

徐常胜^{1,2} 黄晓雯^{1,2} 钱胜胜^{1,2} 方全^{1,2}

基于社会多媒体内容的用户建模应用研究



作者简介:徐常胜,男,中国科学院自动化研究所特聘研究员,博士生导师,国家杰出青年科学基金获得者,国家重点研发计划项目“大数据多模态交互系统关键技术”首席科学家,中国科学院王宽城率先人才计划卢嘉锡国际团队“多媒体计算国际团队”负责人。IEEE Fellow、IAPR Fellow、国际计算机学会杰出科学家(ACM Distinguished Scientist)。担任ACM多媒体专委会中国区主席和中国计算机学会多媒体专委会副主任。担任过国际多媒体期刊 *IEEE TMM* 和 *ACM TOMM* 的 Associate Editor,国际顶级多媒体会议 ACM Multimedia (2009) 程序委员会主席和 2010 年全国多媒体会议程序委员会主席。在多媒体分析、计算机视觉、模式识别、图像处理等领域发表论文 300 多篇,其中 IEEE 和 ACM 汇刊论文 80 篇,发表的论文在 Google Scholar 上被引用超过 13 000 次,H-Index:61。在多媒体国际顶级会议和期刊上获得最佳论文奖 10 余次,获得 2018 年度中国电子学会科学技术(自然科学)一等奖,5 次获得中国科学院优秀导师奖。

E-mail:csxu@nlpr.ia.ac.cn

收稿日期 2019-08-08

基金项目 国家自然科学基金(61432019,61720106006,61572503,61802405,61872424,61702509,61832002);国家重点研发计划(2017YFB1002804)

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
2 中国科学院大学,北京,100190

摘要

随着互联网和社会媒体平台的发展,社会媒体吸引着数以亿计的用户参与其中进行创造和分享信息,产生了海量的文本、图像、音频和视频数据。面对这些数量巨大、异构多源、模态复杂的社会多媒体内容数据,如何对其进行有效的内容理解和知识表示,从而为用户提供更高效、优质的服务,成为实现社会媒体大数据价值的关键。本文对近年来在社会多媒体内容分析、知识提取和表示以及用户建模应用的相关研究展开综述,并针对社会多媒体特征融合、跨模态知识提取与表示,以及基于社会媒体的用户建模相关应用研究三个方面进行详细总结。随后对社会多媒体内容的知识表示和用户建模的研究与应用的发展趋势进行介绍,最后对多媒体知识表示与用户建模研究进行了总结和展望。

关键词

社会多媒体;多模态;知识表示;用户建模

中图分类号 TP391.1

文献标志码 A

0 引言

互联网的发展促使社会媒体的出现和兴盛,社交网络、在线购物网站、视频分享网站等越来越受到人们的欢迎。如图 1 所示,以微博、淘宝、爱奇艺等在线网络平台为代表的社会媒体的出现,使用户可以从这些丰富的网络应用中寻找各自感兴趣的内容或需要的物品。人类已经进入社会多媒体大数据时代,社会媒体作为一种新型的允许人们创造和分享媒体信息的工具和平台,近年来得到飞速的发展,吸引着全球数以亿计的用户参与其中。以用户为中心的网络数据纷繁复杂、包罗万象,除了数据量巨大的特点,数据类型也相当丰富,包括文本、图像、视频、关系、行为等,如何从如此复杂多样的信息中挖掘出有价值的信息是个严峻的挑战。因此,对社会多媒体内容进行有效的知识提取和表示、分析和研究是十分必要的。多媒体内容的知识提取和表示以社会媒体网络上的多模态内容为研究对象,主要目的是通过社会媒体平台上丰富的多模态大数据,提取有价值的信息,构建有效的知识表示,实现对社交媒体用户的理解;另一方面,我们进行有效的用户建模研究,从而为用户提供更优质的个性化服务,例如用户人口属性推断、用户关系标注、个性化推荐等,同时也为社会媒体网络系统的在线广告、异常行为监测等方面提供帮助。



图 1 社交媒体平台高速发展

Fig. 1 Rapid development of social media platforms

社交媒体使得用户可以随时随地获取和分享信息,用户产生的文本、图像、音频、视频等各种网络内容数据呈爆炸式增长,产生了海量的社会多媒体内容数据.据 IDC2013 年初的研究报告指出,预计 2020 年全球数据总量将达到 40 ZB,也就是 40 万亿 GB,平均每人 5 200 GB,年均增长率超过 40%.大数据的潜在价值已经逐步被人们认可并重视.

在我们为大数据的潜在价值欣喜的同时,社交媒体大数据的复杂特性也为知识提取和表示带来了巨大的挑战.社交媒体大数据的复杂特性主要体现在以下几个方面:

1) 大规模性.社交媒体的高速发展带来了数据的爆发式增长,并且涨势迅猛、体量非常庞大.以时下最热门的短视频平台之一快手为例,快手发布的 2018 年度数据报告显示,2018 年,1.9 亿用户在快手发布作品,点赞数逾 1 400 亿,使用总时长突破 500 万年,库存短视频数量百亿级,每日新增视频千万级.

2) 多模态性.社交媒体涵盖了各种各样的网站平台,包括新闻网站、论坛贴吧、微博微信公众号、图片视频分享网站、点评网站、百科地图、知识社区、短视频/直播网站等.同一实体的知识内容数据会以文本、音频、图片、视频以及新媒体文件如 3D 等描述呈现.

3) 多源性、异构性.互联网上的多媒体数据可能由官方或者个人发布上传,并存在于新闻、博客、播客、论坛、视频分享等不同的网站上.从类型看,除了基本的图像、视频和语音,出现了很多新媒体形式,如图片微博、语音图片、带地理位置的视频等.同时,用户通常活跃于各种不同类型的社交媒体平台上,由于各类社交媒体平台的功能不同,用户在各类平台上产生的行为具有异构性.

4) 价值密度低.从产生机制看,开放式网络环境的用户贡献机制下,产生了大量重复、低质量的数据.从需求特点看,在每日上传的 10 万 h 的 YouTube 视频和 180 万张 Flickr 照片中,满足特定查询需求或个性化偏好的只有很小一部分.

另一方面,社交媒体网络以用户为中心,绝大多数的社会多媒体数据都是用户生成内容 (User Generated Content, UGC).用户是社交媒体网络发展的核心竞争力,如何通过海量的 UGC 社交媒体大数据内容挖掘有效的信息,从而服务于用户,给用户更好的社交体验至关重要.因此,用户建模成为了社交媒体持续有力发展的关键技术.但面向社交媒体的用户建模包含几个关键问题:

1) 社交媒体平台的用户行为信息是异构、冗余的.用户的行为对象不尽相同,如文本、图像、视频、音频等.即使对于同一行为对象,用户行为模式也多

种多样,如上传、收藏、分享、评论等。此外,在不同平台的这些用户信息之间可能是冗余甚至是彼此对立的。如何去冗余信息,有效融合异构信息是社会媒体平台用户建模的重要内容。

2)用户在网络上的行为有着丰富的上下文信息,如时间、地点、天气、事件等。在不同的上下文信息下,用户有着不同的行为模式。因此,通过用户在社会媒体网络上的行为信息,有效提取行为内容的知识表示,学习用户在不同上下文情境下的用户表示,有助于更深刻理解用户行为,实现更精准的用户个性化服务。

3)用户行为具有动态特性,用户在社会媒体网络上的行为随着时间的推移会发生变化。但这种行为变化并不是随机产生的,而是具有很强的时序依赖性,即用户在时间节点前的行为信息对时间节点上的行为会产生较大的影响。用户行为的动态特性给用户建模带来了极大的挑战,克服这种挑战有利于我们有效捕捉用户的短期兴趣,使得用户建模更加全面精准,有利于提升个性化服务品质。

综上所述,由于社会多媒体内容的大规模、多模态、多源异构等特性,以及用户行为的异构、上下文依赖、动态变化等特性,基于社会多媒体内容知识表示的用户建模研究极具挑战性。如何对用户生成的多模态内容进行有效的整合,获得优质的知识表示,并设计有效的用户建模模型来深入理解用户,促进用户和社会媒体平台协同发展成为社会多媒体领域一个关键的研究问题。本文对近年来在社会多媒体内容分析、知识提取和表示以及用户建模应用的相关研究展开综述,并针对社会多媒体特征融合、跨模态知识提取与表示以及基于社会媒体的用户建模相关应用研究三方面进行详细总结。

1 研究现状

本章围绕社会多媒体特征融合、跨模态知识提取与表示,以及基于社会媒体的用户建模相关应用研究三方面进行回顾与总结。

1.1 社会多媒体特征融合

近年来,随着网络技术的发展,社会多媒体资讯载体逐步从传统的纯文本内容,转变为富媒体(包含更加丰富的多媒体素材)内容信息。同时,随着大数据技术的发展推广,目前互联网上已经积累了海量的文本、图像、视频、音频等多模态耦合的媒体内容,在此环境下,传统的基于单模态(如文本信息等)的

特征表示技术已经难以满足现有需求,因此,近年来,学术界提出了基于多模态协同挖掘的内容理解技术,以解决这一痛点问题。在社会多媒体内容理解技术中,主要包括了单模态特征表示和多模态特征融合两个部分。

1.1.1 单模态特征表示

社会多媒体内容理解任务中,依赖了自然语言处理、计算机视觉、语音识别等单模态的学习技术,以形成单一模态的特征向量表示,进而支撑跨模态的协同计算。典型地,文本信息可通过词袋模型表示成词频、TF-IDF 特征向量,或使用主题模型,将文档与主题关系建模为矩阵形式,学习潜在的特征表示信息,如 LSI、pLSI、LDA 等。近几年,词向量表示成低维稠密的实值向量的方法^[1],因高效实用而得到大量关注。图像可以提取不同的底层特征表示,如全局性特征:颜色直方图、颜色矩、纹理信息特征、形状信息特征和场景信息特征 GIST^[2]等;局部性特征: SIFT^[3]。音频常用 MFCC^[4]等基于内容的特征描述。对于视频,则在视觉特征外,还需考虑时空信息特征。

以往的特征学习过程中,严重地依赖了人工特征的选择,并且在不同的场景下,均需要定制化地设计出一套与之对应的特征规则,非常不利于数据特征的泛化使用。同时,在现今大数据技术广泛应用的基础上,算法场景逐渐面临着大规模、多模态、异构、非结构化等特点,这使得传统方法的应用显得越来越困难。深度学习技术的出现,有效地解决了这一痛点问题。

深度学习的主要思想是:通过神经网络模拟人脑的多层抽象机制,来实现对数据的抽象表达,进而构建出一种“端到端”的学习模型。该技术的目的是通过神经网络技术,在大规模训练集上,通过梯度下降等优化算法,自动地学习出各类数据的特征表示,通过不断优化模型参数,使模型具备一种高性能的非线性映射能力,从而挖掘出数据中所蕴含的复杂模式,以支撑下游任务。

从 2006 年 Hinton^[5]首先提出“深度学习”概念开始,深度学习当前已经在学术界和工业界引起了广泛的关注。例如 Google 公司发布的 BERT 模型,在机器阅读理解水平测试数据集 SQuAD 上,测试结果全面超越人类,同时在 11 种不同的 NLP 测试中均取得当前最佳的性能表现。同时近期的 GPT-2、ERNIE 等,更是不断取得更好的表示性能。微软雷德蒙研究

院的俞栋博士及其合作者提出使用深层神经网络对数以千计的神经元直接建模,形成一个成功用于大词汇量的语音识别系统的上下文相关的深层神经网络-隐马尔可夫混合模型^[6].2012年,Krizhevsky等运用深度神经网络在ImageNet数据上取得85%的分类准确率,相比2011年的74%提高了11个百分点^[7].2016年微软亚洲研究院的何恺明等^[8]提出的深度残差网络在ImageNet上取得3.57%的错误率,超过了人类的识别水平.由上可见深度学习在图像、语音以及自然语言的数据特征学习及应用任务上都获得了显著的性能提升,取得极大成功.

1.1.2 多模态特征融合

在多媒体信息理解中,仅通过单一模态特征的简单叠加,往往难以取得很好的算法效果.一方面,多模态的特点给相关的研究带来很大的挑战.不同模态的数据服从不同的统计特性^[9].比如,文本常常被表示为单词的数目统计(word count),而图像则是被表示成像素或者一些计算机视觉学者们设计的特征,很难通过一种方法找到它们潜在的相关性.而另一方面,相对于单模态数据的语义理解研究,多模态数据之间的互补性,也为多媒体的内容理解提供了另一种途径.多模态的信息之间可以相互提供补充,使得某些单模态情形下难以理解的数据,在多模态下出现了新的希望.例如用户在Flick上传图片的同时,往往会添加自己的文本标签标注,这种图像与文本标签是强语义关联的,而对于新闻文档或者微信公众号上的文章,这些图像视频与周围环境文本具有较强的语义关联,但与其他位置的文本有时是不相关的.

针对以上特点,研究者们提出了跨模态学习模型,在有效考虑强弱语义关联的同时,将多模态的数据信息统一表示于单一向量空间之中.Frome等提出了相似性模型,目标是如何最小化协同空间中的模态间距离.举例来说,在相似性模型中,单词“dog”和狗的图片,要比车的图片距离更近^[10].Weston等^[11-12]提出了WSABIE(Web Scale Annotation By Image Embedding)模型,它为图像及其标注构建了一个联合空间,此模型构建了一个简单的从图像到文本特征的线性映射,从而使得图像表示和相关的标注区域更加相近.Kiros等^[13]通过使用LSTM模型和成对排序(pairwise ranking)损失协同特征空间,并将此扩展到句子和图像的协同表示.Socher等^[14]将语言模型扩展到依存树RNN上,以实现语义单元的整

合.Xu等^[15]使用主题、动词、宾语组合语言模型,同时引入深层视频信息,构建视频和句子之间的协同学习空间,进而将该表示用于跨模态检索和视频描述任务.Verdrov等^[16]和Zhang等^[17]提出了一种结构化协同表示模型,实现图像和语言的序列嵌入,在该模型中实施了一种非对称的差异性度量.

1.2 跨模态知识提取与表示

1.2.1 多模态知识提取

网络多媒体内容可以为大数据理解提供关键的数据支持,但如何理解网络多媒体内容数据从中提取知识元素,构成知识图谱结构,变成高度结构化的知识信息,仍面临着巨大的挑战.从提取内容上划分,可包括知识实体概念的提取、语义类提取、属性和属性值提取和关系提取等.在知识抽取技术的发展过程中,逐渐从基于规则的匹配方法,朝着基于深度学习的抽取方法转变,并逐渐落地应用.

文献[18]对近几年的自动和半自动的知识抽取方法做了较为详细的介绍,但方法主要集中在基于规则的模板匹配技术上,处理的数据对象多面向半结构化的百科类或垂直站点网站,难以满足非结构化数据的知识提取工作.社交媒体内容数据更多的是非结构类型数据,对其的知识提取是一项挑战性的任务.

1) 实体抽取部分.文献[19]将实体抽取的方法分为三种:基于规则与词典的方法、基于统计机器学习的方法以及面向开放域的抽取方法.基于规则的方法通常需要为目标实体编写模板,然后在原始语料中进行匹配^[20];基于统计机器学习的方法主要是通过机器学习的方法对原始语料进行训练^[21],然后再利用训练好的模型去识别实体.

2) 属性抽取.Google提出Knowledge Vault^[22],以一种概率性的知识融合方法来处理网络文本、HTML表格、标注等数据来抽取知识三元组以构建知识图谱.Carlson等^[23]提出了一个可以利用互联网Web文本信息,自动构建知识库的系统.文献[24]提出基于规则与启发式算法的属性抽取方法,能够从Wikipedia及WordNet的半结构化网页中自动抽取相应的属性名称与属性值.

3) 关系抽取.实体之间关系抽取解决实体间语义链接的问题,早期的关系抽取主要是通过人工构造语义规则以及模板的方法识别实体关系.随后,实体间的关系模型逐渐替代了人工预定义的语法与规则,但是仍需要提前定义实体间的关系类型.因此部

分研究者提出了基于马尔可夫逻辑网、基于本体推理的深层隐含关系抽取方法.文献[25]提出了一种无监督学习模型 StatSnowball,不同于传统的 OIE,该方法可自动产生或选择模板生成抽取器.相比大量的研究工作主要处理文本类型的数据,而对多模态的网络内容数据的语义理解工作仍比较有限. Yang 等^[26]设计了领域自适应算法,即在目标领域视频数量有限的情况下,如何从已有领域的模型获得目标领域的模型. Fang 等^[27]提出一种概率图模型来自动挖掘实体的多模态主题特征以及相应的用户观点. NEIL^[28]是一个不停学习的系统,自动从搜索引擎提取语义概念、关系以及图像实例来构建图像视觉知识库.中国科学院自动化研究所^[29]提出一种系统性的框架,利用社区用户产生标签图片数据来自动完成视觉知识图谱的构建,从图片分享网站 Flickr 收集了 2.4 亿张标签图像,构建框架包括三步:概念发现、概念关系提取、概念层级结构建立,构建的多模态知识库能很好地应用在图像识别以及检索上,并有显著的性能提升.

1.2.2 跨模态知识表示

传统的知识表示方法主要是以 RDF 三元组 SPO 来符号性描述实体之间的关系.这种表示方法通用简单,受到广泛认可,但是其在计算效率、数据稀疏性等方面面临诸多问题.近年来,以深度学习为代表的表示学习技术取得了重要的进展,可以将实体的语义信息表示为稠密低维实值向量,进而在低维空间中高效计算实体、关系及其之间的复杂语义关联,对知识库的构建、推理、融合以及应用均具有重要的意义^[30-33].知识表示学习的代表模型有翻译距离模型、关系矩阵语义匹配模型、神经网络语义匹配模型等几大类别.

1) 翻译距离模型通过建模关系变换将一个实体变换到另一个实体的表示空间,通过距离函数得到评分函数.代表方法包括 TransE^[33] 及其扩展方法 TransH^[34]、TransR^[35]、TransD^[36]、TranSparse^[37]、ManifoldE^[38]、TransG^[39]、KG2E^[40] 模型等.

2) 关系矩阵语义匹配模型通过矩阵分解关系约束相关方法匹配实体表示和关系表示的隐语义,其中的典型代表是文献[41]提出的 RESCAL,及其扩展方法 TATEC^[42]、DistMult^[43]、HoLE^[44]、ComplEx^[45]、ANALOGY^[46] 等.

3) 神经网络语义匹配模型,使用神经网络建模匹配实体表示和关系表示的隐语义,包括神经张量

模型 NTN^[47]、神经关联模型 NAM^[48] 等.

近年来,随着神经网络的发展,研究者们提出一些基于神经网络的方法,建模更加复杂的知识语义关系. Dettmers 等^[49]将实体和关系的语义匹配计算建模为一个二维卷积神经网络,得到的特征图通过全连接层变换为尾部实体表示的预测值. Schlichtkrull 等^[50]将图卷积神经网络扩展为可计算多关系卷积的方式,将知识图谱中更加丰富的邻域结构信息引入知识表示学习,在 DistMult 框架下计算实现. Guan 等^[51]提出了显式建模头实体预测任务和尾实体预测任务共享表示的神经网络,使用映射性质的自适应损失函数.

使用神经网络在多模态知识表示也做了一系列的研究和探索. Mousselly-Sergieh 等^[52]提出了翻译距离模型架构结合语言和图像的多模态表示方法,使用交叉表示的方式融合多模态特征. Pezeshkpour 等^[53]提出了一种神经网络编码模型,将实体文本和图像的信息编码到实体表示中,并建立了基于条件生成式对抗网络的解码模型,可以生成缺失的属性值以及实体相关的文本和图像. An 等^[54]提出了引入关系 mention 和实体描述信息,通过注意力机制增强实体表示和关系表示的学习,缓解实体和关系的多意性.

1.3 基于社交媒体行为的用户建模应用

基于社交媒体的用户建模指的是通过从用户在社会媒体行为中提取有效信息,以表示用户差异化的属性、偏好等.通过对用户在社会媒体上的行为内容进行精细化处理,如多模态特征融合、知识提取与表示等,有利于构建完整准确的用户画像,进而更好地分析和理解用户,为用户提供个性化的服务.下文我们将阐述几类基于用户社会媒体行为的典型的应用.

1.3.1 用户属性推断

如何更好地分析和理解用户,为用户提供个性化的信息服务,成为社交媒体的主要任务和挑战.用户人口统计属性,包括年龄、性别、婚姻状况和职业等,是理解和进行用户画像的基础.用户在社会媒体网络中产生的海量多媒体内容数据与丰富的用户行为信息,隐含地揭示了关于用户个人信息的重要线索,为解决社交网络中用户人口统计属性的缺失与稀疏问题提供了解决途径.社会媒体海量的数据为学者们提供了丰富的原材料,目前对用户属性推断的工作主要有对用户性别^[55-65]、年龄^[56,59-66]、政治倾

向^[60-62,65,67-68]、地理位置^[61,63,69-71]、种族^[59,65,68]、宗教信仰^[60,63]、职业^[63,72]、学历^[59,65]等方向.社交媒体行为信息如用户对话内容^[55-56]、博客文章内容^[55,58,66]、用户的搜索查询词^[59-60]、Twitter 上的推文内容^[61-62,67-68,70-71]、传记^[63,72]、头像图片^[64]、社交关系^[65]等都被应用于用户人口统计属性的研究中.

用户人口统计属性推断主要有两类方法:基于规则的和基于统计学习的.基于规则的方法主要思想是从文本中提取用户的人口统计属性信息.例如, Garera 等^[63]从传记文本信息中提取出传记事实,包括了生日、职业、国籍等属性.Zhou 等^[73]将鉴别和提取传记信息作为了一个摘要任务.Yu 等^[74]使用了一个串联的信息提取框架从简历中提取出用户的个人信息.Mann 等^[75]利用了上下文模式的学习方法提取出了具体事实,比如出生地等.Bergsma 等^[76]提出了使用概念类的属性预测出社交媒体用户的隐性传记属性.虽然基于规则或者模式匹配的方法能够有效地提取出用户人口统计属性的信息,但所学到的模板却受限于特定的人口统计属性的提取,缺乏扩展性和普适性.基于统计学习的属性推断方法被应用于解决这类问题.许多研究工作从社交媒体数据中提取出用户的特征并学习相应的模型来预测用户的人口统计属性.用户的人口统计属性在很大程度上会影响用户在社会媒体网络上的行为内容.文献^[77-78]分析了用户的个人信息和社会媒体行为的一致性,并统计性地证明了利用用户社会媒体行为数据进行人口统计属性推断的可行性.利用社会媒体行为信息提取用户属性特征,结合分类器可以进行简单有效的属性推断.Garera 等^[56]扩展了 N 元模型来提取语言学特征,并利用了线性支持向量机模型在对话和电子邮件中有效提取出用户的个人信息.Rao 等^[61]利用 Twitter 上的用户行为数据进行了用户的人口统计属性推断.van Durme^[79]提出了一个流式的框架推断出参与者的属性信息.Pennacchiotti 等^[68]试图通过聚合用户人口统计属性来对用户进行分类.另外, Xiang 等^[80]在现有的用户人口统计属性的研究基础上,利用用户人口统计属性的关联性和稳定性,对用户人口统计属性推断工作进行了进一步的深入研究.

1.3.2 用户关系标注

社交媒体中的社会关系作为一种渠道,在信息传播中起着重要的作用.有效的社会关系可以帮助用户对信息爆炸的社交媒体网络上的内容取其精华

去其糟粕.当前的用户关系标注问题主要基于从多个通信通道的交互数据中提取的特征或对关系的描述.Gilbert 等^[81]确定了 74 个 Facebook 变量作为潜在的预测因素,并将社交媒体的联系映射为强弱关系.Tang 等^[82]基于某些社交网络属性和有限的预先已知信息来识别社交网络用户之间的关系,并且通过选择用于营销的少量种子来最大化传播.Sun 等^[83]提出利用智能手机独特的多模式交互数据,在电话、物理位置/邻近、电子邮件和在线社交网络四种通信渠道中对社会关系进行分类.Yang 等^[84]同时捕捉用户的行为、社交互动以及两者之间的关联,将社交网络图的边缘标记为正相关或负相关关系.He 等^[85]提出了一种基于决策树的随机游走模型,该模型不仅考虑了全局网络结构,而且很好地利用了局部用户行为,从而识别了多个异构社会网络之间的关系类型.此外,大多数现有作品总是根据不同的规则将每个用户的用户关系划分为固定的类型,例如关系类型(包括家庭、工作和朋友等)^[82-83,85]、关系程度(包括强、弱等)^[81,84,86].

以 Twitter 平台为例, Twitter 的列表功能于 2009 年 11 月向 Twitter 公众公布.由于 Twitter 列表中的元数据中蕴含着丰富而有价值的语义线索,它引起了越来越多的研究者的关注.Kim 等^[87]使用 Twitter 列表来推断用户的特性,尤其是关于其兴趣的特性,并确认在用户感知特征方面,列表是 Twitter 用户的良好分组.Yamaguchi 等^[88]建议从列表名称中提取标记(tags),通过使用 Twitter 列表为用户发现合适的主题.Ghosh 等^[89]挖掘 Twitter 列表信息,以建立在 Twitter 中查找专题专家的系统,并强调 Twitter 列表是 Twitter 中未来内容或专家搜索系统潜在的宝贵信息来源.Rakesh 等^[90]提出了一种新的框架,用于推荐比用户目前订阅的清单更受欢迎的辅助名单,方法是合并几个共同反映用户个人兴趣的特征.Huang 等^[91]从用户自身出发,根据用户之间的相似性来提取特征,从 Twitter 列表中提取丰富的多模态行为和社会交互来进行个性化的社交关系标注.

1.3.3 个性化推荐

随着信息的爆炸式增长,用户容易迷失在系统提供的大量商品中.推荐系统(Recommendation System, RS)用作克服这种信息过载的通用解决方案,旨在从压倒性的在线内容和服务(例如电影、新闻和产品)中找到一组相关商品以满足用户的个人兴趣.个性化推荐基于社交媒体上用户与项目的历

史交互,以及用户和项目的属性信息来评估用户对项目的偏好。

1) 传统推荐方法

传统的推荐策略通常分为三类^[92]:协同过滤(CF)、基于内容的推荐策略和混合推荐策略。本节主要介绍基于协同过滤的推荐方法。CF通过从用户-项目历史交互、显式反馈(例如评论和打分)或隐式反馈(例如浏览和点击)来探索对目标项目的用户偏好。基于记忆的CF方法^[92-94]直接从用户项目交互记录驱动用户和项目相似性矩阵,并且因此在用户和目标项目之间产生估计分数。矩阵因式分解(MF)是最简单、最有效的潜在因子模型之一,它用一个潜在的向量来描述一个用户,将用户显式反馈(如打分)建模为其潜在向量的内积。在MF的启发下,研究者们提出了许多变体,如SVD^[95]、因子分解机(FM)^[96]、Localized MF^[97]、Social MF^[98]等模型。FM目前取得了比较好的效果,因为它可以建模任意数量实体之间的特征交互,而MF只对用户和项目之间的交互进行建模。随着神经网络模型的广泛应用,近年来基于深度学习的用户建模研究开始高速发展。深度学习在图像、文本领域取得的巨大成功表明深度学习模型可以很好地学习隐含的特征表达,挖掘知识表示的深层关联,进而有利于辅助基于社交媒体行为的用户属性/偏好挖掘,从而在用户建模任务中取得较好的效果。从多层感知机(MLP)和自编码器(AE)^[99]到卷积神经网络(CNN)和递归神经网络(RNN),各种类型的深度学习技术都应用广泛。现有的基于深度学习技术的研究丰富了图像和文本等辅助数据中用户和项目的表示方式^[100-101],增强了特征交互功能^[102-103]。文献[102]中提出了Wide&Deep方法,其中Wide部分学习一些明确的特征依赖关系,Deep部分在特征嵌入向量的级联上采用MLP来揭示隐式特征之间的相互作用。深度交叉(deep crossing)^[104]将一个MLP和多个残差单元叠加在单个特征的嵌入上,以自动生成组合特征。神经协同过滤(NCF)^[103]利用前馈神经网络代替MF的内积并对用户-项目交互函数进行参数化,是一种通用的深度推荐解决方案。最近,作为FM的扩展,神经因子分解机(NFM)^[105]提出了一种双线性交互池化操作,在用户和项目的嵌入向量上采用元素级乘积,然后叠加MLP来捕获用户和项目之间的非线性关系。以上所述这些用户建模的研究工作都取得了不错的效果,但往往都忽略了一个问题,即用户行为是具有动

态特性的,用户在社会媒体网络上的行为随着时间的推移会发生变化。但这种行为变化并不是随机产生的,而是具有很强的时序依赖性的,即用户在时间节点前的行为信息对时间节点上的行为会产生较大的影响。为了解决这个问题,许多研究者开始着手于构建用户动态兴趣模型的工作。由于用户行为的时序性,用户的动态兴趣变化通常通过建模用户的序列行为来完成,因此序列推荐任务对捕捉用户动态偏好,全面理解用户兴趣方面有重要意义。

2) 序列推荐

序列推荐问题通常被作为序列预测问题来解决。大多数现有的方法都集中在基于马尔可夫链(MC)的方法和基于神经网络的方法上。可扩展的序列模型通常依赖MC来捕获序列模式^[106-107],其中 L 阶马尔可夫链根据以前的 L 个行为作出推荐。然而,基于MC的模型的一个主要问题是,所有的成分都是独立组合的,表示它在多个因素^[108]之间做出了强烈的独立性假设。基于矩阵分解(MF)的方法同样可以用于序列行为建模,MF将从当前项到下一项的转移概率矩阵分解为潜在因子^[109]。然而,由于现实世界中的数据通常服从幂律分布,MF备受稀疏问题的困扰^[110]。在矩阵分解的强大力量的启发下,分解个性化马尔可夫链(FPMC)^[107]结合MF和MC的能力,对底层MC上的转移矩阵进行分解,为推荐建立个性化的序列行为模型。FPMC及其变体^[111]通过将转移矩阵分解为两个潜在的和低秩的子矩阵来改进该方法。最近,递归神经网络(RNN)方法在序列建模方面取得了很大的成功^[112]。它已经被成功地应用于诸如句子建模任务^[112-114]、视频建模^[115]、序列点击预测^[116]、多行为序列预测^[117]和位置预测^[118]等多个应用。虽然它是编码用户上下文的一种非常有效的方法,但它仍然有一些难以克服的缺点,例如难以并行化、耗时、难以保持长期依赖关系等。最近,基于卷积神经网络(CNN)的编码方法在许多序列处理任务中也取得了与RNN相当的性能^[119-120]。要了解项目和上下文的相关性是很有挑战性的。引入注意力机制^[121],提供在解码器中动态引用特定行为记录的能力,近年来在阅读理解^[122-123]、广告推荐^[124-125]和计算机视觉^[126]等方面已经取得了很大的成功。自注意力模型(self-attention)研究了编码端项目之间的内在联系^[122-123,127]。ATRank^[128]模型仅基于自注意力网络,通过将所有类型的行为投影到多个潜在的语义空间,从而提升推荐的效果。另外,Huang等^[129]

提出了一种完全基于自注意力的用户行为建模框架,其自注意力机制在特征层上工作,同时模型中使用位置编码矩阵对动态上下文依赖进行建模,在提高了推荐准确率的同时也加快了模型的训练速度.

3) 可解释性推荐

可解释的推荐算法旨在解决这样的问题:不仅向用户提供合适的建议,而且为用户解释为什么系统推荐这些项目^[130].由于知识图谱中包含了丰富的用户和项目的外部结构信息,近年来知识图谱被广泛地应用于可解释性推荐任务中.将知识嵌入到推荐系统中的方法大致可以分为两类:基于知识图谱嵌入(Knowledge Graph Embedding, KGE)的方法和基于路径的方法.基于KGE的方法通常将项目本身的内容表示与知识感知嵌入结合起来,以便为项目生成更好的表示^[131-133].这些方法的缺点是,尽管推荐的准确性可以提高,但很难解释为什么将该项目推荐给用户,因为引入的知识图谱嵌入表示是隐式的.对路径连接实体语义关系的忽视导致了推理能力的缺乏.因此,许多研究将基于路径的实体相似度扩展到一般推荐模型,这些路径通过知识图谱中不同的语义来表示两个实体之间的联系.元路径(meta-path)是图中连接对象对的一种典型方式,它是一种关系序列,广泛用于提取结构特征,为推荐捕获相关的语义^[134].以前的一些工作已经将连接模式引入到推荐系统中^[134-140].但是基于元路径的方法在很大程度上依赖于手工构建的特性和所选元路径的质量,这就要求研究人员需要一定的领域知识.最新的一种方法是通过模型自动捕获语义关联.知识图谱的实体对之间的合格路径被自动挖掘,然后通过递归网络进行编码.在网络末端无缝集成推荐层,该推荐层可以以端到端的方式进行训练,以将语义结构知识结合到推荐任务中^[141-142].

2 发展趋势

移动互联网时代尽管已经发展了很多年,但随着通信技术的迅猛发展,比如即将到来的5G时代,社交媒体必定会进入另一个更活跃更丰富多彩的新时期.用户的行为逐渐从文本时代过渡到图片时代,进入到高速网络下的视频时代,未来更有可能进入全民直播时代等.可以预见社交媒体的功能、品类、关系网络只会越来越复杂.因此,如何解决与日俱增的庞大多模态数据与用户越来越独特鲜明的个性化需求之间的矛盾,是未来基于社会多媒体内容用户

建模应用的关键研究和发展趋势.总结来说,当前基于社会多媒体内容的用户建模应用研究领域有以下几个发展趋势:

1) 跨模态社交媒体内容的知识表征

跨模态数据的异构复杂性和演化动态性使得传统的数据表达和融合方法难以实现跨模态大数据的鲁棒、高效、有判别力的特征表达,以及知识的有效组织与融合.未经组织和处理的文本、声音、图像和视频等数据价值密度较低,而知识是一种包含了结构化的经验、关联信息、定量关系以及专家见解等要素的动态组合,是更为抽象的本质的描述.因此,如何从复杂的跨模态数据中提取有价值信息,全面构建知识,实现知识的可表征是多媒体内容理解的一个关键问题.

2) 基于知识图谱的社交媒体深度分析与决策

社交媒体深度分析与决策的核心是将数据抽象为可表征和可推理的知识,从而跨越数据到决策的鸿沟,为后续的用户建模等应用提供知识层面的辅助决策.然而,近年来社会多媒体内容理解的研究热点主要集中于数据层面,即对社交媒体网络上的多媒体大数据进行关联挖掘,利用数据之间的关联关系发现事物发展的潜在规律,进而进行统计推理与预测.然而数据不代表知识,数据是未经组织和处理的文本、声音、图像和视频等,价值密度较低,而知识是一种包含了结构化的经验、价值观、关联信息以及专家见解等要素的动态组合,比数据更有价值是因为它更贴近行动.但当前主要有两个挑战:一是知识的动态关联和增强;二是知识具有不确定性.如何解决这些难题,综合基于知识图谱的社交媒体深度分析与决策实现多媒体应用分析还需要研究者足够的重视.

3) 基于社交媒体内容的用户个性化服务

在用户属性推断应用方面,目前研究大多集中在人口属性推断.但用户物理世界的属性包含很多维度,人口属性推断只是用户物理世界的属性中很小的一部分.目前基于社交多媒体行为的物理世界用户属性研究尚处于起步阶段,有很多属性并未被深入研究.因此,未来的工作应考虑利用社交多媒体行为进行更多的用户物理世界属性的研究,例如用户的心理属性、真实社会关系等.用户关系标注方面,目前的工作大多基于已标注的数据集上进行研究,标注数据获取困难,人工标注也需要耗费大量的人力物力.未来的发展趋势应该借助各类无标注的

样本进行无监督学习,提高模型的可用程度.个性化推荐方面,利用辅助信息是推荐任务中的一项重要任务.未来的工作可以从两个方向进行扩展:一是信息来源方向,不仅考虑单一的用户属性或单模态特征,进一步可以通过考虑更复杂的多模态融合技术,另外引入复杂关系网络或外部知识图谱来作为上下文新辅助进行用户建模和个性化推荐;另一个是模型算法方向,当前的可解释性推荐的相关工作主要集中在考虑用户静态行为的 top-K 推荐任务中,如上文所述,用户行为在动态变化中,如何通过用户的时序行为捕捉用户动态偏好是一个重要的研究方向.未来在可解释性推荐工作的道路上,可解释性序列推荐必定是个值得深入研究的有意义的方向.

3 总结与展望

本综述围绕社会多媒体特征融合、跨模态知识提取与表示,以及基于社会媒体的用户建模相关应用研究三个方面介绍国内外在社会多媒体分析和应用领域的研究进展,总结国内外现有的多媒体内容理解方法和用户建模应用技术,分析国际学科发展趋势.总之,近年来国内外学者在多媒体内容分析和用户建模的三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内容分析其他领域的探索,如跨模态社会媒体内容的知识表征、社会媒体深度分析与决策以及可解释用户个性化服务等.另一方面,在移动互联网、大数据、社交媒体背景下,仍需要在方法创新以及原创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] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality [J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119
- [2] Torralba A, Murphy K P, Freeman W T, et al. Context-based vision system for place and object recognition [C] // Proceedings Ninth IEEE International Conference on Computer Vision, 2003, DOI: 10.1109/ICCV.2003.1238354
- [3] Lowe D G. Object recognition from local scale-invariant features [C] // IEEE International Conference on Computer Vision (ICCV), 1999, DOI: 10.1109/ICCV.1999.790410
- [4] Davis S, Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1980, 28 (4): 357-366
- [5] Hinton G E. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313 (5786): 504-507
- [6] Dahl G E, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20 (1): 30-42
- [7] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C] // Proceedings of the 25th International Conference on Neural Information Processing Systems (Volume 1), 2012: 1097-1105
- [8] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778
- [9] 古曰国. 多媒体、多模态学习剖析 [J]. 外语电化教学, 2007 (114): 3-12
GU Yueguo. On multimedia learning and multimodal learning [J]. Technology Enhanced Foreign Language Education, 2007 (114): 3-12
- [10] Frome A, Corrado G S, Shlens J, et al. DeViSE: a deep visual-semantic embedding model [C] // Proceedings of the 26th International Conference on Neural Information Processing Systems, 2013: 2121-2129
- [11] Weston J, Bengio S, Usunier N. Largescale image annotation: learning to rank with joint word-image embeddings [J]. Machine Learning, 2010, 81 (1): 21-35
- [12] Weston J, Bengio S, Usunier N. WSABIE: scaling up to large vocabulary image annotation [C] // Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (Volume 3), 2011: 2764-2770
- [13] Kiros R, Salakhutdinov R, Zemel R S. Unifying visual-semantic embeddings with multimodal neural language models [J]. arXiv preprint, arXiv: 1411. 2539, 2014
- [14] Socher R, Karpathy A, Le Q V, et al. Grounded compositional semantics for finding and describing images with sentences [J]. Transactions of the Association for Computational Linguistics, 2014, 2: 207-218
- [15] Xu R, Xiong C M, Chen W, et al. Jointly modeling deep video and compositional text to bridge vision and language in a unified framework [C] // AAAI Publications, Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015
- [16] Vendrov I, Kiros R, Fidler S, et al. Order-embeddings of images and language [J]. arXiv preprint, arXiv: 1511. 06361, 2015
- [17] Zhang H, Hu Z T, Deng Y T, et al. Learning concept taxonomies from multi-modal data [C] // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1), 2016, DOI: 10.18653/v1/P16-1169
- [18] 史树明. 自动和半自动知识提取 [J]. 中国计算机学会通讯, 2013, 9 (8): 65-73
SHI Shuming. Automatic and semi-automatic knowledge extraction [J]. China Computer Federation

- Communication, 2013, 9(8):65-73
- [19] 孙镇,王惠临.命名实体识别研究进展综述[J].现代图书情报技术,2010(6):42-47
SUN Zhen, WANG Huilin. Overview on the advance of the research on named entity recognition[J]. New Technology of Library and Information Service, 2010(6):42-47
- [20] Chinchor N, Marsh E. Muc-7 information extraction task definition[C] // Proc of the 7th Message Understanding Conf. Philadelphia: Linguistic Data Consortium, 1998: 359-367
- [21] Liu X H, Zhang S D, Wei F R, et al. Recognizing named entities in tweets[C] // Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (Volume 1), 2011: 359-367
- [22] Dong X, Gabrilovich E, Heitz G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion[C] // Proceedings of the 20th ACM SIGKDD, 2014:601-610
- [23] Carlson A, Betteridge J, Kisiel B, et al. Toward an architecture for never-ending language learning[C] // Twenty-fourth AAAI Conference on Artificial Intelligence, AAAI Press, 2010
- [24] 杨博,蔡东风,杨华.开放式信息抽取研究进展[J].中文信息学报,2014,28(4):1-11,36
YANG Bo, CAI Dongfeng, YANG Hua. Progress in open information extraction[J]. Journal of Chinese Information Processing, 2014, 28(4):1-11, 36
- [25] Zhu J, Nie Z Q, Liu X J, et al. StatSnowball: a statistical approach to extracting entity relationships[C] // Proceedings of the 18th International Conference on World Wide Web, 2009:101-110
- [26] Yang J, Yan R, Hauptmann A G. Cross-domain video concept detection using adaptive SVMs[C] // Proceedings of the 15th International Conference on Multimedia, 2007, DOI: 10. 1145/1291233. 1291276
- [27] 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
- [28] Chen X, Shrivastava A, Gupta A. NEIL: extracting visual knowledge from web data[C] // Proceedings IEEE International Conference on Computer Vision, 2013: 1409-1416
- [29] Fang Q, Xu C S, Sang J T, et al. Folksonomy-based visual ontology construction and its applications [J]. IEEE Transactions on Multimedia, 2016, 18(4):702-713
- [30] Bengio Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127
- [31] 刘知远,孙茂松,林衍凯,等.知识表示学习研究进展[J].计算机研究与发展,2016,53(2):247-261
LIU Zhiyuan, SUN Maosong, LIN Yankai, et al. Knowledge representation learning: a review[J]. Journal of Computer Research and Development, 2016, 53(2): 247-261
- [32] 徐增林,盛泳潘,贺丽荣,等.知识图谱技术综述[J].电子科技大学学报,2016,45(4):589-606
XU Zenglin, SHENG Yongpan, HE Lirong, et al. Review on knowledge graph techniques[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4):589-606
- [33] Wang Q, Mao Z D, Wang B, et al. Knowledge graph embedding: a survey of approaches and applications [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12):2724-2743
- [34] Wang Z, Zhang J W, Feng J L, et al. Knowledge graph embedding by translating on hyperplanes[C] // Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014:1112-1119
- [35] Lin Y K, Liu Z Y, Sun M S, et al. Learning entity and relation embeddings for knowledge graph completion[C] // Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015:2181-2187
- [36] Ji G L, He S Z, Xu L H, et al. Knowledge graph embedding via dynamic mapping matrix [C] // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015:687-696
- [37] Ji G L, Liu K, He S Z, et al. Knowledge graph completion with adaptive sparse transfer matrix[C] // Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016:985-991
- [38] Xiao H, Huang M L, Zhu X Y. From one point to a manifold: knowledge graph embedding for precise link prediction[C] // Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016: 1315-1321
- [39] Xiao H, Huang M L, Hao Y, et al. TransG: a generative mixture model for knowledge graph embedding[J]. arXiv preprint, arXiv:1509. 05488, 2015
- [40] He S Z, Liu K, Ji G L, et al. Learning to represent knowledge graphs with Gaussian embedding[C] // Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015, DOI: 10. 1145/2806416. 2806502
- [41] Nickel M, Tresp V, Kriegel H P. A three-way model for collective learning on multi-relational data[C] // International Conference on International Conference on Machine Learning, 2011:809-816
- [42] Garcia-Durán A, Bordes A, Usunier N. Effective blending of two and three-way interactions for modeling multi-relational data[M] // Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014:434-449
- [43] Yang B, Yih W, He X, et al. Embedding entities and relations for learning and inference in knowledge bases[J]. arXiv preprint, arXiv:1412. 6575, 2014
- [44] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs[C] // Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016: 1955-1961
- [45] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction[C] // International Conference on Machine Learning, 2016:2071-2080
- [46] Liu H X, Wu Y X, Yang Y M. Analogical inference for

- multi-relational embeddings [C] // Proceedings of the 34th International Conference on Machine Learning(Volume 70),2017:2168-2178
- [47] Socher R, Chen D Q, Manning C, et al. Reasoning with neural tensor networks for knowledge base completion [C] // Proceedings of the 26th International Conference on Neural Information Processing Systems (Volume 1), 2013:926-934
- [48] Liu Q, Jiang H, Evdokimov A, et al. Probabilistic reasoning via deep learning; neural association models [J]. arXiv preprint, arXiv:1603.07704, 2016
- [49] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings [J]. arXiv preprint, arXiv:1707.01476, 2018
- [50] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks [M] // The Semantic Web. Cham: Springer International Publishing, 2018:593-607
- [51] Guan S P, Jin X L, Wang Y Z, et al. Shared embedding based neural networks for knowledge graph completion [C] // Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, DOI:10.1145/3269206.3271704
- [52] Mousselly-Sergieh H, Botschen T, Gurevych I, et al. A multimodal translation-based approach for knowledge graph representation learning [C] // Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, 2018:225-234
- [53] Pezeshkpour P, Chen L Y, Singh S. Embedding multimodal relational data for knowledge base completion [C] // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, DOI: 10.18653/v1/D18-1359
- [54] An B, Chen B, Han X P, et al. Accurate text-enhanced knowledge graph representation learning [C] // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies, Volume 1 (Long Papers), 2018, DOI:10.18653/v1/N18-1068
- [55] Boulis C, Ostendorf M. A quantitative analysis of lexical differences between genders in telephone conversations [C] // Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005:435-442
- [56] Garera N, Yarowsky D. Modeling latent biographic attributes in conversational genres [C] // Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Volume 2, 2009:710-718
- [57] Herring S C, Paolillo J C. Gender and genre variation in weblogs [J]. Journal of Sociolinguistics, 2006, 10 (4) : 439-459
- [58] Sarawgi R, Gajulapalli K, Choi Y J. Gender attribution: tracing stylometric evidence beyond topic and genre [C] // Proceedings of the Fifteenth Conference on Computational Natural Language Learning, 2011:78-86
- [59] Weber I, Castillo C. The demographics of web search [C] // Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2010, DOI:10.1145/1835449.1835537
- [60] Bi B, Shokouhi M, Kosinski M, et al. Inferring the demographics of search users; social data meets search queries [C] // Proceedings of the 22nd international conference on World Wide Web, 2013:131-140
- [61] Rao D, Yarowsky D, Shreevats A, et al. Classifying latent user attributes in twitter [C] // Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents, 2010, DOI:10.1145/1871985.1871993
- [62] Zamal F A, Liu W, Ruths D. Homophily and latent attribute inference; inferring latent attributes of twitter users from neighbors [C] // Proceedings of the International AAAI Conference on Web and Social Media, 2012:387-390
- [63] Garera N, Yarowsky D. Structural, transitive and latent models for biographic fact extraction [C] // Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, 2009, DOI: 10.3115/1609067.1609100
- [64] Liu L, Preotiuc-Pietro D, Samani Z R, et al. Analyzing personality through social media profile picture choice [C] // Proceedings of the International AAAI Conference on Web and Social Media, 2016:211-220
- [65] Culotta A, Ravi N K, Cutler J. Predicting twitter user demographics using distant supervision from website traffic data [J]. Journal of Artificial Intelligence Research, 2016, 55:389-408
- [66] Burger J D, Henderson J C. An exploration of observable features related to blogger age [C] // AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, 2006:15-20
- [67] Conover M D, Goncalves B, Ratkiewicz J, et al. Predicting the political alignment of twitter users [C] // Passat/socialcom, Privacy, Security, Risk and Trust, 2011:192-199
- [68] Pennacchiotti M, Popescu A-M. Democrats, republicans and starbucks aficionados; user classification in twitter [C] // Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011, DOI:10.1145/2020408.2020477
- [69] Cheng Z Y, Caverlee J, Lee K. You are where you tweet: a content-based approach to geo-locating twitter users [C] // Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010:759-768
- [70] Eisenstein J, O'Connor B, Smith N A, et al. A latent variable model for geographic lexical variation [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2010:1277-1287
- [71] Kulkarni V, Perozzi B, Skiena S. Freshman or fresher? Quantifying the geographic variation of language in online social media [C] // Proceedings of the International AAAI Conference on Web and Social Media, 2016:615-618
- [72] Filatova E, Prager J. Occupation inference through detection and classification of biographical activities [J]. Data & Knowledge Engineering, 2012, 76/77/78:39-57
- [73] Zhou L, Ticea M, Hovy E. Multi-document biography

- summarization [C] // Proceedings of Empirical Methods Natural Language Process, 2004:434-441
- [74] Yu K, Guan G, Zhou M. Resume information extraction with cascaded hybrid model [C] // Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005:499-506
- [75] Mann G S, Yarowsky D. Multi-field information extraction and cross-document fusion [C] // Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, 2005:483-490
- [76] Bergsma S, Durme B V. Using conceptual class attributes to characterize social media users [C] // Meeting of the Association for Computational Linguistics, 2013:710-720
- [77] Zheleva E, Getoor L S. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles [C] // Proceedings of the 18th International Conference on World Wide Web, 2009:531-540
- [78] Mislove A, Viswanath B, Gummadi K P, et al. You are who you know: inferring user profiles in online social networks [C] // Proceedings of the Third ACM International Conference on Web Search and Data Mining, 2010:251-260
- [79] van Durme B. Streaming analysis of discourse participants [C] // Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012:48-58
- [80] Xiang L C, Sang J T, Xu C S. Demographic attribute inference from social multimedia behaviors: a cross-OSN approach [M] // MultiMedia Modeling. Cham: Springer International Publishing, 2016:515-526
- [81] Gilbert E, Karahalios K. Predicting tie strength with social media [C] // Proceedings of the 27th International Conference on Human Factors in Computing Systems, 2009:211-220
- [82] Tang S J, Yuan J, Mao X F, et al. Relationship classification in large scale online social networks and its impact on information propagation [C] // 2011 Proceedings IEEE INFOCOM, 2011:2291-2299
- [83] Sun D Y, Lau W C. Social relationship classification based on interaction data from smartphones [C] // 2013 IEEE International Conference on Pervasive Computing and Communications Workshops, 2013:205-210
- [84] Yang S H, Smola A J, Long B, et al. Friend or frenemy? Predicting signed ties in social networks [J]. Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2012:555-564
- [85] He J N, Liu H Y, Lau R Y K, et al. Relationship identification across heterogeneous online social networks [J]. Computational Intelligence, 2017, 33(3):448-477
- [86] Zein M, Adl A, Hassanien A E, et al. A social relationship modifiers modeller [C] // 2015 3rd International Conference on Computer, Information and Application, 2015:33-37
- [87] Kim D, Jo Y, Moon I C, et al. Analysis of twitter lists as a potential source for discovering latent characteristics of users [C] // ACM CHI Workshop on Microblogging, 2010
- [88] Yamaguchi Y, Amagasa T, Kitagawa H. Tag-based user topic discovery using twitter lists [C] // 2011 International Conference on Advances in Social Networks Analysis and Mining, 2011, DOI: 10.1109/ASONAM.2011.58
- [89] Ghosh S, Sharma N, Benevenuto F, et al. Cognos: crowdsourcing search for topic experts in microblogs [C] // Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2012:575-590
- [90] Rakesh V, Singh D, Vinzamuri B, et al. Personalized recommendation of twitter lists using content and network information [C] // The International AAAI Conference on Web and Social Media (ICWSM), 2014
- [91] Huang X W, Xu C S, Xiang L C, et al. Social relationship labeling based on multimodal behaviors and social interactions [J]. IEEE Multimedia, 2018, 25(4):68-78
- [92] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions [J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6):734-749
- [93] Linden G, Smith B, York J. Amazon. com recommendations: item-to-item collaborative filtering [J]. IEEE Internet Computing, 2003, 7(1):76-80
- [94] Desrosiers C, Karypis G. A comprehensive survey of neighborhood-based recommendation methods [M] // Recommender Systems Handbook. Boston, MA: Springer US, 2010:107-144
- [95] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model [C] // Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008:426-434
- [96] Rendle S. Factorization machines [C] // 2010 IEEE International Conference on Data Mining, 2010:995-1000
- [97] Zhang Y F, Zhang M, Liu Y Q, et al. Localized matrix factorization for recommendation based on matrix block diagonal forms [C] // Proceedings of the 22nd international conference on World Wide Web, 2013, DOI: 10.1145/2488388.2488520
- [98] Zhao Z, Lu H Q, Cai D, et al. User preference learning for online social recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(9):2522-2534
- [99] Liang D, Krishnan R G, Hoffman M D, et al. Variational autoencoders for collaborative filtering [C] // Proceedings of the 2018 World Wide Web Conference on World Wide Web, 2018:689-698
- [100] Zhang F, Yuan N J, Lian D, et al. Collaborative knowledge base embedding for recommender systems [C] // The 22nd ACM SIGKDD International Conference, 2016:353-362
- [101] Wang S, Wang Y, Tang J, et al. What your images reveal: exploiting visual contents for point-of-interest recommendation [C] // The 26th International Conference, 2017:391-400
- [102] Cheng H, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems [C] // Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, 2016:7-10

- [103] He X, Liao L, Zhang H, et al. Neural collaborative filtering [C] // Proceedings of the 26th International Conference on World Wide Web, 2017: 173-182
- [104] Shan Y, Hoens T R, Jiao J, et al. Deep crossing: web-scale modeling without manually crafted combinatorial features [C] // ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2016, DOI: 10.1145/2939672.2939704
- [105] He X N, Chua T S. Neural factorization machines for sparse predictive analytics [C] // Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, DOI: 10.1145/3077136.3080777
- [106] Feng S S, Li X T, Zeng Y F, et al. Personalized ranking metric embedding for next new POI recommendation [C] // International Joint Conference on Artificial Intelligence, 2015: 2069-2075
- [107] Rendle S, Freudenthaler C, Schmidt-Thieme L. Factorizing personalized Markov chains for next-basket recommendation [C] // Proceedings of the 19th International Conference on World Wide Web, 2010: 811-820
- [108] Wang P F, Guo J F, Lan Y Y, et al. Learning hierarchical representation model for next basket recommendation [C] // Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2015: 403-412
- [109] Chou S Y, Yang Y H, Jang J S R, et al. Addressing cold start for next-song recommendation [C] // Proceedings of the 10th ACM Conference on Recommender Systems, 2016: 115-118
- [110] Hu L, Cao L B, Cao J, et al. Learning informative priors from heterogeneous domains to improve recommendation in cold-start user domains [J]. ACM Transactions on Information Systems, 2016, 35(2): 1-37
- [111] Cheng C, Yang H, Lyu M R, et al. Where you like to go next: successive point-of-interest recommendation [C] // International Joint Conference on Artificial Intelligence, 2013: 2605-2611
- [112] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model [C] // Eleventh Annual Conference of the International Speech Communication Association, 2010: 1045-1048
- [113] Mikolov T, Kombrink S, Burget L, et al. Extensions of recurrent neural network language model [C] // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011: 5528-5531
- [114] Mikolov T, Kombrink S, Deoras A, et al. RNNLM-recurrent neural network language modeling toolkit [C] // Proceedings of the 2011 ASRU Workshop, 2011: 196-201
- [115] Du Y, Wang W, Wang L. Hierarchical recurrent neural network for skeleton based action recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1110-1118
- [116] Zhang Y Y, Dai H J, Xu C S, et al. Sequential click prediction for sponsored search with recurrent neural network [C] // Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014, 1369-1375
- [117] Liu Q, Wu S, Wang L. Multi-behavioral sequential prediction with recurrent log-bilinear model [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(6): 1254-1267
- [118] Liu Q, Wu S, Wang L, et al. Predicting the next location: a recurrent model with spatial and temporal contexts [C] // Thirtieth AAAI Conference on Artificial Intelligence, 2016: 194-200
- [119] Yoon K. Convolutional neural networks for sentence classification [J]. arXiv preprint, arXiv: 1408.5882, 2014
- [120] Tang J X, Wang K. Personalized top-N sequential recommendation via convolutional sequence embedding [C] // Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018: 565-573
- [121] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. arXiv preprint, arXiv: 1409.0473v2, 2014
- [122] Cui Y M, Chen Z P, Wei S, et al. Attention-over-attention neural networks for reading comprehension [J]. arXiv preprint, arXiv: 1607.04423, 2016
- [123] Lin Z H, Feng M W, Santos C N D, et al. A structured self-attentive sentence embedding [J]. arXiv preprint, arXiv: 1703.03130, 2017
- [124] Zhai S F, Chang K H, Zhang R F, et al. DeepIntent: learning attentions for online advertising with recurrent neural networks [C] // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 1295-1304
- [125] Zhou G R, Gai K, Zhu X Q, et al. Deep interest network for click-through rate prediction [C] // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018: 1059-1068
- [126] Xu K, Ba J, Kiros R, et al. Show, attend and tell: neural image caption generation with visual attention [C] // International Conference on Machine Learning, 2015: 2048-2057
- [127] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C] // Advances in Neural Information Processing Systems, 2017: 6000-6010
- [128] Zhou C, Bai J Z, Song J S, et al. ATRank: an attention-based user behavior modeling framework for recommendation [C] // Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018
- [129] Huang X, Qian S, Fang Q, et al. CSAN: contextual self-attention network for user sequential recommendation [C] // 2018 ACM Multimedia Conference on Multimedia Conference, 2018: 447-455
- [130] Zhang Y F, Chen X. Explainable recommendation: a survey and new perspectives [J]. arXiv preprint, arXiv: 1804.11192, 2018
- [131] Huang J, Zhao W X, Dou H J, et al. Improving sequential recommendation with knowledge-enhanced memory networks [C] // The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018: 505-514
- [132] Wang H W, Zhang F Z, Xie X, et al. DKN: deep knowledge-aware network for news recommendation [C] // Proceedings of the 2018 World Wide Web Conference on

- World Wide Web, 2018; 1835-1844
- [133] Zhang F Z, Yuan N J, Lian D F, et al. Collaborative knowledge base embedding for recommender systems [C] // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, DOI: 10.1145/2939672.2939673
- [134] Sun Y Z, Han J W, Yan X F, et al. PathSim: meta path-based top-*k* similarity search in heterogeneous information networks [C] // Proceedings of the VLDB Endowment 4, 2011: 992-1003
- [135] Gao L, Yang H, Wu J, et al. Recommendation with multi-source heterogeneous information [C] // Twenty-Seventh International Joint Conference on Artificial Intelligence, 2018, DOI: 10.24963/ijcai.2018/469
- [136] Heitmann B, Hayes C. Using linked data to build open, collaborative recommender systems [C] // The 2010 AAAI Spring Symposium, 2010
- [137] Hu B B, Shi C, Zhao W X, et al. Leveraging meta-path based context for top-*N* recommendation with a neural co-attention model [C] // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018; 1531-1540
- [138] Shi C, Zhang Z Q, Luo P, et al. Semantic path based personalized recommendation on weighted heterogeneous information networks [C] // Proceedings of the 24th ACM International Conference on Information and Knowledge Management, 2015; 453-462
- [139] Sun Y Z, Han J W. Mining heterogeneous information networks [J]. ACM SIGKDD Explorations Newsletter, 2013, 14(2): 20-28
- [140] Yu X, Ren X, Sun Y, et al. Personalized entity recommendation: a heterogeneous information network approach [C] // Proceedings of the 7th ACM International Conference on Web Search and Data Mining, 2014; 283-292
- [141] Sun Z, Yang J, Zhang J, et al. Recurrent knowledge graph embedding for effective recommendation [C] // Proceedings of the 12th ACM Conference on Recommender Systems, 2018; 297-305
- [142] Wang X, Wang D, Xu C, et al. Explainable reasoning over knowledge graphs for recommendation [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 5329-5336

Review of user modeling applications based on social multimedia content

XU Changsheng^{1,2} HUANG Xiaowen^{1,2} QIAN Shengsheng^{1,2} FANG Quan^{1,2}

¹ National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190

² University of Chinese Academy of Sciences, Beijing 100190

Abstract With rapid development of the Internet and social media platforms, hundreds of millions of users have created and shared information as a large quantities of text, image, audio, and video data. Given this large volume of heterogeneous, multi-source, and complex social multimedia content, the challenge of providing effective content understanding and knowledge representation to users with more efficient and high-quality service has become the key to realizing the value of big social media data. This paper reviews research on social multimedia content analysis, knowledge extraction and representation, and on user modeling applications in recent years. In addition, research on social multi-modal feature fusion, cross-modal knowledge extraction and representation, and user modeling based on social media has been presented in detail. We then discuss development trends in social multimedia content knowledge representation, and review research into and application of user modeling. Finally, we summarize likely research directions for multimedia knowledge representation and user modeling.

Key words social multimedia; multi-modality; knowledge representation; user modeling