



## 在线社交网络的可视化分析\*

文 / 黄河燕  
北京理工大学计算机学院 北京 100081

**【摘要】** 随着在线社交网络规模的指数性增长,可视化分析成为理解社交媒体及用户的重要手段。文章针对社交媒体、社交用户、社交媒体与用户的交互3个角度,从社交媒体主题可视化建模、社交用户偏好可视化建模、社交网络个性化检索可视化3个方面,分析社交网络可视化技术的研究成果。提出时空融合可视化、“三网”融合可视化等未来国内研究方向。以期利用在线社交网络,为互联网服务、国家安全提供有力支撑。

**【关键词】** 社交网络,可视化建模,社会计算,主题分析,用户个性化

DOI 10.16418/j.issn.1000-3045.2015.02.010

### 1 科学背景

在线社交网络可视化分析,是一种利用可视化技术对社交网络服务中所产生数据进行定量分析的方法。随着在线社交网络的普及,规模庞大的社交用户在社交媒体中表达、交流和传播观点,产生海量高度多样化社交信息。在线社交网络可视化分析作为大数据环境下信息挖掘的重要手段,已逐渐成为高校和企业的研究热点。

目前,在线社交网络的可视化分析技术,在互联网服务和国家安全等方面,已逐步得到应用。社交媒体为用户提供了互动

信息平台,具有更好的交互性和开放性。用户使用微博、微信、人人网、新闻客户端、知乎等媒介,发布、传播和评论信息。面向社交媒体的可视化分析将海量异构信息自动构建成多样性主题,并揭示主题的演化特性。面向用户的可视化分析技术从海量用户行为中发掘用户的个体偏好,群体特性及其演变过程。面向媒体与用户交互的可视化分析,依据用户个性化偏好来进行相关社交媒体匹配,提供更好的社交搜索及推荐。互联网企业,监测、分析及挖掘社交媒体信息的传播路径,精确定位消费者人群,节省营销成本。苹果公司在2013年12月收购能分析推特(Twitter)完整数据流的社交媒体分析公司Topsy Labs。Topsy Labs主要研究

\* 基金项目:国家重点基础研究发展计划(“973”)项目(2013CB329605)

修改稿收到日期:2015年2月1日



中国科学院

推特上的数据,它的工具可以破译某个关键词在推特上出现的次数、找到在特定主题中具有影响力的人物以及衡量某次事件或活动的曝光频率。雅虎公司在2014年3月收购了社交媒体数据可视化创业公司 Vizify。Vizify 公司帮助用户基于在线分享数据创建互动信息图表和视频,从而生成某种类型的互联网个人档案,推动雅虎公司对在线社交媒体可视化技术和用户体验方面的发展。

在国家安全支撑方面,在线社交网络可视化分析可根据媒体传播特点及用户行为规律,发现社交网络异常事件,及时监控危害国家安全的信息及行为在社交网络中的传播。社会信息生产过程已由专业化大众传媒机构垄断,转换成大规模用户的主动行为。在线社交媒体可视化分析技术,将原本隐匿、分散的信息转向公开、汇聚,把握事件发展脉络,评估事件影响力,防止网络舆情被“幕后推手”控制,维护网络信息的安全。2015年1月上海踩踏事件发生后,在微博、新闻客户端等网络社交媒体中爆发性发布、转发、评论,分析这些信息可以挖掘民众对事件的反应与态度,为政府提供舆论的风向,提醒政府注意可能的危机、冲突。另外,通过对抗议、示威和行动等事件的最早现场报告进行检测,可发现潜在危机,提供早期预警,预测公共事件走向。

## 2 社交网络可视化建模研究进展

社交网络可视化研究面向3个方面:社交媒体、社交用户、社交媒体与用户之间的交互(图1)。其中,面向社交媒体的可视化建模针对社交媒体主题进行提取,分析社交媒体多样化,以及

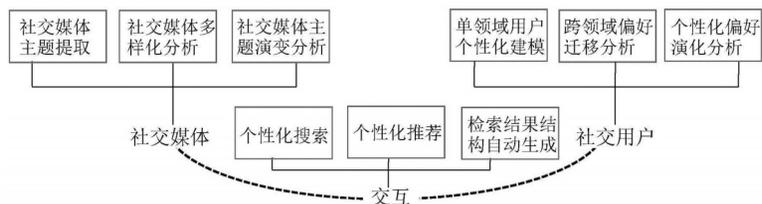


图1 社交网络可视化建模研究概况

社交媒体的主题演化。面向社交用户的可视化建模对用户进行个性化建模,研究社交用户跨领域偏好迁移,以及用户个性化偏好的演化规律。面向媒体与用户交互的研究多集中于社交网络个性化检索的可视化研究,分析社交用户个性化搜索与个性化推荐的有效方法,并自动生成检索结果。

### 2.1 社交媒体主题的可视化建模

社交媒体主题的可视化建模是利用自然语言处理技术与可视化技术,对来自推特、微博、新闻客户端等社交平台的在线资源进行主题提取、多样化分析和主题演变分析。典型的应用场景包括挖掘微博热门话题、预测选举结果以及获取网民对某事件的情绪变化等。图2给出了城市交通主题的可视化案例。从摇号买车困难,到O2O租车服务的兴起,再到出租车司机在多地罢工。该例通过词云主题图来可视化社交媒体主题的演化过程。



图2 社交媒体主题可视化(以城市交通为例)

社交媒体主题提取技术是指利用概率潜语义分析(PLSA)、潜在狄利克雷分配(LDA)等主题模型,结合传统基于关键字的主题提取方法,以及基于字典学习的概率矩阵分解方法,对社交网络信息进行主题提取,挖掘媒体的潜在语义。目前社交媒体主题检测主要有3种方法。第一种是使用字典学习的方法挖掘主题词之间的关联;第二种是使用传统事件检测的方法预测新的主题关键词;第三种是使用主题

模型的方式获取动态词分布。北京航空航天大学陈燕、新加坡国立大学的Hadi Amiri等人<sup>[1]</sup>提出微博未来热门主题词的预测方法,针对银行、大学、政府部门等组织,跟踪与特定组织相关的主题演化,预测组织相关的新主题的产生。这种方法收集包含固定关键字、新兴关键字、公众账号与组织关键用户的实时信息,使用SVM剔除无关信息,建立机构的用户网络,并用主题模型进行分析,然后使用增量聚类算法进行实时监测。

社交媒体多样化分析是指在识别主题的基础上进一步划分主题的不同类型。例如,有的主题属客观事实描述,有的主题是主观感受。目前,美国康奈尔大学蒋云与Ashutosh Saxena<sup>[2]</sup>提出

一种主题分解建模方式 (Factored Topic Models),使用多个平行的主题因素 (Topic Factors)对主题进行建模。在该模型中,每个词的生成过程取决于一组特征向量,而不是已有工作的单变量。向量中的每一种元素代表一类特定的主题类型。与传统的LDA、层次DP等算法相比,该算法所发掘的主题更加简洁,而且有更多实际意义。

社交媒体主题的演变分析是研究社交媒体主题的发展、传播、交互的动态演变过程。信息交互的形式是多种多样的,如社交媒体中一个话题可以通过压制另外一个话题以提高本话题的用户关注度(话题竞争),同样话题之间也可以互相协同一起来吸引更多用户的关注(话题合作)。微软亚洲研究院吴英才等人<sup>[3]</sup>提出了一种基于时间的可视分析方法EvoRiver,对社交媒体大数据主题之间的合作关系进行建模和分析,旨在通

过可视化的形式揭示信息在社交媒体这一虚拟空间上的交互过程。EvoRiver可用于可视化和理解话题在话题领袖作用下复杂的、动态的竞争合作情况,展示话题在社交媒体中动态变化的过程。

## 2.2 社交用户偏好可视化建模

社交用户偏好可视化建模是对社交用户在社交媒体中的行为数据进行挖掘,对用户进行单领域个性化建模,跨领域偏好迁移分析,及个性化偏好演化分析。图3左边给出了社交用户偏好可视化的案例。该年轻女性的偏好随时间变化,由婚庆产品,到孕期卫生,再到婴儿服务。每个阶段的词云表示该用户在该阶段的信息偏好。

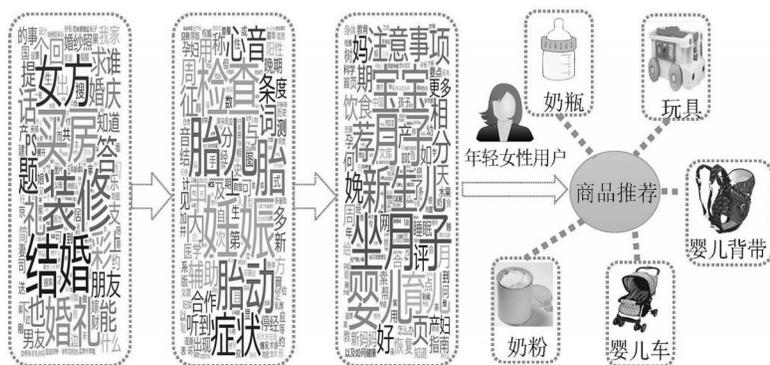


图3 社交用户偏好可视化建模及个性化检索可视化建模案例

单领域用户个性化建模是指根据用户在单领域网站的反馈,对用户的信息偏好进行描述。例如,根据用户对电影的反馈,来对其电影偏好进行建模。典型方法主要是基于协同过滤的用户建模方法。在2014年国际知识发现与挖掘会议上(KDD 2014),新加坡管理大学的刁启明<sup>[4]</sup>提出一种融合客观用户评分和主观用户评论的个性化建模方法。一方面,借助于传统概率矩阵分解的方法描述客观用户评分的低秩特性;另一方面,通过语言模型挖掘主观评论中的评价对象及评价正负倾向。两者结合之后,不仅可以知晓用户对某电影的综合评价,也可以知



道用户给出该评价的原因,如喜欢某电影演员或喜欢某电影情节。后者通过自然语言表达,对用户的偏好给出了可视化建模。

跨领域偏好迁移分析是根据用户在不同领域之间的反馈进行信息偏好的迁移学习与分析。该问题的背景主要来源于单网站用户个性化建模的数据稀疏问题。由于用户在单网站的反馈数呈幂律分布,大部分评论集中在少数用户手中,而大部分用户反馈量极少。因此,迁移分析是描述用户在多个网站之间的信息偏好关系。旨在通过用户在某领域的密集反馈,来推理其在其他领域的信息偏好。例如,某用户在电影网站的评论较多,但是在书籍评价网站的评论较少,那么迁移学习便可根据其电影偏好,推知其对书籍的偏好。在2013年国际互联网大会(WWW 2013)上,上海交通大学胡亮等人<sup>[5]</sup>提出了一种基于张量的迁移协同过滤模型。在传统矩阵分解模型的基础上,该工作扩展了一层新的数据维度,用来表示某领域的特征,使得每个领域有一个独立的特征向量。在判断某用户对某网站的评价时,该特征向量会连同用户特征向量及信息项特征向量一块计算。通过挖掘跨领域的迁移关系,缓解数据稀疏问题。

个性化偏好演化是指对用户信息偏好随时间的变化产生迁移的过程进行描述分析。例如,某用户在单身的时候喜欢科幻电影,在结婚后喜欢怀旧电影,在有了孩子后关注儿童电影。如果能够可视化某用户的偏好演变,则可根据其当前的信息偏好为其推荐产品,提高准确性。在2013年国际信息检索(SIGIR 2013)大会上,新加坡国立大学陈伟等人<sup>[6]</sup>提出一种RTM(Receptiveness over Time Model)的概率生成模型,对用户随着时间推移所产生的兴趣分布变化进行分析,其主要依据来源于用户在不同时期会与不同类型的朋友交互,而这些朋友的兴趣点反映了该用户在此时间段的兴趣。例如,在用户有了孩子以后可能更多与儿童教育相关的专家进行交互,那么该用户当前的兴趣点,便可通过这些专家的信息偏好推知。

## 2.3 社交网络个性化检索的可视化

社交网络个性化检索的可视化将个性化检索技术与可视化技术相结合,对社交用户在社交网络的交互进行个性化搜索分析、个性化推荐分析与检索结果自动生成。实现对用户面临海量信息进行个性化过滤,辅助信息的生产者将信息更快地传递到合适的用户。图3右边给出了个性化检索可视化的案例。在描述出用户对婴儿产品的偏好后,与结构化的婴儿服务社交媒体数据进行匹配,并将最相关的婴儿用品展示给用户。

个性化搜索分析,是在传统信息搜索结果的基础上进行再加工,根据用户喜好把相关的检索结果返回。它既需要根据用户的历史查询信息,对用户的搜索意图做出准确理解,又需要准确匹配相应的搜索结果。在WWW 2013上,美国微软研究院的White等人<sup>[7]</sup>提出了基于协同任务的个性化搜索方法。当用户给出查询后,所提出方法会借助于协同过滤的思想,在数据库中查找类似用户的类似查询任务。根据其他用户在该查询任务的历史点击记录,来提高个性化搜索分析的准确性。

个性化推荐分析是建立用户的个性化描述模型后,主动向用户推荐信息,让信息主动寻找用户。可视化个性化推荐的结果主要是从多样性的角度考虑,防止推荐结果千篇一律。在2013年国际联合人工智能大会上(IJCAI),清华大学的琴丽晶等人<sup>[8]</sup>提出一种基于最大熵的推荐方法。利用熵模型来描述推荐结果的多样性。通过最大熵的优化目标与传统信息推荐的矩阵分解优化目标相结合,使得推荐的信息既能与用户的兴趣相关,同时所推荐的信息项之间具有较大差异性。

检索结果结构自动生成是对个性化搜索或推荐的结果进行视图层面再加工,以结构化的形式展现给用户。所涉及的研究内容包括个性化检索的多样性分析、结构学习以及主题建模。在2014年国际知识发现与挖掘会议上(KDD 2014),荷兰阿姆斯特丹大学的梁尚松等人<sup>[9]</sup>提出了一种监督

式检索结果结构化方法。在该方法中,作者设计了一种概率图主题模型来描述用户兴趣结合检索结果所共同生成的主题。融合结构化支持向量机来做检索结果结构化时,考虑检索结果多样性、用户兴趣点一致性两个约束。通过实验证明,该方法相比非监督式的结构化方法会更加稳定。

### 3 国内研究基础分析

近年来,国内以新浪微博、腾讯QQ为代表的社交媒体已经融入了人们的日常生活,国民开始习惯在社交平台中查看热点事件、表达个人观点。各个行业对大规模社交数据处理和深度分析的需求,推动国内对社交媒体主题可视化建模、社交用户偏好的可视化建模、社交网络个性化检索可视化的研究不断发展。

在社交媒体主题的可视化建模方面,北京航空航天大学许可课题组开发了针对中文微博的在线情感分析系统 MoodLens,通过对新浪微博的情感分类和波动分析,实现物理世界异常或突发事件的监测。MoodLens 系统将情绪细分为高兴、愤怒、低落和厌恶4类,在传统搜索机制中加入情绪维度,提供基于文本情感分析的搜索服务。在解决大规模数据处理的问题上,课题组设计了一个快速分类器处理超过350万条具有情感标注的微博语料,通过增量学习等策略解决了诸如新词产生、词汇感情色彩演化等带来的问题。利用该分类器,研究小组对发布于2011年的近7000万条微博进行情感分析,根据各类情绪比例的波动提出一个快速的异常点发掘算法,并发现2011年全年发生的一些典型的异常或突发事件,均可以被有效地检测出来,甚至可以观察事件带来影响的持续时间等,有效地论证了基于微博情感理解实现事件监测的可信性。北京

大学袁晓如课题组<sup>[10]</sup>开发微博可视分析工具(WeiboEvents),能够快速浏览和分析微博事件中的人和事,把握微博传播的脉络,挖掘深层的人人、人事关系。WeiboEvents利用可视化技术将一条条独立的微博连接起来,提供给用户多种可视化形式展现微博事件的传播方式,方便用户快速浏览和分析微博事件中的人和事,挖掘事件中的信息。WeiboEvents提供面向专家与公众用户的在线可视分析系统与离线的专家分析系统,利用信息可视化技术清晰地呈现出一个事件中的关键人物、关键微博、重要观点。

在社交用户偏好的可视化建模方面,为捕获用户长期和短期的偏好,中科院和IBM中国研究院的项亮和袁泉等人<sup>[11]</sup>提出了基于会话的时态图STG(Session-based Temporal Graph)来刻画用户随时间变化的长期、短期偏好,并基于STG模型提出了新颖的时序推荐算法IPF(Injected Preference Fusion),有效提升了用户行为建模的准确度。在用户行为建模应用中,清华大学的唐杰<sup>[12]</sup>提出名为Confluence的概率模型对用户行为进行建模和预测,研究社交用户的从众行为。通过对社交网络关系进行深入研究,对网络用户的行为和用户交互进行挖掘,实现自动识别出用户之间的社交网络关系。在处理大规模社交网络问题上,使用分布式的学习方法实现对构造影响模型的效率的近似线性的提高。通过构造融合模型区分用户的个体从众性和群体从众性,并在多个社交网络数据集上的实验验证了该方法的有效性。腾讯公司与香港科技大学,提出一种ComSoc方式对复合社交网络用户行为进行建模<sup>[13]</sup>。这种方法使用Map Reduce算法来处理大规模数据,利用用户在不同的社交媒体中的行为的相互影响,解决目前根据用户的历史行为日志数据行为建模产生的数据



中国科学院

稀疏问题,提高用户建模的准确性。通过在腾讯数据集上进行验证,ComSoc方法对用户行为的预测有显著的提高。

在基于社交网络的个性化检索的可视化方面,清华大学的唐杰<sup>[14]</sup>提出跨领域话题学习方法。该方法解决了跨领域推荐的3个关键难点:链接稀疏性、知识互补性和话题偏斜性,提高了交叉领域合作者推荐的精度。李国良等人<sup>[15]</sup>也研究了基于位置的推荐方法,提出将结构和内容相结合的基于R-tree的索引方法提高数据索引的效率。清华大学的王建勇利用用户兴趣模型提高推特数据中的命名实体识别精度,利用异构网络建模的结果提高个性化的标签推荐精度,其基本思路是利用有导随机游走模型学习不同类型关系和不同类型节点对标签推荐的重要性。北京大学的崔斌等人<sup>[16]</sup>提出名为LCARS的推荐模型对社交数据中的位置、内容以及用户兴趣同时进行建模。在传统的协作过滤系统中,用户只访问距离很近的场所User-item的矩稀疏。当用户到一个没有历史记录的新城市时,这个问题变得尤为严峻。LCARS的推荐模型同时考虑用户的个人兴趣与地域偏好,为特定用户推荐合适的场所、活动,并且在用户经常活动区域与新到地区都有很好的效果。LCARS的推荐模型包含在离线建模LCA-LDA和使用TA算法的在线推荐两部分。LCARS的推荐模型在国内豆瓣网络的数据上进行了实验验证,得到了更高的推荐精度。

#### 4 未来研究展望

近年来,国内在线社交网络可视化分析理论研究为我国在线社交网络的利用与研究奠定了坚实的基础,但在应用于国民服务、国家安全方面仍有很多问题亟待解决。本文提出2点研究展望。

(1)融合时空的社交媒体可视化研究。在线社交网络时空数据提取、计算与可视化技术,为制定公众服务策略提供依据。在线社交网络中蕴含丰富的时空数据,可以清晰地描述用户在某个时

间的地理位置。通过用户与社交媒体的交互、用户行为分析技术推测用户的意图与行为,借助海量数据处理与计算,对社交网络时空数据进行可视化,为公共策略的制定提供依据。在景区人流控制、人群密度分析、环境污染检测等方面,依靠时空数据可视化制定策略,可以更好地利用公共资源,防范旅客滞留、踩踏等事件的发生,在提高公众服务方面具有长远的意义。增强时空社交媒体主题可视化研究,推动建立社交媒体信息透明化机制。当前社会,公众期待建立媒体信息透明化机制,真正地参与公众事件,而不是被媒体态度所牵引。但目前海量社交媒体数据处理、信息透明化规则等问题尚未解决,严重阻碍社交媒体透明化机制建立的进程。社交网络时空个性化推荐的研究,对不同的推荐任务制定推荐策略。综合用户意图识别建模、内容建模、社交建模,从不同粒度对用户模型进行更新,制定推荐策略提升推荐的准确度。提供社交媒体个性化推荐的平台,为信息定向发布提供依据。

(2)传统网络、移动互联网、物联网“三网”融合的可视化研究。社交网络强调用户体验,具备聚合互联网、移动互联网和物联网的条件,可推动“三网”融合实现个人化的全媒体播放模式。在“三网”融合中,用户是内容的制造者也是市场的推动者,根据用户多元化、个性化需求提供相应的服务,提升市场竞争力。目前Facebook已经整合了短信、聊天及电邮系统,开拓在线视频服务领域,推出在线电影租售服务。“三网”融合应用广泛,遍及智能交通、环境保护、政府工作、公共安全、平安家居等多个领域,涉及的广电业、电信业和互联网产业都是技术和知识密集型产业,“三网”融合已经上升为国家战略的高度,对推进调整产业结构和发展电子信息产业有着重大的意义。在“三网”融合的进程中,在线社交网络可视化分析技术可以根据媒体传播特点和用户行为规律,监控网络舆情、及时发现情报(图4)。国内的社交平台拥有更加庞大规模社交用户,在获得全面的

社交数据上具有优势。所以我国更应利用好在线社交网络可视化分析成果,凸显其在维护国家安全、打击网络谣言、发掘幕后推手等方面的作用。完善提取社交数据中与国家安全相关的“敏感主题”的方法,客观、全面地挖掘“敏感主题”,监控“敏感主题”的演化过程。建立网络内容分级和过滤系统,对社交网络中的不良信息进行澄清,增强主流信息公信力。研究海量社交数据提取情报的方法,建立系统获取网络情报的体系,为维护国家安全服务。在线社交网络作为信息传播最快、最广泛的媒体,在国家安全领域的作用引起全世界关注。如美国国防部、情报机构利用社交媒体的海量信息,从博客、新闻站点和实时社交平台(如 Twitter)中提取重要的认知和传播主题,提取情报。提高社交媒体主题演化分析与可视化效果,利用可视化手段监察舆情的产生、传播、发展过程,建立高效的社交网络监察机制。及时打击操纵网络舆情的幕后黑手,消弱网络“水军”对网民意见的绑架,掌握舆论控制权,净化网络环境。

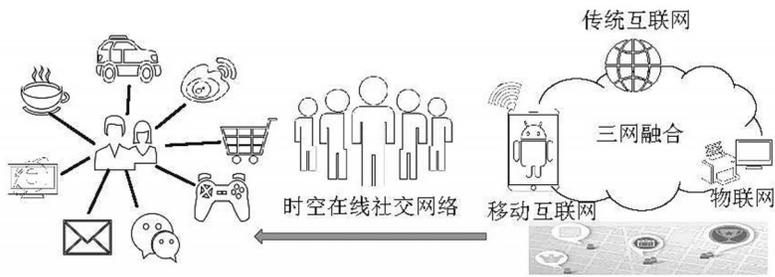


图4 基于时空的“三网”融合在线社交网络

综上所述,国内专家、学者要更加重视社交网络可视化分析技术在推动知识传播、提升公共服务方面的作用,重视其在维护国家安全方面的地位。积极开展针对国民服务、国家安全方面的研究课题,推动在线社交网络可视化分析走向产业化,切实服务于国家与人民。

参考文献

- 1 Chen Y, Hadi A, Li Z J et al. Emerging topic detection for organizations from Microblogs. SIGIR' 13, 2013, 43-52.
- 2 Jiang Y, Ashutosh S. Discovering different types of topics: factored topic models. Proceedings of the 23 international joint conference on artificial intelligence, 2013, 1429-1436.
- 3 Sun G D, Wu Y C, Liu S X et al. EvoRiver: visual analysis of topic coepetition on social media. IEEE Transactions On Visulization And Computer Graphics, 2014, 20 (12): 1753-1762.
- 4 Diao Q M, Qiu M H, Wu C Y et al. Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). ACM SIGKDD, 2014, 193-202.
- 5 Hu L, Cao J, Xu G D et al. Personalized recommendation via cross-domain triadic factorization. Proceedings of WWW 13, 2013.
- 6 Chen W, Wynne H, Mong L L. Modeling user's receptiveness over time for recommendation. SIGIR' 13, 2013: 373-382.
- 7 White R W, Wei Chu, Awadallah A H et al. Enhancing personalized search by mining and modeling task behavior. Proceedings of WWW 13, 2013, 1411-1420.
- 8 Qin L J, Zhu X Y. Promoting diversity in recommendation by entropy regularizer. Proceedings of the 23 international joint conference on artificial intelligence, 2013, 2698-2704.
- 9 Liang S S, Ren Z C, Maarten D R. Personalized search result diversification via structured learning. ACM SIGKDD, 2014, 751-760.
- 10 Ren D H, Zhang X, Wang Z H et al. WeiboEvents: A crowd sourcing weibo visual analytic system. IEEE



中国科学院

- Pacific Visualization Symposium, 2014, 330-334.
- 11 Xiang L, Yuan Q, Zhao S W et al. Temporal recommendation on graphs via long-and short-term preference fusion. ACM SIGKDD'10, 2010, 723-732.
- 12 Tang J, Wu S, Sun J M. Confluence: conformity influence in large social networks. ACM SIGKDD'13, 2013, 347-355.
- 13 Erheng Z, Fan W, Wang J W et al. ComSoc: adaptive transfer of user behaviors over composite social network. ACM SIGKDD'12, 2012, 696-704.
- 14 Tang J, Wu S, Sun J M et al. Cross-domain collaboration recommendation. ACM SIGKDD'12, 2012, 1285-1293.
- 15 Li G L, Wang Y, Wang T et al. Location-aware publish/subscribe. ACM SIGKDD'13, 2013, 802-810.
- 16 Yin H Z, Sun Y Z, Cui B et al. LCARS: a locationcontent-aware recommender system. ACM SIGKDD'13, 2013, 221-229.

## Visual Analysis of Online Social Networks

Huang Heyan

(School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China)

**Abstract** With the exponential growth of online social networks, the visual analysis technique has become an important research issue for understanding social media and social network users. In this article, targeting at social media, social network users, and their interactions respectively, we provide a detailed survey on research work of social media topic analysis, social user modeling, and personalized social information retrieval. We hope these techniques can be utilized to improve the Web user experience and the national security service. Recent research work in visualizing social media are divided into: (1) topic analysis; (2) topic diversification; and (3) topic evolution. Topic analysis means to construct a group of themes from a set of documents, and assign each document with a theme or a group of themes. Typical methods include the Probabilistic Latent Semantic Indexing (PLSI), the Latent Dirichlet Allocation (LDA), etc. Topic diversification refers to distinguish different types of topics. For example, some topics are objective contents, and other topics are subjective sentiments. In order to achieve this, supervised topic models are proposed to identify different topic types. Topic evolution is to connect topics in consecutive time intervals through topic similarity measures, through which one can see the topic change, the topic merging, and the topic separation. Generally, KL-divergence is utilized as the similarity measure. The work in visualizing social network users are classified into: (1) single-domain user preference modeling; (2) cross-domain preference transformation; and (3) user preference evolution. Single-domain user preference modeling is to describe a user's personal bias on items in a single domain. For example, to model a user's preference in a movie review system. Typical solutions are to combine the review aspect modeling with traditional collaborative filtering, in which the aspect is utilized for preference description. Cross-domain preference transformation is to learn the user preference pattern through multiple domains. In a typical scenario, a user has some feedbacks in a movie review site, but has very few feedbacks in a book site. By transferring learning techniques, the movie feedback is utilized to infer the book preference. User preference evolution refers to distinguishing a user's interests in different time periods. For example, a user may pay attention to furniture when he/she moves to a new house, but he/she may be interested in milk powder when he/she has a baby. By modeling the evolution patterns, the interests in the next time period can be predicted. The work in visualizing the interactions between social media and users are grouped into: (1) personalized search; (2) personalized recommendation; and (3) search/recommendation results reconstruction. Personalized search means to return search results by consid-

ering a user's search history. Thus a user model is constructed in advance, and the displayed search results are required to match the user's intent. Personalized recommendation is to retrieve information without queries. Usually, collaborative filtering models are utilized to match a user's interest and the relevant items. In visualizing the items to the user, the diversity among a group of displayed items needs to be considered. Therefore, some entropy-based objectives are proposed to be joint with the traditional collaborative filtering objectives. Search/recommendation results reconstruction refers to convert unstructured retrieval results into structured data automatically. This will make both conveniences for user experiences and search engine indexing. Some future work directions are also discussed in the end, such as visualizing temporal-spatial properties of social network, visualizing the combination of internet of things, mobile internet, traditional internet, etc.

**Keywords** social network, visualization, social computing, topic analysis, user personalization

**黄河燕** 北京理工大学计算机学院院长,教授,博士生导师。1989年获中科院计算技术研究所计算机科学与技术博士学位。北京市海量语言信息处理与云计算应用工程技术研究中心主任,中-德语言信息处理联合实验室(BIT-DFKI)主任。主要研究方向为自然语言处理和机器翻译,社交网络与信息检索,智能处理系统等。主持承担了多项国家级重要科研项目和国际合作项目。先后获得中科院科技进步奖一等奖、国家科技进步奖一等奖等多项奖励。E-mail: hhy63@bit.edu.cn.

**Huang Heyan**, is currently Professor and Dean of School of Computer Science & Technology, Beijing Institute of Technology. She received her Ph.D. degree in Computer Science and Technology in 1989 from Institute of Computing Technology, Chinese Academy of Sciences. Her current research interests include machine learning, information retrieval, and natural language processing. She has published over 90 research papers in high-profile journals and conferences. She serves on the editorial boards of *International Journal of Advanced Intelligence* and *Journal of Computer Research and Development*. She has conducted 20 more research projects including National "863" Project of China, National "973" Project of China, Key Project of National Natural Science Foundation of China, etc. Huang Heyan got the first Prize of National Scientific and Technological Progress of China in 1995, the first Prize of Scientific and Technological Achievements of Chinese Academy of Sciences in 1993, and the second Prize of Scientific and Technological Achievements of Chinese Academy of Sciences in 2001. E-mail: hhy63@bit.edu.cn



中  
國  
科  
學  
院