



广东外语外贸大学
GUANGDONG UNIVERSITY OF FOREIGN STUDIES

网络舆情分析（课程设计）

题 目： 知乎-微博平台下有关接种新冠疫苗的舆情分析以及
 引导建议

成 员： 吴棉滨 20181002919

 罗佳 20170401382

 游畅 20181003005

学 院： 信息科学与技术学院

指导老师： 王连喜

完成时间： 2021. 06. 10

目录

摘要.....	3
Abstract.....	3
1 项目目标.....	4
1.1 选题背景.....	4
1.2 分析目的.....	4
2 国内外相关工作.....	5
3 核心思想和算法描述.....	5
4 系统主要模块流程.....	7
4.1 爬取数据.....	7
4.2 数据观察.....	9
4.3 数据预处理.....	10
4.4 文本分词.....	13
4.5 文本特征提取.....	14
4.6 模型建立与训练.....	14
4.7 提取关键词与生成词云.....	15
5 实验结果及分析.....	16
5.1 平台分布.....	17
5.2 用户情绪分析.....	17
5.3 热词图.....	20
5.4 话题分布.....	24
6 总结与体会.....	27
6.1 各平台舆情.....	27
6.2 总体舆情.....	28
6.3 建议.....	28
6.4 总结思考.....	28
6.5 小组体会.....	29
7 参考文献.....	29

摘要

新冠病毒的出现与广泛传播给国民与世界其他各地人民的生活都带来了翻天覆地的影响，而直至今日，我国也已经推出了针对新冠的疫苗，并开始在全国各地大规模推广接种。本文以新浪微博与知乎为文本爬取平台，结合文本摘要、关键词提取和深度学习等技术，针对“接种新冠疫苗”这一话题的进行舆情分析报告，而研究的最终目的在于改善宣传方式，提升国内新冠疫苗接种率。

关键词：新浪微博、知乎、新冠肺炎、新冠肺炎疫苗、网络舆情分析、文本摘要、关键词提取、文本情感分析、深度学习。

Abstract

COVID-19 turn the world upside down and spread to the people of all parts of the world. It has brought about a huge impact on the lives of the people of other parts of the world. Up to now, China has launched a vaccine against the new crown and has begun to popularize vaccination in all parts of the country. In this paper, Sina Weibo and Zhihu are used as text crawling platforms, combined with text summary, keyword extraction and deep learning technology, to analyze the public opinion on the topic of "vaccinating the new crown vaccine". The ultimate goal of the research is to improve the propaganda methods and enhance the domestic vaccination rate of the new crown vaccine.

Key words: Sina Weibo, Zhihu, COVID-19, COVID-19 vaccine, Internet public opinion analysis, Text summary, Keyword extraction, Text sentiment analysis, Deep learning.

1 项目目标

1.1 选题背景

自 2020 年年初新冠疫情在全球范围爆发以来，人们的生产生活受到了巨大的影响。为了赢得防控阻击战，保护人民群众的生命财产安全，尽最大可能降低损失，尽最早回到原本的运转轨道，人们的生活方式也发生了许多的改变。在经历了停产停工、上网课、线上毕业答辩、线上面试等一系列防控带来的“小麻烦”之后，疫情渐渐得到控制，我们的生活似乎也逐渐回到原来的轨道了。

随着时间向前推进，世界各国都有条不紊地推进疫苗研发进度。在 2021 年 1 月，也就是距新冠疫情爆发一周年的时候，我国宣布新冠病毒疫苗将为全民免费接种，费用由医保基金和财政共同负担。尽管为全民免费接种加上巨大的人口基数意味着我国政府将承担极高的成本，然而我国政府并没有因为承担的成本高、压力大，就把负担转嫁给人民。我国政府的行动不但诠释了什么是“一切为了人民”，什么是“人民至上”，什么是真正的“执政为民”，也从正面保证了我国研发的新冠疫苗良好的安全性。

时间来到 21 年的 5 月份，经过钟南山院士、张文宏医生等国内防疫专家们的研究，提高国内接种率以形成群体免疫，对日后恢复生产生活，甚至在最坏情况下抵御国外疫情输入而言是至关重要的。因而在国内疫情逐渐平稳之后，各地都在有序推进疫苗接种工作。时下广州大学城各所高校也陆续进行着为全体师生接种新冠疫苗的准备。

1.2 分析目的

令人奇怪的是，度过了最紧张严峻的疫情期之后，人们对接种疫苗的热度却并不似意料中的高涨。在 2021 年的 1 月份，国内接种总剂量尚不到一千多万剂，大家对于接种疫苗有各种各样的顾虑，甚至认为国产疫苗不安全等，尤其是在部分知识分子群体中的抵触较大。网络上有人不停发问“接种疫苗会有哪些副作用”、“环境安全可以不接种疫苗吗”、“不接种疫苗会不会有哪些副作用”等问题。

随着“Web 2.0”时代的到来，新媒体资讯平台的崛起使得互动交流型的平台成为人们获取实时信息的窗口。然而，在这样的交互模式中，用户很容易受到某些片面观点输出的冲击，产生不良的心理反应。有研究表明，当个体处于这种感到“有威胁”的环境中，容易自动引发应激反应，例如产生紧张，焦虑的情绪。而负面性的情绪状态又将影响个体的生理过程和行为模式。因此，这样的环境带来的影响是我国疫苗接种比例较低，较难形成有效稳定的群体免疫效应。

针对这个问题，传统的心理学问卷法、实验法等手段在当下防疫背景中受限严重，难以提供实效性、真实性兼备的心理指标数据。互动交流型的社交平台涵盖较为真实丰富的网络用户行为数据：除了获取最新资讯与热点信息，用户可以在微博上发表动态和评论。评论文本能在一定程度上展现用户撰写评论时的情绪状态和主观的认知态度，合理运用用户的评论数据，通过数据挖掘，文本分析，深度学习等方法，能够获取用户的心理指标，进而展开有效及时的研究。

本次研究选取的是国内最大的两个互动交流平台：新浪微博与知乎，旨在通过综合基于 Python 的网络爬虫、数据挖掘以及文本情感分析技术，对近期网民发表的相关接种疫苗的评论进行分析，得出可视化数据，以便学校、社区等基层

宣传口更好地了解舆情，对症下药，优化接种疫苗宣传方式。并且加深对民众心理倾向的了解，发挥宣传主导作用，通过网络推送解答疑惑，正确疏导大众情绪，推动社会舆论氛围。

2 国内外相关工作

2.1 国内外舆情研究发展比较

查阅文献资料，发现国外在舆情分析领域中的研究工作开展得较早，最初的研究对象相对集中在社交媒体平台信息搜索机制上，之后经历了几年的发展，逐渐从社交媒体信息搜索机制转向了舆情情感分析、隐私访问控制、隐私披露意愿影响因素、多类情感词舆情分析上。国内相关领域的研究开展较晚，最初主要集中在微博热点话题的主题词建模和分析上，但是随着微博、微信等社交媒体软件在国内的大量使用，学者们的研究开始转向了大数据驱动的网络舆情治理、网络舆情情感倾向分析与研判、网民的社交媒体倦怠行为、舆情反转分析与预测上[1]。

2.2 国内新冠疫苗舆情研究

在知网上检索，看到的研究新冠疫情的文章大多是和危机治理及疫情状态下社会变化相关的。少数与新冠疫苗相关的文章，在内容上是更侧重于疫苗本身而非民众对疫苗的想法和情绪。使用文本分析、深度学习等技术手段去分析新冠疫苗舆情的论文则更少。

3 核心思想和算法描述

3.1 核心思想

本次项目的关键地方在于利用深度学习模型——Bi-LSTM[2]，对爬取到的微博和知乎文本内容进行情感分类，为最后的分析提供信息基础。在此过程中，还会利用到 TextCNN[3]与 Bi-LSTM 进行比较，进而论证我们模型分类的结果更有可靠性。有关 Bi-LSTM 与 TextCNN 算法的详细描述，见 3.2 小节。

3.2 算法描述

3.2.1 TextCNN

3.2.1.1 TextCNN 算法概述或框架图

TextCNN 模型是由 Yoon Kim 提出的 Convolutional Neural Networks for Sentence Classification[3]一文中提出的使用卷积神经网络来处理 NLP 问题的模型。TextCnn 在文本分类问题上有着更加卓越的表现。从直观上理解，TextCNN 通过一维卷积来获取句子中 n-gram 的特征表示。TextCNN 对文本浅层特征的抽取能力很强，在短文本领域如搜索、对话领域专注于意图分类时效果很好，应用广泛，且速度快，一般是首选；对长文本领域，TextCNN 主要靠 filter 窗口抽取特征，在长距离建模方面能力受限，且对语序不敏感。

3.2.1.2 TextCNN 算法流程图

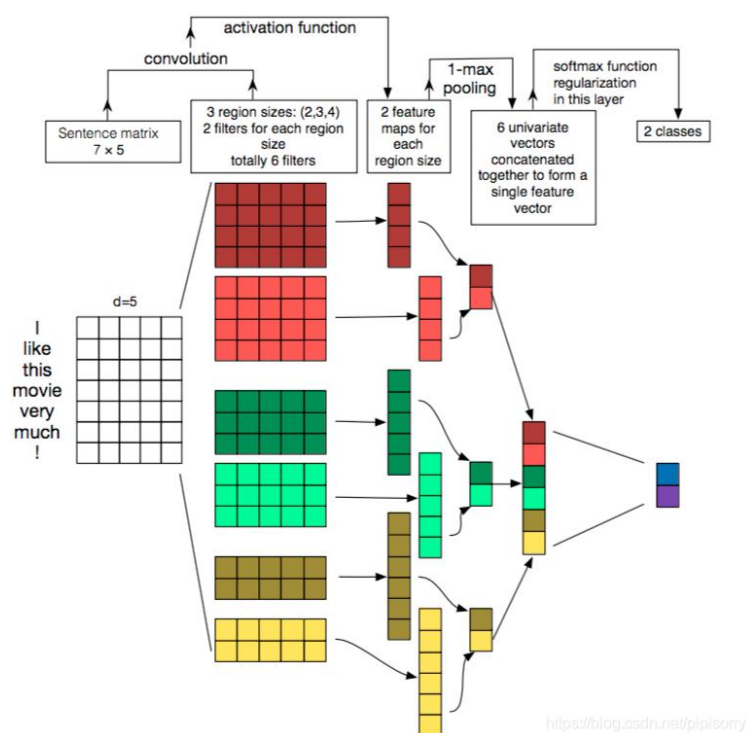


图 1 TextCNN 模型图

3.2.1.3 算法细节

第一步：Embedding：第一层是图中最左边的 7 乘 5 的句子矩阵，每行是词向量。

第二步：Convolution：然后经过 $\text{kernel_sizes}=(2,3,4)$ 的一维卷积层，每个 kernel_size 有两个输出 channel。

第三步：MaxPolling：第三层是一个 1-max pooling 层，这样不同长度句子经过 pooling 层之后都能变成定长的表示。

第四步：FullConnection and Softmax：最后接一层全连接的 softmax 层，输出每个类别的概率

3.2.2 Bi-LSTM

3.2.2.1 Bi-LSTM 算法概述

LSTM 的全称是 Long Short-Term Memory，它是 RNN（Recurrent Neural Network）的一种。LSTM 由于其设计的特点，非常适合用于对时序数据的建模，如文本数据。Bi-LSTM 是 Bi-directional Long Short-Term Memory 的缩写，是由前向 LSTM 与后向 LSTM 组合而成。两者在自然语言处理任务中都常被用来建模上下文信息。

3.2.2.2 Bi-LSTM 算法流程图

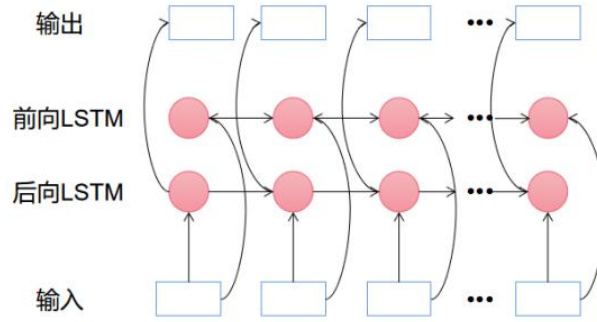


图 2 Bi-LSTM 模型图

3.2.2.3 算法细节

可以学习长期依赖信息，它通过特殊的门结构来去除或增加信息到细胞状态的能力，门是一种让信息选择式通过的方法，LSTM 有三个门。

第一层是忘记门，决定我们会从细胞状态中丢弃什么信息。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

下一步是确定什么样的新信息会被存放在细胞状态。

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

下一步是更新细胞状态

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

最后输出信息。

现在我们采用的是图 2 所示的 Bi-LSTM。该模型由前向的 LSTM 和后向的 LSTM 组合而成，能够很好的处理双向数据的序列信息。一个 LSTM 层式从左往右处理，而另一个是从右往左处理。

4 系统主要模块流程

4.1 爬取数据

4.1.1 平台分布

新浪微博是基于用户关系的社交媒体平台，用户可以简单地通过手机、PC 等多种移动端，以文字、视频、图片等方式公开实时发表信息。作为继搜索、门户之后的新一代互联网入口，微博改变了信息的传播方式，实现了信息的即时分享。根据新浪微博 2020 年 Q1 财报，其月活跃用户已达 5.11 亿人，是国内首屈一指的大型社交平台。

由于新浪微博有着用户众多、时效性强的特点，成为了本次舆情分析的对象平台之一。

知乎则是一种社区式的问答平台，和微博一样拥有用户众多，时效性也极强的特点。同微博不同之处在于，知乎的“信息”单元并非单一的一条回答或者一

个问题，而是一个问题与其下所有回答的集合：不同角度的回答以及各种观点的比例可以一窥人们对相关问题的具体态度分布情况。

因此，选取具体的分析对象时有必要考虑整个问题的关注人数等。

4.1.2 获取文本内容

4.1.2.1 微博

在微博我们选择的获取对象是某个 tag 下的评论。选择的 4 个 tag 分别为：

①#国内灭活疫苗对多数变异株有效#；②#疫苗#；③#一起苗苗苗苗苗#；④#新冠疫苗知识点#。

截至获取文本信息之前，4 个 tag 的讨论热度均在 2000 以上。



图 3 微博爬取的四个话题

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
rd	用户名称	微博等级	微博内容	微博转发量	微博评论量	微博点赞	发布时间	话题名称																					
0	西咸新区泾河新城	蓝v	#常态化疫情防控#【转发周知！#新冠疫苗知识点#@西咸新区泾河新城发布：【#新冠疫苗知识点#】一针、两针、三针，转发	1	1	1	3小时前	#新冠疫苗知识点#																					
1	河南师范大学	蓝v	[cp]#微分享#【#新冠疫苗知识点#】一针、两针、三针，转发	1	1	1	3小时前	#新冠疫苗知识点#																					
2	健康四川	蓝v	#新冠疫苗知识点# #新冠疫苗接种问答#【备孕者、转发	1	1	1	4小时前	#新冠疫苗知识点#																					
3	金陵四小	蓝v	转发周知！#新冠疫苗知识点#@鼓楼微讯【转发周知！转发	1	1	1	4小时前	#新冠疫苗知识点#																					
4	黄河科技学院	蓝v	#关注新冠肺炎#【转发周知！#新冠疫苗知识点#】一针、转发	1	1	1	4小时前	#新冠疫苗知识点#																					
5	河南师范大学	普通用户	#师大分享#【转发周知！#新冠疫苗知识点#】一针、转发	1	1	1	4小时前	#新冠疫苗知识点#																					
6	杭州市居民	蓝v	转发周知！#新冠疫苗知识点#@杭州城投发布：转发周知！转发	1	1	1	4小时前	#新冠疫苗知识点#																					
7	西安市人民	蓝v	【转发周知！#新冠疫苗知识点#】@黑龙江-平安黑龙江转发	1	1	1	5小时前	#新冠疫苗知识点#																					
8	闵行四小	蓝v	【转发周知！#新冠疫苗知识点#】@今日闵行：【转发周知！转发	1	1	1	5小时前	#新冠疫苗知识点#																					
9	宁波江北岸	蓝v	【转发周知！#新冠疫苗知识点#】截至5月23日，#我国转发	1	1	1	5小时前	#新冠疫苗知识点#																					
10	大庆检察	蓝v	【转发周知！#新冠疫苗知识点#】@黑龙江-平安黑龙江转发	1	1	1	5小时前	#新冠疫苗知识点#																					
11	青白江住建	蓝v	【转发周知！#新冠疫苗知识点#】一针、两针、三针，转发	1	1	1	6小时前	#新冠疫苗知识点#																					
12	安阳市居民	蓝v	#安阳微推# #新冠疫苗知识点# #我国新冠病毒疫苗接种超1	2	1	1	6小时前	#新冠疫苗知识点#																					
13	安阳热电	蓝v	转发周知！#新冠疫苗知识点#@安阳城投发布：转发周知！转发	1	1	1	7小时前	#新冠疫苗知识点#																					
14	上城商会	蓝v	【转发周知！#新冠疫苗知识点#】@上城发布：【转发周知！转发	1	1	1	10小时前	#新冠疫苗知识点#																					
15	巴中民政协	蓝v	【转发周知！#新冠疫苗知识点#】一针、两针、三针，转发	1	1	1	10小时前	#新冠疫苗知识点#																					
16	唐东法院	蓝v	【#新冠疫苗知识点#】全国新冠病毒疫苗接种迎来高峰，【转发	1	1	1	11小时前	#新冠疫苗知识点#																					
17	景德镇李	蓝v	#小暑趣分享#【#新冠疫苗知识点#】#我国新冠病毒接种【转发	1	1	1	11小时前	#新冠疫苗知识点#																					
18	汉中中院	蓝v	【#新冠疫苗知识点#】全国新冠病毒疫苗接种迎来高峰，【转发	1	1	1	12小时前	#新冠疫苗知识点#																					
19	绵阳政宣	蓝v	【转发周知！#新冠疫苗知识点#】@民政微语：【转发周知！转发	1	1	1	12小时前	#新冠疫苗知识点#																					
20	西安中院	蓝v	【#新冠疫苗知识点#】全国新冠病毒疫苗接种迎来高峰，【转发	1	1	1	13小时前	#新冠疫苗知识点#																					
21	济南局党	蓝v	#党车微视界#【#新冠疫苗知识点#】全国新冠病毒接种【转发	1	1	1	昨天 21:04	#新冠疫苗知识点#																					
22	南铁九江	蓝v	#九车微视界#【#新冠疫苗知识点#】全国新冠病毒接种【转发	1	1	1	5-26 20:11	#新冠疫苗知识点#																					
23	郑东郑道	普通用户	#郑东郑道#【转发周知！#新冠疫苗知识点#】截至5月，转发	1	1	1	5-26 19:18	#新冠疫苗知识点#																					
24	模县财政	蓝v	【转发周知！#新冠疫苗知识点#】@模县发布：【转发周知！转发	1	1	1	5-26 18:19	#新冠疫苗知识点#																					
25	西铁资讯	蓝v	#西铁微分享#【#新冠疫苗知识点#】全国新冠病毒接种【转发	1	1	1	5-26 18:12	#新冠疫苗知识点#																					
26	大兴区医	蓝v	【转发周知！#新冠疫苗知识点#】@健康大兴：【转发周知！转发	1	1	1	5-26 18:11	#新冠疫苗知识点#																					

图 4 微博爬取的内容

4.1.2.2 知乎

与微博相似地，选择了四个问题，并对相应的回答进行爬取。四个问题分别是：

①在校大学生学校免费接种新冠疫苗，要不要去打？你接种新冠疫苗了吗？（3973 回答）；②如何看待山东接种疫苗后可升级为金色健康码？你解锁了吗？（520 回答）；③如何看待张文宏称害怕打疫苗是人性问题？（1310 回答）；④新冠疫苗不打好对后面的生活会有影响吗？（500 回答）



图 5 爬取知乎的话题示例

A	B
rid	text
1	其实很早就有网友提出，把接种过疫苗的人升级为金色健康码可以提高疫苗
2	表明持有此金色健康码的人受到新冠疫苗的保护那个画面跟武侠世界里面的
3	凉心四十八天的视频凉心四十八天的视频凉心四十八天的视频
4	用这个带着金色的健康码，对比下某大象国度形如地狱般的黑红色调，真是
5	起码，能够提高一些人接种疫苗的积极性挺好的，值得肯定山东的很多社会
6	所以，金色健康码这样看似“土”的办法，恰恰可能是很实用的办法——增强
7	最好是做成那种满屏都是金色，出示健康码时都晃眼的效果如果能够选择静
8	如果深圳搞个接种疫苗，可获得【金色微信头像框】你看看年轻人疫苗接种
9	你求爷爷告奶奶都不一定有苗，这次推行这个黄金二维码也是为了给鼓励大
10	当然还有最后一招(也最接近现实)，疫苗限时免费，限免期后短暂恢复“原价”
11	记得最开始新冠疫苗200一针，很多人还要排队去打，现在全民免费了，又不
12	当然，很多人不去打疫苗，也是由于对疫苗有顾虑，借此时机，也向大家科
13	随着疫苗接种竞赛进入全面高潮，肯定还会解锁更多的东西请仔细留心看看
14	我们防疫科主任说，这叫金钟罩铁布衫，百毒不侵。
15	从武汉爆发到封城切断传播链阻断蔓延，到各地拉紧门户，宁愿牺牲经济、
16	假如说以后要打针对新型变异（比如印度的）加强版疫苗，能不能把整个健
17	怎么，你咋不说这叫区别对待，这叫歧视不打疫苗的儿童有的人就不配拥有
18	这是好事啊。但我还没解锁。倒不是因为我没接种新冠疫苗，而是因为贵州
19	就很帅！过几天就该打第二针了，坐等云南来个七彩边框，四角有大象、
20	满城尽带黄金甲！我的也升级啦哈哈！希望以后不会被印度的变异病毒破防
21	想起金色盾牌，热血铸就虽然接种疫苗没有警察和医生那么大贡献，但至少
22	我也想要金色健康码。。。但是我深知这个手写的，无防伪，无加密，无公
23	建议全国推广 我们两针都打了的小白鼠，整个七彩玛丽苏健康码不过分吧（
24	不过这一切的前提依然是有足够的疫苗供应，而且对确实疫苗过敏和有禁忌

图 6 爬取的知乎内容

4.2 数据观察

此次项目采用的情感数据集是 weibo_senti_100k 数据集，一共有 119988 条带情感标注的新浪微博评论，其中正负向评论均为 59994 条。

在进行实验之前，我们小组进行了对数据集、爬取的微博文本以及知乎文本的长度与对应的数量进行了统计，大致结果如图所示。

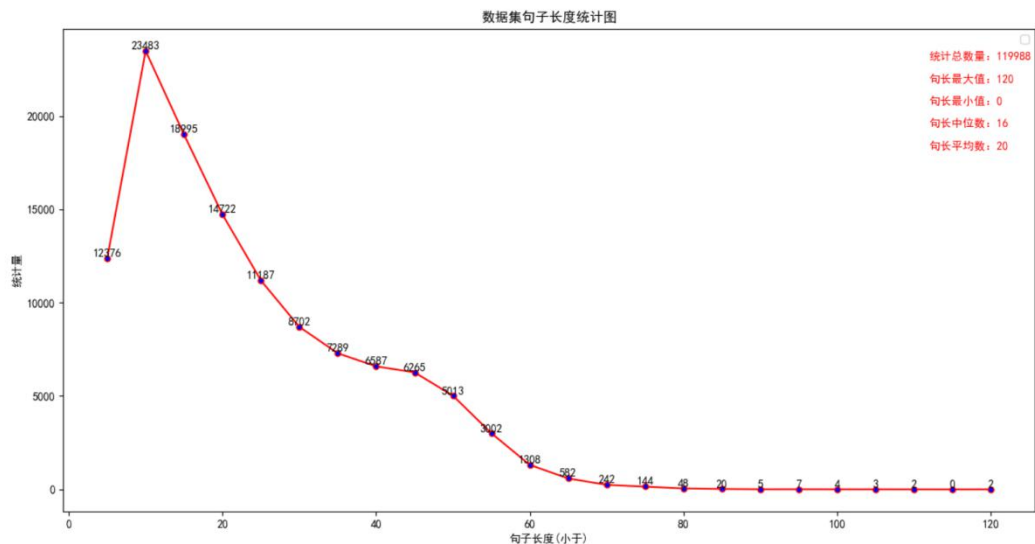


图 7 数据集句子长度统计图

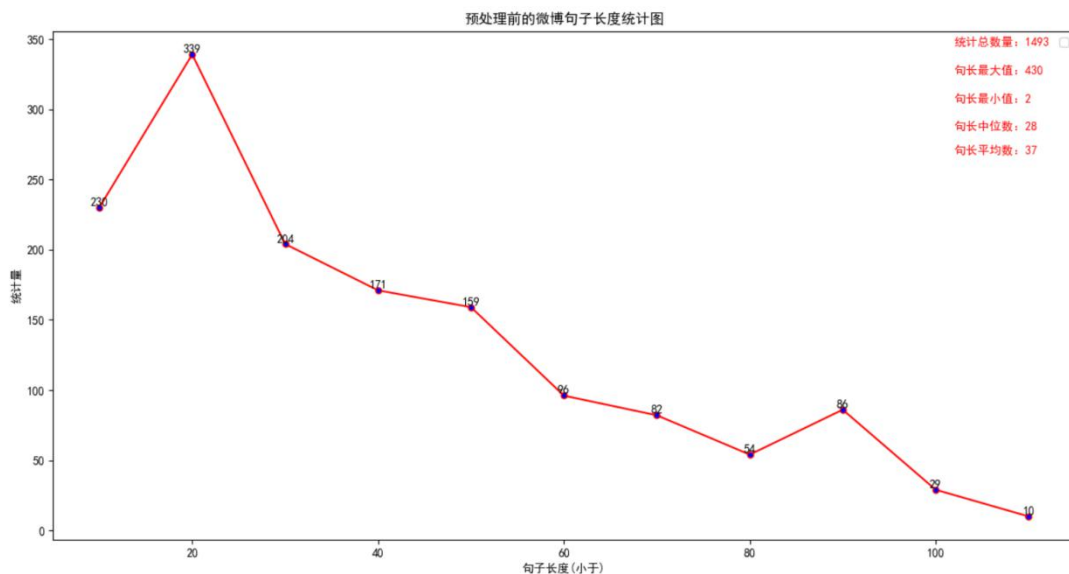


图 8 预处理前的微博句子长度统计图

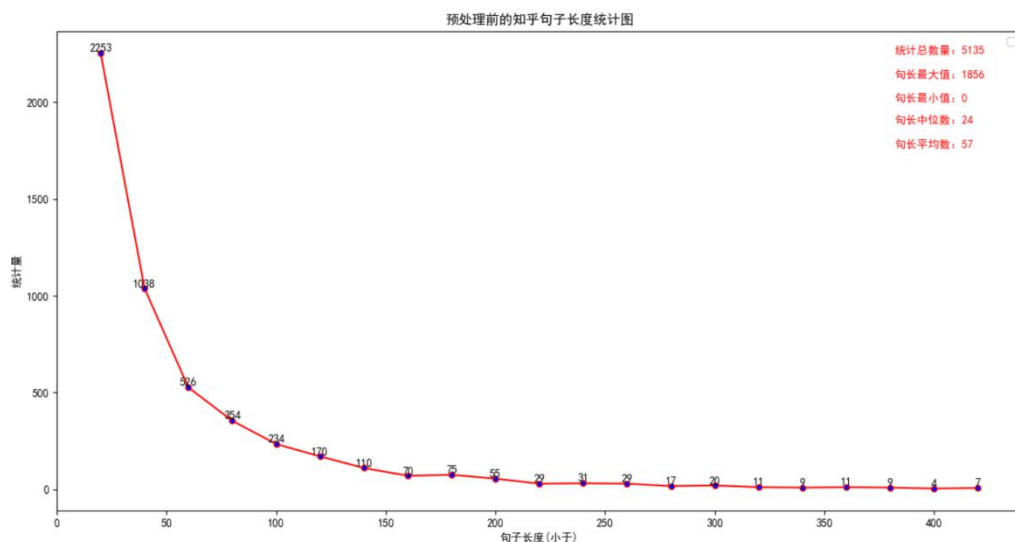


图 9 预处理前的知乎句子长度统计图

对句子进行分词、去除停用词后得到的词语个数，即为图中的句子长度所表达的意思。

通过这三个图我们不难发现，我们用于训练模型的数据集文本的长度相对较短，句长中位数仅为 16，平均数仅为 20。而我们爬取的知乎文本的句长要远高于数据集的文本长度的文本数量较多，因此，为避免后续模型对知乎文本的分类效果不佳，我们在后续的数据预处理操作增加了对句长较大的知乎文本进行摘要提取的工作，具体见 4.3 小节的内容。

4.3 数据预处理

为避免一些对文本情感分类没有任何帮助的文本内容对最终的实验结果造成干扰，需要对其进行清除操作。而由于爬取得的微博文本和知乎文本存在着

不同的问题，因而需要分开进行预处理操作。

4.3.1 微博文本

下表是我们爬取到的关于话题“一起苗苗苗苗苗”的部分微博文本内容，其中 rid 为微博文本对应的序号。

表 1 话题“一起苗苗苗苗苗”的部分微博文本内容

rid	微博内容
	已接种预防性 HPV 疫苗的妇女还需要进行子宫颈癌筛查吗？#hpv#疫苗
19	目前，约有 30%的子宫颈癌不能通过接种 HPV 疫苗预防。由于接种 HPV 疫苗后仍有可能发生宫颈癌，所以 25~64 岁有性生活的所有女性，即使接种过预防性 HPV 疫苗，仍需定期接受子宫颈癌筛查。 #一起苗苗苗苗苗##九价 hpv# #hpv# 0 讨论 0 阅读
43	#一起苗苗苗苗苗# 下午终于把疫苗给打上了，没有任何不适症状。打的人挺多的，早日建立群体免疫吧！ 动图 #一起苗苗苗苗苗#
53	跑了三个点终于在排队两小时后打上 来世还生种花家 祖国抗疫工程真伟大 广州
158	#一起苗苗苗苗苗# 夫妻双双把苗打 滁州·定远县 重点在社区！！ #一起苗苗苗苗苗# #新冠病毒疫苗# @央视新闻：【 #基层医疗卫生机构是疫情第一道防线# 】辽宁、安徽都有一个似曾相识的细节：基层诊所和社区卫生服务站接诊了有疑似症状的病人，却没上报，没能尽早发现病例。目前已有多人被追责，但避免重蹈覆辙才是关键！基层医疗卫生机构是疫情防控的第一道防线，把好这个小门，才能守好疫情防控的大门。值得警惕！... 全文
164	166 万次播放 1:18

由于我们这次项目的主要任务就是在这些话题下进行舆情分析，即获取在这些话题下的情感占比情况，所以不该利用话题本身的文本内容作为特定的文本情感特征用于模型训练，因而将其除去。同时，像“动图”、“0 讨论 0 阅读”、“166 万次播放”、“1:18”这些非用户输入的文本信息以及“广州”和“滁州·定远县”这些地址信息，并不能为文本的情感分类提供帮助，因而将其去除。最终清洗后的微博文本句子长度统计图如下。

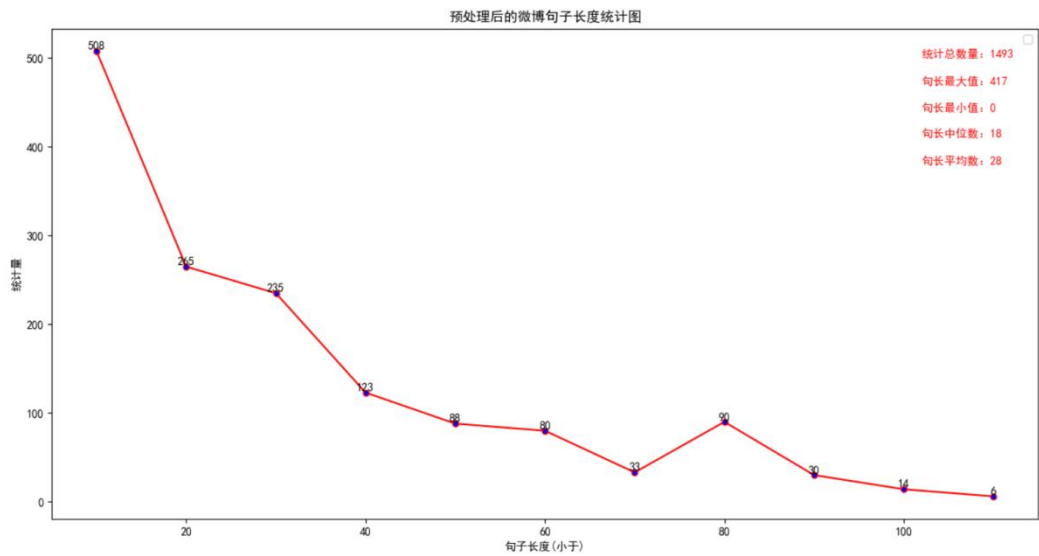


图 10 预处理后的微博句子长度统计图

4.3.2 知乎文本

下表是我们爬取到的关于问题“如何看待山东接种疫苗后可升级为金色健康码？你解锁了吗？”的部分知乎回答内容，其中 rid 为知乎文本对应的序号。

表 2 部分知乎回答内容

rid	回答内容
	非常有创意，这健康码看起来非常给力，等级也看起来高许多，金色比紫色要强大。顺便支持下我的家乡，如果再给拥有金色健康码的人送个 888rmb 的大红包，就更好了，就更完美了。
	边框稍微突出了下
	金色象征着高贵、光荣、华贵和辉煌。金色可以象征着闪耀、光辉、光明的代表，所以通常用金色制作衣服。在许多国家，因为黄金的颜色是金色，所以金色代表至高无上。
	预计后续健康码上可以继续做文章了，比如后面可以推出电子版的基于健康码的疫苗护照等。
4	人在东南亚，开始想家了，毕竟一年多没回了，就等着健康码升级到疫苗护照，来回不用隔离，我就可以回家了。
	回家的感觉真好。
	用这个带着金色的健康码，对比下某大象国度形如地狱般的黑红色调，真是强烈的视觉震撼。
	某大象国的色调
	祖国大地上的 5.1 节日，人潮人海但无疫情之忧。
	遥祝祖国更强大，更美好，大家更幸福，更安全。
	想方设法鼓励接种，当然很看好这种方式。
	在过去的一年多时间里，全国以巨大的经济代价为牺牲，保住了绝大多数人的健康利益，如今面对即将到来的“九月或十月，世界将重新开放，我们绝大多数人不打疫苗就没有免疫力”的状况，
51	每个适合接种的人都有义务完成，只有足够高的接种（成功）比例才能完成群体免疫，达到阻断传播的目的。而现在的实际情况却是很多人不主动接种，迫使一个保护措施变成了政治任务，让很多人想尽办法吸引人来打针，我们的教育力度中，科普教育尤其薄弱，这种没谱的状况下，搞一些小窍门也算是被逼出来的办法了。
391	看了其他人的回答，有一部分在说想要布灵布灵满天星特效的闪光金码，我之前也想那样的好看，

但又一想好像不行，原因如下：

1. WX 技术上或者不支持。众所周知，wx 是无法发送动图的，我去试了一下朋友圈和聊天框都是无法发送动图的，如果想要布灵布灵的动图 wx 方需要改动技术支持（这条是我的猜测，也许可以，也许不用改动也有办法）
2. 在金码的加持下仍需要红黄绿三码。众所周知，打了疫苗不是一定 100%安全的，健康码提供的是你的行程轨迹，如果在风险地区回去的还是要变成黄码的，如果直接全金码的话会直接变成全黄码，那就体现不出疫苗的作用了，毕竟咱是有一层加持保护的嘛。
3. 有一定概率会把金码看成黄码。这个其实是可以改动的，把黄码饱和度调高让它显示出亮黄色，金码是暗金色的或者如 1 所说金码布灵布灵闪光，是可以克服这个问题的，但是 2 的问题还是问题。

第一个吃螃蟹的人总会被无知的人骂，期待其他省的创新，也期待我大山东的进步。

提个建议：两针的疫苗，可以第一针是银色边框，第二针是金色边框；三针的疫苗，第一针是紫色第二针是银色第三针是金色。毕竟没全打完也是有抗体的，而且可以很好地提醒自己打到第几针了。

期待我的健康码也有金边保护的那一天，期待全民免疫的那一天！

结合 4.1 小节的内容，我们不难发现，需要对知乎文本的预处理操作主要是进行文本摘要。而这里我们主要利用到的是 `textrank4zh` 这个 Python 的第三方模块。

最终清洗后的知乎文本句子长度统计图如下。

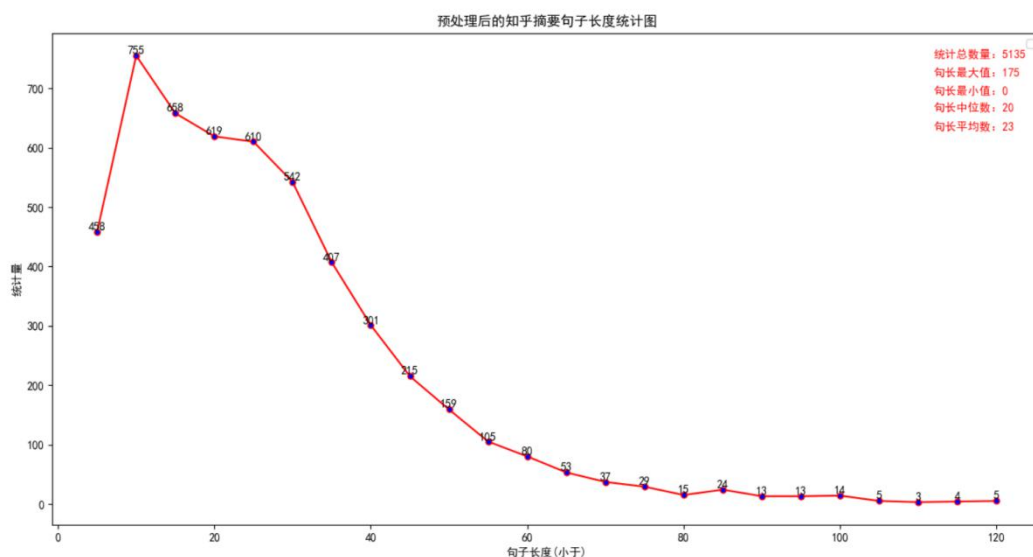


图 11 预处理后的知乎摘要句子长度统计图

4.4 文本分词

中文分词是中文文本处理的一个基础步骤，也是中文人机自然语言交互的基础模块。不同于英文的是，中文句子中没有词的界限，因此在进行中文自然语言处理时，通常需要先进行分词，分词效果将直接影响词性、句法树等模块的效果。当然分词只是一个工具，场景不同，要求也不同。

在人机自然语言交互中，成熟的中文分词算法能够达到更好的自然语言处理效果，帮助计算机理解复杂的中文语言。在构建中文自然语言对话系统时，结合语言学不断优化，训练出了一套具有较好分词效果的算法模型，为机器更好地理

解中文自然语言奠定了基础。

中文分词根据实现原理和特点，主要分为以下 2 个类别。

4.4.1 基于词典分词算法

也称字符串匹配分词算法。该算法是按照一定的策略将待匹配的字符串和一个已建立好的“充分大的”词典中的词进行匹配，若找到某个词条，则说明匹配成功，识别了该词。常见的基于词典的分词算法分为以下几种：正向最大匹配法、逆向最大匹配法和双向匹配分词法等。

基于词典的分词算法是应用最广泛、分词速度最快的。很长一段时间内研究者都在对基于字符串匹配方法进行优化，比如最大长度设定、字符串存储和查找方式以及对于词表的组织结构，比如采用 TRIE 索引树、哈希索引等。

4.4.2 基于统计的机器学习算法

这类[4]目前常用的是算法是 HMM[5]、CRF、SVM、深度学习等算法，比如 stanford、Hanlp 分词工具是基于 CRF 算法。以 CRF 为例，基本思路是对汉字进行标注训练，不仅考虑了词语出现的频率，还考虑上下文，具备较好的学习能力，因此其对歧义词和未登录词的识别都具有良好的效果。

常见的分词器都是使用机器学习算法和词典相结合，一方面能够提高分词准确率，另一方面能够改善领域适应性。如双向 LSTM+CRF 实现中文分词。

对于本次项目，我们小组采用了结巴分词，并且在原本的语料上加载了搜狗语料库、百度语料库、腾讯语料库、盘古语料库以及自定义的语料库，以此来提升分词的准确率。

4.5 文本特征提取

本次项目便是利用 word2vec[6]生成特征向量用于模型训练。

Word2vec，是一群用来产生词向量的相关模型。这些模型为浅而双层的神经网络，用来训练以重新建构语言学之词文本。网络以词表现，并且需猜测相邻位置的输入词，在 word2vec 中词袋模型假设下，词的顺序是不重要的。训练完成之后，word2vec 模型可用来映射每个词到一个向量，可用来表示词对词之间的关系，该向量为神经网络之隐藏层。

通过 Word2Vec 算法得到每个词语的高维向量（词向量，Word Embedding）表示，词向量把相近意思的词语放在相近的位置。我们只需要有大量的某语言的语料，就可以用它来训练模型，获得词向量。而在此次项目中，我们采用的训练语料就是 weibo_senti_100k 数据集的文本。

4.6 模型建立与训练

本次项目选用了两种深度学习的模型进行数据训练，分别为 Bi-LSTM 和 TextCNN 模型。我们将数据集按照 8:2 的比例分为训练集和测试集，并且，两者采用的损失函数都是交叉熵损失函数，都使用 Adam 进行优化。

4.6.1 Bi-LSTM

本次项目构建的 Bi-LSTM 模型大致结构如下：词嵌入-双向 LSTM-Dropout-全连接。

最终模型评估效果如下图所示。

表 3 Bi-LSTM 模型评估表

	precision	recall	f1-score	support
负面	0.99	0.98	0.98	12001
正面	0.98	0.99	0.99	11997
micro avg	0.99	0.99	0.99	23998
macro avg	0.99	0.99	0.99	23998
weighted avg	0.99	0.99	0.99	23998
samples avg	0.99	0.99	0.99	23998

4.6.2 TextCNN

本次项目构建的 TextCNN 模型大致结构如下：词嵌入-卷积池化*3-拼接-全连接-Dropout-全连接。

最终模型评估效果如下图所示。

表 4 TextCNN 模型评估表

	precision	recall	f1-score	support
负面	0.92	0.62	0.74	12024
正面	0.71	0.95	0.81	11974
micro avg	0.78	0.78	0.78	23998
macro avg	0.82	0.78	0.78	23998
weighted avg	0.82	0.78	0.78	23998
samples avg	0.78	0.78	0.78	23998

由于时间问题，我们小组未能将 TextCNN 模型调至最优参数，因而可能导致了 TextCNN 的分类效果不够理想。所以后续对我们事先爬取并清洗好的文本，我们选择了效果更好的 Bi-LSTM 进行情感评估，具体评估结果见第 5 部分——实验结果及分析。

4.7 提取关键词与生成词云

在自然语言处理中，TF-IDF 和 TextRank 算法是较为常见的两种关键词提取算法，算法大致介绍如下。

4.7.1 TF-IDF

TF-IDF (Term Frequency/Inverse Document Frequency) [7] 是信息检索领域非常重要的搜索词重要性度量；用以衡量一个关键词 w 对于查询 (Query, 可看作文档) 所能提供的信息。词频 (Term Frequency, TF) 表示关键词 w 在文档 D_i 中出现的频率：

$$TF_{w,D_i} = \frac{\text{count}(w)}{|D_i|} \quad (5)$$

其中， $\text{count}(w)$ 为关键词 w 的出现次数， $|D_i|$ 为文档 D_i 中所有词的数量。逆文档频率 (Inverse Document Frequency, IDF) 反映关键词的普遍程度——当一个词越普遍 (即有大量文档包含这个词) 时，其 IDF 值越低；反之，则 IDF 值越高。IDF 定义如下：

$$IDF_w = \log \frac{N}{\sum_{i=1}^N I(w, D_i)} \quad (6)$$

其中, N 为所有的文档总数, $I(w, D_i)$ 表示文档 D_i 是否包含关键词, 若包含则为 1, 若不包含则为 0。若词 w 在所有文档中均未出现, 则 IDF 公式中的分母为 0; 因此需要对 IDF 做平滑 (smooth):

$$IDF_w = \log \frac{N}{1 + \sum_{i=1}^N I(w, D_i)} \quad (7)$$

关键词 w 在文档 D_i 的 TF-IDF 值:

$$TF-IDF_{w, D_i} = TF_{w, D_i} * IDF_w \quad (8)$$

4.7.2 TextRank

TextRank[8] 是基于 PageRank 的, 所以首先简要介绍下 PageRank 算法。

4.7.2.1 PageRank

PageRank 通过网页之间的超链接来确定页面的重要性, 它将整个互联网可以看作是一张有向图, 网页是图中的节点, 网页之间的链接就是图中的边。根据重要性传递的思想, 如果一个大型网站 A 含有一个超链接指向了网页 B, 那么网页 B 的重要性排名会根据 A 的重要性来提升。

构造完图后, 使用下面的公式来计算网页 i 的重要性 (PR 值):

$$S(V_i) = (1-d) + d * \sum_{j \in \text{In}(V_i)} \frac{1}{|\text{Out}(V_j)|} S(V_j) \quad (9)$$

其中 V 表示网页, S 表示每个网页的 score, S 越大表示网页的重要程度越高。 d 是阻尼系数, 一般取 0.85。

$\text{In}(V_i)$ 表示存在指向网页 i 的链接的网页集合。 $\text{Out}(V_j)$ 表示网页 j 中的链接指向的网页的集合; $|\text{Out}(V_j)|$ 是集合中元素的个数。

4.7.2.2 TextRank

TextRank 在构建图的时候将节点由网页改成了句子, 并为节点之间的边引入了权值, 其中权值表示两个句子的相似程度, 本质上构建的是一个带权无向图, 其计算公式如下:

$$WS(V_i) = (1-d) + d * \sum_{j \in \text{In}(V_i)} \frac{w_{ji}}{\sum_{k \in \text{Out}(V_j)} w_{jk}} WS(V_j) \quad (10)$$

其中 w_{ji} 是节点 j 到节点 i 的权重。

在本次项目中, 我们首先利用了 TF-IDF 提取每一个话题文件下的每一句话的关键词, 再利用 TextRank 提取整个话题的关键词, 最终生成对应的词云。具体结果见第 5 部分——实验结果及分析。

5 实验结果及分析

5.1 平台分布

新浪微博是基于用户关系的社交媒体平台，用户可以简单地通过手机、PC 等多种移动端，以文字、视频、图片等方式公开实时发表信息。作为继搜索、门户之后的新一代互联网入口，微博改变了信息的传播方式，实现了信息的即时分享。根据新浪微博 2020 年 Q1 财报，其月活跃用户已达 5.11 亿人，是国内首屈一指的大型社交平台。

由于新浪微博有着用户众多、时效性强的特点，成为了本次舆情分析的对象平台之一。

知乎则是一种社区式的问答平台，和微博一样拥有用户众多，时效性也极强的特点。同微博不同之处在于，知乎的“信息”单元并非单一的一条回答或者一个问题，而是一个问题与其下所有回答的集合：不同角度的回答以及各种观点的比例可以一窥人们对相关问题的具体态度分布情况。

因此，选取具体的分析对象时有必要考虑整一个问题的关注人数等。

5.2 用户情绪分析

5.2.1 舆情背景

因受新冠疫情影响，公众对于新冠疫苗的接种问题成为舆论场上的焦点话题，而随着国家卫生健康委副主任、国务院联防联控机制科研攻关组疫苗研发专班负责人曾益新确认新冠病毒疫苗将为全民免费提供，关于疫苗接种的各类话题热度便在社交网络中居高不下。

纵览两大平台收集到的内容，我们不难发现用户给予了接种疫苗这个话题极大的关注度。以新浪微博为例，自 2021 年 1 月 4 日@人民日报 发文：“【新冠疫苗全民免费！#20 问了解新冠疫苗接种要点#】1 月 1 日起，重点人群开始接种我国首批新冠病毒疫苗。哪些人不适合接种新冠疫苗？#接种新冠疫苗注意事项#有哪些？接种疫苗需要打几针？近期接种过其他疫苗怎么办？就新冠疫苗的安全性、好处、优先接种人群、接种地点、禁忌人群、年龄要求、注意事项、留观、不良反应及处理、接种针数、产生抗体时间、疫苗效果、抗体可持续时间、与 HPV 疫苗是否能同时接种、与其他疫苗间隔时间、接种所需证明、接种疫苗后的口罩佩戴问题、接种证明能否代替核酸检测报告”，对等网友最为关心的问题进行了解答，博文被大量转发传播。

此外，@央视新闻 的发文【直播！#央视探访新冠病毒疫苗生产基地#】让公众更加贴近疫苗的生产情况。@北京日报、@天津交通广播 @人民网 等媒体账号对北京、天津等地新冠疫苗接种情况进行了报道，@中国新闻网、@环球网、@新华视点、@新京报、@环球时报、@每日经济周刊、@财经网 等媒体都对新冠疫苗接种相关问题进行了解答报道，助推舆情热度不断升高，于 1 月 4 日达到峰值。

舆情动向轨迹：急剧升高，逐渐降低，平稳波动。

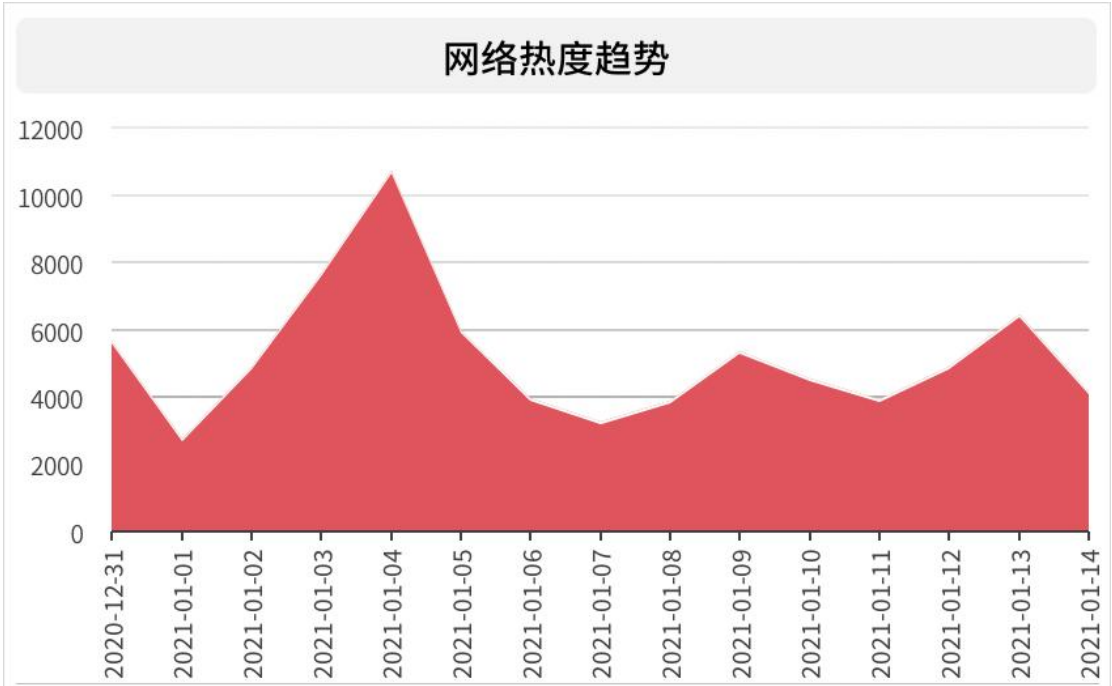


图 12 网络热度趋势图

在信息传播的过程中，新闻类的网站也发挥了重要的作用。搜狐新闻、今日头条等纷纷对接种相关问题及注意事项进行了报导。从消息传播次数之多和话题之广可以看出这一利好消息的受欢迎程度。

热门文章		
[新闻客户端] 新冠病毒疫苗接种，一组图带你全看懂	搜狐新闻 2021/01/14 10:55	传播9999次，163个网站 ↓
[新闻客户端] 重磅！新冠病毒疫苗全民免费	网易号 2021/01/07 09:55	传播7515次，128个网站 ↓
[新闻客户端] 新冠疫苗全民免费接种，全部费用政府出！	搜狐新闻 2021/01/13 07:30	传播3714次，126个网站 ↓
[新闻客户端] 好消息！新冠病毒疫苗，全民免费	今日头条 2021/01/12 07:49	传播3625次，44个网站 ↓
[新闻客户端] 来了！新闻早班车	58同镇 2021/01/14 11:21	传播2890次，15个网站 ↓
[新闻] 新冠疫苗开始接种！这份指南请查收→	贵阳网 2021/01/14 14:03	传播2479次，57个网站 ↓
[新闻客户端] 还在纠结打不打新冠病毒疫苗？看完就明白了	搜狐新闻 2021/01/14 12:47	传播2214次，48个网站 ↓

图 13 热门话题图

5.2.2 舆情背景下双平台用户情绪分析

5.2.2.1 微博的评论为短评居多，通常表现的是博主对接种疫苗一事的感想。或是描述接种完自身的感受，有抱怨排队困难的，有表达忧虑的，也有乐观正面的。以微博话题“一起苗苗苗苗苗”的部分评论为例分析：

中午有个辖区企业联系人找我打疫苗，听说他们企业同事都没打，我扔下筷子开着车就去接人。仿佛一个月底冲业绩的可怜中介。路上还发了一条朋友圈：c80周边打疫苗找我，车接车送！车到企业楼下，接种点传来消息：排队，排一百多号人了。收割蓝湾计划泡汤，只好改期。回来的路上还顺便发了个会议通知。车开得如同一个王八蛋，俩马路司机底下窗户骂我

图 14 偏负面

为什么他们打疫苗都没啥反应啊 我每次打疫苗胳膊都得肿好几天 走路都不能幅度太大 睡觉还不能侧着睡 我这是什么体质呢 害！

图 15 偏负面

两个人还担心各种副作用，结果一点都不疼，啥事也没有

图 16 偏正面

完成两针打疫苗大计 北京生物加成都生物 感谢伟大的祖国

图 17 偏正面

钟南山近日接种了新冠疫苗，并表示“感觉很好”。他还呼吁大家尽快接种新冠疫苗，因为这是非常好的实现群体免疫的方法。我国接种率较低，要争取尽快达到70%—80%的接种率：“希望大家一起努力，尽快接种疫苗，为国家、世界、也为个人预防新冠尽一份额。” m微风的微博视频

图 18 偏正面

图 14 至图 15 是相关话题被分类为情绪负面的评论，推测文字表达内容可知，图 1 可能是由临时通知导致日程安排冲突，也可能是接种现场秩序相对混乱；图 2 可以看出，是个人自身的体质差异导致的接种后反应较大。

图 16 至图 18 是相关话题被分类为情绪正面的评论，由文字表达内容可知，三条评论皆表示了自身接种之后“感觉良好”。

5.2.2.2 知乎平台获取的评论主要针对的是某个问题，因此也需要对问题的情绪进行分类。这里以一个情绪正面和情绪负面的问题下的回答为例，做初步简单的分析：

① “不打新冠疫苗对后面的生活会有影响吗？”

这个问题本身在情绪分析中被判定为负例，但结合题主对问题的描述（题主目前在吃药，不符合接种需求），不难理解提问者本身的态度以及立场是支持接种的。

不知道为什么大家还会有这么严重的疑虑，我们单位已经基本都打完了，最早的一批也有几个月了最严重的副作用就是呕吐，嗜睡，过去了也就没什么感觉了我是第二针打完，打疫苗的地方有点儿疼，大概疼了2-3天，之后也没事儿了

图 19 偏正面

昨天刚打了第一针，除了有点口渴，晚上睡得香外，没一点感觉，当天晚上因为天气热出汗多，还洗了个澡，同行三个人都没有啥感觉。等待三周后的第二针。灭活疫苗还是比较安全的。

图 20 偏正面

打完两三天了，一点儿影响都没有，就是变得能吃了

图 21 偏正面

图 19 至图 21 皆表示了评论者恢复良好，但是与主题并不十分契合；

会，只要有一个新冠患者跟你接触，你传染上的几率就高于50%，所有跟你接触过的人都会增加得病率，打了至少可以把这个概率降低，毕竟疫苗只是让你不得重病不是完全不让你得病

图 22 偏负面

第一针打完后续干干嘛，身体倍棒，吃嘛嘛香，第二针打完饭后量多了，晚自习还睡了半小时

图 23 偏正面

打，肯定要打，对于即将来临的北京冬奥会，我们有责任有义务打，到是大量外国人进入中国，有可能病毒携带过来，如果到时没有形成全民免疫，中国将再次面临疫情的洗礼。

图 24 偏正面

图 22 至图 24 前一者分类为负面，后两者分类为正面。前者表示若未接种疫苗感染病毒的几率较高；后两者表示的是对接种的支持。

②问题二“山东接种疫苗后可升级为金色健康码”

这个问题本身在情绪分析中被判定为正例。

你求爷爷告奶奶都不一定有苗，这次推行这个黄金二维码也是为了给鼓励大家都打疫苗，把自己的健康段位升上去啊，用心了~这个真心不错，我记得刚开始打疫苗的时候，是求爷爷告奶奶求你去，但是担心的人多，打的人少——针黄金，两针铂金，三针翡翠，四针钻石

图 25 偏正面

满城尽带黄金甲！我的也升级啦哈哈！希望以后不会被印度的变异病毒破防！很漂亮的健康标识有没有！4月1日接种的第一针4月30日接种的第二针两针之间间隔3到8周都OK

图 26 偏正面

不过真的感觉到我们国家体制的优越性了，现在就盼着南边的雷千万别炸到我们这次算是打开了个新思路，对上了年纪的大爷大妈就送纸送鸡蛋，对年轻人就给金色加持来提高积极性刚看到消息就立刻从支付宝打开健康码了(平时一直用的是支付宝里的健康码)，可是支付宝还没有更新，弄得我还以为是假消息，从知乎上翻了几个回答以后才确信，立马就去微信申请了个微信的健康码

图 27 偏正面

新冠本来就是人类的不幸，更不幸的是一个病毒没有把人的肉体毁灭却把人类重新带回了额头刺字的时代，健康码不能用于划分人的不同，金色健康码更是对人类尊严的践踏，人不能因为打不打疫苗而被贴上标签，这是一种歧视，国家相关部门必须马上介入制止这种荒唐的做法

图 28 偏负面

图 25 至图 27 皆是偏正面的回答，表示对这种富有新意、年轻化有活力的鼓励接种疫苗方式的支持。图 28 则明确表示反对这种人们凭借是否接种疫苗而被打标签区分的做法。

5.3 热词图

(1) 国内灭活疫苗对多数变异株有效。无疑，在疫苗相关话题构成的词云

当中疫苗二字是绝对的中心地位，其次大家最关心的是来自国外的变异病毒与疫苗的接种显示。第三高的关注度是国内安徽的疫情。



图 29 国内灭活疫苗对多数变异株有效话题的热门词

(2) 新冠疫苗知识点。我们可以从这份词云中看出，网友们关注的接种注意事项有：不同人群适应度、专家观点、有效期、是否与其他疫苗接种、异地接种如何衔接、保护期有多久等。



图 30 新冠疫苗知识点话题的热门词

(3) 一起苗苗苗苗苗。我们可以从这份词云中看出，网友们纷纷表达的是社区基层接种点井然有序，志愿者们组织良好，感激医护人员的辛苦付出，也说明了基层接种点包括社区、学院等具体组织类型。

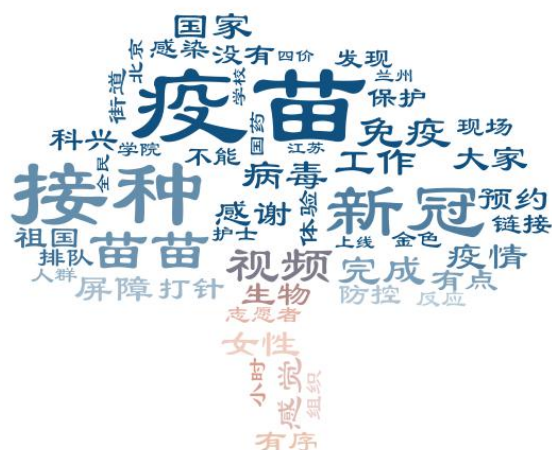


图 31 一起苗苗苗苗苗话题的热门词

(4) 疫苗。我们可以从这份词云中看出，网友们的关注点倾向于疫苗接种的流程，例如排队、小时、截至、注射、社区等词，以及国内、外的疫情现状，例如广州、英国、感染者等词。此外就是一些疫苗相关的信息，例如国家发布、科兴、号召、免费等词。



图 32 疫苗话题的热门词

(5) 如何看待张文宏称害怕打疫苗是人性问题。由词云可以看出，大家关于张文宏的说法的讨论点集中于是否需要接种疫苗。尽管回答的角度五花八门，但形成词云后可以发现是符合我们最初的选择目的的。

图 35 新冠疫苗不打对后面的生活会影响吗话题的热门词

（8）在校大学生学校免费接种新冠疫苗。在这个话题下大家的回答相对比较集中，都在反馈自身接种后的状况以及注意事项。从问题本身限定来看，回答者是大学生可能性比较大。



图 36 在校大学生学校免费接种新冠疫苗话题的热门词

5.4 话题分布

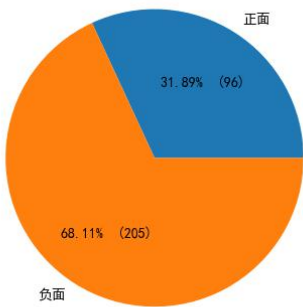


图 37 国内灭活疫苗对多数变异株有效话题的情感占比图

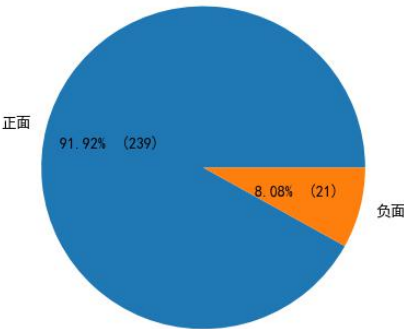


图 38 新冠疫苗知识点话题的情感占比图

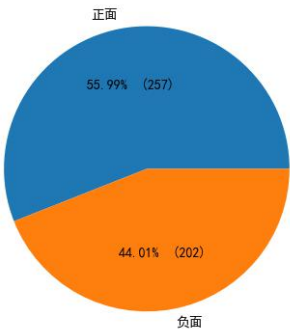


图 39 一起苗苗苗苗话题的情感占比图

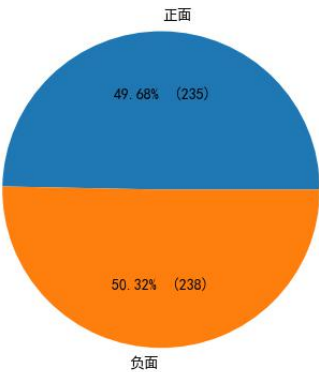


图 40 疫苗话题的情感占比图

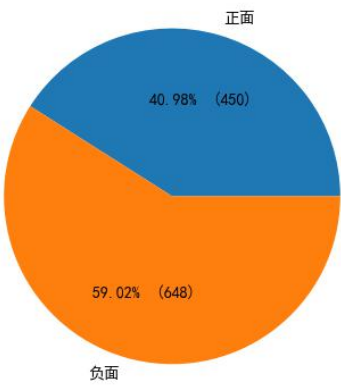


图 41 如何看待张文宏称害怕打疫苗是人性问题话题的情感占比图

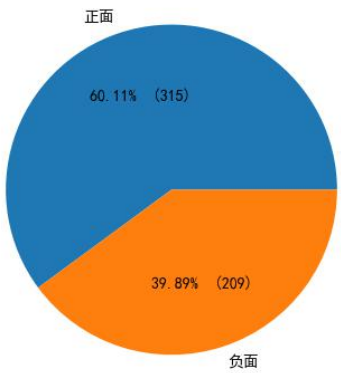


图 42 山东接种疫苗后可升级为金色健康码话题的情感占比图

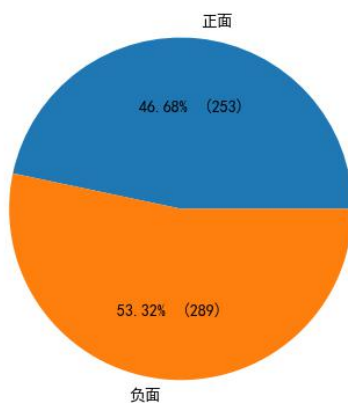


图 43 新冠疫苗不打对后面的生活会影响吗话题的情感占比图

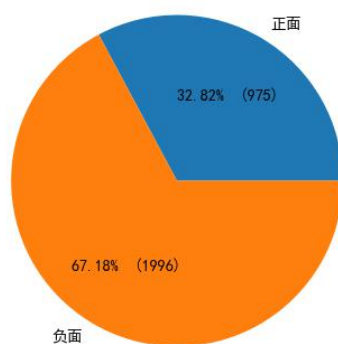


图 44 在校大学生免费接种新冠疫苗话题的情感占比图

6 总结与体会

6.1 各平台舆情

6.1.1 微博舆情分析

由于微博的信息表现形式更多的倾向于传播而非人与人之间的直接交互，由第三部分的话题分布可以看出：针对接种疫苗这一话题，微博上持正面情绪的消息更多，负面情绪的消息更少。唯一一个负性比例大于正性的话题是 tag “国内疫苗对多重变异株有效”，推测原因在于话题本身可能引发网友对外国疫情的焦虑情绪更大，因此负性情绪的占比更多。

此外，在微博平台上官方账号传播的力度较大，所收到的反馈也较多，因此可以认为官方账号在微博平台起到的宣传作用与引导作用是比较大的。

6.1.2 知乎舆情分析

与微博的信息表现形式不同，从“用户提问-用户解答”和“用户解答-用户评论-用户更新解答”这两个交互模式来看，知乎的互动性更为强大一些。

然而这样导致的问题则是网民之间容易发生意见分歧，并且很容易偏离问题的中心。不同价值观的正面碰撞也易于导致社区戾气加重，无利于讨论问题。

本次舆情分析的目的介绍中我们提到，在这样的交互模式中，用户很容易受到某些片面观点输出的冲击，产生不良的心理反应。当个体处于这种感到“有威胁”的环境中，容易自动引发应激反应，例如产生紧张，焦虑的情绪。因此在知乎开展宣传工作所需付出的努力是相对微博要大的，但也是非常有必要的，同时成效也将是可预见的明显的。

6.2 总体舆情

两个平台对接种疫苗表现出的态度不一而同，总体而言偏向于支持接种的观点占比较多。但是鉴于目前的接种率仍然未达到目标，因此仍有宣传引导的价值。

6.3 建议

尽管在表现形式上有所区别，但总体不难看出大家对接种新冠疫苗的关心程度极高，因此仍有宣传引导的价值。

新浪微博与知乎同为国内数一数二的流量平台，其受众的广泛性、观点的多样性所带来的影响是巨大的。把握好这两个宣传阵地，有利于安定社会舆论场，有助于提高疫苗的接种率，尽早打到群体免疫屏障的标准，恢复原有的生产生活秩序。

以下是本次舆情分析所得出的有关宣传接种新冠肺炎疫苗的几条建议。

①基层可以使用简单有趣，朗朗上口的标语，淡化接种疫苗的仪式感，减少人们的心理斗争；

②针对社交平台上抱怨接种后身体反应大的现象，官方可以请权威的医疗工作者科普介绍接种疫苗后有反应属于正常现象，不应妖魔化。同时为民众打上心理上的“预防针”；

③适当地公开数据，公开透明化使民众安心；

④对有失偏颇的、过于偏激的评论进行限流，降低负面影响；

⑤线上志愿者详细解释接种流程以及注意事项，或者开通人工助理线上办理接种申请；

⑥利用名人效应，着重宣传接种；

⑦统一发布接种注意事项，如可否洗澡，可否运动，接种适应症等等，给出一个金标准，降低民众做出选择的成本，帮助民众“偷懒”获取正确的信息；

⑧适当公开疫情数据，提升接种紧迫感；

⑨走软路线宣传，而非强打、滥打温情牌。呼吁向民间借脑，通过民间的暖心呼吁提升接种的认同感；

⑩各品牌疫苗效果相同的情况下应避免过度宣传介绍疫苗品牌，防止民众扎堆选择某一品牌造成该品牌短缺而另一品牌无人问津等不必要的问题。

6.4 总结思考

当前，新冠疫苗接种工作依然在积极有序地推进。通过对现阶段的舆论检视分析，有助于优化下一步的舆情动员工作。

首先，宣传口应把握好民众的需求为何。在社交网络上，民意多是社会问题的投射。面对疫苗接种，民众们在发出千百个问题时收到的答案却是不清晰的。在本次收集到的评论内容中，相当一部分网友表达了对“朝令夕改”、“前后矛盾”等问题的诟病；同时，也有部分网友固执地持“最好的宣传即是公布实验结果”等偏见和误解，久未得到纠正。如此一来容易损害政府公信力。

其次，舆论动员最重要的一环是引导最作用的发挥。民意宜疏不宜堵，相关部门应积极采取措施引导公众正确参与意见表达，尤其是事关疫苗接种这样的话题。媒体应在舆论动员过程中发挥好主流话语权优势，疏导网民一味沉浸于负面情感宣泄的不利局面，谨防舆论被少数人带偏，避免恶性舆论动员带来的不良后果。本次舆情分析中，这一点在知乎上体现得格外明确。

最后，尽管本次舆情分析针对的是网络平台的线上宣传内容，但线下宣传内容在整体的宣传工作中的作用也是不容忽视的，例如可以选择线下张贴接种须知的海报，在社区、学校内拉挂趣味横幅等。此外还有一个例子，针对“个别地方疫苗接种一刀切、简单化”“强制要求全员接种”的情况，国家卫健委表示“必须坚决予以纠正”，这一线下存在的问题在官方表态下舆论风向才迅速转变。由此可见，自上而下舆论动员的协作协调和线上线下内在一致性显得尤为重要。

6.5 小组体会

吴棉滨：在完成整个项目之后，个人认为对于自身的代码能力和项目经验都是有所帮助的。从数据观察到数据清洗，从文本分词到特征提取，从建立模型到训练调参，以及关键词提取和文本摘要等工作，都让自己对于这些自然语言处理领域所常见或必要的技术有了更深的了解。我目前也比较倾向于在后续学习中去尝试更多这方面的研究，所以这次项目，包括整个学期的课程知识，都为此提供了较大的帮助。最后，也希望自己后面能有机会在这方面有更多的收获。

罗佳：进行舆情分析需要大量获取网民对舆情的看法，这些数据能够帮助我们更准确地把握民众对舆情的情绪方向。实际上，分析及解决其它问题也和做舆情分析一样，不能偏听偏信，而是要获取大量信息再从中筛选出有价值的去使用。

游畅：对于互动程度高网络社区而言维持讨论氛围与讨论秩序是一项非常有挑战的工作，尤其是富有争议的话题。反过来看，当我们抱着信息收集的目的浏览这样的话题时，便需要小心地收集观点，以免偏听偏信。

7 参考文献

- [1]封超,杜娟.国内外舆情治理研究热点及趋势分析[J].陕西行政学院学报,2021,35(01):21-27.
- [2]Nguyen N K , Le A C , Hong T P . Deep Bi-directional Long Short-Term Memory Neural Networks for Sentiment Analysis of Social Data[C]//International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making. Springer International Publishing, 2016.
- [3]Kim Y . Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 2014.
- [4]刘武. 基于统计机器学习算法的汉语分词系统的研究[D]. 北京邮电大学.
- [5]李强,余祥,朱峰,等. 一种基于多层隐马模型的分词方法及系统.
- [6]CHURCH, Ward K . Word2Vec[J]. Natural Language Engineering, 2017,

23(01):155-162.

[7]El-Khair I A . TF*IDF[M]. Springer US, 2009.

[8]Mihalcea R, Tarau P . TextRank: Bringing Order into Text. [J]. Emnlp, 2004:404-411.