引用格式：王妍，杨妹，王传彪，崔鸣宇. 数据驱动下的传播动力学：前沿探索与趋势分析[J]. 中国传媒大学学报（自然科学版），2024，31（04）:01-09+54.
文章编号：1673-4793（2024）04-0001-10

数据驱动下的传播动力学：前沿探索与趋势分析

王妍[1,2]，杨妹[1,2]，王传彪[1,2*]，崔鸣宇[1,2]
（1. 中国传媒大学数据科学与智能媒体学院，北京 100024；
2. 中国传媒大学融合与传播国家重点实验室，北京 100024）

摘要：随着互联网和社交媒体的迅猛发展，信息传播和舆情形成机制变得日益复杂。本文系统地回顾了信息传播动力学舆情仿真领域的研究进展，重点探讨了不同类型的真实数据如何与传播动力学模型结合，机器学习在参数估计和传播规律预测中的应用，以及传播动力学与大模型的结合在信息传播研究中展现出的广阔前景。本文通过分析大量传播动力学文献，提出了优化模型复杂性与计算成本、提升数据质量、实时数据处理等方面的建议。
关键词：传播动力学；复杂网络；信息传播；机器学习；大模型
中图分类号：TP391.4　**文献标识码：**A

Communication dynamics driven by data: frontier exploration and trend analysis

WANG Yan[1,2], YANG Mo[1,2], WANG Chuanbiao[1,2*], CUI Mingyu[1,2]
(1.School of Data Science and Media Intelligence, Communication University of China, Beijing 100024, China; 2. State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024, China)

Abstract: With the rapid development of the Internet and social media, the mechanism of information dissemination and public opinion formation has become increasingly complex.This article systematically reviews the research progress in the field of information dissemination dynamics and public opinion simulation, focusing on how different types of real data can be combined with communication dynamics models, the application of machine learning in parameter estimation and propagation law prediction, and the broad prospects of combining communication dynamics with large models in information dissemination research. This article proposes suggestions for optimizing model complexity and computational costs, improving data quality, and real-time data processing through an analysis of a large number of communication dynamics literature.
Keywords: propagation dynamics; complex network; information dissemination; machine learning; large language model

1　引言

随着互联网和社交媒体的迅猛发展，信息传播和舆情形成机制变得日益复杂。信息的快速传播和舆情的迅速扩散不仅对社会治理和舆论引导提出了新的挑战，也为信息传播动力学与舆情预测研究提供了

基金项目：国家自然科学基金（72274182）
作者简介（*为通讯作者）：王妍（1979-），女，博士后，教授，博士生导师，主要从事复杂网络与信息计算、媒体大数据分析与评价、舆情动力学建模与仿真研究。Email：wy@cuc.edu.cn；杨妹（1999-），女，硕士研究生，主要从事信息传播动力学研究。Email：2022200810j5002@cuc.edu.cn；王传彪（1984-），男，博士，副教授，硕士生导师，主要从事动力系统研究。Email: cbwang@cuc.edu.cn；崔鸣宇（2000-），男，博士研究生，主要从事信息传播动力学研究。Email：myCui@cuc.edu.cn

广阔的发展空间。信息传播动力学不再局限于在人造网络进行仿真,而是结合真实数据和大语言模型进行拟合和预测,更贴合实际地探讨信息在网络中的传播规律、舆情形成机制及其对社会的影响,以及如何利用这些规律和机制进行舆情预测和干预,可以使动力学模型发挥更多实际价值。

本文旨在对信息传播动力学舆情仿真及预测方法的研究进展进行全面梳理和综述。通过回顾现有研究,分析传播动力学研究面临的挑战,提出一些切实可行的建议。

研究问题如下:

(1)不同类型真实数据如何与传播动力学模型结合?

(2)如何利用机器学习处理已有数据进行参数估计及传播规律预测?

(3)传播动力学与大模型结合的未来前景如何?

本文针对以上问题,梳理分析了42篇文献形成综述,以期对在数据应用范围内进行的信息传播动力学的仿真研究有所帮助。首先,从信息传播动力学建模和舆情仿真的基本理论出发,介绍传播动力学模型的仿真分析流程,探讨真实数据如何用于建模拟合。其次,详细介绍了机器学习与信息传播动力学结合在舆情预测中的使用,总结了大模型对传播场景拟真与预测方面的应用。文末对当前信息传播动力学舆情仿真、拟真及预测方法存在的问题和挑战进行总结讨论,阐述了传播动力学与数据结合的一些潜在领域,并展望未来的研究方向和发展趋势。

本文以下章节内容如下:第2章介绍了仿真实验的方法。第3章介绍了真实数据的具体应用。第4章介绍了使用机器学习结合传播动力学模型做数据拟合及预测的应用方法。第5章介绍了使用大语言模型模拟及预测信息传播的应用方法。

2 舆情传播动力学仿真方法

在线社交网络上的信息传播过程分析和规律挖掘是研究网络信息传播规律的重要基础。数学模型构建后通常利用计算机程序来实现数学模型的仿真。这些程序可以模拟个体或群体的行为,以及它们之间的相互作用。

模型仿真首先需要考虑用户和关系形成的网络结构,根据模型适用场景选择合适的网络模型,常用的网络模型主要包括随机图模型、小世界模型和无标度网络模型。随机图模型(Erdős-Rényi模型[1]):节点之间的边是随机连接的,用于模拟人们之间随机互动的情况。小世界模型(Watts-Strogatz模型[2]):具有高聚类系数和短路径长度,适合模拟社会关系(如朋友、同事)形成的网络中的信息传播。无标度网络模型(Barabási-Albert模型[3]):节点度分布服从幂律分布,适合模拟社交媒体平台(如微博、Twitter)的用户和他们之间的连接(如好友关系、关注关系)形成的网络中的信息传播。网络中包含节点和连边,节点代表用户,连边表示关系。以边的权重和属性代表用户之间各种各样的关系以及信息的传递。边的权重可以表示传播的概率或强度,不同权重的边可以影响传播路径和速度。边的方向可以表示为信息传递的方向、关注关系、好友请求等含义。不同的网络结构(如社交网络、交通网络)对传播过程有显著影响,仿真时需首先考虑网络的拓扑特性。

其次,研究者需要将模型机制、状态转移规则写入程序中,而后根据研究问题进行数学推导结果验证、敏感性分析、情景分析、导控效果评估等方面分析,将仿真结果以图表等形式可视化,以便于直观理解和规律挖掘。数学推导结果验证即将基本再生数、稳态分析、最优控制等计算结果用仿真模拟验证是否正确。敏感性分析即评估模型对不同参数变化的敏感性,以确定哪些因素对传播过程影响较大。情景分析即通过改变模型中的某些参数或假设,模拟不同的传播情景,以预测可能的结果。导控效果评估即利用仿真结果来评估不同干预措施的效果,如强制隔离、定期科普、真相宣传等。

3 真实数据建模拟合

使用真实数据建模拟合不仅是让舆情传播模型仿真过程符合实际情况的方法,也是验证舆情传播模型可靠性、有效性的重要手段。真实数据的应用包含以下几种情况:在仿真过程中代入真实舆情数据计算,在仿真过程中代入真实网络数据计算,在仿真过程中仅用真实数据量化某个或某些参数,将仿真的结果与真实的结果比较分析。数据相关信息如表1所示。

3.1 真实舆情数据

在传播动力学研究中,将真实舆情数据直接代入仿真模型进行计算,是提高研究可信度和结果实用性的关键策略。这种做法通过捕捉真实世界中的舆情动态,使得仿真结果更贴近实际情境,增强了研究的外部有效性。通过采集和整理社交媒体评论、新闻报道等多渠道的真实数据,能够更全面地理解信息传播

表1 真实数据拟合研究

拟合方式	数据来源	数据类型	数据真实性	年份	作者	参考文献
真实舆情数据	微博	舆情数据	真实	2012	Zhao等	[4]
	微博等多个平台	舆情数据		2017	夏一雪等	[5]
	微博	传播曲线		2019	黄远等	[6]
		舆情数据（点赞、评论、转发数据）	谣言	2021	Yu等	[7]
		舆情数据（用户名、微博内容和发布时间）	真实	2021	庄文英等	[8]
	Twitter	舆情数据	谣言	2021	Jiang等	[9]
真实网络数据	Facebook	网络数据（用户关系数据集）	真实	2015	Xia等	[11]
	Twitter			2015	尹熙成等	[12]
	Facebook			2017	刘泉等	[13]
				2018	Yang等	[14]
				2022	Jing等	[15]
	Twitter			2022	Wang等	[16]
真实参数数据	微博	用户发送信息时间间隔数据	真实	2012	Yan等	[10]
	其他	关键参数值		2013	丁金珠等	[17]
	其他	参数初始值	谣言	2018	项权等	[18]
	微博	关键参数值		2021	Wang等	[25]
		部分参数值	真实	2022	Geng等	[19]
与真实数据比较	微博	传播曲线	谣言	2014	Zhang等	[20]
	百度	百度指数	真实	2015	陈福集等	[21]
	微博			2015	尹熙成等	[12]
		传播曲线		2017	崔金栋等	[22]
		评论趋势		2019	黄远等	[6]
		传播曲线		2021	Yu等	[7]
		意见演化趋势		2021	庄文英等	[8]
	Twitter	传播曲线		2022	Eminente	[23]

的机制和影响因素。Zhao等[4]采集上海新闻可信度的问卷调查数据以及流行社交网络Ifeng网站、非官方网络上“日本的核危机引发并加剧了中国购买碘盐的恐慌”事件的数据，并将其代入仿真过程。夏一雪等[5]通过舆情监测软件统计6个网络传播平台涉及“高考”的网络舆情数据并用这些数据进行建模分析。黄远等[6]把“杭州公交纵火案”事件的信息、心理、主流观点的真实数据代入模型中，实现舆情传播的仿真。Yu等[7]把微博上“大象踩踏游客”事件发生后每小时点赞、评论、转发的真实数据代入仿真过程，模拟出该谣言的演变过程。庄文英等[8]通过Python软件采集了微博“中行原油宝”事件的用户名、微博内容和发布时间等数据代入仿真过程，计算出该事件的实际演化趋势并模拟出该事件的模型仿真演化趋势。Jiang等[9]收集Twitter上哈维飓风期间，“在允许你进入避难所之前，必须检查移民身份”的谣言数据，代入模型的仿真过程中。Yan等[10]收集新浪微博用户发送的信息数据，研究评论数量和转发次数与用户发布信息的时间间隔的关系。

3.2 真实网络数据

代入真实网络结构数据进行计算是提高仿真可靠性和符合实际情况的有效手段。相较于简单地带入网络模型，直接使用真实网络结构数据更能反映出网络的真实结构特征，从而增强了仿真过程的现实性和可信度，使得结果更具有说服力。在这个过程中，可以更全面地考虑网络中节点之间的真实交互关系，捕捉信息传播的更为精细的路径和影响。这种方法不仅关注节点之间的直接联系，还能够揭示网络中的隐含关系和复杂结构，为仿真结果提供更深刻的理解。Xia等[11]在仿真过程中，代入Facebook的真实网络数据组成子网络，研究在真实网络上SEIR（susceptible exposed infectious removed）模型的动力学。尹熙成等[12]在仿真过程中，代入Twitter用户关系数据集，对其剪枝处理后，将该子网作为拟真实验的载体网络。刘泉等[13]集中Twitter网络数据，选取部分构成子网络并进行仿真分析。Yang等[14]在真实网络上对竞争信息模型进行了仿真。真正的复合网络由两个

大型Facebook网络组成。Jing等[15]模拟谣言和事实在人工网络和两个真实网络feather-lastfm-social和ego-Facebook中的传播。Wang等[16]在仿真过程中,使用社交媒体聚合程序Friendfeed获得的Friendfeed和Twitter耦合网络数据。

3.3 真实参数数据

建立丰富的模型机制和有效的引导策略是至关重要的,但这涉及到多个参数的调控。在某些情况下,一些参数可能没有明确的参考标准,因此需要通过在仿真过程中代入真实舆情数据进行计算,以确定这些参数的值,使得舆情演变过程更贴合真实情况。在调整参数的过程中,可以关注特定的舆情传播特征,如传播速度、影响范围等,以确保模型与真实情况更为一致。Yan等[10]收集微博数据,发现微博信息量和频率在12点和24点达到峰值,因此用12量化参数T。丁金珠[17]在仿真过程中代入南京地区五种媒介的公信力数据量化模型中的媒介网络公信力参数。项权等[18]通过借鉴其他文献对"日本地震后的抢盐风波"事件的调查数据,设置了部分模型参数的初始值。Wang等[16]在"双黄连"事件的仿真过程中代入真实的外部干预策略时间来量化模型中的参数t_0。Geng[19]抓取"4·19特斯拉车主维权事件"微博数据,使用TextCNN进行情感分析,将正面和负面情绪帖子的总数用于计算部分参数值。

3.4 与真实数据比较

为了验证舆情传播模型的有效性和合理性,可以通过采集真实舆情事件的相关数据,与仿真结果进行比较,以评估模型对舆情传播的模拟能力。将这些真实数据代入仿真模型,进行曲线拟合,并计算仿真结果与真实数据之间的差异。一种常见的评估指标是平均相对误差,即模型仿真结果与真实数据之间相对差异的平均值。通过采用平均相对误差等指标,可以细致地检查模型的预测与实际数据之间的一致性,并识别模型可能存在的偏差或不足。这样的验证过程不仅有助于确认模型的可靠性,还能提供深入洞察模型的机制局限性和改进方向。Zhang等[20]对比微博一个实际情况的谣言传播(福岛核事故后有盐抢购的谣言)与包含官方谣言反驳的ICSAR模型的曲线,变化趋势非常相似。陈福集等[21]采集"上海外滩踩踏"事件中感染传播人群的百度指数,其趋势基本与模拟的该事件的感染传播人群变化趋势相同。尹熙成等[12]采集"马航MH370"事件的百度指数(包括全国搜索指数和媒体指数),发现事件发生的一个月间,搜索指数和媒体指数都经历了多次高峰,与仿真结果相符合。崔金栋等[22]对微博"北大学子在美失联"话题2017年6月12日至2017年6月21日传播数据进行抓取,发现仿真数据与真实数据的曲线大致形同,并通过计算真实数据与仿真数据之间的平均相对误差分析模型可靠性。黄远等[6]通过抓取分析"杭州公交纵火"微博评论数,得到不同观点的真实每日分布趋势,对比可知其与仿真所得结果高度一致,验证了模型的有效性。Yu等[7]采集微博上"大象踩踏游客"事件的真实数据,发现真实数据与仿真结果曲线比较吻合,即仿真结果能很好地反映谣言传播的演变过程。庄文英等[8]对照实际微博意见演化趋势以及模型仿真演化趋势,发现演化趋势基本一致。Eminente等[23]将模型演化趋势与推特上的"阿斯利康疫苗辩论"真实数据对比,利用机器学习为模型选择合适参数,证明模型中广播激活的时刻与真实数据中人们开始对阿斯利康辩论产生兴趣的时刻一致。

模型对真实案例的选择均以该模型的类型及特点为基础,数据主要选取微博、Twitter、Facebook、百度等平台的特定热点事件数据进行分析。

4 机器学习的应用

真实舆情传播情况具有高度的不确定性和复杂性[24],传播动力学模型在辨别随时间变化的复杂关系的泛化能力(即模型对新样本的适应能力)和鲁棒性(模型本身结构和参数扰动下的稳定性)方面存在局限性,因此传播动力学模型与真实数据的结合需要借助机器学习。机器/深度学习方法具有更高的泛化能力,不仅可以用于传播预测,也可以应用于数据处理和参数估计,具体方法如图1所示。

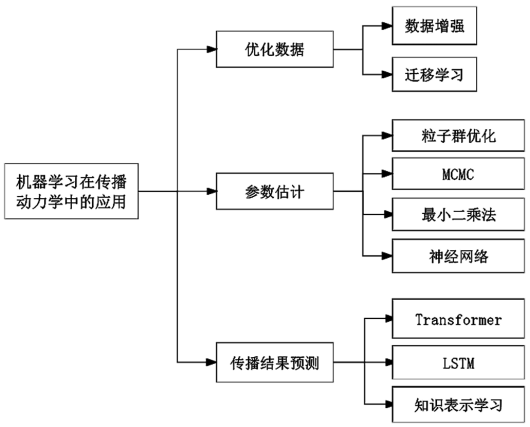

图1 机器学习在传播动力学中的应用

4.1 数据处理

机器学习模型输入需要采用合适的数据，但时常出现数据的质量和数量不符合要求的情况。深度学习适合数据密集型应用，但缺乏从有限数量的数据样本中学习的能力。其中一种解决方法是数据增强，通过对现有数据进行各种变换和处理来生成新的数据样本，用于增加训练数据集多样性，例如Wang等[24]通过生成对抗网络（GAN，generative adversarial network）综合生成数据来解决数据稀缺问题，生成器与称为鉴别器的对抗网络进行对抗，以获得可以堪比真实数据分布的生成数据。

另一种解决方法是迁移学习，将一个领域中学到的知识应用到另一个相关但不同的领域中。通过迁移学习，可以在数据量有限的情况下，利用源任务中学习到的知识，迁移到目标任务中使用。例如Roster等[25]使用基于实例（源域的某些部分数通过重新加权在目标域中进行学习）的迁移方法和基于参数传递（使用源疾病模型的参数并仅在目标疾病的模型上进行微调）的迁移方法，利用地方病数据进行学习用于新疾病的预测。类似地，目标舆情的传播规律预测可以使用相似舆情事件的数据进行迁移学习。

4.2 参数估计

为了使传播动力学模型能够准确反映实际情况，学者们通常会利用真实数据对模型参数进行准确估计。随着数据量和模型复杂度的增加，机器学习方法在参数估计中的应用日益受到关注，常用方法包括最小二乘法、马尔可夫链蒙特卡洛、粒子群优化、神经网络等。

最小二乘法（LS，least squares method）的核心思想是通过最小化模型预测值与实际观测数据之间的误差，来估计传播动力学模型的参数值。利用初始参数和数值方法模拟预测值，与实际观测数据进行比较，定义误差函数量化预测误差，进而通过优化算法调整参数值，最小化误差函数。从而找到最优参数，使模型仿真结果尽可能贴近实际数据。Yin等[26]使用LS通过将MNE-SFI（multiple negative emotional susceptible-forwarding-immune）模型与微博的真实数据进行拟合，估计模型参数和初始易感人群S_0。Zhu等[27]使用最小二乘法和两个真实微信群的实际数据对微信群传播模型中的各参数进行拟合。

马尔可夫链蒙特卡洛（MCMC，markov chain monte carlo）方法是一种常用的贝叶斯统计方法，可以用来估计模型参数的后验分布。首先，为模型中需要估计的参数设定先验分布，使用传播动力学模型构建似然函数。该函数将模拟模型生成的数据与真实数据进行比较，以评估参数值的可能性。之后，选择一个初始参数值组合，使用MCMC采样从参数的后验分布中抽取样本。在每个迭代中，根据当前参数值和似然函数，生成新的参数值，并根据一定的接受准则决定是否接受该值。Shi等[28]通过MCMC方法，使用香港的疫情数据拟合模型参数，用平均值、标准差、MCMC方法产生的误差、MCMC方法的延迟接受时间等指标评价参数拟合的准确性。由于最小二乘法对多峰后验分布或复杂的非线性模型，可能会陷入局部最优解，因此有学者将MCMC与最小二乘法结合，提高估计值的准确性。Yin等[29]使用LS的计算结果作为MCMC估计的初值$\Theta(0)$，以微分方程中未知参数的后验分布为不变极限分布构造马尔可夫链，利用后验参数的协方差矩阵估计模型仿真迭代1000次的参数分布。

粒子群优化（PSO，particle swarm optimization）的核心思想是利用群体中个体之间的信息共享，使整个群体的运动在问题解空间中由无序向有序演化，从而获得问题的最优解。在传播动力学中，以需要估计的参数作为粒子位置，定义一个适应度函数，用于衡量模型模拟结果与观测数据之间的差异。常见的适应度函数包括均方误差（MSE，mean squared erro）、均绝对误差（MAE，mean absolute error）等。根据适应度函数改变目标方向和速度迭代寻求粒子的局部最优位置，达到最大迭代次数后获得全局最优参数组合。例如Dong等[30]构建最优控制的目标函数作为适应度函数以寻找控制强度K、控制策略触发参数θ和收敛参数ρ。Eminente等[23]采用MSE作为适应度函数，用推特上的阿斯利康疫苗辩论数据以模拟参数。

神经网络估计参数的核心思想是利用神经网络的强大非线性逼近能力，通过训练神经网络从实际观测数据中学习这些参数的最佳值。具体过程包括设计适当的神经网络架构、定义损失函数、通过最小化损失函数迭代优化网络参数。例如Soures等[31]在SIRNET模型中使用循环神经网络（RNN，recurrent neural network）架构，用国家、州和县三级的流动数据分别拟合这些区域内的接触率。Li等[32]采用了基于物理信息神经网络（PINN，physics informed neural network）架构和真实谣言传播结合的新型谣言信息神经网络（RINN，rumor-informed neural network）拟合多个谣言传播模型参数。

最终选择哪种方法取决于具体的应用场景和需求,各参数拟合方法优劣如表2所示。最小二乘法适合简单、线性的模型,而MCMC、PSO和神经网络则适合处理更复杂的传播动力学模型。MCMC能够处理复杂的非线性模型,但计算成本高;PS0具有强大的全局搜索能力,但对参数选择敏感;神经网络适用于大规模复杂数据,但训练时间长且解释性差。研究者应根据实际情况和资源限制综合考虑这些方法的优缺点来选择最合适的参数估计方法。

表2 参数拟合方法优劣

目的	方法	核心思想	优点	缺点
拟合参数	最小二乘法	最小化模型预测值与实际观测数据之间的误差	简单易用	对于非线性模型效果较差
	MCMC	构建马尔可夫链,从后验分布中采样参数值,使得参数值的分布逐渐逼近目标后验分布	适用广泛,能够处理复杂的非线性模型	计算成本高,调参需确保收敛
	粒子群优化	模拟一群粒子在搜索空间中的移动,逐步找到能够最小化误差函数的最佳参数值	适用性广,全局搜索能力强	计算量大,参数选择敏感
	神经网络	训练神经网络从观测数据中通过最小化误差函数来学习和提取模型参数	高精度,适合处理大规模数据	训练时间长,易过拟合,解释性差

4.3 传播结果预测

了解信息如何传播,预测其未来趋势,对于政府、企业、学术界乃至个人都至关重要。信息的传播不仅仅是简单的信息发布,而是一个复杂的交互系统,涵盖了参与人群、传播渠道、传播方式等多个维度。因此,要准确预测信息传播的规律,需要深入理解这些维度之间的关系以及它们随时间的变化。深度学习方法为传播规律预测提供了更好的方法。

长短期记忆网络(LSTM,long short-term memory networks)的主要思想是通过内部的记忆单元捕捉序列数据中的时间依赖关系,以学习和预测时间序列中的模式和规律。LSTM是RNN的变体,改良了传统RNN梯度爆炸和消失的问题。它可以接受信息传播的序列数据作为输入,捕捉到序列数据中的时间依赖关系,通过其内部的记忆单元来保存和更新过去的信息,也可以自动学习序列数据中的特征,如用户的行为特征、信息的内容特征等,并将其作为输入序列的一部分,从而提高预测的准确性。同时,需要明确定义预测的目标,例如预测未来某一时刻的信息传播量、预测某一信息在网络中的传播路径、预测某一用户的信息传播行为等。根据具体的预测目标,可以设计相应的模型结构和损失函数来进行训练和优化。Xie等[33]在LSTM的基础上引入隐半马尔可夫模型(HSMM,hidden semi-Markov models)的思想,提出一种基于显式持续时间循环网络(EDRN,explicit duration recurrent network)的微博信息传播模型,用于流行信息的早期检测。Gautam等[34]将迁移学习应用于LSTM网络,使用N个LSTM网络单元学习过去N天的数据,预测未来N+X天的传播趋势。Soures等[35]将LSTM模型同传统的SEIR模型结合,将流行病学模型嵌入到神经网络中,编码易感(S,susceptible)、暴露(E,exposed)、感染(I,infectious)和恢复(R,recovered)的动态变化。

Transformer模型核心思想是通过自注意力机制捕捉序列中各元素之间的长距离依赖关系,实现并行化操作,以高效地学习和预测序列中的模式和规律。自注意力机制通过计算每个位置与其他位置的相关性,从而捕捉到序列中各个元素之间的关系。Transformer的架构由编码器和解码器两部分组成,编码器用于处理输入序列并提取特征,解码器生成输出序列。每个编码器和解码器层包含多个自注意力和前馈神经网络单元。Wang等[24]采用Transformer模型预测时间步长τ的结果,并与没有应用基于Transformer的预测模型进行对比分析。

知识表示学习(KRL,knowledge representation learning)通过将知识图谱中的实体和关系嵌入到低维向量空间,捕捉复杂的关系和语义信息,从而为传播规律预测提供特征支持。在传播规律预测中,首先构建知识图谱,包括用户、信息、时间节点等实体及其相互关系。通过KRL方法将这些实体和关系转换为向量表示,保留其语义和结构信息。这些向量可以作为输入特征,结合深度学习模型或图神经网络进行传播规律建模。Xiang等[36]采用知识表示算法,通过稀疏矩阵将谣言主题网络的元素映射到不同的关系空间,得到用户特征向量表示和消息特征向量表示。利

用图神经网络（GraphSAGE，graph sample and aggregate）进行谣言热度预测。

类似地，每种方法各有利弊，选择哪种方法也取决于具体的应用场景和模型设置，各预测方法优劣如表3所示。长短期记忆网络（LSTM）通过其强大的时间序列处理能力，能够捕捉信息传播中的时间依赖关系。Transformer模型则凭借其自注意力机制，更好地捕捉文本中的长距离依赖关系和语义信息，适用于复杂的传播模式预测。知识表示学习（KRL）通过将知识图谱中的实体和关系嵌入低维向量空间，为传播规律预测提供了丰富的上下文信息和语义关联，进一步提升了模型的预测能力。这些技术的结合和应用提供了更精准的传播规律预测工具，帮助更好地理解和预测信息传播的未来趋势。随着技术的进一步发展，传播规律预测有望在更多领域中发挥重要作用，推动各类信息传播的科学管理和有效控制。

表3 传播规律预测方法优劣

目的	方法	核心思想	优点	缺点
传播预测	LSTM	通过记忆单元捕捉序列数据中的时间依赖关系，以学习和预测时间序列中的模式和规律	适合处理具有时间依赖性的传播数据，以有效传递状态信息	训练时间长，内存消耗大，难以并行训练
	Transformer	通过自注意力机制捕捉序列中各元素之间的长距离依赖关系，实现并行化操作	可并行化处理，适合大规模的数据和计算，可处理长序列数据，易扩展	计算资源需求高，不适合稀疏数据
	知识表示学习	通过将知识图谱中的实体和关系映射到低维向量空间，以捕捉和表示它们之间的复杂关系和语义信息	提供额外的语义信息和约束，理解数据背后的语义信息	知识表示的质量取决于外部知识的准确性和完整性，可能只特定领域表现良好

5 大语言模型的应用

在信息传播研究中，大语言模型（LLMs，large language models）展示了其强大的应用潜力和广泛的适用性。本节介绍了LLMs在社交运动模拟、假新闻传播、流行病建模和新闻传播中的多种应用场景。通过将LLMs与主体模型（ABMs，agent-based models）结合，研究人员能够更加准确地模拟和预测用户行为及其传播规律。尽管当前研究展示了LLMs在捕捉人类行为复杂性、行为决策、评估信息影响力度以及预测未来事件方面的潜力，但其计算复杂性和高成本仍然是主要挑战。以下几方面展示了LLMs在信息传播研究中的使用，为与传播动力学结合提供了有价值的参考。

（1）复杂行为模拟

大模型被用于模拟个体或智能体（agent）的复杂行为和决策过程，利用大模型处理个性化的提示和上下文信息，为每个智能体提供基于其特征和记忆的推理能力，包括日常活动选择、健康决策、社会互动等。Williams等[37]使用大模型模拟了在流行病传染过程中智能体根据个性、健康状况、感知风险等因素做出是否外出的决策。Mou等[38]利用大模型驱动核心用户的模拟，配置模块设置用户人设，记忆模块驱动用户进行反思并更新观点，动作模块驱动用户进行发布、转发、回复等行为。

（2）记忆和学习机制

在模拟中实现短期和长期记忆机制，使智能体能够根据历史互动和新信息更新其观点和行为。Mou等[38]设置了记忆模块，包括让每个智能体通过观察将当前经历写入记忆，从记忆中提取与当前观察相关的信息，定期对记忆进行反思。Liu等[39]模拟的每个智能体都配备了一个短期记忆来捕捉当天的互动和一个长期记忆捕捉更广泛的上下文，通过交流以更新对假新闻的观点。

（3）交互与动态连接

大模型模拟智能体之间的交互和连接的动态变化，以及这些变化对信息传播的影响。Williams等[37]使每个离家智能体单独与等于模型接触率的多个独特智能体进行交互，导致易感者和感染者之间的疾病传播。Mou等[38]模拟的普通用户具有选择功能，可以与更相似的agent进行交互（内部因素），也可以由平台的推荐算法进行交互（外部因素）以改变对社会运动的看法。Liu等[39]设置每天每个智能体都将与其他c个智能体随机交互，更新自己对假新闻的观点。DeBuse等[40]构建时变网络，考虑了社交网络图中agent之间连接的共同进化。

（4）情感和态度分析

大模型能够分析文本数据中的情感倾向，如社交媒体帖子的情绪得分，用于理解公众情绪和态度。Gujral等[41]通过LLM处理社交媒体数据，识别其中各个政党的情感倾向，从而预测选举结果。Mou等[38]通过态度评分来观察社会运动中的集体态度。

（5）干预策略评估

在模拟中引入官方智能体或其他形式的干预，使

用大模型评估这些干预对信息传播、态度变化或行为模式的影响。Liu等[39]用大模型模拟了个体对假新闻的态度动态,并评估了官方智能体引入官方驳斥对假新闻传播的影响。DeBuse等[40]用大模型研究了三类代理人(固执代理人、受欢迎的代理人、策略代理人)在时变社会网络中控制意见传播的策略。Mou等[38]评估了核心用户对普通用户态度变化的影响,这可以被视为一种社会运动干预策略的评估。

(6)大规模数据集处理

大模型可以处理和分析大规模数据集,从数据中提取有用的信息和模式。Zhou等[42]引入拥有113k个问答对的外部知识源,agent生成帖子时会根据帖子主题从外部数据集中检索角色应该拥有的知识,从而为动作提供支持。

(7)预测和建模

利用大模型进行预测建模,如预测选举结果、流行病传播趋势或社会运动的发展。Gujral等[41]用大模型分析推特数据,根据用户情绪、文本情绪预测了印度旁遮普邦和北方邦的选举结果,展示了LLM在政治选举预测中的应用。Williams等[37]使用大模型构建基于生成式智能体的流行病模型,模拟了个体在不同条件下对流行病的反应和行为,进而预测疾病传播的动态。

综上所述,LLMs在信息传播研究中的应用不仅体现了其处理复杂数据和模拟人类行为的能力,也突显了其在社会现象预测和干预策略评估中的潜力,这些研究成果为进一步探索大语言模型在不同领域的应用潜力奠定了基础。然而,随着计算复杂性和成本的增加,如何高效、经济地利用LLMs,减少算力成本,以及如何解决与数据隐私、模型偏见相关的问题,将是未来研究需要重点关注的方向。此外,随着各家厂商对大模型的持续改进,LLMs解释能力和泛化性的提高将进一步推动LLMs在信息传播研究中的应用和发展。

6 总结与展望

本文系统地探讨了数据驱动下的传播动力学模型及其与多个领域结合的发展和前景。主要内容涵盖了仿真方法介绍、真实数据拟合、机器学习的应用、以及大语言模型在信息传播研究中的作用。通过对现有研究成果的梳理,总结了以下几点结论和未来研究方向。

(1)现有传播动力学分析以仿真分析为主,初始参数及网络结构多为人为设定,缺乏理论依据,基于数据拟合的研究以真实数据和仿真结果验证为主,或利用真实数据结合机器学习对模型部分初始参数进行改进,带入数据进行全过程的拟合分析非常有限。数据驱动是未来舆情传播动力学研究的重点。

(2)机器学习在传播动力学中的应用为数据驱动动力学模型提供了技术支持。随着数据量和模型复杂度的增加,传统的方法在处理多维数据和非线性关系时的局限性日益显现。机器学习,尤其是深度学习,通过提高模型的泛化能力和鲁棒性,为复杂数据分析提供了适用于多种应用场景的新方法。

(3)大语言模型(LLMs)的引入展示了其在信息传播研究中的巨大潜力。LLMs能够通过处理自然语言文本、模拟个体的复杂行为和决策过程,为信息传播研究提供了更为细致的模拟手段。尽管大模型具有丰富的使用方式,但尚未出现与传播动力学密切结合的研究。

已有大量研究揭示了传播动力学模型和大数据技术在信息传播中的重要性,但仍存在一些挑战和未来研究方向:

(1)模型的复杂性与计算成本:随着模型复杂度的增加,计算成本也显著上升。未来的研究需要在保持拟合准确性的同时,优化算法以降低计算成本。

(2)数据质量问题:在使用真实数据进行建模和预测时,需要关注所采数据的事件与模型的匹配度,保证数据收集的及时性和完整性,特别是在快速传播的事件中。同时需要注意数据量和无效数据的清洗。

(3)跨学科融合:传播动力学研究需要更多地融合社会学、心理学等学科的理论和方法,以便更全面地理解信息传播的多维度影响,提供更贴合实际的人群行为和传播场景搭建。

(4)实时数据处理与分析:未来的研究应关注实时数据的处理和分析,开发高效的实时监测和预警系统,注重动态网络中的信息传播。

综上所述,本论文通过对传播动力学模型及其应用的深入探讨,揭示了动力学结合数据应用和人工智能在信息传播研究中的广阔前景。未来的研究需要进一步提升模型的精确性和实用性,为社会治理和公共政策提供更加科学的决策支持。

参考文献:

[1] Erdős P, Rényi A R. On random graphs I[J].Publicationes Mathematicae, 1959, 6: 290-297.

[2] Watts D J，Strogatz S H. Collective dynamics of ‘small-world’networks[J]. Nature，1998，393：440-442.
[3] Barabási A L，Albert R. Emergence of scaling in random networks[J]. Science，1999，286(5439)：509-512.
[4] Zhao L，Wang Q，Cheng J，et al. The impact of authorities’ media and rumor dissemination on the evolution of emergency[J]. Physica A：Statistical Mechanics and its Applications，2012，391(15)：3978-3987.
[5] 夏一雪，兰月新，刘冰月，等. 大数据环境下网络舆情信息交互模型研究[J]. 现代情报，2017，37(11)：3-9+16.
[6] 黄远，刘怡君. 多层多属性舆情传播网络的仿真研究[J]. 系统工程学报，2019，34(06)：844-854.
[7] Yu S，Yu Z，Jiang H，et al. The dynamics and control of 2I2SR rumor spreading models in multilingual online social networks[J]. Information Sciences，2021，581：18-41.
[8] 庄文英，许英姿，任俊玲，等. 突发事件舆情演化与治理研究——基于拓展多意见竞争演化模型[J]. 情报杂志，2021，40(12)：127-134.
[9] Jiang M，Gao Q，Zhuang J. Reciprocal spreading and debunking processes of online misinformation：a new rumor spreading-debunking model with a case study[J]. Physica A：Statistical Mechanics and its Applications，2021，565：125572.
[10] Yan Q，Yi L，Wu L. Human dynamic model co-driven by interest and social identity in the MicroBlog community[J]. Physica A：Statistical Mechanics and its Applications，2012，391(4)：1540-1545.
[11] Xia L L，Jiang G P，Song B，et al. Rumor spreading model considering hesitating mechanism in complex social networks[J]. Physica A：Statistical Mechanics and its Applications，2015，437：295-303.
[12] 尹熙成，朱恒民，马静，等. 微博舆情话题传播的耦合网络模型——分析话题衍生性特征与用户阅读心理[J]. 情报理论与实践，2015，38(11)：82-86.
[13] 刘泉，荣莉莉，于凯. 考虑多层邻居节点影响的微博网络舆论演化模型[J]. 系统工程学报，2017，32(06)：721-731.
[14] Yang D，Chow T W S，Zhong L，et al. The competitive information spreading over multiplex social networks[J]. Physica A：Statistical Mechanics and its Applications，2018，503：981-990.
[15] Jing W，Kang H. An effective ISDPR rumor propagation model on complex networks[J]. International Journal of Intelligent Systems，2022，37(12)：11188-11213.
[16] Wang Y，Wang J，Zhang R，et al. Enhanced by mobility? effect of users’ mobility on information diffusion in coupled online social networks[J]. Physica A：Statistical Mechanics and its Applications，2022，607：128201.
[17] 丁金珠. 基于复杂网络的广告信息传播研究[D]. 南京：南京邮电大学，2013.
[18] 项权，于同洋，肖人彬. 突发事件网络舆情演化与干预[J]. 计算机应用，2018，38(S2)：97-102.
[19] Geng L，Zheng H，Qiao G，et al. Online public opinion dissemination model and simulation under media intervention from different perspectives[J]. Chaos，Solitons & Fractals，2023，166：112959.
[20] Zhang N，Huang H，Su B，et al. Dynamic 8-state ICSAR rumor propagation model considering official rumor refutation[J]. Physica A：Statistical Mechanics and Its Applications，2014，415：333-346.
[21] 陈福集，游丹丹. 基于系统动力学的网络舆情事件传播研究[J]. 情报杂志，2015，34(09)：118-122.
[22] 崔金栋，郑鹊，孙硕. 基于改良SEIR模型的微博话题式信息传播研究[J]. 情报科学，2017，35(12)：22-27.
[23] Eminente C，Artime O，Domenico M D. Interplay between exogenous triggers and endogenous behavioral changes in contagion processes on social networks[J]. Chaos，Solitons & Fractals，2022，165：112759.
[24] Wang H，Tao G，Ma J，et al. Predicting the epidemics trend of COVID-19 using epidemiological-based generative adversarial networks[J]. IEEE Journal of Selected Topics in Signal Processing，2022，16(2)：276-288.
[25] Roster K，Connaughton C，Rodrigues F A. Forecasting new diseases in low-data settings using transfer learning[J]. Chaos，Solitons & Fractals，2022，161：112306.
[27] Yin F，Xia X，Pan Y，et al. Sentiment mutation and negative emotion contagion dynamics in social media：a case study on the Chinese Sina Microblog[J]. Information Sciences，2022，594：118-135.
[27] Zhu H，Jin Z. A dynamics model of knowledge dissemination in a WeChat Group from perspective of duplex networks[J]. Applied Mathematics and Computation，2023，454：128083.
[28] Shi L，Chen Z，Wu P. Spatial and temporal dynamics of COVID-19 with nonlocal dispersal in heterogeneous environment：modeling，analysis and simulation[J]. Chaos，Solitons & Fractals，2023，174：113891.
[29] Yin F，Pan Y，Tang X，et al. An information propagation network dynamic considering multi-platform influences[J]. Applied Mathematics Letters，2022，133：108231.
[30] Dong Y，Zhao L. An improved two-layer model for rumor propagation considering time delay and event-triggered impulsive control strategy[J]. Chaos，Solitons & Fractals，2022，164：112711.
[31] Soures N，Chambers D，Carmichael Z，et al. SIRNet：understanding social distancing measures with hybrid neural network model for COVID-19 infectious spread[DB/OL]. arXiv：2004.10376，2020.
[32] Li D，Zhao Y，Deng Y. Rumor spreading model with a focus on educational impact and optimal control[J]. Nonlinear Dynamics，2024，112(2)：1575-1597

（下转第54页）