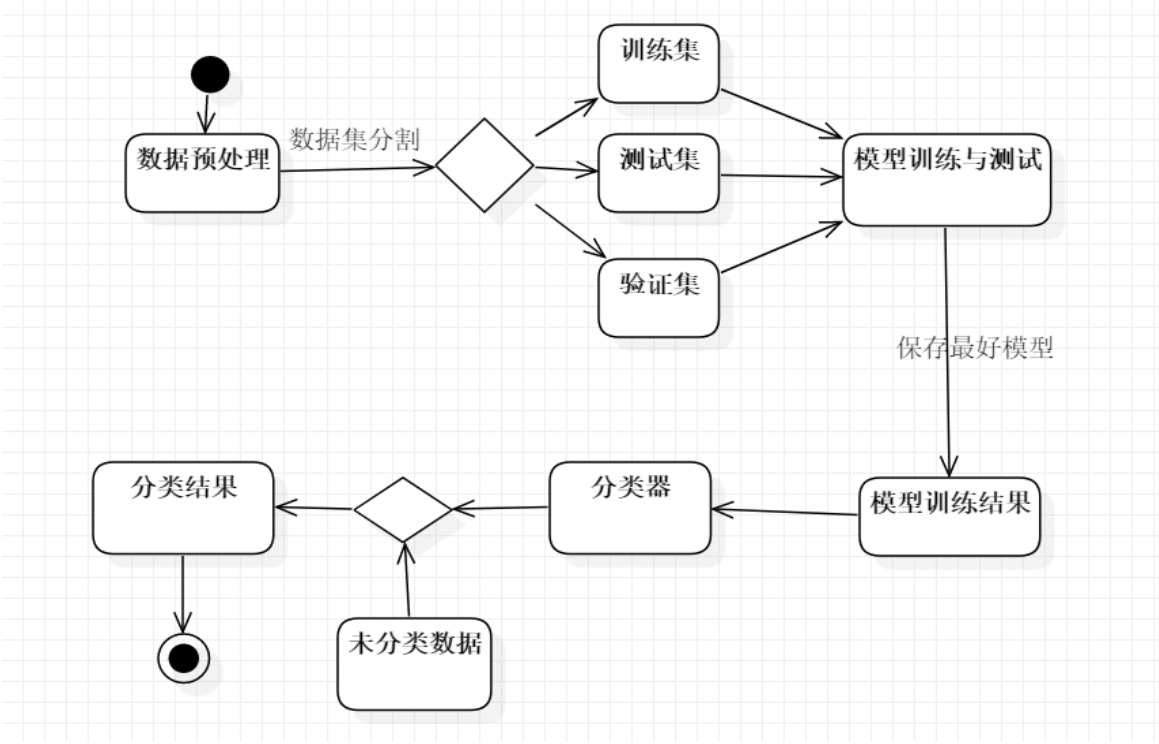


图 1-1-1 问题 1 流程图

1.1.2 数据预处理

题目所给数据中，数据格式不统一，数据存放形式不符合模型训练等问题，需要对数据进行预处理与一般清洗。比如训练模型需要提取的是数据中的留言主题、留言内容和一级分类标签，需要读取文件，逐项提取和保存。对于类型，需要建立对应的标签，以数字代替标签，即建立“分类标签-数字”的标签对映射，使用 0-6 代替原有的 7 个标签。另外，由于所给数据的类别仅有 7 个，与附件 1 中的 15 类不同，需要进行甄别与注意。防止类别错误引起分类模型错误。针对附件 1 的一级分类以及附件二数据，得到 class.txt 文件存放数据的分类标签以及对于数字标号，train.txt, test.txt, dev.txt,分别为训练集，测试集和验证集，三者数据占比为 8:1:1。其他对语句的处理均在模型内完成。

1.1.3 模型的训练与处理基本原理

本题的分类模型使用的是 ERNIE 语言处理模型。该模型通过多任务学习获取语料中的更多信息，包括：词、句共现，词法，句法，语义等。所以模型中基于成熟的 Word2Vec, GloVe 词语识别与理解技术，结合机器学习与神经网络，从数据中先训练一个最简单的模型，然后

增加一个训练任务，在上一个任务的参数的基础进行训练,上一个任务也继续参加训练，得到更好的语言分类模型，如此循环往复，得到一个预训练模型。在此模型的基础上，结合本题的数据，再进行训练，得到具有本题数据特征的训练模型。如图，模型数据训练过程简单示意图。

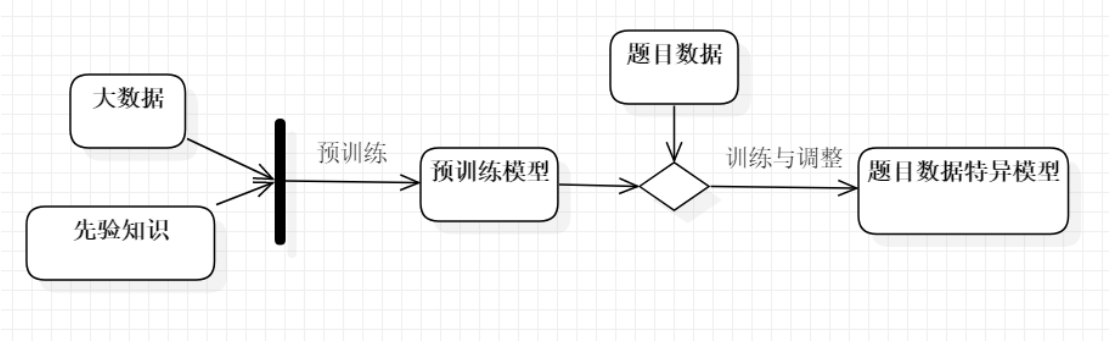


图 1-3-1 本题模型流程示意图

1.1.4 模型策略与分析

语言学习模型的发展大多需要引用或者改进现有技术，得到新的处理方式，提高了处理的效果，处理速度得到提升等。而本模型也有基于 Bert 模型的思想，并且在模型中得到了改进，更适合处理任务数据。

(一) MASK 策略

MASK 指的是掩盖一句话中的部分信息，让模型学习其中主体之间的关系。如路遥是《平凡的世界》的作者。当掩盖“路遥”时，模型就可以根据书名，联系到该书的作者，进而实现知识学习。

模型的 MASK 分三阶段进行。第一阶段为基础级，随机 MASK 一句话中的一个字。第二阶段将词语或词组搭配进行 MASK，然后让模型去猜测这些被掩盖的信息。这个过程，会将这些词组等编码进任务共享的 word embedding，后续任务可以得到这些已经学习的知识。第三阶段会将命名实体，如地名、物品名、人的称谓等词进行 MASK，故而模型在训练完成后，也学会了这写实体的信息。

通过三阶段的学习，模型可以将数据信息中的实体名词、事件词汇等信息，与题目中提供的一级分类标签建立对应关系，获得初步的分类能力。

(二)持续学习的策略

在本模型中，由大数据训练得到的预训练模型，在此基础上再加如题目数据进行训练。修正预训练模型中的一些参数，使得模型更适合题目的数据。

由于训练的任务是分批次进行，所以会面临的问题是，在修改参数的过程中，是否会将预训练模型得到的知识遗忘。针对这个问题，模型建立了持续学习的结构。任务划分为字层级、句层级和语义层级，针对新输入的训练数据可以划分不同层次的学习。通过使用大规模数据和先验知识持续构建无监督预训练任务，即构建不同词语、句式和语义学习的任务，所有任务均为自动感知与识别，不需要人工标注信息。

此外，模型不采用一般的一个个任务递进式学习，而是将所有任务进行多轮学习，多轮训练之后，得到的模型会同时具有先前学习的知识，有新学会的知识，避免了新学习的模型对先前数据的遗忘。前者虽然对新输入的数据处理效果好，但是使用先前数据对于模型测试，则会显现较差的效果，即出现了学了新的忘记了旧的

知识。而后者在这方面由多任务同时学习体现的效果较好，能够结合先前学习的知识持续学习，故而对前后数据处理的效果都很有卓越表现。比如本题数据加入预训练模型后，由于模型不会大量遗忘预训练模型的数据，对于本题提供的训练数据中没有的一些词汇但是先前学习过，在使用未知数据测试时，遇到先前遇到的语句，依然可以得到良好的识别分类效果。

(三)后续增强学习与拓展

得益于持续学习的策略，在训练完成的模型中，也可以加入其他预训练模型进行训练，使得已有模型可以学习更多新知识，具有良好的拓展可能性。

例如本题中，附件提供的一级分类有 15 类，而附件二的数据仅有 7 类。即现有模型只能对这 7 类的数据进行分类，其他类型的数据则会错误分类。当有其他类型的数据时，将现有模型当作预训练模型，加入新类型的数据进行训练，得到新的模型，获得了对新加入数据的识别分类能力，并且对先前数据分类能力不会出现大幅度下降，使得本模型具有良好的拓展性。

1.1.5 模型评价标准

本模型使用 F-Score 为主要评价标准。F-Score 定义为：

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

其中P_i为第 i 类的查准率，R_i为第 i 类的查全率(召回率)。

准确率（查准率）针对预测结果，表示的是预测为一个类的样本与全部样本的数量比，其的定义为：

$$\text{查准率} = \frac{\textit{分类正确的样本}}{\textit{样本总数}}$$

查全率(召回率)针对数据样本，表示的是预测为该类的样本与实际述语该类的样本的数量比，其定义为：

$$\text{查全率} = \frac{\textit{分类正确的样本数}}{\textit{数据集中该种类的样本总数}}$$

本题使用 F1-Score，认为精准度与查全率同样重要，即取得是精准率与查全率两者的调和平均数作为最终结果。该结果可以客观真实反映该模型对于一个数据分类的准确性，以及是否真的可以正确对该样本进行分类。

1.2 问题 2 分析方法与过程

1.2.1 流程图

为了实现对热点问题的挖掘，我们采用了以下分析思路：

- 采用 k-means 算法对附件 3 中的留言详情进行聚类。
- 统计各个类簇中的数据条数。
- 统计各个类簇中的点赞总数、反对总数。
- 使用 Bert 模型训练出命名实体识别模型(NER),对数据中的地址/人群信息

进行抽取。

- 使用各个类簇中的数据条数、点赞数总数、反对数总数来衡量热度指数。

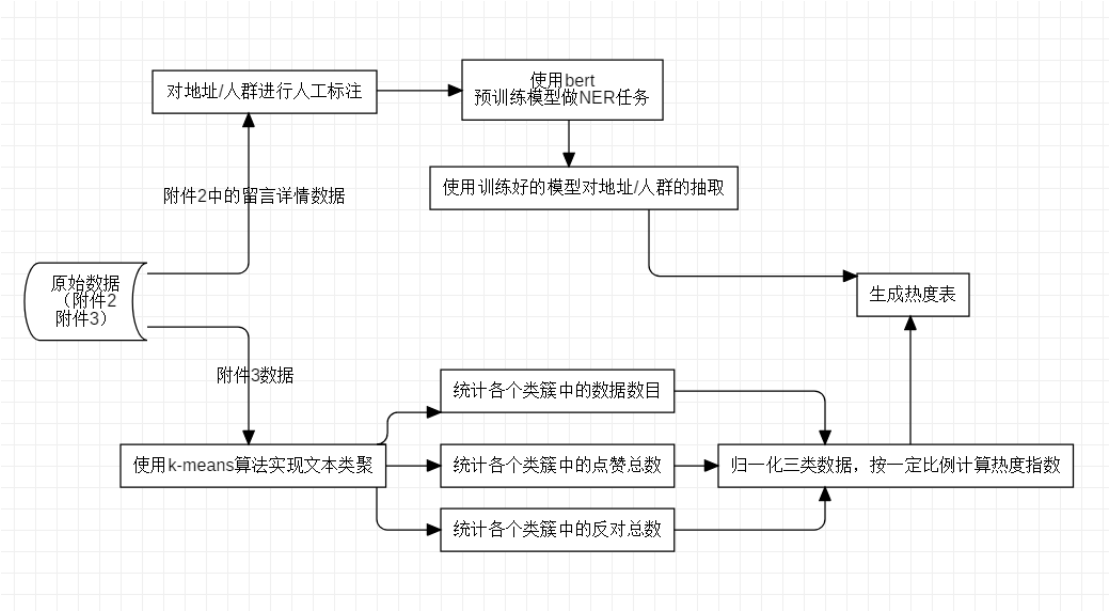


图 1-2-1 问题 2 处理流程图

1.2.2 数据预处理

1.2.2.1 Bert-NER 模型的训练数据

从附件 2 数据抽取留言详情并去空 ，使用标注工具对数据进行 BIO 标注：（1）B-NP：名词短语的开头（2）I-NP：名词短语的中间（3）O：不是名词短语。标注 LOC(地址)、PER（人群），标注示例结果如附件 annotation.data 所示。

1.2.2.2 k-means 算法聚类数据

需要对附件 3 数据进行去重、去空，存为需要使用的文件格式

1.2.3 Bert 预训练模型进行命名实体识别

由于数据中存在类似 A 市、B 市的虚拟数据，使用现有的库无法很好的抽取出地址、人群信息，因此需要使用 Bert 模型训练出一个用来抽取地址、人群信息的 bert 模型。Bert 模型是由谷歌 AI 团队发布的预训练模型，它在 NLP 领域中是里程碑式的进展，使用 Bert 预训练模型完成 NER 任务准确率非常高，并且训练速度也非常快。

1.2.4 K-means 算法

使用 k-means 算法对题目所给数据进行类聚。K-means 算法原理如下：

k-means 算法需要事先指定簇的个数 k，算法开始随机选择 k 个记录点作为中心点，然后遍历整个数据集的各条记录，将每条记录归到离它最近的中心点所在的簇中，之后以各个簇的记录的均值中心点取代之前的中心点，然后不断进行迭代，直到收敛。

算法具体描述如下：

1. 随机选取 k 个聚类质心点（cluster centroids）为

$$\mu_1, \mu_2, \cdots, \mu_k \in \mathbb{R}^n$$

2. 重复以下过程直至收敛{

对每一个样例 i，计算其归属的类：

$$c^{(i)} := arg_j^{min} \left\| x^{(i)} - \mu_j \right\|^2$$

对于每一个类 j，重新计算质心：

$$\mu_j := \frac{\sum_{i=1}^{m} 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^{(i)} = j\}}$$

}

1.2.5 归一化与热度计算方法

归一化就是把一组数（大于 1）化为以 1 为最大值，0 为最小值，其余数据按百分比计算的方法。

归一化的计算步骤如下：

1.找出一组数里的最小值和最大值，然后就算最大值和最小值的差值；

2.数组中每个数都减去最小值，再除去差值 r

热度计算方法：

1.计算聚类后各类簇中的数据条数

2.计算各类簇中数据的总点赞数、总反对数

3.归一化前面步骤的三类数据，使得数据被限定在一定的范围内（比如[0,1]），从而消除数据指标间的差异。

4.热度指数按一定比例选取上述归一化后的三类数据进行求和，作为热度值。（本文使用的是：各类簇内数据条数：各类簇中数据的总点赞数：各类簇中数据的总反对数=2:3:5 的比例）

综合以上方法，得到各个类簇的评分，通过评分排行确定最后的结果。

1.3 问题 3 分析方法与过程

1.3.1 流程图

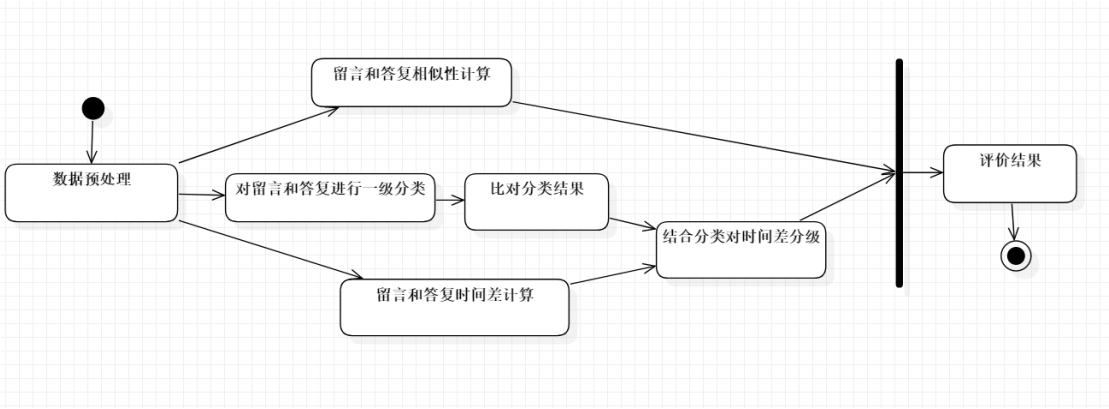


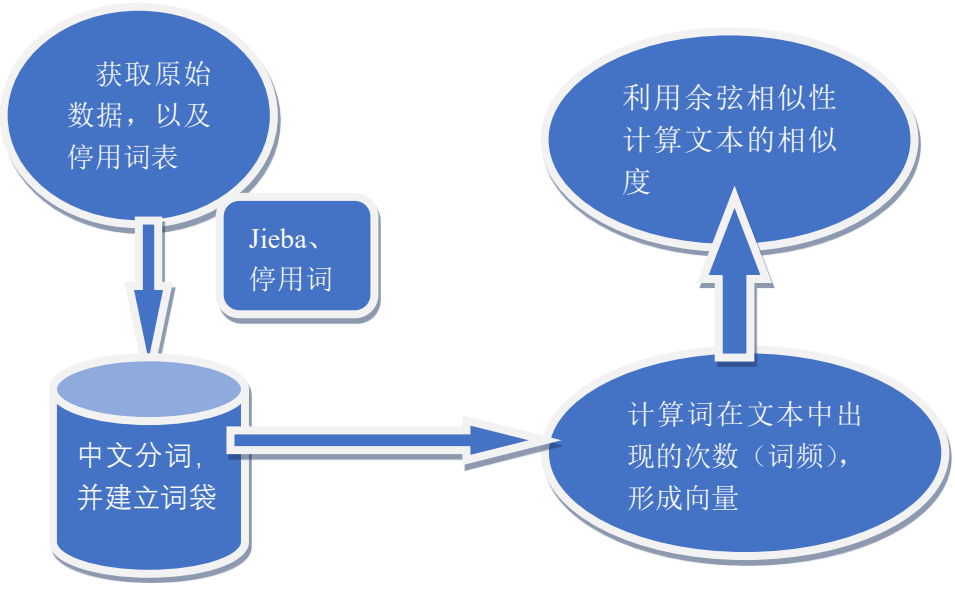
图 1-2-1 问题 3 处理流程图

1.3.2 数据预处理

对于附件四提供的数据，先进行了数据预处理。首先对时间进行处理，所提供的时间数据格式混合，解析得到有 str、time、datetime 类型，从表格中提取处理后进行统一化处理，统一处理成 datetime 的 Year-month-day 格式。对于留言与留言答复，先综合留言主题与留言内容，然后分别将留言与答复存储，存为 tsv 格式，等待输入分类器以及相似度模型。

1.3.3 留言与答复相似性计算

1.3.3.1 流程图



1.3.3.2 对留言与答复中文分词，去停用词

在进行中文分词之前，需要进行简单的去空处理，以提高数据的准确性。在获取数据之后，留言与答复的描述内容会分别存储在字符串中。而在 python 中的 strip()函数就可以直接为字符串去掉两边的空字符。

利用计算机处理数据，必须将数据转化为计算机可以识别的信息。在此之前应将留言与答复的描述内容的一长串中文文本字符，分割成规模更小的中文词语。而这里采用 python 中文分词包 jieba 进行分词。Jieba 分词工具采用了基于前缀词典实现高效的词图扫描，生成句子中汉字所有可能成词情况所构成的有向无环图 (DAG)；采用了动态规划查找最大概率路径, 找出基于词频的最大切分组合；对于未登录词，采用了基于汉字成词能力的 HMM 模型，使用了 Viterbi 算法。

在进行中文分词的过程中，去除了停用词——以避免不必要的检索，提高信息处理的效率。处理该问题的停用词包括中文的功能词等。

1.3.3.3 计算词频，生成词向量，计算相似度

对留言与答复进行分词之后，生成了一对词语列表，但还需要将它们转化为向量。词频选择该词在每篇文本中出现的次数。

$$\mathrm{TF} = \frac{\textit{词在文中出现的次数}}{\textit{文中词的总数}}$$

在统计词频的同时，使用词频形成两个词向量。

获得了向量之后，利用余弦相似性，将每一对向量进行相似度计算。两个向量间的余弦值可以通过使用欧几里得点积公式求出：

$$\mathrm{a}\cdot\mathrm{b} = |a|\cdot|b|\cdot\cos\langle a,b\rangle \qquad (1)$$

给定两个属性向量，A 和 B，其余弦相似性θ由点积和向量长度给出，如下所示：

$$\text{Similarity} = \cos\theta = \frac{\mathrm{A}\cdot\mathrm{B}}{|\mathrm{A}||\mathrm{B}|} = \frac{\sum_{i=1}^{n}\mathrm{A}_i\times\mathrm{B}_i}{\sqrt{\sum_{i=1}^{n}\mathrm{A}_i^2}\times\sqrt{\sum_{i=1}^{n}\mathrm{B}_i^2}} \qquad (2)$$

这里的分别代表向量 A 和 B 的各分量。例如计算两个句子向量：

句子 A(1,1,2,1,1,1,0,0,0)和句子 B(1，1,1,0,1,1,1,1,1)的向量余弦值来确定两个句子的相似度

$$\cos\theta = \frac{\mathbf{1\times1+1\times1+2\times1+1\times0+1\times1+1\times1+0\times1+0\times1+0\times1}}{\sqrt{\mathbf{9}}\times\sqrt{\mathbf{8}}} = 0.81$$

1.3.4 留言与回复的一级分类

1.3.4.1 分类器构建与使用

分类器的构建使用的模型是问题 1 中得到的模型。通过调用保存的模型，对预处理后的数据进行分词，然后进行类别匹配。通过多次分类，得到可信度最高的分类结果，排除单次分类出现的错误，以及排除单次训练可能出现的意外，即出现不在模型已有分类中的数据类型的误判。综合多次分类得到较好的分类结果，具有较好的可信度。

由于分类模型首先会对语句进行关键词提取，即得到一组可以表述语句主要意思的词

组，进而得到该语句的分类。而一句留言或答复的关键词可以认为是一个类型标签所对应的关键词的一个子集。当两个子集的重合度较高时，即表现出同一个分类。所以将留言分类与答复分类的结果进行比较，如果两者类被相同，则可认为答复与留言具有相关性。而关键词重复越多，则认为留言的回复越能对留言进行解释与说明。

将两者分别分类，得到分类结果，分别存储于 MessageType.txt, ReplyType.txt。再将两者写回清洗后的数据文件。然后通过脚本进行比较，最终在数据文件中新建一列 is_equal, 值为 0 则表示两者不同，为 1 则表示相同。

1.3.4.2 时间与分类的处理

分别读取附件 4 中的留言时间和回复时间。先前已经对该类数据进行清洗，统一为 datetime 格式。然后遍历两者做差，即可得到留言与回复消息两者的时间差。

由于对留言的处理时间会因类别对于的部门所需时长因素的影响，以及不同留言人员对于留言回复的期望值不同，回复时间对于评分的影响需要综合两者的需求。通过阅读相关文件，咨询一级分类下相关部门的工作人员，以及了解留言群众对于得到回复的平均期望时间，初步建立留言处理时间的分级表。由于不同分类对应的部门对于留言内容所需要的处理时间不同，以及留言内容所要求的工作量上有差别，下表仅作为本模型的评价标准之一，若与实际情况有较大误差，可进行修正。

满意度 类别 时间/天	较快，满意	一般	较差、不满意
城乡建设	0-15	15-30	30+
环境保护	0-7	7-15	15+
交通运输	0-3	3-15	15+
教育文体	0-3	3-7	7+
劳动和社会保障	0-3	3-7	7+
商贸旅游	0-5	5-15	15+
卫生计卫	0-5	5-15	15+

表格 1-3-1，其中 A-B 表示大于 A 天小于 B 天，C+表示大于 C 天。该表反映该分类对于留言内容进行处理以及完成留言需求或可以答复留言所需的平均时长与留言人员诉求期望得到回复的平均时间长综合而得。

1.3.4.3 评价结果

综合以上多中因素，针对留言回复所需的相关性，可解释性以及完整性，得出综合评价模型。评价先基于分类，一级分类相同则认为两者具有相关性，否则认为回复与留言不相关。同时还依据文本相似度，最终两者共同评价是否相关。其次，取留言相似度的中位数，大于该值即认为可解释性与完整性好，得分越高则最终评价得分也越高。最后综合时间得分，按照表 1-3-1 的标准进行得分。最后综合三个评价标准分别得分以及该标准的权重占比，归一化后得到评分。依据分数高低得到不同的评价结果。

2. 结果分析

2.1 问题 1 结果与分析

2.1.1 模型训练结果

通过使用题目同一批数据对多个模型的训练与修正，对训练优化参数的微调以及对于训练次数等参数的调整，得到多个模型的训练结果。并且进行结果比较，得出最优模型。

模型通过 F-Score 进行评判，本题使用的模型为 ERNIE，其他模型为测试模型，用于进行分类效果的比对。

表 2-1-1：模型训练结果表

评价指标 / 模型	Accuracy	F1-Score
ERNIE, Chinese	91.45%	95.95%
ERNIE-Tiny, Chinese	89.85%	95.65%
Bert-Base, Chinese	90.92%	93.85%
Bert-wwm, Chinese	89.10%	92.49%
RoBERTa-wwm-ext, Chinese	90.81%	95.35%

由表格数据可见，ERNIE 模型在准确率与 F1-Score 上表现均优于其他模型。在同样的数据进行训练与验证得到的结果，对于同样的验证集，ERNIE 的分类准确度接近 96%，而其他模型虽然也有接近 96%的 F1-Score，在准确率上仍不如 ERNIE 模型。

2.1.2 模型的修改与优化

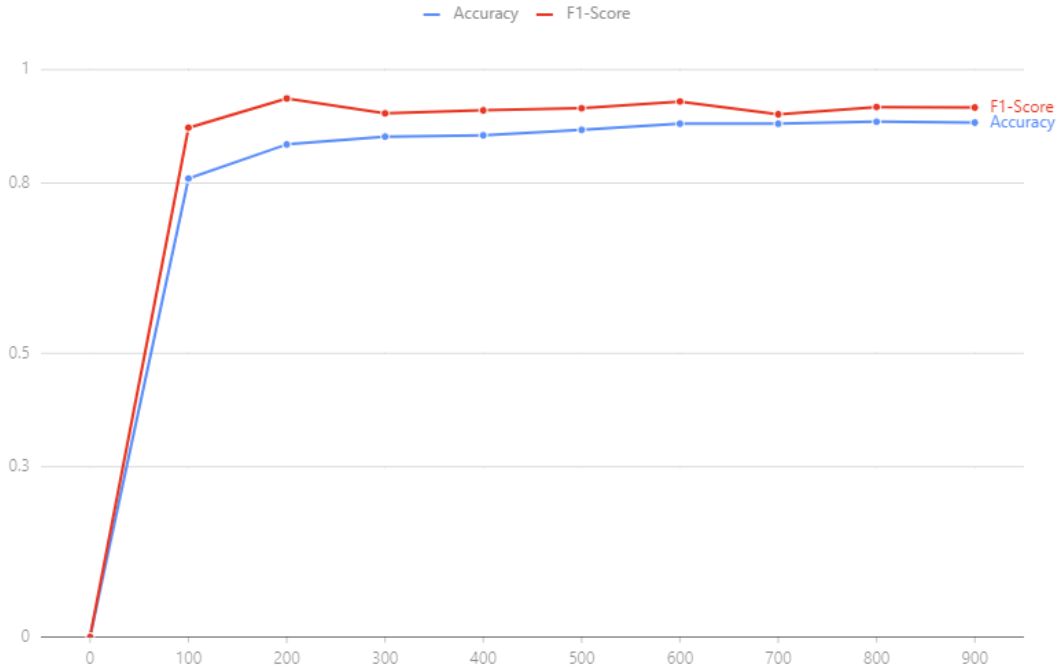
1. 学习率的修改。模型初始学习率为 4e-5,如此得到的模型效果 f1 为 92.3%，后续通过修改为 5e-5 得到更优的模型。训练得到当前最优模型。

2. Fine-Tune 优化策略和衰减策略的选择。模型支持 linear_decay 和 noam_decay 两种衰减策略可选，即线性衰减与多项式衰减，选用的是线性衰减曲线。优化策略有 AdamWeightDecayStrategy、ULMFiTStrategy、DefaultFinetuneStrategy 等。选用的是 AdamWeightDecayStrategy 优化策略。

3. 原模型用于二分类，而本题数据为多分类。所以在训练前，将模型改为多分类模型，对数据输入以及种类映射、数据输出都需要做修改。

4. 每隔 100step 即在验证集上评估一次模型性能，模型性能变化如下图。

图表 2-1-2 模型性能变化曲线



2.2 问题 2 的结果与分析

2.2.1 bert-NER 模型训练结果

运行模型，在验证集上的 F-1 值在训练了 15 个 epoch 时就已经达到了 87.49%，并在测试集上达到了 89.35%，这是多次调整模型参数得到的较好结果了。

由于使用人工标注的方式非常耗费时间，我们最终标注的数据只有 2000 多条，最终模型的预测效果并不是非常理想，对地址、人群信息的抽取不能够满足我们的需求。

如果能够对附件 2 中 9210 条留言详情数据完成标注，模型的准确率是会达到一个更加理想的模型效果。当模型的标注数量足够大时，新输入数据可以不需要进行标注，再进行测试时，也能得到较好的效果。

2.2.2 k-means 算法聚类结果

对 4320 条数据原始数据完成聚类，我们设置聚类数目为 500，得到聚类效果如下（详细见附件 热点问题留言明细表.xls）：

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
0	250461	A00012981	咨询 A 市计算机技术与	2019/7/2 20:33:50	我是 A 市事业单位职工，目前在医院	2	0

			软件专业技术资格与职称对应与评定问题		信息科，从事计算机专业技术工作…..		
0	242358	A00037760	在A7县注册义工团体需要准备哪些材料？	2019/3/31 15:10:51	需要在A7县注册一个服务组肿瘤患者的专业义工团体，不知道与哪里联系和需要准备哪些材料。	0	0
…	…	…	…	…	…	…	…

表 3-2-1：热点问题留言明细表

由热点问题留言明细表（见附件-热点问题留言明细表.xlsx），我们可以看出，使用留言详情做文本类聚，每个类簇代表同一类事件，在某些效果明显的类簇中，它所展示的就是同一事件，对热点事件的挖掘有很大的意义。

对上述热点问题留言明细表进行计算热度指数，得到热度排名，并使用类簇中点赞数最多的一项作为该类簇的簇中心（获取问题描述），热度排名表如下（详细见附件-热点问题表.xls）：

热度排名	问题ID	热度指数	时间	问题描述	地址/人群
1	17	5.3	2019/2/21 18:45:14	请书记关注A市A4区58车贷案	A市A4区
2	29	4.9	2019/8/19 11:34:04	A市A5区汇金路五矿万境K9县存在一系列问题	A市A5区
3	4	4.0	2019/4/11 21:02:44	反映A市金毛湾配套入学的	A市家长

				问题	
4	19	3.4	2019/7/8 9:50:45	建议将渝长厦 A9 市站选择在 A9 市西南角设站	A9 市市民
5	126	2.8	2019/7/8 9:50:45	A7 县星沙凉塘路东七线到东十线段今年会启动建设吗	A7 县

表 4-2-2　热点问题排名前五

2.2.3 评价结果

由于用于训练 Bert NER 的数据量较少，因此对地址人群的抽取效果不够理想，通过加大训练数据样本，能够提升模型的准确度。

使用 k-means 模型完成文本的聚类，在数据预处理过程中，将清洗过的数据使用 jieba 进行分词、去停用词。其中分词的时候，载入预先设置好的自定义词典(见附件 location.txt)，提高分词精度。

在热度指数的计算上，各类簇内数据条数、各类簇中数据的总点赞数、各类簇中数据的总反对数对热度指数的影响程度是不同的，因此需要在对其归一化后，按一定比例选取并求和计算热度指数，通过多次尝试，我们发现比例为 2:5:1 的时候，热度指数的代表意义比较突出。我们总结出原因如下：

3. 聚类的效果受较多因素的影响，比如 jieba 的分词效果、簇中心数目大小、聚类算法的选取等，聚类最终结果未必能够很好的将同一事件划分到同一类簇中。
4. 点赞数与反对数是比较能够反映人们对事件的关注程度的，对于热度指数的影响相对较大。

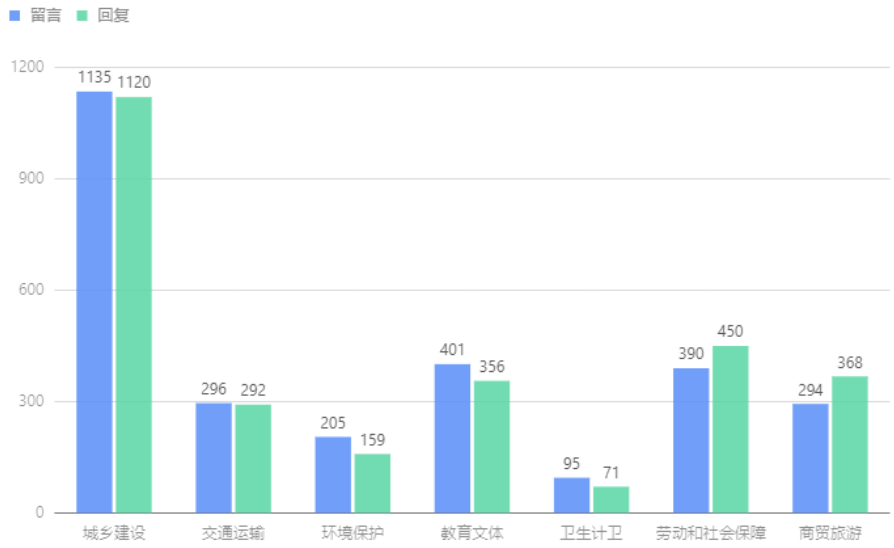
综上，我们认为，如果能够提升地址/人群信息的抽取效果，使用地址/人群信息替代留言详情完成文本聚类，或许能够将发生在同一地点、同一时间段的热点事件挖掘出来，另外，调整点赞数、反对数、聚类数在热度指数中的影响程度，能够获得更好的热度指标。

2.3 问题 3 的结果与分析

2.3.1 数据分类结果

通过分类器对附件四留言以及回复的分类，得到各个分类情况如下表。

图表 1-3-1 留言及回复分类结果



由图表 2-3-1 可以看出，在环境保护、教育文体和商贸旅游三大类三，都有大约 50 的数量差。即表明，在这些分类下，答复的内容与留言内容相差程度较多，答复对于留言需求的回答不完整，不能覆盖到相应的分类下，留言与答复的相关性不高。说明这三个分类的留言答复质量还有待提高。

2.3.2 相似度计算结果

由下列部分相似性表的数据中可以看出，当答复与留言关键词的匹配度低时，得到的相关度较低，即回复不能对留言需求做出解释，答复不完整。而相关度越高，则表示回复对留言的解答越详细，能够满足留言的需求，可以完整回答好留言的问题。

用 户 ID	留言主题	留言详情	留言回复	相似度
U U0081 320	咨询打狂犬役苗报销比例是多少	请问领导，农合费用增加了，打狂犬役苗报销比例是多少。盼回音。先谢了！	已收悉	0
U U008 971	关于 625 新政导致 30 多万首付被套问题的再反映	尊敬的网友：您好！住建委网签系统没有限制未缴纳首付的不得网签，只要购房资格审核通过即可进网签。但是为了落实调控政策，对房价 8 千到一万以上的项目，实行了限网签制度，每个项目每天限签 1-2 套。另外，关于您提到的公积金贷款问题，建议您找公积金中心咨询！感谢您对我中心工作的关注与理解！市住建委（A 市房屋产权登记管理中心）我的上次投诉所说的是网审，不需要交清首付，不是网签。现在个人能接受的处理措施有: 1.因为公积金贷款无希望，个人无经济能力购买，请求政府	“UU008971”您好！关于退房退款问题，购房事件为您与开发之间的购买行为，请您按照您与开发商之间的合同约定，进行协商处理。特此答复！如还有疑问可拨打 0000-00000000 咨询。 2018-07-20	0.125

		督促开发商退房退款谢谢！		
A00031618	请加快提高A市民营幼儿园老师的待遇	地处省会A市民营幼儿园众多，小孩是祖国的未来，但民营幼儿园教师一直都是超负荷工作且收入又是所有行业最低，甚至连养老和医疗金都没交，在国家大力倡导普惠型幼儿园的同时更是加大了教师的工作压力，在降低成本的同时还增加了学生数量，让本来就喘不过气的教师更是雪上加霜，希望市委市政府加快提高民办幼儿园教师工资待遇水平和降低工作压力有何具体政策和行动？	市民同志：你好！您反映的“请加快提高民营幼儿园教师的待遇”的来信已收悉。现回复如下：为了改善和提高民办幼儿园教师待遇，根据2019年1月8日出台的《中共A市委A市人民政府关于学前教育深化改革规范发展的实施意见》长发〔2019〕2号文件精神，对于学前教育教师的培养和待遇问题做出了明确要求。一是在提高教师待遇方面，依法保障民办幼儿园教职工待遇，民办幼儿园聘任教职工要依法签订劳动合同，依法缴纳城镇企业职工养老保险、医疗保险、生育保险、工伤保险、失业保险和住房公积金，民办园要参照当地公办园教师工资收入水平，合理确定相应教师的工资收入。二是加强监管协同推进，加强对民办幼儿园的日常监管和质量管理，保障民办幼儿园教师待遇，在完善人事（劳动）、工资待遇、社会保障和职称评聘等方面继续推进。谢谢您对我市学前教育的关注和支持！	0.46
UU008	请求落实K4县学前教育巡	"	""UU008802”您好！您在	0.635

802	回支教志愿者应有的待遇	我们是2012年K4县学前教育巡回支教志愿者，在乡村服务期为2年，在支教期间我们严格遵守单位的规章制度，两年均考核合格，为K4县的教育事业贡献了自己的力量。根据中央、省、市县的规定，支教人员应当享有相应的政策优惠。到目前为止，我们支教期满已经四年多了，在这期间，K4县每年都招录大量的教师及学前教育教师，但都没有按照国家规定，给予我们支教人员相应的政策优惠。作为K4县首批支教志愿者，我们经过了层层筛选，通过了政府组织的考试，在支教期间我们积累了丰富的教学经验，支教结束后我们仍然服务于学前教育一线。目前K4县学前教育师资力量缺口大。同时，国家关于“三支一扶”人员的政策规定对于服务期满两年考核合格、符合用人单位岗位要求并愿意继续在农村基层工作的，在乡事业单位编制内新增工作人员时，可免于参加统一招考，由接收单位报县人事部门办理事业单位工作人员聘用核准相关手续。我们拿着对“三支一扶”的优惠政策到K4县教育局去反映，不但未给我们明确回复，却给了我们一个失望的回复，说我们不符合条件，不符合政策，这一回复，明显与国家发布的政策不一致，让我们无偿参加支教的志愿者心灰意冷；因此，希望能够组织相关部门核实相关支教政策，督促K4县教育局落实相关政策，优先考虑安排我们这一批巡回支教志愿者在编工作。报告人:全体支教老师2019年7月25日	《问政西地省我们查阅了当时的录用资料以及国家、省、市关于巡回支教方面的相关政策，就您反映的问题回复如下： 1.2012年K4县被省教育厅选定为学前教育巡回支教试点县，根据省教育厅给K4县的计划，制定了《K4县2012年开展学前教育巡回支教工作志愿者招聘工作方案》，根据方案，2012年K4县招募了巡回支教志愿者60人，并与他们签订了《K4县2012年学前教育巡回支教志愿者服务合同》（以下简称《服务合同》）。在志愿者的招聘、安排、管理、考核等方面严格执行省教育厅的文件要求，支教工作取得了圆满成功。 2.《服务合同》第三条规定了“志愿者在服务期间考核合格的，可享受省、市人数部门招聘教师所规定的优惠政策”，同时也在第五条注明了“合同期两年，合同期满后，双方聘用关系自行解除，面向社会招聘的巡回支教志愿者自主择业”。支援者作为社会人员，按照教师招聘工作方案，巡回支教志愿者年龄不足30周岁的，有幼儿教师资格	

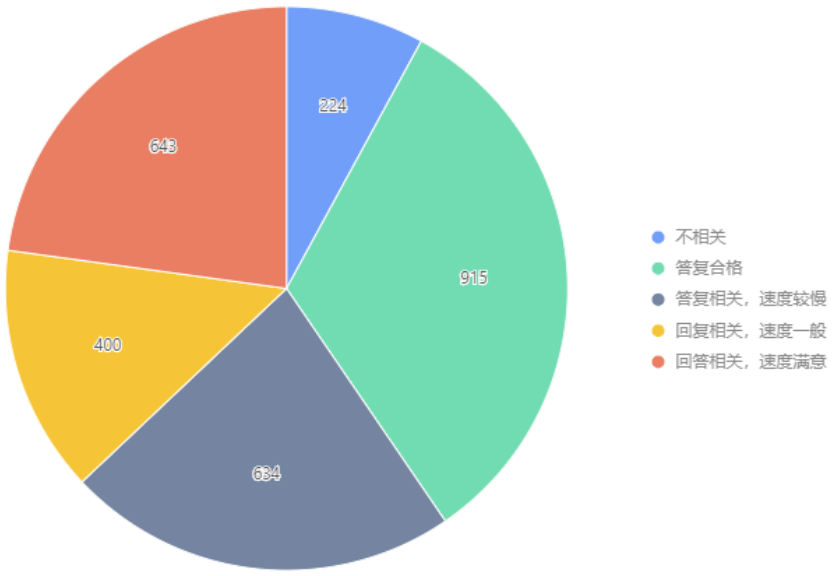
			证的可以享有我县教师招聘的报考资格。目前我县 2019 年的教师招聘工作已经结束，支教老师们所要求的放宽招聘条件，今年无法承诺兑现。明年的教师招聘中，只要符合《西地省事业单位公开招聘人员办法》规定和国家、省、市其他有关规定的，经请示县委、县政府同意后，与县人社部门协商，可以适当考虑支教老师的诉求。2019 年 7 月 29 日"	

使用 EXCEL 进行统计，得到平均数为 0.2622，中位数为 0.25。这里取中位数为评价标准，即相似度大于 0.25 的答复可以认为和留言有较高相关性。由于不相关的数据相似度为 0，数量有 107 个，对于平均数影响较大，使得平均数的可信度下降。所以取中位数较合适。

2.3.3 评价结果

由模型得出各个留言的评价，建立下表 2-3-2，由表格可以看出，由于相关性较低，导致答复与留言分类不一致，以及文本相似性低的原因，在总共 2817 条数据中，约 1/3 的数量被评价为答复合格，即相似度低于中位数。

图表 2-2-2 评价结果分布



在相似度以及分类表现都良好的答复上，基于给出答复所需时间进行进一步分级评价，评价为一般以上的有 1034 条，占总数据量的 36.7%，可以认为约 40%的留言都得到了较快的处理，而且处理的效果即内容相关性，可解释性也有良好的效果。

答复与留言最终评价为不相关的内容有 107 个，少于分类结果不相同的 268 个。即认为同时满足留言与答复分类不同和留言相关性为 0 两个条件的，才认为两者确实不相关。

综上所述，经由模型得到的评价结果，具有良好的基于数据的判别能力和可信度。

3. 结论

对政务文本进行文本挖掘以及信息提取，减少人工的工作量，能够极大地提高政府政务的处理效率，提升政府的的管理水平和人民对政府服务的满意度。政务的文本挖掘是当今自然语言处理的重要应用领域之一，当然，文本挖掘的也存在其发展的瓶颈，本文使用 ERNIE 作为留言分类的模型，在分类的效果上是能够满足我们的需求的；而使用 k-means 算法对热点问题进行归类，并使用 bert-NER 的模型抽取问题中的相关信息，对于挖掘热点问题有很不错的效果；最后使用文本相似度等指标作为答复意见的质量指标，能够合理地评价答复意见是否能够对留言做出合理的解答。

通过文本的分类，能使政府快速、高效、专注解决某一类的政务，减少服务等待时间，提高效率；对热点问题的挖掘与分析，可以让政府及时关注群众需求、群众问题，了解群众生活现状和关注的中心等问题；对于群众留言的答复意见的质量评判，可应用于政府部门的自我监督，提升对人民群众的服务水平有促进作用。

4. 参考文献

[1] Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, Hua Wu. ERNIE: Enhanced Representation through Knowledge Integration [R]: 2019

[2] Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Hao Tian, Hua Wu, Haifeng Wang. ERNIE 2.0: A Continual Pre-training Framework for Language Understanding [R], 2019:11-21

[3] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Ziqing Yang, Shijin Wang, Guoping Hu. Pre-Training with Whole Word Masking for Chinese BERT [R]: 2019

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT. Pre-training of Deep Bidirectional Transformers for Language Understanding [R]: 2019

[5] 翟东海，鱼江，高飞，于磊等。最大距离法选取初始簇中心的 K_means 文本聚类算法的研究 [R].西南交通大学.2014

[6]王千，王成，冯振元，叶金凤.K-means 聚类算法研究综述.2012

[7] [美]Steven Bird Ewan Klein Edward Loper. Natural Language Processing With Python. 人民邮电出版社, 2014

[8] 涂铭，刘祥，刘树春. Python 自然语言处理实战. 机械工业出版社.2018 :53, 59-61, 85-95, 166-168