



多媒体社会事件分析的研究与展望

摘要

随着互联网的快速发展,近年来出现了越来越多的社交媒体网站.用户可以非常方便地在这些网站上分享想法、图片、帖子和从事其他相关活动.当一个流行事件发生时,它可以在不同社交媒体网站中快速地进行传播,同时产生大量的多媒体数据.因此,研究多媒体社会事件分析是非常必要的,能够让人们及时地了解社会事件随时间演变的发展趋势.本文对近年来多媒体社会事件分析的相关研究进行了综述,并针对多媒体社会事件表示、多媒体社会事件检测和跟踪、多媒体社会事件动态演变和多媒体社会事件主题观点分析这4个具体任务进行了详细总结.随后,对多媒体社会事件的发展趋势进行了介绍.最后,对多媒体社会事件分析方面未来可能的研究课题进行了展望.

关键词

多媒体;社会事件;多模态;跨平台

中图分类号 G201;G206

文献标志码 A

收稿日期 2017-08-28

资助项目 国家自然科学基金(61225009,61303173,61572498,61432019)

作者简介

钱胜胜,男,助理研究员,主要研究方向是多媒体数据分析及应用. shengsheng.qian@nlpr.ia.ac.cn

徐常胜(通信作者),男,研究员,博士生导师,主要研究方向是多媒体数据分析及应用. esxu@nlpr.ia.ac.cn

0 引言

社会事件是发生在特定的地点和时间的特定行为,它由许多随时间发生的子事件组成.热点社会事件分析是伴随互联网发展起来的技术,特别是社交媒体网络的盛行,它不仅能减少人们浏览庞大的互联网信息所耗费的精力和时间,协助人们在浩瀚的网络数据中快速、准确地了解所关注的社会事件,而且能够为国家宏观把握、监控网络舆论走势提供技术支撑.如图1所示,如果能在地图上可视化整个事件随时间发展的演变过程,用户更易于理解整个事件的发展趋势,不需要花费大量的时间查看海量的事件数据.因此,对社会事件的分析研究是十分必要的.多媒体社会事件分析技术以社会多媒体事件为研究对象,其主要目的是通过研究现有的社会事件多媒体数据,实现发现子事件、梳理事件的发展脉络、实时监控事件的演变动态、挖掘不同平台的事件观点等.在当今时代,互联网作为信息传递的媒介早已在世界经济、政治和社会建设中发挥至关重要的作用.互联网同时也是社交媒体事件的传播器和放大器,互联网中的社会热点事件更容易受到公民的广泛关注和参与.在此影响下,多媒体社会事件分析的研究显得尤为重要.

根据 Google 统计,在 2014 年全球网站数量就突破 10 亿大关,到 2016 年,全球互联网用户数已经超过 30 亿.在国内,据《中国互联网络发展状况统计报告》报道,至 2016 年 6 月,中国网民规模已达 7.10 亿,网民每天平均上网时间接近 3.8 h.另外,随着移动设备和移动互联网的普及,促使新型社交媒体网站呈现爆发式增长,诸如 Facebook、Flickr、YouTube、Google News 等新型社交网络平台往往拥有大量的用户基础,通过这些媒体平台,用户能够方便地发表评论、分享经验和获取新闻资讯.因此,当一个社会流行事件发生时,在社交网络上每个人都是事件的传播者和评论者,社会热点事件会在大量网络用户的交互参与下迅速发酵,同时会产生大量的多媒体数据.如图 1 所示,2011 年的“英国暴乱”这一事件在爆发之后短短 4 d 时间,在网络上迅速演化、蔓延,在互联网中产生了大量相关的多媒体数据.

然而,由于社会多媒体数据主要来自互联网中不同的社交网络平台,数据呈现出跨平台(不同的社交媒体平台,例如 Facebook、Flickr 和 YouTube)、多模态(例如文本、图像和音频)、大规模、噪声大等特点.对于这 4 大特点,传统社会事件分析在检测精度和监控效率

1 中国科学院自动化研究所 模式识别国家重点实验室,北京,100190



图1 “英国暴乱”事件蔓延趋势
Fig. 1 Spread of the event “Britain Riots”

方面面临更高的要求和挑战. 社会事件数据的复杂性主要体现在如下4个方面:

1) 社会事件数据具有跨平台特性

社交媒体数据的跨平台特性也称数据多源性, 是指同一社会事件对应的多媒体数据会分布在不同的社交媒体平台中, 需要分别获取不同社交媒体平台中的媒体数据进而分析社会事件. 当前社交媒体网站数量呈现爆炸式增长, 不同社交媒体网站可能会在内容上提供类似的事件信息, 在形式上呈现出差异化表示. 例如 Google News 和 Flickr. 这些相关的数据有不同的视角, 在 Google News 上的信息是官方的, 在 Flickr 上的是信息是非官方的, 其中有个人的评论和有趣的照片. 对于同一社会事件, 由于这些不同平台是从不同角度表现数据的, 所以简单地对某一个平台的数据进行单独的分析处理很难全面分析相应的社会事件. 如微博是当前流行的社交媒体平台, 其主要特点是用户群体对新闻热点进行即时评论和转发, 一般是用户简短的观点; 博客的博文更偏向于日记的形式, 用户通过较多的篇幅详细介绍一个主题. 由此可知, 不同平台的数据有所侧重, 各有特点, 可以利用跨平台数据的不同视角互补从而更全面准确地描述社会事件, 因此, 跨平台社会事件的研究是非常必要的. 但是, 由于不同平台的数据结构

和描述重点不同, 使得跨平台社会事件分析面临重大挑战, 如何设计通用的跨平台社会事件分析方法变得更加重要.

2) 社会事件数据具有多模态特性

在互联网发展的初期, 文本通常是网络中的数据的主要表现形式, 伴随着移动设备的普及和微博等社交媒体平台的蓬勃兴起, 社会事件往往会产生大量网络媒体数据, 包括文本、图像、视频、音频等多媒体数据模态. 如“美国总统竞选”事件, 在竞选期间, CNN 新闻、推特、Facebook 等媒体平台上不仅会发布和分享很多关于总统竞选的文本信息(如演讲词、新闻评论等), 而且会上传大量图片、视频等数据(如反对者游行图像、辩论视频等). 图片、视频等数据模态能生动直观地使用户明白事件的概况, 帮助用户了解事件, 文本信息能让用户更加详细深刻地研究事件的细节内容. 对于同一个社会事件, 尽管不同的用户在媒体平台上表述的文本内容可能不同, 但是其平台上的视觉信息可能是相同或者相似的. 因此, 虽然不同模态的数据的表现力有所差别, 但是这些多模态信息可以相互补充, 能够帮助用户全方位、深层次理解事件. 但是, 由于网络环境十分复杂, 社交媒体数据多模态化的同时也造成传统的文本分析技术不适用于现阶段的社会事件分析任务. 对于

多模态数据,如何对文本数据以外的其他数据特征进行合理表示、如何有效建立2种模态的语义关联、如何设计统一的多模态模型等,都是目前急需解决的技术难题。

3) 社会事件数据具有大规模特性

在移动互联网和 Web 2.0 发展的浪潮下,社交媒体网站蓬勃发展,互联网用户的数量急剧增加,导致社交媒体数据呈现大规模或超大规模的特性。据统计,2016 年 Twitter 月活跃用户 3.10 亿,Facebook 的月活跃用户达到 16.5 亿。在中国,2016 年腾讯 We-Chat 第二季度月活跃用户达 8.06 亿,QQ 月活跃用户也超过 8 亿,而且这些数字还在呈增长的趋势。社交媒体网站具有大量的用户群体,用户成为媒体数据的产生者和传播者,用户产生的数据成为媒体主流势必使社交媒体数据的数量呈现爆炸式增长。另外,智能手机、监控摄像头等智能设备是数据的主要来源,用户通过智能手机在微博、Twitter 等社交网络中分享和传播信息,而监控摄像头则更是无时无刻不在制造视频数据。由上述描述可知,大规模数据的产生是不可避免的,社会事件分析技术和大数据处理技术正是在此背景下应运而生的。大规模的数据分析非常有用,对某条公路长期监控录像分析,可以获得汽车流量和某一时段路况等信息,为公民出行提供道路建议。对“美国总统选举”事件,可以根据 Twitter 数据或新闻媒体数据等统计分析,分析候选人支持率从而推断获胜概率等。大规模数据信息给社会事件分析带来很大的好处,同时也面临诸多严峻的挑战。

4) 社会事件数据具有噪声大、信息冗余的特性

社交网络平台中的数据大多是基于用户分享的,而由于用户本身产生的数据往往有结构不规则、信息不完整、描述偏向口语化、评论趋于感情化等因素,其产生的数据很大程度上充满了噪声。例如,对于“朴槿惠亲信门”和“美国总统选举”事件,网友产生的文本信息可能只是“支持、反对、发起抗议”等没有辨识力的短语或短句,用户上传的图片信息也可能包含很多图像表情等,不能描述事件的特点。另外,网络数据普遍存在重复、冗余的特点,当用户浏览互联网数据的时候,可能只是倾向于了解最近发生了什么事、某事最近的发展情况等,然而噪声干扰搜索引擎只能返回大量相关数据,显示的信息重复而且冗余。又如,在监控视频中,有价值的信息可能仅有几秒,但由于传感技术的限制,必须要同时存储

和处理大量的噪声。结果是,用户必须反复地来回切换,以便完全理解事件的主体。通常情况下,用户浏览这些庞大的文件是非常耗时的,几乎不可能捕捉到整个事件的演化过程。除此之外,由于互联网环境复杂,数据噪声可能会含有大量虚假内容,这不仅增加社会事件分析的难度还会降低分析精度。所以,互联网中的社交媒体数据噪声大、信息冗余的特性,成为社会事件分析中亟待解决的难题。

综上所述,由于社会事件数据来自不同的网站,具有多跨平台、多模态、大规模、噪声大等特点,所以基于社会事件的分析研究非常具有挑战性。如何对上述社交媒体数据特点进行处理,研究社会事件分析方法、设计有效的社会事件分析模型成为未来社会事件分析研究的关键问题。基于社交媒体数据的研究也从未间断,自 1996 年美国国防高级研究计划署提出话题检测与跟踪项目(Topic Detection and Tracking, TDT)以来,多媒体社会事件分析技术逐渐成为多媒体领域内的研究热点,越来越受到国内外学者的关注。从 1998 年开始,在 DARPA 的支持下,美国国家标准技术研究所(NIST)每年都要举办专门 TDT 国际会议,进行相应的系统评测。虽然国内学者对相关工作的开展起步较晚,但发展速度较快并取得了骄人的成绩。本文总结多媒体社会事件分析领域已有的技术,尽可能全面地分析已存在的方法的优缺点。本文重点介绍多媒体社会事件分析领域中典型的研究方向:多媒体社会事件表示、多媒体社会事件检测、多媒体社会事件跟踪与演变分析、多媒体社会事件主题观点挖掘,从这 4 个方向入手总结近几年国内外学者在多媒体社会事件领域的主要研究进展,分析国内外主要机构的研究特色和差距,并对未来的发展趋势进行展望。

1 研究现状

多媒体社会事件分析研究顺应移动互联网和社交媒体网站快速发展的趋势,可用于社会热点事件分析、社会舆情分析和预测等领域,并涉及到自然语言处理、数据挖掘、机器学习、模式识别等多个研究领域的知识。本章将主要介绍社会事件分析中的多媒体社会事件表示、多媒体社会事件检测与跟踪、多媒体社会事件动态演变分析和多媒体社会事件主题观点分析 4 个方面。

1.1 多媒体社会事件表示

多媒体社会事件表示是指利用社会事件对应的

多媒体数据信息获取具有良好判别力的社会事件表示,从而进一步执行其他相关社会事件分析任务.多媒体社会事件研究强调社会事件的载体是互联网多媒体数据,多媒体数据包括文本、图像、视频等多模态信息.传统的社会事件表示主要是基于文本进行研究,随着互联网的发展,多模态数据的普及,基于图像等其他模态的社会事件分析技术逐渐被提出,基于多模态信息的社会事件分析越来越流行并且受到研究者的重视.

1) 文本特征表示

文本特征表示是将文档内容转换为计算机可识别的信息,因为必须将其转化为一个简洁的、统一的、能够被学习算法和分类器识别的结构化形式,从而应用到其他文本处理中.目前最常用的文本表示方法是将文本表示为向量的形式.词袋模型(Bag-of-Words, BoW)^[1]是最常用文本表示方法之一,但是BoW模型只包含了单词在词典中的索引和词频信息,而忽略了文本的其他信息,实质上造成了文本语义信息的缺失,文本的表示向量一般为字典长度,造成向量维度灾难.词袋模型很长一段时间是文本表示应用的主要方法,在传统词袋模型的基础上,有2个主要研究问题:一个是特征选择问题,另一个是计算特征权重问题.常见的特征选择方法有:特征频率、互信息、期望交叉熵等,目前最常用的方法是特征频率^[2],这是一种基于统计的方法,把度量值小于给定阈值的特征过滤掉,大于给定阈值的特征认为是有效特征.以上方法是基于文本的向量表示方法,文献^[3]提出了一种基于图的文本表示方法,这个方法首先根据自己所定义的特征把文档转化为特征文本,然后在特征文本上构图,并且定义了计算图之间相似度的度量公式,这样就可以通过计算图的相似度计算出文档的相似度,这种方法弥补了上述方法的不足,考虑了文本特征之间的有序关系,但是由于构建图的过程中涉及到过多的参数,其实验效果远不如向量空间模型好.

随着研究的逐渐深入,人们对文本表示的研究逐渐向基于语义的文本表示方法转移,从原来的特征选择转变为语义特征学习研究^[4-6].语义特征学习的方法获得的文本特征表示可以认为是文本的深层表示.这种方法弥补了词袋模型和向量空间模型的缺陷,在一定程度上考虑了文本的语义信息.主题模型LSI(Latent Semantic Indexing)^[4]、PLSI(Probabilistic Latent Semantic Indexing)^[5]、LDA(Latent

Dirichlet Allocation)^[6]等模型,通过无监督的学习,从文本中提取语义信息,也就是平时所说的“主题(Topic)”.其中,LDA模型将每一篇文档表示为一系列主题组成的概率分布,把每一个主题表示为一列单词所组成的概率分布.这个模型能够识别大规模文档集中潜在的语义信息.LDA模型凭借其在文本表示中表现出的优良性能,吸引了很多研究者对其进行深入的研究,并在此基础上,提出了大量的改进算法.文献^[7]在LDA模型的基础上,提出了一种有监督的隐含狄利克雷分布模型,利用文档语料库中的监督信息优化学习过程,从而获得更好的文本表示.近年来随着深度学习发展的日益成熟,深度学习^[8-9]在文本表示领域也获得了巨大的成功.词向量(word embedding)是深度学习用于文本表示的核心技术,也是深度学习应用在自然语言处理领域(NLP)中的关键技术之一.Word embedding是由Hinton^[10]提出的一种词向量表示方法,区别于one-hot表示方法^[11],把文本中的每一个单词对应向量中的每一维,易造成维数灾难的情况,word embedding将每个词映射为一个低维的实数向量,所有这些向量构成词的向量空间,每个词可以看作是向量空间中的一个点,这样,语义相似的词在向量空间中的距离就会更近.Mikolov等^[12]提出的Skip-gram模型是一种对文本集进行快速训练获得word embedding的模型,该模型主要思想是用当前词来预测其上下文,具有良好的时间性能.随着word2vec^[12]工具的提出,研究者开始越来越重视词嵌入模型的探索^[13-14].循环神经网络(Recurrent Neural Network)由Elman等^[15]在1990年首次提出.该模型的核心是通过循环方式逐个输入文本中的各个词,并维护一个隐藏层,保留所有的上文信息.但是循环神经网络的语义都会偏向文本中靠后的词.因此,循环神经网络很少直接用来表示整个文本的语义,由于其能有效表示上下文信息,因此被广泛用于序列标注任务.

2) 图像特征表示

图片是除文本以外的最常见的社交媒体数据的表现形式,传统的图像表示方法是采用类似于文本表示的词袋模型(BoW)^[1,16],其主要步骤为:提取局部特征、构建词典、特征编码、特征池化.虽然传统词袋模型(BoW)常被用于图片检索领域,但是其忽略了以下问题:首先,图像局部特征不同于文本中的单词能够表示一个独立的语义单元信息,其次,在图像词袋模型中,大部分方法为了提高编码效率,都采用

K 均值聚类方法对局部特征描述子进行聚类从而得到词典^[17-18],进而将图片表示为这些视觉单词的特征向量,这使得大部分局部特征描述子丢失了其判别性.其聚类中心是由出现最多的局部特征描述子决定的,而不一定是包含图片最多的图片内容信息的描述子.由于上述基于图像词袋模型的特征表示方法一定的问题,也有其他研究人员提出其他图像编码方法.文献[18]提出一种新的非参数朴素贝叶斯最邻近(Naive-Bayes Nearest-Neighbor, NBNN)编码方法,克服了 BoW 中局部特征描述子判别性差的缺点,使局部特征描述子不需要进行编码量化,但是这种方法在执行效率方面并不尽如人意.近年来,随着深度学习发展的日益成熟,深度学习在图像特征表示领域的优势越来越明显,特别是基于监督学习的深层卷积神经网络在图像分类中取得的重大突破^[19],此后更多的深层神经网络相继被提出^[20-23],这些深层次的神经网络结构都取得了很好的效果.

3) 多模态特征表示

移动互联网时代,网络中不仅存在大量的文本数据,而且存在大量的其他类型的数据.图片、视频等数据中往往富含大量的信息,并对文本信息有一定的补充作用,然而,传统的基于单模态的特征表示方法忽略了事件的多模态属性.如何充分利用社交媒体多模态信息已成为社会事件分析的研究热点问题之一.本部分主要探讨近年来不同模态数据结合表示的相关方法.早期的多模态数据表示常采用相关性分析方法^[24-27],如典型相关性分析(Canonical Correlation Analysis, CCA)^[24].典型相关分析是研究 2 组变量之间相关性的一种多元统计分析方法,这种方法利用变量之间的相关关系来反映 2 组指标之间的整体相关性.典型相关性分析方法没有考虑变量组内部的诸变量之间的关系,仅仅考虑变量之间的相关性.在此基础上,文献[25]用典型相关性分析方法把文本和图像映射到同一特征空间,然后对图像提取语义特征表示.这种方法把文本和图片 2 种模态之间的相关性以及用逻辑斯蒂回归模型获得的语义特征表示结合起来.文献[28]使用典型相关性分析方法提取概念之间的语义关系,其数据集包含标注的图像.除此之外,文献[26]把多核学习(Multiple Kernel Learning, MKL)应用在多模态数据表示中,对文本和图片标注信息分别建立核函数,通过结合不同的核函数达到特征融合的目的.文献

[29]通过学习视觉特征和听觉特征之间的关联性作为互补信息,并用于对图像-音频数据集进行聚类分析.近年来,基于主题模型的多模态特征融合方法受到更多研究者的注意.文献[30]提出 Corr-LDA 模型,捕捉图片及其注释在主题上的关系.文献[31-33]提出一种多模态隐含狄利克雷模型(mm-LDA),同样也使用主题模型对文本和图片进行建模.上述 2 种方法都是在主题空间上对文本和视觉 2 种模态进行特征表示,2 种方法都是假设文本和图片 2 种模态的主题是一一对应的.文献[32]研究了图像和视频的注释 2 种模态,考虑了不同主题之间的关系.文献[34]针对社会事件数据的跨平台多模态特性,提出非参贝叶斯的协同学习框架,通过采用共享域先验和共享模态先验的方式,实现跨平台多模态事件数据的语义层次关联.类似的,文献[31]还提出多模态概率模型挖掘用户的主题信息,通过对用户的文本注释信息和用户上传的图像信息进行建模,并将该多模态主题模型用于协同推荐和个性化图像搜索中.文献[35]提出对多媒体文档中的文本和图像内容同时建模,分析多模态的语义主题关联,并应用于跨媒体检索问题.文献[36]认为,社会多媒体数据类型不限于模态多样,也包括各种类型的媒体数据,比如社会链接、地理信息等.文献[37]提出利用社会媒体的属性信息如标签、空间和时间等特征构建融合的事件特征表示,并解决缺省属性值的问题,最终通过分类性能验证了提出的特征表示的有效性.最近,基于 word embedding 的多模态数据表示成为新的研究方向.基于主题模型和 word embedding 思想的混合模型被提出^[38-39].近年来,深度学习也被用于多模态特征表示学习中^[40-41],主要思想就是用不同模态的数据同时训练多层自编码器(autoencoders)或者受限玻尔兹曼机(RBM),从而得到共享的特征表示来对不同模态数据进行统一表示.

1.2 多媒体社会事件检测与跟踪

多媒体社会事件检测与跟踪是指利用事件多媒体数据进行分析,进而实现发现新事件、检测不同种类的事件以及跟踪各个子事件等任务.社会事件检测与跟踪起源于话题检测与跟踪(Topic Detection and Tracking, TDT)^[42],从 1998 年开始,在 DARPA 的支持下,美国国家标准技术研究所(NIST)每年都要举办专门 TDT 国际会议,进行相应的系统评测.话题检测主要包括文本的分类任务和聚类任务,最早的文本分类的思想是美国 IBM 公司的 Luhn 在 20 世

纪50年代所做的研究,到1960年,Maron^[43]最先提出基于关键词的自动分类技术,Maron利用贝叶斯公式自动对文本进行分类,在文本分类领域做出了巨大贡献.之后提出的大量的分类算法可用于文本分类,包括朴素贝叶斯方法^[44-45]、决策树(Decision Tree)算法^[46]、K-最邻近(KNN)算法^[47]、支持向量机(SVM)^[48]等.聚类方法是一种无监督的机器学习问题,大致可把聚类算法分为层次聚类算法^[49]、划分式聚类算法^[50]、基于密度和网格的聚类算法^[51].层次聚类算法使用数据的连接规则,通过一种层次架构方式,反复将数据进行分裂和聚合.划分式聚类算法需要预先规定聚类数目,经过反复迭代,逐步降低目标函数误差值,当目标函数收敛最终聚类完成.除此之外,在聚类算法领域,文献[52]早期做了很多研究工作,分别从理论、算法和应用3个层次来讨论聚类和分类技术,全面分析了聚类和分类算法的关键技术,总结了在统计、机器学习和模式识别等领域的聚类、分类算法.

传统事件检测基于上述基础分类算法和聚类算法,一定程度上实现了相关功能,但是检测效果不尽如人意.近年来,随着社会事件数据在互联网各个社交媒体平台上的大量增长,研究者提出了一系列方法^[37,53-58]对社会事件数据进行建模,以期达到更好的检测效果.这些方法主要分为基于单模态分析和多模态分析这两类.在单模态数据分析中,存在的方法主要采用文本信息(比如,事件名字、事件时间、事件位置、标题、标签和描述等)或者视觉信息(比如,图像和视频)^[55-56]来建模社会事件.文献[55]通过分析和事件相关的Twitter推文信息,对事件可视进行研究分析.Makkonen等^[56]提取有效的语义特征,比如名字、时间和位置等,然后学习了一个基于单个簇分区的相似性度量.Becker等^[57]利用和社会媒体数据关联的丰富的上下文信息并且使用一个新颖的聚类算法进行事件识别.清华大学的朱军等^[59]通过利用Max-margin的思想将SVM分类器思想融合到传统主题模型中,利用新的损失函数训练模型,大大提高了主题模型在多标签分类任务的准确度.然而,这些单模态的方法忽略了事件的多模态属性,不会好于多模态的方法.为了解决上述问题,许多研究者提出了多个不同特征相结合(比如,时间、标签、位置特征、图像和视频等)来进行社会事件数据的表示^[37,53-54,58].文献[37]探究事件媒体数据的不同特征以及如何处理社会媒体数据中的缺失值.文献

[58]计算事件文档的相似性,是基于单个不同的特征,比如时间、标签和位置信息.文献[53]通过使用机器学习方法训练的模型和多个特征建模文档的相似性,媒体文档被分类到相关的事件中.文献[54]对每类媒体事件构建了一个特征独特性度量的组合,应用到事件识别框架.文献[60]利用主题模型建模多模态社会媒体数据,为处理大规模数据难以建模的问题,将模型融合到一个Boosting算法中,构建了大数据集下的多媒体社会事件检测框架.

传统的话题跟踪(Traditional Topic Tracking, TTT)主要包括基于知识和基于统计的2种研究方向.前者的核心问题是分析报道内容之间的关联与继承关系,通过特定的领域知识将相关报道串联成一体.后者则根据特征的概率分布,采用统计策略裁决报道与话题模型的相关性.基于社会媒体的社会事件跟踪是社会事件分析领域传统且富有挑战性的话题,目前已经有大量的社会事件跟踪相关的研究^[61-63].事件跟踪主要有2个挑战性的因素.首先,社交媒体包含了大量的非结构化的多模态媒体数据,它不同于传统的只有单一模态的数据,比如只有文本数据.在不同的社交媒体平台,社会事件具有丰富的多模态信息形式,比如文本、图片、视频.这些多模态信息形式可以互相补充,能够更好地帮助用户理解事件^[64-65].例如,对于同一个社会事件,尽管不同的用户在媒体平台上表述的文本内容可能不同,但是其平台上的视觉信息可能是相同或者相似的.然而,目前的一些工作可能都是集中研究某一种形态比如文本或者图片^[56,66],而我们需要综合考虑多模态数据进行社会事件的分析.其次,在社会事件的跟踪过程中,各个媒体网站可能具有相似的事件报道,比如“占领华尔街”、“美国总统选举”,它们都具有相似的主题,类似“美国,政府,总统”.但是,由于很多信息都是由用户产生的,所以有可能会存在一些噪音.例如,评论和相应的图片可能表述的事件没有关系.这些问题很可能导致在线的社会事件跟踪产生模型漂移.所以设计一个多模态融合策略和避免模型漂移问题的在线跟踪方法对于社会事件跟踪是非常有必要的.

针对上述挑战,目前已有学者提出相应的解决方案.有些方法采用视觉信息(例如图像和视频)或文本信息(例如名称、时间参考、位置、标题、标签和描述)^[57,66-67]对事件进行检测并跟踪建模事件数据.文献[68]提出了一种新颖的主题检测算法,主要想

法是首先将传入的新闻分类成预定义的类别,然后使用启发式的条件式主题来识别新的事件.文献[57]利用丰富的上下文相关的社交媒体数据,使用聚类算法来识别事件.在社会事件分析中,很少有工作聚焦于一个统一的方式来分析多模态数据.当前,在不同的社交媒体平台中,社交媒体事件具有丰富的多模态信息,例如文本、图像和视频,它们有助于社会事件分析^[64-65].微软亚洲研究院的学者^[69]假设社会媒体网络之间存在中间层主题空间,提出把社会流媒体网络 Twitter 的推文信息和视频分享网络 YouTube 中的视频标签信息投影到共同主题空间中进行数据关联和聚合.最近,多模态的社会事件分析已经受到相当大的关注. Kender 等^[70]研究手动标注的视觉概念(例如网站、人和对象)和文本主题标注之间的相关性,然后利用图形切割技术对话题进行聚类. Zhai 等^[71]提出了一种概念跟踪方法,通过相关文本和关键帧匹配来连接不同电视频道的新闻报道. Zhang 等^[63]提出了一种用于跨域多事件跟踪的 CO-PMHT 算法,它可以通过使用跨域知识来跟踪事件,并随时间获取其摘要信息.

1.3 多媒体社会事件动态演变分析

多媒体社会事件动态演变分析是指根据已有的事件多媒体数据训练相应的模型,从而对即将到来的多媒体数据进行新事件的发现和与历史事件的关联,并通过一定的可视化手段将事件时间节点上发生的子事件进行可视化展示,从而达到方便用户了解事件发展脉络、把握事件发展动态的目的.互联网中含有大量的媒体数据,当人们浏览网页或是搜索感兴趣的社会事件时,往往会得到一系列杂乱无序的相关事件,不利于用户方便地理解社会事件.时序性事件动态演变分析主要针对某个社会事件,方便快捷地可视化出该事件的时序性主题相关信息,使人们方便地理解事件的前因后果从而达到事件时序性动态演化分析的目标.时序性事件动态演化分析的内容是,给定一个社会事件起始时间的数据信息,首先要对后续的社会多媒体事件进行识别,判断该事件的相关数据内容,并提取其主要的主题信息,从而进一步将事件的时序性发展进行可视化.时序性社会事件动态演变分析是社会事件分析的重要的研究领域,与社会事件检测和社会事件跟踪密切相关.许多基于主题模型的时序性主题演变方法被提出来进行动态演化分析.文献[72]考虑了时间动态上的主题模型,获得每个时间片的主题分布;文献[73]组

合传统 LDA 和 SLDA 模型,提出一种跨时空的事件发现方法,为事件动态演变分析提供基础.如何检测大规模流文本的主题,并对主题的演变进行模拟,也是近年来的研究重点之一.文献[74]提出一种在线的主题模型,用于顺序分析文档集中主题的时序性演化,通过提出长时间和短时间等不同尺度的依赖关系,学习鲁棒的时序性主题模型.文献[68]提出了连续时间动态主题模型(cDTM),利用时序观测文本的稀疏性处理离散时间点的数据,但随着事件粒度的增加,模型复杂性也迅速升高;文献[75]利用主题模型分析消费者购买行为随时间的演变过程,以及消费者兴趣随时间的变化,其模型可以根据当前数据和历史估计结果自适应地跟踪兴趣的趋势变化;文献[63]为应对高效地组织和监控多媒体社会事件面临的挑战,提出一种跨域的多个事件跟踪模型,利用不同域的多媒体数据协同合作提高事件的跟踪性能.社会事件动态演化在实际场景中也有重要的应用,如突发事件监控对维护社会安定有极其重要的作用.文献[76]通过研究应急领域 Web 数据,着重解决突发事件不确定性对事件跟踪和演化的影响;文献[77]为应对海量时序性社会多媒体数据中社会事件分析的难题,提出一种基于多模态主题模型的社会事件跟踪和演变框架,不仅能有效捕捉多模态社会事件主题,同时也能够获得社会事件的演变趋势.

话题发现是事件演变分析的前提.文献[78]使用字典学习识别新兴主题,并在 Twitter 时序数据流上进行实验;文献[79]提出一种分组主题模型,改进了传统主题模型主题发现能力.社会事件的发展是有一定的生命周期的,包括事件产生、发展和消亡,其相关话题同样也有一定的生命周期^[80].事件动态演变分析首先需要对社会事件发展过程中的主题进行主题关联,但是由于不同时刻的事件主题变化可能较大,需要对事件演变过程产生的数据进行聚类.文献[81]提出一种演变聚类方法,为保证邻近时刻之间主题不会发生太大的偏差,将时间平滑度整合到聚类算法中,从而使模型对短期噪声不敏感.文献[82]使用 DP 方法自动确定聚类的数目.文献[83-84]将事件依赖性结合 DP 方法,用以建模相邻时间的数据.文献[85]通过在相邻时刻添加时间依赖关系,发现不同时刻的事件演变模式,并在多重文本语料库中进行实验.文献[80]采用分层狄利克雷(HDP)模型发掘事件内容的演变模式,并用可视化

系统呈现实验结果. Wang 等^[86]没有专注于事件的发展演变,转而研究事件主题的消亡行为.文献[87]提出一种主题热度演化模型,引入基于内容和基于连接的热度计算方法,最后通过热度演化模型预测出新兴热点话题.

事件摘要及时序性事件演变可视化的基础工作,多文档摘要可以有效地解决信息过载问题,在过去几十年中备受关注.Gong 等^[88]提出了一个通用的文本摘要方法,从原始文档中提取句子进行排序获得摘要. Haghghi 等^[89]提出了一个新颖的产生式概率图模型用于多文档摘要的分析,利用一个层次 LDA 模型来表示多个语义层次的主题.Zhou 等^[90]提出了一种 2 层摘要生成框架,来总结多个与灾难相关的文档.Wang 等^[91]提出了一种基于句子级语义的多文档摘要分析框架,主要是利用对称非负矩阵分解模型.哈尔滨工业大学的学者对社会事件的跟踪进行了比较深入的研究^[92],通过结合事件抽取、时序分析、数据挖掘等技术,抽取主题事件中的关键信息,并进行时序分析.

1.4 多媒体社会事件主题观点分析

多媒体社会事件主题观点分析是通过分析社会事件多媒体数据,挖掘出不同组织或个人对热点社会事件的观点,也是网络舆情分析的必需技术.观点挖掘应用广泛,不止用于用户观点意见倾向分析,也可用于其他观点分析、观点预测方向.主题观点挖掘是多媒体最重要的任务之一,有着广泛的应用,如基于产品评论的主题观点挖掘^[93]、跨不同的新闻媒体的社会事件主题分析^[34,94-95]、政治争议性观点挖掘^[96]和基于电影评论的主题观点挖掘^[97].但是,传统的观点挖掘方法在准确率和应用范围上仍然有很大的不足.为提高观点分析的准确度,文献[98]提出统一的细粒度观点分析框架,在预测精度和召回率上达到更好的性能.但是其研究观点挖掘主要针对的是文本数据,无法适应社会多媒体数据的观点挖掘任务.文献[99]从实用性应用出发研究客户语音情感和来电意图,通过对声音数据的分析挖掘其情感类别.文献[100]通过综合考虑社会事件多媒体数据的多模态和多视角属性,通过多媒体数据多模态属性互补的特点充分挖掘社会事件的主体,通过分析不同视角的多媒体数据挖掘不同平台的观点差异,分析结果有助于指导用户理解不同媒体的观点倾向.

上面提出了多媒体社会事件分析面临的多模态

问题,另外,社会事件的主体学习在多个新闻媒体中也具有多视角属性.这是因为不同的新闻媒体对于某个热事件可能有相似报道,然而他们也在同一主题中有些主题差异性.通常,来自不同媒体的报告内容信息可以被分为 2 部分:来自所有数据集合(新闻媒体)的共同主题(Common topics)和来自每个集合的特定主题(Collection-specific topics).现实场景中事件观点也具有多视角属性.通过上述主题学习的主题不仅包含描述主题的词,而且包含表达关于主题观点的词.换句话说,我们需要在不同集合中这些学习得到的多模态主题自动识别其观点倾向并发现不同集合中的多视角差异.因此,多视角主题观点挖掘的目的是通过利用多个跨域数据集,联合学习代表性的主题和对应的观点.我们以多个社交新闻媒体的社会事件分析的主题观点挖掘为例,在政治辩论、报纸和许多社交媒体网站(例如 YouTube、Facebook 和 Google News)上有许多政治上有争议的社会事件,这些事件有可能影响成千上万的人,例如“利比亚危机”、“叙利亚内战”和“阿富汗战争”.对于这些事件,每个新闻媒体对一个有争议的话题不仅有许多不同的方面,而且有自己的观点倾向,并且在这些话题中没有绝对的对与错.因此,联合主题观点挖掘是很重要的,能够自动理解具有争议性的社会热点事件的主体以及不同新闻媒体对这个事件的观点倾向.而且,主题观点挖掘可以从社交新闻媒体中大量的多个信息源发现集体和主观的信息,并且挖掘结果可以用于许多应用,例如多视角检索^[96]、观点挖掘和情感分析^[101],以及多视角关联可视化^[102].

如何在不同数据集合上进行多视角的主题观点挖掘已经吸引了很多研究者的兴趣.基本上,主要有 2 个主要的研究课题:

1) 跨数据集合的细粒度主题分析,主要是通过采用扩展的跨数据集合的主题模型来发现所有数据集合中的共同主题,以及建模跨多个集合的相似性和差异.例如,Paul 等^[103]提出了一种跨数据集合主题挖掘算法来学习 2 个不同的主题分布:共同主题和特定主题,并且应用于跨文化分析.

2) 主题和观点的组合分析,主要是通过采用扩展主题模型方法或其他隐变量模型,同时学习主题和对应的意见.例如,文献[97,100-101]提出了一种基于主题的舆论挖掘方法并应用于在线客户评论的产品分析. Qiu 等^[95]通过联合建模 3 个重要因素:视

点特定主题偏好、用户身份和用户交互,提出了一个潜在主题模型,用于论坛帖子的观点挖掘。

这2个研究话题主要聚焦于文本领域中的主题观点挖掘,通过利用一个隐变量模型来建模跨数据集的数据。

近年来,基于概率主题模型的主题观点挖掘成为领域内研究的热点。在文献[96]中,作者提出了一个新颖的相互依赖的主题模型,能够从在线产品评论中学习潜在的主题及其评分信息。Moghaddam等^[101]综合分析了一系列基于LDA变形的主题模型在观点挖掘方面的有效性。在社会事件分析中,也有许多的观点挖掘的研究工作^[95,104]。在文献[95]中,通过联合建模3个重要因素:视点特定主题偏好、用户身份和用户交互,作者提出了一个潜在变量模型用于观点发现。Fang等^[104]提出了一种多模态主题观点挖掘模型,通过联合建模主题和观点来提取文本和视觉2种模态之间语义的相关性以及主题和观点之间的相互依赖关系。Fang等^[96]提出了一种跨视角的主题模型,他们将词汇表中词拆分成不同生成过程的主题术语和观点术语。Gutierrez等^[105]提出了一种多语言主题模型,从多语言非并行数据中同时学习一组共同主题,并自动发现跨语言社区对这些主题的观点差异,用于理解多视角的文化差异。近年来基于社会多媒体事件的舆情分析吸引了很多学者的注意,文献[106]通过研究网络突发事件舆情的发展态势和规律,将舆情演变生命周期划分为孕育、扩散、变换和衰减4个阶段,并分别对每个阶段的议题展开研究。微博是互联网信息传播的重要渠道,是监控公众舆论的重要平台,文献[107]以微博数据为研究对象,开发出一个基于神经网络的微博舆情趋势预测系统。

2 发展趋势

随着移动互联网时代的到来和社交媒体的蓬勃发展,互联网多媒体数据将迎来进一步的爆炸式增长。智能手机和4G网络为用户的生活带来了极大的方便,使用户可以随时随地分享身边发生的事情,用户从信息的获取者转变成了信息的发布者和传播者。大数据背景下的多媒体数据成为互联网信息的主要载体和社会事件分析的主要研究对象。一方面,多媒体大数据集为社会事件分析提供了充足的数据基础,为更加准确地挖掘分析社会事件提供了可能;另一方面,多媒体大数据集是大数据背景下进行社

会事件分析的一个严峻挑战。不仅如此,当今互联网用户成为信息发布和上传数据的主体,使得互联网数据没有统一的结构和规范,大数据集也会造成多媒体数据的多模态、跨平台、数据噪声大、信息冗余等难题。因此,当前多媒体事件分析的一个研究趋势是如何针对大数据难以训练的挑战提出新的方法,或者是改进现有的方法使之适应大规模数据集。总结来说,当前多媒体社会事件分析领域有以下几个发展趋势:

1) 基于深度学习的多媒体社会事件分析研究

随着深度学习技术在学术界和工业界掀起巨大的浪潮,研究者越来越意识到深度学习的优势,如卷积神经网络在图像特征表示方面、循环神经网络对语言的描述能力等。目前,深度学习在图像处理和计算机视觉等领域应用广泛,但是在社会事件分析领域中仍没有系统的方法。所以,将深度学习引入到社会事件分析的各个任务中是我们努力的方向。例如,利用卷积神经网络和词嵌入技术对图像和文本特征进行表示学习,利用递归神经网络考虑事件的时间信息进而研究事件跟踪问题等。因此,将现有深度学习方法应用于社会事件分析任务,并研究有效神经网络结构解决社会事件分析难题是我们进一步要研究的问题。

2) 事件预测相关理论探索。

事件预测是根据历史事件演变过程和当前的多媒体媒体数据进行分析,了解事件的发展脉络和发展速度,从而对事件的发展趋势进行预测,如预测事件未来的热度、关注度等。虽然目前已经有学者开始研究社会事件预测的相关问题,并取得了一定的研究成果,但是大部分研究大多是基于统计分析和基于因果关系的预测方法,没能提出一个统一的有效社会事件预测模型。另外,不同源的社会事件数据存在一定的联系,同时研究多源事件的协同学习,挖掘出相似事件的共性,为社会事件预测提供了一种思路。因此,如何高效地学习多源事件的协同主题,利用不同源数据促进预测分析是下一阶段我们研究的重点问题之一。

3) 基于二元空间的事件关联分析

移动互联网时代,现实世界发生的事件在网络空间中也会有相应的数据呈现。物理空间和网络空间中的事件往往是相互依存、相互补充的。目前大多数事件关联分析研究都是基于网络空间数据,忽略了网络空间和物理空间事件的一致性和关联性。因

此,利用二元空间数据的互补性和共生性,解决社会热点事件的协同关联分析等问题是值得深入研究的课题.另外,二元空间事件关联分析面临一些难题,如多源空间数据异构,物理空间和网络空间数据分别具有多模态性,二元空间数据不同步等问题.如何解决这些难题,综合二元空间的数据进行联合建模实现社会事件关联分析还没有引起研究者足够的重视.

3 需求与展望

在多媒体社会事件分析领域的发展研究中,国内外的研究者提出了很多高效的模型和方法,取得了令人瞩目的成就,逐步形成了4个研究方向:多媒体社会事件表示、多媒体社会事件检测与跟踪、多媒体社会事件动态演化分析和多媒体社会事件主题观点分析,并提出了一系列方法应对多媒体数据的跨平台、多模态、大规模、噪声大和信息冗余等挑战,从而对多媒体数据进行高效、准确的分析,从中发现有价值的信息.

虽然在过去几十年科研工作者在多媒体社会事件分析领域取得了很多骄人的成绩,但是在新时期的发展背景下仍然有很多亟待解决的问题.首先,随着可穿戴设备(如智能眼镜等)和监控设备的普及,越来越多的实际场景的社会媒体数据可被用于多媒体社会事件分析,因此研究实用性更强的多媒体社会事件分析系统需要被提上日程.其次,当今计算机硬件的性能增长速度不能满足多媒体数据的增长速度,在利用大规模多媒体数据的同时,不能忽视大规模和超大规模社会媒体数据的冗余和噪声,设计良好的数据过滤算法是提高数据质量、减轻计算负荷的重点.最后,多媒体社会事件分析的研究不仅仅是局限于单领域的研究,必须结合语言学、社会学、计算机科学等多领域知识,只有如此才能开发出切合实际的社会事件分析系统,为国家政治经济决策、社会舆论监控提供有价值的指导.

本文围绕多媒体社会事件表示、多媒体社会事件检测与跟踪、多媒体社会事件动态演化分析、多媒体社会事件主题观点分析4个方面介绍国内外在多媒体社会事件分析领域的研究进展,总结国内外现有的社会事件分析技术方法,分析国际学科发展趋势以及国内的研究进展和研究特色.近年来国内外学者在多媒体社会事件分析的4个方面进行了广泛的研究并取得了骄人的成果,另外研究者并未停止

对多媒体社会事件分析其他领域的探索,如社会事件预测、社会舆情分析等.尽管如此,在移动互联网、大数据、社交媒体背景下,仍需要在方法创新、应用拓展以及基础理论研究等方面进一步加强,并注重加强学术界到工业界从技术到产品的转换以及交叉学科的互补研究.

参考文献

References

- [1] Salton G, Wong A, Yang C S. A vector space model for automatic indexing [J]. *Communications of the ACM*, 1975, 18(11): 613-620
- [2] Debole F, Sebastiani F. Supervised term weighting for automated text categorization [C] // *ACM Symposium on Applied Computing*, 2003: 784-788
- [3] Jin W, Srihari R K. Graph-based text representation and knowledge discovery [C] // *ACM Symposium on Applied Computing*, 2007: 807-811
- [4] Deerwester S C, Dumais S T, Landauer T K, et al. Indexing by latent semantic analysis [J]. *Journal of the American Society for Information Science*, 1990, 41(6): 391-407
- [5] Hofmann T. Probabilistic latent semantic indexing [C] // *International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999: 50-57
- [6] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3: 993-1022
- [7] Blei D M, McAuliffe J D. Supervised topic models [J]. *Advances in Neural Information Processing Systems*, 2010, 3: 327-332
- [8] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507
- [9] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. *Neural Computation*, 2006, 18(7): 1527-1554
- [10] Hinton G E. Learning distributed representations of concepts [C] // *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, 1986: 1-12
- [11] Song F X, Liu S H, Yang J Y. A comparative study on text representation schemes in text categorization [J]. *Pattern Analysis and Applications*, 2005, 8(1/2): 199-209
- [12] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality [J]. *International Conference on Neural Information Processing Systems*, 2013: 3111-3119
- [13] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data [C] // *International Conference on Neural Information Processing Systems*, 2013: 2787-2795
- [14] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations [C] // *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014: 701-710

- [15] Elman J L. Finding structure in time [J]. *Cognitive Science*, 1990, 14(2): 179-211
- [16] Csurka G, Dance C, Fan L X, et al. Visual categorization with bags of keypoints [C] // *Workshop on Statistical Learning in Computer Vision*, 2004: 1-22
- [17] Behmo R, Marcombes P, Dalalyan A, et al. Towards optimal naive Bayes nearest neighbor [C] // *European Conference on Computer Vision*, 2010: 171-184
- [18] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2008: 1-8
- [19] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C] // *International Conference on Neural Information Processing Systems*, 2012: 1097-1105
- [20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv e-print*, 2014, arXiv: 1409. 1556
- [21] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1-9
- [22] Srivastava R K, Greff K, Schmidhuber J. Highway networks [J]. *arXiv e-print*, 2015, arXiv: 1505. 00387
- [23] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778
- [24] Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical correlation analysis; An overview with application to learning methods [J]. *Neural Computation*, 2004, 16 (12): 2639-2664
- [25] Rasiwasia N, Pereira J C, Coviello E, et al. A new approach to cross-modal multimedia retrieval [C] // *ACM International Conference on Multimedia*, 2010: 251-260
- [26] Guillaumin M, Verbeek J, Schmid C. Multimodal semi-supervised learning for image classification [J]. *Computer Vision and Pattern Recognition*, 2010, 119(5): 902-909
- [27] Theil H, Chung C F. Relations between two sets of variates; The bits of information provided by each variate in each set [J]. *Statistics & Probability Letters*, 1988, 6 (3): 137-139
- [28] Katsurai M, Ogawa T, Haseyama M. A cross-modal approach for extracting semantic relationships between concepts using tagged images [J]. *IEEE Transactions on Multimedia*, 2014, 16(4): 1059-1074
- [29] Zhang H, Zhuang Y T, Wu F. Cross-modal correlation learning for clustering on image-audio dataset [C] // *ACM International Conference on Multimedia*, 2007: 273-276
- [30] Blei D M, Jordan M I. Modeling annotated data [C] // *International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003: 127-134
- [31] Sang J, Xu C S. Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications [C] // *ACM International Conference on Multimedia*, 2012: 19-28
- [32] Putthividhy D, Attias H T, Nagarajan S S. Topic regression multimodal latent dirichlet allocation for image annotation [J]. *Computer Vision and Pattern Recognition*, 2010, 238(6): 3408-3415
- [33] Ramage D, Heymann P, Manning C D, et al. Clustering the tagged web [C] // *ACM International Conference on Web Search and Web Data Mining*, 2009: 54-63
- [34] Qian S S, Zhang T Z, Hong R C, et al. Cross-domain collaborative learning in social multimedia [C] // *ACM International Conference on Multimedia*, 2015: 99-108
- [35] Yu J, Cong Y H, Qin Z C, et al. Cross-modal topic correlations for multimedia retrieval [C] // *International Conference on Pattern Recognition*, 2012: 246-249
- [36] Sang J T, Xu C S, Jain R. Social multimedia mining: From special to general [C] // *IEEE International Symposium on Multimedia*, 2017: 481-485
- [37] Liu X L, Huet B. Heterogeneous features and model selection for event-based media classification [C] // *ACM International Conference on Multimedia Retrieval*, 2013: 151-158
- [38] Das R, Zaheer M, Dyer C. Gaussian LDA for topic models with word embeddings [C] // *Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, 2015: 795-804
- [39] Wan L, Zhu L, Fergus R. A hybrid neural network-latent topic model [C] // *International Conference on Artificial Intelligence and Statistics*, 2012: 1287-1294
- [40] Ngiam J, Khosla A, Kim M, et al. Multimodal deep learning [C] // *International Conference on Machine Learning*, 2011: 689-696
- [41] Srivastava N, Salakhutdinov R. Multimodal learning with deep Boltzmann machines [J]. *Journal of Machine Learning Research*, 2012, 15(8): 1967-2006
- [42] Allan J. Detection as multi-topic tracking [J]. *Information Retrieval Journal*, 2002, 5(2/3): 139-157
- [43] Maron M E. Automatic indexing; An experimental inquiry [J]. *Journal of the ACM*, 1961, 8(3): 404-417
- [44] Lewis D D. Naive (Bayes) at forty: The independence assumption in information retrieval [C] // *European Conference on Machine Learning*, 1998: 4-15
- [45] McCallum A, Nigam K. A comparison of event models for naïve bayes text classification [C] // *AAAI-98 Workshop on Learning for Text Categorization*, 1998: 41-48
- [46] Sebastiani F. Machine learning in automated text categorization [J]. *ACM Computing Surveys*, 2001, 34(1): 1-47
- [47] Altman N S. An introduction to kernel and nearest-neighbor nonparametric regression [J]. *American Statistician*, 1992, 46(3): 175-185
- [48] Joachims T. Text categorization with support vector machines; Learning with many relevant features [J]. *European Conference on Machine Learning*, 1998, 1398: 137-142
- [49] Fred A L N, Leitao J M N. Partitional vs hierarchical clustering using a minimum grammar complexity approach [C] // *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, 2000: 193-202
- [50] Marques de Sa J P. Pattern recognition: Concepts, methods and applications [M]. Berlin: Springer-

- Verlag, 2001
- [51] Zhao Y C, Song J D. GDILC: A grid-based density-isoline clustering algorithm [C] // International Conferences on Info-Tech and Info-Net, 2001; 140-145
- [52] 卜东波. 聚类/分类理论研究及其在文本挖掘中的应用 [D]. 北京: 中国科学院计算技术研究所, 2000
BU Dongbo. Clustering/classification theory and its application in text mining [D]. Beijing: Institute of Computing Technology, Chinese Academy of Sciences, 2000
- [53] Becker H, Naaman M, Gravano L. Learning similarity metrics for event identification in social media [C] // ACM International Conference on Web Search and Data Mining, 2010; 291-300
- [54] Reuter T, Cimiano P. Event-based classification of social media streams [C] // ACM International Conference on Multimedia Retrieval, 2012; 22
- [55] Diakopoulos N, Naaman M, Kivran-Swaine F. Diamonds in the rough: Social media visual analytics for journalistic inquiry [C] // Visual Analytics Science and Technology, 2011; 115-122
- [56] Makkonen J, Ahonen-Myka H, Salmenkivi M. Simple semantics in topic detection and tracking [J]. Information Retrieval Journal, 2004, 7(3/4): 347-368
- [57] Becker H, Naaman M, Gravano L. Event identification in social media [C] // International Workshop on the Web and Databases, 2009; 291-300
- [58] Firan C S, Georgescu M, Nejdil W, et al. Bringing order to your photos: Event-driven classification of flickr images based on social knowledge [C] // ACM International Conference on Information and Knowledge Management, 2010; 189-198
- [59] Zhu J, Chen N, Perkins H, et al. Gibbs max-margin topic models with data augmentation [J]. Journal of Machine Learning Research, 2013, 15(1): 1073-1110
- [60] Qian S S, Zhang T Z, Xu C S, et al. Social event classification via boosted multimodal supervised latent dirichlet allocation [J]. ACM Transactions on Multimedia Computing Communications and Application, 2015, 11(2): 27
- [61] Kumaran G, Allan J. Text classification and named entities for new event detection [C] // International ACM SIGIR Conference on Research and Development in Information Retrieval, 2004; 297-304
- [62] Merler M, Huang B, Xie L X, et al. Semantic model vectors for complex video event recognition [J]. IEEE Transactions on Multimedia, 2012, 14(1): 88-101
- [63] Zhang T Z, Xu C S. Cross-domain multi-event tracking via CO-PMHT [J]. ACM Transactions on Multimedia Computing Communications & Applications, 2014, 10(4): 31
- [64] Wu X, Ngo C W, Hauptmann A G. Multimodal news story clustering with pairwise visual near-duplicate constraint [J]. IEEE Transactions on Multimedia, 2008, 10(2): 188-199
- [65] Kalamaras I, Drosou A, Tzovaras D. Multi-objective optimization for multimodal visualization [J]. IEEE Transactions on Multimedia, 2014, 16(5): 1460-1472
- [66] Yang Y M, Zhang J, Carbonell J, et al. Topic-conditioned novelty detection [C] // Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002; 688-693
- [67] Allan J, Wade C, Bolivar A. Retrieval and novelty detection at the sentence level [C] // International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003; 314-321
- [68] Wang C, Blei D, Heckerman D. Continuous time dynamic topic models [J]. arXiv e-print, 2012, arXiv: 1206.3298
- [69] Roy S D, Mei T, Zeng W J. Bridging human-centered social media content across web domains [C] // Human-Centered Social Media Analytics, 2014; 3-19
- [70] Kender J R, Naphade M R. Visual concepts for news story tracking: Analyzing and exploiting the NIST TRECVID video annotation experiment [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005; 1174-1181
- [71] Zhai Y, Shah M. Tracking news stories across different sources [C] // ACM International Conference on Multimedia, 2005; 2-10
- [72] Griffiths T L, Steyvers M. Finding scientific topics [J]. Proceedings of the National Academy of Sciences, 2004, 101(sup1): 5228-5235
- [73] Pan C C, Mitra P. Event detection with spatial latent dirichlet allocation [C] // Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries, 2011; 349-358
- [74] Iwata T, Yamada T, Sakurai Y, et al. Online multiscale dynamic topic models [C] // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010; 663-672
- [75] Iwata T, Watanabe S, Yamada T, et al. Topic tracking model for analyzing consumer purchase behavior [C] // International Joint Conference on Artificial Intelligence, 2009; 1427-1432
- [76] Wu Q H, Lv J H, Ma S L, et al. EET: Efficient event tracking over emergency-oriented web data [C] // International Joint Conference on Neural Networks, 2015; 1-8
- [77] Qian S S, Zhang T Z, Xu C S, et al. Multi-modal event topic model for social event analysis [J]. IEEE Transactions on Multimedia, 2016, 18(2): 233-246
- [78] Kasiviswanathan S P, Melville P, Banerjee A, et al. Emerging topic detection using dictionary learning [C] // ACM International Conference on Information and Knowledge Management, 2011; 745-754
- [79] Wang X R, Mohanty N, McCallum A. Group and topic discovery from relations and text [C] // Proceedings of the 3rd International Workshop on Link Discovery, 2005; 28-35
- [80] Gao Z J, Song Y Q, Liu S X, et al. Tracking and connecting topics via incremental hierarchical dirichlet processes [C] // IEEE International Conference on Data Mining, 2011; 1056-1061
- [81] Chi Y, Song X D, Zhou D Y, et al. Evolutionary spectral clustering by incorporating temporal smoothness [C] //

- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2007:153-162
- [82] Ahmed A, Xing E. Dynamic non-parametric mixture models and the recurrent Chinese restaurant process; With applications to evolutionary clustering [C] // SIAM International Conference on Data Mining,2008:219-230
- [83] Caron F, Davy M, Doucet A. Generalized Polya urn for time-varying dirichlet process mixtures [J]. arXiv e-print, 2012, arXiv: 1206.5254
- [84] Ren L, Dunson D B, Carin L. The dynamic hierarchical dirichlet process [C] // International Conference on Machine Learning,2008:824-831
- [85] Zhang J W, Song Y Q, Zhang C S, et al. Evolutionary hierarchical dirichlet processes for multiple correlated time-varying corpora [C] // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2010: 1079-1088
- [86] Wang X H, Zhai C X, Hu X, et al. Mining correlated bursty topic patterns from coordinated text streams [C] // ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2007:784-793
- [87] Chen J F, Yu J J, Shen Y. Towards topic trend prediction on a topic evolution model with social connection [C] // IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 2013: 153-157
- [88] Gong Y H, Liu X. Generic text summarization using relevance measure and latent semantic analysis [C] // International ACM SIGIR Conference on Research and Development in Information Retrieval,2001:19-25
- [89] Haghighi A, Vanderwende L. Exploring content models for multidocument summarization [C] // Human Language Technologies: The 2009 Conference of the North American Chapter of the Association for Computational Linguistics,2009:362-370
- [90] Zhou W B, Shen C, Li T, et al. Generating textual storyline to improve situation awareness in disaster management [C] // IEEE International Conference on Information Reuse and Integration,2015:585-592
- [91] Wang D D, Li T, Zhu S G, et al. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization [C] // International ACM SIGIR Conference on Research and Development in Information Retrieval,2008:307-314
- [92] 李风环. 主题事件挖掘及动态演化分析研究 [D]. 哈尔滨: 哈尔滨工业大学计算机科学与技术学院, 2015
LI Fenghuan. Research on topic event mining and dynamic evolution analysis [D]. Harbin: School of Computer Science and Technology, Harbin Institute of Technology, 2015
- [93] Moghaddam S, Ester M. ILDA: Interdependent LDA model for learning latent aspects and their ratings from online product reviews [C] // International ACM SIGIR Conference on Research and Development in Information Retrieval,2011:665-674
- [94] Yang X S, Zhang T Z, Xu C S. Cross-domain feature learning in multimedia [J]. IEEE Transactions on Multimedia,2015,17(1):64-78
- [95] Qiu M H, Jiang J. A latent variable model for viewpoint discovery from threaded forum posts [C] // Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics, 2013: 1031-1040
- [96] Fang Y, Si L, Somasundaram N, et al. Mining contrastive opinions on political texts using cross-perspective topic model [C] // International Conference on Web Search and Web Data Mining,2012:63-72
- [97] Liu B, Zhang L. A survey of opinion mining and sentiment analysis [M] // Aggarwal C C, Zhai C X. Mining text data. Springer,2012:415-463
- [98] Wang H, Zhang C, Yin H Z, et al. A unified framework for fine-grained opinion mining from online reviews [C] // Hawaii International Conference on System Sciences, 2016:1134-1143
- [99] Li P J, Yan Y H, Wang C M, et al. Customer voice sensor: A comprehensive opinion mining system for call center conversation [C] // IEEE International Conference on Cloud Computing and Big Data Analysis, 2016: 324-329
- [100] Qian S S, Zhang T Z, Xu C S. Multi-modal multi-view topic-opinion mining for social event analysis [C] // ACM on Multimedia Conference,2016:2-11
- [101] Moghaddam S, Ester M. On the design of LDA models for aspect-based opinion mining [C] // ACM International Conference on Information and Knowledge Management, 2012:803-812
- [102] Keller K L. Conceptualizing, measuring, and managing customer based brand equity [J]. Journal of Marketing, 1993,57(1):1-22
- [103] Paul M, Girju R. Cross-cultural analysis of blogs and forums with mixed-collection topic models [C] // Conference on Empirical Methods in Natural Language Processing,2009:1408-1417
- [104] Fang Q, Xu C S, Sang J T, et al. Word-of-mouth understanding: Entity-centric multimodal aspect-opinion mining in social media [J]. IEEE Transactions on Multimedia, 2015,17(12):2281-2296
- [105] Gutierrez E D, Shutova E, Lichtenstein P, et al. Detecting cross-cultural differences using a multilingual topic model [J]. Transactions of the Association for Computational Linguistics, 2016,4:47-60
- [106] 方付建. 突发事件网络舆情演变研究 [D]. 武汉: 华中科技大学公共管理学院, 2011
FANG Fujian. Research on the evolution of online public opinion in unexpected events [D]. Wuhan: College of Public Administration, Huazhong University of Science and Technology, 2011
- [107] Chen N Y, Liu Y, Zhang Z J. A forecasting system of micro-blog public opinion based on artificial neural network [J]. Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2014, 16(6):999-1004

Review on multimedia social event analysis research

QIAN Shengsheng¹ ZHANG Tianzhu¹ XU Changsheng¹

¹ National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

Abstract In recent years, with the rapid development of Internet, more and more social networking sites appear and allow users to conveniently share their ideas, pictures, posts, and activities. Therefore, when a popular event is happening around us, it can spread very fast in different social media sites with substantial amounts of multimedia data including images, videos, and texts. Therefore, it is important and necessary to conduct the research of multimedia social event analysis to know the evolutionary trend of social event over time automatically. This paper provides a survey and summarizes major progresses in multimedia social event analysis. We focus on four areas: (1) multimedia social event representation; (2) multimedia social event detection and tracking; (3) multimedia social event evolutionary analysis; and (4) multimedia social event topic-opinion analysis. Then, the development trend of multimedia social event analysis is highlighted. Finally, the possible future research topics in multimedia social event analysis are prospected.

Key words multimedia; social event; multi-modality; multi-domain