系列危害公共安全事件分析建模报告

队伍名称:

Thursdaynight

指导老师: 刘晓娟 ¹ 北京师范大学政府管理学院

队伍成员:

高奇² 李思敏² 夏茂庚² 肖泽东³ 许景平² (队长)

- 1 北京师范大学政府管理学院
- 2 北京师范大学认知神经科学与学习国家重点实验室
- 3 北京师范大学信息科学与技术学院

赛题:系列危害公共安全事件的关联关系挖掘及预测

任务背景

各类危害公共安全的事件给广大人民群众的生命和财产带来了严重损害,极大地影响着社会稳定和民族团结。尤其是诸如系列公交车爆炸事件、极端民族主义导致的系列暴恐事件、幼儿园砍伤事件等,影响恶劣,人民安全受到极大威胁。系列事件的发生并非偶然,有些是有组织有预谋的群体性破坏行动(近期越来越呈现离散化发生趋势),有些可能是经由某些社会因素影响(诸如媒体大规模报道、网民舆论传播带来的启发和情绪影响等)发酵形成的个体行为,个体行为导致的事件一旦形成模式,危险性不亚于群体性事件。所以,了解这些危害公共安全事件在互联网上的触发、传播机理,找到相关事件间的影响关系和共性,是意义重大的研究课题。可以为事件发生的预测提供支持,如通过对系列公车爆炸事件的分析(或不同系列事件的分析),预测北京近期发生同类事件的可能性等。

竞赛任务

基本任务:

- 1. 通过数据清洗,剔除杂质或无关数据;
- 2. 根据模型和算法的维度需求,自定义标签。

核心任务(可完成部分核心任务):

- 1. 挖掘同系列事件间(如各地的公车爆炸事件)的相互触发关系和传播影响因素:
- 2. 挖掘不同系列事件间的可能关系和影响因素(或共性);(跨事件关联)
- 对国内近期涉及这三类公共安全事件发生的可能性进行预测(选一类即可)。

说明:提交较为详细的分析报告(包含模型和算法的自评标准)、程序源码等。

摘要

近年来,各类危害公共安全的事件频发,对社会造成了恶劣影响。对危害公共安全事件的关联关系进行挖掘,并对其发生进行预测,具有重要的意义。本文基于海量公司提供的近三年的三类事件(暴恐、校园砍杀、公交爆炸)的新闻和微博数据,进行了挖掘和分析。对新闻数据,提取事件信息,建立了三个模型分别进行事件预测,传播拟合和心理动机分析;对微博数据,分析各事件在公众中引起的情感变化。

为了从新闻数据集中提取事件信息,本文使用向量空间模型(VSM)对每则新闻进行量化,然后对三类事件进行去噪和中心聚类,共去除了53443条噪声数据,去除噪声的抽样准确率达100%。根据案发后该事件的报道量突增的特点,本文从去噪后的新闻数据集中提取事件列表;并使用关键字匹配(TF-IDF)的方法,划分事件列表中与各事件相关的新闻集合。对微博数据,本文采用半监督和逻辑回归分类器结合的方法,为其添加了情感(积极/消极/中性)标签。

本文对三类事件特点进行基本分析之后,首先,根据暴恐事件发生具有在同一地点连续发生的特点,建立了概率图转移模型,以较大概率(P=0.31)预测了2014年5月份的乌鲁木齐暴恐的发生。其次,考虑到事件连续发生时,前一事件对后一事件有所影响,本文建立传播模型拟合了这种互相影响。然后,本文还用心理学模型分析三类事件,揭示"无差别杀人"这一共性及其在季节和年龄上与自杀率的关系。最后,对发生最为频繁、影响也最恶劣的暴恐事件,本文通过情感分析入手,深入挖掘了其在舆论传播中的特点。

目录

摘要	5	III
— ,	前言	1
	1.1 任务背景	1
	1.2 分析思路	1
_,	数据预处理	2
	2.1 新闻预处理	2
	2.2 微博预处理	6
三、	事件提取和归并	8
	3.1 暴恐事件	8
	3.2 校园砍杀事件	10
	3.3 公交车爆炸事件	11
	3.4 新闻集归并	13
四、	事件分析	13
	4.1 暴恐事件	14
	4.2 校园砍杀事件	17
	4.3 公交车爆炸事件	20
五、	微博情感分析	22
	5.1 微博数量分析	23
	5.2 微博极性分析	24
	5.3 微博情感回归分析	26
	5.4 微博分析小结	29
六、	建模分析	29
	6.1 暴恐预测模型	29
	6.2 传播模型	32
	6.3 无差别犯罪心理模型	34
七、	总结与展望	36
八、	参考文献	38
九、	附录	39
	附录 1. 统计显著最有差异的 300 个词	39
	附录 2.暴恐事件统计表	40
	附录 3.校园砍杀事件统计表	41
	附录 4.公交车爆炸事件统计表	41
	附录 5.微博情感极性分类抽样结果	42

一、前言

1.1 任务背景

近年来公共安全事件频发,给广大人民群众的生命和财产带来了严重损害, 影响着社会稳定和民族团结。尤其是诸如系列公交车爆炸事件、极端民族主义导 致的系列暴恐事件、幼儿园砍伤事件等,影响恶劣。

公共安全事件的发生,从个体案例来看属于偶然。然而从整体出发,系列事件的发生背后很可能隐藏规律。这种规律的重要性显而易见,因此一直是研究的热点。但回顾过去,以往对个体案例的研究为时空环境所限制;而大数据时代的到来——更大的数据存储处理技术、更有效的学习分析算法、更强大的计算能力支撑——让我们有机会跨越时空的界限,从更加宏观的角度来观察这些看似孤立的案例之间隐藏的联系。

每一个体案例的发生都会被新闻详细的记述,同时对大众产生的影响被微博所记录。对这些事件记述和记录进行深入的挖掘,可以找到事件内在和交互关系的规律,为事件发生的预测提供支持,从而建立起公共安全预警机制,保障人民生命财产安全和社会的安定。

1.2 分析思路

海量公司提供的数据包括新闻数据和相应的微博数据。时间从 2011 年 4 月到 2014 年 4 月,新闻数据共有 540204 条,微博数据共有 178649 条。共包括 18 个字段,基本描述了一条资讯的各种特征。对待新闻,本文的目的,首先是从这些新闻信息中整合出真实的事件,并抽取出对真实事件特征的描述;然后通过事件特征,对同类事件之间的关联进行建模研究。而对待微博,本文的目的,是挖掘微博的情感极性,进行事件舆论传播的归纳。

根据目标,本文按照如下流程对海量提供的数据进行了分析。

第一步是对新闻、微博数据进行预处理(数据清洗)。将新闻信息向量化表征,离事件信息数据集中心较远的数据去除。然后按照微博、新闻和三类公共安

全事件将数据划分为六个数据集。对微博数据添加了情感标签。

第二步是事件抽取,得到每一类的事件列表及对应的事件特征描述。可以根据前面划分出的新闻数据集绘制出三个信息量曲线图。通过设定阈值,找到波峰,然后在波峰对应日期查看新闻的方式找到了所有三类安全事件的集合,并抽取出了事件特征描述。

第三步是事件分析。分析事件发生的时间、地点、犯罪手法、死伤人数、被害者类型、罪犯人数以及新闻传播量等等特征,找到系列事件发生的规律,为建模分析做准备。

第四步是微博的情感分析。本文对事件对微博数量、事件对微博的情感极性、以及事件对舆论情感极性的影响进行分析,得出一系列舆论传播有关的结论。

第五步是建模与分析,根据事件分析得到的规律,本文建立了暴恐预测模型,新闻传播模型,心理发展模型对事件关系进行预测、拟合、解释。整个过程如图 1-1 所示

最后对整个工作进行了总结和展望。



图 1-1 分析思路图

二、数据预处理

为了得到可靠的数据方便后续处理,本文进行了数据的预处理。新闻数据比较系统和正式,本文希望通过新闻整合出真实的事件,分析建模。所以新闻预处理中本文主要进行了去噪和重新分类。而微博数据较为碎片化和情绪化,本文希望通过微博探索事件的舆论反应,所以添加微博情感标签。

2.1 新闻预处理

对新闻预处理包括对新闻数据向量化:基于向量化的数据,剔除与暴恐事件、

校园砍杀与公交车爆炸事件无关的新闻数据;根据聚类的方法对划分有错的新闻重新划分到暴恐事件、校园砍杀与公交车爆炸事件三类新闻集合中。

2.1.1 向量化

在剔除新闻的错误分类工作之前,首先基于向量空间模型(Vector Space Modal, VSM)对新闻数据进行向量化表示。向量空间模型,首先定义一组关键词集作为新闻的特征,然后再确定每篇新闻在这些特征的权重形成一个基于关键词的特征向量[1]。

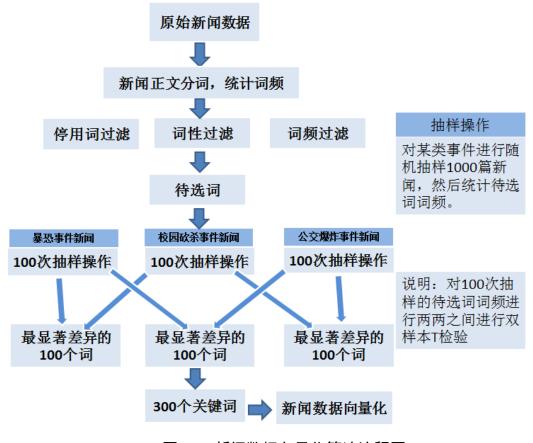


图 2-1 新闻数据向量化算法流程图

定义关键词集:用统计的方法提取最可能表征三类事件特点的词,具体算法如下:

- (1) 找出数据库新闻类所有文章出现的词进行词频统计,出现的词进行词性和停用词过滤,保留下的词作为待选词集
 - (2) 分别对两类型新闻信息(C1, C2)进行随机抽样,每次抽取 M 篇作为

- 一个样本. 总共抽取 N 次 (M=1000. N=100)
 - (3) 对每个样本的待选词集进行词频统计
- (4) 对两类文档进行 t 统计检验, 词频显著最有差异的词留下作为关键词集
 - (5) 选取前 K 个最显著有差异的词作为关键词集(结果见附录 1) 抽取出具体的关键词集,接下来对新闻数据进行向量化表示

权重确定:分别统计每条新闻的关键词集的词频,如果词频大于0,权重为1,词频为0,则权重为0。

对向量化的新闻数据进行 PCA 降维能够很明显区别出三类事件,这进一步说明向量化后的结果能够很有效表征三类事件。

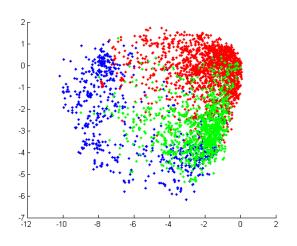


图 2-2 向量化的新闻数据经 PCA 降维

2.1.2 去噪

为了保证后期处理的数据可靠性,本文对新闻数据进行了噪声过滤,根据新闻数据的特点,字数少的基本是关于这三类事件的图片和视频新闻,字数过多的新闻基本是垃圾广告新闻,字数在 250~10000 间的同时包含噪声和三类事件的新闻。针对不同的数据,做不同的处理,如图 2-2 所示:

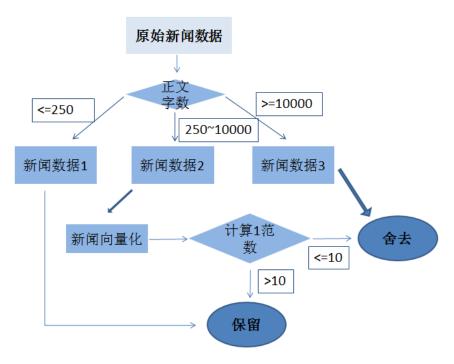


图 2-3 去噪算法流程图

对流程图中的新闻数据 2 的新闻,基于向量化后新闻,计算每篇新闻的距离原点的距离一阶距,小于阈值 10 的认为是噪声。

经过上述操作总共找出了 53443 篇噪声数据,在认为是噪声的文章中抽出 1000 篇,进行了噪声验证,去噪正确率为 100%。

2.1.3 数据分类

题目所给数据有些新闻数据没有归类到单一的某类事件,为了确定新闻信息的分类,本文对数据做了聚类分析。首先计算向量化后三类事件新闻的数据中心,对有争议的新闻数据根据其距离三类事件的中心点的欧式距离的远近程度进行聚类,将其归类到距离最近的类别中。

之后按照新闻、微博和三类事件,将整个数据集划分为六个子集。六个子集分别为暴恐事件新闻、校园砍杀事件新闻、公交车爆炸事件新闻、暴恐事件微博、校园砍杀事件微博、公交车爆炸事件微博。

2.2 微博预处理

微博因为其自身简短、多变的特点,因而处理有别于新闻,不能进行直接划分到具体事件。然而微博不同于新闻,首先它时效性更短,这使得发出微博的时间就能基本锁定最相关的事件;其次微博由个人发出,其统计量便能够代表公众舆论倾向。因而要想获得舆论对事件的反应,最好的途径便是微博。根据需求,给微博打上情感标签,供后续分析使用。

对于微博情感分析,最常见的方法是统计微博中包含感情的词汇、表情等内容,训练分类器进行情感极性判别。更加高级的方法是识别出微博内容中的命名实体,整理命名实体和情感词汇的联系来判别微博情感^[2]。

考虑只需要统计上的舆论情感倾向,并不需要精确判别单条微博的情感,因而采用了开源工具 lingpipe^[3]提供的 logistic regression 分类器,用半监督的思路迭代训练,得到微博情感标签。

2.2.1 分类器训练

LingPipe 是一个自然语言处理的 Java 开源工具包,由 alias 公司开发,有 很丰富的功能,包括主题分类、命名实体识别、中文分词、情感分析、语言辨别 等等。功能强大,文档详细,每个模型甚至连参考论文都列出来,使用非常方便。

Lingpipe 对文本基本极性的分类采用 logistic regression 的方法。首先调用其已完成的分词和特征抽取 API,将输入文本转化训练 logistic regression 使用的数值向量形式,然后将数值向量输入分类器得到分类结果。

在分类器的训练上,采取了半监督学习的思想。算法流程如下:

- (1) 首先用 1000 条标记过极性的微博(积极约 200 条, 消极约 200 条, 中性约 600 条) 作为训练集。
- (2) 训练分类器。用分类器对暴恐(校园砍杀/公交)事件微博进行情感判别。
- (3) 从新得到极性标签(积极/消极/中性)的微博中抽取一定数量(<5000) 加入训练集。
 - (4) 重复(2) 到(4) 步(到测试集测试结果正确率超过80%)

2.2.2 分类结果测试

训练出的分类器在测试集上的结果如下图,训练集 10324 条微博,测试集 1110 条微博,正确率可以达到 89%。测试结果如图 2-3 所示。

Training.

- # Training Cases=10324
- # Training Chars=1643275

Evaluating.

- # Test Cases=1110
- # Correct=994
- % Correct=0.8954954954954955

图 2-4 微博情感分类测试结果

2.2.3 分类人工检测

在微博有了情感标签之后,本文首先进行了人工抽样检测,查看积极/消极/中性的微博是否已经正确的分类。结果(随机抽样结果见附录)显示微博的极性基本分类结果具有如下特征:

消极: 悲伤、缅怀、恐惧、哀伤、愤怒…等

中性:事件事实陈述,情感因素不强的个人观点…等

积极:希冀,祈福,事件处置…等

抽样结果表明,微博的情感分类结果比较明确,积极/消极/中性可以基本代表舆论的情感倾向,满足做舆论分析的需求。

三、事件提取和归并

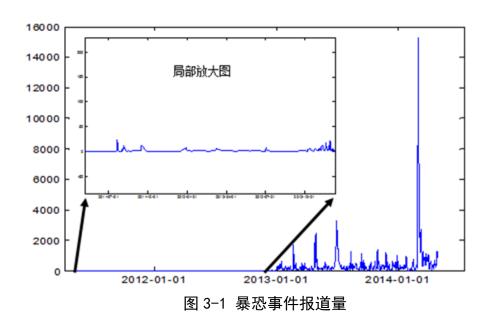
事件抽取部分中,目的是抽取出所有真实事件。经过数据预处理,本文把新闻集合中的无关数据去噪,并根据"暴恐事件"、"校园砍杀事件"和"公交爆炸事件"将去噪后新闻划分到不同的新闻集合。因为在纷繁复杂的新闻中找出时空唯一的单体事件,对某一类新闻的定量、定性描述非常关键,所以在事件抽取这一步,本文进一步的对数据进行统计和抽取。最终得到了三个分类下的事件及其特征描述的列表。

对三类事件处理方式相似,都是归纳出事件并抽取事件特征详细的描述。而 后用 TF-IDF 算法将描具体某一事件的新闻归纳到一起,这样每一个具体事件都 有其新闻集合。

3.1 暴恐事件

3.1.1 归纳事件

首先对去噪后的、标记有"暴恐"标签的新闻数据,进行了分天的新闻数目统计。统计结果如图 3-1 所示,横轴是时间,纵轴是新闻条目数量。



可以看到新闻数目随时间变化的曲线有非常多的尖峰。应该注意到的是, 2013年之前的新闻数目相比于 2013年之后的新闻数目有数量级级别的差距,因 而在一张图上不好显示,其局部放大如图 3-1 所示。

基于事实经验,对某一暴恐事件的发生做出如下假设:若暴恐事件发生,则当天以及事件发生接下来几天内都会有事件相关的报道。根据事件的严重程度的不同,报道数量会有所不同。一般而言,有关该事件的报道数量越多,那么可以间接的反映出该事件越严重、影响越大。

基于上述假设,可以得到如下结论,即:如果报道量在某一天及紧接而来的几天突然增加,即在图上表现为一个新闻数目尖峰,那么很可能是这一段时间发生的暴恐事件造成了报道量增多,这样就可以锁定事件发生的时间邻域,即尖峰最高点附近。

根据上面的假设推断,找出尖峰对应的时间邻域,并抽取"暴恐"关键词向量内积最高的 10 篇新闻进行人工的总结,抽取具体事件。表 3-1 是抽出的暴恐事件列表。

表 3-1 暴恐事件表

	A C C A STORY
时间	事件
2011-7-18	新疆和田派出遭恐怖袭击
2011-7-31	新疆喀什市汽车冲撞人群暴力恐怖案件
2012-2-28	新疆叶城暴恐行凶
2012-6-29	新疆劫机案
2013-4-23	新疆喀什巴楚县暴力恐怖案件
2013-6-26	新疆鄯善县暴力恐怖事件
2013-8-28	喀什团伙暴力恐怖案
2013-10-28	吉普車天安門爆炸
2013-12-15	新疆袭击事件
2013-12-30	暴徒冲击新疆莎车县公安局
2014-1-25	新疆新和县暴恐案

2014-3-1

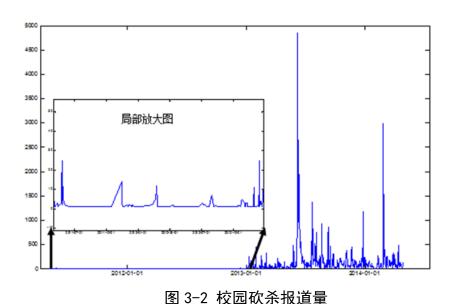
3.1.2 抽取事件特征

人工抽取该事件的关键信息描述见附录 2。具体的信息还包括事件的发生时间、发生地点、作案方式,警察伤亡人数,平民伤亡人数,恐怖分子伤亡被捕逃跑人数等等。

3.2 校园砍杀事件

3.2.1 归纳事件

我们也同样绘制了校园砍杀新闻集合内,精确到天的新闻条目变化。如图 3-2 所示。横轴是时间,纵轴是新闻数量。



校园砍杀事件的新闻数目(图 3-2)表现出了和图 3-1 相似的特性,即 2013年之后的新闻数目和 2013年之前的新闻数目不在一个数量级,因而同样的放大了 2013年前的局部。

同样和处理暴恐事件新闻数目的方法一致,归纳出校园砍杀事件如表 3-2 所

示。

表 3-2 校园砍杀事件表

时间	事件
2012-9-21	广西平南精神病砍杀小学生
2013-1-16	新化北渡小学惨案
2013-1-30	浙江小伙持刀砍幼儿园
2013-2-1	开封发生校园伤害案
2013-5-18	湖南衡阳校园血案
2013-5-22	广东茂名精神病人持刀伤害小学生
2013-8-7	癫痫病患砍杀学生
2013-9-10	桂林校门口爆炸案
2014-4-18	安康一中学学生被砍案

3.2.2 抽取事件特征

抽取的校园砍杀事件相应的特征描述见附录 3。

3.3 公交车爆炸事件

3.3.1 归纳事件

处理公交爆炸事件新闻集合的方法和处理前两种方法的一致,首先是新闻数目随时间变化如图 3-3 所示。

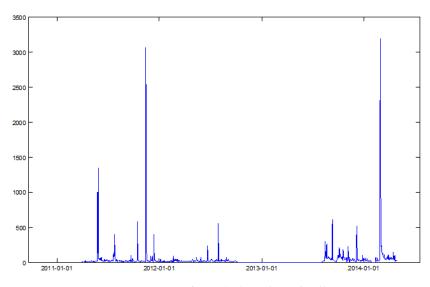


图 3-3 公交爆炸案件新闻报道量

这里应该注意到,公交车爆炸案的新闻数目,并没有在 2013 年前后产生数量级的变化。虽然没有前面两类新闻条目跳变的特性,但是依然可以用之前的办法处理。经过分析,得到了公交车爆炸案的案件列表 3-3。

表 3-3 公交车爆炸案件表

时间	事件
2011-5-4	温州公交车场起火爆炸
2011-9-7	哈尔滨公交行驶中自燃
2012-7-28	北京市9路公交车纵火案
2013-1-8	郑州 92 路公交自燃
2013-1-28	哈尔滨公交车起火
2013-3-11	桐城公交车行驶中失火
2013-4-4	杭州公交车自燃
2013-5-27	广东东莞公交车爆炸
2013-6-7	厦门 BRT 公交纵火案
2013-7-9	杭州城区公交车自燃
2013-9-19	成都 98 路公交车上点火炮
2013-10-8	北京公交自燃
2013-10-18	湖北襄阳公交车失火

2013-10-25	广东韶关公交车起火
2013-12-10	公天津交车起火
2014-2-27	宁波公交车起火
2014-2-27	贵阳公交车纵火案
2014-3-11	南京公交车起火

3.3.2 抽取事件特征

人工抽取到的公交爆炸案事件特征描述见附录 4。

3.4 新闻集归并

得到了具体的时间列表后,发现新闻是互相混杂的,如 A 事件发生 10 天之后 B 事件的报导就会和 A 事件的后续报导混杂起来。

为了细致得研究单独事件,需要确定属于具体事件的新闻集合。在这里,采用了 TF-IDF 算法将确定属于某一具体事件的新闻划分到同一集合。对某一具体事件确定其新闻集合的算法步骤如下:

- 1. 给定三篇相关报导作为种子新闻
- 11. 定义该事件的时间邻域
- 111. 使用 TF-IDF 算法获得该事件的关键词集合
- IV. 在事件时间邻域内,对所有新闻进行关键词匹配,将满足需求的新闻加入该事件的新闻集合

具体事件的新闻集合参考附件。

四、事件分析

事件分析部分,本文的目的是找出每类事件的特点和规律。当得到具体事件 表之后,可以进一步收集每个事件的具体信息,从而挖掘的时间地点特征和事件 之间的关联规律。筛选出了每个事件的时间,地点,犯罪者信息,被害者信息, 新闻报道量,微博情感特征等事件特征,对其进行了分析

对根据造成的死亡人数对危害公共安全事件进行了评级,标准如下:

表 4-1 事件危害等级表

死亡人数	危害等级
0	一级
1~10	二级
10~20	三级
20~30	四级
30 以上	五级

4.1 暴恐事件

对暴恐事件的发生地点,作案方式,作案团伙,受害者类型进行了分析。发 现暴恐事件的发生具有时间连续性和空间连续性等特征。

4.1.1 暴恐事件的地点分布

按暴恐事件的地域分布绘制了暴恐事件地域图 4-1。

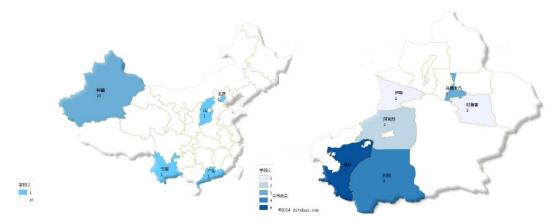


图 4-1 暴恐事件地域分布图

根据图中可以看到,暴恐事件集中在新疆地区,在疆外其他地域也有发生。 而在新疆地区中,以喀什地区最为严重,有向外辐射的趋势。距离喀什越近辐射 越强,同时会倾向于在省会,发达城市发生。

4.1.2 暴恐事件的作案情况

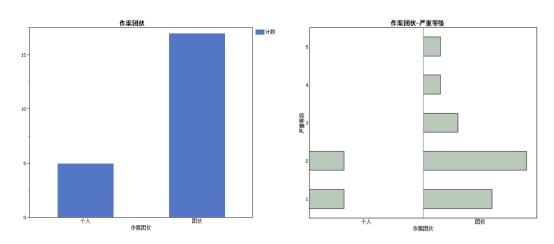


图 4-2 暴恐事件团伙与严重等级图

对恐怖事件作案人员是团伙或者非团伙进行了分析,发现大部分暴恐事件是集体作案,而且绝大部分3级以上恐怖事件都是团伙作案作为。

4.1.3 暴恐事件的作案地点与严重等级关系

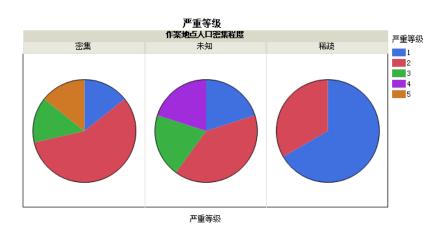


图 4-3 暴恐事件作案地点与严重等级图

对恐怖事件发生地点人口密集程度的统计发现,人口稀疏地区一般不会发生3级以上恶性恐怖事件,恶性恐怖事件一般发生在人口密集的街道,火车站等地。

4.1.4 暴恐事件的受害人类型与严重等级关系

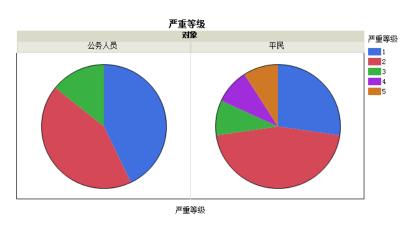


图 4-4 暴恐事件受害人与严重等级图

通过对恐怖事件受害人统计,发现当恐怖事件受害者为警察、公务人员你的时候,死伤人数较少,而当恐怖事件受害者为普通平民,死伤人数剧增。

4.1.5 暴恐事件的作案方式分布

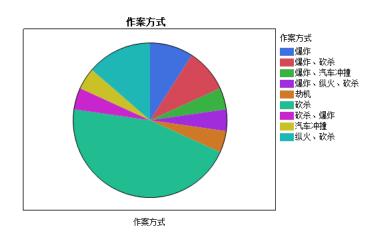


图 4-5 暴恐事件作案方式

通过对恐怖事件作案方式的分析,绝大部分恐怖事件采用了砍杀或者多种作案方式综合的手法。可能因为砍杀会对人心里造成更大的恐惧,具有更大的恐怖效果。而多种手法综合更容易造成更大的死伤人数。

4.2 校园砍杀事件

针对校园砍杀事件,本文分析了几个方面包括时间、地点、死伤人数以及发生原因。时间局部性、地点的边境分布以及精神病人无差异砍杀的具体分析如下。

4.2.1 校园砍杀事件的时间分布

从校园砍杀事件发生的时间分布图 4-6 看,在开学月份(3月、4月、9月、10月)以及邻近放假月份(12月、1月、7月)发生的校园砍杀事件较多,而中间的 5、6、11月份发生的校园砍杀事件较少。发生校园砍杀事件较多主要是返校和离校的月份,学生心理活动较不稳定,容易发生一些心理冲动的行为。

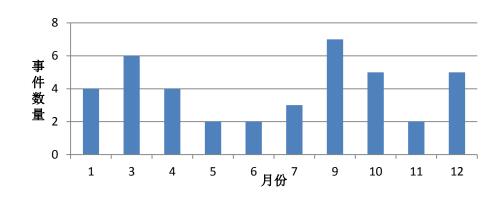


图 4-6 校园砍杀事件时间分布

4.2.2 校园砍杀事件的地点分布

从省份来看,如图 4-7 所示,分布在广东的最多,有 6 起;在云南有 4 起;在安徽、广西、海南、湖南各有 4 起;在河南、江苏、江西、陕西、上海、四川各有 2 起;在北京、黑龙江、湖北、甘肃、山东、浙江最少,各有 1 起。

从校园砍杀的省份分布可以看出存在3起以上的都是南方的省份,南方存在校园砍杀的省份占所有存在校园砍杀省份的61.11%(11/18),南方发生的数量占所有数量的72.5%(29/40),无论从北方发生省份的数量还是校园砍杀事件的数量,南方都要明显多于北方。此外从发生校园砍杀发生数量较多的省份更多的是边疆省份例如广东、云南、广西、海南等。

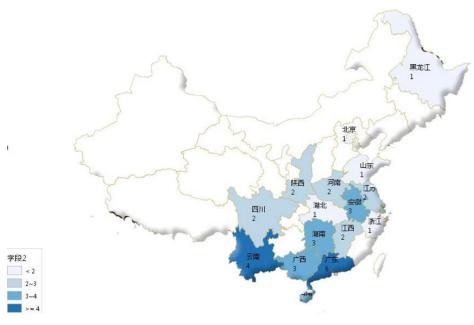


图 4-7 校园砍杀事件省份分布

4.2.3 校园砍杀事件的学校分布

校园砍杀事件的学校分布如图 4-8 所示,中学 16 起,大学 10 起,小学 7 起,幼儿园 4 起,其他 3 起。从校园发生校园的分布可以看出,中学的校园砍杀事件最多,原因一是中学的时间跨度最大,包括初中和高中 6 年,原因二是中学阶段,学生叛逆心理比较严重,容易冲动,导致砍杀事件最多。第二多的是大学,大学生离开家乡,各种行为更加自由,恋爱也使得感情更为复杂,导致砍杀事件较多。小学幼儿园砍杀事件少,主要原因是小学幼儿园的学生发生砍杀的能力较弱,主要是社会上人员造成的。其他的很少部分主要包括职高、专科学校存在的校园砍杀事件。

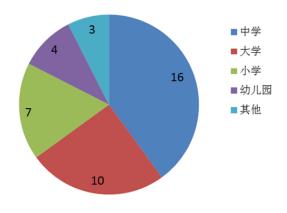


图 4-8 校园砍杀事件学校分布

4.2.4 校园砍杀事件死亡以及受伤人员统计

图 4-9 和图 4-10 分别为校园砍杀事件死亡人数和受伤人数的事件数目的统计,从图中可以看出,随着死亡和受伤人数的增加,相应的事件数目在减少,说明重大校园砍杀事件发生的少,轻微校园砍杀事件发生的多。在受伤人数最多的那起,共有 23 人受伤,这次事件的原因是,临近所谓的"世界末日",发生的一起无差别砍杀事件,这次砍杀没有目的,作案者目的是砍杀而没有具体砍杀目标,因此受伤人数特别多。相对无差别砍杀事件,具有明确砍杀目标的砍杀事件一般死亡人数和受伤人数特别少,一般只有一人受到伤害或者死亡。

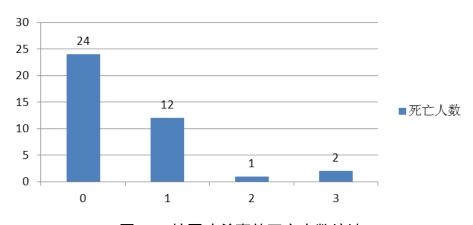


图 4-9 校园砍杀事件死亡人数统计

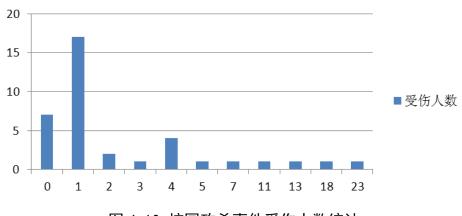


图 4-10 校园砍杀事件受伤人数统计

4.2.5 校园砍杀事件原因

校园砍杀的原因总结如图 4-11 所示,主要包括以下几种:矛盾纠纷、精神

病(无差别砍杀)、其他(因阻挡砍杀、多年宿怨,砍错人等)、原因不明等。其中砍杀原因不明的事件最多,主要是该事件正在调查中,事件起因还不明确。因 矛盾纠纷而起的事件有 15 起,无差别砍杀主要是因精神病等造成的

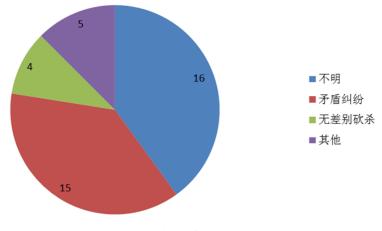


图 4-11 校园砍杀事件原因

4.2.6 校园砍杀事件总结

从上述分析中可以看出校园砍杀事件的发生主要存在几个特点:时间上,返校离校较多;地理上,南方较多;严重的较少,轻微的较多;发生的原因无差别砍杀较多。

此外结合死伤人数以及校园砍杀事件起因两个方面来看,在发生较多死伤人数的事件中,起因更多的为无差别砍杀。其中最大的一起校园砍杀事件是,2012年12月14日,河南光山学生被砍案,该案共砍伤学生23人,是作案人受世界末日论影响造成的一起校园砍杀事件。另外还有2012年9月21日广西平南一名精神病患者闯入托儿所砍杀小学生事件,死亡3人,受伤13人;2013年5月21日,广东茂名6名小学生被砍凶手有精神病史事件。

4.3 公交车爆炸事件

2011年5月至2014年4月,从收集到的新闻集合统计,三年间共发生公交起火或者爆炸事件18起,统计表见附录3。

4.3.1 公交车爆炸事件的地点分布

统计公交起火爆炸事件在全国各省市的分布,见图 4-12,此类事件主要分布 在中国东部和南部省市,这与中国东部和南部省市人口比较密集,公共交通比较 发达,公交车辆较多是密切相关的。



图 4-12 全国各省市公交起火爆炸事件分布图

4.3.2 公交车爆炸事件原因

此类事件中,由于人为原因(如纵火、点火炮等)引发的有4起,占比22%,由非人为原因(如自燃、机器故障等)引起的有14起,占比78%。其中自燃引起的有5起,占比28%,见图4-15。自燃等非人为因素引起的14事故中,只有一起(事件4)造成1人受伤。但是,在三起纵火事件中,厦门BRT公交纵火案造成47人死亡,34人受伤;贵阳公交车纵火案造成6人死亡,35人受伤。由此可见,非人为原因造成的公交起火爆炸带来微小的人员伤亡,而人为的公交纵火事件造成了严重人员伤亡,伤亡人数统计见图4-13。因此,为了保障社会公共安全,应该严厉管控公交纵火之类事件,加强监察携带易燃易爆危险品上车的力度。在财产损失方面,起火爆炸事件都造成了事故公交的损毁。

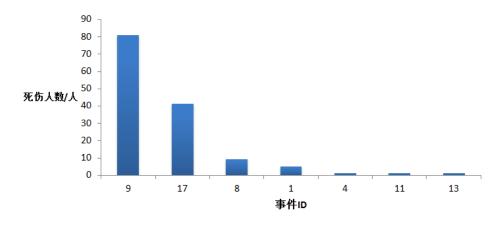


图 4-13 伤亡人数统计

附录统计表 1.1 中,通过作案人数,可以发现,公交纵火的一般都是个人行为。对非人为纵火公交起火事件发生的时间季节性进行分析,见图 4-14,可以发现自燃之类的公交起火事件并不是集中发生在天气炎热的夏季,反而在春季和秋季的发生事件数相对更多,由于时间事件数较少,从中发现季节性规律也是不可靠的。

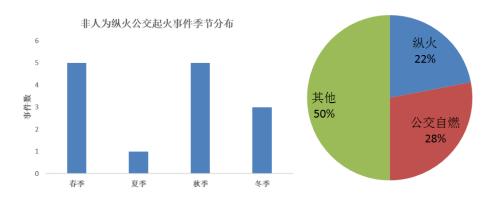


图 4-14 非人为纵火公交起火事件季节性分布图 4-15 公交起火爆炸事故原因统计

五、微博情感分析

微博作为信息时代的最活跃的大众信息交互发布平台之一,在社会事件发生时,既可以看作媒体对象,又可以作为公众舆论对象。考虑到微博"微"和"多"的特点,进行微博具体内容的深入挖掘,投入产出比不高。而进行微博极性(积极、消极、中性)的分析,能够充分利用数量优势在统计上得到舆论情感倾向。

基于以上原因,本文在预处理阶段给微博打上了情感标签,在这一章用极性 微博结合微博数量等基本特征,对暴恐事件进行了较为深入的分析。

在这一部分,首先对暴恐事件和微博数量、舆论情感倾向的关系进行了分析,对特例进行阐述,尝试对统计结果进行解释。紧随其后,分析了暴恐事件发生之后,情感平复的过程,也发现了非常有趣的现象,并尝试对现象进行了解析。

5.1 微博数量分析

首先统计了每天有关暴恐事件的微博的数量,并和整理出的事件进行对比,得出了微博尖峰和事件具有一致性的结果。在给定数据的 1125 天内,每日有关暴恐的微博数量变化如图所示,蓝色的"x"标出了之前通过对新闻挖掘获得的事件。

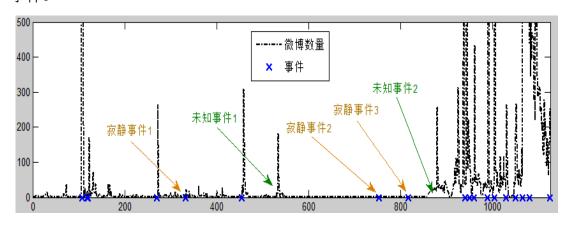


图 5-1 每日微博数量和事件对照

从图 5-1 中可以看到,之前找到的事件大部分都激起了微博数量的尖峰。但是,依然有没有标记事件的尖峰(本文称之为"未知事件");或事件发生却没有伴随微博数量尖峰的情况(本文称之为"寂静事件")。下面把这些事件罗列出来。

寂静事件:

寂静事件 1: 2012. 2. 28 新疆喀什市叶城暴力恐怖事件

寂静事件 2: 2013. 4. 24 新疆巴楚县发生严重暴力恐怖案件

寂静事件 3: 2013. 6. 26 新疆鄯善发生暴力恐怖袭击案件

根据所得结果分析, "寂静事件"产生的原因, 很可能是政府主动施加影响的结果。

未知事件:

未知事件 1: 2012.9 中国各地爆发反日游行

未知事件 2: 2013. 8. 27 微博内容大致为"新疆行动击毙十五人,警 方牺牲一人。官媒没报道"

分析微博数据得知,未知事件1是因为各地爆发的反日游行,在网络上人们拿这件事和暴恐事件类比,因而产生含有关键词"暴恐"的微博尖峰。而未知事件2则是因为,人们对之前有关暴恐事件的报道隐藏了信息非常不满,所以互相转发造成了微博数量尖峰。

综上,事件和微博数量尖峰具有一致性,即可以说事件激发微博数量尖峰。而不满足这一特性的原因有两点,一是因为舆论外部强制力对舆论的干预,二是舆论内部对外部强制力的反抗。

5.2 微博极性分析

5.2.1 积极和消极情感的爆发对称性

统计随时间变化的极性微博的数量和事件的发生进行对照,发现暴恐事件的发生激发微博数量尖峰的同时,也造成了极性微博数量的爆发。我们统计了数据覆盖的 1125 天内,情感微博数量的随时间的变化,得出了积极和消极情感同时爆发的特性,本文称之为情感爆发对称性。统计结果如图 5.1.2 所示,数轴上方是积极微博条数(绿线),下方是消极微博条数(红线虚线)。同样的,用蓝色的"x"在时间轴上标记出了事件。

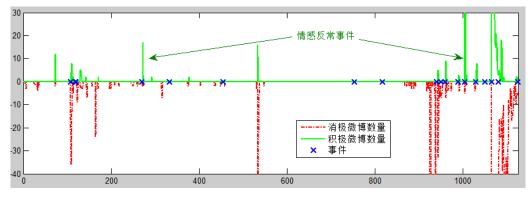


图 5-2 每日极性微博数量和事件对照

从图 5-2 可以看到, 恐怖事件的发生导致积极/消极微博数量产生尖峰, 且爆

发时间基本一致,即上下的尖峰总是几乎同时出现。

5.2.2 消极情感主导

从图 5-2 也可以看到,同时爆发的积极和消极微博数量尖峰对,虽然几乎同时爆发,但是消极微博数量总是远大于积极微博数量。因此可以说恐怖事件的发生,导致舆论被消极情感主导。

5.2.3 无辜人员伤亡影响感情倾向

在图 5-2 中,有两个事件(已标出:情感反常事件)不满足消极情感主导的特点,事件发生后积极微博数量远远大于消极微博的数量,因而对其进行了关注。 下面列出积极情感主导的情感反常事件:

情感反常事件 1: 2011. 12. 29 新疆公安机关打掉一暴恐团伙

情感反常事件 2: 2013. 12. 30 新疆喀什市莎车县暴力恐怖袭击事件

对比了两个事件,发现两个暴恐事件的共性是由团伙作案,且警方取得重大胜利。但对细节的分析揭示出了更多的线索。

情感反常事件 1 由 11 人作案,警方当场击毙 7 人,4 人被捕。警方 1 人遇难, 1 人受伤。媒体的报道口径正面例,如"公安机关打掉一暴恐团伙"。

情感反常事件 2 由 9 人作案,警方当场击毙 8 人,1 人被捕。警方没有人员伤亡。媒体报道口径中性,如"莎车县发生暴力恐怖袭击事件"。

协同其他负面情感微博数占优但不占绝对优势的事件分析,基本可以归纳出 无辜人员伤亡数同积极情感微博数量成反比的结论。因而进一步的,无辜人员伤 亡影响舆论情感倾向。

5.2.4"报道口径"影响情感波动

在得出上述结论后,放大了情感反常事件 1 和情感反常事件 2 的极性微博数量变化图局部,其变化如图 5-3,左侧是报道口径正面的事件 1,可以看到在事件发生的当天,极性微博数量为 0;事件发生第二天,积极微博数为 0,消极微博数有小尖峰;事件发生第三天,消极微博数为 0,而与此同时积极微博数达到

一个大尖峰。图 5-3 右侧是事件 2 的情感变化曲线,同事件 1 形成鲜明对比的是, 在事件发生的当天,积极和消极微博的数量就达到了尖峰。

对比两事件,可以从报道口径推测原因。对事件1采用正面口吻,因为有民警遇难,想要积极的舆论情感导向;而对事件2采用常规中性口吻,则是因为其不需要报道口径的帮助就可以得到积极的舆论情感回馈。

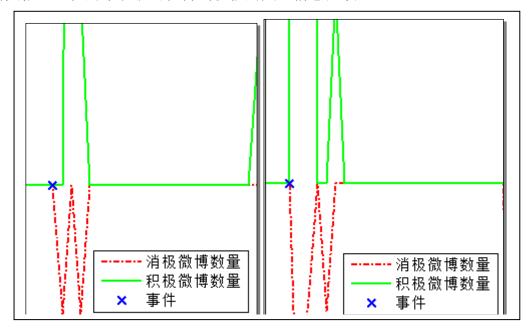


图 5-3 情感反常事件 1 和 2 的极性微博数量变化细节

5.3 微博情感回归分析

因为分析得到的结果显示,暴恐事件里无辜人员的死亡会影响舆论情感倾向, 所以按照恐怖事件的无辜人员死亡人数,将事件划分成三个层次,研究 了事件发生之后,舆论情感如何回归常态的过程。

根据无辜人员死亡人数定义三种暴恐事件等级:

一级:事件中没有无辜人员死亡

二级:事件中无辜人员死亡的个数少于10人

三级:事件中无辜人员死亡的个数大于10人

划分事件等级之后,本文分析了从事件发生的当天起 40 天内,舆论感情渐渐回归常态的过程(极性情感的数量趋于 0)。下图(图 5-4)是三种等级的暴恐事件,从事件发生开始 40 天内的情感微博数量变化。

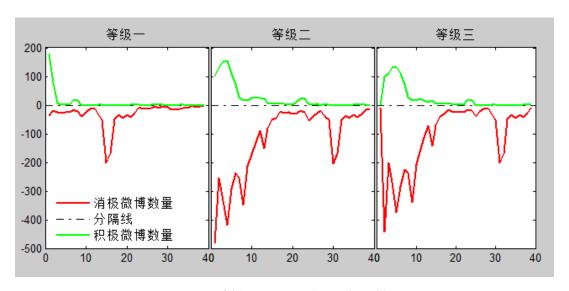


图 5-4 三种等级暴恐事件对应的情感平复曲线

对比图 5-4 左图和中图,可以看到暴恐事件有无无辜人员死亡,对舆论情感的影响巨大。没有无辜人员死亡的暴恐事件发生,大众舆论以积极情感(祈福,表达希冀,关注事件处置进展)为主导,两周左右后有一次消极情感尖峰,应该是事件的后续处理带动舆论对暴恐事件进行谴责。

对比图 5-4 中图和右图,可以看到暴恐事件死亡人数多和少对舆论情感波动的影响并不明显。如果死亡人数众多,那么事件发生的开始微博会受到管控(可以看到图 5-4 右图,在事件发生的当天甚至没有积极或消极的微博出现)但是情感后续恢复常态的过程和死亡人数较少的暴恐事件差距不大。

进一步的支持本文的结论,又统计了事件发生后,情感恢复常态过程中积极 (消极)微博在极性微博总数中占有比重的情况,如图 5-5 所示。

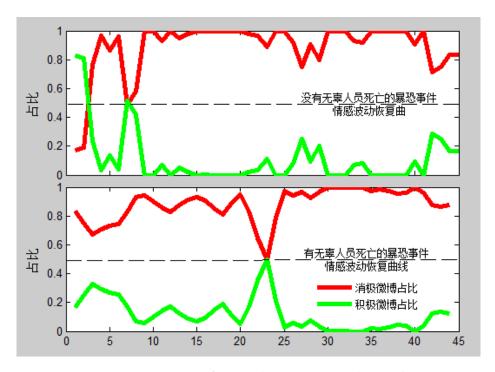


图 5-5 积极和消极微博占有极性微博比重变化

从图 5-5 依然可以得到一样的结论,即暴恐事件中有无无辜人员死亡对舆论 情感倾向波动及恢复有显著影响。

在图 5-5 中还可以观察到另外一个特点,即在事件发生一段时间后(图 5-5 中,上图为一周左右,下图为三周左右),积极情感微博总会达到和消极微博一样的比重,之后又会回到消极微博占优的状态。本文猜测这个波动的出现可能有以下原因:

原因一,事件的处置。因为没有无辜人员死亡的暴恐事件相对规模较小,造成的伤害和影响都比较小,或者是由警方抢先控制了局面。这样的事件相对来说处置的时间非常短,因而在一周左右的时间进行处置信息公布,造成积极情感微博增多。而相应的,较为严重暴恐事件需要的处置时间更长,因而这个对冲点到达的时间更晚。

原因二,影响力。结合图 5-4 可以看到在相应的时间段(无无辜人员死亡为一周左右,有无辜人员死亡为三周左右)内,积极和消极微博的数量都下降到接近 0 点,因而这个对冲点,也说明了时间的影响力。

5.4 微博分析小结

在本节中本文首先分析每日微博数量变化和事件的相关性,发现了微博数量尖峰和事件的一致性;相应的分析每日极性(消极/积极)微博数量的变化和事件的相关性,发现微博积极消极情感爆发具有对称性,且暴恐事件的舆论由消极情感主导;进一步的,通过分析数据趋势和案例归纳,发现无辜人员死亡和媒体的报道口径影响舆论的情感变化。在得到这一结论后,又具体的研究了,有或没有无辜人员死亡的暴恐事件发生后,群众情感回归常态的过程;发现有或没有无辜人员死亡对舆论情感回归曲线影响巨大,并且无无辜人员死亡的暴恐事件影响力只持续一周,而有无辜人员死亡的暴恐事件影响力可以持续三周。

六、建模分析

这一部分本文的目的是建立模型,对暴力恐怖事件进行地点预测、对暴恐事件影响传播进行拟合,以及对"无差别杀人"进行心理学模型解析。

6.1 暴恐预测模型

对新疆暴恐事件发生地点按照行政地区划分,得到如表 6-1

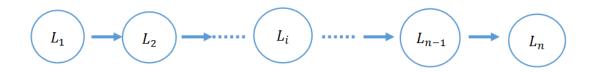
表 6-1 新疆暴恐事件时间地点表

事件地点P _i	时间T _i	间隔 $T_i - T_{i-1}$
4.11 短火1	H1 IH1 I	天
和田	2011/7/18	0
喀什	2011/7/30	12
喀什	2011/7/31	1
和田	2011/12/28	150
喀什	2012/2/28	62
和田	2012/6/29	122
喀什	2013/4/23	298

吐鲁番	2013/6/26	64
喀什	2013/11/16	143
喀什	2013/12/15	29
喀什	2013/12/30	15
阿克苏	2014/1/24	25
阿克苏	2014/2/14	21
伊犁哈萨克自治州	2014/3/2	16
乌鲁木齐	2014/3/17	15
乌鲁木齐	2014/4/30	44

6.1.1 建立模型

建立概率图转移模型, L_i 表示第i个事件发生的地点



和田→喀什→喀什→和田→喀什→和田→喀什→吐鲁番→喀什→喀什→→喀什 →阿克苏→阿克苏→伊利哈萨克自治州→乌鲁木齐→乌鲁木齐

每一个事件作为链的一个节点,节点的值 V_i 由以下公式 6.1 定义

$$V_{i} = \begin{cases} 1 & \text{if } \exists j \ (j < i) \& (L_{i} = L_{j}) \& \left(\left(T_{i} - T_{j}\right) < 90\right) \\ 0 & \text{others} \end{cases}$$

$$\sharp \Phi: \tag{6.1}$$

 V_i :第i个事件节点的值

 L_i :第i个事件节点发生地点

 T_i : 第i个事件节点发生时间

定义某暴恐事件发生后,三个月内会在同一地域(行政地区)发生暴恐事件

的概率为 P_a , 计算方法如公式 6.2。

$$P_a = \frac{\#(V_i = 1)}{n} \tag{6.2}$$

其中:

 P_a 三个月内发生在同一地域发生暴恐事件的概率

= $(Number\ of)$, # $(V_i = 1)$ 表示 $(V_i = 1)$ 的节点点数

n节点总数,即新疆所有暴恐事件数

6.1.2 模型求解

根据以上模型定义,得到各事件对应的值 V_i ,见表 6-2 表 6-2 事件节点取值表

L_i	T_i	V_{i}
和田	2011/7/18	0
喀什	2011/7/30	0
喀什	2011/7/31	1
和田	2011/12/28	0
喀什	2012/2/28	0
和田	2012/6/29	0
喀什	2013/4/23	0
吐鲁番	2013/6/26	0
喀什	2013/11/16	0
喀什	2013/12/15	1
喀什	2013/12/30	1
阿克苏	2014/1/24	0
阿克苏	2014/2/14	1
伊犁哈萨克自治州	2014/3/2	0
乌鲁木齐	2014/3/17	0
乌鲁木齐	2014/4/30	1

三个月内会在同一地域(行政地区)发生暴恐事件的概率为 P_a

$$P_a = \frac{\#(V_i = 1) = 5}{n = 16} = \frac{5}{16} = 0.31$$

6.1.3 模型验证

通过网络新闻报道,收集得到 2014 年 5 月—2014 年 7 月新疆发生的暴恐事件,如表 6-3:

暴恐事件	地点	时间
5・22 乌鲁木齐爆炸案	乌鲁木齐	2014/5/22
2014年和田市砍杀事件	和田	2014/6/15
2014年叶城公安局袭击事件	喀什	2014/6/21
7 • 28 新疆喀什市莎车县暴恐袭击案	喀什	2014/7/28

表 6-3 2014 年 5 月—7 月新疆发生的暴恐事件表

2014年4月30日,在乌鲁木齐发生了一件暴恐事件,因此可以预测,三个月内乌鲁木齐再发生一起暴恐事件的概率为0.31,准确预测了"5•22乌鲁木齐爆炸案"。

此模型结果反映了,在某地发生暴恐事件后,政府需要特别关注此地的公共安全,有 0.31 的可能在案发后的三个月内同一地区再发生暴恐事件。

6.2 传播模型

为了进一步了解一个群体事件在网络的传播量的影响因素,特地的研究了暴恐事件的传播特地,并建立了传播模型。通过数据发现暴恐数据的新闻的传播量,不仅仅和某件事件本身的严重程度有关,还和之前事件的传播量有关。

假设一个事件的报道量不仅仅和当前事件的严重性还和前一个事件的严重 性和报道量有关。因此建立一个回归模型。

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + Y + \varepsilon$$

符号说明:

Y: 表示当前暴恐事件的新闻传播量,在此模型用该事件的新闻报道量表示。

X₁: 前一件暴恐事件的严重程度

X₂: 当前暴恐事件的严重程度

X₃: 前一件暴恐事件的新闻报道量

ε: 误差项

β 。: 表示偏置项

Y: 为了防止过拟合的正则化项

表 6-4 暴恐类事件传播量信息

事件序	前一件事	当前事	前一件事件	当前事件的				
号	件的严重	件严重	的报道量	报道量				
	程度	程度						
1	2	2	25	91				
2	2	2	91	93				
3	2	1	93	46				
4	1	3	46	38				
5	3	1	38	18				
7	1	3	18	4982				
9	3	4	4982	8566				
12	4	2	8566	1770				
13	2	2	1770	285				
14	2	2	285	305				
15	2	1	305	1198				
16	1	2	1198	655				
17	2	1	655	1132				
18	1	5	1132	10029				
19	5 1		10029	511				

用本文抽取的暴恐信息数据,对模型进行求解,得到结果 $\beta_0\beta_1\beta_2\beta_3$ 的值分别为: -602.8, -905.0, 1868.1, 0.4, 说明了暴恐事件类型在网络上的新闻传播

量和暴恐事件本身的严重程度及上一次暴恐事件的传播量成正相关,本次暴恐事件严重程度越大,上次暴恐事件传播量越大,本次传播量越广。然而,本次暴恐事件的传播量和上次暴恐事件的严重程度成反相关,上次暴恐事件越严重,反而降低本次暴恐传播量。

6.3 无差别犯罪心理模型

题目中的三类恐怖事件,校园砍杀和公交车爆炸案主要产生于"无差别杀人事件"[4][5]。虽然其他原因也会造成校园砍杀和公交爆炸,但其中"无差别杀人"占比大,影响恶劣。

一些心理学的分析认为,"无差别杀人事件"是属于"反常自杀"的一种。"反常实际上产生一种激怒和厌烦的状态,这种状态根据不同的情况可以转而针对自己或针对他人:在前一种情况下会引起自杀,在后一种情况下会引起杀人"[6]

简单说,人在遇到巨大压力或者困境会进入一种反常状态,处于这个状态的 人可能会选择自杀,或者杀人。这是一个简单的模型。

无差别杀人的手段形式不仅有公交车爆炸和校园砍杀,马路砍杀和开车撞人也是常见的手段。由于其"犯罪工具选择任意性强"^[4]的特点,我认为如果孤立的看待各类事件,可能更加难以发现规律。所以如果能有驾车撞人和马路砍杀的数据,可能分析起来更加完整。

有两个现象很有趣:

第一是季节分布。文献称夏季为自杀高发季节,自杀构成占到了 60[~]80%^[7],而根据本文的统计,公交车爆炸案件 60%分布在 6[~]8 月。

第二是年龄规律。在 38 起无差别杀人案件中,犯罪人年龄处于 30-50 岁这一阶段的共有 28 起,占 73.68%^[7]。而根据 95⁹9 中国自杀年龄统计,30⁵⁰ 是明显的自杀率低谷^[8]。

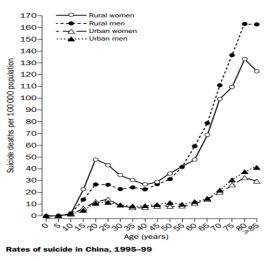


图 6-1 不同年龄自杀论分布图

两个现象提供了一个线索,如果根据前面提到的"正常状态→反常状态→自 杀 or 杀人"的模型。季节和年龄两个因素可能分别影响到了模型中的两个阶段。 (如下图所示)

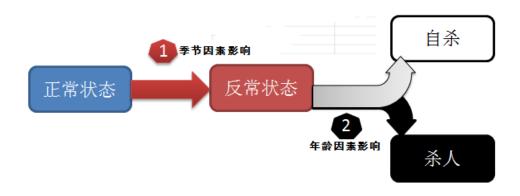


图 6-2 反常状态示意图

本文很感兴趣什么样的因素会增加进入反常状态的人数,什么因素会影响人选择杀人还是自杀。这里,本文希望能够得到一些月份级的自杀率,他杀率,离婚率,死亡率,失业率的数据,希望能够找到他们与"无差别杀人事件之间的联系"

对危害公共安全事件的预警和追踪,对于保护人民群众生命财产安全,维护社会稳定意义重大。几类公共安全事件之中有其共性与规律。

通过对数据的清洗,查找,描述。挖掘了事件性质,对事件时空特性进行了

描述,对新闻量,微博情感进行了分析。建立了统计模型。

未来的工作中,可以试着丰富数据类型,通过更加广泛的数据来挖掘危害公共安全事件与社会安定程度,人民幸福程度,国际环境,自然气候之间的关系,或许能够为预警恐怖事件提供更加有效的信息。

七、总结与展望

近年来各类危害公共安全的事件发案呈逐年上升趋势,给广大人民群众的生命和财产带来了严重损害,极大地影响着社会稳定和民族团结。了解各类危害公共安全的事件的发生特点及其在网络的传播效应对预防事件的发生和降低事件的危害性有着重要的作用。本文利用海量公司提供的新闻和微博数据针对 2011年到 2014年5月之间的暴恐事件、校园砍杀,公交车爆炸事件三类危害公共的安全事件进行了对发生时间、地点、诱因、犯罪模式、危害性等方面进行了分析。

通过对最近三年的这三类事件分析发现暴恐事件发生地点主要发生在新疆地区,而且暴恐事件的发生具有在同一地点连续发生的特点。根据这一特点,本文建立预测模型,以较大的概率预测了5月份的乌鲁木齐暴恐的发生。然而校园砍杀和公交车爆炸的发生模型地点呈现离散化,从数据驱动的角度上不具有可预测性。进一步对校园砍杀和公交爆炸这两类事件的犯罪原因分析发现,当罪犯属于无差别杀人时,造成案件的严重性远大于其他原因。"无差别杀人"的罪犯多是被边缘化的"失败者",值得注意的是大规模社会舆论的传播也会诱发这类犯罪,比如说2012年"河南光山校园杀人案"受"世界末日论"的影响。同时本文也对特定暴恐事件在网络的新闻传播量的影响因素进行了分析,发现这类事件的新闻传播量主要受事件本身的严重性,同时前一件的暴恐事件的严重程度和报道量也会对当前暴恐事件的新闻传播量影响。对微博情感分析的结果表明:从时间维度看,事件会激发微博数量尖峰;极性(消极/积极)微博数量的爆发具有对称性,且暴恐事件的舆论由消极情感主导;进一步的,发现媒体的报道口径影响舆论的情感变化,且案件中有无无辜人员死亡严重影响舆论情感的平复过程。

总之,通过本文的研究,主要得到如下结论:

1,在暴恐事件发生后,短期内该地发生暴恐的可能性很大。相关部门尤其 得注意严防暴恐事件二次发生的情况,从而降低暴恐事件的危害。

- 2,大规模社会舆论很容易诱发无差别杀人犯罪的产生。政府在社会发生重大创伤性事件,大规模恐慌期间尤其要注意疏导社会舆论,以降低无差别杀人的产生几率。
- 3,校园砍杀事件、公交爆炸事件的案发特点成地点离散化,不具有可预测性。针对此类事件,理应从加强安防措施来降低此类案件的影响。
- 4,事件在网络的传播量,不仅仅和事件本身的严重程度有关(严重程度越重网络传播量越大),而且还受前一个同类事件的严重程度(反相关)和传播量有关(正相关)。这一特点有助于政府在事件发生后提前估量其在网络的传播规模。
- 5,微博能够直接地反映大众针对危害公安安全事件情感态度。政府可以从 微博上了解社会的舆论,有助于疏导民众的不满情绪,从而达到事后干预降低事 件造成的影响。

最后,现有的自然语言处理工具虽然已经有了长足的进步,但是依然不能满足深入挖掘舆情的需求,同时考虑到本次赛题提供的数据源也比较有限,本文的工作还有很大的改进空间。未来的工作中,可以试着丰富数据,尝试新的机器学习方法,来挖掘危害公共安全事件背后更深层次的规律,为预警提供更加有效的信息。

八、参考文献

- [1] "结巴"中文分词, https://github.com/fxsjy/jieba
- [2] 周胜臣, 瞿文婷, 石英子, 等. 中文微博情感分析研究综述[J]. 计算机应用与软件, 2013, 30(3):161-164.
- [3] Alias-i. 2008. LingPipe4. 1. 0. http://alias-i.com/lingpipe (accessed October 1, 2008)
- [4] 无差别杀人是一种反常自杀,http://www.fyfzw.cn/a/fangyuan/20130902/2013/1010/51181.html
- [5] 张小虎. 我国无差别杀人犯罪的现实状况与理论分析[J]. 江海学刊, 2011, 1: 020.
- [6] 爱米尔. 杜尔凯姆[J]. 自杀论, 1988, 166: 23.
- [7] 吕琳, 曾小燕. 自杀的流行病学研究[J]. 中国现代医学杂志, 2002, 12(1): 104-105.
- [8] Phillips M R, Li X, Zhang Y, et al. Suicide rates in China[J]. The Lancet, 2002, 359(9325): 2274-2275.

九、附录

附录 1. 统计显著最有差异的 300 个词

暴恐事件与校园砍杀事件最显著有差异的 100 词:

反,追,家人,民族,感情,报警,全国,倒,严厉,巨响,中学,活动,门口,残忍,孩子,原因,放学,大局,死,祖国,刺,杀人案,烟,大学,斗争,家庭,打开,当晚,火车站,冒,打,上课,充分,对方,拿,焚烧,上车,校门,性格,财产,救援,实现,尸体,浓烟,关系,当前,和谐,同班,消防,死刑,带,良好,能力,进,项目,父亲,干部,了解,自治区,矛盾,就读,车,一起,还,新,想到,贡献,男生,寻找,暴徒,女,击伤,病史,听到,认为,停车,态势,闻到,人民法院,部署,暴恐,直到,代表,6月,头部,放火,知道,思想,校方,暴行,新闻办,以为,人类,给,只是,窗户,害怕,逃离,局面,少年

暴恐事件与公交爆炸事件最显著有差异的 100 词:

反,追,家人,民族,感情,报警,全国,倒,严厉,巨响,中学,活动,门口,残忍,孩子,原因,放学,大局,死,祖国,刺,杀人案,烟,大学,斗争,家庭,打开,当晚,火车站,冒,打,上课,充分,对方,拿,焚烧,上车,校门,性格,财产,救援,实现,尸体,浓烟,关系,当前,和谐,同班,消防,死刑,带,良好,能力,进,项目,父亲,干部,了解,自治区,矛盾,就读,车,一起,还,新,想到,贡献,男生,寻找,暴徒,女,击伤,病史,听到,认为,停车,态势,闻到,人民法院,部署,暴恐,直到,代表,6月,头部,放火,知道,思想,校方,暴行,新闻办,以为,人类,给,只是,窗户,害怕,逃离,局面,少年,悲剧,基础,刑事拘留,车身,遂,挑战,检察院,想,发生,见,说法,强烈,却,涉嫌,合作,闯入,车尾,精神,当,犯罪分子,赶到,看到,火灾,16日,内向,车子,晚上,发展,相信,东西,37分,国务院,剩下,运营,力量,打架,争吵,规模,中院,

住,驾驶员,心,经,遭到,出,党委,嚣张气焰,紧急,高中,犯罪嫌疑人,傍晚,外出,预谋,搏斗,窗,师傅,怀疑,战略,母亲,保持,冒烟,两会,4月,深化,造成,后门,批准,反对,截至,见到,和平,逮捕,不是,令人发指,得知,做,让,陈,遇难者,读,会议,学,团伙,抓,衣服,水,学院,民生,伤亡,抓获,勘查,受害人,同志,判决,任务,坚定,打工,领域,教室

公交爆炸事件与校园砍杀事件最有显著差异的100词:

学校,恐怖,学生,暴力,公交,乘客,杀人,稳定,小学,司机,行 凶,起火,砍伤,共同,扑灭,同学,维护,家长,砍,警方,恐,爆 炸,下午,群众,燃烧,火势,暴,坚决,杀害,国际,案发,时,打 击,大火,人民,平时,判处,校园,车上,持,曾,势力,纵火,化 名,行驶,作案,没有,事故,形势,审理,老师,国家,伤员,疏散, 指出,地区,统一,案情,杀死,调查,3月,车辆,觉得,分裂,杀, 离家,击毙,事发,教育,找到,纵火案,校长,上学,凶手,现场,父 母,附近,开庭,突然,中级,当时,男子,发现,过,自燃,小学生, 1日,分子,车厢,逃生,快速,找,反恐,称,行径,儿子,破坏, 但,主席,烧,接到

附录 2.暴恐事件统计表

事件工	作案时间	事件名称	事件地点	警员死亡	警员受伤	歹徒死亡	平民死亡	平民受伤	作案人對	逃离歹徒	对象	作案方式	作案地点	资讯数量
1	2011/7/18	7・18新疆和田暴力恐怖事件	新疆和田	1	2	14	5	5	18	0	公务人员	砍杀	派出所	25
2	2011/7/30	新疆喀什市汽车冲撞人群暴力恐怖案件	新疆喀什市	0	0	1	7	28	2	0	平民	砍杀	街道	91
3	2011/7/31	新疆喀什市恐怖分子袭击餐厅	新疆喀什市	0	3	5	4	12	7	2	平民	砍杀	餐厅	93
4	2011/12/28	新疆公安机关打掉一暴恐团伙	新疆和田皮山县	1	1	7	0	0	11	0	平民		Щ⊠	46
5	2012/2/28	2 · 28新疆喀什市叶城暴力恐怖事件	新疆喀什市叶城	1	4	8	15	14	9	0	平民	砍杀	街道	38
6	2012/6/6	新疆和田爆炸案	新疆和田	0	0	0					平民	爆炸		
7	2012/6/29	新疆劫机案	新疆和田	0	4	0	0	0	6	0	平民	劫机	飞机	18
8	2013/3/7	新疆库尔勒市袭击事件	新疆库尔勒市	0	0	0	4	8			平民	砍杀		
9	2013/4/23	新疆巴楚县发生严重暴力恐怖案件	新疆喀什市巴楚县	12	2	6	0	0	14	0	公务人员	纵火、砍杀	居民家	4982
10	2013/4/25	新疆和田中学保安被杀纵火焚尸案	新疆和田	0	0	0	2	0			平民		中学	
11	2013/6/20	新疆墨玉县杀害施工人员	新疆墨玉县	0	0	0					平民		工地	
12	2013/6/26	新疆鄯善发生暴力恐怖袭击案件	新疆鄯善	2	0	11	24	21	16	1	公务人员、平民	纵火、砍杀	镇政府、民工工地	8566
13	2013/10/28	10·28北京冲撞金水桥暴力恐怖袭击案	北京	0	0	0	2	40	3	3	平民	汽车冲撞	街道	1770
14	2013/11/16	11·16新疆喀什巴楚暴力恐怖袭击	新疆喀什巴楚	2	2	9	2	2	9	0	公务人员	砍杀	派出所	285
15	2013/12/15	?・15新疆喀什市疏附县暴力恐怖袭击事	新疆喀什疏附	2	0	14	0	0	16	0	公务人员	爆炸、砍杀	居民家	305
16	2013/12/30	?・30新疆喀什市莎车县暴力恐怖袭击事	新疆喀什刹车	0	0	8	0	0	9	0	公务人员	ま炸、纵火、砍 ź	公安局	1198
17	2014/1/24	新疆新和县爆炸案	新疆新和县	0	1	12	1	2	17	0	平民	爆炸	街道	655
18	2014/2/14	新疆乌什县袭警案	新疆乌什县	0	2	11	0	2	12	0	公务人员	爆炸、砍杀	街道	1132
19	2014/3/1	3·1昆明火车站暴力恐怖案	昆明	0	0	4	31	141	8	0	平民	砍杀	火车站	10029
20	2014/3/2	3·02新疆伊宁市故意杀人案	新疆伊宁市	0	0	0	1	1	1	1	平民	砍杀	社区	
21	2014/3/17	乌鲁木齐发生袭警事件	新疆乌鲁木齐	0	1	0	0	1	1	0	公务人员	砍杀	社区	
22	2014/4/30	4·30乌鲁木齐火车南站暴力恐怖袭击案	新疆乌鲁木齐	0	0	2	1	79	2	0	平民	砍杀、爆炸	火车站	
23	2014/5/6	5·6广州火车站暴力袭击事件	广州	0	0	0	0	6	1	0	平民	砍杀	火车站	
24	2014/5/22	5・22乌鲁木齐爆炸案	新疆乌鲁木齐	0	0	4	39	94	5	1	平民	爆炸	街道	
25	2014/6/15	2014年和田市砍杀事件	新疆和田市	0	0	2	0	4	3	0	平民	砍杀	室内	
26	2014/6/21	2014年叶城公安局袭击事件	新疆叶城	0	3	13	0	0		公	务人员、警务设	爆炸、汽车冲撞	公安局	
27	2014/7/28	7・28新疆喀什市莎车县暴恐袭击案	新疆喀什莎车县			59	37	13	215		公务人员	纵火、砍杀	派出所、镇政府	
28	2014/9/21	2014年轮台县爆炸案	新疆轮台县	4		40	10	54	42	0	公务人员、平民	爆炸、砍杀	派出所、农贸市场	
29	2014/10/12	10·12巴楚暴力恐怖袭击事件	新疆喀什巴楚				22		4		平民、公务人员	砍杀	街道	

附录 3.校园砍杀事件统计表

作案时间 🕶 首次报道时 🔻	事件名称 ▼	事件地点▼死	亡/▼ 受伤.	从 ▼ 6	▼幹領∓	作案人类▼	对1~	žþ: 🔻	作案地▼	省份▼	时间
2007-7-20 2012_04_27	疯汉闯幼儿园连砍19人 女老师斗歹徒受伤	广东顺德	0	18	成年人	校外精神病	学生	不明	幼儿园	广东	7
2010-3-30 2013_01_11	川大学生因丑杀人案重审 是否患精神病成焦点	四川成都	1	2	成年人	学生	学生	因丑遭	大学	四川	3
2011-4-11 2011_04_12	老师砍女学生数刀 自己跳楼身亡	淮安市	0	1	成年人	老师	学生	不明	中学	江苏	4
2011-4-17 2011-4-18	北师大珠海分校发生校园血案 大三学生砍伤同学	广东珠海	0	1	成年人	学生	学生	矛盾纠	大学	广东	4
2011-6-7 2011_06_10	云南高考生砍人事件	云南曲靖	0	1	未成年	学生	学生	不明	中学	云南	6
2011-9-15 2011_09_16	一女子持刀进幼儿园砍伤一老师砍人者与被砍者	云南昆明	0	1	成年人	女子	女老师	不明	幼儿园	云南	9
2011-9-21 2011_09_23	小学生校园被砍致眼睛遭摘除 砍人者或是其邻	宿州符簋	0	1	成年人	邻居	学生	家庭矛	小学	安徽	9
2011-9-29 2011_09_30	"变态男"夜闯女生宿舍砍人	广西	0	4	成年人	校外人员	学生	感情	大学	广西	9
2011-10-12	新州镇中学6名学生遭到一伙男子持刀追砍	海南省儋州市	0	4	未成年	校外,团伙	学生	不明	中学	海南	10
		南京			未成人	校外人员	学生	不明	中学	江苏	10
2011-10-23 2011-10-24	年仅15岁的儋州初二学生邓某开在新州镇中心学	海南省儋州市	1		未成年	团伙		不明	中学	海南	10
2011-10-28 2011_10_30	广州发生校园砍人案 大三男生持菜刀砍死室友	广州	1	0	成年人	学生	学生	室友矛	大专	广东	10
2011-11-19	,临高县一名16岁在校学生为躲避他人追砍,趴	海南省儋州市	1		未成年	团伙		不明	校外	海南	10
2012-3-6 2012_03_06	阿城第三中学现校园暴力 两月内初一学生砍伤:	黑龙江省哈尔	0	1	未成年			不明	中学	黑龙江	11
	家庭纠纷女子在幼儿园发狂 持刀砍伤儿童被捕		0		成年人				幼儿园	江西	3
2012-4-10 2012_04_19	光天化日翻墙进校园 七八人在南昌县职高砍伤-	南昌	0				学生		职高	江西	4
2012-7-12 2013_10_21	17岁少年校内割喉杀人 称想扮杀手追求刺激	北京	1	0	成年人	学生	学生	想扮杀	大学	北京	7
	广西平南一名精神病患者闯入托儿所砍杀小学生		3					精神病		广西	9
2012-9-26 2012_09_27	兰州一大学数十人夜闯宿舍行凶 多名学生被砍1	兰州	3		成年人				大学	兰州	9
	血流满地! 武汉大一女生遭前男友割喉身亡: 行	武汉	1			校外人员			大学	湖北	11
2012-12-11 2012_12_26		珠海	0			社会青年			大学	广东	12
2012-12-14 2012_12_14		河南光山县	0		成年人			世界末	小学	河南	12
		安徽合肥	1		成年人				大学	安徽	12
	歹徒翻墙入校砍伤多名学生 教师挺身勇擒凶		0			社会青年				山东	12
		上海	0		成年人			平时有		上海	1
	湖南新化发生一起校园惨案 校长遇难一学生受1		1			社会青年				湖南	1
	男子持刀冲进幼儿园 女园长将其抱住被砍两刀		0						幼儿园	浙江	1
	河南发生校园伤害案 两名14岁嫌犯进校砍伤学		0				学生			河南	1
	高一学生在校门口被砍成重伤 目前未脱离危险		0				学生		高中	云南	3
	上海奉贤一男子砍杀两人后袭击小学 多人受伤		2			社会青年			小学	上海	3
	湖南衡阳8名青少年持斧闯进校园砍伤4名学生		0						中学	湖南	5
		广东茂名	0				学生+3			广东	5
	高中男生毕业典礼上持刀砍伤同窗 曾有精神病!		0					精神病		广东	6
	因多年前宿怨 广西永福—中学生在校园杀害校-		0		成年人		校长			广西	7
		南充市仪陇县	0		成年人		老师			四川	9
	高二男生在校门口被砍十余刀 疑被误当成他人		0		成年人			误当成		安徽	9
		云南富民县	1		未成年				初中	云南	12
		湖南怀化	1			社会青年			大学	湖南	3
	陕西一中学高二男生刺死女同学 警方介入调查		1		未成年			不明	中学	陕西	3
2014-4-15 2014-4-18	安康一中学十余学生放学路上遭陌生人刀砍 1死	陕西安康	1	2	成年人	社会青年	学生	不明	中学	陕西	4

附录 4.公交车爆炸事件统计表

公交起火爆炸事件统计表											
事件ID	时间	事件名称	事件地点	原因	损毁车辆	死亡人数	受伤	作案人数			
1	2011/5/4	温州公交车场起火爆炸	温州		3	0	0				
2	2011/9/7	哈尔滨公交行驶中自燃	哈尔滨	自燃	1	0	0				
3	2012/7/28	北京市9路公交车纵火案	北京	纵火	1	0	0	1			
4	2013/1/8	郑州92路公交自燃	郑州	自燃	1	0	1				
5	2013/1/28	哈尔滨公交车起火	哈尔滨		1	0	0				
6	2013/3/11	桐城公交车行驶中失火	桐城		1	0	0				
7	2013/4/4	杭州公交车自燃	杭州	自燃	1	0	0				
8	2013/5/27	广东东莞公交车爆炸	东莞	零件老旧	1	0	9				
9	2013/6/7	厦门BRT公交纵火案	厦门	纵火	1	47	34	1			
10	2013/7/9	杭州城区一公交车自燃	杭州	自燃	1	0	0				
11	2013/9/19	成都98路公交车上点火炮	成都	点火炮	0	0	1	1			
12	2013/10/8	北京公交自燃	北京	自燃	1	0	0				
13	2013/10/18	湖北襄阳公交车失火	襄阳		1	1	0				
14	2013/10/25	广东韶关公交车起火	韶关		2	0	0				
15	2013/12/10	公天津交车起火	天津		2	0	0				
16	2014/2/27	宁波公交车起火	宁波		1	0	0				
17	2014/2/27	贵阳公交车纵火案	贵阳	纵火	1	6	35	1			
18	2014/3/11	南京公交车起火	南京	机车故障	1	0	0				

附录 5.微博情感极性分类抽样结果

中性

- 1.[人一定要想清三个问题],第一你有什么,第二你要什么,第三你能放弃什么。"对于多数人而。。。
- **2.**[拜拜]其实就是国内的傻逼宣泄不满找了个理由罢了//@littlesoo: 動不動就聖戰的傻逼! //@VV 腰__: 说实在的真不关权志龙什么事==。尽管我不喜欢他==//@_木土 O_O 木土_: 土豪我觉得你说的对!。。。
- 3.【警惕"东突"暴恐新动向】在"东伊运"极端思想的煽动、人员渗透下,"东突"暴力恐怖活动愈发呈现出"基地"组织色彩证据显示,昆明"3?01"事件是一起"东突"恐怖势力经过严密策划······
- **4**【人大开幕会为昆明暴恐事件遇难者默哀】3月5日上午9时,第十二届全国人民代表大会第二次会议在 人民大会堂开幕,全体与会人员为3月1日晚在云南昆明火车站发生的严重暴力恐怖事件中的······
- **5**【新疆乌什县发生袭警案件 3 名嫌犯自爆而亡】14 日 16 时许,阿克苏地区乌什县发生袭警案件。犯罪嫌疑人驾驶车辆携带爆燃装置,手持砍刀,袭击公安巡逻车辆,致 2 名群众和 2 名民警受伤……
- **6** 新浪头条新闻今日 16 时许,新疆乌什县发生袭警案件,2 名民众和 2 名民警受伤,警方击毙 8 人抓获 1 人,3 名疑犯在犯罪时发生自爆死亡。天山网
- 7//@央视新闻:【外交部:美发言人罔顾事实无法自圆其说】美发言人就我国新疆莎车暴恐事件表态称,希望"中国政府允许表达自由的权力"等。对此外交部回应称,这是一起暴力恐怖案件,证据确凿。.。。
- 8//@西华师大文学院 2013 级 7 支部: /@环球新闻广播:#焦点后续#【重症伤员急诊手术已全部完成】目前, 云南昆明 "3•01" 严重暴力恐怖事件中重症伤员急诊手术已全部完成,部分重症伤员将根据
- **9** 转 // #我在看新闻# 《云南 200 村民打砸 30 余公务车》晋宁县警方传唤嫌疑人引发村民聚集,27 名警方人员受伤。 http://t.cn/zRJATwY 来自@腾讯新闻客户端
- **10**【云南晋宁发生群体性事件 27 名警察受伤 30 辆公车被砸】云南网消息,22 日下午,晋宁县公安机关 在晋城镇依法传唤2名嫌疑人后,引发村民聚集,现场工作人员和途经该地的公务人员遭村民堵截.。。

消极

- 1. 你们楼下说新疆的煞笔!? 这你麻痹和新疆人有你麻痹的关系? 这你麻痹的暴徒……
- 2. 自从学生上学以来,世界就再也没有太平过。第一周:东莞扫黄;第二周:昆明暴恐;第三周:马航 失联;第四周:长沙暴恐;第五周:省长违法;那么,第六周会发生什么呢?
- **3.** 刚得知,老妈的好朋友,一个孃孃得白血病…??为生命祈福!??从昆明暴恐事件,到马航失联,再到身边人得白血病,突然觉得生命好脆弱。??珍惜身边人吧,朋友们…?
- 4.【今日,我们与昆明同在】鲜血、哭喊、颤栗…29条鲜活生命瞬间被夺,数十家庭永失所爱。光天化日,

滥杀无辜,丧心病狂!暴徒挑战的是人类共同的底线。挥向平民的屠刀,制造了血腥与恐怖……

- **5.**《默哀!痛定思痛!我们需要更多的正能量》春天,春城,昆明。昨夜,走过一场恶梦!他们中,有一起外出打工的夫妻,有一起回老家的亲人,还有年幼孩子的父母······
- **6.** 这几天总在加班,真的很累,偶尔看看新闻,看到昆明后续事件,感觉让人心寒,并不是说谴责那些新疆的暴徒,只是替这其中的那个未成年少女感到悲哀。在她以后的生活里,要终日面对冰冷的高墙……
- **7.** "3.01"昆明暴力恐怖案被击伤女暴徒已清醒并供罪,年龄只有 16 岁,我还差一个星期也 16 岁,想到她拿着刀到处砍人,才 16 岁为何如此疯狂,我感到不寒而栗——可怜之人必有可恨之处,国家悲哀……
- 8. #我有话说#【张春贤谈昆明暴恐案哽咽:曾自己在房里静静思考】 http://t.cn/8FFXzQb
- 9. 春城被染上了不该有的血红色让人悲痛。假设昨晚有一台心灵净化仪把暴徒的恶念苗头消去,那么现在火车站里面的: A 一家子已经到达丽江,享受着玉龙雪山的极乐幸福; B 君已经出差回到家中……
- **10.** 昆明暴恐,马航失联,长沙砍人……可不可以给努力在找工作和写论文的夹缝里求生存的我们一点正能量?

积极

- 1. 《9 暴徒袭击新疆莎车公安局扔 8 人被击毙》 http://t.cn/8kBKMx7
- 2. #祈福昆明# 顺便说一句,暴徒操你妈详情:
- 3. 新疆莎车县处置一起暴恐袭击案件(分享自@新浪新闻)http://t.cn/8kBJE8Z
- 4. #祈福昆明#没有人性的暴徒祈福昆明我在:乙十四路……
- 5. #我是新疆人# 人民不是羊, 暴徒不是狼! 详情:我是新疆人
- **6.**请不要给我们新疆乱贴标签。暴徒没有资格代表我们新疆!千万不要向其它新疆人维吾尔人渲泄愤怒,那就正中分裂分子下怀。56个民族,不论什么民族,都是好人多!看看新疆多美,快来旅游吧~
- 7. #我是新疆人#虽然我是四川人,也从来没有去过新疆,但姥爷以前在新疆兵团工作过,所以对新疆有种特别的情愫。暴徒代表不了新疆,民族之间没有歧视。大盘鸡和哈密瓜都是中国的美食。我是新疆人
- 8. 专题:《昆明暴力恐怖事件》 http://t.cn/8FrGXs9
- 9. 把新疆的石油给内地停供一周.把新疆的天然气给内地停送一周.把新疆的煤给内地停运一周.把新疆的电给内地停输一周.把新疆的稀土矿停产一周.把罗布泊炸的那二十颗原子弹放到内地,就没有各种癞瓜子再唧唧歪歪了!暴徒≠新疆,新疆也≠暴徒,懂吗?!既然你们说新疆人都是暴徒······
- 10. 新疆打掉一暴恐团伙击毙 7 暴徒 1 名干警牺牲 http://t.cn/SxVZaE