面向主题的微博热门话题舆情监测研究*

——以“北京单双号限行常态化”舆情分析为例

张瑜[1]，李兵[1]，刘晨玥[1]

（1.对外经济贸易大学 信息学院，北京 100029）

摘要： 社交媒体舆情监测是社交媒体分析的热点研究问题，学界和工业界取得了很多研究成果。但目前针对热门话题舆情监测研究中，往往只在整体上关注事件舆情趋势，而没有对事件内部不同的讨论主题进行分析。鉴于此，本研究将主题分类模型引入到舆情监测中来，并在此基础上，以时间为脉络进行面向主题的情感分析。并以“北京市单双号限行常态化”这一微博话题为例进行实证研究，通过各个时段 “北京市单双号限行常态化”这一微博话题群体情感倾向变化的分析，为舆情的监测提供对象和时点选择的参考建议。

关键词：巴斯模型；舆情监测；短文本情感分析

中图分类号：TP391　　**文献标识码：**A

Research on Topic-oriented Supervision of Public Sentiment towards Heated Weibo Events

——A Case Study of “ implementing 'Odd-even' Vehicle Restriction on a Regular Basis”

Zhang Yu　Li Bing　Liu Chenyue

(School of Informaion and Technology, University of International Business and Economics, Beijing 100029)

Abstract: The supervision of public sentiment is a popular theme in the study of social media where a myriad of research concentrated on the general trend of public sentiment towards certain event has been done. However, few of them has analyzed the public sentiment towards various topics beneath the event. Therefore, this paper would focus a topic-oriented sentiment analysis on temporal term. And the weibo over the event of implementing ″odd-even″ vehicle restriction on a regular basis is taken as the empirical data of our work. By observing the sentimental trend of the different topics beneath this event, we attempt to offer feasible suggestions for public sentiment monitoring.

Keywords: Naïve Bayes; public sentiment monitoring; text analysis

*收稿日期： 2015 年 6 月 1 日　　定稿日期：2015 年 8 月 10 日

基金项目: 北京市自然科学基金项目“基于多源信息融合的北京公共危机事件情境感知研究”（项目编号： 9142014）研究成果之一。

1 引言

微博以碎片化的信息形式渗入到人们生活的方方面面，已经成为互联网舆论演化的重要平台。网民对网络话题所持观点的演化过程是舆论演化的重要内容，并对舆论的发展趋势有重要影响。而网络舆情的突发性、广泛性也使得网络舆情极易造成社会恐慌，如果舆情没有得到及时有效的引导，往往造成社会舆论爆发、谣言出现等，严重的还将导致社会情绪低落和失控，甚至会危及社会稳定，因此研究微博舆论观点演化的规律对科学地引导舆论具有重要意义。

微博热门话题事件是微博强大信息传播能力的一个代表，也是微博舆情监控的一个主要阵地。持续性敏感话题是指一些长期受公众关注的话题或事件，由于并没有最终的结论或结果而长时间的处于亚沸点状态，一旦有相关事件进展发生，很容易就会触发舆论热潮。针对于这类事件或话题的研究相对较少，而且以往针对热门事件舆情传播及其监测的研究往往仅是从整体上关注事件舆情趋势，而没有对事件内部不同的讨论主题进行分析。这导致对于舆情的监测往往失却着力点，忽略事件内部的主要矛盾问题而不能达到有效监测舆情的目的。因此，本课题的研究意义就在于，通过关注持续性敏感性话题内部不同主题的民众情感倾向，结合时间发展变化，寻找大多数民众最关心、最敏感的议题方向。为舆情的监测提供有效的参考建议。

2 研究现状

2.1 国内研究现状

就网络舆情的类型而言，“网络舆情”是一个宽泛而模糊的概念，目前学界对于网络舆情的分类有多种不同的方法，大多数的分类从舆情的内容性质
出发。如谢耕耘（2012）[1]按内容将舆情分为食品安全舆情、环境舆情、医疗业舆情、教育舆情、反腐倡廉舆情、官员人事任免舆情、交通舆情、涉警涉法舆情、企业及企业家舆情等。但也有文献从不同的角度对网络舆情进行分类，中宣部情信息局(2009)[2]分别按形成过程，分为自发网络舆情和自觉网络舆情; 按构成，分为事实性信息和意见性信息; 按境内外，分为境内网络舆情和境外网络舆情。也有文献对“事件主体”和“传播媒介”等进行了研究。王国华（2013）[3]根据刺激性事件的主体行为将舆情划分为政府类行为事件和非政府类行为事件。根据传播媒介的不同将舆情划分为个体传播为主的舆情事件和媒体传播为主的舆情事件。

从网络舆情的监控角度来看，当前，我国网络舆情问题涉及人数众多，信息量巨大，影响力空前。因此，网络舆情预警研究成为网络舆情研究的热点，目前，主要的网络舆情监控研究有以下几种模式：“指标体系+ 评价模型”模式的网络舆情预警研究，通过按照一定的科学方法确定关键指标、指标维度、指标层次、指标量化方法，建立预警指标系统，根据不同的评价模型对网络舆情进行监控[5]-[6]；基于情感态度分析的网络舆情预警研究，即通过群体情感倾向性分析(包括“赞同”、“反对”、“中立”三种态度)，利用计算机对网络文本进行分析，关注舆情的发展状态[7]；基于数据挖掘技术的网络舆情监控研究，即通过对网络数据进行数据特征提取、聚类、关联规则挖掘等，得到相关数据，然后通过数据分析对网络舆情进行监控[8]。

2.2 国外研究现状

国外的相关研究多基于推特话题探测、公众意见分析、传播机制三个方向。

从推特话题探测角度入手的学者多提出话题探测方法并用实际数据进行验证。Mario Cataldi(2010)[9]等人基于推特内容老化理论对推特特征词词语进行提取，利用用户的关系网络计算用户权威度，并通过主题连通图联结关键词，以探测新兴话题的产生。Ana-Maria Popescu(2010)[10]等人则构建由话题目标、话题持续期、与话题相关的推特构成的三元组，通过回归机器学习模型计算每个话题的争议程度大小，从而探测公众情感两级分化较大的推特话题。

从公众意见分析角度入手的专家学者，采取将推特文本情感分析得到的结果与实际公众情感对比，探究了使用推特文本进行公众情感检测的可行性与准确性。Brendan O'Connor(2010)[11]等人以 08 到 09 年的

政治观点和消费者信心的调查为例，将以投票方式测度的公众意见与以文本测度的公众情感相连接，测度两者之间的相关系数，认为后者可以在一定程度补充或替代前者。

而 Daniel M.Romero(2011)[12]等学者则从信息的传播机制入手，通过分析一微博用户在频繁接触该话题后参与该话题的可能性，探究不同推特话题在传播过程中的差异；并探究导致该差异的原因是以为推特事件的“影响”为主还是以人们保持“同质性”的倾向为主。而 Hsia-Ching Chang(2010)[13]利用创新扩散理论，基于 Logistic Model 和 Bass Model 提出推特信息传播机制的一般性结论，分析信息传播过程中“模仿”因子和“创新”因子扮演的角色。

2.3 研究现状总结

以往研究多针对于及时探测突然性话题和监测网络舆情对于政府决策的意义，然而除突发性热门话题及事件外，还有一类话题长期受社会关注而处于敏感状态，一旦有事件触发就会立刻爆发舆论潮。我们将这种话题称为持续性敏感话题。

先前的研究在缺乏对话题性质的辨别，因而在监测舆情时无法确切把握其着力点；以往舆情研究多数关注的是总体情感倾向，而针对持续性敏感话题内部议题情感倾向变化对总体情感倾向变化的解释作用的研究却很少。

本文在判定微博话题性质的基础上，关注内部不同议题的民众情感倾向，寻找大多数民众最关心、最敏感的议题方向。为舆情的监测提供有效的参考建议。

3 研究方法

3.1 理论基础与方法

(1) 巴斯模型

网络舆情的演变离不开微博使用者的“创新”与“模仿”。“创新”，即微博使用者受到话题事件进一步发展的外部影响而撰写微博的行为。“模仿”，即微博使用者受到其他微博使用者的内部影响而参与到该微博话题中的行为。在不同话题的舆情演变过程中，“创新”与“模仿”两者的相对重要性不同。掌握舆情演化的主要影响因素，是对网络舆情进行监控的基础。

本文使用巴斯模型对以微博为代表的网络舆情演化机制进行探究。Frank.M.Bass[14]提出，在新产品上市的时候，每个人只购买一个单位的新产品，不存在重复购买。在该假设下，以新产品推出之时为 0 时刻，则在 t 时刻未采用而即将采用新产品的潜在市场份额是 t 时刻之前采用的消费者的线性函数。

其基本形式为：

$$\frac{f(t)}{1-F(t)} = p + \frac{q}{M}[A(t)]$$

其中 t 代表从新产品上市开始所经历的时间，三个巴斯模型系数 p、q、M 分别为外部影响系数（即创新系数）、内部影响系数（即模仿系数）、总体潜在消费者数。F（t）代表在 t 时刻之前累计消费者占总体潜在消费者的比率，f（t）是 F（t）的导数，表示 t 时刻消费者占总体潜在消费者的比率。A（t）代表在 t 时刻之前累计消费者数，a（t）是 A（t）表示 t 时刻消费者数。

以微博为代表的的网络舆情演化机制在一定程度上与新产品的市场扩散类似，其相似性表现在：首先，对于某一微博话题而言，微博用户对该话题的一次性关注的可能性远高于重复或持续关注该话题的可能性。因此可以认为基于微博的网络舆情演变过程满足了巴斯模型的基本假设。其次，微博用户可以分为创新者和模仿者。创新者即为受到外部影响如话题事件发生而主动撰写微博的人群；模仿者即为受到内部影响，如因他人转发相关话题的微博而参与该话题讨论的人群。

鉴于以上两点，考虑到本文的数据抓取时间单位为天，本文将使离散形式的巴斯模型（即 Srinivasan-Mason[15]形式）对网络舆情演化情况进行分析。将 A(t)=MF(t),a(t)=Mf(t)代入巴斯模型的基本形式中即可得到：

$$\mathrm{a(t)} = \mathrm{Mp} + [\mathrm{q}-\mathrm{p}]\mathrm{A(t-1)} - \frac{q}{M}A(t-1)^2$$

其中，创新系数 p 为在外部影响下微博用户主动撰写相关微博的可能性，模仿系数 q 为在内部影响下微博用户可能参与该话题讨论的可能性，A（t-1）代表第 0 至第 t-1 天的累计参与该话题的微博用户数；a（t）表示第 t 天该话题参与者数目。受内外影响共同作用而参与该话题讨论的微博用户数 a（t）即为该话题的舆情情况。

Hsia-Ching Chang 指出，对于一般性话题，创新系数并不显著表明微博用户缺乏内在驱动参与该话题，而显著的模仿系数表明微博用户极有可能受到其他微博用户的影响来参与到话题中。本文将沿用这一思路，根据创新系数和模仿系数是否显著来判断持续性敏感话题的舆情情况。

(2) 基于特征向量空间模型和朴素贝叶斯分类器的议题划分

该微博话题下的议题因为后续事件的发展而不断改变，而该话题内部不同议题的情感倾向变化无疑对该话题的总体情感倾向变化具有解释作用。因此基于巴斯模型对以微博为代表的网络舆情演化机制进行探究后，本文利用特征向量空间模型对微博进行特征提取，使用朴素贝叶斯分类器对微博所属议题进行划分，具体步骤如下。

- 微博文本分词

对于经过预处理的微博数据，本文使用中国科学院计算机所软件室开发的中文分词工具 ICTCLAS 进行分词操作，以便下一步提取微博文本的特征词。

对于微博语料口语化的特殊性，本文对 ICTCLAS 的用户词典添加了相应的网络用语，以提高其分词准确性。

- 阈值确定

对于微博数据的分词结果计算每个词的 TF-IDF 值，并据此降序排列，结合微博数据实际情况和数据处理经验，词汇大约在 36.5%基本丧失特征性，因此选取 36.5%处对应的 TF-IDF 值作为阈值。

- 微博文本特征提取

对于分词结果，本文采用特征向量空间模型[16]对微博文本进行特征提取，通过微博文本的特征项和其相应权值来替代微博文本。

$$d = d\big((t_1, w_1), (t_2, w_2) \dots\dots (t_n, w_n)\big)$$

每个微博短文本可以看成由若干特征项t_i组成，每个特征项具有权值w_i.为了保证选取的特征项具有高代表性和高区分度，以特征项t_i的 TF-IDF 值作为该特征项的权重w_i。并选取 TF-IDF 值超过阈值的词语作为微博文本特征项。

如上分词后的微博文本以权值向量的形式表示如下：

$$d = d\Big(\big(\textit{老百姓},\ 0.8796857\big), \big(\textit{车主}, 0.4159038\big) \dots\dots\Big)$$

- 议题词典构建

在对 TFIDF 超过阈值的词语进行统计收集后，我们以人工头脑风暴交流讨论的方式对收集到的词汇进行了分类。构建本事件内部微博议题类别y_i(i=0,1,2)：环境保护、公共交通、公民权利保障。

基于此，本文构建了议题词典，如上述表格中的“所有权”、“听证会”、“法治”等词被归为公民权利保障类议题的关键词，“生态”、”高能耗“等词语被归为环境保护类的关键词，”公路”、“机动车”等词语被归为公共交通类的关键词。

- 微博文本议题划分

本文使用朴素贝叶斯分类器[17]对微博文本特征词所属于的议题进行划分。

对于对 TFIDF 超过阈值特征词集合 t，本文选择部分特征词构成训练集t_r，其余特征词则作为实验集t_e（即待分类的特征词集合）。对于出现在训练集t_r中的属于微博文本的特征词t_k，可计算其在各议题y_i对应的词典中出现的概率$P(t_k|y_i)(k = 1,,2,..v)$。同时，根据训练集t_r，可计算得到不同议题y_i出现的概率$P(y_i)$。

将训练集得到的先验概率应用于剩下的微博文本中，利用朴素贝叶斯分类器进行议题分类。当特征词t_k在实验集t_e中再次出现时，凭借训练集计算得到的先验概率，我们可以使用如下公式计算得到来自微博文本d_j（$d_j = d(t_{j1}, t_{j2}, t_{j3} \dots\dots t_{jv})$）的一个特征词$t_{jk}$属于议题$y_i$的概率值$P(y_i|t_{jk})(k = 1,,2,..v)$。

$$P(y_i|t_{jk}) = \frac{P(t_{jk}|y_i)P(y_i)}{P(t_{jk})}$$

上式中，y_i(i=0,1,2)代表不同议题，t_{jk}表示微博文本d_j（$d_j = d(t_{j1}, t_{j2}, t_{j3} \ldots\ldots t_{jv})$）的一个特征项。通过计算特征项$t_{jk}$属于议题类别$y_i$的最大后验概率$P(y_i|t_{jk})$，即可判断特征项$t_{jk}$所属于的议题类别。最后选择该微博文本$d_j$的特征项中权值最大的特征项所属于的议题作为该微博文本所属于的议题。

其中，在计算$P(t_k|y_i)(k = 1,,2,..v)$时，为了防止其值为零，使用拉普拉斯概率来完成条件概率的计算。本文中采取频次类型的计算方法：

$$P(t_k|y_i) = \frac{1 + TF(t_k|y_i)}{v + \sum_{k=1}^{v} TF(t_k|y_i)}$$

其中$TF(t_k|y_i)$代表特征项t_k在类y_i下出现的微博条数的频次统计,v 代表特征项的总数。

分类结束后，选取大小为 200 的样本，对分类结果进行了 5 次抽样验证，最终的结果如表中所示,训练数据的精度为 Accuracy=83.25%,测试数据的精度为 Accuracy=75.23%，相比起英文文本情感分析，中文语言更加复杂，多样。再加上微博语言的不规范性造成的一定干扰。所以要准确预测情感极性很困难，75.23%的准确度是可以被采纳的。

4 实证研究

3.1 数据采集和数据处理

本文使用网络数据采集工具MetaSeeker来抓取网页微博的数据，数据采集使用了分时段采集的方式。采集了包括用户id、微博id、本微博内容、被转微博内容等特征的数据作为研究对象。

在对预料进行分词之前，在数据库中对数据进行了清洗处理。去掉各种文本交互信息及新闻、广告等垃圾信息的干扰。并对如“的”，“明日”，“他”等高频但无研究意义的词汇进行清洗处理。

之后，利用自行开发的词频统计工具进行统计操作。经过词频统计操作后，将用户微博数据和热门事件数据按其词语词频的降序顺序排列存储。

3.2 研究对象

本文研究微博热点事件内部不同主题方向舆情随时间的变化发展。

本文选取话题“#单双号限行常态化#”下的微博作为样本语料数据，于 2014 年 11 月 26 日至 12 月 30 日对新浪微博“#单双号限行常态化#”热门话题微博进行了采集。在对数据进行去重过滤、去不相关处理后共得新浪微博关于北京单双号限行事件有效微博 3983 条。本次数据采集内容包括微博用户 id、微博 id、本微博内容、被转微博内容、微博发布时间等。

3.3 数据特征描述

为了发现舆情演变整个时间段(11 月 26 日-12 月 27 日)中政府、新闻媒介、网民的特点，我们利用收集到的关于单双号限行常态化事件的数据，对其进行统计分析，得到不同主体的微博数量如图 1：

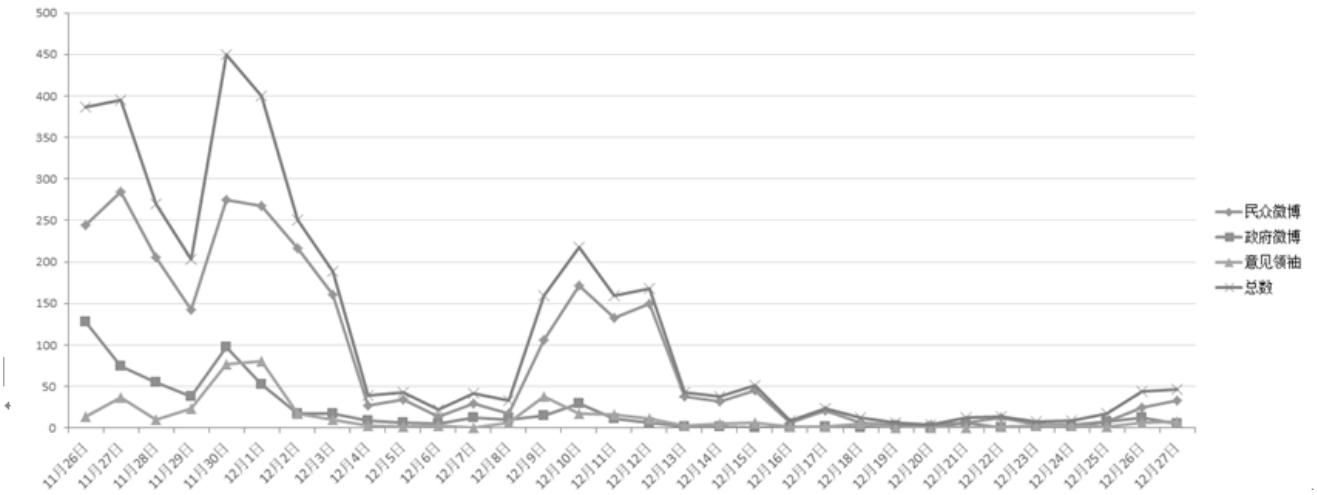

图 1 事件期间各角色主体发博数量

3.4 网络舆情演化结果呈现

由于本文的数据为以天为单位进行抓取，为离散数据，因此将使用 Srinivasan-Mason 离散形式的巴斯模型进行非线性回归，以观测单双号限行常态化这一微博话题的网络舆情演化情况。

$$a(t) = c + bA(t-1) + aA(t-1)^2$$

其中，A（t-1）代表第 0 天至第 t-1 天的累计参与该话题的微博用户数；a（t）表示 第 t 天该话题参与者数目。

通过最小二乘法进行非线性回归，我们可以确定参数 a，b，c 的值如下

	Coefficients	标准误差	t Stat	P-value
c	430.6262337	77.74904	5.538669	6.38E-06
a	-5.71638E-07	1.56E-05	-0.03662	0.971051
b	-0.110760847	0.073418	-1.50863	0.142596

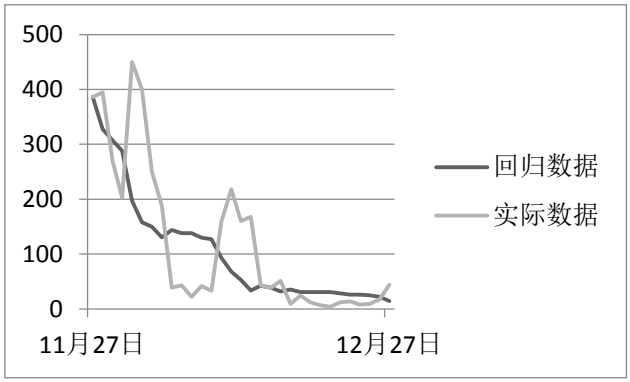

图 2 回归数据与实际数据对比

当 a(t)=0，此时 0 至 t-1 时期的累计微博话题参与者数 A(t)达到最大，即达到了潜在话题参与者的最大值 M。因而，M 可以通过下式计算得到

$$M = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

创新系数 p 可由下式得到

$$p = \frac{c}{M}$$

模拟系数 q 可由下式得到

$$q = p + b$$

	Coef..	标准误差	t Stat	P-value
Innovative effect(p)	0.11294	0.019053	5.927738	2.22E-06
Imitation effect(q)	0.00217	0.059526	0.036616	0.971051

得到的创新系数 p 为 0.113 在 95%的置信度下显著（p 值=2.2E-06<0.05），然而模仿系数 q 在统计意义上并不显著。从以上统计数据中，结合单双号微博话题的演变过程，我们可以看到，外部影响如政府做出回应、人大代表发表评论等触发事件，相比内部影响如其他微博用户的转发评论，对单双号限行这一网络舆情的演变产生了更大的影响。在舆情传播过程中，这样长期存在、受事件影响更大的话题，本文将其称为持续性敏感话题。

在持续性敏感微博话题的舆情传播过程中，与该话题相关的后续事件在舆情演变的过程中扮演着不可忽视的角色：该微博话题下的议题因为后续事件的发展而不断改变，而该话题内部不同议题的情感倾向变化无疑对该话题的总体情感倾向变化具有解释作用。

3.5 网络舆情生命周期划分结果呈现。

观察得到的统计数据，很容易看到整个网络舆情的演变呈现落差式分布。根据既有研究对网络舆情演化过程的分段，并结合本案具体情况，笔者划分舆论生命周期。观察数据的时间分布特征，结合事件的发展历程。可以明显发现在事件进展的关键转折时点，微博数据的发帖量也有明显的增长。本文就根据事件的几个转折点将舆情划分为 6 个阶段,分别对应各舆情演化阶段,如表 1 所示。

阶段	关注度	时间节点	事件转折点
潜伏	386	11.26	网民微博报道单双号常态化新闻
成长	289.3	11.27-11.29	消息传播
爆发	425	11.30-12.1	新华社评论《新华社：单双号限行常态化突破法治红线？》；政府作出回应表明单双号常态化有待讨论
衰退	88.14	12.2-12.8	单双号限行常态化讨论暂时搁置
波动	179	12.9-12.11	人大教授发表评论称单双号限行为违法行为，再掀波澜
死亡	33.7	12.11-12.27	全国人大常委会委员审议时表示，单双号限行常态化侵犯公民财产权利，建议删除。不能随便给单双号限行常态化“开口子”

表 1 舆情阶段划分及各阶段转折点

3.6 网络舆情议题划分结果呈现。

- 微博文本分词结果

对于经过预处理的微博数据。分词结果举例如下。

我们/rr也/d要/v顺带/d鄙视/v央/vg视/vg好像/v什么/ry都/d懂/v似的/uyy给/v单/b双/m号/q限/v行/ng找/v数据/n支持/vn，/wd殊不知/v央/vg视/vg的/ude1数据/n时/ng怎么/ryv编造/v的/ude1。/wj闭/v口/n不/d谈/v粗/a钢/n产量/n，/wd至少/d有/vyou三/m个/q省/n在/p全球/n排名/vi前/f五/m名/q，/wd拿/v车/n说/v事/n那/rzv是/vshi把/pba屎/n盆子/n往/p广大/b老百姓/n车主/n身上/s扣/v。/wj官员/n和/cc富人/n谁/ry会/v在乎/v那/rzv100/m块/q钱/n的/ude1罚款/n，/wd而且/c还/d不/d见/v得/ude3一定/d被/pbei抓/v限/v行/vi///xs

- 词语 TFIDF 文本表征

为了通过程序对大批量的微博文本进行议题的归类，本文首先提取了 TFIDF 权值超过规定阈值的词语作为用于议题归类的关键词，如表 2。

	TF	IDF	权值		TF	IDF	权值
所有权	0.0938	5.9610	0.559	**高峰**	0.0795	4.4543	0.354
生态	0.0667	6.6542	0.444	**车牌**	0.0769	4.5747	0.352
市政府	0.0909	4.7005	0.427	**专用道**	0.0588	5.9610	0.351
听证会	0.0769	5.5530	0.427	**高能耗**	0.0476	7.3473	0.350
车主	0.1000	4.1589	0.416	**大跃进**	0.0455	7.3473	0.334
化工厂	0.0526	7.3473	0.387	**法治**	0.0556	5.7366	0.319
公路	0.0625	5.9610	0.373	**元凶**	0.0556	5.7366	0.319
机动车	0.1250	2.8904	0.361	**宪法**	0.0526	5.9610	0.314

表 2 选取的关健词及其相应 TFIDF 权值

- 议题划分结果

本文从 3983 条数据中选取 400 条构成训练样本集，应用如下表 3 的数据展示了各阶段网民关注议题的比例数量变化， 图 3 是表 3 数据的图表表示。

阶段	环境保护		公共交通建设		公共政策制定	
	个数	比例	个数	比例	个数	比例
潜伏期	212	54.8%	53	13.8%	121	31.4%
成长期	377	43.4%	181	20.8%	311	35.8%
爆发期	290	34.1%	108	12.7%	452	53.2%
衰退期	189	30.7%	84	13.6%	344	55.7%
波动期	147	27.3%	51	9.5%	339	63.2%
死亡期	112	22.2%	45	8.9%	349	68.9%

表 3 各阶段网民关注各议题比例变化

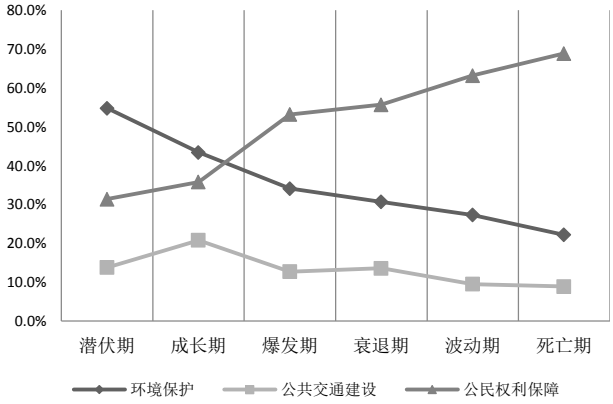

图 3 各阶段网民关注各议题比例变化折线图

3.7 情感极性演变结果呈现与分析

本文使用 Semantria 情感分析软件对微博文本进行情感极性的分析，得到结果如下图 4 所示。

Document Sentiment	Document Polarity	Data Column A
-0.725000024	negative	[/wkz民景/n新闻/n:/wp北京/ns官方/n：/wp正/d论证/v
-0.112499997	negative	论证/v单/b双/a号/q限/v行/vi？/ww又/d开始/v蒙/b很子
-0.649999976	negative	2014-11-26/m，/wd[新闻直播间]/xm北京/ns将/d论证/v
-0.800000012	negative	现在/t广播/n整天/d为/v汽车/n单/b双/a号/q限/v行/ng
-0.449999988	negative	我/rr觉得/v，/wd什么/ry时候/n你们/rr这些/rz所谓/v
-0.65625	negative	分享/v自/p石述思/nr《/wkz北京/ns“/wyz单/b双/a号/q
0.016666671	neutral	北京/ns总结/vapec/n期间/f治理/v雾/n霾/n的/ude1经验
0.024999976	neutral	null/a单/b双/a号/q限/v行/ng，/wd公交/b地铁/gjtgj减
0.524999976	positive	今天/t北京/ns空气/n质量/n很/d好/a，/wd是/vshi优/ag
-0.725000024	negative	小汽车/n要/v单/b双/a号/q限/v行/ng，/wd公交/b地铁/g
0.5	positive	null/a单/b双/a号/q限/v行/ng就/d没/d发现/v一个/mq开
-0.300000012	negative	北京/ns副/b市长/n：/wp正/d论证/v单/b双/a号/q限/v行
0.449999988	positive	null/a单/b双/a号/q限/v行/ng跟/p车/n有/vyou毛/nr1关
0.600000024	positive	北京/ns单/d双/a号/q限/v行/ng常态/n化/k，/wd车王/nf
-0.120833337	negative	null/a单/b双/a号/q限/v行/ng对于/p客/v希望/v于/p减/v

图 4 情感分析结果

通过情感极性分析结果进行统计，得到结果如下。图 5,6,7 是三个议题情感极性变化的分别展示。

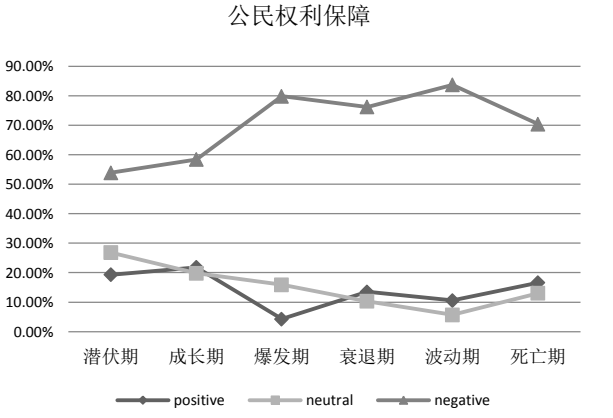

图 5 议题—“公民权利保障”各阶段情感极性变化

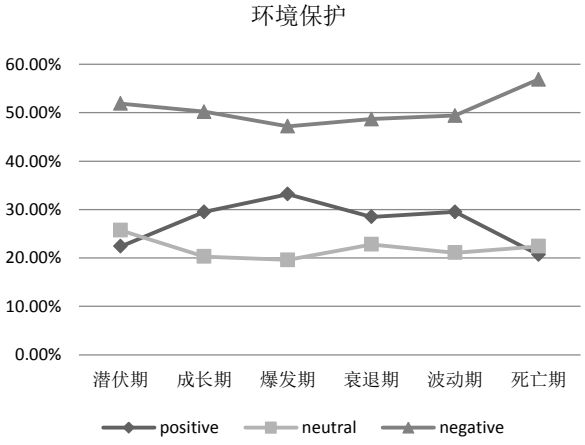

图 6 议题—“环境保护”各阶段情感极性变化

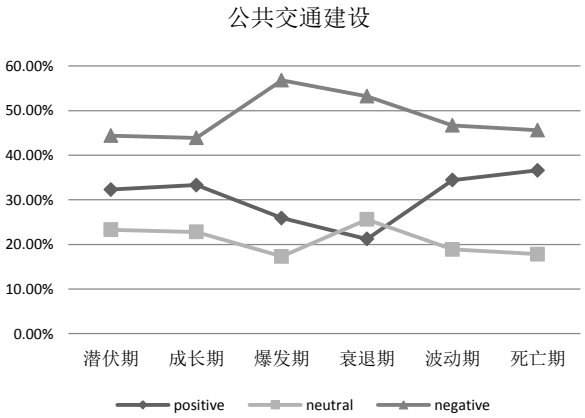

图 7 议题—“公共交通建设”各阶段情感极性变化

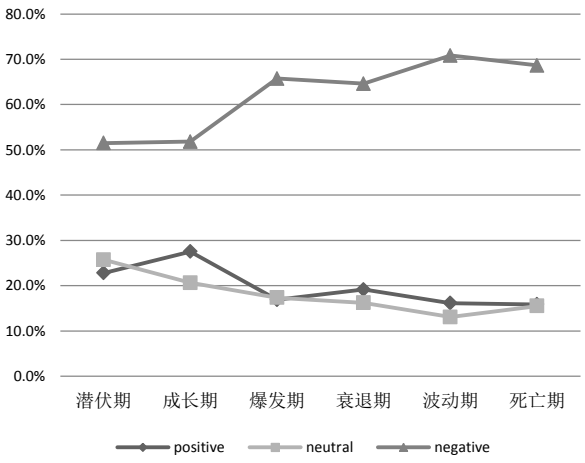

图 8 单双号限行常态化事件总体情感极性分布

可以看出，在舆情演化的不同阶段，群体的情感极性变化是有鲜明的变化趋势的。下文我们将“公民权利保障”，“环境保护”，“公共交通建设”分别称为“话题一”，“话题二”，“话题三”。

具体分析如下：

(1) 潜伏期（11 月 26 日）：2014 年 11 月 26 日下午，因某政府官员表示要“听取和论证”单双号限行的建

议，但因为语言表达具有一定的歧义，大众认为单双号限行的政策已经开始制定并将要实施，引起舆论传播的热潮。通过图 3，我们可以看到，因 APEC 刚过，良好环境质量的影响犹在，潜伏期超过半数群体主要关注的议题是单双号常态化对环境保护的作用，其次比较关注的方面是交通限行政策的制定是否触及公民的权利，其余的 13%左右的群体关注公共交通建设的问题；观察图 5，6，7 三图，“环境保护”和“公民权利保障”两议题的消极倾向都超过了半数，受此影响事件所有微博的总体情感倾向也有超过半数的群体偏向消极。潜伏期一个值得注意的问题是，在 26 日上午的言论造成舆论争议后，政府在下午就立刻采取了行动在主要的媒体微博澄清表示：论证但不一定会实施。一定程度上消减了之前言论的刺激性影响，但整个事件并没有因此而冷却下来。

(2) 成长期（11 月 27 日-11 月 29 日）27 日到 29 日是事件的传播期。这一阶段，并没有特别的触发事件发生。但之前政府的澄清微博没能完全消解事件影响，事件相关微博依然在传播，但发帖量相比潜伏期前期的爆发，数量已经有显著降低。成长期内相对潜伏期，群体关注的议题发生了一定改变。“公共交通建设”和“公民权利保障”相关议题得到了更多的关注，“环境保护”议题的影响力开始下降。观察群体情感极性，本阶段触发事件和意见领袖的缺失使得整体上和议题内部的情感变化都比较微小，舆论情感倾向并没有发生很大的变化。

(3) 爆发期（11 月 30 日-12 月 1 日） 从 29 日下午起，媒体相关意见领袖开始逐渐发声。意见领袖集中于讨论单双号限行常态化的可行性情况，带动事件热度提高。直到 11 月 30 日下午，新华视点，人民日报先后在微博上推送“【单双号限行常态化突破法治红线】”提出专家意见。权威媒体传播专业人士对政策合法性质疑这一触发事件的发生，使得整各事件舆论评述进入爆发期。相关微博被大量转载，我们关注的话题下微博发帖量也迅速增加。值得注意的是，因这一阶段的触发事件多关注单双号限行常态化这一政策的合法性讨论，在本阶段，“公民权利保障”相关议题得到了极大的关注，观察图 3 可以发现在爆发期，已有超过 50%的群体关注该话题。“环境保护”和“公共交通建设”议题关注度则有所下降。情感极性方面，受触发事件影响，话题一和话题三的情感极性消极方面占比迅速提高，话题一的消极比例更是达到了 79%左右。话题二的情感极性中反而是消极情感比例略有下降，积极情感比例上升。但因为占微博总体数量比例小，话题一、三微博数量较多，总体上在爆发期，整个群体消极情绪爆发，积极情感降低。

(4) 衰退期（12 月 2 日-12 月 8 日） 微博巨大的信息量在带来巨量舆论信息流的同时，也加快了事件降温的速度。因没有触发事件的继续刺激，群体情感逐渐冷静，进入一个衰退期。发文数量和情感极性都趋向平稳。本时期内群体关注议题情况与爆发期情况基本一致。议题一三的消极情感比例略有回落，总体情感中性和积极情感比例小有上升，消极情感比例下降。

(5) 波动期（12 月 9 日-12 月 11 日）衰退期中，舆情的发展比较稳定，消极情感极性并没有继续发展，但若有刺激性触发事件出现，舆情会产生波动。12 月 9 日微博账号中国之声发表微博：“人大教授：北京单双号若常态化将违宪”。此次触发事件的主人公权威性较高，并涉及到“违宪”。于是在衰退期后，迎来了这次舆论波动。因为多次事件触发皆关注议题一，所以在波动期，议题一相关微博比例继续上升。其余两个议题的比例仍下降。整个事件过程中，政府除 26 日下午的信息澄清外，并没有明确的回应，这也使得群体情绪一直在波动中朝消极方向发展。议题一群体舆情消极比例创出新高，受其影响，整个事件微博的总体舆情的消极情感倾向达到顶峰。

(6) 死亡期（12 月 12 日-12 月 27 日）12 月 26 日全国人大常委会委员审议时表示，单双号限行常态化侵犯公民财产权利，建议删除，不能随便给单双号限行常态化“开口子”。事件有了最终结果。受此触发事件影响，事件经历了最后的发帖热潮，此阶段最受关注的议题依然是议题一，议题一三的消极情感比

例皆有所下降。但单双号限行讨论的搁置不利于环境的保护，议题二的消极情感比例有所上升。但总体上，群体情感得到了安抚，满意度上升。

4 讨论与结论

4.1 持续性敏感话题长期处休眠于状态，一旦有相关事件触发就可能会产生剧烈爆发，引起舆论潮。单双号限行政策从 2007 年第一次实施起就存在着一定的争议。之后该政策扩散到西安、武汉等城市时都曾成为一段时间内的地方性热门话题，该话题本身长期处于一种温热的亚沸点状态。本次同样是被事件触发形成热门话题，因此注意识别潜在的持续性敏感话题并进行一定的跟踪监督十分必要。

4.2 在同一话题下，不同的人群因个人社会地位、相关利益的不同，会关注不同的议题方面，也会因为社会环境和舆论环境的变化而发展。本例中，潜伏期的首次触发事件发生在 APEC 刚刚结束后，此时“APEC 蓝”仍被人们津津乐道，大气环境污染问题仍是网络舆论空间里的一个重要的话题，所以在话题预热的最初，人们相对关注的话题仍旧是单双号对于环境保护的意义。但随后的意见领袖对于此话题的关注集中于单双号常态化的合法性及对公民权利的影响，多次触发事件的影响加上意见领袖的诱导，导致群体对于本话题下的议题关注比例发生了明显的变化。这表明对于话题内部的不同议题的引导是必要且有意义的。

4.3 话题内部主要议题的群体情感极性对最终整个话题的群体情感状态起决定性作用。观察图5,6,7和图8，可以明显的发现图 5 与图 8 的图形相似度极高。在本例中因为我们研究的是舆情的监控，所以主要关注群体舆情的消极方面。对“公民权利保障”议题相关微博的消极情感微博数据和总体消极情感微博数据进行相关性分析可得两者的相关系数达 0.955。这表明在对整个话题舆情的监测过程中，寻找到多数群体关注的议题，掌握该群体的情感极性对于针对性的调控整个话题舆情的发展具有重要意义。

4.4 网络舆情对于政府决策机制具有指导作用，同时，政府的决策反馈也可以起到安抚舆情的反向作用。政策应以民意为向导，通过掌握网络大量群体对于某项政策的情感倾向和意见指向，为有效得根据民意制定政策提供了一种方法。在本例中，群体对于北京单双号限行常态化使得该项建议的论证得到了极大的关注，最终得到了良好的讨论和处理，事件最终结果同时也安抚了群体情绪，扭转了舆情消极倾向不断发展的态势。

4.5 政府相关部门应重视政务信息传播的准确性，并关注微博上相关话题的传播发展，适当建立政府的官方政务微博。本次话题的预热起源于对政府官员的语言信息的缺漏性歧义传播，该官员的原意表达的是：将对单双号限行常态化这一建议进行论证。但在随后的新闻媒体传播过程中，缺失了“建议”二字，使得部分网友产生了单双号限行常态化的政策已经进入立法阶段的错觉，并引发了政府“懒政”的嫌疑，整个话题热度徒升。虽然该官员在当天下午就通过媒体澄清了愿意，但因为该媒体没有权威而强有力的话语权，虽然在一定程度上消弭了事件的影响，但并不能完全消除话题热度。事件的触发效应已经发生，将持续性敏感话题从休眠状态唤醒，进入了一次舆论潮。因此，政府掌握建立一个强有力的权威政务微博是有必要的。

5. 总结与未来工作

持续性敏感话题是指一些长期受社会关注而处于休眠状态的社会话题，一旦有事件触发就会立刻爆发舆论潮。我们通过巴斯模型对此类事件进行识别，并通过关注持续性微博敏感性话题内部不同主题的民众情感倾向，结合时间发展变化，寻找大多数民众最关心、最敏感的议题方向为舆情监测供调控对象和时点选择的建议。

由于持续性敏感性话题，往往会在相当的一段时间内得到持续关注，这一点与巴斯模型 “不存在重复购买”的假设有所偏离，我们会在未来的工作中会进一步完善我们的研究，未来的工作包括修正模型，以

及针对类似事件的实证研究等。

参考文献

[1]谢耘耕．中国社会舆情与危机管理报告[M] 北京：社会科学文献出版社，2012

[2]中共中央宣传部舆情信息局． 网络舆情信息工作理论和实务[M]．北京：学习出版社，2009：9-12

[3] 王国华，冯伟，王雅蕾．基于网络舆情分类的舆情应对研究[J] 情报杂志，2013,32(5)：1-4

[4] 杨娟娟，杨兰蓉，曾润喜，张韦．公共安全事件中政务微博网络舆情传播规律研究:——基于“上海发布”的实证[J] 情报杂志，2013,32(9)：11-15

[5] 曾润喜，徐晓林． 网络舆情突发事件预警系统、指标与机制[J] 情报杂志，2009， 28(11) ：52-54

[6]王青，成颖，巢乃鹏． 网络舆情监测及预警指标体系构建研究[J]． 图书情报工作，2011(4) ：54-57

[7] 刘全超，黄河燕，冯 冲．基于多特征微博话题情感倾向性判定算法研究[J] 中文信息学报 2014,28(4)：123-131

[8]吉祥．基于观点挖掘的网络舆情信息分析[J]． 现代情报，2010(11)：46-49

[9] Mario Cataldi,Luigi Di Caro, ClaudioSchifanella．
Emerging Topic Detection on Twitter based on Temporal and Social TermsEvaluation[C].Proceedings of the 10th International Workshop on Multimedia Data Mining.New York:ACM，2010:4．

[10] Ana-Maria Popescu, Marco Pennacchiotti,Detecting Controversial Events from Twitter[C].Proceedings of the 19th ACM International Conference onInformation and Knowledge Management.New York:ACM, 2010(16):1873–1876

[11]Brendan O'Connor,Ramnath Balasubramanyan,Bryan R.Routledge,Noah A.Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series[J].Computer and Information Science.2010(5):122-129

[12]Daniel M.Romero, Brendan Meeder,Jon Kleinberg. Differnces in Mechanics of Information Diffusion Across Topics: Idiom,Political Hashtags,and Complex Contagion on Twitter[C]. Proceedings of the 20th International Conference on World Wide Web. Hyderabad, India, 2011: 695-704

[13]Hsia-Ching Chang.Rehashing Information Architecture:Explorint Human-Information Interaction of Collaborative Tagging Using Twitter Hashtags[D].
New York:University at Albany,State of University of New York.2010:47-57

[14]The Bass Model[EB/OL]．[2015-04-20]．
http://www.bassbasement.org/BassModel.

[15] Srinivasan, V. Seenu and Charlotte Mason. 1986. Nonlinear least squares estimation of new product diffusion models. Marketing Science, 5 (2), 169–178.

[16] Salton, G., Wong, A., Yang, C.S. 1975. A vector space model for automatic indexing. Communications of the ACM 18: 613-620

[17] Naive Bayes[EB/OL]．[2010-06-07]．
http://group.cnblogs.com/topic/40112.html

作者简介

作者一：张瑜（1995-），女，本科，研究方向：电子商务，Email：zhangyubut@foxmail.com

作者二（通讯作者）：李兵（ 1970 － ），男，博士，教授，研究方向：社会网络分析和数据挖掘，Email：lb0501@126.com

作者三：刘晨玥（1994-），女，本科，研究方向：信息管理，Email：liuchenyue1617@sina.com