

“中印边境事件” 舆情分析报告

黄思佳 朱浅 李泽康 王淇营

2017.12.08

摘要

北京时间 6 月 18 日印度边防人员非法越界进入中国洞朗地区，引发中印边境局势持续紧张。本报告围绕“中印边境事件”，对 2017 年 6 月 18 日到 2017 年 8 月 28 日期间，互联网上采集到的 1522740 条信息进行了全方位的全网舆情分析。主要分析全网舆情特点、各大媒体平台舆情特点、微博舆情和传播特点、事件整体舆情特点，并对政府的舆情控制给出建议。其中关键使用 LSTM 深度学习模型对微博内容进行情感分析，以及使用 LDA 主题概论模型对微博、今日头条、微信公众号和知乎的事件相关数据进行文本聚类与典型意见挖掘。

关键字：热点事件发现 中印边境对峙 LSTM 模型 LDA 主题模型 Word2vec 词向量分析 情感极性分析 文本聚类 典型意见挖掘 微博传播分析 政府舆情控制

目录

一、 事件简介及走势	3
1.1 事件简介	3
1.2 事件走势回顾	4
二、 爬取数据说明	7
2.1 爬取数据说明	7
2.2 具体说明	8
三、 关键技术	12
3.1 基于 Gibbs 采样算法的 LDA 主题分类模型	12
3.1.1 模型介绍	12
3.1.2 模型建立	12
3.1.3 算法求解	12
3.2 基于 LSTM 模型的情感分析	13
3.2.1 数据标注	13
3.2.2 数据处理	14
3.2.3 基于 LSTM 模型的情感分析	14
3.3 其它算法	17
四、 全网舆情分析	17
4.1 全网关注度概况	17
4.1.1 全网热度走势	17
4.1.2 全网文章和评论热度随时间变化	19
4.1.3 全网关注度来源分布比例	20
4.1.4 微博热度随时间变化	21
4.2 全网关注点分析	22
4.2.1 全网关键词词云	22
4.2.2 全网关联词分析	24
五、 微博舆情分析	26
5.1 微博传播分析	26
5.1.1 核心传播用户	26
5.1.2 核心传播用户的影响力排名	27

5.2 微博网民群体画像分析	28
5.2.1 微博关注用户的性别渗透率.....	28
5.2.2 微博关注用户的地域渗透率.....	29
5.2.3 微博海内外关注用户的参与统计.....	30
5.3 微博内容情感分析	30
5.3.1 微博内容的情感极性时间变化图.....	31
5.3.2 微博内容的情感地域分布.....	33
六、 主流媒体平台的舆情特点对比分析.....	35
6.1 主流媒体平台热度(声量大小)的变化趋势对比.....	35
6.2 主流媒体平台的典型意见挖掘——基于 LDA 主题模型.....	39
6.2.1 媒体平台文章的主流观点对比.....	39
6.2.2 媒体平台网民评论的主流观点对比.....	42
七、 舆情总结与启示.....	45
7.1 事件分析	45
7.1.1 印度为何态度如此强硬?	45
7.1.2 印度最终为何选择和解?.....	45
7.1.3 国际各国对此事件的态度是怎样的?	45
7.1.4 中国的态度?	45
7.2 舆情整体特点总结	45
7.3 政府启示	46

一、事件简介及走势

1.1 事件简介

2017 中印军队洞朗对峙事件 [1] 是中华人民共和国与印度共和国的边防部队于 2017 年 6 月 18 日至 8 月 28 日在洞朗地区进行的军事对峙，起因是印度边防人员非法越界进入中国洞朗地区。

2017 年 6 月 16 日，中方在洞朗地区施工时，遭到印军越线阻拦，印方公然派军队越过双方承认的边界线进入中国境内，严重损害中国领土主权。期间中印双方多次交涉无果，在国内和国际上都引发了广泛热议，相关舆情量在 7 月 5 日达到顶峰。8 月 28 日印方将越界人员和设备全部撤回，至此中印边境冲突平息。

1.2 事件走势回顾

下表是利用 LSI(Latent Semantic Index) 文本相似度索引算法、基于微博数据得到的事件走势回顾表，以时间轴顺序完整展示“中印边境事件”从发生到平息的各个节点的典型微博及其相同文章数量。

6 月 18 日	印度边防人员非法越界进入中国洞朗地区，引发中印边境局势持续紧张。【相同文章数：1】
6 月 26 日	就印度边防部队在中印边界锡金段越界时间答记者问。耿爽说，近日，印度边防人员在中印边界锡金段越过边界线进入中方境内，阻挠中国边防部队在洞朗地区的正常活动。中方已采取相应应对措施。【相同文章数：2】
7 月上旬	印陆军参谋长宣称，印度在为“2.5 线战争”做准备。所谓的“2.5 线战争”，即同时应对中国、巴基斯坦和国内的安全威胁，国内算半个。【相同文章数：58】
8 月 11 日	印军进入高等级战备状态。【相同文章数：170】
8 月 16 日 -17 日	中国外交部两度回应“中印问题”。中方敦促印方恪守两国间有关协议协定，切实遵守 1959 年实控线，维护两国边境地区的和平稳定。【相同文章数：104】

8月18日 ● 中国外交部回应“中印问题”。针对印媒报道的“日本驻印度大使表示支持印度在洞朗对峙事件上的立场，认为不应用武力改变现状”和有媒体称“美国呼吁中印进行直接对话”，中方表示洞朗地区不存在领土争议；企图通过非法越界制造事端、改变现状的是印度，不是中国；中方要求印方立即无条件撤回所有非法越界的人员和设备。同日，印报称日本政府支持印度立场，日驻印使馆否认。印度媒体18日报道称，日本驻印度大使平松贤司表示日本政府支持印度立场。然而，日本驻印度大使馆随即否认了相关报道内容。【相同文章数：41】

8月22日 ● 中国网民高度评价“一脚哥”。左侧是解放军，右侧是阿三部队。此时，一个使阴招的阿三士兵占据制高点位置，向解放军猛砸石头。解放军一脚哥绕后突然出现，一脚猛踹，直接把阿三踹飞了出去。同日，中国外交部回应“中印问题”。针对印度内政部长辛格昨天称“印中洞朗对峙很快就会有解决方案，希望中方采取积极行动；印度没有攻击其他国家，也没有野心扩张其边界”的言论，中方表示中国坚定维护和平和坚决捍卫国家主权和领土完整，希望印方立即撤回所有非法越界的人员和装备。针对印媒“21日，印度不具名官员称，近期，印度就中印洞朗对峙问题与俄罗斯进行了讨论，印方关于中国修路单方面改变现状并对印安全带来严重影响”的立场，希望寻求俄方支持”的报道，中方表示，洞朗地区是毫无争议的中国领土。印方以所谓关切中方修路为借口非法越界，理由荒谬、性质恶劣、事实清楚。解决此事的唯一前提和基础是印方立即无条件撤回所有非法越界的人员和设备。【相同文章数：53】

- 8 月 23 日 ● 中国外交部回应“中印问题”。针对“金砖国家领导人第九次会晤将在 10 天后举行。印度总理莫迪是否已经明确确认将出席？中方是否担心由于洞朗对峙事件他可能不来？”的问题，中方申，印军立即无条件地撤出所有非法越界的人员和设备是解决这次事件的前提和基础。针对“尼泊尔总理德乌帕今天开始访问印度。有印度媒体称，印方担心尼泊尔已从中国“一带一路”倡议中获得巨额投资，印度面临“拿出大单抗衡中国”的压力”的问题，中方表示尼泊尔、印度都是中国的要邻国，我们乐见尼印关系健康发展，同时当然也希望印方平和地看待中尼关系的发展。欢迎各国在自愿的基础上，秉持共商共建共享的原则积极参与“一带一路”建设。【相同文章数：49】
- 8 月 24 日 ● 中国驻印度大使馆网站消息，中国驻印度使馆提醒在印度中国公民注意安全。同日，中国外交部回应“中印问题”。针对印媒“近日，印度内政部批准建设一条从马尔斯米克拉到热泉的公路。马尔斯米克拉到班公湖西北端仅 20 公里。修建这段公路可以把偏远地区的印度安全部队与中印边界争议地区连接起来，以确保印度边防部队的优势”的报道，中方表示，印方在西段实控线附近修建具有明显军事目的的基础设施，无助于维护中印边境西段的和平稳定，无助于缓和当前两国边境地区的局势。【相同文章数：34】
- 8 月 25 日 ● 中印班公湖冲突后，印度在数公里外举行军演。【相同文章数：20】
- 8 月 28 日 ● 14 时 30 分许，印方已将越界人员和设备全部撤回到边界印方一侧，中方现场人员对此进行了确认。中国军队将保持高度戒备，坚定捍卫国家领土主权。【相同文章数：154】

二、爬取数据说明

2.1 爬取数据说明

我们爬取了 2017 年 6 月 18 日至 8 月 28 日期间新浪微博、知乎、微信公众号、今日头条这四大主流社交媒体平台上与“中印边境事件”相关的数据，数据分类及条数见下表：

数据分类	数据条数	数据分类	数据条数
微博	63837	知乎评论	20033
微博评论	535050	知乎用户	30667
微博转发	322383	微信文章	557
微博用户	459301	头条文章	161
知乎提问	100	头条评论	43118
知乎回答	2439	头条回复	5171
知乎专栏	38	头条用户	39885

2.2 具体说明

表 1 微博内容

字段	说明
id	微博 id
url	微博 url
content	微博内容
time	微博发布时间， 用 UTC 时间戳表示
author	发表这篇微博的用户的 id
comments_count	微博评论数
reposts_count	微博转发
likes_count	微博点赞数

表 3 微博评论

字段	说明
id	评论 id
content	评论内容
user	发表评论的用户的 id
time	评论发表时间，用 UTC 时间戳表示
like	点赞数
target_type	评论对象的类型，一种是“weibo_post”， 表示针对原微博的评论， 一种是“weibo_comment”， 表示回复某一个评论的评论
target_id	如果 target_type 是“weibo_post”， 则表示微博 id，如果 target_type 是 “weibo_comment”，则表示评论 id

表 2 微博转发

字段	说明
id	转发 id
content	转发时发表的内容
like	点赞数
user	转发微博的用户的 id
time	转发时间，用 UTC 时间戳表示
origin_post	原微博的 id

表 4 微博用户

字段	说明
id	用户 id
url	用户主页 url
name	用户昵称
gender	用户性别，1 ~ 男， 0 ~ 女，-1 ~ 未知
location	用户所在地
description	用户简介
verified_reason	认证信息，为空表示没有认证
follow_count	关注人数
follower_count	粉丝数
status_count	发表的微博数

表 5 知乎用户

字段	说明	字段	说明
id	用户 id	articles_count	文章数
url_token	可以唯一标识用户的一个字符串	question_count	提问数
url	用户主页 url	columns_count	专栏数
name	用户名	logs_count	参与公共编辑次数
gender	用户性别, 1 ~ 男, 0 ~ 女, -1 ~ 未知	voteup_count	用户获得的赞同数
locations	用户所在地	thanked_count	用户获得的感谢次数
educations	用户教育经历	follower_cont	粉丝数
employments	用户职业经历	following_count	关注的人数
headline	用户个性签名	favorite_count	收藏数
answer_count	回答数	favorited_count	被收藏数

表 6 知乎回答

字段	说明
id	用户 id
url	回答 url
author	回答者的 url_token
content	回答内容
time	回答时间, 用 UTC 时间戳表示
upvote	赞同数
target_type	回答对象类型, “question” 表示对问题的回答, “answer” 表示对某个回答的回答
target_id	如果 target_type 是 “question” 则表示问题 id, 如果 target_type 是 “answer” 则表示回答 id

表 7 知乎提问

字段	说明
id	问题 id
author	提问者的 url_token
title	问题标题
content	问题详细
time	提问时间，用 UTC 时间戳表示

表 8 知乎专栏

字段	说明
id	专栏 id
author	作者的 url_token
title	专栏标题
content	专栏内容
url	专栏 url
time	专栏发表时间，用 UTC 时间戳表示
like	点赞数

表 9 头条用户

字段	说明
id	用户 id
name	用户名
followers_count	粉丝数
followings_count	关注人数
verified_count	认证信息

表 10 头条回复

字段	说明
id	评论 id
reply_id	回复 id
text	回复内容
name	发表回复的用户名
user_id	发表回复的用户 id
digg_count	点赞数
create_time	回复时间

表 11 头条文章

字段	说明
id	文章 id
media_name	媒体名称
title	文章标题
content	文章内容
datetime	发表时间
comments_count	评论数
digg_count	点赞数
bury_count	踩数
favorite_count	收藏数

表 12 头条评论

字段	说明
id	评论 id
article_id	文章 id
user_id	发表评论的用户 id
user_name	发表评论的用户名
text	评论内容
score	头条提供的影响力分数
digg_count	点赞数
bury_count	踩数
reply_count	回复数
create_time	评论时间

**表 13 微信公众号
文章**

字段	说明
id	文章 id
author	公众号名称
url	文章 url
title	文章标题
content	文章内容
time	发表时间

三、关键技术

3.1 基于 Gibbs 采样算法的 LDA 主题分类模型

3.1.1 模型介绍

LDA 是一种典型的无监督、基于统计学习的词袋模型，一种文档主题生成模型，也称为一个三层贝叶斯概率模型。生成模型即我们认为一篇文章的每个词都是通过“以一定概率选择了某个主题，并从这个主题中以一定概率选择某个词语”。

LDA 模型包含词、主题和文档三层结构。即它认为一篇文档是由一组词构成的一个集合，词与词之间没有顺序以及先后的关系；一篇文档可以包含多个主题，文档中每一个词都由其中的一个主题生成。主题模型通过分析文本中的词来发现文档中的主题、主题之间的联系方式和主题的发展，通过主题模型可以使我们组织和总结无法人工标注的海量电子文档。

LDA 是一种**非监督机器学习**技术，本报告利用该模型识别微博、微信、今日头条和知乎四大媒体平台上的文章和评论中**潜藏的主题信息，并将文章与评论进行分类，并分别用若干个关键词来描述每个主题，从中提取出舆情的主流观点。**

3.1.2 模型建立

LDA 模型基于贝叶斯网络结构，在 LDA 模型中一篇文档生成的方式如下：

- 从狄利克雷分布中取样生成文档 i 的主题分布 θ_i
- 从主题的多项式分布中取样生成文档 i 第 j 个词的主题 $z_{i,j}$
- 从狄利克雷分布中取样生成主题的词语分布 $\phi_{i,j}$
- 从词语的多项式分布 $\phi_{i,j}$ 中采样最终生成词语 $w_{i,j}$

3.1.3 算法求解

整个模型中所有可见变量以及隐藏变量的联合分布是：

$$P(w_i, z_i, \theta_i, \Phi | \alpha, \beta) = \prod_{j=1}^N P(\theta_i | \alpha) P(z_{i,j} | \theta_i) P(\Phi | \beta) P(w_{i,j} | \theta_{z_{i,j}})$$

最终一篇文档的单词分布的最大似然估计可以通过将上式的以及进行积分和对进行求和得到：

$$P(w_i | \alpha, \beta) = \iint \sum_{z_i} P(w_i, z_i, \theta_i, \Phi | \alpha, \beta)$$

根据上述 $p(w_{i,j} | \cdot)$ 的最大似然估计，通过吉布斯采样方法估计出模型中的参数，具体过程如下：

- 首先对所有文档中的词遍历一遍，为其都随机分配一个主题， $z_{m,n} = k \sim Mult \frac{1}{K}$ ，其中 m 表示第 m 篇文档，n 表示文档中的第 n 个词，k 表示主题，K 表示主题的总数； $n_m^{(k)} + 1, n_m + 1, n_k^{(t)} + 1, n_k + 1$ 分别表示 m 文档中 k 主题出现的次数，m 文档中主题数量的和，k 主题对应的 t 词的次数，k 主题对应的总次数
- 对下述操作进行重复迭代。
- 对所有文档中的所有词进行遍历，假如当前文档 m 的词 t 对应主题为 k，则 $n_m^{(k)} - 1, n_m - 1, n_k^{(t)} - 1, n_k - 1$ ，即先拿出当前词，之后根据 LDA 中 topic sample 的概率分布采样出新主题，在对应的 $n_m^{(k)}, n_m, n_k^{(t)}, n_k$ 上分别 +1。

$$P(z_i = k | z_{-i}, w) \propto (n_{k,-i}^{(t)} + \beta_t)(n_{m,-i}^{(k)} + \alpha_k) / \sum_{t=1}^V n_{k,-i}^{(t)} + \beta_t$$

- 迭代完成后输出主题-词参数矩阵和文档-主题矩阵

主题 k 中词 t 的概率分布： $\phi_{k,t} = (n_k^{(t)} + \beta_t) / (n_k + \beta_t)$

文档 m 中主题 k 的概率分布： $\theta_{m,k} = (n_m^{(k)} + \alpha_k) / (n_m + \alpha_k)$

3.2 基于 LSTM 模型的情感分析

传统情感分析方法需要提前建立好情感词典，而情感词典的建立不仅需要数据专家，而且需要语言学家的参与，这就带来了很大的问题。并且情感词典只是具有普适性，针对于特定问题并没有很好的效果。机器学习方法例如 SVM，则不需要建立情感词典，但是它在分析句子时会把各个词语当成独立的单元去处理，并没有上下文关系，显然我们在语言中是需要上下文的。语言数据属于时间序列数据，每个词语的出现都依赖于它的前一个词语和后一个词语。由于这种依赖的存在，我们使用 LSTM（长短期记忆网络）来进行情感分析。

LSTM(长短期记忆神经网络) 是 RNN 网络的拓展，具有时序建模的作用。LSTM 通过输入门，遗忘门，输出门保持和更新细胞状态，可以判断哪些信息是有用的，哪些是没用的，并把有用的信息在 LSTM 中保存。

3.2.1 数据标注

由于国内暂时无法找到此类政治/军事事件的情感标注数据集，并且尝试其他领域情感标注数据集之后发现效果并不是很好，与预期相差较大，故我们决定进行对爬取的微博数据构建人工标注数据集。由于微博文章存在多种多样的情感较难使用语言描述，我们将其划分为 4 个情感等级，分别为 0, 1, 2, 3，从 0 到 3 情感依次从平和到愤慨。我们人工标注了 6000 条左右的数据。

3.2.2 数据处理

微博爬取到的数据有很大一部分与中印对峙无关，比如我们爬取数据时使用的“中印”关键词会爬到如“莞中印象”等无关数据。首先，我们采用标注好的数据中词频较多的关键词来进行粗筛，比如“中印”、“对峙”、“边境”、“战争”、“洞朗”等。之后我们采用 TF-IDF 进行词频统计，并进行停用词过滤，将与标注数据相似的文章提取出来。大概可以筛掉一半以上的数据，大大减小了噪声和训练成本。

3.2.3 基于 LSTM 模型的情感分析

本报告的情感分析框架如下：

1. 数据分析

通过 1 可知，中印对峙微博评论情感较为强烈。

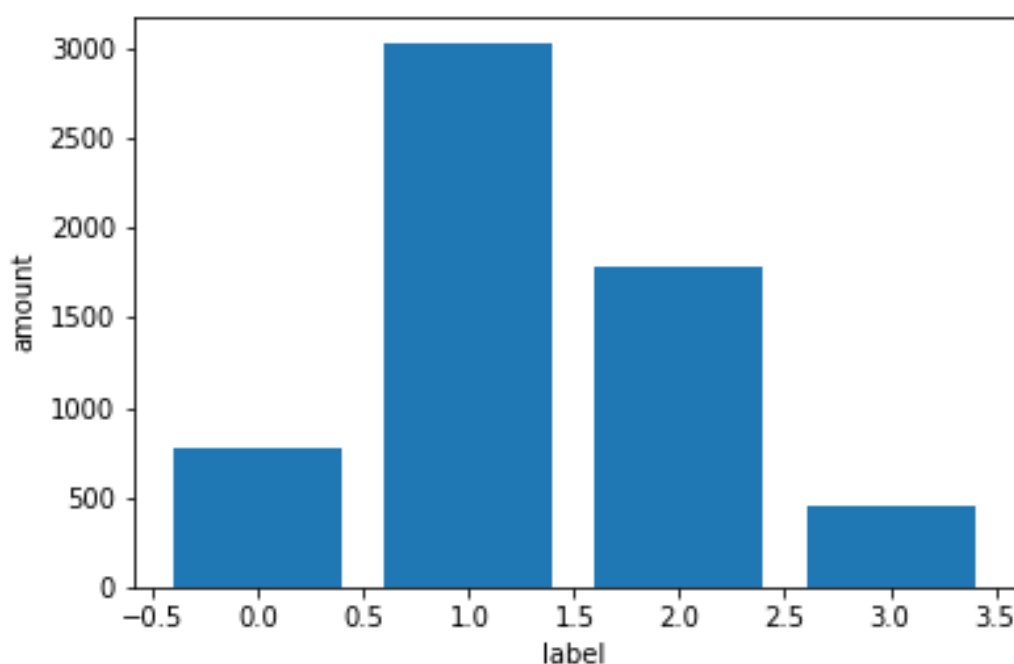


图 1 可视化人工标注的微博情感数据

2. 模型选择

在文本向量化部分，我们选择了 word2vec 模型。情感是和上下文相关的，word2vec 模型可以很好的根据上下文语境推断出每个词的词向量，如果两个词在上下文的语境中，可以被互相替换，那么这两个词的距离就非常近。

在情感分类模型部分，我们采用了 LSTM 模型来训练。LSTM(长短期记忆神经网络)是 RNN 网络的拓展，具有时序建模的作用。LSTM 通过输入门，遗忘门，输出门保

持和更新细胞状态，可以判断哪些信息是有用的，哪些是没用的，并把有用的信息在 LSTM 中保存。LSTM 单元如2所示：

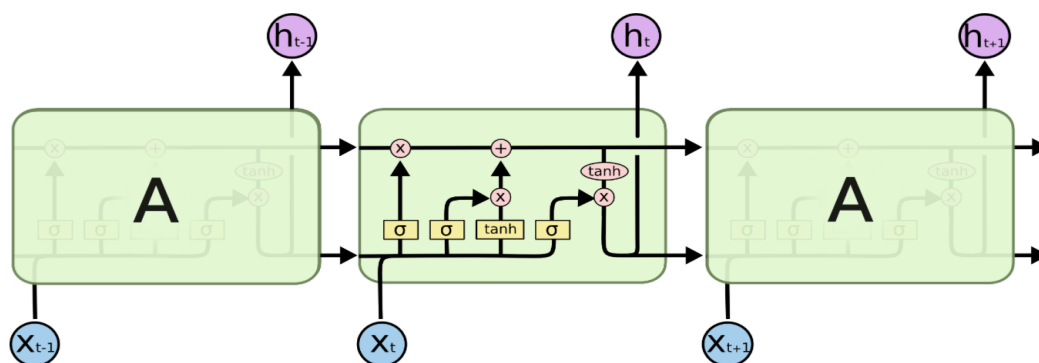


图 2 LSTM 单元

3. 模型训练与结果

我们采用 jieba 分词，对微博训练样本进行分割，并采用 Google 开源的 Word2Vec 训练，将分词词语表示为词向量，之后将每个微博样本向量化作为 LSTM 网络的输入。我们统计了每个微博样本的单词数（图3）。

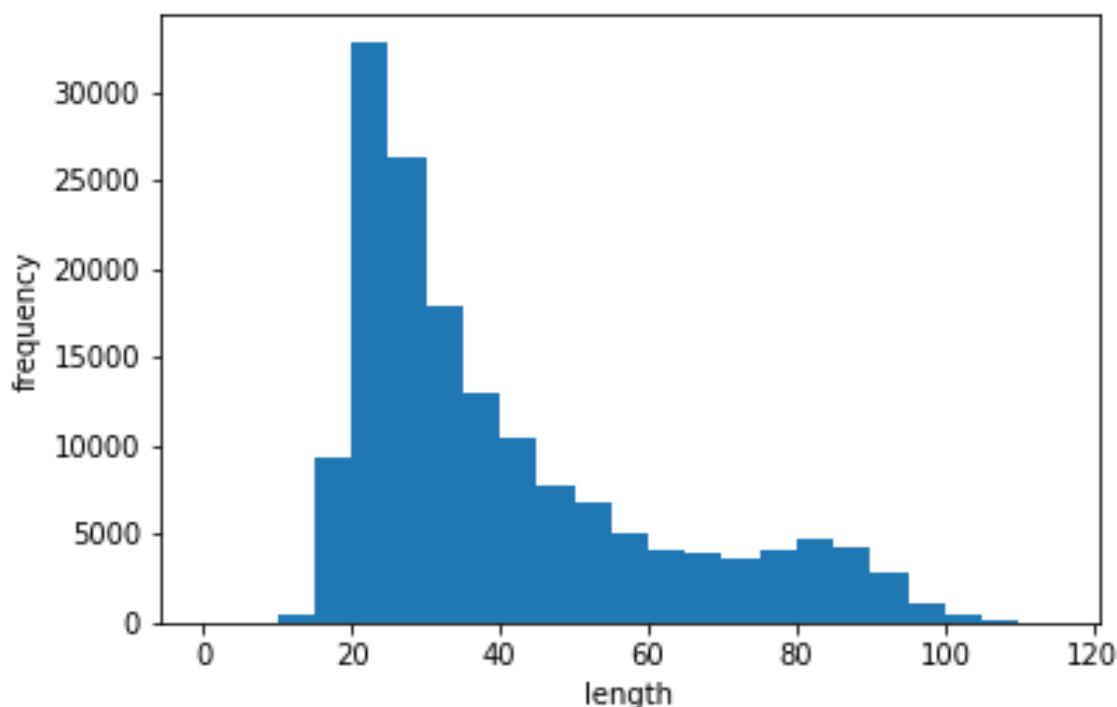


图 3 微博样本的单词数

根据直方图，我们选择将句子最大长度设置为 85，之后我们将数据转化为索引矩

阵，得到我们的输入数据，转换流程如图4：

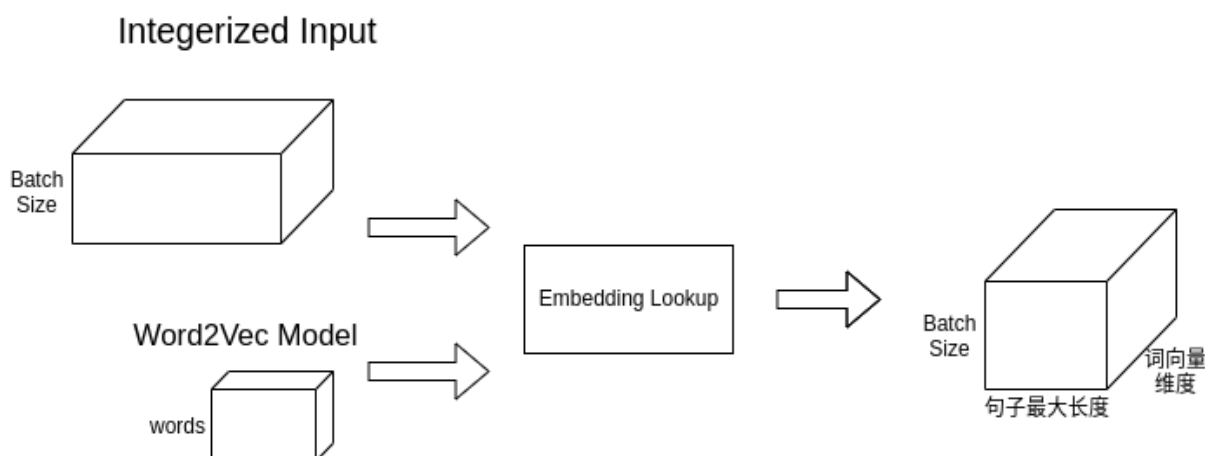


图 4 LSTM 单元

首先，我们尝试进行二分类情感分析：即将标注数据的 0,1 归为一类，2,3 归为一类，放入图5的神经网络中进行训练。

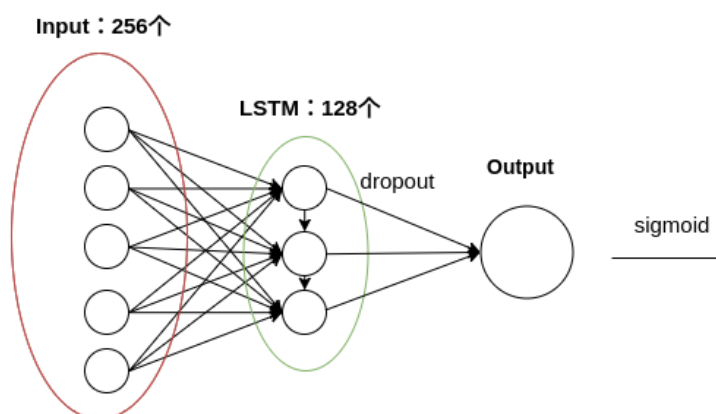


图 5 二分类的 LSTM 神经网络模型

单个 GTX960 显卡大概需要 5 分钟左右时间完成训练，训练准确率约为 99%，测试集准确率约为 90%

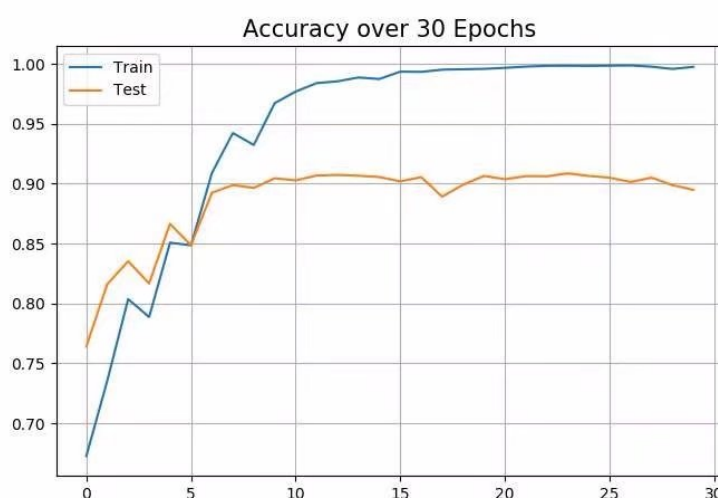


图 6 二分类的 LSTM 神经网络模型

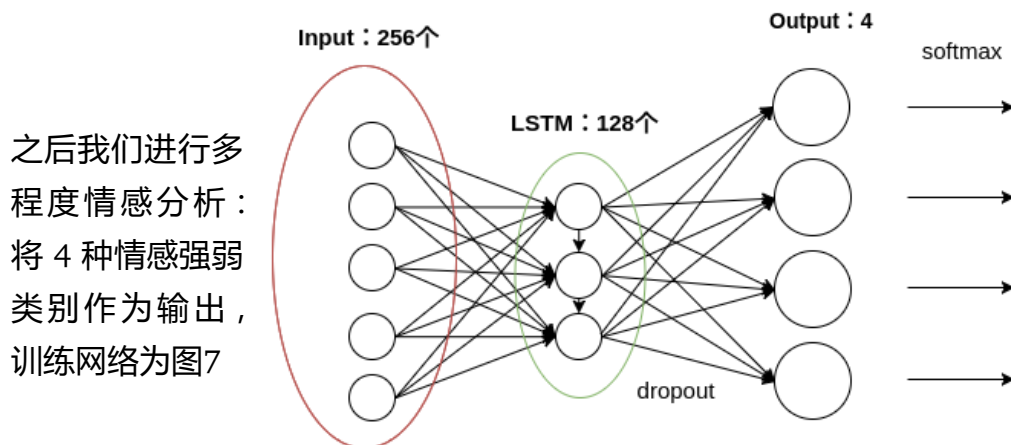


图 7 多分类的 LSTM 神经网络模型

我们最终得到**训练集准确率在 83% 左右，测试集准确率在 77% 左右**，效果不算很好。但是通过比对，测试集真实值与预测值在情感程度划分上差别较小，可能与数据标注的误差有关系。我们最终选择了使用多分类的情感模型。

3.3 其它算法

分析对象	算法
关联词分析	Word2vec 词向量分析
全网事件走势回顾 (相同主题文章数量分析)	LSI(Latent Semantic Index) 文本相似度索引算法

四、全网舆情分析

4.1 全网关注度概况

4.1.1 全网热度走势

图8是“中印边境事件”在整个周期内的全网热度走势图。自事件发生起，其**全网热度**总体呈波动上升趋势，出现十余个大小峰值，且在 8 月 11 日达到最高峰。对比本报告 1.2 节中的「事件走势回顾图」，易发现**全网热度峰值时间点与此事件的每一步进展的时间点紧密吻合**，说明此事件的热度增减与事件进展紧密相关。每当此事件相关新闻出现都会引发网民的舆论热议，由此可见**网民对此类涉及国家安全和领土完整的政治问题是长期密切关注和参与的**。

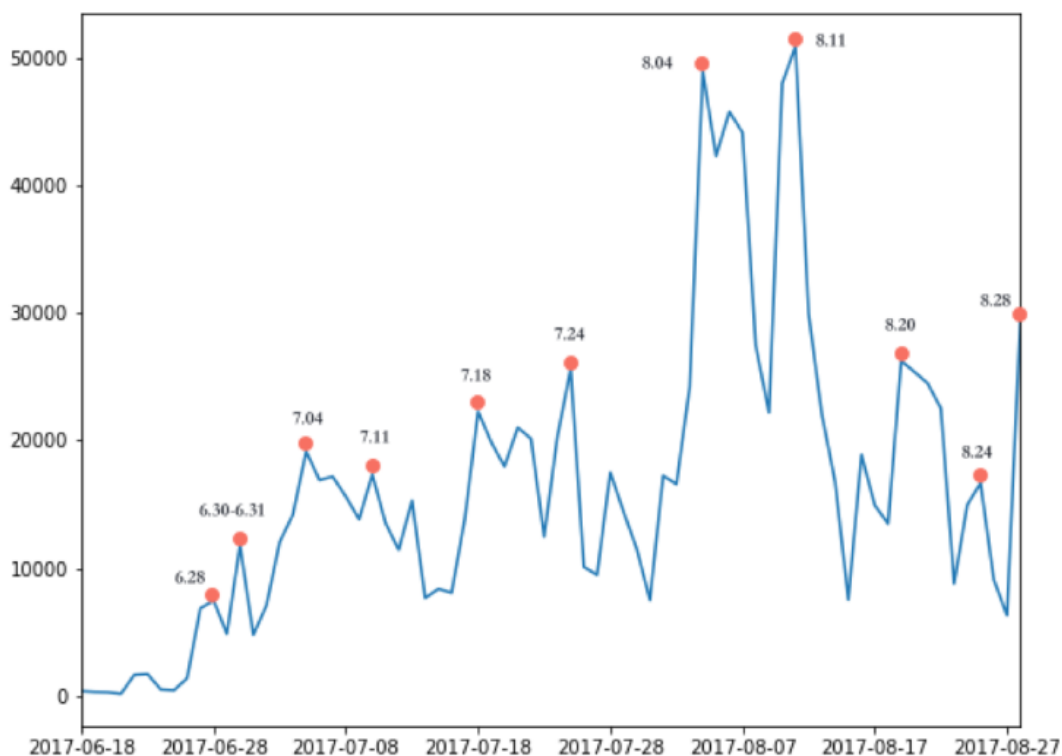


图 8 全网热度走势

8月4日热度达次高峰，是由于“8月3日上午至4日凌晨，新华社、解放军报、外交部、国防部、中国驻印度大使馆以及人民日报这中国6个国家部委和机构先后就印方越界事件发声，披露印方非法越界的性质，并强调中国将采取一切必要措施维护自己的正当合法权益”。由此可知，中方对待此事件的强硬态度引起网民巨大的舆论热议，可见网民的爱国之心被广泛激发。说明**国家在面对国家领土和主权问题时态度强硬、立场坚定，能广泛振奋人心、给民众以信心和力量，激发民众的爱国之心。**

8月11日热度达最高峰，是由于“印军进入高等级战备状态”的新闻爆出，这似乎是印度战争部署的预警。由此可见，**中印边境事件紧张局势的进一步升级刺激到网民的敏感神经，引发了整个周期中最大的舆论热议，这体现网民在国家安全可能受战争威胁下的群体不安和“舆论骚动”。**

8月28日热度达第二次高峰，是由于“印方将越界人员和设备全部撤回，中印边境冲突平息”。由此可见，**政治事件的和平解决也会引发大量舆论关注。**

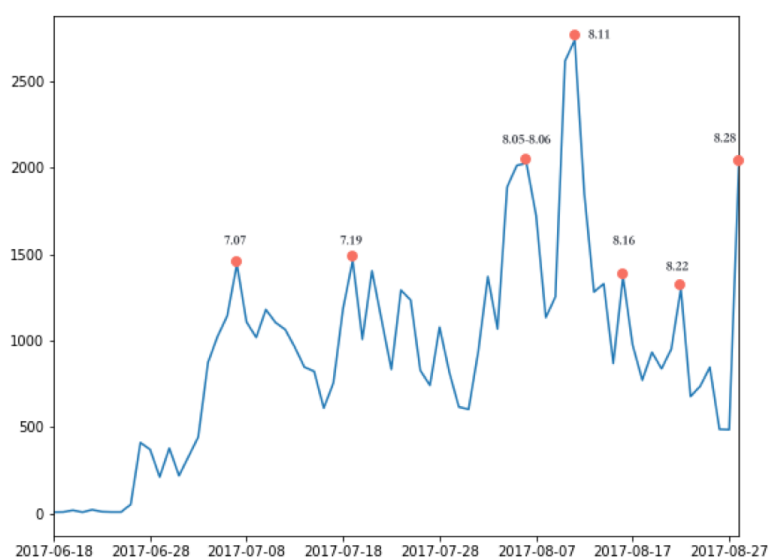


图 9 全网文章热度随时间变化

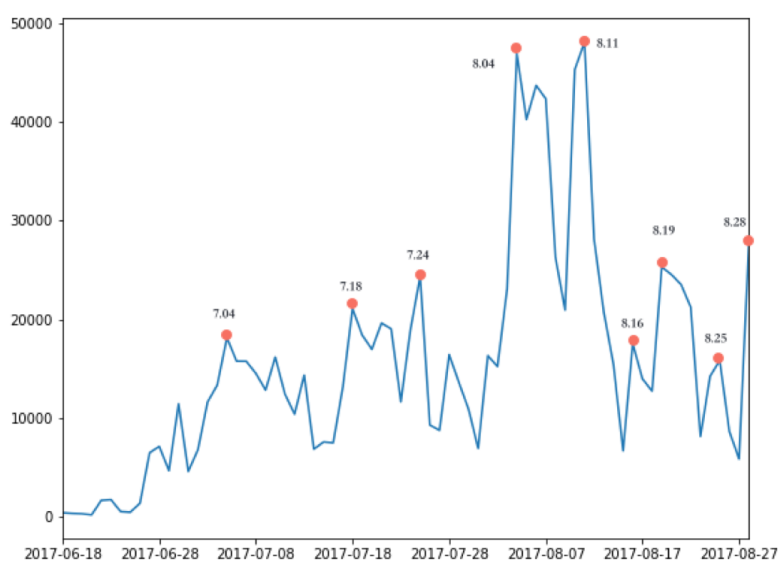


图 10 全网文章热度随时间变化

4.1.2 全网文章和评论热度随时间变化

对比「全网文章热度随时间变化」「全网评论热度随时间变化」与「全网热度走势」三图，易发现三图的整体趋势十分相似，但全网评论的热度峰值时间点比全网文章的更接近于全网热度峰值时间点，这可能一是由于全网评论主要来源于微博，而微博相

较于其他媒体平台及时性更高，二是相较于 UGC 特征强的微博，**新闻文章从采编到发布到被阅读有较长的时间差，故导致全网文章的热度峰值普遍略迟。**

4.1.3 全网关注度来源分布比例

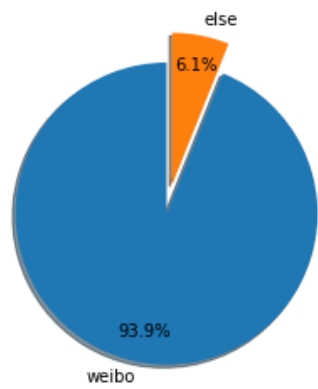


图 11 全网关注度来源分布比例

图11显示在全网对此事件的关注度来源中微博占比高达 93.9%，而来自其他媒体平台的信息量只占 6.1%，可见基于微博庞大的用户量 (能产生庞大的 UGC 内容) 和不断凸显的社交媒体平台属性，微博至少在新闻舆论方面保持绝对影响力。

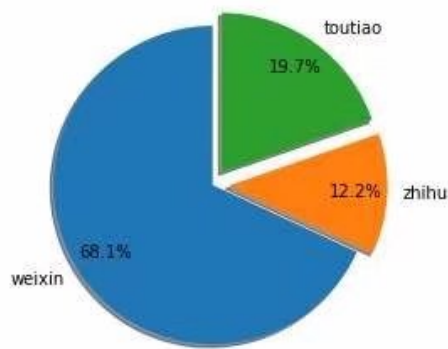


图 12 除微博外其他媒体平台的报道量比例

由图12可看出，除微博外，微信和今日头条对此事件的信息量贡献最多。

微信的信息量大，一是基于微信庞大的公众号内容生产体系，二是得益于微信在今年新增的「搜索资讯」功能，使用户能够主动搜索关键词，浏览热点资讯、好友关注的资讯以及自己感兴趣的文章。这体现微信在社会化媒体内容生产方面的重要性日渐突出。

今日头条的信息量大，一是由于其强大的数据挖掘能力源源不断为平台输送内容，二是由于其庞大的用户基量产生巨大的资讯阅读需求和评论回复信息量 (《2016 移动资讯行业细分报告》[2] 显示:截止 2016 年 12 月底，今日头条用户量达 7.0 亿，日活跃用户量达 7800 万)。

4.1.4 微博热度随时间变化

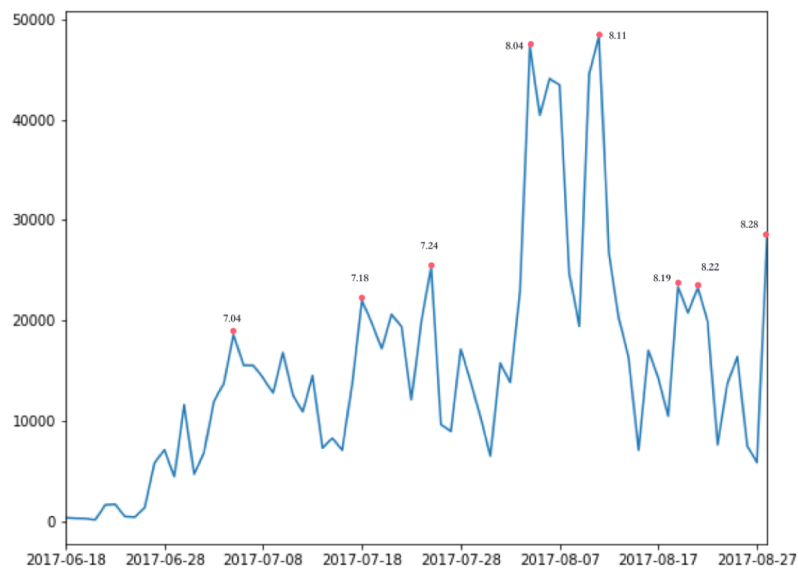


图 13 微博热度随时间变化

对比「微博热度随时间变化」与「全网热度变化」，易发现二者走势极为相似，这是由于此事件的微博信息量在全网信息量中占比极大，反映**微博平台的强大舆论影响力**。

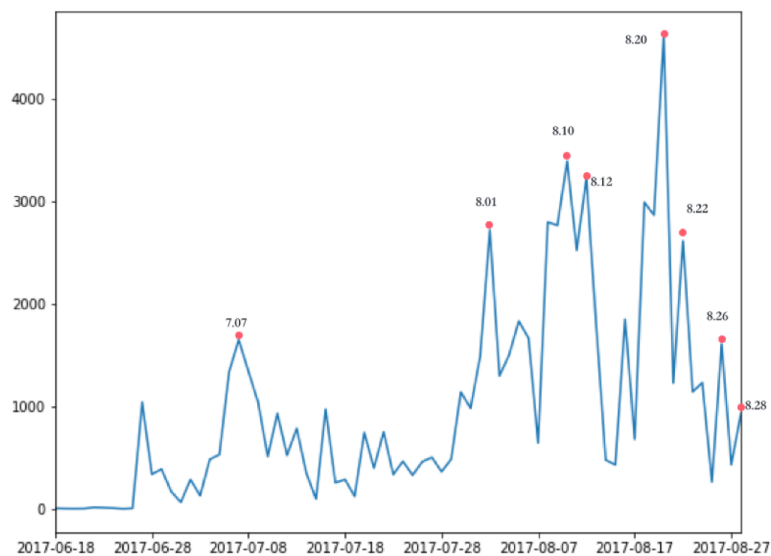


图 14 除微博外媒体热度随时间变化

对比「除微博外媒体热度随时间变化」与「微博热度随时间变化」可发现：二者的

旋”、“双边及多边联合军演”等活动，反映美国可能存在拉拢印度以抗衡中国的意图以及印度可能存在寻求美国合作/援助的意图。**若美印联手，则必将加剧中印边境冲突的紧张局势。**

“日本”：事件期间**据部分印媒报道日本政府支持印度立场。但日本驻印度大使馆随即否认了相关报道内容。因历史、领土争端等因素，中日关系素来易引起中国网民的关注。**而此次中印边境事件亦涉及领土主权问题，故易让人联想到日本。

“越南”：中印边境争端持续之际，**印度和同样与中国存在领土争端的越南越走越近**（越南总理访问印度）。很难不让网民猜测其用意。

“不丹”：因为**印度称自己在这场对峙中代表的是不丹**。故在“中印对峙事件”中不丹的态度可能影响中、印、不丹三国关系。

“台湾”：因**“中印对峙波及台湾，台两学生被印拒绝入境”**新闻，触动中国网民关于台湾问题这一**“历史遗留问题”**的敏感神经。

“经济”：可能是**担心中印冲突会影响两国经济合作**，以及若发生战争会影响中国的经济建设和国家整体发展。

以上关键词反映此事件牵涉的相关主体和网民关注的焦点。可以看出，**网民对事件背后的国家间博弈以及事件可能对本国经济造成的影响尤为关注。**

4.2.2 全网关联词分析

下面是分别对核心词“中印”和“中国”的关联词分析¹。

中印		
核心词	关联词	关联度
中印	持续时间	41.3%
	中朝	38.8%
	退却	37.2%
	持久战	37.0%
	两军	36.7%
	现状	35.7%
	长期	33.9%
	增兵	33.8%
	军队	33.6%
	阅兵	33.1%
	洞朗	32.8%
	转折点	32.6%
	摩擦	32.1%
	晚打	32.0%
	天赐良机	31.6%

图 16 核心词“中印”的关联词分析

图16显示与核心词“中印”相关度最高的词语为：持续时间 (41.3%)、中朝 (38.8%)、退却 (37.2%)、持久战 (37.0%)。反映**网民舆论较为关注中印对峙已经和将要持续的时间，且网民已有打“持久战”的猜测或心理准备。**

其中“中朝”与“中印”相关度高是由于**中印对峙期间中国突然增兵朝鲜边境，故引发网民舆论热议讨论。**中国突然增兵朝鲜边境行为背后的含义其一可能是警告美国，显示出中国军事维护半岛“无战、无乱、无核”的原则和决心；其二可能是防止万一中印开战美军搞突然袭击。

¹关联分析：与主题相关的关联词分析，即与核心词共现频率（相关度）最高的词。关键词 A 和关键词 B 的相关度：同时包含关键词 A 和关键词 B 的文章的阅读数/包含关键词 A 的文章的阅读数

中国		
核心词	关联词	关联度
中国	印度	45.2%
	扩张	45.1%
	中国政府	39.2%
	武力	38.7%
	出兵	38.1%
	威胁	38.0%
	入侵者	37.6%
	安全隐患	37.4%
	实力	37.2%
	对手	37.0%
	动武	36.1%
	让步	35.5%
	单方面	35.4%
	修筑道路	34.8%
	撤出	34.6%

图 17 核心词“中国”的关联词分析

图17显示与核心词“中国”相关度最高的词语为：印度 (45.2%)、扩张 (45.1%)、中国政府 (39.2%)、武力 (38.7%)、出兵 (38.1%)、威胁 (38.0%)、入侵者 (37.6%)、安全隐患 (37.4%)。反映媒体报道和网民舆论十分关注中印关系，认为“中印边境事件”反映作为“入侵者”的印度扩张边境的野心，以及担忧其可能对我国西南边境造成安全隐患，并可能进行了是否使用武力解决冲突的讨论。

五、 微博舆情分析

5.1 微博传播分析

5.1.1 核心传播用户

机构	头条新闻
	紫光阁
	环球时报
	人民日报
	环球网
	中国之声
	央视新闻
	澎湃新闻
媒体人	防务君（军事视频自媒体）
	二战那些事（军事视频自媒体）
用户	局座召忠
	文艺泽
	打奥特曼de小怪兽
	陌曦-Autumn
	风之子FC

图 18 核心传播用户：机构/媒体人/网民

核心传播机构中头条新闻的核心传播作用可能由于其庞大的用户基量，而紫光阁、环球时报、人民日报等机构则可能由于其官媒属性强，以及因有国家颁发的新闻采写资格证而掌握稀少的新闻采写资源。两个核心传播媒体人则主要因其军事视频自媒体的内容属性。核心传播用户中局座召忠也是军事报道与分析的自媒体。由此可见，**在这类政治军事事件中的核心传播用户的共同特点是擅长军事属性内容创作或是有官方属性。**

5.1.2 核心传播用户的影响力排名

表 14 影响力排名表

排名	1	2	3	4	5
媒体	局座召忠	头条新闻	人民日报	环球时报	防务君

从表14可以看出，局座召忠作为一个专注于军事内容创作的自媒体，在“中印边境事件”中的传播影响力最强。

5.2 微博网民群体画像分析

针对关注和参与“中印边境事件”的微博网民的用户画像，本报告从性别、国内地域分布、海内外参与比例三个数据维度进行分析，分析用户的不同群体特征对他们在舆论事件中的关注和参与度的影响。

5.2.1 微博关注用户的性别渗透率

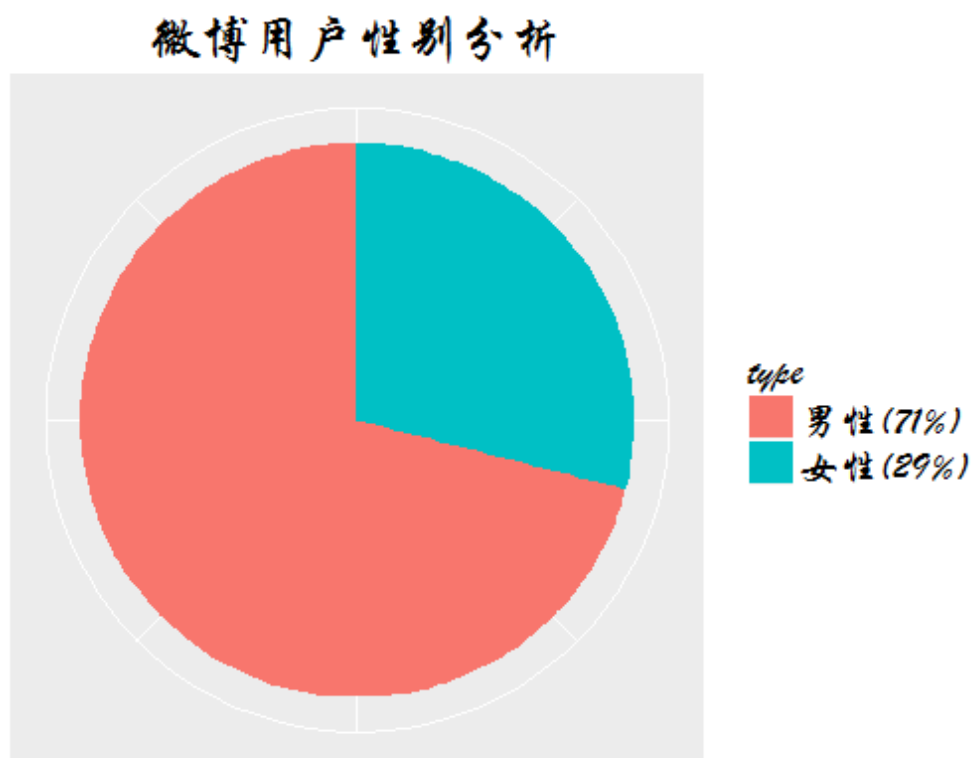


图 19 微博关注用户的性别渗透率^{1 2}

图19数据显示，微博关注此事件的男性网民与女性网民比例是 71%：29%，可看出：**男性网民对此事件的关注度远远高于女性网民**。而根据微博数据中心发布的《2016 微博用户发展报告》，2016 年微博活跃用户中男性用户以 55.5%：44.5% 略高于女性用户，可知男性用户比例本身只是略高于女性用户。导致关注此事件的男性网民比例超出寻常地高于女性用户，可能是由于**不同性别网民的兴趣差异，因为通常来说男性网民对“中印边境事件”这类国际、军事、时政、历史等性质的话题的关注度要远高于女性网民**。

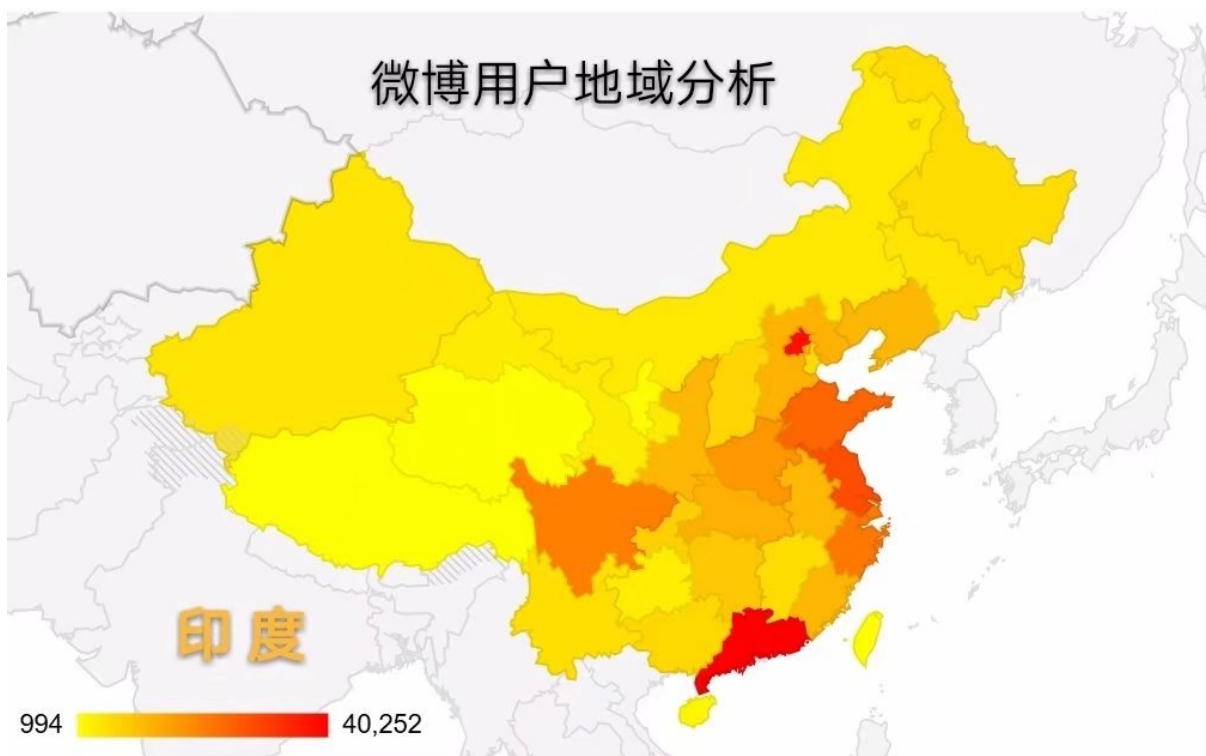


图 20 微博用户地域分析³

5.2.2 微博关注用户的地域渗透率

由图20可看出，**北京和广东的微博网民对话题的关注度最高，其次是江苏、上海、山东、河南、云南等。**

北京经济发达，人口众多，网民数量多，且作为我国政治中心，**网民对政治事件的关注度和敏感度较高**，故北京微博网民对“中印边境事件”的关注度最高。

广东省是**中国的经济大省和改革开放前沿阵地，与印度的经贸合作一直十分紧密**（去年粤印贸易总额达 144 亿美元，占中印贸易总额的 1/5）；对印度来说广东的经济地位也十分突出，广东省与印度的古吉拉特邦已缔结友好省邦协议；再加上因互通贸易、人员流通等因素，广东省与印度文化上也有很大的渊源；且随中印双方更加紧密的发展伙伴关系的构建和“一带一路”建设的深入推进，广东和印度的经贸合作在未来会有更大的发展。所以**担心“中印边境事件”会影响广东与印度的经济合作**，可能是导致广东的微博网民对此事件的关注度最高的重要因素之一。

云南省作为中国通往南亚、东南亚开放的前沿省份，是**中国面向印度的陆路窗口**，与印度有地缘政治下的区域经济合作优势，且改革开放以来云南与印度的经济合作总体处于较快发展状态；以及在“一带一路”倡议和建设面向南亚、东南亚辐射中心的利好条件下，云南与印度的经济合作发展前景广阔。而且地域毗邻下若中印发生军事冲

¹渗透率：表示用户对特定事件关注度的比例。

²性别渗透率：某性别用户对关键词的关注度 / 该性别用户总关注度

³地域渗透率：某省份或城市用户对包含特定关键词文章的关注度 / 该省份或城市产生的总关注度

突，会首先影响到云南。故可以推测，**担心中印边境冲突可能会影响云南的安定以及其与印度的经济合作**，可能是导致云南的微博网民的对此事件关注度次高的重要因素之一。

据相关报道可知，**河南为兵员大省**，所以其微博网民对此事有较高的关注度。

5.2.3 微博海内外关注用户的参与统计

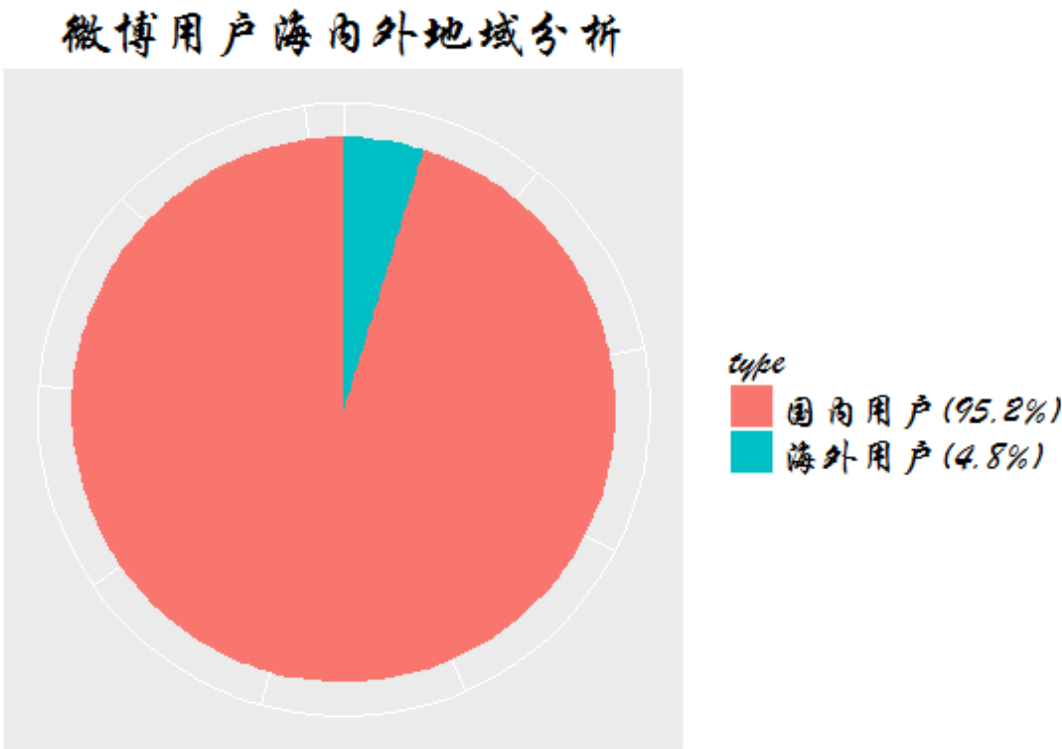


图 21 微博海内外关注用户的参与统计

根据微博数据中心发布的《2016 微博用户发展报告》[3]，港澳台及海外的微博用户占全体微博用户的 2%。而上图显示，微博海内外关注用户中海外用户占比 4.8%，高于 2%。这说明海外微博用户对此事件较为关注。这可能是由于**“中印边境事件”**的国际属性，以及其事件本身在国际上就引起了较大的舆论争议。

5.3 微博内容情感分析

本报告利用 LSTM 模型对事件相关的微博内容进行情感分析，将情感强度划分为四个情感等级，分别为 0，1，2，3，从 0 到 3 情感依次从平和到愤慨。

5.3.1 微博内容的情感极性时间变化图

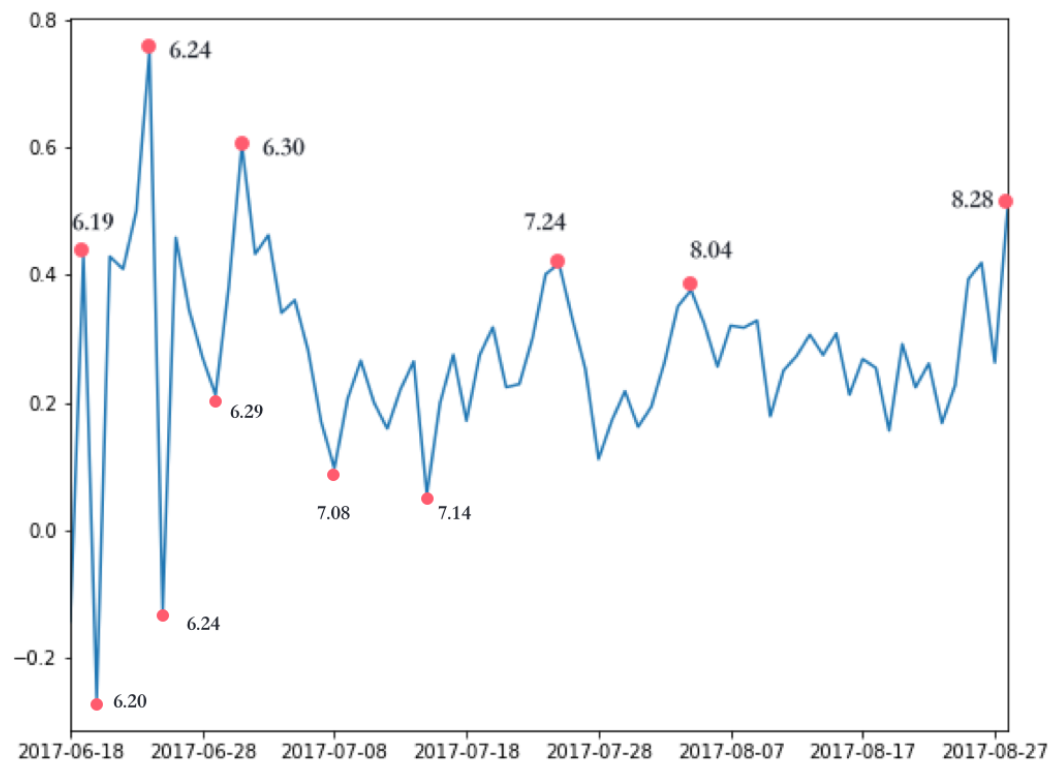


图 22 微博文章的情感极性时间变化图

对比图22与本报告 1.1 节的「事件走势回顾图」，易发现**微博文章的情感极性峰值和谷值时间点与整个事件进展的关键时间节点十分吻合**。分析上图可知：

- 事件进展的最新新闻 (无论正面还是负面新闻) 极易调动微博网民情绪
- 事件周期内微博网民绝大部分时候处于愤慨情绪状态
- 此事件中负面新闻 (如：6 月 26 日前印军阻挠中国边防部队在洞朗地区的正常活动；6 月 29 日外交部展示印度边防人员非法越过中印边境进入中国领土照片) 的爆出一层层刺激网民的敏感神经，接连不断引发网民的愤慨情绪峰值
- 正面新闻的报道则会引发网民平和情绪谷值

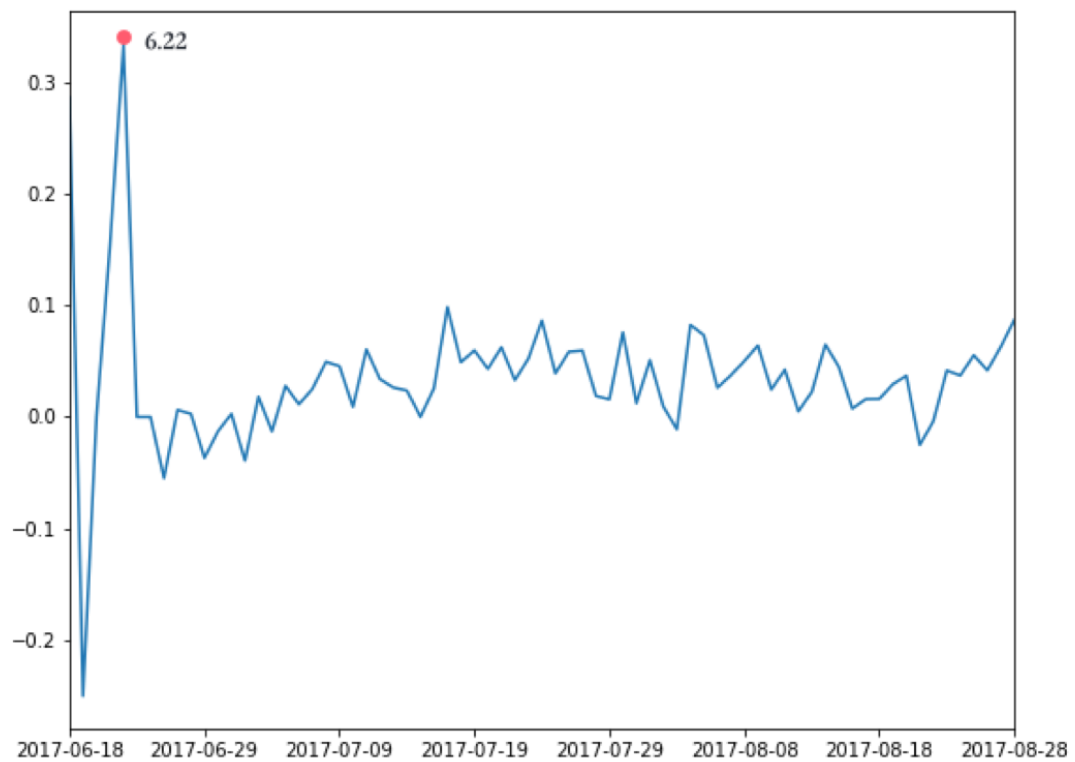


图 23 微博评论的情感极性时间变化图

对比图22和23，易发现微博评论的情感极性变化走势与微博文章十分相似，但波动幅度比微博文章小，峰值谷值没那么明显，且绝大部分时候处于愤慨情绪状态。

5.3.2 微博内容的情感地域分布

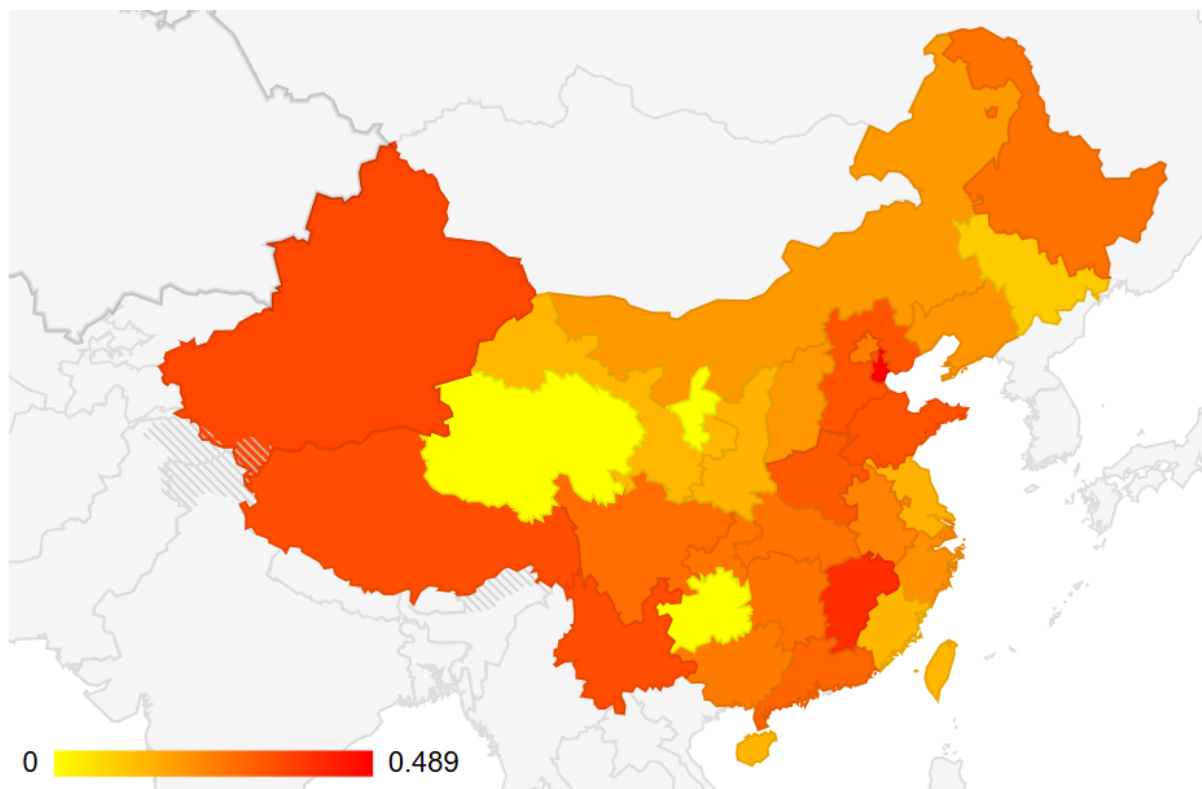


图 24 微博文章的情感地域分布

图24显示“中印边境事件”下微博文章的情感强度的地域分布。由图可看出，新疆网民的微博文章的情感强度最高，表明最愤慨，其次是西藏、江西、云南等。

云南：由于印度、云南地域毗邻，故若中印发生军事冲突可能会首先影响到云南。故可以推测，担心中印边境冲突影响云南安定可能是云南微博网民较为愤慨的原因。

新疆：与此次事件争议的位于西藏的锡金段类似，绝大部分属于新疆管辖的阿克塞钦地区也是中印边界争议地区，且历史上中印多次边界冲突事件都牵涉到此地区。故**印度与中国新疆的边界争议和曾爆发的多次冲突可能是引起新疆微博网民较为愤慨的原因。**

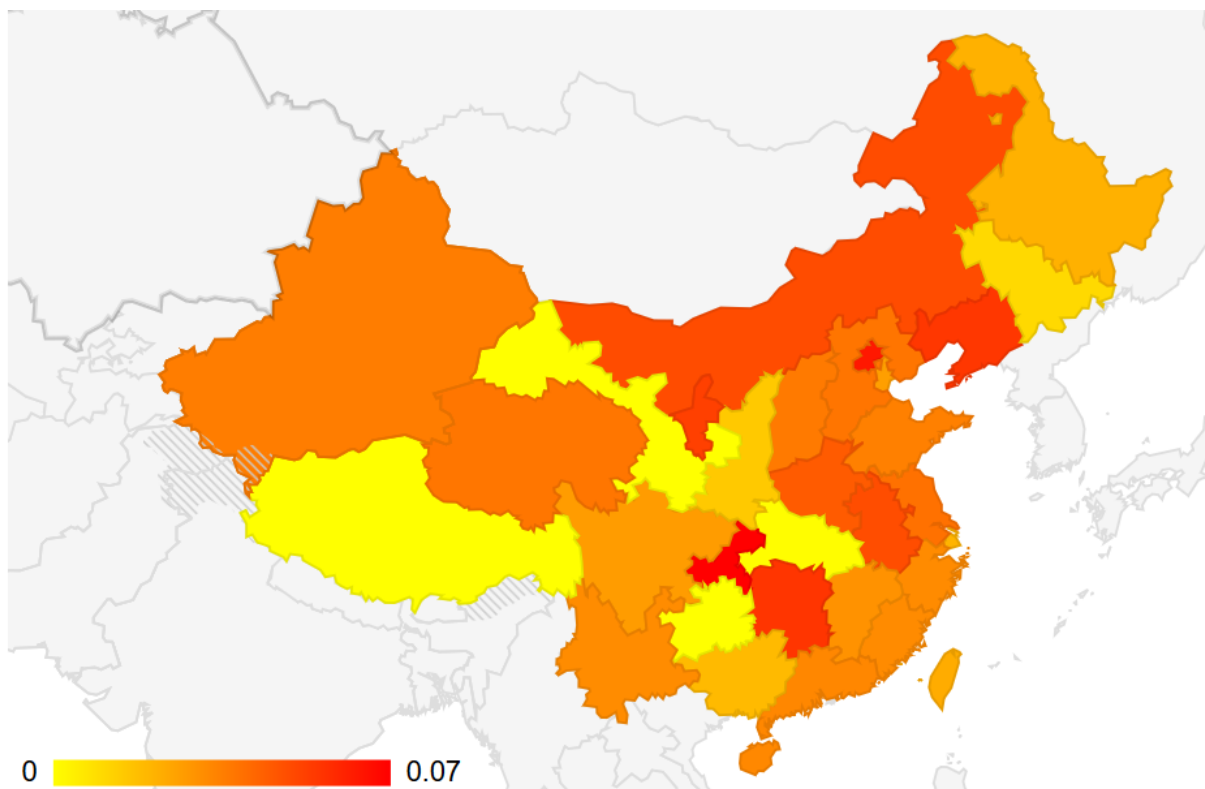


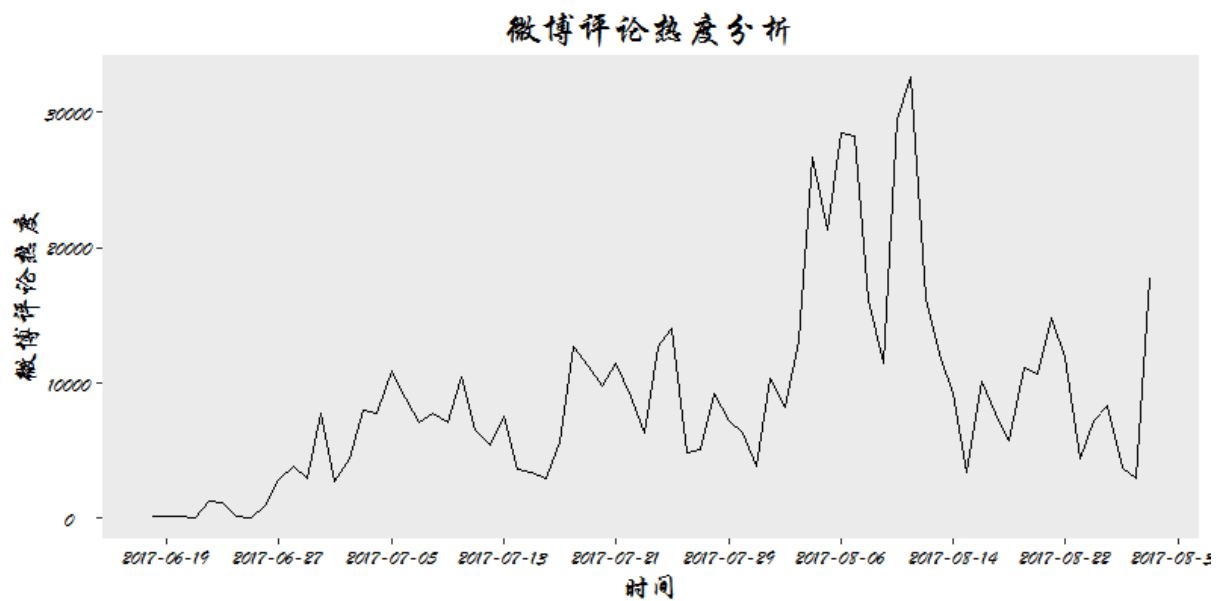
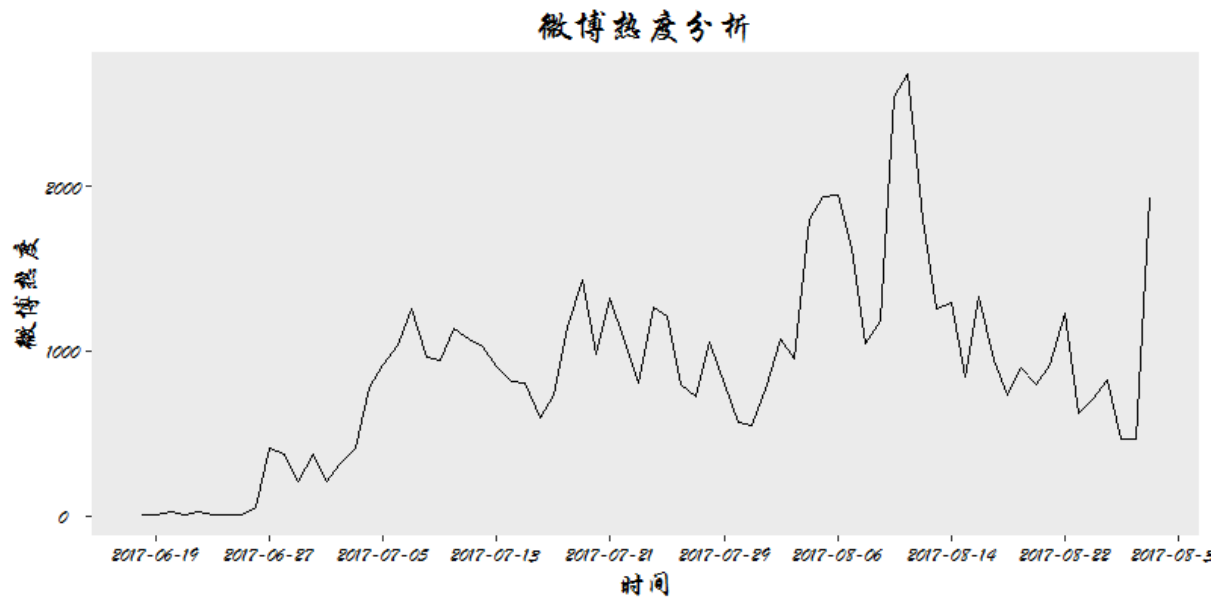
图 25 微博评论的情感地域分布

图25显示此事件下微博网民评论的情感强度的地域分布。由图可看出，北京和重庆的微博网民的微博评论的情感强度最高，表明最愤慨，其次是辽宁、湖南、宁夏等。其中，**由于北京首都和全国政治中心的性质，故北京微博网民对这类政治事件的敏感度相较更高、反应更为强烈**，这可能是导致北京微博网民评论体现出最为愤慨的情感的原因。

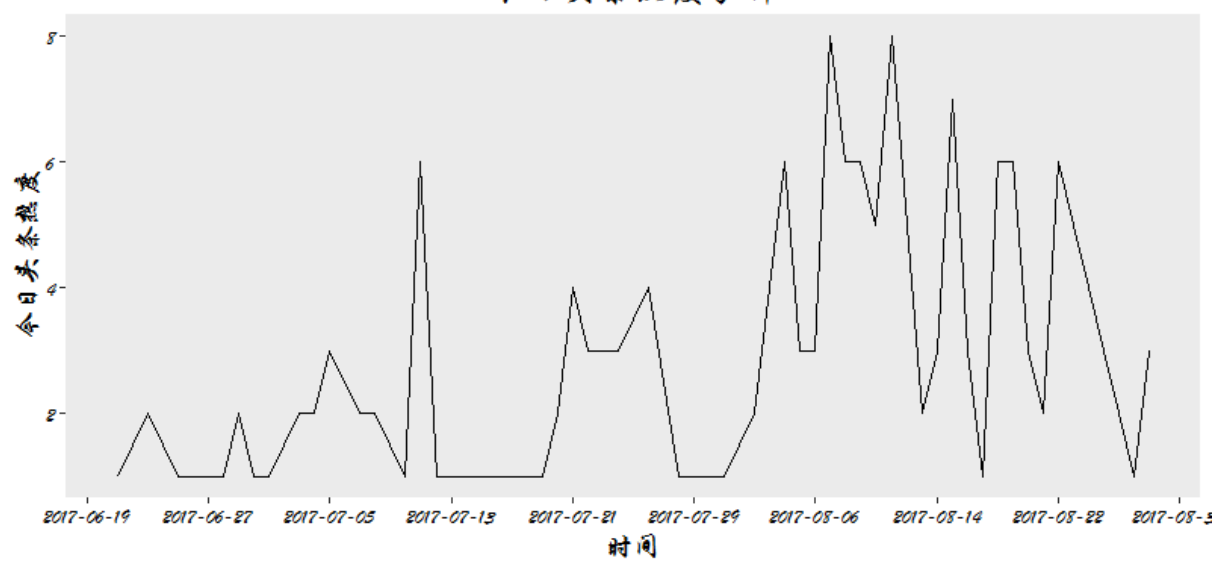
通过对比上面两表，易发现二者特征差异很大。比如：微博文章的情感地域分布中，沿中印边界线的省份的微博网民情绪较为愤慨；而微博评论的情感地域分布中，中国中部和北方地区、东北地区的微博网民情绪较为愤慨，沿中印边界线的省份的微博网民情绪反而较为平和。

六、主流媒体平台的舆情特点对比分析

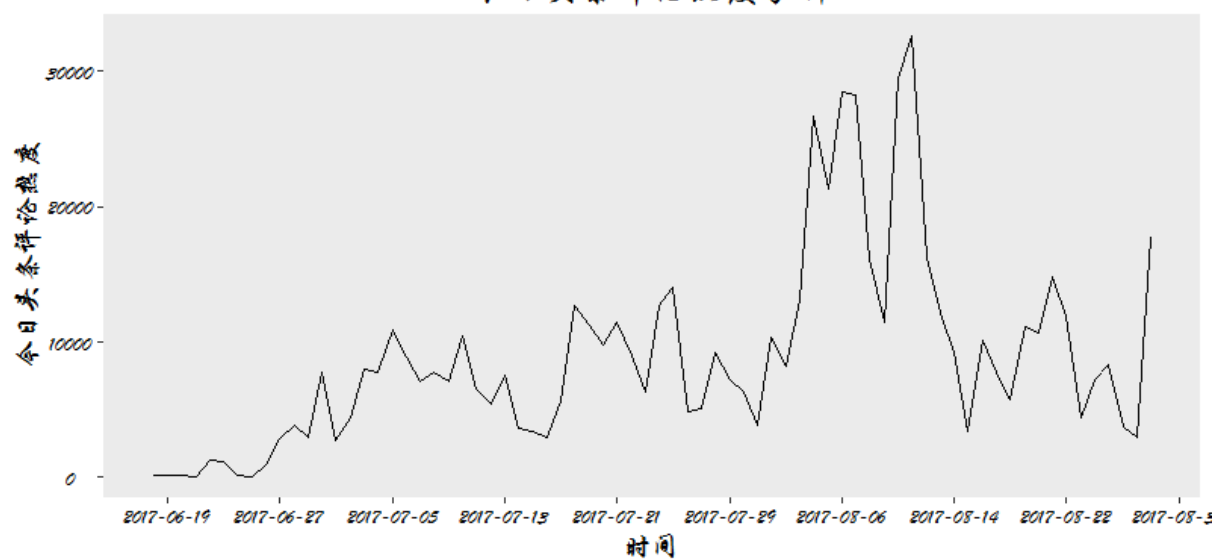
6.1 主流媒体平台热度 (声量大小) 的变化趋势对比



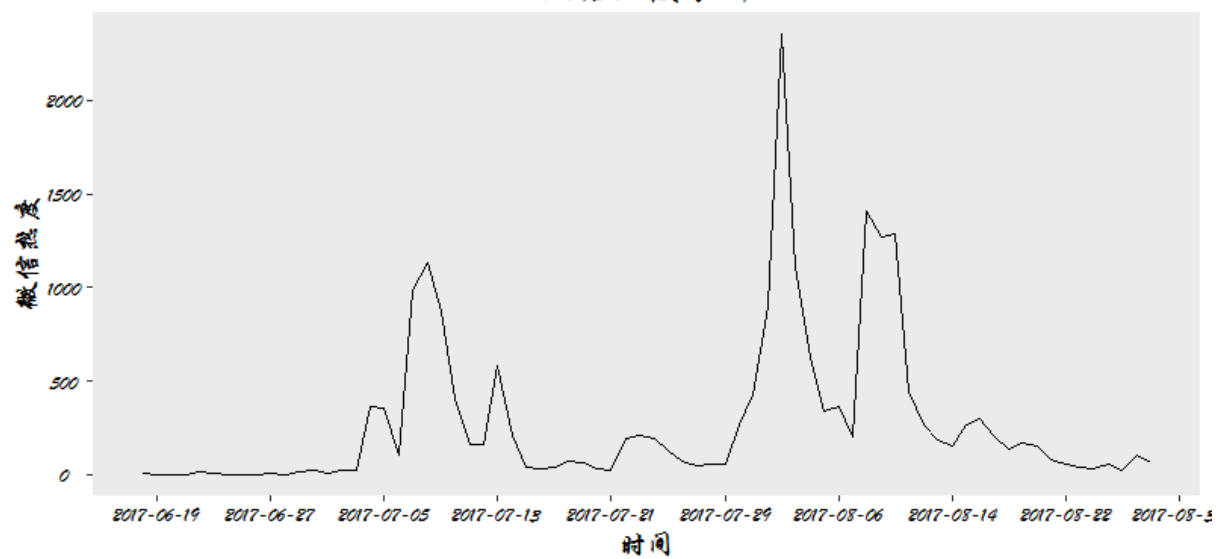
今日头条热度分析



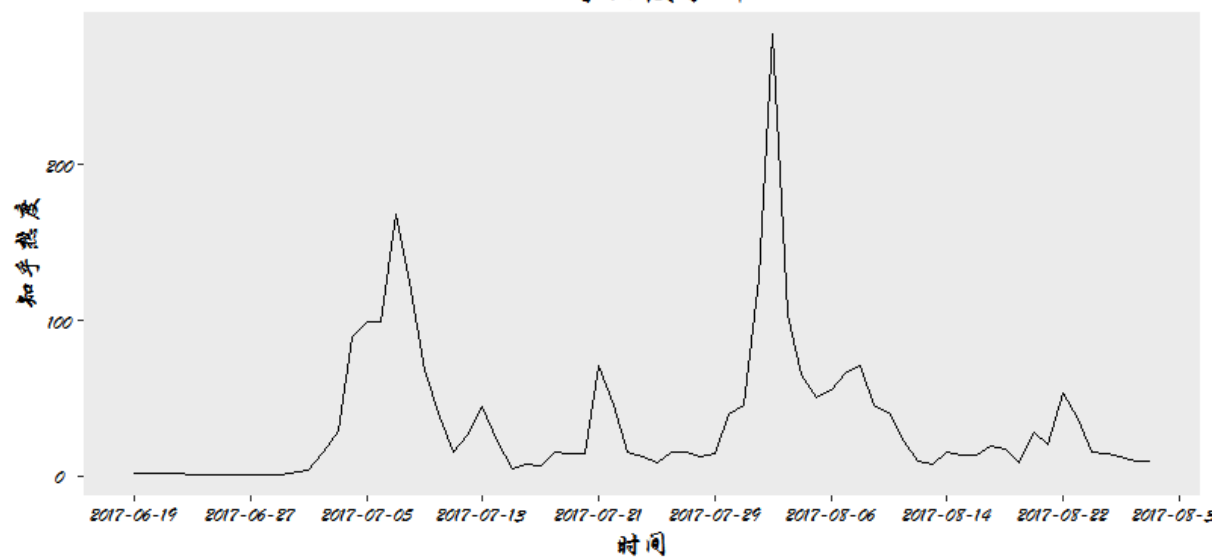
今日头条评论热度分析

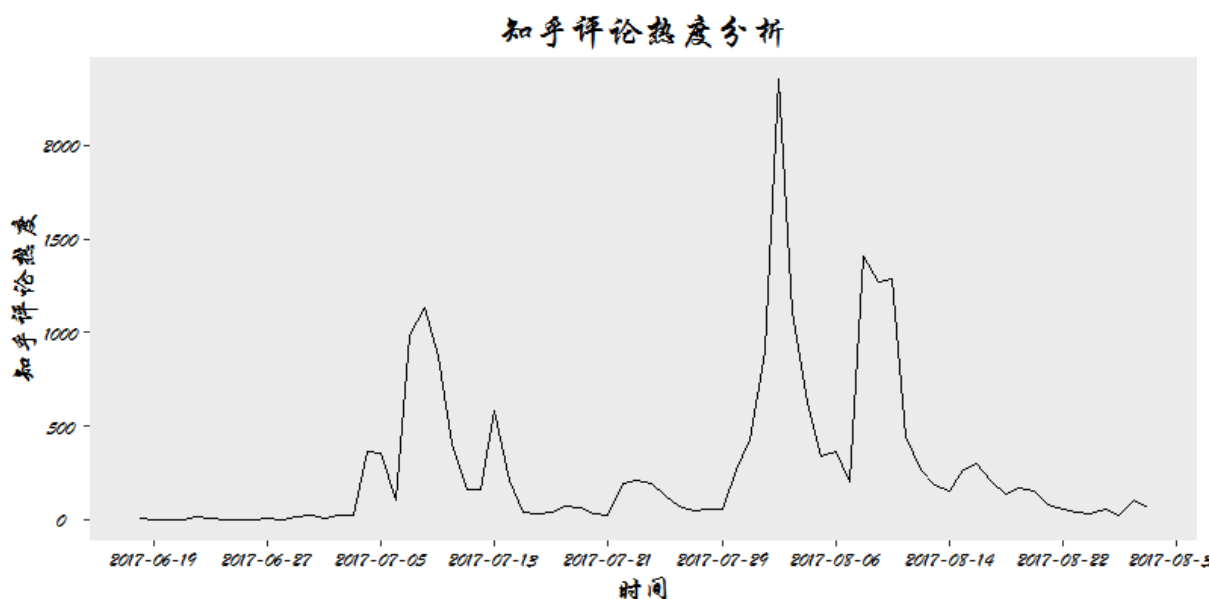


微信热度分析



知乎热度分析





分析上面 7 张图表可知，在“中印边境事件”中：

- **舆论热度大小比较：微博 > 今日头条 > 微信 > 知乎**
- **舆论热度的波动大小比较：今日头条 > 微博 > 微信 ≈ 知乎**
- 微博舆论热度最高，声量最大，其中微博评论热度要远高于其文章热度
- 微博文章与微博评论、头条文章与头条评论、知乎回答与知乎评论这三对的热度变化趋势各自十分相似
- 今日头条与微博的舆论热度变化趋势较为相似，且波动都较大；而微信和知乎热度总体波动较小
- 四个媒体平台的热度最高峰出现的时间段不同，今日头条与微博的最高峰都是在 8 月 11 日（报道：印军进入高等级战备状态），而微信和知乎的最高峰都是在 8 月 03-04 日（报道：中国六大部门集体就印方越界事件发声）。

由上可总结：在此事件中，**微博与今日头条的舆论热度特征相似，微信与知乎的舆论热度特征相似**。原因可推测：可能由于今日头条与微博的新闻即时性、社交化媒体属性、信息传播性更强，且内容都偏向于快餐式，故二者舆论热度特征相似；而微信和知乎相对来说新闻即时性、社交化媒体属性、信息传播性偏弱，且内容都偏向于深层次，故二者舆论热度特征相似。

而热度最高峰的时间点的差异，**实质上反映出这四个媒体平台对同一事件的关注点以及同一事件不同阶段进展的关注程度不同**。头条和微博上“印军进入高等级战备状态”热度最高，可能是因它涉及国家安全，极具争议性和传播性，易在这两个平台上引起广泛传播和全民讨论；而微信和知乎上“中国六大部门集体就印方越界事件发声”热度最高，可能是由于它展现出中国对待这一事件不同以往的强硬态度和坚定立场，值得深入挖掘和讨论。

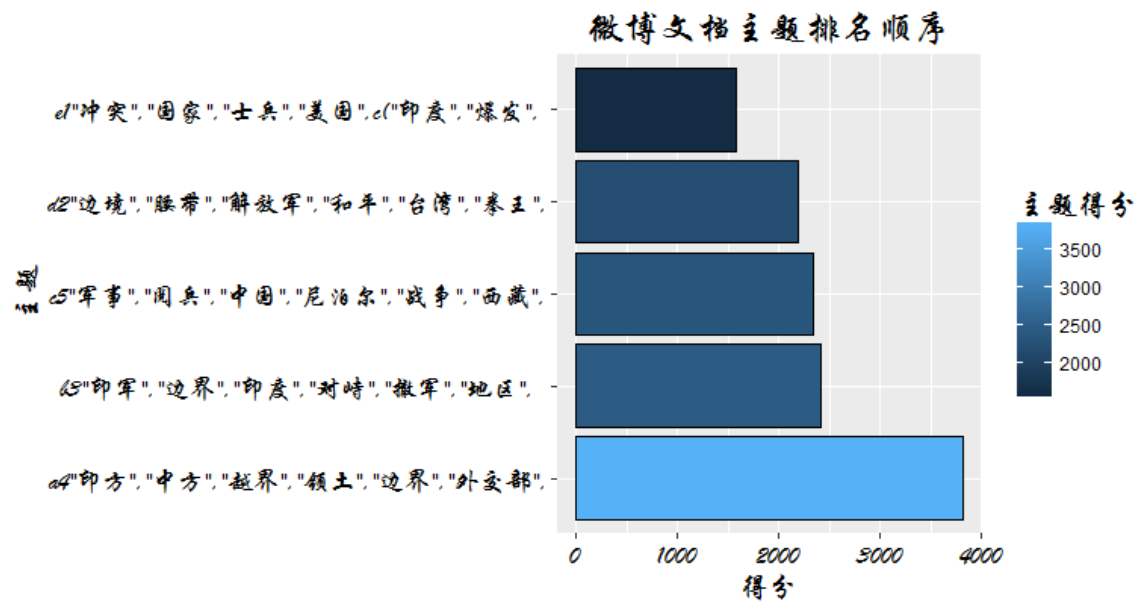
6.2 主流媒体平台的典型意见挖掘——基于 LDA 主题模型

LDA 是一种非监督机器学习技术，本报告利用该模型识别微博、微信、今日头条和知乎四大媒体平台上的文章和评论中潜藏的主题信息，并将文章与评论进行分类，并分别用若干个关键词来描述每个主题，从中提取出舆情的主流观点。

下列主题排名顺序表中：主题即反映主流观点；主题得分越高，说明其意见占比越大。

6.2.1 媒体平台文章的主流观点对比

1. 微博文档主题排名顺序



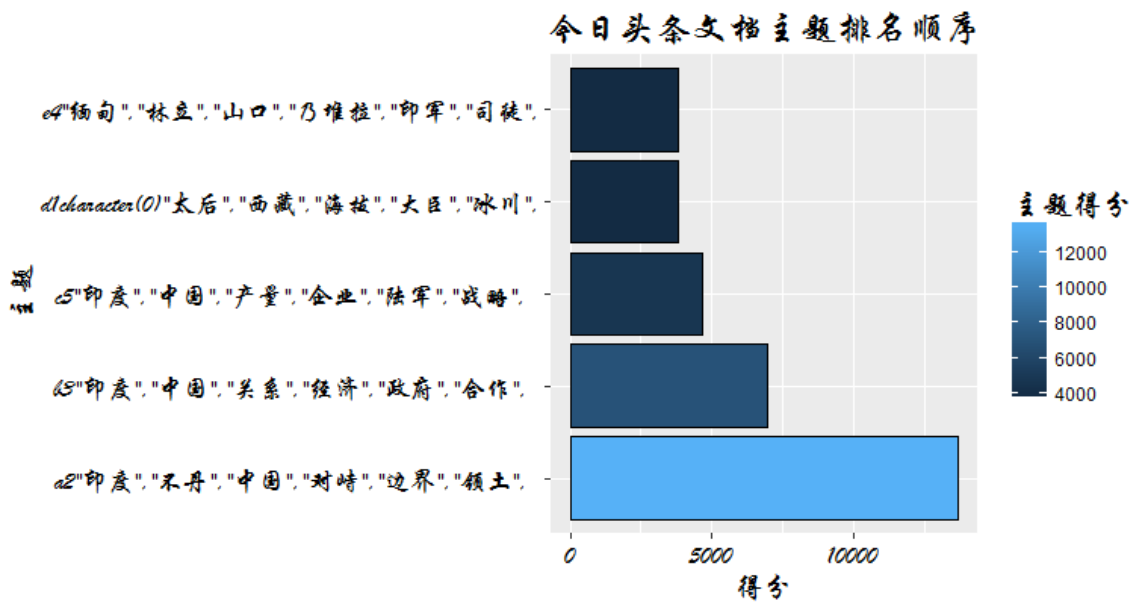
上图显示“中印边境事件”中微博文档的主流观点，其中“印方”，“中方”，“越界”，“领土”，“边界”，“外交部”主题得分最高 (3800 分)。由此可知，**微博文档倾向于报道中印洞朗边界冲突发展状况，和中国外交部的多次回应。**

这反映微博网民和媒体密切关注中印边境事件的最新进展。而密切关注中国外交部对此类政治敏感事件的表态，可能一是因为一般来说国家外交部对政治敏感事件的表态直接反映出国家解决事件的对策倾向，二是网民很在意且希望国家表现出强硬和坚定的维护国家主权的立场，这可给国民增加民族自信心和底气。

此外，微博文档还倾向于印度撤军、中国军事阅兵、尼泊尔立场等主题。

2. 今日头条文章主题排名顺序

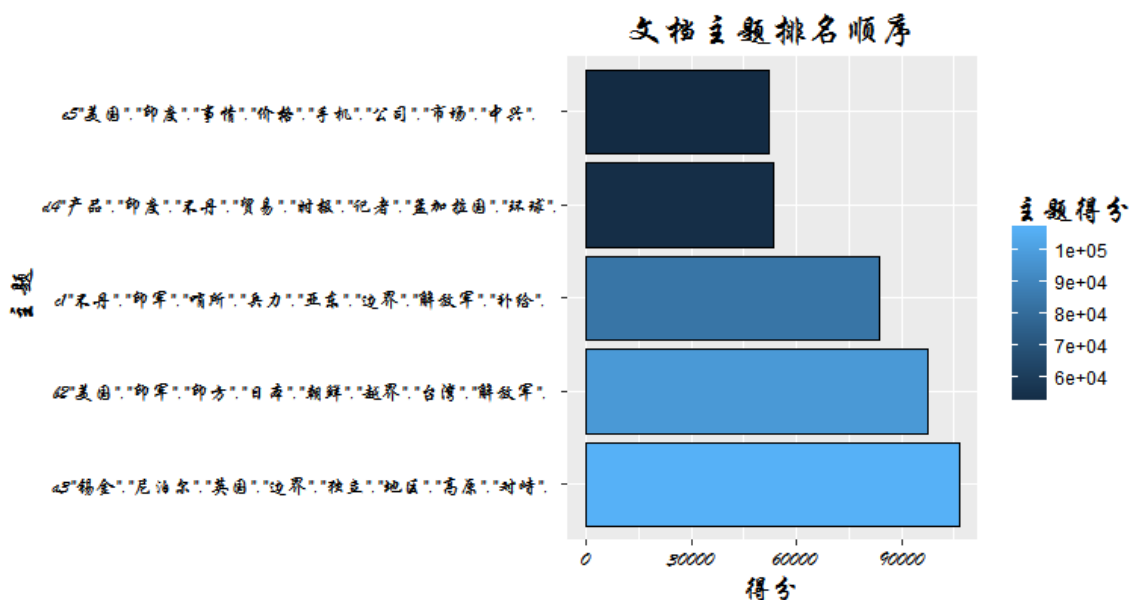
上图显示此事件中今日头条文章的主流观点，其中“印度”，“不丹”，“中国”，“对峙”，“边界”，“领土”主题得分最高 (12800 分)。由此可知，**今日头条文章倾向于**



报道中印对峙暴露出的中、印、不丹三国领土边界争议，反映出今日头条文章在此事件中对这一历史遗留问题（边界争议）的关注。这可能是由于头条的文章普遍倾向于分析“中印边境事件”背后反映出的历史因素。

此外，今日头条文章还倾向于关注中印对峙下的中印经济合作问题等。

3. 微信文章主题排名顺序



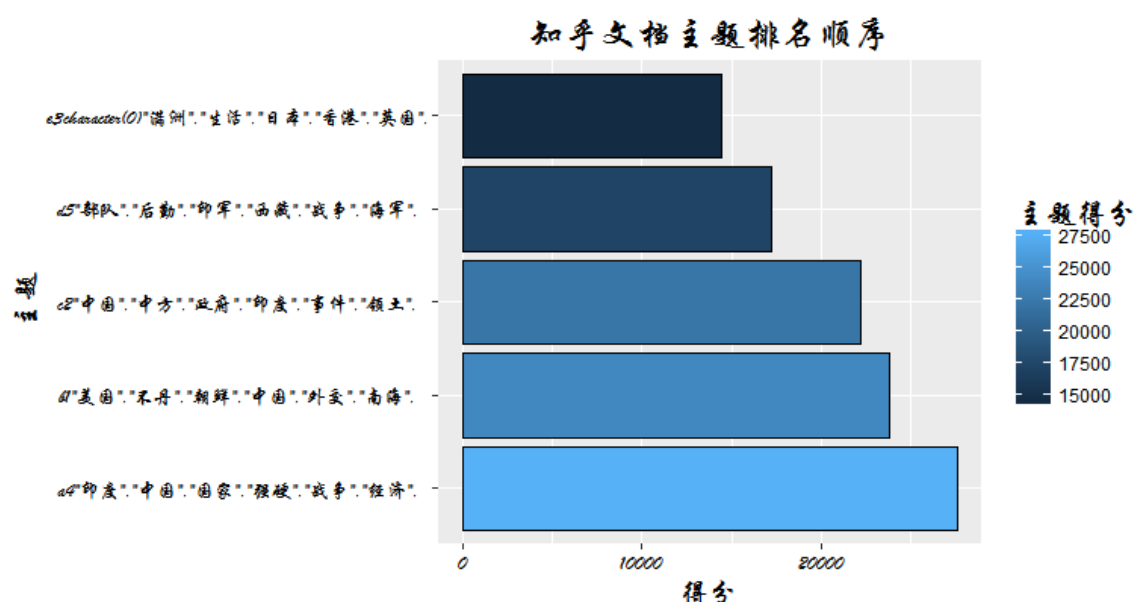
上图显示此事件中微信文章的主流观点，其中“锡金”，“尼泊尔”，“边界”，“独立”，“地区”，“高原”，“对峙”主题得分最高（110000分）。由此可知，微信文章

倾向于关注中印锡金段边界对峙、尼泊尔立场、英国立场。其中英国立场受较大关注的可能原因，一是历史上印度与英国曾是殖民地与宗主国的关系，且印度现为英联邦成员国；二是英国表态不会介入中印争端。

此外，微信文章还倾向于分析在此事件中美国、日本、朝鲜、台湾、不丹、孟加拉国的态度和立场，解放军的力量，印军的亚东军事部署，中兴等中国手机公司在印度市场受影响等主题。这反映出**微信文章将关注焦点更多聚焦于中印事件背后的国际政治关系博弈，以及事件对中印经济贸易的影响。**

其中**中兴等中国手机公司在印度市场受影响**的主要原因，一是由于近年中兴、小米等中国手机品牌集体出征印度市场，印度市场对这些公司的海外市场战略很重要；二是由于事件期间印度国民出现的反华情绪和抵制“中国货”运动可能为这些中国手机厂商增加了变数。

4. 知乎回答主题排名顺序

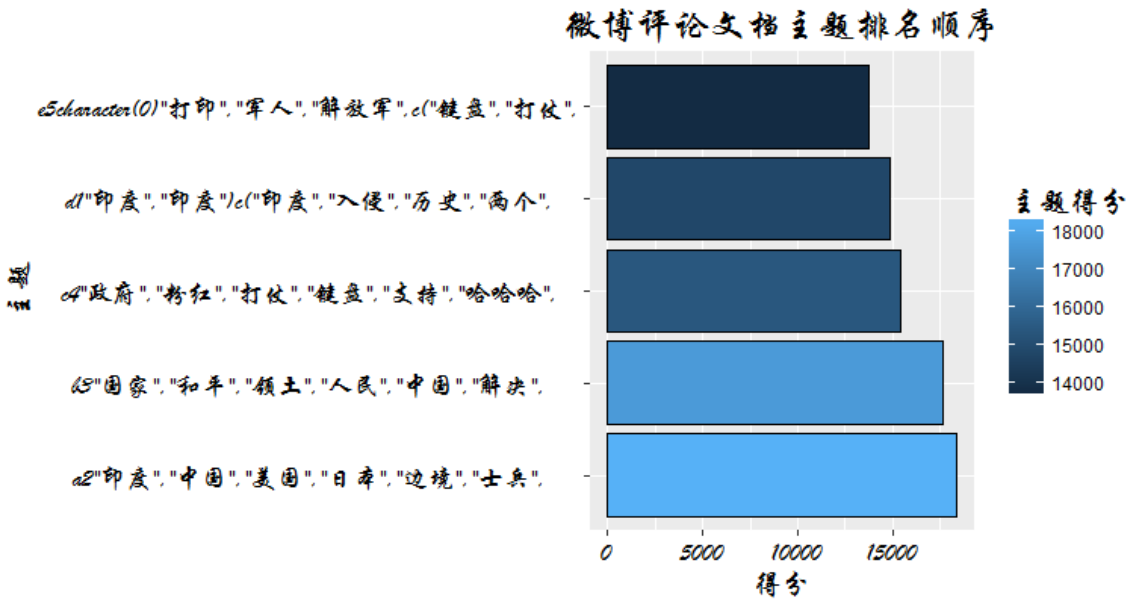


上图显示此事件中知乎回答的主流观点，其中“印度”，“中国”，“国家”，“强硬”，“战争”，“经济”主题得分最高 (28000 分)。由此可知，**知乎回答的主流观点是：对印态度要强硬，中印经济发展，是否诉诸战争。**

此外，知乎回答还倾向于谈论在中印边境事件中美国、不丹、朝鲜、南海的角色，中国的战略优劣势 (后勤补给、海军、军队部署) 等主题。由此可见，**知乎回答更倾向于分析战争、军事战略层面，反映网民对此事件下国际多方立场的关注，对中国是否该诉诸武力、以及若发生战争中国的战略优劣势的思考。**

6.2.2 媒体平台网民评论的主流观点对比

1. 微博评论主题排名顺序



上图显示“中印边境事件”中微博网民评论的主流观点，其中“印度”，“中国”，“美国”，“日本”，“边境”，“士兵”主题得分最高 (18700 分)。由此可知，**微博网民评论更倾向于讨论中印边境事件中中、印、美、日各方的国际关系。**

此外，微博评论还倾向于传播和平解决中印争端、印度非法入侵、小粉红、主战打印等主题。

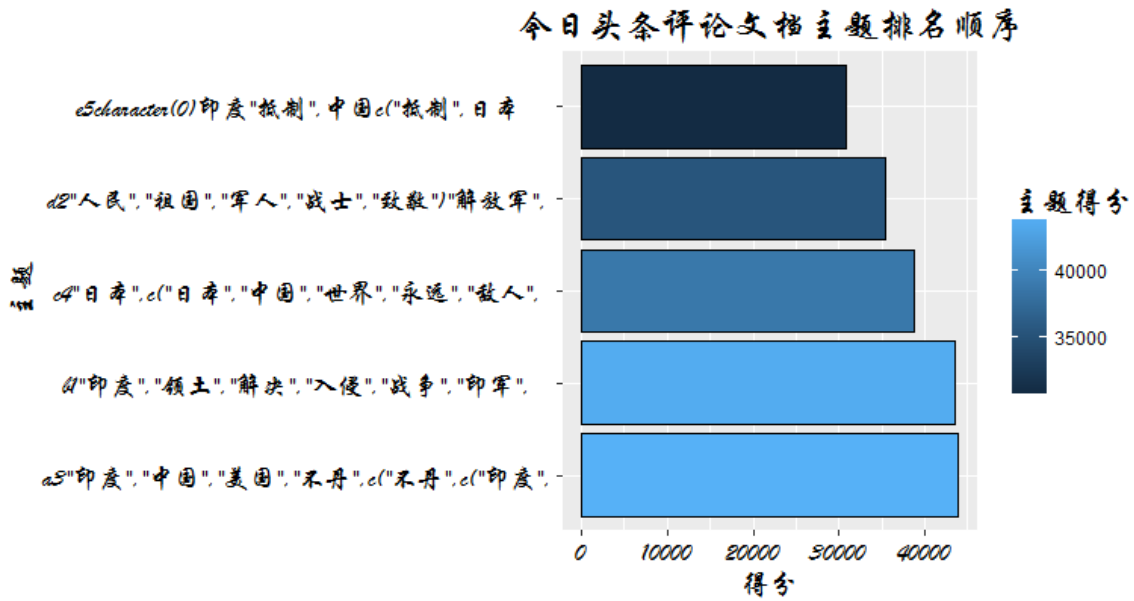
以下是对“小粉红 [4]”受到较大关注的分析：

小粉红是中国大陆自由派社交网络用户用以贬低中国大陆民族主义社交网络用户的负面标签。这部分群体具有简单粗暴的爱国主义情感和执行简单粗暴的爱国主义行为，如“抵制日货运动”。媒体对其历来褒贬不一。此次“中印边境事件”亦出现大量“主战派”“小粉红”，他们不满中国政府迟迟不肯开战，将印军入侵称为“非法越界”，主张诉诸武力捍卫国格与军威。

如：网友 @ 透过本质看真相：别国的部队都过界了，平时那么强硬的媒体砖家还有爱国的小粉红去哪里了？

但与以前不同，此次网上并未有对其的大量贬低、讽刺之声。由此可推测，**对这类国家领土被非法入侵和国家安全受威胁的政治事件，诉诸武力是网民的考虑选项之一，支持者不在少数，而不是被看做“小粉红式”的“简单粗暴的爱国主义想法”。**

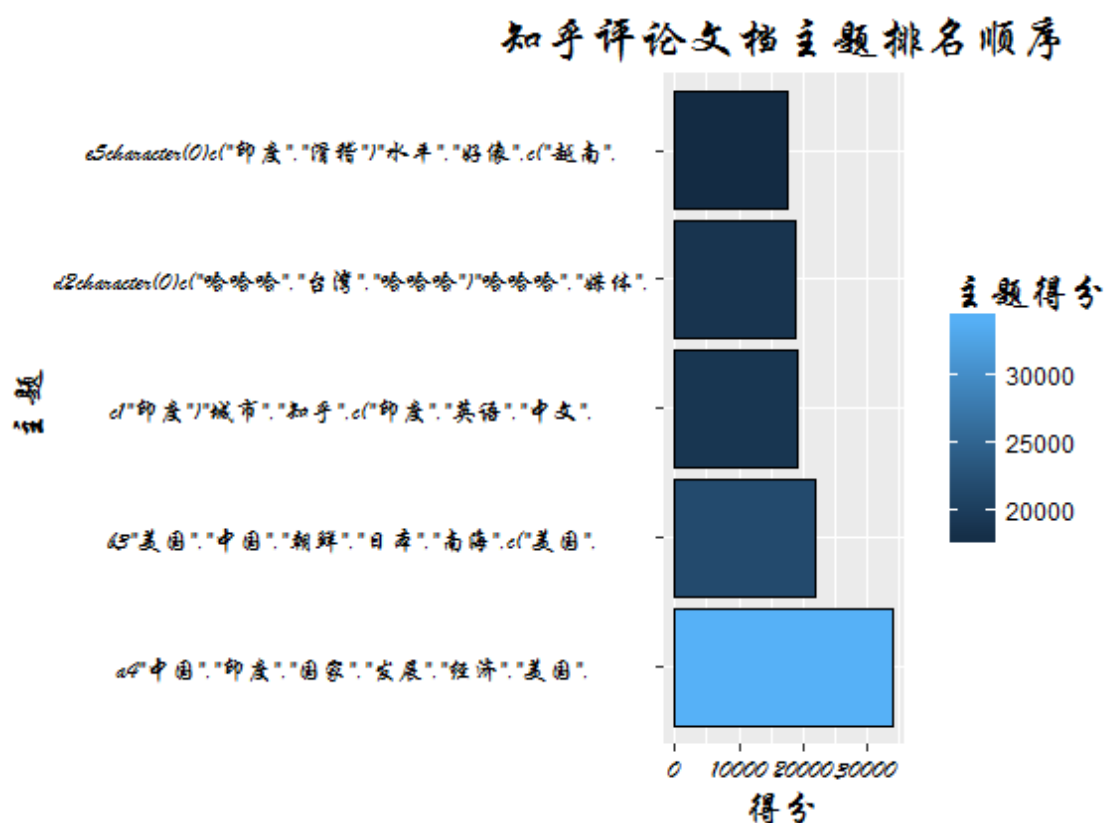
2. 今日头条评论主题排名顺序



上图显示此事件中今日头条网民评论的主流观点，其中“印度”，“中国”，“美国”，“不丹”主题得分最高（44000分）。由此可知，**今日头条网民评论更倾向于谈论此事件下中、印、美、不丹四国的立场。**

此外，今日头条评论还倾向于谈论印度抵制中国货，致敬中国解放军，印度非法入侵中国领土等主题。

3. 知乎评论主题排名顺序



上图显示此事件中知乎网民评论的主流观点，其中“中国”，“印度”，“国家”，“发展”，“经济”，“美国”主题得分最高 (34000 分)。由此可知，知乎评论更倾向于谈论此事件下中印国家关系和中印经济发展等主题，这反映**知乎网民评论中对此事件对中印国家关系和经济发展、经贸往来的影响的思考**。

此外，知乎网民评论还倾向于分析此事件下中、美、朝、日、南海五者的角色和关系，以及所涉台湾问题 (因中印冲突，台两学生被印拒绝入境)。

七、舆情总结与启示

7.1 事件分析

7.1.1 印度为何态度如此强硬？

印度态度如此强硬，背后必然有着更深层次的原因。

- 一方面印度外债高筑，经济发展形势并不乐观，政府有无法掌控局势的趋势，这时候挑起与中国的冲突能有效转移国内矛盾；
- 另一方面，随着中国“一带一路”的推进，势必会改变沿线地区的地缘政治形势，并与其他大国的利益发生冲突，中东这里就有美国的核心利益石油；而“中巴经济走廊”项目将通过克什米尔地区更是让印度痛心疾首，在围堵中国的共同利益下，印度必然得到以美国为首的相关国家大力支持，更加有恃无恐。

7.1.2 印度最终为何选择和解？

- 一是印度军事实力弱于中国；
- 二是印度现今经济情况不容乐观，经济情况可能无法承受战争的负担。

7.1.3 国际各国对此事件的态度是怎样的？

各国均表示希望中印双方和解。

7.1.4 中国的态度？

中国希望和平崛起，但在涉及领土的核心利益上，也决不会做任何妥协。从长远来看，军事冲突的代价对于双方都是难以承受，双方如果开战，伤亡在所难免，在谈判桌上和平解决争端是最好的选择。中国的在整次事件中的做法是在保证足够军事威慑的基础上，和平解决此次边境问题。

7.2 舆情整体特点总结

基于对微博、微信公众号、今日头条和知乎四大媒体平台的事件相关数据的分析，我发现了此事件有如下舆论特点。

- 全网热度峰值时间点与此事件的每一步进展的时间点基本上紧密吻合，不同平台有所差别。
- 微博的舆论声量最大，今日头条、微信、知乎分别为第二、三、四。微博对信息量的贡献占绝大部分，可以说是中国第一大“社会公共舆论场”。

- 网民关注的焦点主要是：事件进展、中国外交部的回应、事件背后的国际政治关系博弈、战争对中国经济的影响、军事战略等，不同平台用户的关注焦点有所差别。--
- 4. 微博关注网民以男性居多，主流情感是愤怒，主战派声量较大且受到很多网友的赞同。微博的事件核心传播用户基本上都是官媒或军事自媒体背景，传播层级和覆盖面非常广，且传播周期短、效率高。

7.3 政府启示

微博是中国第一大“社会公共舆论场”，网民首选且重要的舆论参与平台。微博的每一部移动终端和微博空间的每一个网民，在其本质上都演绎成整个民间公共信息系统的释放点、采集点、延伸点、链接点；微博是网民在日常社交过程中通过碎片化传播随时随地参与完成对政府社会治理、公共事务和公共服务评议的“融合讯息通道”。

政府可以从这次舆情事件分析中得到什么启示？下次类似舆情事件发生，政府该如何应对？

- 充分发挥微博大V的舆论影响力，借助相关大V来助力舆论引导，如利用官媒和军事自媒体的舆论影响力为官方舆论服务；
- 快速响应：热门微博的传播有时效性，政府需要密切关注原创发出的后几小时；
- 建立网络舆情预警机制，防止恶性信息的进一步扩散和传播，将负面舆情第一时间扼杀在摇篮中，及时疏导网络舆情，做到对其的有效应对；
- 创新正能量传播，多账号“抱团”垂直行业内部主动策划舆论传播，形成一定传播声量；
- 建立常态互动机制：主动回应舆情，从单向管理转向双向互动。

参考文献

- [1] 2017年中印军队洞朗对峙事件 <https://zh.wikipedia.org/wiki/2017%E5%B9%B4%E4%B8%AD%E5%8D%B0%E5%86%9B%E9%98%9F%E6%B4%9E%E6%9C%97%E5%AF%B9%E5%B3%99%E4%BA%8B%E4%BB%B6>
- [2] 今日头条算数中心.2016 移动资讯行业细分报告 <http://www.199it.com/archives/559343.html>
- [3] 微博数据中心.2016 微博用户发展报告 <http://data.weibo.com/report/reportDetail?id=346>
- [4] 维基百科. 小粉红 <https://zh.wikipedia.org/wiki/%E5%B0%8F%E7%B2%89%E7%B4%85>