

基于5W传播模型的技术体系:计算传播技术综述

王续澎, 何洪波, 王闰强

中国科学院计算机网络信息中心, 北京 100083

摘要

海量的数据和人工智能技术为传播学的研究提供了很大的便利,可以有效地解决传播效果测量的问题。该领域现有的研究主要关注计算传播学的研究范式、场景应用和发展路径,缺少对计算传播中新型计算方法的梳理和归纳。采取文献调研法,提出了基于5W传播模型的计算方法归类框架,以提升传播效果为目标,较系统地梳理了传播过程中的新型计算方法,阐述了不同计算方法的思路、所需的数据集以及各项评价方法。分析了计算传播技术面临的挑战,展望了未来可能的研究方向。

关键词

计算传播; 5W传播模型; 计算传播技术; 计算方法; 社交网络

中图分类号: TN92

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2025020

Technology system based on the 5W communication model: a review of computational communication technology

WANG Xupeng, HE Hongbo, WANG Runqiang

Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China

Abstract

Massive data and artificial intelligence technology provide great convenience for communication research, which can effectively solve the problem of communication effect measurement. Existing research in this field mainly focuses on the research paradigm, scenario application and development path of computational communication, and lacks the sorting and categorization of novel computational method in computational communication. This paper adopted the literature research method, proposed a framework for categorizing computational method based on the 5W communication model, and systematically compiled the new computational method in the communication process with the goal of improving the communication effect, describing the ideas of the different computational method, the required data sets and the evaluation methods. This paper analyzed the challenges faced by computational communication technology and looked forward to possible future research directions.

Key words

computational communication, 5W communication model, computational communication technology, computational method, social network

0 引言

21世纪以来,网络科学和大数据的蓬勃发展极大地便利了对海量社交网络数据的分析研究,与此同时,随着社交媒体在人们日常生活中的日益普及,研究者对人类传播行为的研究也越来越便捷与深入。在这种发展趋势下,数据科学、网络科学和传播学逐渐融合成为新的学科——计算传播学。

Lazer等^[1]提出计算社会科学的概念,标志着计算传播学的兴起,也标志着该领域系统性研究的正式启动。Salganik^[2]将计算社会科学看作数据科学与社会科学的融合,系统地介绍了大数据时代对社会研究带来的便利和机遇。祝建华等^[3]更多地关注计算社会科学在新闻传播学中的应用。而计算传播的概念是王成军^[4]首次提出的,计算传播是指数据驱动,借助于可计算方法进行的传播过程,分析计算传播现象的研究领域就是计算传播学,他通过网飞公司邮寄光盘的商业模式和谷歌通过PageRank算法来评估网页内容传播价值的例子,强调了计算的重要性。Van等^[5]提出了计算传播的4点机遇:从自我报告到真实行为、从实验室研究到展示社交环境、从小数据到大数据、从单独研究到合作研究。巢乃鹏等^[6]总结了计算传播研究现状:国外学术界最关注的是健康传播议题,其次是计算方法,再次是计算范式;而国内学术界最关注的是计算范式,其次是研究特征,再次是计算方法。张敬玮^[7]通过文献计量分析、研究、探讨了从2008年

至2020年计算传播学的发展趋势,从作者角度、国家地区角度、关键词角度和文献共引角度清晰地还原了十多年来的研究成果。借助关键词聚类分析方法总结得出,计算传播学是一门以社交媒体为平台,通过对大数据进行分析,研究受众行为和信息传播规律的学科。吴晔等^[8-9]连续两年对相关论文进行综述,与前几年相比,研究议题呈多维度发展,并出现计算宣传和计算视觉等崭新的议题,逐渐形成学科建制完善化、关键议题多元化、研究方法体系化。刘庆振^[10]明确计算传播学可弥补传播学中传播效果在量化方法上存在的缺陷,实现用户场景内容的精准匹配,提升传播效果。目前研究计算传播的有两大组织:一个是国际传播学会计算方法分会,由美国密歇根州立大学Winson Peng教授等发起;另一个是国内的计算传播学研究委员会,由祝建华教授倡导,联合国内28所高校发起,全面推进我国计算传播学领域的教学与科研。

赵甜芳^[11]基于文本挖掘法和网络分析法认为本土计算传播研究处于理论先行的阶段,实证研究较少,传统量化方法研究丰富,新型计算方法偏少。韩少卿等^[12]总结计算传播学中的计算路径,通过大数据计算、解决复杂性问题。通过计算传播学发展可凝练出计算传播的三要素:数据驱动、可计算方法和传播过程。数据驱动和可计算方法是人工智能必要的组成部分,传播过程是从传播模型中凝练出的要素。传播模型由Larswell首次提出,并从5个问题来分析传播:“Who?”“Says What?”“In Which Channel?”“To Whom?”“With What Effect?”^[13]。从这些问题中

提炼出的5个基本要素构成了传播过程，分别是信息传播者、内容、渠道、受众和效果。信息传播者是发起信息的源头，没有传播者，之后的传播过程也无从谈起，它可以是个体、团体或者组织。一个信息传播者的经验、态度、知识、技能和观点都会影响信息的传播。因此在社交媒体中如何识别有影响力的节点和如何使节点影响力最大化成为人们对信息传播者的研究方向。内容是一系列有意义的符号组成和信息组合。如果想要让内容传播得更广、讨论热度持续的时间更长，离不开信息传播者对内容的修饰和表达，同时也与受众对内容的兴趣息息相关，还与信息传播者在什么渠道去传播密不可分。不难看出，这是传播过程各部分共同作用的结果。如果人们仅针对内容本身展开讨论，会发现社交媒体中的热点是一个有意义的研究话题。接收者是传播过程中的受众，在社交网络中如何让受众接收到自己想要看到的内容，肯定离不开对接收者用户画像的构建，也离不开推荐系统技术。有了这两项技术就可以定点给用户推送他们想要的内容和潜在想要的内容。渠道也是在传播过程中很重要的一环，好的内容在不同渠道的发表可能直接影响它的传播效果。效果是传播过程的结果，它并不像前几个维度是具象的，而是基于前几个维度得出的结论，同时也是由多方面因素共同影响的结果。

为了更好地促进该领域的发展，已有研究者梳理了近几年来关于计算传播的相关工作并形成综述，但这些综述主要关注计算传播的研究范式、场景应用和发展路径，并没有结合大数据时代蓬勃涌现的各种计算方法进行梳理。基于以上讨论，本文对有关将智能计算方法融入传播过程基本要素的各种工作进行了系统的总结，包括技术类型、所用数据集和评价指标，构建基

于5W传播模型的计算传播技术体系（如图1所示），并讨论和展望未来的研究方向。

1 相关文献和研究现状

关于计算传播技术体系，本文依照传播过程的5个维度，分别从每个维度选取研究角度，以提升传播效果为目标整理目前所应用的技术现状，梳理相关研究工作，部分文献的计算传播研究目标及其技术方法见表1。

1.1 信息传播者

关于通过计算方法研究信息传播者，根据研究目的可以将信息传播者分为如何识别关键传播者和如何使传播影响力最大化。本节将从这两个方面梳理相关研究工作。

信息传播者是一个传播过程的开端，在社交网络中，每个用户都是潜在的信息传播者。如何从众多信息传播者中发现关键的信息传播者，并通过这些信息传播者使其传播的内容在网络中影响力最大化是笔者感兴趣的研究方向。关键的信息传播者就是有影响力的用户，挖掘这些用户有助于笔者了解该社交网络的结构，并进行社交网络分析。

1.1.1 关键节点识别

(1) 主要算法

定义一个网络拓扑图 $G=(V,E)$ ，包含 $|V|$ 个节点和 $|E|$ 条边。一般将邻接矩阵定义为 A ， a_{ij} 表示节点 v_i 到节点 v_j 是否有连边， $a_{ij}=0$ 代表节点 v_i 到节点 v_j 无连边。 $a_{ij}=1$ 代表节点 v_i 到节点 v_j 有连边。可以利用网络拓扑图结构来刻画节点的重要程度。

计算传播技术体系							
信息发起层		信息中间层		信息接受层		用户行为改变层	
传播者	关键节点识别	内容	热点事件发现	受众	推荐系统	效果	直接影响
	节点紧邻 路径排序 结构洞 熵		话题检测 事件聚类 贝叶斯概率 状态转换机制		混合推荐 用户行为推荐 会话 深度学习		语义网络共识 情感分析
	影响力最大化	渠道	特有渠道信息扩散		用户画像	间接影响	
	贪心算法 启发式算法		社交平台特性 信息扩散树		基本信息构建 数据采集 用户标签提取	构建指标体系	
信息传播						用户态度	

图1 计算传播技术体系

在20世纪中期就有科学家提出了多种中心性指标，如度中心性^[14]、介数中心性^[15]、接近中心性^[16]等用于衡量节点关键性，但是都有明显的劣势。度中心性认为一个节点邻居越多，其影响力越大。由于不同规模的网络中相同度值有不同的影响力，一般定义节点 v_i 的归一化度中心性如式（1）、式（2）所示。

$$DC(i) = \frac{K_i}{n-1} \quad (1)$$

$$K_i = \sum_j a_{ij} \quad (2)$$

其中， K_i 是节点 i 的度数， n 是网络中总节点数。

虽然度中心性有简单、易理解的优势，但是它仅关注局部信息，欠缺对节点周围环境的考虑。介数中心性一般指最短介数

表1 计算传播研究目标及其技术方法

研究对象	研究目标	研究方法
信息传播者	关键节点识别	度中心性 ^[14] 、介数中心性 ^[15] 、接近中心性 ^[16] 、改进度中心性 ^[17] 、局部中心性 ^[18] 、 K 壳分解法 ^[19] 、映射熵 ^[20] 、结构洞 ^[21] 、Tsallis熵 ^[22] 、基于节点度值和弱连接法 ^[23] 、结合度中心性和共同邻居数量的方法 ^[24]
	节点影响力最大化	CELF ^[25] 、CELF+ ^[26] 、Degree Discount ^[27] 、前缀排除最大影响树枝算法 ^[28] 、IRIE ^[29] 、核覆盖 ^[30] 、MTIM ^[31]
内容	热点事件发现	NED ^[32] 、传播热度模型 ^[33-35] 、词语共现的生成式概率模型 ^[36] 、有意义字符串聚类 ^[37] 、多任务图分割的聚类算法 ^[38] 、关联链接网络聚类 ^[39] 、概率的子网选择 ^[40] 、自适应自动回归 ^[41] 、SimBEEP网络结构模型 ^[42] 、隐马尔可夫方法 ^[43]
	消息流行度预测	特征提取方法 ^[44-46] 、点过程建模方法 ^[47-49] 、深度学习预测方法 ^[50-52]
	假新闻检测	人工特征构造 ^[53] 、发文内容分析 ^[54] 、多域视觉神经网络 ^[55] 、时间对抗神经网络 ^[56] 、多模态变分自动编码器 ^[57]
受众	用户画像	网络爬虫、本体的构建 ^[58] 、规则定义 ^[59-60] 、主题模型 ^[61-62] 、聚类算法 ^[63]
	推荐系统	混合推荐 ^[64] 、相似主题推荐 ^[65] 、用户历史行为推荐 ^[66] 、用户兴趣挖掘 ^[67-68] 、语义分析的位置感知 ^[69]
渠道	平台研究	信息扩散树 ^[70] 、社交平台特性研究
传播效果	指标体系	基于效果定义的指标体系 ^[71]
	情感分析	基于情感字典、基于机器学习、基于深度学习

中心性，一个节点的影响力大小可以通过计算经过该节点的最短路径数量来衡量。经过这个节点的最短路径数量越多，节点影响力越大。虽然最短介数中心性考虑了整个网络，但却导致计算时间复杂度高，在实际网络中难以应用。节点 v_i 的接近中心性如式 (3) 所示。

$$d_i = \frac{1}{n-1} \sum_{j \neq i} d_{ij} \quad (3)$$

其中， n 是网络中总节点数， d_{ij} 是节点 v_i 到节点 v_j 的距离。

接近中心性认为一个节点距离其他所有节点的平均路径越短，节点影响力越大，这种度量通过计算节点与网络中其他节点的平均距离来评估节点的重要程度。然而，这种方法同样存在计算时间复杂度高的劣势。基于以上基本指标，众多研究者对其进行改进。

王建伟等^[17]提出基于节点度和节点邻居的度的方法，解决了度中心性全局应用的问题，同时降低了计算复杂度。Chen 等^[18]提出基于局部中心性的方法，通过计算节点邻居的所有一阶邻居个数及二阶邻居的个数弥补度中心性利用节点信息少的缺陷，并降低了计算复杂度。Kitsak 等^[19]提出基于节点位置信息的 K -shell 算法，通过节点中心性的分层，剥离外层度数小的节点，内层的节点拥有更高的重要性。Nie 等^[20]构造 Mapping Entropy，通过考虑节点度和节点邻居的度来计算节点重要性。Yang 等^[21]借助结构洞的概念，提出基于局部信息的启发式算法。杨松青等^[22]提出利用 Tsallis 熵的方法，结合节点的结构洞特征和 K 壳中心性，考虑到节点和邻域节点信息，降低了时间复杂度。Ruan 等^[23]通过弱连接思想提出一种基于节点邻居数量和拓扑重叠的节点度值和弱连接强度指标。朱敬成等^[24]

通过考虑节点及其邻域之间的网络拓扑重合度，提出一种结合节点度中心性和共同邻居数量的 DCN (degree centrality and neighbor) 评估指标。**表 2** 为相关指标、计算式及参数含义。

(2) 数据

一般地，研究人员会选取已有真实网络或者人工生成网络进行节点重要性排序计算式的比较和分析。比较常见的真实网络有爵士音乐家网络、Slavko 在 Facebook 上的朋友圈网络、人群感染网络、邮件网络、科学家合作网络、有关蛋白质之间相互作用的 Yeast 网络和海豚社交网络等^[24]。这些网络的结构特征属性（网络中的节点总数、网络中节点之间的连边数量、网络节点的平均度大、网络集聚系数、平均最短路径长度、网络直径、网络经过 k 壳分解后最后剥离的节点所属的层级）基本相同。为了保证识别关键节点方法的普适性，研究人员通常会选取多个网络进行实验，因为不同网络的统计特征值存在差异。常见的人工生成网络依据不同的结构类型可分为 3 种^[72]。ER 随机图模型 (erdős-renyi random graph model) 是将一堆顶点随机连接，相连是根据某个概率值来确定的。BA (barabási-albert) 模型的典型特征是在网络中的大部分节点只与很少的节点连接，而有极少的节点与非常多的节点连接。小世界模型 (small-world model) 是一个由大量顶点构成的图，其中任意两点之间的平均路径长度均比顶点数量小得多。**表 3** 为社交网络数据集。

(3) 评价方法

目前主流节点影响力排序的评价标准有两个。一个是基于网络的鲁棒性和脆弱性的方法，另一个是基于网络的传播动力学模型的方法。主要思想是通过考察节

表2 关键节点识别公式

指标名称	公式	参数含义
度中心性	$DC(i) = \frac{K_i}{n-1}$	K_i 是节点 <i>i</i> 的度数, n 是网络中总节点数
介数中心性	$BC(i) = \frac{2}{(n-1)(n-2)} \sum_{i \neq s, t \neq i, s \neq t} g_{st}^i$	g_{st} 为从节点 v_s 到 v_t 的所有最短路径的数目 g_{st}^i 为从节点 v_s 到 v_t 的 g_{st} 条最短路径中经过节点 <i>v</i> 的最短路径的数目
接近中心性	$CC(i) = \frac{n-1}{\sum_{j \neq i} d_{ij}}$	d_i 为节点 v_i 到其他节点的平均最短距离
改进度中心性	$W(i) = \frac{w_i}{\sum_{j \in N} w_j}$	w_i 为节点 v_i 所连边的权值之和
局部中心性	$Q(u) = \sum_{w \in \Gamma(u)} N(w)$ $C_L(v) = \sum_{u \in \Gamma(v)} Q(u)$	$N(w)$ 为节点 v_w 一阶邻居和二阶邻居数量之和 $\Gamma(u)$ 为节点 v_u 的邻居
Tsallis熵	$Cnc(i) = \sum_{j \in \Gamma(i)} IC(i)$ $TSM(i) = \sum_{j \in \Gamma(i)} Cnc(i)$	$Cnc(i)$ 节点 v_i 的邻域核心中心性 Γ_i 为节点 v_i 的邻居集合
结构洞	$DSHC_i = \sum_{j \in \Gamma_i} ((\frac{1}{k_i} + \frac{1}{k_j}) * \frac{1}{1 + \Delta_{ij}})^2$	Γ_i 为节点 v_i 的邻居集合 Δ_{ij} 为节点 v_i 和节点 v_j 的结构洞数量
映射熵	$ME_i = -DC_i \sum_{j=1}^M \log DC_j$	DC_j 为节点 v_i 邻居的度中心值
基于节点度值和弱连接法	$S_{ij} = \frac{ N(i) \cap N(j) }{\sqrt{k_i \cdot k_j}}$ $DWT_i = \sum_{i \in N(i)} (\frac{1}{k_i} + \frac{S_{ij}}{k_i})^2$	k_i 为节点 v_i 的度值 $v(i)$ 为节点 v_i 的邻居
结合度中心性和共同邻居数量的方法	$DCN(i) = \sum_{j \in \Gamma(v_i)} (\sum_{w=1}^{\text{ceil}(\log_2^i)} \frac{1}{s(v_i, v_j) + w})$	Γ_i 为节点 v_i 的邻居集合 $s(v_i, v_j)$ 为节点 v_i 和节点 v_j 的共同邻居数量 w 为调整参数 K_i 是节点 <i>i</i> 的度数

表3 社交网络数据集

数据集名称	节点数量/个	连边数量/条	节点平均度	网络聚集系数
Jazz	198	2 742	27.70	0.618
Dolphins	62	159	5.12	0.259
Infectious	410	2 765	13.48	0.456
Email	1 133	5 451	9.62	0.220
NetScience	379	914	4.82	0.694
Yeast	2 375	11 693	9.84	0.306
powergrid	4 941	6 594	2.66	0.080

点对网络结构的变化趋势来衡量节点排序结果的准确性。

① 基于网络的鲁棒性和脆弱性的方法

在讨论网络的鲁棒性和脆弱性时,通常采用最大连通子图比例和网络效率来衡量网络遭受破坏的程度。网络被破坏的趋势变化越明显,越快到达最小值,说明该节点越关键。在移除节点的过程中会形成一个互不连通的网络团体,其中最大的网络团体被称为网络巨片,而网络巨片包含的节点数与整个网络的节点数的比值则为最大连通子图比例。

$$\bar{S} = \frac{S}{n} \quad (4)$$

其中, S 指删除一定节点后网络巨片的节点数, n 为整个网络节点数^[24]。

网络效率以节点之间的路径长度衡量网络连通性强弱, 删除一定节点后, 导致节点对之间的路径变长甚至无法到达, 阻碍网络中节点连, 如式(5)所示。

$$\varphi = \frac{1}{n(n-1)} \sum_{i,j \in V} \varphi_{ij} \quad (5)$$

其中, n 为整个网络节点数, $\varphi_{ij} = \frac{1}{d_{ij}}$, d_{ij} 为节点 v_i 到 v_j 的最短路径长度^[24]。

② 基于网络的传播动力学模型的方法

基于网络的传播动力学模型的方法主要有传染病模型 SI (susceptible infected)、SIS (susceptible infected susceptible) 和 SIR (susceptible infected recovered) 模型^[73]。节点在传染病模型中有3种可能的状态: 易感态 (susceptible, S)、感染态 (infected, I)、免疫态 (recovered, R)。在 SI 模型中, 以单个节点为感染源, 以 p 的概率感染其他节点, 通过观察不同时间感染节点的数量, 可以衡量节点的传播速度, 并将其作为该节点的传播能力。在 SIS 模型中, 一个节点的传播能力是在稳态条件下该节点被感染的概率。SIR 模型则通过计算一个节点的平均传播范围来确定其传播能力。

1.1.2 节点影响力最大化

节点影响力最大化问题可描述为在网络中选取指定大小的种子节点集合, 使该集合在一定时间内传播覆盖更多的节点。在一个网络拓扑图 $G = (V, E, W)$ 中, V 代表节点集, E 代表有向边集, W 代表有向边对应权值集合, $W(u, v)$ 代表节点 u 激活节点 v 的概率。对于给定网络 G , 参数 k , 影响力最大化问题旨在找出种子集 $S \subseteq V$ 且

$|S| = k$, 使种子集 S 影响范围最大化。激活节点总数记为 $\sigma(s)$ 。

节点影响力最大化问题最早由 Domingos 等^[74]引入社交网络领域展开研究。研究方法的演变经历了最初的贪心算法, 到启发式算法, 再到目前结合两种方法的优势而形成的混合算法。这种融合算法不仅适用于多种传播模型, 也适用于大规模社交网络, 它能够在保证一定的影响范围的同时, 缩短算法的运行时间。

(1) 主要算法

在 Domingos 等^[74]提出影响力最大化算法之后, Kempe 等^[75]首次提出形式化 k 个节点影响力最大化问题, 并证明此问题在不同传播模型下均是一个 NP-hard 问题。他们采用一种贪心算法使影响范围达到最优解的 63%。但由于每次迭代寻找最有影响力节点都需要对非种子节点进行影响力计算, 运行时间长, 难以应用于大规模网络。Leskovec 等^[25]提出通过影响传播的子模性质, 减少冗余计算的 CELF (cost effective lazy forward) 算法, 使贪心算法在维持原有影响范围下运行时间提高了 700 倍。Goyal 等^[26]在 CELF 算法的基础上设计出复杂度更优的 CELF++ 算法。虽然科学家们一步步地优化算法, 但运行时间仍为数小时。Chen 等^[27]首先尝试改进贪心算法提出一种 Newgreedy 算法, 运行时间只比 CELF 算法缩短了 15%~34%, 并没有在数量级上有较大差别。于是他们提出了一种启发式算法的策略, 旨在显著提高计算速度, 同时确保影响范围不受影响。在此基础上, 他们还设计了基于节点度的 Degree Discount 算法, 这种算法基于节点度进行优化, 以实现更高效的节点影响力最大化。算法基本思想是当节点 v 的邻接节点中已有被选作初始传播节点时, 应对节点 v 的度数打折处理。实验结果和贪

心算法的精度接近并将运行时间减少至数秒。Chen 等^[28]利用节点的局部树结构近似传播影响,克服了 Degree Discount 算法在均匀独立级联模型中的局限性,提出一种前缀排除最大影响树枝算法,在非均匀的独立级联模型下也能很好地平衡运行时间和传播影响范围。Jung 等^[29]提出了一种集成影响力排序和影响力估计的 IRIE (influence ranking and influence estimation) 算法,该算法适用于独立级联 (independent cascade, IC) 模型以及包括负面意见传播的 IC-N 模型。相比于前缀排除最大影响树枝算法,IRIE 算法占用了更少的内存资源,并且在处理百万级的大规模网络时,其运行速度比前缀排除最大影响树枝算法快了两个数量级。此外,IRIE 算法还可以移植到并行图计算平台上,进一步提升其处理效率。曹玖新等^[30]基于核数的思想,提出一种核覆盖算法 (core covering algorithm),通过引入覆盖距离参数 d 使种子节点保持一定距离,从而克服影响重叠问题,此算法在运行时间和影响范围方面均优于其他启发式算法。曹玖新等^[76]将启发式算法和贪心策略相结合,首先利用启发式算法选出候选种子节点集合,随后再利用贪心算法找出较优解。虽然这一融合方法的运行时间从秒数量级提升到小时数量级,但相较于传统的贪心算法需要以天为单位的计算时间,仍有显著提升。从影响范围角度来看,该方法也优于其他的启发式算法,并能适用于百万节点的大规模网络。此外,Borgs 等^[77]基于反向影响采样技术对节点进行筛选,随后结合贪心算法来识别影响力最大的种子集。在此基础上,Tang 等^[78]引入切尔诺夫边界思想估计采样次数,使算法更稳定、高效。王璿等^[31]提出 MTIM (mixed three-stage influence

maximization) 算法,该算法融合了节点度筛选策略、边界约束策略、影响力剪枝策略,并能够在 IC 传播模型和线性阈值 (linear threshold, LT) 传播模型上同时运行。通过结合网络拓扑结构和采样技术。MTIM 算法不仅缩短了计算时间,还扩大了影响范围,尤其适用于大规模网络环境下的应用。上述模型均把社交网络抽象为静态图考虑。

节点影响力最大化问题实质上是对关键节点的进一步筛选,这一过程不仅将研究视角从个体节点拓展至群体组合,以符合真实网络节点间相互影响的特性,同时通过排除冗余的传播者来优化传播效率。这种复合式传播的方法首先依据关键节点识别指标计算节点影响力,然后依赖特定的传播模型模拟传播过程,并将最终的传播覆盖范围作为传播效果测量指标。

在解决节点影响力最大化问题时,通常需要贪心算法的优化和网络拓扑结构的节点初筛选两项技术。研究方向上也从单一网络扩展到多社交网络,探索如何通过自传播性建立不同网络之间的联系。同时,也有研究聚焦于在同个社交网络的多条信息竞争环境下最大化影响力。此外,吴安彪等^[79]将社交网络抽象化为动态图,探讨在大规模时序图下的影响力最大化问题,为这一领域带来新的研究维度和深入见解。

(2) 数据

为了保证实验结果的普适性,数据一般选取多个结构不同的真实网络分别模拟传播过程。前文提到的真实网络数据均可用于节点影响力最大化方法的检验。

(3) 评价方法

在节点影响力最大化问题中评价指标为:①节点影响范围 R ,即在规定的种子节点数目 N_0 下的最终影响节点数目 N_t ;②算

法运行时间 T ，即传播结束后算法的运行时间。实验通常采用多种不同规模的数据集，基于某个或多个传播模型在不同的传播概率下对不同种子节点数量进行研究，对比不同算法的影响范围和运行时间。如果种子集的影响范围越广，运行时间越短，则算法性能越好。

1.1.3 小结

本节主要讨论了关键节点识别和节点影响力最大化问题的研究方法。这两个问题是紧密的、有联系的。首先，从网络图中用不同的方法寻找关键节点。其次，利用这些已有的关键节点在传播模型中模拟传播过程，传播过程结束后得到种子节点集的影响范围即覆盖节点数。为了使影响力最大化问题适用于大规模网络，笔者倾向使用启发性算法，给定节点影响范围并缩短运行时间。目前来看，针对节点影响力这一问题从网络结构角度研究比较充分，但是利用节点自身其他特征（如转发次数、提及次数、有效用户数量等）的分析较少。这种情况也是计算传播学在研究个体时的简化假设与社会科学强调社会意义相互冲突的一个具体体现^[80]。

1.2 内容

内容在传播过程中代表不同用户表达的信息。随着网络媒体的快速发展，人们能够从网络上获取纷繁复杂的信息，有些信息在传播中成为大家讨论的热点，有些则缺少讨论热度。随着 ChatGPT 等生成模型的不断发展，假新闻也越来越多。本节选取热点事件发现、消息流行度预测和假新闻检测 3 个方面讨论对应技术的发展现状。

1.2.1 热点事件发现

(1) 主要算法

热点事件是社会中引起广泛关注及讨论、激起民众情绪、引发强烈反响的，发生在特定时间和地点的事情。热点事件发现技术最早源于话题检测与跟踪（topic detection and tracking, TDT）体系中的新事件检测（new event detection, NED）技术，一般采用文本特征计算与每个事件模型的相似度，并根据预测的先验阈值来判断该报道是否为新事件。

基于新闻的热点事件发现问题可以用 TDT 体系中的 NED 技术作为支撑，加入热度定义计算式，即可解决此问题。NED 技术的主流方法来自 Allan^[32]提出的在线识别系统检验事件是否为新出现的事件。主要思路是计算事件与每个已知事件模型的相似度，并根据先验阈值判断是否为新事件。对于早期针对长文本、信息量充足、表达语言规范的新闻数据，刘星星等^[33]通过空间向量模型特征化文本，将语料按时间分组，对语料进行 Single-pass 聚类得到事件列表，将事件的有效报道天数、事件报道数量频率和事件的平均相似度作为特征计算事件热度。

$$DR_i = \frac{\sum_{j=1}^n \frac{df_{ij}}{df_j}}{n} \quad (6)$$

$$W_i = TR_i \times DR_i \times F_i \quad (7)$$

罗亚平等^[34]根据 TF-IDF 思想构造媒体关注度特征，利用用户阅读、回复的比例构造用户关注度特征，提出基于媒体关注度和用户关注度的加权方法计算热点事件热度指数。徐佳俊等^[81]基于隐含狄利克雷分布（latent Dirichlet allocation, LDA）话题模型，通过话题抽取结果，构造话题支持率计算式识别热点话题。

后来随着微博、推特等社交平台的兴起,内容呈现出短文本、碎片化的新特点,传统方法并不适用特征稀疏的文本信息。由于事件的语义特征会随着时间推移而变化,所以新事件检测技术并不能满足在动态网络资源上的应用要求。杨亮等^[82]提出情感分布语言模型,通过比较相邻时段情感分布语言模型差异的方法来发现热点事件。Bai等^[36]提出了一种基于突发词语、词语共现的生成式概率模型,如果某个词汇在一段时间内出现频率极高,并且该词汇和其他突发词语有接近的特征分布,说明它们共同组成了一个热点。贺敏等^[37]提出将有意义字符串作为文本特征,通过此特征再利用 Bisecting K -means 聚类算法发现话题,并利用话题相关文档数量来反映话题热度。这种有意义字符串由一系列词语组成,表达意义较完整,内部结构稳定,解决了特征稀疏的问题。Nikolaos等^[38]考虑了社交媒体的动态性和时序性,

$$g_3(k, w) = \frac{\sum_{m=1}^{N(k)} p_m^f(k) \cdot p_m^r(k) \cdot i_m(w, k)}{\sum_{m=1}^{N(k)} p_m^f(k) \cdot p_m^r(k)} \cdot \log \frac{\sum_{j=1}^p \sum_{m=1}^{N(k-j)} p_m^f(k-j) \cdot p_m^r(k-j)}{\sum_{j=1}^p \sum_{m=1}^{N(k-j)} p_m^f(k-j) \cdot p_m^r(k-j) \cdot i_m(w, k-j)} \quad (10)$$

其中, $N_k^{(w)}$ 指时间间隔 k 内包含词 w 的推文数量, N_k 指时间间隔 k 内的推文数量, C_k 指时间间隔 k 内所有单词的数量, $C_k^{(w)}$ 指时间间隔 k 内词 w 出现的数量, $f_m(k)$ 指在时间 k 内第 m 条推文的粉丝数, $r_m(k)$ 指在时间 k 内第 m 条推文的转发数。

Liu等^[39]提出在线事件检测框架,通过弹性阈值把文档分类为事件话题,并基于 Association Link Network Cluster 的聚类方法发现新的热点。Shen等^[40]提出一种基于概率的子网选择方法,先筛选出候选节点子集,再通过候选节点子集根据事件参与的概率判断是不是热点。这种方法会受到候选节点筛选条件的影响,是一种通过节点信息进行热点检测的应用。Tong

提出基于模糊表示的推特文本表示方法,结合 3 种 TF-IDF 改进方法形成一个时序信号,并考虑一个词会同时出现在不同事件的情况,提出一种多任务图分割的聚类算法,通过词与词之间的相似度对图进行分割,以实现事件的聚类。3 个改进文本的表示如式 (8)、式 (9)、式 (10) 所示。式 (8) 加入时间因素,用时间间隔内包含特定单词的推文数量代替单词在文档中出现的次数。式 (9) 加入时间因素,用时间间隔内特定单词出现的数量代替单词在文档中出现的次数。式 (10) 考虑了在时间间隔内粉丝数和转发数衡量用户的信用度和内容重要程度。

$$g_1(k, w) = \frac{N_k^{(w)}}{N_k} \cdot \log \frac{\sum_{i=1}^p N(k-i)}{\sum_{i=1}^p N^{(w)}(k-i)} \quad (8)$$

$$g_2(k, w) = \frac{C_k^{(w)}}{C_k} \cdot \log \frac{\sum_{i=1}^p C(k-i)}{\sum_{i=1}^p C^{(w)}(k-i)} \quad (9)$$

等^[41]提出基于自适应自动回归方法,能够判断事件在一段时间内的热度变化趋势,根据设定阈值判断从某一时刻起该事件成为热点事件的可能。Ma等^[42]提出 SimBEEP 网络结构模型,通过考虑事件频率平均改变速率、事件频率加速度、最长推文传播长度和好友网络中的社区数量,基于贝叶斯模型预测热点事件。Nie等^[35]通过提取关键词,并基于网络本体语言 (web ontology language, OWL) 方法对话题进行聚类,监测一段时间内的话题数量得到可能的热点事件,再借助用户关联网络,利用用户被激活的数量反映事件传播热度程度。Liu等^[43]采用隐马尔可夫方法中的状态转移机制模拟话题状态演变过

程，建立事件时效模型，通过前向概率预测算法预测下一步可能的状态和相关文章数量。

(2) 数据

基于新闻的热点发现问题通常会爬取不同门户网站中不同领域的新闻，而基于社交媒体的热点发现问题通常会爬取新浪微博或者 Twitter 上的发文，并通过人工对发文标注热点。

(3) 评价方法

基于不同问题的构造形式，热点发现问题评价方法可分为基于聚类问题的评价方法和基于分类问题的评价方法。

① 基于聚类问题的评价方法

互信息衡量真实聚类与预测的一致性。NMI (normalization mutual information) 被定义为聚类分配与数据集已有标签之间的互信息，用经验边际最大可能熵的算术平均值进行归一化。V-measure 使用基于条件熵方法计算同质性和完整性的调和平均值，以计算检测的事件和实际事件的相似性。完全同质的聚类是每个聚类具有属于同一类标签的数据点的聚类，完全完整的聚类是将属于同一类的所有数据点聚类到同一个聚类中^[83]。

② 基于分类问题的评价方法

分类评价指标见表 5，其中 TP 表示被正确划分为正例的个数，TN 表示被正确划分为负例的个数，FP 表示被错误地划分为正例的个数，FN 表示被错误地划分为负例的个数。

1.2.2 消息流行度预测

社交网络信息纷繁复杂，从海量的消息中预测并识别未来的热门消息，可帮助人们从信息过载的困境中解脱出来。消息流行度预测大体分为基于特征提取方法、基于点过程建模方法、基于深度学习的流

行度预测方法。

(1) 基于特征提取方法的流行度预测方法

将流行度预测描述为一个回归或分类问题，通过人工定义抽取结构、内容、时序和用户等特征进行预测。但是这些特征通常是从启发式算法获取的，模型最终的预测性能又极度依赖这些特征的质量。Szabo 等^[44]通过研究 Digg 上的新闻数据和 Youtube 的视频数据，发现内容早期的流行度和内容后期的流行度存在对数线性关系。Pinto 等^[45]将流行度的增长序列作为特征，使用多元线性回归模型预测单条微博的流行度。高金华等^[46]提出了一种基于相似消息的流行度预测方法，首先使用 LDA 主题模型表示待预测消息 j 的传播模式，然后选出历史消息相似度较高的 K 个消息，将 K 个较相似的历史消息在该时间点的流行度预测作为消息 j 在该时间点的预测。这种方法考虑到消息的特异性并且利用了历史消息，能够显著提升流行度预测效果。

(2) 基于点过程建模方法的流行度预测方法

通常只利用待预测消息自身的信息，通过学习观测窗口内信息传播过程随时间变化的规律，对传播速率建模，从而进行预测。Shen 等^[47]提出了一种增强泊松过程的 RPP (reinforced poisson process) 模型，在传播速率建模中引入了消息本身的

表 4 分类问题评价方法

指标名称	英文名称	公式
准确率	Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$
精确率	Precision	$\frac{TP}{TP + FP}$
召回率	Recall	$\frac{TP}{TP + FN}$
F1 值	F1 score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

吸引力、时间衰减效应和富者愈富机制。Gao等^[48]在RPP模型基础上使用时间衰减函数，扩展增强泊松过程。Mishara等^[49]利用自激励点过程，使传播速率考虑历史转发的累计效应来预测内容流行度。但是传播速率函数会依赖某些假设，在真实情况下并不能论证这些假设成立。

(3) 基于深度学习的流行度预测方法

该类方法通过深度学习构造特征从而进行更精准的预测。对于文本内容，研究者提出利用循环神经网络和卷积神经网络捕捉短文本序列特性，对于新闻等长文本卷积神经网络无法描述长文本结构特性，Liao等^[50]加入层级注意力机制对句子中不同词的权重以及长文本不同句子权重进行学习，有效刻画长文本结构及语义特性。Zhang等^[51]提出一种用户引导的分层注意力网络，将非结构化文本和视觉模式的表征学习与流行度预测整合到一个模型里。Xu等^[52]提出一种结合视觉、文本、用户和时间空间特征的流行度预测框架，建立特征与流行度得分关系，并引入关注机制，在推理阶段为指定模态分配更大的权重。

1.2.3 假新闻检测

(1) 计算方法

假新闻检测早期是对新闻文本内容差异性特征进行分析，比如标点符号数量、负面词比例等。Castillo等^[53]将特征分为信息、用户、主题和传播四大部分，利用J48决策树算法分类检测新闻的真实性，但手工特征需要花费大量时间探索其有效性。后来研究者在内容特征基础上加入用户评论、用户发文历史偏好等辅助信息。Giachanou等^[54]提出了基于长短期记忆(long short-term memory, LSTM)网络的Emocred方法，通过结合文本内容情

感信号判断新闻内容的真伪。为了增强假新闻的可信度，发布者在文本内容基础上加入图片形式的信息。Jin等^[84]首次在新闻检测任务中系统地探索图像特征，并提出5种视觉特征和7种统计特征进行检测。假新闻图片可分为篡改图片和误导性图片两类，篡改图片是将真实图片进行修改误导大众，误导性图片是将旧图片放到新发生的新闻里。Qi等^[55]提出一种多域视觉神经网络(multi-domain visual neural network, MVNN)，通过视觉信息借助卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)提取融合像素域特征和频域特征，并加入注意力机制进行分类。为了识别新出现的假新闻，Wang等^[56]提出了一种事件对抗神经网络(event adversarial neural network, EANN)，其包含一个时间检测器和假新闻检测，文本内容部分将词向量作为输入，利用CNN生成文本特征表示，图像表征从ImageNet预训练的VGG(visual geometry group)-19模型中获取。通过挖掘事件不变的特征预测新出现的假新闻。Khattar等^[57]提出了多模态变分自动编码器模型，通过使用LSTM+VGG-19的双模态提取特征，结合变分自动编码器来执行检测任务。Tahmasebi等^[85]通过数据增强的方式创造虚假样本让模型有更好的泛化能力。Dun等^[86]提出了基于知识图谱的检测方法，构造了一个知识感知注意力网络框架，将实体及实体上下文知识纳入检测。

(2) 数据

在假新闻检测领域，有两个被广泛使用的数据集，见表5。一个是MediaEval验证多媒体使用基准中的Twitter数据集，其包括文本数据和附带的图片相关信息；另一个是微博上由真实的新闻和经过

微博辟谣系统验证的假新闻组成的数据集。

1.2.4 小结

不论是挖掘热点、消息流行度预测还是假新闻检测，研究方法初期大多仍以构造文本特征进行分类预测，如突发词语、情感词汇的变化、关键词出现频率的变化。也可以对传统 TF-IDF (term frequency - inverse document frequency) 算法进行改进，考虑更多维度全面地表示文本。随着多模态和深度神经网络技术的发展，研究者开始挖掘图像特征和视频特征，对内容的不同模态进行全方位建模分析。

1.3 受众

受众是传播过程中接收信息的一方。在这个时代，数据正以爆炸式的速度增长，为人们带来了源源不断的信息。然而，这种信息的泛滥也导致了严重的信息过载。为了使人们更精准地接收信息，推荐系统技术和用户自身画像构建技术变得越来越有意义。推荐系统技术能够针对不同用户的需求挖掘其感兴趣的项目；用户自身画像构建技术能够特征化表示用户，提取用户的重要特征，服务于个性化推荐，提高用户的满意度和黏性。

1.3.1 用户画像

用户画像这一概念最早是由交互设计之父 Cooper^[87]提出的，他认为用户画像不是一个实体，而是从数据层面提取的虚构代表。之后科学家们逐渐完善了这一概念，认为用户画像是一个可以描述用户需求、偏好和兴趣的用户信息集合。一个完整的用户画像模型构建可分为三大技术领域，

表5 假新闻检测领域中被广泛使用的数据集

数据集	假新闻数量/条	真实新闻数量/条	图片数量/个
Twitter 数据集	7 898	6 026	514
微博数据集	4 749	4 779	9 528

包括用户基本信息特征构建技术、数据采集技术和用户标签提取技术。

(1) 基本信息特征构建技术

基本信息特征构建技术的意义在于明确需要用户哪些方面的信息，本质上是对用户信息的筛选和归类，构建不同维度的用户信息，从而服务于后续标签化用户，全面地刻画用户特征。赵雅慧等^[88]在参考 Lainé-cruze 等^[89]的观点后，提出了将用户画像的构建维度分为用户和情境相关维度两大类。其中用户相关维度包括自然属性维度、兴趣维度、社交维度、能力维度、行为维度、信用维度等。情境相关维度由用户位置信息和与这些位置相关的搜索历史记录组成。王凌霄等^[90]从用户资历、用户参与度、用户回答质量和用户发展趋势 4 个方面构建了社会化问答社区用户画像。徐海玲等^[91]将豆瓣网电影作为对象，在社交媒体领域进行构建，从用户的自然属性、用户的行为特征属性和用户的需求属性 3 个方面划分，并能够基于不同社交媒体的特性调整 3 个方面的特征。

(2) 数据采集技术

数据采集技术分为线上采集技术和线下采集技术。线上采集技术又可分为采集开放数据、采集第三方平台数据和 App 数据。研究者可以通过自己编写的爬虫程序或者已有爬虫程序（如八爪鱼），获取想要的的数据。线下采集技术又可分为问卷调查和问卷访谈，但在数据量比较大的时候，获取统计并不方便。

(3) 用户标签提取技术

用户标签提取技术主要分为基于本体的构建方法、基于规则的定义方法、基于主题的方法和基于聚类的方法。

① 基于本体的构建方法

本体构建可理解为基于知识资源的构建，基于本体的构建方法如图2所示。刘海鸥等^[58]通过用户基本信息、用户行为信息、用户情境信息构建了旅游情境化推荐的用户画像概念模型。不同场景下的本体构建有很大差异，不具有普适性，需要详细理解业务需求才能完整、准确地构建用户画像。

② 基于规则的定义方法

基于规则的定义方法需要对数据的特点和特定的理论进行自定义提取标签规则，计算相应指标。汪强兵等^[59]基于手势行为在用户兴趣维度构建画像，基于用户浏览微博过程中产生的手势行为种类与次数，计算用户对微博的兴趣度。Tahar等^[60]改进TF-IDF方法计算用户历史检索文档中词语的权重和用户对文档标注的标签的权重，提出一种基于社交维度和情景维度的用户画像方法。

③ 基于主题的方法

基于主题的方法利用用户的行为表现

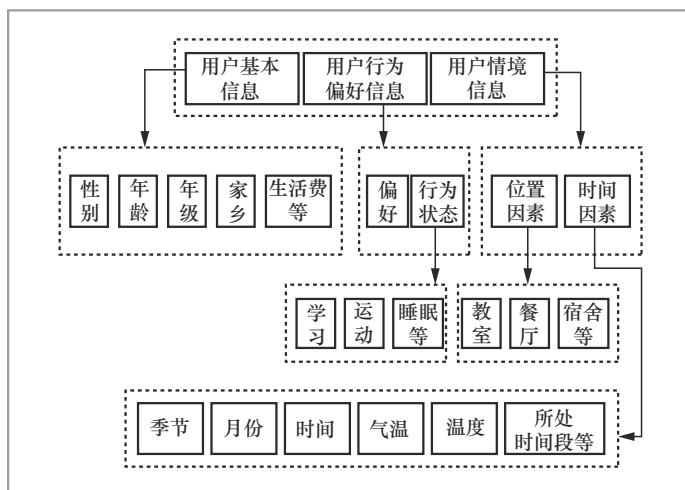


图2 基于本体的构建方法

和语义信息得到不同主题的权重，但缺少自身属性维度的构建。林燕霞等^[61]运用LDA主题模型方法，针对移动应用程序日志数据以及微博文本数据进行分析，得到用户对不同主题的偏好特征，构建基于微博用户的兴趣画像。张亚楠等^[62]提出一种L-LDA有监督的主题模型，针对科研人员每个时间点的科研行为数据进行分析，得到科研人员画像标签在每个时间点的分布，抽取主题概率作为画像标签权重。

④ 基于聚类的方法

基于聚类的方法能够发现用户的群体特征，有助于了解用户的兴趣分布，但无法发现每个用户的特征。张艳丰等^[63]使用K-Medoids聚类方法，对移动社交媒体倦怠用户画像进行无监督分类，并将其分为4个群体类型。K-Medoids算法通过绝对误差准则函数衡量每个对象与之对应的代表对象之间的相异度总和是否最小。

$$E = \sum_{j=1}^k \sum_{p \in c_j} |p - o_j| \quad (11)$$

其中， E 是数据集中所有对象间绝对误差之和， p 是簇 c_j 中一个给定的对象， o_j 为簇的中心点， k 是最终要得到的簇的数目。

用户画像可以依据研究领域的特点、研究平台的业务流程以及研究目的来确定合适的维度，从多个不同的维度构建的特征尽量要全面、详细地刻画用户，有助于标签提取。事实上，这是计算传播学新范式和大众传播学传统范式之间的根本性差异，通过精确地对用户画像进行描述来预测用户下一秒的信息需求^[92]。

1.3.2 推荐系统

(1) 主要算法

目前推荐系统在电子商务、信息检索、社交网络、位置服务、新闻推送等多个领域有良好的应用。本节基于传播过程中内

容为载体、受众为目标的个性化新闻推荐这个领域，梳理相关研究工作。

传统新闻推荐方法主要包括协同过滤、基于内容的推荐方法和混合推荐方法，前两种方法都存在冷启动的问题。而个性化推荐方法需要考虑用户的兴趣偏好和新闻的内容、位置、类型相匹配的问题。孟祥福等^[93]总结了个性化新闻推荐的3个特点（多样性、时效性和流行性），并构建了新闻-用户表示学习框架，如图2所示，能够从大量数据中学习新闻和用户特征，提供高效的推荐服务。Zhu等^[94]提出基于用户行为和新闻流行度的动态个性化用户分析方法，该方法根据用户的阅读行为和新闻的受欢迎程度，赋予每个历史新闻相应的权重，通过计算阅读行为在不同新闻类型下的概率记录阅读偏好。同时该方法加入用户长期偏好特征和短期偏好特征，解决一些用户偏好容易随时间改变的问题。Darvishy等^[64]提出了一种混合推荐算法，该算法结合了基于用户阅读行为相似度、阅读频率以及用户偏爱文章的平均热度这3个特征的协同过滤方法与基于内容构造命名实体识别的阅读特征推荐方法；其提出新的聚类算法有序聚类，能够根据新闻特征和用户阅读行为对新闻和用户进行聚类。王嵘冰等^[65]基于相似主题和HITS（hyperlink-induced topic search）算法，在微博主题相似度中引入用户的权威度，并依照不同用户类别进行推荐。这种方法融合了用户自身影响力特征，在内容特征之外新增了一个维度，提高了推荐准确率。Okura等^[66]提出一种基于嵌入的方法，分布提取特征。首先利用降噪编码器提取文章特征，然后将用户的浏览历史作为输入序列，通过RNN来生成用户表示，最后采用点乘的方式将文章的隐表示和用户隐表示计算生成推荐列表。De等^[67]

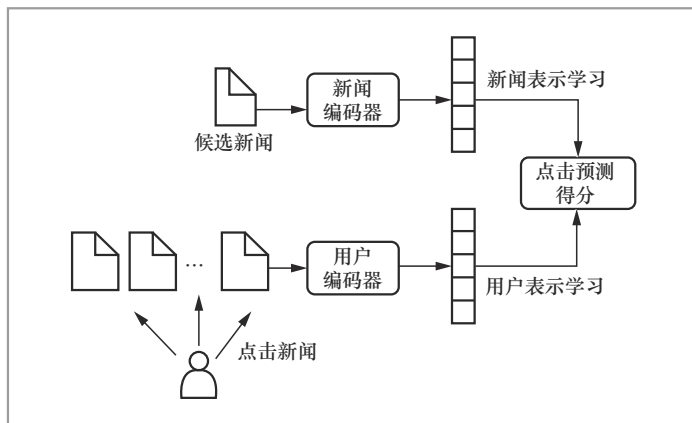


图2 新闻-用户表示学习框架

提出一种基于深度学习的Chameleon元架构推荐系统，它包含文章内容表征模块和下一篇文章推荐模块，并且可以根据给定问题设置的特殊性，以各种方式进行实例化；随后在Chameleon元架构基础上使用深度学习技术实现学习文章表征，并利用用户会话预测推荐排名^[95]。刘树栋等^[68]通过注意力机制区分用户对不同兴趣维度的影响，缓解新闻推荐长尾效应，提高了模型对用户兴趣挖掘的能力。张玉朋等^[96]通过图神经网络捕捉用户和新闻的高阶关系，建立用户长期兴趣和候选新闻表示，并利用Transformer对用户短期兴趣建模，解决LSTM或GRU（gate recurrent unit）方法认为每条新闻同等重要的问题。Chen等^[69]提出了一种具有明确语义分析的位置感知的个性化新闻推荐方法。这种方法不仅基于用户兴趣而且加入位置信息这一特征，并利用深度神经网络提取出密集的、抽象的、低维的、有效的用户、新闻和地点的特征表示。黄振华等^[97]将增量学习引入新闻推荐任务，避免大量的新闻数据和用户交互数据导致长时间的训练。

目前基于深度学习的推荐算法逐步取代了传统的推荐方法，一是因为构造了更复

杂的特征代替人工提取特征，以提高模型的泛化能力，二是因为深度学习被应用在社交网络图结构数据上能更好地挖掘用户新闻之间的交互信息，但是图结构数据训练时间长，占用大量计算资源是待解决的问题。

(2) 数据

通常数据来源于公开的数据集和从公共新闻门户网站或者社交网站爬取的网络数据集。

公开数据集有3个比较有名的数据集。

①Globe.com 数据集。Globe.com 是巴西的媒体公司，该网站有超过8 000万名独立用户，每月发布超过10万篇新文章；②SmartMedia Adressa 数据集包含来自挪威新闻门户网站的约2 000万次阅读记录；③MIND 数据集由微软新闻网站采集的100万用户的新闻日志构成。

爬取的网络数据集涉及娱乐、体育、军事、游戏、时尚、政治和文化等不同领域不同话题的用户互动和发表文章的信息的数据，对于社交网站而言，可以爬取推文中新闻文章的网址获取新闻数据，数据量为5万~10万条，并收集2 000个左右的用户信息。这种爬取的方法会产生脏数据或数据冗余的情况，需要清洗数据，否则会对实验造成误差。

(3) 评价方法

新闻推荐系统评价方法主要分为离线评估和线上评估。

离线评估在推荐算法模型开发与选型的过程中对推荐算法进行评估，通过评估具体指标来选择合适的推荐算法模型，将选择好的推荐算法模型部署上线，为用户提供推荐服务。具体可以评估的指标见表6。

线上评估指推荐模型部署到线上提供真实的推荐服务后的评估。离线评估指标线上也是可以计算的，此外还有每天使用

次数、人均使用时长、点击次数和点击通过率等。

目前大部分文献仍在通过构造不同的特征提升推荐效果，在离线模型评估中较基线模型取得了更好的效果。但很少有将模型部署到线上进行实验的验证，并不能证明在真实系统中有同样良好的效果。

1.3.3 小结

本节主要讨论了社交媒体的用户画像构建技术和新闻领域的推荐系统。通过成熟的用户画像构建技术，可以标签化用户特征，分析单个用户的行为特征和群体的行为特征，得到用户的兴趣偏好列表。推荐系统帮助人们在传播过程中更精准地找到用户可能感兴趣的内容，在真正意义上完成信息的接收。推荐系统可完善用户的长期画像和短期场景，使用户获取感兴趣的信息，实现内容的精准传播。

1.4 渠道

渠道是用来传达信息的手段，为了使内容有效地传达给用户，需要仔细选择渠道。国内的社交平台主要包含微博、抖音、哔哩哔哩、小红书、知乎等。这5个社交平台目前比较有代表性。微博是国内最大的综合社交媒体，具有话题类型覆盖广、用户量大、内容形式多样的特点。抖音是国内最大的短视频平台，在国外也备受推崇，用户群体多为年轻人。短视频更符合现代人快节奏的生活，瀑布式的快速推荐方法也受到人们喜爱。哔哩哔哩则是国内最大的长视频平台，相比于短视频，长视频吸引人们的点在于内容深度和视频制作，它比短视频更能使人们共情。小红书是一个分享自己的生活、购物、美食、美妆、旅游等各种经验和心得的一个以女性用户

表6 推荐系统评估指标

指标名称	英文名称	作用	公式
精确率	Precision	推荐列表占预测推荐列表的比例	$\frac{TP}{TP+FP}$
召回率	Recall	推荐列表占真实推荐列表的比例	$\frac{TP}{TP+FN}$
F1值	F1 score	精确率和召回率的调和平均值	$2 \times \frac{Precision \times Recall}{Precision + Recall}$
微观平均 精确率	Micro Averaged Precision	计算每个用户的精确率	$\frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FP_i}$
微观平均 召回率	Micro Averaged Recall	计算每个用户的召回率	$\frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FN_i}$
微观平均 F1值	Micro Averaged F1 Score	计算每个用户的F1值	$2 \times \frac{Micro\ Averaged\ Precision \times Micro\ Averaged\ Recall}{Micro\ Averaged\ Precision + Micro\ Averaged\ Recall}$
宏观平均 精确率	Macro Averaged Precision	用户精确度的平均值	$\frac{\sum_{i=1}^n Precision_i}{n}$
宏观平均 召回率	Macro Averaged Recall	用户精确度的召回率	$\frac{\sum_{i=1}^n recall_i}{n}$
宏观平均 F1值	Macro Averaged F1 Score	用户精确度的F1值	$2 \times \frac{Macro\ Averaged\ Precision \times Macro\ Averaged\ Recall}{Macro\ Averaged\ Precision + Macro\ Averaged\ Recall}$
新闻 多样性	Diversity (News)	新闻集N多样性	$\frac{2}{p(p-1)} \sum_{n_i \in N} \sum_{n_j \in N, n_j \neq n_i} (1 - \text{sim}(n_j, n_i))$
平均均值 精度	Mean Average Precision	计算每个用户在推荐列表中的 平均精度均值	$mAP_u = \frac{\sum_{i=1}^k (P_{u_i} * rel_i)}{\min(k, N_u)}$ $\frac{1}{ U' } \sum_{u \in U'} mAP_u$

为主体，图片形式和视频形式的社交平台。知乎是国内最大的问答式社交平台，可以获得较高专业性的答案，但获取周期长，相比其他平台并不适合内容传播。

实际上，用户往往同时拥有多个社交平台的账号，不仅可以在一个平台上接收和扩散信息，还可以扮演信息传播的角色，将信息从一个平台扩散到另一个平台。由于信息扩散主要基于用户的行为，忽略了相同用户行为在不同社交平台的差异，所以缺少针对特定平台的研究。比如基于不同的特定平台对比信息的扩散深度和规模

的异同，比如来自哪些平台的信息更容易传播等。目前研究者的主要研究对象仍局限于微博平台，其他平台鲜有研究。

微博平台的研究方法通过平台上的真实信息扩散数据，基于信息级联的规模和深度符合幂律分布或正态分布的宏观特性，利用信息扩散树的指标进行研究。每一条原创微博及其传播路径构成一棵信息扩散树，它是一个有向网络，节点为微博，边为转发关系。没有入度节点的为种子节点，即原创微博。有入度节点和出度节点的为病毒节点。没有出度节点的是子节点。一

般地，扩散规模、扇出系数、可传递性、基本再生数、扩散速度和扩散深度是信息扩散树的基本指标。表7为信息扩散树相关指标及定义。

王玉等^[70]利用含有不同类URL的微博扩散深度和扩散规模的互补累积分布进行对比分析，利用假设检验的数理统计方法，比较微博单一平台和跨平台扩散规模差异。发现来自微信公众号、微博视频、微博文章和新浪新闻的信息在扩散深度和规模上相比其他类信息更有优势。

在今天，媒介即服务^[98]，人们利用信息扩散树的方法可以比较针对不同类型的信息媒体渠道的扩散异同，找到最合适的媒体终端为用户提供信息服务。

1.5 传播效果

传播效果研究的核心是借助媒介对他人的观念、态度、认知、行为等造成短期或长期的影响^[99]。根据定义可知，造成短期或长期的影响的研究本身难以量化，因此目前并没有能够直接判断效果的算法，取而代之的是通过传播学的经典理论构建传播效果指标体系。

表7 信息扩散树相关指标及定义

指标	含义
N_s	种子节点数量
N_p	叶子节点数量
N_v	病毒节点数量
N	扩散规模 $N = N_s + N_p + N_v$
k_s^{out}	种子节点出度
r_v	扇出系数 $r_v = \frac{N - N_s - k_s^{\text{out}}}{N_v}$
λ	可传递性 $\lambda = \frac{N_v}{N - N_s}$
r_0	基本再生数 $r_0 = \lambda r_v$
d	扩散深度
s	扩散速度

汤景泰等^[71]从“历时性”角度考虑议题在生命周期的5级指标：曝光、触达、互动、社群、演化。曝光指标衡量议题相关信息的数量与扩散程度，触达指标衡量议题相关信息在多大程度上被受众接收且转化为被动反馈行为，互动指标衡量议题相关信息在被接收后多大程度转化为主动反馈行为，社群指标衡量议题所计划社群的结构特征，演化指标衡量议题周期与演化模式特征。郑丽勇等^[100]以麦奎尔态度改变模型为基础，从广度、深度、强度、效度4个维度指标衡量传播效果，并听取专家意见对4个维度赋予权重。广度指受众规模；深度指受众接受的传播量大小；强度指传播的保持能力，取决于内容本身和渠道的影响；效度指使受众态度发生变化并外化于行为，取决于受众个人的决策力、消费力和二次传播力。虽然不同的研究人员有不同的指标构建方法，但是主体思路大体是从传播者、内容、渠道、受众4个方面选取指标作为评判效果的方法，再结合相关性分析或是变异系数等统计学方法优化指标，合并或剔除相关性较高的指标，实现理论体系和现实数据的适配。另外，直接判断效果的方法是利用自然语言处理技术和情感分析对比用户接收信息前后的情感态度的变化^[101-102]，或是通过比较网络论坛用户语义网络相似度测量用户之间达成共识的程度^[103]。如何全面、多维评价最终的传播效果，提升其科学性和实用性仍需要研究人员继续探索。

2 挑战与未来展望

2.1 传播过程技术应用缺陷

在现有的研究工作中，内容和受众这两

个方面已经有一些可计算方法能够被应用在传播学的研究当中，而信息传播者、渠道和传播效果的可计算方法仍需研究人员继续探索。

信息传播者的研究没有结合复杂网络理论和社交网络特性。传播者的研究主要基于网络拓扑学的方法，虽然在社交网络中有很好的普适性，但缺少社交网络自身特性，如用户行为特征和发文内容特征。通过网络结构特征可以构建不同的判断节点重要性的方法，通过用户的社交特征和内容特征可以细致地划分领域性节点重要性。如何结合已经成熟的网络拓扑学方法和社交网络独有特征共同建立用户影响力排名是研究人员可以研究的方向。

渠道方面的研究目前偏薄弱，对不同社交平台的特性掌握不足，无法对内容进行平台推荐，让内容找到最优渠道传播。在社交媒体发达的今天，各渠道有其传播的优劣势，选择最优渠道，基于特定渠道定点传播可能会成为研究人员深入探索的一个重要方向，或者充分利用不同渠道的特点，融合多渠道进行传播。文献[104]已经针对多社交网络展开影响力最大化研究，但仍缺少对不同渠道影响效果异同的评价。

传播效果的研究是复杂的、多维度的，每一个维度都可能影响传播效果。可以从传播过程的前4个维度基于信息传播者、内容、受众和渠道各部分的影响提升传播效果，也可以基于传播效果定义，利用情感变化、共识程度等直接分析。如何合理地量化效果评价指标，仍需要研究者进行更深入的思考。

2.2 大模型与计算传播学相结合的发展机遇

目前，大部分研究人员将自然语言处理技术应用于计算传播学，但传播内容是

多元化的，图像音频视频等都值得深入研究，计算视觉研究方法也开始慢慢被关注^[9]。随着数据、算力和算法的提升，大模型逐渐广泛兴起，旨在通过自监督学习从大量的数据中训练并应用于一系列下游任务中^[105]。目前大模型在自然语言处理、计算机视觉和多模态方向中都有应用，这也为计算传播领域添加了更多新的可能。通过用户描述的需求实现在社交网络中更精准的用户推荐，充分利用社交网络中文本、图像和视频多模态数据描述用户喜好并进行分类，学习不同领域的用户数据生成高品质内容，成为主题领域内优质创作者。此外，研究人员还讨论了基于图的大模型，尝试补充大语言模型中缺失的图结构信息，但在数据质量、学习范式、模型评价方法方面仍需进一步探索^[106]。目前还未出现如ChatGPT一样成功的图大模型，未来或许可以通过借鉴大语言模型的方法，在之后几年出现一款图大模型的应用，能够被成功应用于社交网络。

3 结束语

计算传播学发展至今仍未超过20年，在大数据时代的推动下，借助人工智能技术和传播学知识应用于多个领域，对于理解和引导公众意见、优化信息发布策略等都具有重要的社会价值。近年来虽然有众多研究者研究计算传播学这个领域，但都是从宏观角度研究计算传播学，以计算传播学范式研究、场景应用和发展路径为重点进行综述^[6, 12]。而笔者从计算机技术角度入手，通过梳理已有的相关工作，按照传播过程的5个维度，较系统地介绍了基于不同维度的计算方法，阐述了不同计算方法的思路和需要的数据集以及各个计算方

法的评价方法,提出了一种计算传播技术体系架构。本综述与计算传播相关领域综述最大的不同在于聚焦计算传播过程中的技术层面,跳出对计算传播的理论研究、场景应用和发展路径的探索,结合大数据时代蓬勃涌现的各种计算方法技术梳理相关技术点。希望新的研究者在已有研究工作的基础上,在不同传播过程维度中提出新的研究方法,建立更有效的模型,更好地发挥计算机技术在传播学的应用。

参考文献:

- [1] LAZER D, PENTLAND A, ADAMIC L, et al. Computational social science[J]. *Science*, 2009, 323(5915): 721-723.
- [2] SALGANIK M J. Bit by bit: social research in the digital age[M]. Princeton: Princeton University Press, 2019.
- [3] 祝建华, 彭泰权, 梁海, 等. 计算社会科学在新闻传播研究中的应用[J]. *科研信息化技术与应用*, 2014, 5(2): 3-13.
ZHU J H, PENG T Q, LIANG H, et al. Computational social science in communication research[J]. *E-science Technology & Application*, 2014, 5(2): 3-13.
- [4] 王成军. 计算传播学的起源、概念和应用[J]. *编辑学刊*, 2016(3): 59-64
WANG C J. Origins, concepts and applications of computational communication[J]. *Editors Bimonthly*, 2016(3): 59-64.
- [5] VAN ATTEVELDT W, PENG T Q. When communication meets computation: opportunities, challenges, and pitfalls in computational communication science[J]. *Communication Methods and Measures*, 2018, 12(2/3): 81-92.
- [6] 巢乃鹏, 吴兴桐, 黄文森, 等. 计算传播学研究现状与前沿议题[J]. *全球传媒学刊*, 2022, 9(1): 19-40.
CHAO N P, WU X T, HUANG W S, et al. Research status and frontier issues of computational communication research[J]. *Global Journal of Media Studies*, 2022, 9(1): 19-40.
- [7] 张敬玮. 文献计量视域下计算传播学发展现状及未来发展趋势思考[D]. 西安: 陕西师范大学, 2021.
ZHANG J W. Reflections on the current situation and future development trend of computational communication from the perspective of bibliometrics[D]. Xi'an: Shaanxi Normal University, 2021.
- [8] 吴晔, 李小军. 2022年计算传播学研究综述[J]. *教育传媒研究*, 2021(1): 41-45.
WU Y, LI X J. 2022 Review of research in computational communication[J]. *Journal of Education and Media studies*, 2021(1): 41-45.
- [9] 李华, 吴晔, 陈艳培, 等. 审视本体、扩容议题与创新方法: 走向“学科化”——2023年计算传播学研究综述[J]. *教育传媒研究*, 2024(1): 44-49.
LI H, WU Y, CHEN Y P, et al. Review - ing ontologies, expanding issues, and innovating methods: towards “disciplin - arisation” - a review of computational communication research in 2023[J]. *Journal of Education and Media studies*, 2024(1): 44-49.
- [10] 刘庆振. 计算传播学: 智能媒体视阈下传播学研究的新范式[J]. *教育传媒研究*, 2018(6): 21-25.
LIU Q Z. Computational communication: a new paradigm of communication re - search from the perspective of intelli - gent media[J]. *Journal of Education and Media Studies*, 2018(6): 21-25.
- [11] 赵甜芳. 计算传播研究的知识地图与实证框架——基于12991篇中外核心文献的对比分析[J]. *新闻与传播研究*, 2023, 30(3): 40-59, 126-127.
ZHAO T F. Knowledge map and empiri - cal framework of computational commu -

- nication research - a comparative analysis based on 12991 Chinese and foreign core Literatures[J]. Journalism and Communication, 2023, 30(3): 40-59, 126-127.
- [12] 韩少卿, 巢乃鹏. 复杂性转向: 计算传播学的发生逻辑[J]. 南京社会科学, 2024(3): 109-122.
HAN S Q, CHAO N P. The complexity turn: the generative logic of computational communication science[J]. Nanjing Journal of Social Sciences, 2024(3): 109-122.
- [13] LASWELL H D. The structure and function of communication in society[J]. The communication of ideas, 1948.
- [14] BONACICH P. Factoring and weighting approaches to status scores and clique identification[J]. The Journal of Mathematical Sociology, 1972, 2(1): 113-120.
- [15] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1): 35.
- [16] BAVELAS A. Communication patterns in task-oriented groups[J]. The Journal of the Acoustical Society of America, 1950, 22(6): 725-730.
- [17] 王建伟, 荣莉莉, 郭天柱. 一种基于局部特征的网络节点重要性度量方法[J]. 大连理工大学学报, 2010(5): 822-826.
WANG J W, RONG L L, GUO T Z. A new measure method of network node importance based on local characteristics[J]. Journal of Dalian University of Technology, 2010(5): 822-826.
- [18] CHEN D B, LYV L Y, SHANG M S, et al. Identifying influential nodes in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2012, 391(4): 1777-1787.
- [19] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6: 888-893.
- [20] NIE T Y, GUO Z, ZHAO K, et al. Using mapping entropy to identify node centrality in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 453: 290-297.
- [21] YANG H H, AN S. Critical nodes identification in complex networks[J]. Symmetry, 2020, 12(1): 123.
- [22] 杨松青, 蒋沅, 童天驰, 等. 基于 Tsallis 熵的复杂网络节点重要性评估方法[J]. 物理学报, 2021, 70(21): 273-284.
YANG S Q, JIANG Y, TONG T C, et al. A method of evaluating importance of node in complex network based on Tsallis entropy[J]. Acta Physica Sinica, 2021, 70(21): 273-284.
- [23] RUAN Y R, TANG J, HU Y L, et al. Efficient algorithm for the identification of node significance in complex network[J]. IEEE Access, 2020, 8: 28947-28955.
- [24] 朱敬成, 刘辉, 王伦文, 等. 基于网络拓扑重度的关键节点识别方法[J]. 计算机应用研究, 2021, 38(12): 3581-3585.
ZHU G C, LIU H, WANG L W, et al. Method for identifying key nodes based on overlap of network topology[J]. Application Research of Computers, 2021, 38(12): 3581-3585.
- [25] LESKOVEC J, KRAUSE A, GUESTRIN C, et al. Cost-effective outbreak detection in networks[C]//Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2007: 420-429.
- [26] GOYAL A, LU W, LAKSHMANAN L V S. CELF++: optimizing the greedy algorithm for influence maximization in social networks[C]//Proceedings of the 20th International Conference Companion on World Wide Web. New York: ACM, 2011: 47-48.
- [27] CHEN W, WANG Y J, YANG S Y. Effi-

- cient influence maximization in social networks[C]//Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2009: 199-208.
- [28] CHEN W, WANG C, WANG Y J. Scalable influence maximization for prevalent viral marketing in large-scale social networks[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2010: 1029-1038.
- [29] JUNG K, HEO W, CHEN W. IRIE: scalable and robust influence maximization in social networks[C]//Proceedings of the 2012 IEEE 12th International Conference on Data Mining. Piscataway: IEEE Press, 2012: 918-923.
- [30] 曹玖新, 董丹, 徐顺, 等. 一种基于 k -核的社会网络影响最大化算法[J]. 计算机学报, 2015, 38(2): 238-248.
- CAO J X, DONG D, XU S, et al. A k -core based algorithm for influence maximization in social networks[J]. Chinese Journal of Computers, 2015, 38(2): 238-248.
- [31] 王璿, 张瑜, 周军锋, 等. 基于社交网络的影响力最大化算法[J]. 通信学报, 2022, 43(8): 151-163.
- WANG X, ZHANG Y, ZHOU J F, et al. Influence maximization algorithm based on social network[J]. Journal on Communications, 2022, 43(8): 151-163.
- [32] ALLAN J. Introduction to topic detection and tracking[M]. Boston: Springer, 2002: 1-16.
- [33] 刘星星, 何婷婷, 龚海军, 等. 网络热点事件发现系统的设计[J]. 中文信息学报, 2008, 22(6): 80-85.
- LIU X X, HE T T, GONG H J, et al. Design of hot web event detection system[J]. Journal of Chinese Information Processing, 2008, 22(6): 80-85.
- [34] 罗亚平, 王枫, 周延泉. 基于关注度的热点话题发现模型[C]//第七届中文信息处理国际会议论文集. 武汉: 中国中文信息学会, 2007.
- LUO Y P, WANG C, ZHOU Y Q. Study on hot topic discovery model based on attention degree[C]//Proceedings of Seventh International Conference on Chinese Language Information Processing. Wuhan: Chinese Information Processing Society of China, 2007.
- [35] NIE X C, ZHANG W S, ZHANG Y, et al. Method to predict bursty hot events on twitter based on user relationship network [J]. IEEE Access, 2020, 8: 44031-44040.
- [36] BAI J L, GUO J, CHEN G, et al. An efficient algorithm of hot events detection in text streams[C]//Proceedings of the 2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery. Piscataway: IEEE Press, 2010: 321-326.
- [37] 贺敏, 王丽宏, 杜攀, 等. 基于有意义串聚类的微博热点话题发现方法[J]. 通信学报, 2013, 34(S1): 256-262.
- HE M, WANG L H, DU P, et al. Microblog hot topic detection method based on meaningful string clustering[J]. Journal on Communications, 2013, 34(S1): 256-262.
- [38] DOULAMIS N D, DOULAMIS A D, KOKKINOS P, et al. Event detection in twitter microblogging[J]. IEEE Transactions on Cybernetics, 2016, 46(12): 2810-2824.
- [39] LIU Y, LUO X F. A framework for online hot event discovery on the web[C]//Proceedings of the 2013 IEEE 16th International Conference on Computational Science and Engineering. Piscataway: IEEE Press, 2013: 989-996.
- [40] SHEN P, ZHOU Y, CHEN K. A probability based subnet selection method for hot event detection in Sina Weibo

- microblogging[C]//Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York: ACM, 2013: 1410–1413.
- [41] TONG H Q, LIU Y, PENG H, et al. Internet users' psychosocial attention prediction: web hot topic prediction based on adaptive AR model[C]//Proceedings of the 2008 International Conference on Computer Science and Information Technology. Piscataway: IEEE Press, 2008: 458–462.
- [42] MA X, GAO X F, CHEN G H. BEEP: a Bayesian perspective early stage event prediction model for online social networks[C]//Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE Press, 2017: 973–978.
- [43] LIU R F, WANG J, ZHANG M. A HMM_based hot topic lifecycle prediction model[C]//Proceedings of the 2011 International Conference on Electric Information and Control Engineering. Piscataway: IEEE Press, 2011: 2881–2884.
- [44] SZABO G, HUBERMAN B A. Predicting the popularity of online content[J]. Communications of the ACM, 2010, 53(8): 80–88.
- [45] PINTO H, ALMEIDA J M, GONÇALVES M A. Using early view patterns to predict the popularity of youtube videos[C]//Proceedings of the 6th ACM International Conference on Web Search and Data Mining. New York: ACM, 2013: 365–374.
- [46] 高金华, 沈华伟, 程学旗, 等. 基于相似消息的流行度预测方法[J]. 中文信息学报, 2018, 32(11): 79–85.
- GAO J H, SHEN H W, CHENG X Q, et al. Popularity prediction based on similar historical tweets[J]. Journal of Chinese Information Processing, 2018, 32(11): 79–85.
- [47] SHEN H W, WANG D S, SONG C M, et al. Modeling and predicting popularity dynamics via reinforced Poisson processes [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2014, 28(1).
- [48] GAO S, MA J, CHEN Z M. Modeling and predicting retweeting dynamics on microblogging platforms[C]//Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. New York: ACM, 2015: 107–116.
- [49] MISHRA S, RIZOIU M A, XIE L X. Feature driven and point process approaches for popularity prediction[C]//Proceedings of the 25th ACM International Conference on Information and Knowledge Management. New York: ACM, 2016: 1069–1078.
- [50] LIAO D L, XU J, LI G F, et al. Popularity prediction on online articles with deep fusion of temporal process and content features[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 200–207.
- [51] ZHANG W, WANG W, WANG J, et al. User-guided hierarchical attention network for multi-modal social image popularity prediction[C]//Proceedings of the 2018 World Wide Web Conference on World Wide Web – WWW '18. New York: ACM, 2018: 1277–1286.
- [52] XU K L, LIN Z M, ZHAO J Q, et al. Multi-modal deep learning for social media popularity prediction with attention mechanism[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 4580–4584.
- [53] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C]//Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 675–684.
- [54] GIACHANOU A, ROSSO P, CRESTANI F.

- Leveraging emotional signals for credibility detection[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2019: 877-880.
- [55] QI P, CAO J, YANG T Y, et al. Exploiting multi-domain visual information for fake news detection[C]//Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE Press, 2019: 518-527.
- [56] WANG Y Q, MA F L, JIN Z W, et al. EANN: event adversarial neural networks for multi-modal fake news detection[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 849-857.
- [57] KHATTARD, GOUD JS, GUPTA M, et al. MVAE: multimodal variational autoencoder for fake news detection[C]//Proceedings of the World Wide Web Conference. New York: ACM, 2019: 2915-2921.
- [58] 刘海鸥, 孙晶晶, 苏妍嫒, 等. 基于用户画像的旅游情境化推荐服务研究[J]. 情报理论与实践, 2018, 41(10): 87-92.
- LIU H O, SUN J J, SU Y Y, et al. Research on the tourism situational recommendation service based on persona[J]. Information Studies (Theory & Application), 2018, 41(10): 87-92.
- [59] 汪强兵, 章成志. 基于手势行为的社交网络用户兴趣画像构建及应用[J]. 图书与情报, 2019(2): 114-119, 132.
- WANG Q B, ZHANG C Z. Construction and application of user interest profile in social networks based on gesture behavior[J]. Library & Information, 2019(2): 114-119, 132.
- [60] TAHAR R, SAMIR K. A geo-social user profile for a personalized information retrieval[C]//Proceedings of the 2016 8th International Conference on Information Management and Engineering. New York: ACM, 2016: 62-66.
- [61] 林燕霞, 谢湘生. 基于社会认同理论的微博群体用户画像[J]. 情报理论与实践, 2018, 41(3): 142-148.
- LIN Y X, XIE X S. User portrait of diversified groups in micro-blog based on social identity theory[J]. Information Studies (Theory & Application), 2018, 41(3): 142-148.
- [62] 张亚楠, 黄晶丽, 王刚. 考虑全局和局部信息的科研人员科研行为立体精准画像构建方法[J]. 情报学报, 2019, 38(10): 1012-1021.
- ZHANG Y N, HUANG J L, WANG G. A method considering local and global information for constructing stereoscopic and accurate portraits of scientific researchers[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(10): 1012-1021.
- [63] 张艳丰, 彭丽徽, 刘金承, 等. 新媒体环境下移动社交媒体倦怠用户画像实证研究: 基于SSO理论的因果关系视角[J]. 情报学报, 2019, 38(10): 1092-1101.
- ZHANG Y F, PENG L H, LIU J C, et al. An empirical study on mobile social media fatigue user portraits in the new media environment: a causality perspective based on SSO theory[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(10): 1092-1101.
- [64] DARVISHY A, IBRAHIM H, SIDI F, et al. HYPNER: a hybrid approach for personalized news recommendation[J]. IEEE Access, 2020, 8: 46877-46894.
- [65] 王嵘冰, 徐红艳, 冯勇, 等. 基于相似主题和HITS的微博用户推荐算法研究[J]. 中文信息学报, 2019, 33(7): 128-135.
- WANG R B, XU H Y, FENG Y, et al. Microblog user recommendation algorithm based on similar topics and HITS[J].

- Journal of Chinese Information Processing, 2019, 33(7): 128–135.
- [66] OKURA S, TAGAMI Y, ONO S, et al. Embedding-based news recommendation for millions of users[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017: 1933–1942.
- [67] DE S P M G. Chameleon: a deep learning meta-architecture for news recommender systems[C]//Proceedings of the 12th ACM Conference on Recommender Systems. New York: ACM, 2018: 578–583.
- [68] 刘树栋, 张可, 陈旭. 基于多维度兴趣注意力和用户长短期偏好的新闻推荐[J]. 中文信息学报, 2022, 36(9): 102–111.
- LIU S D, ZHANG K, CHEN X. Multi-dimensional interest-attention-based news recommendation with long and short-term user preferences[J]. Journal of Chinese Information Processing, 2022, 36(9): 102–111.
- [69] CHEN C, MENG X W, XU Z H, et al. Location-aware personalized news recommendation with deep semantic analysis[J]. IEEE Access, 2017, 5: 1624–1638.
- [70] 王玉, 许楠楠, 胡海波. 社交媒体中的跨平台信息扩散特征及机制[J]. 复杂系统与复杂性科学, 2022, 19(4): 7–16.
- WANG Y, XU N N, HU H B. Characteristics and mechanisms of cross-platform information diffusion in social media[J]. Complex Systems and Complexity Science, 2022, 19(4): 7–16.
- [71] 汤景泰, 徐铭亮. 指标构建与核心流程: 社交媒体内容传播效果的综合评估[J]. 现代传播(中国传媒大学学报), 2022, 44(4): 37–46.
- TANG J T, XU M L. Index construction and core process: comprehensive evaluation of social media content communication effect[J]. Modern Communication (Journal of Communication University of China), 2022, 44(4): 37–46.
- [72] ADDIS B, ARINGHERI R, GROSSO A, et al. Hybrid constructive heuristics for the critical node problem[J]. Annals of Operations Research, 2016, 238(1): 637–649.
- [73] 韩忠明, 陈炎, 刘雯, 等. 社会网络节点影响力分析研究[J]. 软件学报, 2017, 28(1): 84–104.
- HAN Z M, CHEN Y, LIU W, et al. Research on node influence analysis in social networks[J]. Journal of Software, 2017, 28(1): 84–104.
- [74] DOMINGOS P, RICHARDSON M. Mining the network value of customers[C]//Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2001: 57–66.
- [75] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network[C]//Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '03. New York: ACM, 2003: 137–146.
- [76] 曹玖新, 闵绘宇, 徐顺, 等. 基于启发式和贪心策略的社交网络影响最大化算法[J]. 东南大学学报(自然科学版), 2016, 46(5): 950–956.
- CAO J X, MIN H Y, XU S, et al. Mixed heuristic and greedy strategies based algorithm for influence maximization in social networks[J]. Journal of Southeast University (Natural Science Edition), 2016, 46(5): 950–956.
- [77] BORGS C, BRAUTBAR M, CHAYES J, et al. Maximizing social influence in nearly optimal time[C]//Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2014: 946–957.

- [78] TANG Y Z, XIAO X K, SHI Y C. Influence maximization: near-optimal time complexity meets practical efficiency[C]// Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2014: 75-86.
- [79] 吴安彪, 袁野, 乔百友, 等. 大规模时序图影响力最大化的算法研究[J]. 计算机学报, 2019, 42(12): 2647-2664.
- WU A B, YUAN Y, QIAO B Y, et al. The influence maximization problem based on large-scale temporal graph[J]. Chinese Journal of Computers, 2019, 42(12): 2647-2664.
- [80] 许小可, 胡海波, 张伦, 等. 社交网络上的计算传播学[M]. 北京: 高等教育出版社, 2015: 71-72.
- [81] 徐佳俊, 杨颀, 姚天昉, 等. 基于LDA模型的论坛热点话题识别和追踪[J]. 中文信息学报, 2016, 30(1): 43-49.
- XU J J, YANG Y, YAO T F, et al. LDA based hot topic detection and tracking for the forum[J]. Journal of Chinese Information Processing, 2016, 30(1): 43-49.
- [82] 杨亮, 林原, 林鸿飞. 基于情感分布的微博热点事件发现[J]. 中文信息学报, 2012, 26(1): 84-90, 109.
- YANG L, LIN Y, LIN H F. Micro-blog hot events detection based on emotion distribution[J]. Journal of Chinese Information Processing, 2012, 26(1): 84-90, 109.
- [83] ROSENBERG A, HIRSCHBERG J. V-Measure: a conditional entropy-based external cluster evaluation measure[J]. EMNLP-CoNLL 2007 - Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2007: 410-420.
- [84] JIN Z W, CAO J, ZHANG Y D, et al. Novel visual and statistical image features for microblogs news verification[J]. IEEE Transactions on Multimedia, 2017, 19(3): 598-608.
- [85] TAHMASEBIS, HAKIMOV, EWERTHR, et al. Improving generalization for multimodal fake news detection[C]// Proceedings of the 2023 ACM International Conference on Multimedia Retrieval. New York: ACM, 2023: 581-585.
- [86] DUN Y Q, TU K F, CHEN C, et al. KAN: knowledge-aware attention network for fake news detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(1): 81-89.
- [87] COOPER A. The inmates are running the asylum[M]. Indianapolis: Sams, 1999.
- [88] 赵雅慧, 刘芳霖, 罗琳. 大数据背景下的用户画像研究综述: 知识体系与研究展望[J]. 图书馆学研究, 2019(24): 13-24.
- ZHAO Y H, LIU F L, LUO L. A review of user profile in the context of big data: knowledge system and research prospect[J]. Research on Library Science, 2019(24): 13-24.
- [89] LAINÉ-CRUZEL S, LAFOUGE T, LARDY J P, et al. Improving information retrieval by combining user profile and document segmentation[J]. Information Processing & Management, 1996, 32(3): 305-315.
- [90] 王凌霄, 沈卓, 李艳. 社会化问答社区用户画像构建[J]. 情报理论与实践, 2018, 41(1): 129-134.
- WANG L X, SHEN Z, LI Y. User profiling of social Q & A community[J]. Information Studies (Theory & Application), 2018, 41(1): 129-134.
- [91] 徐海玲, 张海涛, 魏明珠, 等. 社交媒体用户画像的构建及资源聚合模型研究[J]. 图书情报工作, 2019, 63(9): 109-115.
- XU H L, ZHANG H T, WEI M Z, et al. Research on the construction of social media user portrait and resource aggregation model[J]. Library and Information

- Service, 2019, 63(9): 109-115.
- [92] 刘庆振, 于进, 牛新权. 计算传播学: 智能媒体时代的传播学研究新范式[M]. 北京: 人民日报出版社, 2019: 79-82.
- [93] 孟祥福, 霍红锦, 张霄雁, 等. 个性化新闻推荐方法研究综述[J]. 计算机科学与探索, 2023, 17(12): 2840-2860.
- MENG X F, HUO H J, ZHANG X Y, et al. Survey of research on personalized news recommendation approaches[J]. Journal of Frontiers of Computer Science and Technology, 2023, 17(12): 2840-2860.
- [94] ZHU Z L, LI D Y, LIANG J, et al. A dynamic personalized news recommendation system based on BAP user profiling method[J]. IEEE Access, 2018, 6: 41068-41078.
- [95] MOREIRA G D S P, JANNACH D, CUNHA A M D. Contextual hybrid session-based news recommendation with recurrent neural networks[J]. IEEE Access, 2019, 7: 169185-169203.
- [96] 张玉朋, 李香菊, 李超, 等. 基于Transformer与异质图神经网络的新闻推荐模型[J]. 模式识别与人工智能, 2022, 35(9): 839-848.
- ZHANG Y P, LI X J, LI C, et al. News recommendation model based on transformer and heterogenous graph neural network[J]. Pattern Recognition and Artificial Intelligence, 2022, 35(9): 839-848.
- [97] 黄震华, 林威, 吴正洋, 等. 基于最优传输和知识回放的新闻推荐模型增量学习方法[J]. 计算机学报, 2023, 46(10): 2161-2177.
- HUANG Z H, LIN W, WU Z Y, et al. Incremental news recommendation method based on Optimal Transport and knowledge replay[J]. Chinese Journal of Computers, 2023, 46(10): 2161-2177.
- [98] 夏洪波. 全球品牌工程[J]. 中国广告, 2018(5): 34-36.
- XIA H B. Global branding project[J]. China Advertising, 2018(5): 34-36.
- [99] VALKENBURG P M, PETER J, WALTHER J B. Media effects: theory and research[J]. Annual Review of Psychology, 2016, 67: 315-338.
- [100] 郑丽勇, 郑丹妮, 赵纯. 媒介影响力评价指标体系研究[J]. 新闻大学, 2010 (1): 121-126.
- ZHENG L Y, ZHENG D N, ZHAO C. Research on media influence evaluation indicator system[J]. Journalism Research, 2010 (1): 121-126.
- [101] 王春东, 张卉, 莫秀良, 等. 微博情感分析综述[J]. 计算机工程与科学, 2022, 44(1): 165-175.
- WANG C D, ZHANG H, MO X L, et al. Overview on sentiment analysis of microblog[J]. Computer Engineering & Science, 2022, 44(1): 165-175.
- [102] 王璐, 马宏伟, 吕欢欢. 方面级文本情感分析综述[J]. 计算机应用, 2022, 42(S2): 1-9.
- WANG L, MA H W, LYU H H. A summary of aspect-level text emotion analysis[J]. Journal of Computer Applications, 2022, 42(S2): 1-9.
- [103] LIANG H. Coevolution of political discussion and common ground in web discussion forum[J]. Social Science Computer Review, 2014, 32(2): 155-169.
- [104] 李国良, 楚娅萍, 冯建华, 等. 多社交网络的影响力最大化分析[J]. 计算机学报, 2016, 39(4): 643-656.
- LI G L, CHU Y P, FENG J H, et al. Influence maximization on multiple social networks[J]. Chinese Journal of Computers, 2016, 39(4): 643-656.
- [105] BOMMASANIR, HUDSON D A, ADELIE, et al. On the opportunities and risks of foundation models[EB]. arXiv preprint, 2021, arXiv: 2108.07258.
- [106] LIU J W, YANG C, LU Z Y, et al. Towards graph foundation models: a survey and beyond[EB]. arXiv preprint, 2023, arXiv: 2310.11829.

作者简介



王续澎（1998-），男，中国科学院计算机网络信息中心助理工程师，主要研究方向为数据挖掘、机器学习。



何洪波（1970-），男，中国科学院计算机网络信息中心高级工程师、硕士生导师，主要研究方向为新媒体技术应用、互联网数据挖掘和信息推荐。



王闰强（1976-），男，中国科学院计算机网络信息中心正高级工程师、新媒体技术与应用发展部主任，主要研究方向为新媒体科学传播与教育技术、应用、服务研究和实践。

收稿日期: 2024-09-12

通信作者: 王续澎, xpwang@cnic.cn

基金项目: 中国科学院网络安全和信息化专项(No.CAS-WX2022GC-030402)

Foundation Item: Cyber Security and Informatization Project, Chinese Academy of Sciences (No.CAS-WX2022GC-030402)