

DOI: 10.11992/tis.201612007  
网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20170111.1619.002.html>

# 在线社交网络挖掘与搜索技术研究

石磊<sup>1</sup>, 杜军平<sup>1</sup>, 周亦鹏<sup>2</sup>, 叶杭<sup>1</sup>, 赖金财<sup>1</sup>, 何奕江<sup>1</sup>

(1. 北京邮电大学 智能通信软件与多媒体北京市重点实验室, 北京 100876; 2. 北京工商大学 计算机与信息工程学院, 北京 100048)

**摘 要:** 随着在线社交网络的蓬勃发展, 传统的数据挖掘的和搜索方法已经不能完全适用于 Web 2.0 时代的社交网络。社交网络具有社交关系复杂、数据量大、动态更新、数据多模态等特点, 给数据挖掘和搜索的研究来了巨大的挑战。因此, 研究基于社交网络挖掘和搜索的新方法成为学术界和工业界的一项新任务。文章全面分析了社交网络发展的基本情况和存在的问题, 阐述了社交网络结构建模、信息传播机制、社区发现、情感分析、事件监测及社交网络搜索排序技术的主要研究工作, 并基于已有研究工作对社交网络挖掘和网络搜索技术进行了分析和展望。

**关键词:** 社交网络; 数据挖掘; 搜索; 社区发现; 信息传播

**中图分类号:** TP393 **文献标志码:** A **文章编号:** 1673-4785(2016)06-0777-11

中文引用格式: 石磊, 杜军平, 周亦鹏, 等. 在线社交网络挖掘与搜索技术研究[J]. 智能系统学报, 2016, 11(6): 777-787.

英文引用格式: SHI Lei, DU Junping, ZHOU Yipeng, et al. A survey on online social network mining and search[J]. CAAI Transactions on Intelligent Systems, 2016, 11(6): 777-787.

## A survey on online social network mining and search

SHI Lei<sup>1</sup>, DU Junping<sup>1</sup>, ZHOU Yipeng<sup>2</sup>, YE Hang<sup>1</sup>, LAI Jincai<sup>1</sup>, HE Yijiang<sup>1</sup>

(1. Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China; 2. School of Computer Science and Information Engineering, Beijing Technology and Business University, Beijing 100048, China)

**Abstract:** With the vigorous development of online social networks, the traditional technologies of data mining and searching cannot solve the problems of social networks in the Web 2.0 era. Social networks, accompanied by complex social relationships, large amounts of data, dynamic updates, multimodal data, etc. have brought great challenge to the study of data mining and searching. Therefore, the research of novel algorithms of social network mining and searching has become a new task in both academia and industry. This paper summarized the basic situation and problems of social networks, and analyzed structural modeling techniques, information transmission mechanisms, community detection, sentiment analysis, event detection and search ranking techniques of social networks. Based on the analysis of previous researches, the prospect of social network data mining and search technologies was forecasted in this paper.

**Keywords:** social networks; data mining; search; community detection; information transmission

在线社交网络也称社交网络服务 (SNS)<sup>[1]</sup>,

SNS 是由网络上每个独立存在的个体以及个体之间的相互关系所构成的一个社会化媒体网络。随着这种新型网络的出现, 把以前网络仅仅是用户消耗和获取信息, 变成了一个人人参与、人人可以产生信息, 而且用户之间可以进行交流和互动的网络。目

收稿日期: 2016-12-06.  
基金项目: 国家自然科学基金重点项目(61532006); 国家自然科学基金重大国际合作项目(61320106006).  
通信作者: 杜军平. E-mail: junpingdu@126.com.

前存在的社交网络形式可以分为 5 类<sup>[2]</sup>:即时通信类,如腾讯 QQ、MSN、微信等;在线社交平台类,如 FaceBook、LinkedIn、人人网等;社交微博平台类,如 Twitter、新浪微博、腾讯微博等;标签协同平台类,如 Delicious、Flickr 等;科研分享与社交平台,如 ResearchGate、学术圈等。用户通过这些平台可以在线交流、交换信息、分享资源、新闻转发和评论等。另外,随着社交网络的出现,传统的新闻门户网站或各大传统媒体的官方网站也都提供了基于社交网络的分享和转发功能,方便用户快速分享新闻或电子期刊到相关社交网络平台。

随着在线社交网络的蓬勃发展,数据也在快速增长,FaceBook 有超过 16 亿的活跃用户,而中国的新浪微博也有超过 5 亿的活跃用户。社交网络正日益影响着用户的生活习惯,成为用户生活中的一部分。社交网络相比传统网络体现出更加复杂的综合特征,使得传统的挖掘理论和模型难以描述社交网络中的用户行为方式。只有通过有效的挖掘社交网络中的文本、图片等多种信息,分析数据中隐含的特有属性、研究用户的行为特征,才可以弥补传统挖掘和搜索方法在社交网络中的不足,实现满足用户个性化需求的智慧化的搜索。当前在线社交网络挖掘中最具有代表性的研究方向包括社交网络结构建模、信息传播、社区发现、情感分析、事件监测等。

在海量数据中找到自己想要的或者感兴趣的内容和信息,通常被称之为信息搜索,它是指信息按一定的方式组织起来,并根据信息用户的需要找出有关的信息的过程和技术<sup>[3]</sup>。信息搜索主要使用传统的搜索引擎来进行,采用的方法包括基于 Web 1.0 静态网页,通过 BM25 等算法计算内容的相似性。在排序方面主要采用了传统的谷歌 PageRank 和 HITS 算法。但这种传统的搜索算法对 Web 2.0 的应用不能很好地支持,同时没有考虑社交网络的 5V 性<sup>[4]</sup>。社交网络搜索区别于一般搜索的一个重要特点是除了对内容的搜索之外,还可以提供对特定对象如个人、群体、社区的搜索,这就需要挖掘社交网络中社交关系、社区、事件传播、情感分析等内容来弥补社交网络内容搜索的缺陷,因此如何围绕用户和网页作为中心,实现如搜索用户、社会关系、社区发现、事件来源等社会化层面的搜索,并通过理解用户的意图等实现智慧化搜索是社交网络搜索研究的关键。

本文主要从社交网络结构建模、信息传播、社区

发现、情感分析、事件监测、搜索索引及排序等方面对目前社交网络挖掘和搜索相关研究的现状进行了总结和论述,并对社交网络挖掘和搜索的发展趋势进行了展望。

## 1 在线社交网络挖掘的关键技术

社交网络挖掘作为最近几年热门的研究和应用,涉及多种理论和技术,包括了数理统计、数据挖掘技术、矩阵论等。

数据挖掘<sup>[5]</sup>一般是指从大量的数据中通过算法搜索隐藏于其中信息的过程。数据挖掘通常与计算机科学有关,并通过统计、在线分析处理、情报检索、机器学习、专家系统和模式识别等诸多方法来实现。而社交网络分析与挖掘则是通过数据挖掘的方法从社交网络数据中提取信息的过程<sup>[6]</sup>,目前在社交网络分析与挖掘方向主要的研究有社交网络结构建模、信息传播研究、社区发现、情感分析及事件监测等。

### 1.1 社交网络结构建模

社交网络结构建模是社交网络研究的基础,采用结构化方法和数学方法来研究社交网络的内部各种特征和机制。基于社交网络研究的第一步是对其结构进行建模<sup>[7]</sup>,目的是对其内部结构和演化规律进行深化。一般对社交网络结构建模通过统计学习方法来构造,然后分析社交网络的分布规律、关系紧密程度、相识关系的紧密程度,某个用户对于网络中其他用户对之间传递消息的重要程度等诸多统计特性。

社交网络建模由最初的 ER 随机图模型到小世界、无标度模型、六度分割等。Moreno 等首先将图论的方法引入了人类社交关系分析中,随着技术的发展,研究人员发现 ER 随机图已经不能解决重尾分布等问题,因此能够解决这些问题的参数特征小世界网络被提出,这是一个基于统计的模型。Handcock 等提出的隐含位置聚类模型,Watts 等分析和验证了“六度分隔”和小世界模型<sup>[8]</sup>。Kumar 等研究了在线社区的路径长度问题,其长度要大于“六度分隔”。Harary 等提出一种有向图模型来表示社交网络中的单向关系<sup>[8]</sup>。HSU 等<sup>[9]</sup>通过应用可变的向量时钟和权重变化,构建了一个权重耦合的定向链路生成算法,对社交网络群体结构建模。Dong 等<sup>[10]</sup>通过图模型研究了社交网络中个人和相关社会现象之间的关系,实现人口统计推断、链接推荐、社会影响预测等应用。Slaughter 等<sup>[11]</sup>提出基于贝叶斯多层次模型的社交网络结构模型,拓展

了现有的随机图结构,同时允许组内连接结构之间的信息传输。Amato 等<sup>[12]</sup>在超图模型基础上提出了基于 Flickr 社交多模态信息的超图模型结构模型,利用超图建立多媒体内容之间、用户和多媒体内容之间、用户和用户之间的关系,该模型的引入打破了传统方法只能引入简单的社交关系的限制,为深入研究社交网络提供了方法。

基于非图结构的研究方面,Bajaj 等<sup>[13]</sup>提出了一种基于 Agent 的模型,该模型把社交网络结构视为一个外生变量,是一种时间感知的关系模型。Mihoub 等<sup>[14]</sup>模拟一个主体参与的面对面社交行为,提出了基于贝叶斯网络多式联运的行为模型,并通过实验验证了其性能优于马尔可夫模型(HMM)和隐马尔可夫模型(HSMMs)。

由于在线社交网络结构具有社交关系复杂、内容多样、结构动态变化、数据多维等特点,所以对其进行深入研究还需要更多有效的建模和计算方法,而不是仅局限于图的建模方法。

## 1.2 社交网络信息传播

社交网络的分享和转发功能使用户可以随时转发和分享社交信息,其信息数据的传播和扩散的范围迅速扩大,通过对社交网络中信息的传播机制进行研究可以实现社交网络的舆情与态势分析、谣言传播溯源、事件追踪、精准搜索等研究。社交网络中的信息传播分析涉及到社会学、传播学、统计学、计算机科学等,一般的研究只是针对几个热门的研究点进行,这里主要介绍目前学术界聚焦的信息传播建模和社交影响力分析这两个方向。

信息传播主要是通过建立信息传播模型和社交影响力分析等研究社交网络中信息传播的机制。目前的研究可以归纳为两个模型:解释模型和预测模型。解释模型的目的是推导出潜在的级联传递,得出完整的激活序列。Gomez 等<sup>[15]</sup>提出了基于次模函数优化的 NETINF 迭代算法,利用节点间感染次数的联系来推导出级联传递的结构,并假设被激活的节点以特定的概率传递给相邻节点。Jones 等<sup>[16]</sup>扩展了 NETINF 算法,通过解决最大似然问题来推断出两两节点的传输速率和传递图像。Gomez 等<sup>[17]</sup>继续扩展了 NETRATE 算法,提出了基于时间变化的推理算法 INFOPATH,采用随机梯度提供变化的网络中的在线结构和时序变化分析。Sadikov 等<sup>[18]</sup>基于 K 树模型提出了一种可以通过少量的完整被激活序列估计级联传递中的属性的方法。

预测模型的目的是观察并学习过去的传递过程

中时间和空间的变化,来预测社交网络中传递过程的展开方式,分为基于图和基于非图两类模型。

1) 基于图模型。Romero 等<sup>[19]</sup>提出了一个线性阈值模型来预测传递过程的方法,该方法依赖于信息病毒式传播及成对用户之间的影响力和接受信息概率参数等。Saito 等<sup>[20]</sup>提出了 ASIC 和 ASLT 算法,通过提出异步扩展,放宽传统的基于图形的 IC 和 LT 模型的同步性假设。Guille 等<sup>[21]</sup>将传播过程建模为异步独立级联过程,提出了 T-BaSIC 模型。Xu 等<sup>[22]</sup>提出随时间变化的信息扩散模型。该模型主要探讨了随时间变化的信息扩散模型和标准模型的关系。

2) 基于非图模型。Wen 等<sup>[23]</sup>提出了一个随机节点水平传播分析模型,它可以动态地获取信息传播时间和捕获人们的行为差异,并可以分析社会参数对信息传播的影响。Tuarob 等<sup>[24]</sup>提出了一个基于四种网络信号的 OIA-SIRS 模型,用于观察网络节点信息在不同节点中的传递情况。Tambuscio 等<sup>[25]</sup>提出了一个改进的 SIS 模型用于预测错误信息在社交网络中的传递规模。模型依赖于 4 种参数,分别为传递速率、轻信程度、验证可能性和遗忘速度。在同构网、异构网络和真实社交网络的实验中,显示出错误信息检查的可能性的阈值,定量地衡量出根除骗局所必需的最小反应。Wang 等<sup>[26]</sup>通过对新浪微博的信息传播过程进行分析,发现微博社区的信息传播类似于动态模型,故通过数据密集型计算理论对动态模型的算法进行改进,同时对新浪微博数据的各种特征进行了挖掘和建模,并提出了 Seir 模型,取得了较好的效果。

在影响力分析方面。主要有基于网络结构的影响力排序算法、结合网络结构和文本内容的影响力排序算法、异构信息中的影响力算法等。经典的影响力分析算法主要有 PageRank 算法以及相关的改进算法,如 SimRank 和 TwitterRank 等。Rometro 等<sup>[19]</sup>提出了一种基于图的 IP 算法,与 HITS 算法类似,为每一个用户转发信息时分配了一个相对影响力和被动分数;Pal 等<sup>[27]</sup>提出了一种非图的话题敏感模型,通过对节点的集群排名发现最具影响力和权威性的人;Suo 等<sup>[28]</sup>提出了基于超网及其拓扑结构的模型用于分析社交网络中的用户影响力;吴岷辉等<sup>[29]</sup>提出了一种基于用户行为网络的改进的 PageRank 算法,称之为 TopicLeaderRank 算法,将用户属性、网络拓扑及交互信息等特征综合考虑;Suppa 等<sup>[30]</sup>考虑到社交网络数据和图像的快速增长造



成的计算复杂度较大,提出了一种改进的 Brande 算法快速评估大型网络中结点距离,为了便于选择主结点,采用 Louvian 算法进行聚类;Yang 等<sup>[31]</sup>通过用户社会角色之间的相互联系及在信息传递时的影响力调查,提出了一个角色导向信息扩散模型,将社会角色识别和扩散模型集成到一个统一的框架,开发出基于吉布斯抽样的算法应用于该模型;Subbian 等<sup>[32]</sup>提出了一个在社会流量中利用主题和时间敏感性的方法计算用户的实时影响力。

上述社交网络的信息传播机制的研究主要集中在经典的传染病模型的利用和扩展上,没考虑从社交网络作为一个个体自身去考虑,比如考虑传播过程中用户的心理因素、用户扮演的角色等。在影响力分析方面通常只考虑了社交网络用户的全局影响力或者局部影响力,而没有根据实际情况综合去考虑,忽视了社交网络的尺度多样性特征。

### 1.3 社区发现

社交网络的核心是参与其中的用户以及用户之间的关系。因此,学术界通常采用图模型对其进行建模,其中节点表示参与社交网络中的用户,而边则表示社交网络中用户间的关系,同时利用每条边的权重表示用户之间关系强度或亲密程度,权值越大表示关系强度或者亲密程度越大,那些内部连接比较紧密的节点子集合对应的子图叫做社区,各社区节点集合彼此没有交集的称为非重叠型社区,有交集的称为重叠型社区。网络图中包含一个社区的现象称为社区结构。给定一个网络图,找出其社区结构的过程叫做社区发现。一个典型的社区如图 1 所示,图中各个点表示成千上万的用户,边表示用户之间的关系,每个点聚集的区域表示一个社区,同时由于每个人可能有多种爱好,不同的社区可以发生重叠。

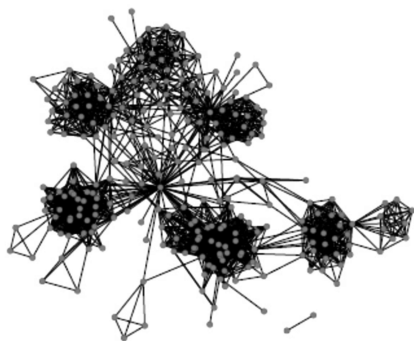


图 1 典型的社区图

Fig.1 Typical community graph

社区发现技术可以发现社交网络中相关的拓扑结构以及兴趣爱好,通常采用不同的数据挖掘算法来研究,目前的研究方法通常集中在图论的相关

算法,例如图分割、图聚类、图的修剪等方法。

图分割的相关工作认为社区的成因是因为网络连边之间存在“强弱连边关系”,主要有两点:“三元闭包”关系下演化出来的、节点之间相互博弈生成的。因此在图分割视角下的科学问题就是“如何识别强弱连边关系”。典型的图分割方法是模块化的方法和 CPM 方法。Su 等<sup>[33]</sup>提出了一个模糊模块最大化(FMM)的社区发现方法,该方法利用了广义 NEWMAN 模块的最大化方法来发现社区,然后采用树的结构局部最优化发现的社区,实验结果表明通过该方法可以高效地发现重叠社区。Kloster 等<sup>[34]</sup>用一个基于热核的算法来标识社区的起始节点,提出了一个确定性的局部算法计算社区的产生,并通过过度加权范数的矩阵指数模型来估计社区。AL-TUNBEY 等<sup>[35]</sup>提出了元启发式模块优化算法,该算法通过优化网络模块化的适应度函数来发现重叠社区,取得了较好的效果。ARAB 等<sup>[36]</sup>提出自下而上的社区检测方法,采用模块化和 NMI 的混合方法从细粒度的社区开始,逐步发现真实社区。Chen 等<sup>[37]</sup>在基于图分割策略的基础上提出了一个局部菲德勒向量中心算法(LFVC)来发现深度社区。模块法方法虽然有不错的效果,但其缺点也很明显,就是存在识别极限的问题。

当前图聚类的方法主要是基于谱聚类思想。它关注的科学问题是“节点的空间映射问题”。谱聚类算法是解决网络生成模型的有效的方法,这种概率生成模型的理论基础也使得其具有广泛的普适效应,成为现今社区发现算法的主要研究方向。Zhang 等<sup>[38]</sup>的研究致力于通过谱聚类算法解决重叠社区发现的问题。Gao 等<sup>[39]</sup>通过图聚类的发放解决了复杂网络的适用性问题。Mahmood 等<sup>[40]</sup>通过线性编码来提高算法运行速率的问题,取得良好的效果。然而,上述的方法都是从建模网络连边密度入手的,没有实际建模网络连边的生成过程。而且上述方法认为每个节点仅仅属于一个社区,忽略了社区中存在的重叠现象,因此,节点表达的思路认为每个节点都是  $K$  个社区的分配的表达。这里的科学问题就是“如何通过观测网络学习得到这种节点的隐式表达”。AIROLDI 等<sup>[41]</sup>提出了混合隶属度随机块模型,这种基于概率统计方法的生成式模型更好地解释了节点之间的边是如何生成的以及整个网络是如何生成的,并通过机器学习方法来学习隐变量得到网络的重叠划分。这种方法对网络的解释性更好,唯一缺点就是优化速度慢,可能会优化到局部最优。

1.4 情感分析

情感计算是 1997 年由 MIT 的 Picard 教授提出的,情感计算是与情感相关,来源于情感或能够对情感施加影响的计算,而随着社交网络的发展,基于社交网络的情感分析再次成为学术界的研究热点,通过情感分析可以对社交网络搜索提供更加精确的理解,提高搜索准确度。

目前学术界基于情感分析的研究方法主要集中在社交网络文本的情感词方法。该方法主要是通过人工整理出程度副词表、否定词表和社交网络中默认表情符号的褒贬分类,然后在情感词语计算的基础上,考虑上下文中否定词和程度词对修饰情感词语的情感倾向和情感强度的影响,同时也设定规则计算表情符号对一条微博的情感倾向判断的作用<sup>[42]</sup>。Marquez 等<sup>[43]</sup>使用相关的情感词典从不同的情绪特征维度出发来提升微博情感分类的精度;Hu 等<sup>[44]</sup>通过加入社会学的方法来提高情感分析的准确性,该方法结合了情绪感染理论到监督学习的过程,并利用稀疏学习实现了微博文本中的去噪;Debashis 等<sup>[45]</sup>通过用户社交网络中的对话,确定用户的情绪,通过一个新的词汇字典和情感的罗素模型识别情感,并利用隐含狄利克雷分配(LDA)生成模型建立主题和情感分布;Sixto 等<sup>[46]</sup>利用 BM25 排序函数与支持向量机相结合的监督学习方法来对 Twitter 进行情感分析,具有较好的效果。

随着社交网络和移动社交 APP 的发展,社交网络中其他媒体如图片、视频、音乐等数据急聚增多,对社交网络中图像、视频的情感分析等相关技术的研究也成为一个重要方向。深度学习的兴起对跨媒体数据分析具有重要意义,You 等<sup>[47]</sup>利用微调的深度卷积神经网络构架训练图片情感分析模型,相比传统的方法具有较好的效果;Chao 等<sup>[48]</sup>利用长期记忆神经循环网络构架和时间池技术对音频和视频情感分析;Poria 等<sup>[49]</sup>针对跨媒体的情感分析进行研究并取得了一定的进展,其主要利用了各自领域的情感分析方法,然后通过特征级和决策级的特征融合来训练情感分析模型,其结果较之单一模态的情感分析方法精度更高。

目前的情感分析方法大多是通过简单的使用一些情感词等基于文本的方法,而忽略了用户用来增强情感的图像及视频等内容,这样很难符合真实社交网络中用户复杂的情感表示,将为社交网络情感分析带来新的挑战。

1.5 社交网络事件监测

社交网络事件监测的目标是对社交内容中的事件和热点话题的自动识别和已知话题的持续跟踪。事件监测的基础方法为计算文档之间的相似性。具体方法是预先设置关键词或者突发词,然后计算词与词之间相似度来监测事件,文档之间相似性常用度量方法为夹角余弦,如式(1)所示:

$$\text{sim}(D_1, D_2) = \cos \theta = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

式中:  $\text{sim}(D_1, D_2)$  表示相似度函数,  $D_1$  和  $D_2$  表示文档内容,而  $A_i$  和  $B_i$  表示两个  $n$  维向量。

该方法仅适用于静态文本语料库分析的传统话题监测技术,而社交网络中不但存在大量的静态文本,同时也存在跨媒体内容,这就涉及语义分析相关内容,因此这种简单计算文本相似性的方法无法直接适用于在线社交网络产生的跨媒体的海量数据中。Kaleel 等<sup>[50]</sup>利用词频逆序文档频率和局部敏感哈希的方法实现热点事件发现,并通过聚类的方法提高事件监测的效率。Andrea 等<sup>[51]</sup>提出了一个基于微博交通事件实时监测系统,根据微博标签和预设的搜索条件,利用支持向量机算法对事件进行分类,最后实现事件监测。Li 等<sup>[52]</sup>提出了基于 Spark 的分布式微博突发事件监测增量时间主题模型,该模型能够利用短文本数据集和时间信息监测突发事件,这种分布式的设计大大提高了监测效率。Zhang 等<sup>[53]</sup>提出了突发事件监测和趋势预测的方法。该方法利用词频和用户的社交关系等信息进行事件监测,并提出了一个扩散模型来预测事件的流行趋势。该方法解决了大多数现有的方法只专注于事件监测,但忽略了预测未来趋势的问题。Zhou 等<sup>[54]</sup>基于图的模型提出了一个监测社会事件的框架 LTT,该框架可以捕捉内容、时间、地点和社交信息,具有良好的适用性。Pohl 等<sup>[55]</sup>提出了社交网络事件自动监测方法,可以高效地实现对 Flickr 和 YouTube 的社交事件和子事件进行监测。上述方法在事件监测方面都取得了良好的效果,但是上述方法侧重于社交网络的文本内容的事件监测,而忽略了社交相关内容。Guille 等<sup>[56]</sup>提出了异常事件监测方法,该方法主要利用动态链接的创作频率。用户动态地在微博上插入需要监测重要事件,并估计对人群的影响程度。Zhang 等<sup>[57]</sup>提出了基于突发词权重的时间窗口内提取突发词方法,然后结合层次聚



类算法,利用突发词监测突发性事件。Yang 等<sup>[58]</sup>利用训练好的社交视频利用基于字典学习的方式实现社交网络事件监测。Abdelhaq 等<sup>[59]</sup>基于 Twitter 提出了一种局部实时社交事件检测方法,首先提取时空关键词,根据时空相似性对关键词聚类,最后对关键词进行评分来监测事件,取得了良好的效果。

随着事件监测研究的不断进步,基于多模态的社交网络事件监测方法成为研究热点。Manos 等<sup>[60]</sup>提出了一种基于时间窗的多模态方法对海量图片进行事件监测。Gao 等<sup>[61]</sup>利用超图分割的思想并引入文本内容、视觉内容、位置信息、时间信息等实现社交多媒体的事件监测,然后利用新浪微博数据集进行实验,取得了不错的效果。该方法的优点是引入了多模态信息和时空特性,为基于时空特性的事件监测提供了解决方案,但是该方法并没给出预测热点事件的方法,针对这个问题 Unankard 等<sup>[62]</sup>提出了基于位置感知的社交网络事件监测方法,把 Twitter 事件和位置信息进行统一表示,将用户位置和事件位置的相关性的强度进行标识来监测新兴事件和预测热点事件。

## 2 社交网络搜索关键技术研究

目前搜索技术的研究热主要集中在传统的搜索引擎技术和社交网络搜索技术。传统的搜索引擎一般采用经典的倒排索引,然后对索引结果根据相关算法进行重排序,典型的算法有 PageRank、HITS 等<sup>[63]</sup>。随着社交网络的发展,搜索引擎提供商也开发了基于社交网络的搜索功能,而且社交网络服务公司也提供了基于本平台的社交搜索引擎。国内的搜狗公司提供了基于微信、知乎等在线社交网络的搜索服务,以提供更加适应于社交网络的搜索技术。特别地,在社交平台搜索方面,FaceBook 开发了备受欢迎的知识图谱搜索技术<sup>[64]</sup>,通过该技术可以对用户发布的图片、内容、地点等进行社会化的搜索和推荐,相比传统的搜索引擎, Twitter 提供的搜索完全是基于社交网络的搜索,提供了基于内容的搜索、社交关系搜索、事件的情感搜索、用户等搜索功能、热点事件搜索等。国内的新浪微博也提供基于社交网络搜索功能,与 Twitter 相比新浪微博的搜索比较简单,仅提供了相关话题推荐、热点事件推荐、用户搜索、内容搜索等功能。

### 2.1 搜索索引研究

社交网络索引的设计对社交网络搜索结果的准确度和搜索效率有着至关重要的影响,不同于传统

的搜索索引,社交网络有字数限制(一般要求不能超过 140 字),所以社交网络的内容相对比较稀疏,噪声数据较多,传统的方法已经不适用于社交网络。目前针对索引的设计主要集中在对倒排索引的改进<sup>[65]</sup>,使之更加适应社交网络。

Chen 等<sup>[66]</sup>针对社交网络搜索以键值存储形式的不可扩展性,设计了摘要引的算法,通过该算法可以避免对邻居结点的穷举查询,提高索引效率,同时提出了 piggyback 策略来消除冗余信息,提高搜索的速度。Chen 等<sup>[67]</sup>提出了社交网络索引的自适应方案,在排序算法基础上设计了基于时间戳的用户和内容联合关系的排序方法,该方法首先对社交网络内容和主题进行分类,然后对其动态排序。该方法很好地解决了传统索引方案实时性不强的问题。Li 等<sup>[68]</sup>提出了一种 3-D 倒排索引技术,引入了社交属性和时间属性用于实现支持词频、社交距离和时间新鲜度的 3 个维度的搜索技术,并应用了 Dijkstra 算法计算社交网络距离来优化算法的冷启动问题。

Zhao 等<sup>[69]</sup>提出了自适应融合策略 (ls-ams) 来处理用户对社交网络实时查询的请求,该策略在倒排索引基础上增加指标包来管理倒排索引,通过这种机制可以减少合并查询的开销,提高查询性能。Huang 等<sup>[70]</sup>提出了一个实时的分布式索引系统,该系统用时间分割索引,在每一个分区中映射一个哈希表,通过这种机制可减少特定范围内查询的时间; Yuan 等<sup>[71]</sup>提出了基于部分索引机制的分类和索引方法以提高检索效率,分类算法将查询分为候选查询和流行查询,并采取单独的存储策略,实验结果表明该分类和索引算法可以减少实时搜索时间,提高搜索效率; Rissola 等<sup>[72]</sup>提出了索引项目失效策略,该策略引入了索引干的概念,索引干负责跟踪基本词汇的演变和删除无效的索引条目,通过该方法大大地提高了索引效率。

### 2.2 排序研究

目前对排序算法的研究主要是基于传统排序算法的改进和引入情感计算、社会影响力等因素,提高搜索排序算法的精度和准确性。

Xie 等<sup>[73]</sup>提出了社交网络的上下文嵌套搜索模型,通过这个模型可以有效地捕获和理解用户目的,同时提出了上下文感知的个性化搜索方法,通过该方法可以动态调整用户信息文件中的标签信息,进而可以根据用户的喜好来搜索上下文的资源,提高了搜索的效率和准确性。Liang 等<sup>[74]</sup>提出了一种改进的混合排序算法 TimeRa,该算法把对微博单独排

序生成的结果进行融合,并根据潜在因素模型推断文件的排序等级和分数,将最终的排序结果返回给用户。Wang 等<sup>[75]</sup>根据与用户回复和转发数进行评分,回复和转发数高的评分越高。为了降低名人微博回复和转发数对评分结果的影响,提出了基于参与用户时间窗的异常检测方法,把用户的原始参与引入评分算法,很好地解决了名人评分问题。Liu 等<sup>[76]</sup>运用 LDA 推断微博的主题分布和用户的兴趣取向,通过实时 Top-k 响应,实时微博查询,然后比较用户取向和微博主题分布之间的相似度判断用户的兴趣。该方法是基于社交网络搜索算法基础上的一个用户个性化的搜索算法,可以应用在社交网络个性化推荐中。卫冰洁等<sup>[77]</sup>将社交网络文本搜索总结为以下三点,时间感知的文本表示、时间感知的文档先验、时间感知的查询扩展,将查询相关文档分布的高峰时刻称为该查询的热门时刻,提出了基于查询热门时刻的 4 个系列语言模型。周霞娟等<sup>[78]</sup>提出了一个用户驱动的可视化微博搜索方法。采用特征词及其权重来对用户兴趣进行建模,在此基础上建立用户特征词直接的关系。搜索微博信息时,首先定位与检索词相关的微博用户,在相关微博用户的微博中筛选与搜索相关的微博。Severyn 等<sup>[79]</sup>把深度学习方法引入社交网络搜索,建立了基于卷积神经网络的短文本对的重排序方法,取得了较好的效果。Chy 等<sup>[80]</sup>提出融合时间特征、用户特征、文本特征的搜索结果重排序方法,并结合上述特征加权的排名模型来估计最后的得分。

上述搜索和排序算法大部分都是基于传统的搜索和排序机制上的改进和扩展,尽管提高了搜索的效率,但是没有从用户角度去感知用户的搜索目的和实现智能化的社交搜索。同时搜索内容都是基于社交网络文本或事件的搜索,而缺少对社交网络中跨媒体内容的索引和搜索机制。

### 3 结论和展望

本文总结了社交网络结构建模、社交网络信息传播、社区发现、情感分析、事件监测及社交网络索引、搜索等技术的研究现状和最近进展。尽管取得了较多成果,但以下方面的研究仍然亟待解决。

现有社交网络挖掘工作一般都是基于宏观层面,很少从微观或者更细粒度层面开展研究,比如情感分析大都只是从积极、消极和中立等情感词分析进行研究,而影响力分析也都只是从全局影响力开展研究,同时在线社交网络用户发表的评论信息

常存在歧义性、模糊性、甚至恶意性等问题,特别是一些影响力较大的用户的言论,会对社会造成较大的影响,因此需要综合利用自然语言处理、统计学习和社交网络挖掘等各种技术进行处理,这将给观点挖掘、情感分析、影响力分析等研究带来挑战。

在线社交网络挖掘领域中的各热点话题都面临着如何研究更有效方法应对在线网络的大规模化、复杂化等带来的效率和质量问题。特别是针对在线社交网络中的海量跨媒体大数据,如何研究高效率、高质量的数据挖掘技术,帮助人们从社交网络中得到更有价值的信息,实现满足用户个性化需求的个性化搜索,仍然是目前该领域重要的研究方向。

在线社交网络搜索技术对特定对象精准搜索的研究还存在不足,不能从超大规模、超高维度、不完全的、有噪声的、模糊的在线社交网络跨媒体大数据中挖掘出与特定对象时空特性、社会属性相关的知识模式,因此需要研究高质量、高效率的跨媒体大数据挖掘算法以及支持时空特性的跨媒体时空数据挖掘体系,根据时空特性、社交特性、用户行为特征,并结合领域本体知识库推理演算和语义查询扩展,构建支持时空特性的在线社交网络对象的精准搜索模型。

随着社交图谱搜索技术的发展,这种与人相关的时间、地点、多媒体、社交内容等进行搜索的技术必将成为社交网络搜索的发展趋势,与此同时,结合深度语义学习、情感立场、事件传播、群体社团关系,实现对用户的社交行为分析、搜索意图理解的智能化、智慧化搜索是一个亟待解决的问题。

由于用户移动端的普及,各个社交网络都提供了手机签到的功能,通过这个功能可以获取用户的时间信息、空间信息,如何结合数据挖掘技术实现基于时空特性的社交网络搜索也是一个需要解决的问题。

### 参考文献:

- [1] 李立耀, 孙鲁敬, 杨家海. 社交网络研究综述[J]. 计算机科学, 2015, 42(11): 8-21, 42.  
LI Liyao, SUN Lujing, YANG Jiahai. Research on online social network[J]. Computer science, 2015, 42(11): 8-21, 42.
- [2] 王大玲, 冯时, 张一飞, 等. 社会媒体多模态、多层次资源推荐技术研究[J]. 智能系统学报, 2014, 9(3): 265-275.  
WANG Daling, FENG Shi, ZHANG Yifei, et al. Study on the recommendations of multi-modal and multi-level resources in social media[J]. CAAI transactions on intelligent systems, 2014, 9(3): 265-275.

- [3] AGRAWAL R, GOLSHAN B, PAPALEXAKIS E. Whither social networks for web search[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2015: 1661–1670.
- [4] 贺超波, 汤庸, 麦辉强, 等. 在线社交网络挖掘综述[J]. 武汉大学学报: 理学版, 2014, 60(3): 189–200.  
HE Chaobo, TANG Yong, MAI Huiqiang, et al. A survey on online social network Mining[J]. Journal of Wuhan university: natural science edition, 2014, 60(3): 189–200.
- [5] SHENG Q Z, VASILAKOS A V, YU Qi, et al. Guest editorial: big data analytics and the web[J]. IEEE transactions on big data, 2015, 1(4): 123–124.
- [6] 唐杰, 陈文光. 面向大社交数据的深度分析与挖掘[J]. 科学通报, 2015, 60(5/6): 509–519.  
TANG Jie, CHEN Wenguang. MAI Huiqiang Deep analytics and mining for big social data[J]. Chinese science bulletin, 2015, 60(5/6): 509–519.
- [7] 许进, 杨扬, 蒋飞, 等. 社交网络结构特性分析及建模研究进展[J]. 中国科学院院刊, 2015, 30(2): 216–228.  
XU Jin, YANG Yang, JIANG Fei, et al. Social network structure feature analysis and its modelling[J]. Bulletin of Chinese academy of sciences, 2015, 30(2): 216–228.
- [8] AGGARWAL C C. Social network analysis[J]. Encyclopedia of social network analysis & mining, 2015, 22(1): 109–127.
- [9] HSU T Y, KSHEMKALYANI A D. Modeling social network topology with variable social vector clocks[C]//Proceedings of 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Paris, France: IEEE, 2015: 584–589.
- [10] DONG Yuxiao. User modeling in large social networks [C]//Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2016: 713.
- [11] SLAUGHTER A J, KOEHLI L M. Multilevel models for social networks: hierarchical bayesian approaches to exponential random graph modeling [J]. Social networks, 2016, 44: 334–345.
- [12] AMATO F, MOSCATO V, PICARIELLO A, et al. Multimedia social network modeling: a proposal[C]//Proceedings of 2016 IEEE Tenth International Conference on Semantic Computing. Laguna Hills, CA, USA: IEEE, 2016: 448–453.
- [13] BAJAJ A, SEN S. Simulating the effect of social network structure on workflow efficiency performance [J]. Social networking, 2014, 3(1): 32–40.
- [14] MIHOUB A, BAILLY G, WOLF C, et al. Graphical models for social behavior modeling in face-to face interaction [J]. Pattern recognition letters, 2016, 74: 82–89.
- [15] RODRIGUEZ M G, BALDUZZI D, SCHÖLKOPF B. Uncovering the temporal dynamics of diffusion Networks [C]//Proceedings of the 28th International Conference on Machine Learning. Bellevue, Washington, USA: ICML, 2011: 561–568.
- [16] JONES S, WEUTHEN T, HARMER Q J, et al. Modeling information propagation with survival theory[J]. Philosophical magazine letters, 2013, 95(2): 85–91.
- [17] RODRIGUEZ M G, LESKOVEC J, BALDUZZI D, et al. Uncovering the structure and temporal dynamics of information propagation[J]. Network science, 2014, 2(1): 26–65.
- [18] SADIKOV E, MEDINA M, LESKOVEC J, et al. Correcting for missing data in information cascades [C]//Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2011: 55–64.
- [19] ROMERO D M, GALUBA W, ASUR S, et al. Influence and passivity in social media [M]//Gunopulos D, Hofmann T, Hofmann D, et al. Machine Learning and Knowledge Discovery in Databases. Berlin Heidelberg: Springer, 2010: 18–33.
- [20] KIMURA M, SAITO K, OHARA K, et al. Speeding-up node influence computation for huge social networks [J]. International journal of data science and analytics, 2016, 1(1): 3–16.
- [21] GUILLE A, HACID H, FAVRE C. Predicting the temporal dynamics of information diffusion in social networks [J]. Computer science, 2013, 144(1): 1145–1152.
- [22] XU Xin, CHEN Xin, EUN D Y. Modeling time-sensitive information diffusion in online social networks [C]//Proceedings of 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Hong Kong, China: IEEE, 2015: 408–413.
- [23] WEN Sheng, HAGHIGHI M S, CHEN Chao, et al. A sword with two edges: propagation studies on both positive and negative information in online social networks [J]. IEEE transactions on computers, 2015, 64(3): 640–653.
- [24] TUAROB S, TUCKER C S, SALATHE M, et al. Modeling individual-level infection dynamics using social network information [C]//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2015: 1501–1510.
- [25] TAMBUSCIO M, RUFFO G, FLAMMINI A, et al. Fact-checking effect on viral hoaxes: a model of misinformation spread in social networks [C]//Proceedings of the 24th International Conference on World Wide Web. New York,



- NY, USA: ACM, 2015: 977-982.
- [26] WANG Ru, RHO S, CHEN Bowei, et al. Modeling of large-scale social network services based on mechanisms of information diffusion; Sina Weibo as a case study[J]. Future generation computer systems, 2016, doi: 10.1016/j.future.2016.03.018.
- [27] PAL A, COUNTS S. Identifying topical authorities in microblogs[C]//Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2011: 45-54.
- [28] SUO Qi, SUN Shiwei, HAJLI N, et al. User ratings analysis in social networks through a hypernetwork method[J]. Expert systems with applications, 2015, 42(21): 7317-7325.
- [29] 吴岷辉, 张晖, 赵旭剑, 等. 基于用户行为网络的微博意见领袖挖掘算法[J]. 计算机应用研究, 2015, 32(9): 2678-2683.
- WU Xianhui, ZHANG Hui, ZHAO Xujian, et al. Mining algorithm of microblogging opinion leaders based on user-behavior network[J]. Application research of computers, 2015, 32(9): 2678-2683.
- [30] SUPPA P, ZIMEO E. A clustered approach for fast computation of betweenness centrality in social networks[C]//Proceedings of 2015 IEEE International Congress on Big Data. New York, NY, USA: IEEE, 2015: 47-54.
- [31] YANG Yang, TANG Jie, LEUNG C W K, et al. Rain: social role-aware information diffusion[C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, Texas, USA: AAAI, 2015: 367-373.
- [32] SUBBIAN K, AGGARWAL C C, SRIVASTAVA J. Querying and tracking influencers in social streams[C]//Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2016: 493-502.
- [33] SU Jianhai, HAVENS T C. Quadratic program-based modularity maximization for fuzzy community detection in social networks[J]. IEEE transactions on fuzzy systems, 2015, 23(5): 1356-1371.
- [34] KLOSTER K, GLEICH D F. Heat kernel based community detection[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM, 2014: 1386-1395.
- [35] ALTUNBEY F, ALATAS B. Overlapping community detection in social networks using parliamentary optimization algorithm[J]. International journal of computer networks and applications, 2015, 2(1): 12-19.
- [36] ARAB M, AFSHARCHI M. Community detection in social networks using hybrid merging of sub-communities[J]. Journal of network and computer applications, 2014, 40: 73-84.
- [37] CHEN Pinyu, HERO A O. Deep community detection[J]. IEEE transactions on signal processing, 2015, 63(21): 5706-5719.
- [38] ZHANG Yuan, LEVINA E, ZHU Ji. Detecting overlapping communities in networks using spectral methods[J]. Physica a: statistical mechanics and its applications, 2014, 405: 1-37.
- [39] GAO Chao, MA Zongming, ZHANG A Y, et al. Achieving optimal misclassification proportion in stochastic block model[J]. Computer science, 2015, 20(3): 88-90.
- [40] MAHMOOD A, SMALL M. Subspace based network community detection using sparse linear coding[J]. IEEE transactions on knowledge and data engineering, 2016, 28(3): 801-812.
- [41] AIROLDI E M, BLEI D M, FIENBERG S E, et al. Mixed membership stochastic blockmodels[J]. The journal of machine learning research, 2008, 9: 1981-2014.
- [42] 赵文清, 侯小可, 沙海虹. 语义规则在微博热点话题情感分析中的应用[J]. 智能系统学报, 2014, 9(1): 121-125.
- ZHAO Wenqing, HOU Xiaoke, SHA Haihong. Application of semantic rules to sentiment analysis of microblog hot topics[J]. CAAI Transactions on intelligent systems, 2014, 9(1): 121-125.
- [43] BRAVO-MARQUEZ F, MENDOZA M, POBLETE B. Combining strengths, emotions and polarities for boosting Twitter sentiment analysis[C]//Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining. New York, NY, USA: ACM, 2013: 2.
- [44] HU Xia, TANG Lei, TANG Jiliang, et al. Exploiting social relations for sentiment analysis in microblogging[C]//Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. New York, NY, USA: ACM, 2013: 537-546.
- [45] NASKAR D, MOKADDEM S, REBOLLO M, et al. Sentiment analysis in social networks through topic modeling[C]//Proceedings of the 10th Edition of the Language Resources and Evaluation Conference (LREC) 2016. Porto-roz: LREC, 2016.
- [46] SIXTO J, ALMEIDA A, LÓPEZ-DE-IPÍÑA D. Improving the sentiment analysis process of spanish tweets with bm25[M]//MÉTAIS E, MEZIANE F, SARAEE M, et al. Natural Language Processing and Information Systems. Switzerland: Springer, 2016: 285-291.
- [47] YOU Quanzeng, LUO Jiebo, JIN Hailin, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks[C]//Proceedings of the

- Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, Texas, USA: AAAI Press, 2015: 381–388.
- [48] CHAO Linlin, TAO Jianhua, YANG Minghao, et al. Long short term memory recurrent neural network based multimodal dimensional emotion recognition [C]//Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge. New York, NY, USA: ACM, 2015: 65–72.
- [49] PORIA S, CAMBRIA E, HOWARD N, et al. Fusing audio, visual and textual clues for sentiment analysis from multimodal content[J]. *Neurocomputing*, 2016, 174: 50–59.
- [50] KALEEL S B, ABHARI A. Cluster-discovery of twitter messages for event detection and trending[J]. *Journal of computational science*, 2015, 6: 47–57.
- [51] D'ANDREA E, DUCANGE P, LAZZERINI B, et al. Real-time detection of traffic from twitter stream analysis[J]. *IEEE transactions on intelligent transportation systems*, 2015, 16(4): 2269–2283.
- [52] LI Jianxin, WEN Jianfeng, TAI Zhenying, et al. Bursty event detection from microblog: a distributed and incremental approach [J]. *Concurrency and computation practice and experience*, 2016, 28(11): 3115–3130.
- [53] ZHANG Xiaoming, CHEN Xiaoming, CHEN Yan, et al. Event detection and popularity prediction in microblogging [J]. *Neurocomputing*, 2015, 149: 1469–1480.
- [54] ZHOU Xiangmin, CHEN Lei. Event detection over twitter social media streams [J]. *The VLDB journal*, 2014, 23(3): 381–400.
- [55] POHL D, BOUCHACHIA A, HELLWAGNER H. Social media for crisis management: clustering approaches for sub-event detection [J]. *Multimedia tools and applications*, 2015, 74(11): 3901–3932.
- [56] GUILLE A, FAVRE C. Mention-anomaly-based event detection and tracking in twitter [C]//Proceedings of 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Beijing, China: IEEE, 2014: 375–382.
- [57] ZHANG Yu, QU Zhiyi. A novel method for online bursty event detection on twitter [C]//Proceedings of the 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS). Beijing, China: IEEE, 2015: 284–288.
- [58] YAN Yan, YANG Yi, MENG Deyu, et al. Event oriented dictionary learning for complex event detection[J]. *IEEE transactions on image processing*, 2015, 24(6): 1867–1878.
- [59] ABDELHAQ H, SENGSTOCK C, GERTZ M. Eventtweet: online localized event detection from twitter[J]. *Proceedings of the VLDB endowment*, 2013, 6(12): 1326–1329.
- [60] SCHINAS M, PAPADOPOULOS S, PETKOS G, et al. Multimodal event detection and summarization in large scale image collections [C]//Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval. New York, NY, USA: ACM, 2016: 421–422.
- [61] GAO Yue, ZHAO Sicheng, YANG Yang, et al. Multimedia social event detection in microblog [M]//HE Xiangjian, LUO Suhui, TAO Dacheng, et al. *MultiMedia Modeling*. Switzerland: Springer International Publishing, 2015: 269–281.
- [62] UNANKARD S, LI Xue, SHARAF M A. Emerging event detection in social networks with location sensitivity [J]. *World wide web*, 2015, 18(5): 1393–1417.
- [63] BOUADJENEK M R, HACID H, BOUZEGHOUB M. Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms [J]. *Information systems*, 2016, 56: 1–18.
- [64] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. *计算机研究与发展*, 2016, 53(3): 582–600.
- LIU Qiao, LI Yang, DUAN Hong, et al. Knowledge graph construction techniques [J]. *Journal of computer research and development*, 2016, 53(3): 582–600.
- [65] 费洪晓, 莫天池, 秦启飞, 等. 社交网络相关机制应用于搜索引擎的研究综述[J]. *计算技术与自动化*, 2014, 33(1): 1–9.
- FEI Hongxiao, MO Tianchi, QIN Qifei, et al. The researches of applying social networking mechanism to search engine: a survey [J]. *Computing technology and automation*, 2014, 33(1): 1–9.
- [66] CHEN Chun, LI Feng, OOI B C, et al. Ti: an efficient indexing mechanism for real-time search on tweets [C]//Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data. New York, NY, USA: ACM, 2011: 649–660.
- [67] CHEN Hanhua, JIN Hai. Efficient keyword searching in large-scale social network service [J]. *IEEE transactions on services computing*, 2015, doi: 10.1109/TSC.2015.2464819.
- [68] LI Yuchen, BAO Zhifeng, LI Guoliang, et al. Real time personalized search on social networks [C]//Proceedings of the 2015 IEEE 31st International Conference on Data Engineering. Seoul, South Korea: IEEE, 2015: 639–650.
- [69] ZHAO Feng, LIU Jun, ZHOU Jingyu, et al. LS-AMS: an adaptive indexing structure for realtime search on microblogs [J]. *IEEE transactions on big data*, 2015, 1(4): 125–137.
- [70] HUANG Haifei, LI Jianxin, ZHANG Richong, et al. Liveindex: a distributed online index system for temporal

microblog data[C]//Proceedings of 2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), the 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICESS). New York, NY, USA: IEEE, 2015: 884-887.

[71] YUAN Jingbo, WANG Bairong, DING Shunli. A real-time search structure and classification algorithm of microblog based on partial indexing[J]. Indonesian journal of electrical engineering and computer science, 2014, 12(3): 2271-2277.

[72] RÍSSOLA E A, TOLOSA G H. Inverted index entry invalidation strategy for real time search[C]//Proceedings of XXI Congreso Argentino de Ciencias de la Computación. Junín: CACIC, 2015.

[73] XIE Haoran, LI Xiaodong, WANG Tao, et al. Personalized search for social media via dominating verbal context[J]. Neurocomputing, 2016, 172: 27-37.

[74] LIANG Shangsong, REN Zhaochun, WEERKAMP W, et al. Time-aware rank aggregation for microblog search[C]//Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. New York, NY, USA: ACM, 2014: 989-998.

[75] WANG Wenbo, DUAN Lei, KOUL A, et al. YouRank: let user engagement rank microblog search results[C]//Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. Palo Alto, California: AAAI, 2014.

[76] LIU Lijun. Research on real-time personalized recommendation algorithm[J]. International journal of u-and e-service, science and technology, 2014, 7(5): 359-368.

[77] 卫冰洁, 王斌. 面向微博搜索的时间感知的混合语言模型[J]. 计算机学报, 2014, 37(1): 229-239.

WEI Bingjie, WANG Bin. Time-aware mixed language model for microblog search[J]. Chinese journal of computers, 2014, 37(1): 229-239.

[78] 周霞娟, 汪飞, 金玲, 等. 用户驱动的微博可视化搜索[J]. 中国图象图形学报, 2015, 20(5): 715-723.

ZHOU Xiajuan, WANG Fei, JIN Ling, et al. User-driven visual micro-blog search[J]. Journal of image and graphics, 2015, 20(5): 715-723.

[79] SEVERYN A, MOSCHITTI A. Learning to rank short text pairs with convolutional deep neural networks[C]//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY, USA: ACM, 2015: 373-382.

[80] CHY A N, ULLAH M Z, AONO M. Combining temporal and content aware features for microblog retrieval[C]//Proceedings of the 2015 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA). Chonburi, Thailand: IEEE, 2015: 1-6.

作者简介: 作者简介:



石磊,男,1986年生,博士研究生,主要研究方向为人工智能和数据挖掘。



杜军平,女,1963年生,教授,博士生导师,博士,中国人工智能学会常务理事,智能服务专业委员会主任,主要研究方向为社交网络分析、数据挖掘、运动图像处理,主持国家“863”、“973”计划项目、国家自然科学基金重点项目、国家自然科学基金重大国际合作项目、北京市自然科学基金重点项目等多项。



周亦鹏,男,副教授,博士,主要研究方向为 Web 挖掘、跨媒体信息检索。