

# 基于微博文本数据分析的社会群体情感 可视计算方法研究

刘翠娟<sup>1,2</sup> 刘箴<sup>1,†</sup> 柴艳杰<sup>1</sup> 方昊<sup>1</sup> 刘良平<sup>1</sup>

1. 宁波大学信息科学与工程学院, 宁波 315211; 2. 浙江万里学院智能控制研究所, 宁波 315100;

† 通信作者, E-mail: liuzhen@nbu.edu.cn

**摘要** 针对已有的情感分析研究多侧重在情感的倾向性方面, 缺乏对各类情感的详细描述, 不能形象直观地反映社会群体的情感变化的问题, 提出一种基于依存句法和人工标注相结合的情感分析方法。该方法采用三维立体的人脸表情进行情感分析, 形象地呈现社会群体的情感变化。对于不同的社会事件, 以可视化方式来展现不同地区微博群体的情感。实验结果表明, 该模型可以有效地描述人群情感, 研究结果为基于大数据的网络舆情分析提供了一种新思路。

**关键词** 文本分析; 微博; 情感分析; 可视计算

**中图分类号** TP391

## Research on the Visualization Method of Social Crowd Emotion Based on Microblog Text Data Analysis

LIU Cuijuan<sup>1,2</sup>, LIU Zhen<sup>1,†</sup>, CHAI Yanjie<sup>1</sup>, FANG Hao<sup>1</sup>, LIU Liangping<sup>1</sup>

1. Faculty of Information Science and Technology, Ningbo University, Ningbo 315211; 2. Intelligent Control Research Institute, Zhejiang Wanli University, Ningbo 315100; † Corresponding author, E-mail: liuzhen@nbu.edu.cn

**Abstract** Existing sentiment analysis focus on the emotional tendency, which are lack of detailed description of all kinds of emotions, they can't intuitively reflect the emotional change of social groups. An emotional analysis method based on the combination of dependency parsing and artificial tagging was proposed. Facial expression animation to present emotions analysis was realized. The microblog crowd's emotion in different areas for different social events was visualized. The experimental results show that the model could closely and effectively simulate the crowd emotion, and it could provide a new way of the analysis of network public opinion based on large data.

**Key words** text analysis; microblog; sentiment analysis; visual computing

随着移动互联网的发展, 社交媒体成为人们表达情感的重要载体。微博作为传播较广泛的社会媒体, 已经成为了解民众情感的重要渠道。面对既庞大又看似杂乱无章的微博数据, 如何有效地通过已有数据提取有价值的信息进而分析网络舆情<sup>[1]</sup>, 并以更加清晰的方式呈现, 成为当前备受关注的重要研究领域。

虽然关于文本的情感倾向性研究已取得不少成果, 但是社交媒体群体情感可视分析方面的相关研究很少见。本文首先对基于微博的文本情感分析和大数据的可视化进行综述。然后给出微博文本的情感计算模型, 提出一种基于依存句法和人工标注相结合的情感分析方法, 并以三维立体的人脸表情呈现群体情感。最后结合微博文本分析实验, 说明群

国家自然科学基金(61373068)、宁波市科技计划项目(2015A610128, 2015C50053, 2015D10011, 2014C50018)、高等学校博士学科点专项科研基金(20133305110004)和浙江省教育厅科研项目(Y201431792)资助

收稿日期: 2015-06-19; 修回日期: 2015-09-15; 网络出版日期: 2015-09-29

体情感分析的思路。本文从社会计算的视角进行群体情感分析,针对特定话题,从微博等社会传感网络获取所关注的社会信号,从中提取感兴趣的社会事件信息。本文研究结果可为深入开展微博群体情感分析研究提供一种新的视角。

## 1 相关研究

关于微博情感分析的研究已经积累了很多工作。基于机器学习的方法把文本情感分析看做分类问题,采用朴素贝叶斯、K 最邻近、支持向量机和条件随机场等算法构造分类器,通过对训练集的特征进行学习来构造模型。从监督微博群体情感的视角,情感分析的计算效率较高。采用机器学习的方法,由于需要训练大量的样本<sup>[2]</sup>,对技术速度要求较高。因此,采用基于情感词匹配的方法并结合句法分析(syntactic parsing)成为群体情感计算的有效途径。Caro 等<sup>[3]</sup>提出一个模型,通过语义理解分析用户的情感,并借助数据可视化展现情感的正负强度。Loia 等<sup>[4]</sup>提出一个提取情感的框架,采用四维情感层次,利用同义词集计算情感的相似度,结合模糊集的语义模式计算情感强度。实验证明,此框架能很好分析文本中情感的极性。Wang 等<sup>[5]</sup>提出一个交互的可视化系统,用于分析互联网上用户的群体情感,采用细胞自动机模型比较时变特性,可以适用不同的社交平台。关于情感原因的检测, Lee 等<sup>[6]</sup>提出一个文本驱动的、以规则为基础的情感原因的检测方法,构建了一个情感原因标注语料库和检测系统,经验证,取得显著效果,有助于分析和语言建模。在情感标注方面, Mohammad 等<sup>[7]</sup>关注词的情感与词的极性相结合,生成一个大规模词-情感的联合词汇,并考虑 8 种基本情绪,使用自动生成的词选择题来进行情感标注,取得较好效果。Francisco 等<sup>[8]</sup>提出一种自动标注情感的方法,该方法考虑情感类别和情感维度,通过情感本体进行分层分类来描述通用情感和特定情感,为情感标注提供了一种有效的手段。冯时等<sup>[9]</sup>提出一种基于句法依存分析技术的算法,考虑情感关系对中的依存距离,并通过建立原型系统实现情感倾向的搜索结果。徐琳宏等<sup>[10]</sup>构造情感词汇本体,从情感类别、强度和极性三方面来描述词汇。

可视化是一种大数据分析的重要方法,可以对数据进行推理,进而展示隐含的数据模式<sup>[11]</sup>。可视化数据的基本原理是让数据用某种可视化方式呈现

(如图形、图表、地图等),使人们能够更加直观地发现庞大数据内部隐藏的规律,进而得出结论,并且能够直接与数据进行交互。Zhang 等<sup>[12]</sup>使用 5Ws 数据维度进行大数据分析和可视化,证明大数据可视化可以揭示数据接收和发送的密度,从密度视角分析大数据的特征。Mukhopadhyay 等<sup>[13]</sup>将可视化应用于生物医学领域,通过多路结构图,从生物文本中提取重要信息,有助于理解基因和疾病的关系。可视界面是直观了解海量社会媒体信息的重要工具。利用可视化工具对微博数据进行可视分析有助于更好地发现微博的数据规律,具有广阔的发展前景。微博数据的可视化方面主要有以下几种方式:标签云与 Wordle、图表、网络、地图等。Zhang 等<sup>[14]</sup>提出一个电子云模型(ECM)来分析微博的情感,将情感映射到电子的稳定性,通过核密度和边界绑定表述情感的空间变化。数据可视化为微博的深入研究开辟了一条新径,具有深远的理论指导意义和实践意义。

综上所述,关于微博情感分析的可视研究虽然已取得可观的成果,但还存在一些问题。1)已有的研究多侧重在情感倾向性上,缺乏对各类情感的详细描述。对于社会事件,人们更希望了解不同负面情感的强度,这方面研究尚不多见。2)在情感分析中,已有研究对于情感词强度没有给出具体的量化方法,仅用级别代替。3)现有的社会媒体可视研究大多停留在线条图形,可视效果欠缺。

鉴于以上不足,本文针对社会事件,从多情感视角分析群体情感。在情感词的强度上,采用模糊集合理论,给出情感词的量化描述;在可视效果上,采用人脸表情呈现情感,实现更加逼真的情感表现方式。

## 2 微博文本的情感计算模型

微博文本的情感计算模型采用三元组表示:  $EModel = \langle W, C, E \rangle$ , 其中  $W = \langle w_1, w_2, \dots, w_n \rangle$  为情感词,  $n$  为情感词个数;  $C = \langle c_1, c_2, \dots, c_6 \rangle$  为情感词的类别,共 6 类;  $E \in (0, 1)$  为情感词的情感强度。微博文本的情感计算模型如图 1 所示。

模型的算法步骤如下。

1) 对于收集的微博文本,首先进行句法分析,对微博文本进行预处理。

2) 抽取并统计情感句数目,根据句中情感词的强度,计算每个情感句的初始值。如句子结尾有感

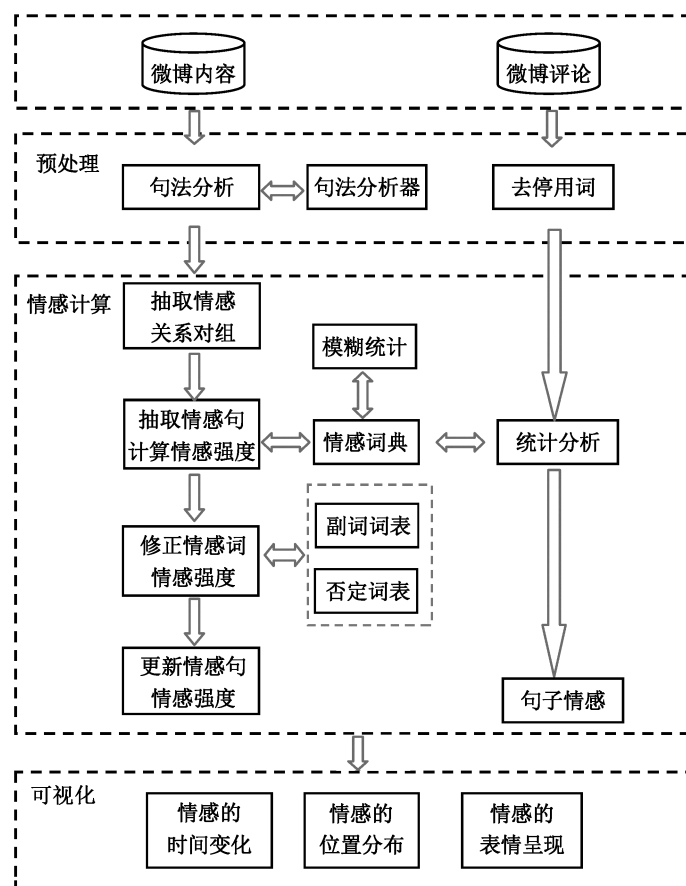


图 1 微博文本情感计算技术路线

Fig. 1 Affective computing technique route of micro blog text

叹号, 则句子的情感强度按照 1.5 倍计算; 如果有疑问号, 则句子的情感强度按照 0.8 倍计算。对于表情符号, 可以采用类似的思路, 并将最终的句子情感强度进行归一化处理。

3) 以情感句中的每一情感词为单位, 抽取关系对组, 并计算出每个情感词的原始值( $e_i$ )。

4) 对于每个情感词的关系对组(包括转化后的 ADV 依存对), 结合程度副词和否定副词修饰强度的变化, 构造情感强度的修正公式, 计算其依存修饰后的情感值。新的公式计算出的情感值仍在 0~1 之间。

5) 结合情感句分析, 计算微博文本中的各类情感强度值, 结果呈现为二元组:  $EGroup = \langle C, E \rangle$ 。由于微博文本通常很短小, 所以本文选择强度最大的情感作为该微博文本的情感类型。最后根据各条微博文本的情感类型和强度, 采用算数平均来计算该种类型的群体情感强度。

## 2.1 情感词的强度人工标注方法

**情感词典的整理和分类标注** 在林鸿飞情感词汇本体<sup>[10]</sup>的基础上, 对微博常用词进行标注, 将 6 种情感类别(快乐、愤怒、厌恶、惊讶、恐惧和悲伤)映射到 6 种基本表情形态。

**情感模糊性** 自然语言所表达的情感强度的变化是连续的, 并不是离散的几个强度点。人们在划分情感强度时会觉得界限模糊, 对于某一类情感, 由于情感强度的边界划分并不清晰, 不同的标注者可能将它划分到不同的级别。例如情感词“提心吊胆”、“惊弓之鸟”、“人心惶惶”和“触目惊心”, 在进行情感强度划分时, 很难给出不同情感级别的明显界限。鉴于语言本身具有模糊性, 认识情感词语强度的模糊性, 可以有效地理解情感强度的模糊特征, 提高情感强度分析的准确性。Kazemzadeh 等<sup>[15]</sup>从概念上生成两个语言计算模型, 将情感词描述为间隔模糊集合, 并运用模糊逻辑, 从效价、激活和支

配 3 个维度描述情感，分析情感的相似性和包容性，最后通过显式的模型描述隐性的情感。

**情感词强度统计** 以往的情感分析中，一般默认采用情感词典的强度，但情感词典中的强度一般采用分级标注，并没有结合上下文信息。我们用统计人工标注的方式，通过人机界面进行统计(图 2)。

结合微博内容，用户可以对情感词的情感强度进行级别标注，针对性更强。本文通过模糊统计实验计算隶属度。将情感强度分为低、中、高 3 个级别，对应的集合为  $S$ 、 $M$  和  $L$ 。利用人工方式对情感词强度进行手工标注，由于不同情感强度之间的界限模糊不清，所以分别统计落入不同强度级别的人数  $N_S$ 、 $N_M$  和  $N_L$ 。不失一般性，赋予情感低、中、高 3 个级别的情感强度值分别为  $V_S$ 、 $V_M$  和  $V_L$ ，则情感词的强度  $E_w$  可以按照式(1)计算：

$$E_w = \frac{(N_S V_S + N_M V_M + N_L V_L)}{(N_S + N_M + N_L)} \quad (1)$$

考虑到微博大多具有情感倾向性，为简便起见，

本文采用最大化原则。统计人数最多的级别则为该情感词的强度，即情感强度 =  $\max \frac{t \in A^*}{n}$ ，其中  $A^*$  为  $S$ 、 $M$  和  $L$  中的一个， $t$  为属于某一情感级别的人数， $n$  为实际参与标注的微博总数目。

2.2 基于依存句分析的情感词强度修饰计算

**句法分析** 句法分析指根据给定的语法规则，对句子中的词语语法功能进行分析，自动识别语句含有的句法单位以及它们之间的关系。句法分析是对自然语言进行语义分析的基础，已成为大多数自然语言处理应用的关键环节。本文利用基于树库的统计句法分析，得到语法树形式的结果，提取情感词所在的句子成分，将其作为统计整个语句的情感特征的权重赋值标准之一。通过对语法树分析，提取具有情感色彩的关系对组，如图 3 所示。

**依存距离** 依存距离指句法分析树中任意两节点  $w_i$  和  $w_j$  遍历先后次序的差值。语法距离即词汇在句法分析树中经过分词分句和词性标注等预处理

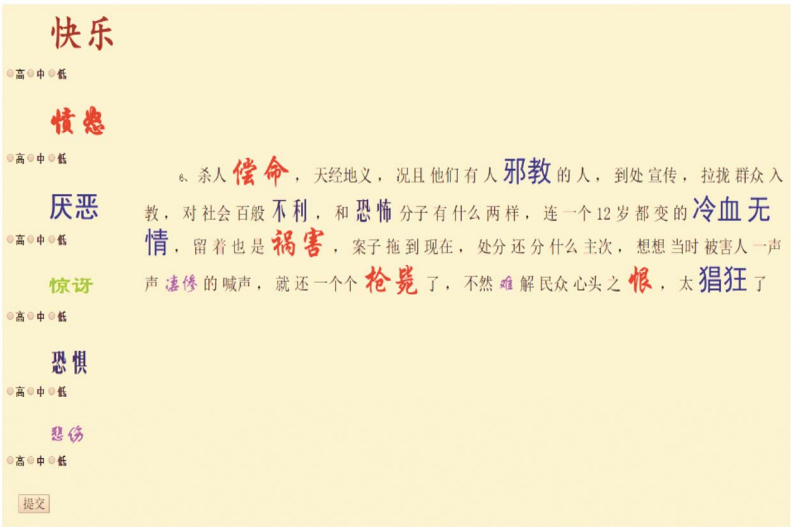


图 2 针对“招远围殴凶杀”事件对北京地区的情感调查  
Fig. 2 Emotional survey in Beijing for the “Zhao Yuan hit homicide” event

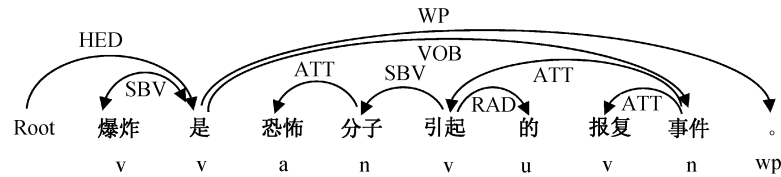


图 3 句法分析依存关系实例  
Fig. 3 Examples of dependency parsing

后, 先序遍历句法分析树时两节点次序之差的绝对值<sup>[9]</sup>, 即  $\text{dist}(w_i, w_j) = |N_i - N_j|$ 。  $N_i$  和  $N_j$  分别表示先序遍历句法树时词语  $w_i$  和  $w_j$  的词号。语法距离值为正整数, 能相对准确地反映词语  $w_i$  和  $w_j$  之间的依存关系和修饰关系。语法距离体现了情感句中修饰词与被修饰的情感词间相互依存的紧密程度。语法距离越大说明词语间依存关系越松散, 情感修饰程度越弱。不同的修饰词对情感的修饰起不同的作用。例如, 有的副词对情感词强度起加强作用, 有的副词对情感词强度起减弱作用, 否定词对情感词的强度起否定作用。我们在情感句子的情感计算中充分考虑副词和否定词的语法距离引起的修饰作用, 使用语法距离来衡量修饰词对句子情感强度的影响, 在语言学的逻辑角度上, 更具有科学性。

**强度修饰** 基于依存句分析的情感计算关键是情感词强度修饰的计算。情感修饰主要考虑副词和否定词的修饰, 并结合修饰词和情感词之间的依存距离。通用的修饰公式为

$$E = \frac{e_i \cdot \text{revi}}{\theta \cdot \sqrt{\text{dist}}}, \quad (2)$$

其中,  $e_i$  为情感词强度,  $\text{revi}$  为修饰强度,  $\theta$  为修正参数(默认为 1),  $\text{dist}$  为修饰词的依存距离。

1) 对于副词修饰的情感, 按照副词强度表, 查询得到副词的修饰强度  $b$  (一般在 1~2 之间)。

2) 对于否定词修饰的情感, 按照否定词强度表, 查询得到否定词的修饰强度  $c$  (一般在 1~2 之间)。

3) 对于既有副词又有否定词修饰的情感, 按照副词强度表和否定词强度表, 分别查询得到副词的修饰强度  $b$  和否定词的修饰强度  $c$ ; 根据依存距离, 副词和否定词对情感词进行加强或减弱修饰。

更新的情感词强度为  $e_i = \min(1, \text{revi} \cdot e_i)$ , 即更新后的情感强度范围为  $[0, 1]$ 。

### 3 情感的可视化呈现

以往对于微博舆情的可视展现大多通过颜色或者线条的变化来进行, 而对于社会事件的群体情感, 则需要一种直观的情感表现手段。本文提出采用虚拟人头来直接呈现某个地区的群体情感状态。采用三维模型工具, 制作 6 种虚拟人的表情形态, 每种表情形态代表情感强度为 1 的状态。几何模型形态顶点向量为  $\mathbf{fa}_i$ , 中性表情的顶点几何向量为  $\mathbf{fa}$ , 采用线性插值的方法计算合成的表情形态。假设各类

基本情感强度为  $E_i$ ,  $\mathbf{face}_i$  代表对应的表情形态几何顶点向量, 则有如下公式:

$$\mathbf{face}_i = (1 - E_i) \mathbf{fa} + E_i \mathbf{fa}_i. \quad (3)$$

采用虚拟人头表情来展现群体情感能够直观反映微博情感的变化, 是实现数据可视化的一种新方法。为了突出不同地区微博数量的差异性, 本文根据不同地区微博数量的相对大小调整虚拟人头的大小, 虚拟人头越大, 表示该地区的微博数量多, 反映群体情感的影响越大。

## 4 实验及分析

本文目的是分析微博中社会事件的群体情感, 评估提取群体情感方法的性能及其对文本情感分析的影响。我们将实验设计为两部分: 一是情感词的标注实验, 给出合理的情感强度; 二是微博情感分析实验, 判断文本的情感类别和强度。

### 4.1 实验设计

**情感词的标注实验** 针对不同的事件, 在全国 34 个省市自治区范围内, 从性别、情感类型和情感强度等方面进行调查。例如, 选取 5 个地区, 每个地区取 200 条微博进行调查, 微博数不足 200 条的地区以实际数目进行统计。每个页面只调查一条微博的情感。在进行标注实验时, 考虑到标注者性别的差异性, 请男生和女生分别标注, 最后将各性别标注的结果取平均。

**微博情感分析实验** 本实验分词及句法分析使用哈尔滨工业大学语言云平台。实验中抽取的情感关系对(依存关系对)是七元组:  $\langle \text{firstid}, \text{firstword}, \text{firstpos}, \text{secondid}, \text{secondword}, \text{secondpos}, \text{rintensity} \rangle$ , 分别表示前项词号、前项词、前项词性、后项词号、后项词、后项词性和初始强度。本实验关注以下形式的关系对: 情感词为后项词, 前项词中出现副词或者否定词的关系对。我们依据式(2)修正该关系对的强度。微博情感可视化编程环境为 Windows 8.1, 采用的工具为 vs2013, 可视化编程接口为 directx 11, 用 3dmax 2009 进行地图制作。

### 4.2 实验数据采集

本文的实验数据包括微博内容和微博评论。微博内容包括“招远围殴凶杀”、“事业单位打破终身制”、“大老虎”、“世界杯”、“高考生夺刀救人”等社会事件。为了验证本文提出的方法, 通过腾讯和新浪微博的 API 接口, 抓取大量社会事件的微博数据。这些数据是去重(过滤掉重复)后的数据, 避免

同一个博主所发信息被多次采集。在调查初始情感强度时,只选择部分去重后的数据集(特定事件的500条)。在分析时,选择特定事件去重后的所有数据集。从实验数据分析看,微博情感随时间的变化与事件本身的特性有关,对于突发事件,通常情感随着时间变化快速达到高潮,然后迅速消退。目前情感词典不可能包含微博上一些口语化的情感词汇,这在很大程度上影响了情感分析的效果。

### 4.3 实验结果分析

#### 4.3.1 情感词的标注实验

下面分别从性别、地区、情感强度和情感类型4个方面对情感词的标注实验进行分析。

1) 性别。在采集的微博数据中,男性网民比女性网民数量多。对于“大老虎”事件,性别不同,对事件的关注度也不同,参与评论的男性普遍多于女性。可见在通过社交媒体参与评论方面,男性比女性更关注社会政治。

2) 地区。图4显示,对于“大老虎”事件,得到的6种情感强度值的比较,从不同的省份(5个省份)得到的调查结果看,情感强度有微弱的差别,也可能是样本数据量不够充分,没有体现出地区间的差异。总体上,不同地区的情感类型和强度差异不大,说明网民对该事件的关注与地区位置无关。

3) 情感强度。对不同事件进行调查,得到情感词的初始强度。从图5可以看出,不同事件中情感词的情感强度差异较大,比如“愤怒”的强度较大,而“惊讶”和“恐惧”的强度较低。说明对于社会事件,网民表现“愤怒”这种情感时比较强烈,而表现“惊

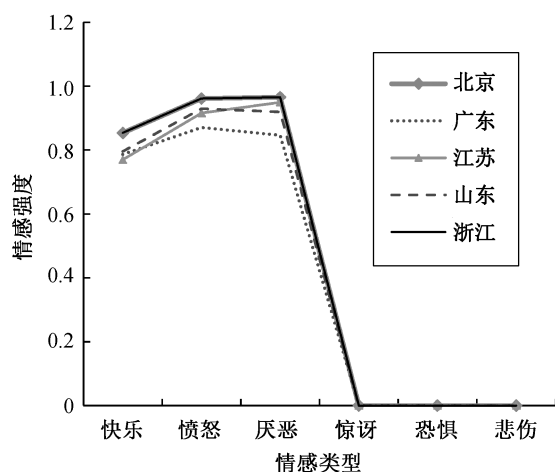


图4 “大老虎”事件中不同省份的情感强度  
Fig. 4 Emotional intensity of “big tiger” event in different provinces

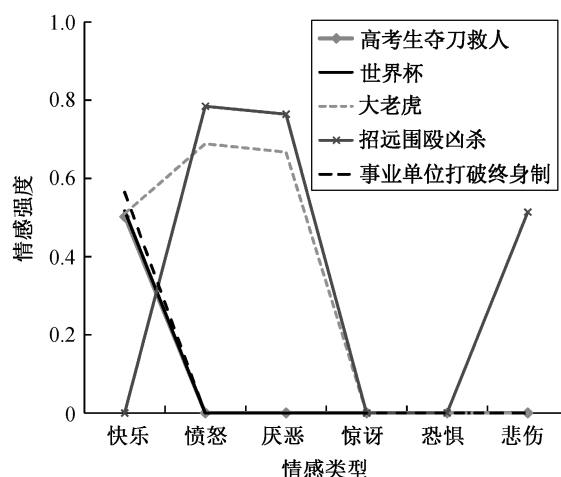


图5 不同事件的情感类型和强度  
Fig.5 Emotion types and intensity of different events

讶”和“恐惧”情感时强度比较柔和。

4) 情感类型。在标注情感词过程中,有些地区可能不出现某种情感强度或者某类情感强度较低,原因可能是发表的微博数量过少。针对这种情况,我们假设该地区该种情感强度为0。

从表1中可以看出,不同事件对情感的类型也有影响。例如,“高考生夺刀救人”事件,大家更多表现出快乐的情感;同一事件可以体现多种情感,例如,“大老虎”事件,大家既表现出快乐的情感,也表现出愤怒和厌恶的情感。

#### 4.3.2 微博情感分析实验

1) 相同事件的分析。从图6和7可以看出,对于“招远围殴凶杀”事件,不同地区网民表现出不同的情感,主要为愤怒、悲伤和厌恶。黑龙江省网民的愤怒和悲伤情感最强;每类情感各地区也有差异,悲伤类差异最小。

针对“招远围殴凶杀”事件,我们对比了本文方法和情感单词匹配方法(在微博文本中搜索情感单词,一旦搜索到,就把该情感词的情感作为该条微博的情感类型)。由于每条微博文本所包含的情感类型较为单一,所以本文方法和情感单词匹配方法在情感类型判断的准确率和召回率相差不大。在测试的样本中,以厌恶情感为例,本文的方法获得的准确率和召回率分别为0.6379和0.5441,而情感词匹配的方法获得的结果为0.6121和0.5221。

2) 不同事件的分析。从图8可以看出,同样是高兴情感,不同的事件引起的情感强度却不相同。各地区的情感图中,“事业单位打破终身制”普

表 1 不同事件的情感类型和强度  
Table 1 Emotion types and intensity for different events

情感类型	高考生夺刀救人	世界杯	大老虎	招远围殴凶杀	事业单位打破终身制
快乐	0.502	0.517	0.510	0.000	0.565
愤怒	0	0	0.688	0.784	0
厌恶	0	0	0.667	0.764	0
惊讶	0	0	0	0	0
恐惧	0	0	0	0	0
悲伤	0	0	0	0.514	0

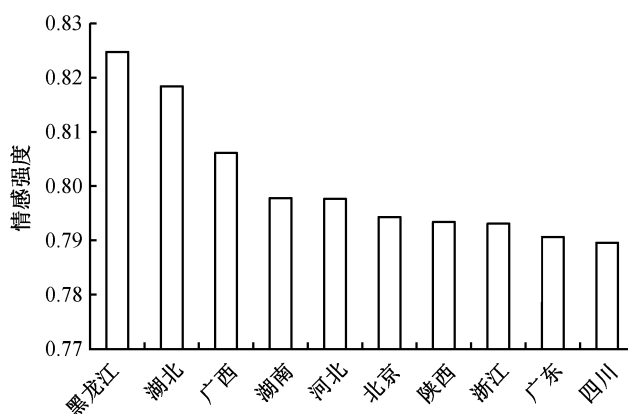


图 6 “招远围殴凶杀”事件中的愤怒情感

Fig. 6 Angry emotion of “Zhaoyuan hit homicide” incident

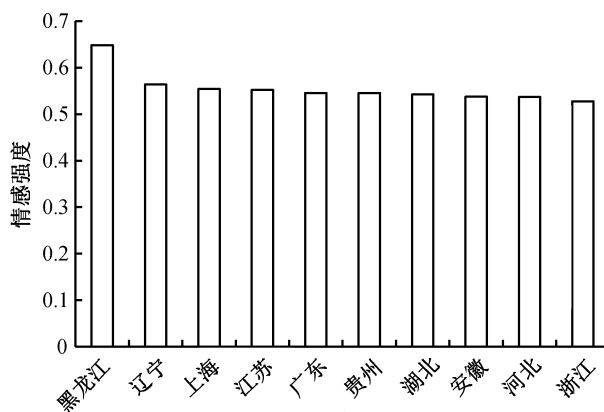


图 7 “招远围殴凶杀”事件中的悲伤情感

Fig. 7 Sadness emotion of “Zhaoyuan hit homicide” incident

遍比“高考生夺刀救人”的情感强度值高，可能是由于前者涉及全社会民众的切身利益，关注的人更全面，从而情感强度更强烈。

3) 情感可视化。本文采用立体的脸谱(6种表情和1种中性表情)与中国地图相结合的方式，对实验结果进行可视化呈现，然后结合不同地区的微博

数目进行整体分析，其中人头越大表示该地区情感影响力越大(考虑情感强度和微博数量)。

从图9可以看出，针对“打破事业单位终身制”事件，主要表现的是“快乐”的情感，说明民众对该事件普遍持支持的态度。广东地区对此事件的反映最强烈，部分地区对此事件没有明显反映，例如台

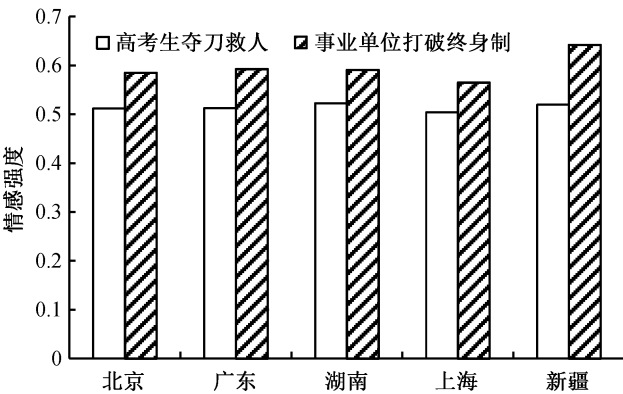


图 8 5 个省市高兴情感的对比  
Fig. 8 Happy feelings comparison of 5 provinces

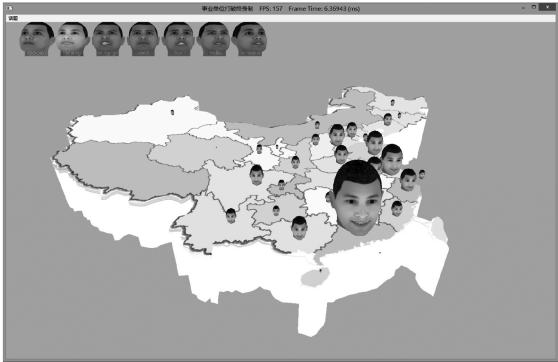


图 9 “打破事业单位终身制”事件中高兴情感的可视化显示  
Fig. 9 Happy emotion visualization in the event of “Breaking tenure institutions”

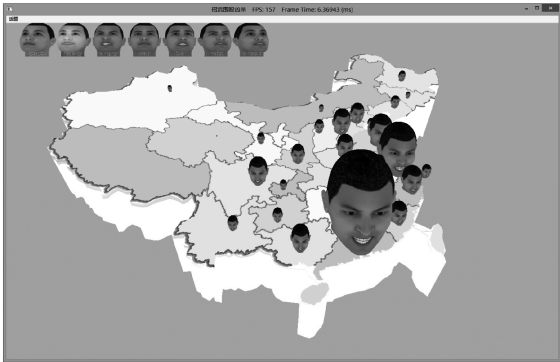


图 10 “招远围殴凶杀”事件中愤怒情感的可视化显示  
Fig. 10 Angry emotion visualization of “Zhaoyuan hit homicide” incident

湾和西藏等。通过可视化显示，可以看出各地参与此类事件讨论的积极程度。

由于该微博文本所包含的情感类型并非单一，情感单词匹配方法容易导致对微博文本真正情感类型的误判，以高兴情感为例，本文方法比情感单词匹配方法在情感类型判断的准确率和召回率更有优势。在测试的样本中，仍以高兴情感为例，本文方法获得的准确率和召回率分别为 0.7831 和 0.5462，而情感词匹配的方法获得的结果为 0.7195 和 0.4958。

从图 10 可以看出，对于“招远围殴凶杀”事件，从微博数量和情感强度综合来看，广东、山东等地的参与程度较高。原因可能是事件发生地在山东，会引起山东网民的积极关注；广东地区可能是由于的网民通过网络参与社会事件的意识较高，更关注社会热点事件。

通过对上述各类实验结果的分析发现，凶杀

类、政治类事件更容易引起民众的愤怒情感，同时伴随强烈的恐惧情感。对于这类事件，通过社会媒体的参与，容易形成人群的负面情感，对社会稳定形成不良影响，应当引起政府足够的重视。

5 结语

随着近年来群体性事件的不断攀升，人群管理正面临各种新的挑战。以微博为代表的社交媒体是了解民众群体情感的重要载体，社交媒体爆发式的信息传播方式容易引发负面情感的迅速扩散，一旦失控，将导致群体性事件变得不可预测和难以控制。如何从微博的海量数据中分析民众的群体情感，一直是政府和学术界关注的热点。本文在已有微博情感分析研究的基础上，提出微博群体情感计算框架，建立了基于人工标注统计的情感单词强度计算方法，并结合依存句法分析，计算微博文本的情感类型和强度，结合地理位置，采用表情人头来



形象呈现群体的情感类型。实验结果表明,该方法能够有效地分析群体情感,通过分析不同事件来反映群体的情感类型。本文方法可用来预测社会事件可能造成的人群负面情感,为和谐社会的人群情感管理提供一种高效可视化方式,为基于大数据分析社会事件提供一种新思路。限于数据资源和水平,本文的工作还需要进一步完善。未来的工作将考虑补充数据资源和情感词典,在微博可视化呈现中加入体感交互虚拟人,实现更真实的可视交互体验。

### 参考文献

- [1] 王飞跃. 社会信号处理与分析的基本框架: 从社会传感网络到计算辩证解析方法. 中国科学: F 辑, 2013, 43(12): 1598–1611
- [2] Ren Fuji, Kang Xin. Employing hierarchical Bayesian networks in simple and complex emotion topic analysis. *Computer Speech and Language*, 2013, 27(4): 943–968
- [3] Caro L D, Grella M. Sentiment analysis via dependency parsing. *Computer Standards & Interfaces*, 2013, 35(5): 442–453
- [4] Loia V, Senatore S. A fuzzy-oriented sentic analysis to capture the human emotion in Web-based content. *Knowledge-Based Systems*, 2014, 58(1): 75–85
- [5] Wang Changbo, Xiao Zhao, Liu Yuhua, et al. SentiView: sentiment analysis and visualization for internet popular topics. *IEEE Transactions on Human-Machine Systems*, 2013, 43(6): 620–630
- [6] Lee S Y M, Chen Ying, Huang Churen, et al. Detecting emotion causes with a linguistic rule-based approach. *Computational Intelligence*, 2013, 29(3): 390–416
- [7] Mohammad S M, Turney P D. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 2013, 29(3): 436–465
- [8] Francisco V, Gervás P. Emotag: an approach to automated markup of emotions in texts. *Computational Intelligence*, 2013, 29(4): 680–721
- [9] 冯时, 付永陈, 阳锋, 等. 基于依存句法的博文情感倾向分析研究. *计算机研究与发展*, 2012, 49(11): 2395–2406
- [10] 徐琳宏, 林鸿飞, 潘宇, 等. 情感词汇本体的构造. *情报学报*, 2008, 27(2): 180–185
- [11] 任磊, 杜一, 马帅, 等. 大数据可视分析综述. *软件学报*, 2014, 25(9): 1909–1936
- [12] Zhang Jinson, Huang Maolin. 5Ws model for bigdata analysis and visualization // 2013 IEEE 16th International Conference on Computational Science and Engineering. Sydney, 2013: 1021–1028
- [13] Mukhopadhyay S, Palakal M, Maddu K. Multi-way association extraction and visualization from biological text documents using hyper-graphs: applications to genetic association studies for diseases. *Artificial Intelligence in Medicine*, 2010, 49(3): 145–154
- [14] Zhang Chenghai, Liu Yuhua, Wang Changbo. Time-space varying visual analysis of micro-blog sentiment // The 6th International Symposium on Visual Information Communication and Interaction. New York: ACM, 2013: 64–71
- [15] Kazemzadeh A, Lee S, Narayanan S. Fuzzy logic models for the meaning of emotion words. *IEEE Computational Intelligence Magazine*, 2013, 8(2): 24–49