

# IP地理定位技术综述：理论、方法与应用创新

刘学婷<sup>1</sup>, 台文鑫<sup>1</sup>, 周帆<sup>1</sup>, 王永<sup>2,3</sup>, 陈凯<sup>2</sup>

(1. 电子科技大学信息与软件工程学院, 四川 成都 610054; 2. 香港科技大学计算机工程学院, 香港 999077;  
3. 郑州埃文科技有限公司, 河南 郑州 450047)

**摘要:** IP地理定位在推断网络空间实体位置方面具有关键作用, 广泛应用于通信系统优化、金融风险防控和网络拓扑测绘等领域。针对近20年IP地理定位领域的研究论文进行系统总结, 分析各类技术的特点、优势与局限性, 并深入调研实际应用场景和潜在发展方向。通过构建综合性IP地理定位技术体系框架, 整合不同方法与范式, 重点分析目前存在的问题, 旨在为后续研究提供理论与方法论的支持, 推动IP地理定位技术进一步发展和创新性应用。

**关键词:** 网络空间实体; IP地理定位; 数据挖掘; 基准点分析; 网络时延测量; 机器学习; 深度学习

**中图分类号:** TN393

**文献标志码:** A

**DOI:** 10.11959/j.issn.1000-436x.2025046

## Survey of IP geolocation: theory, methods, and innovative applications

LIU Xueting<sup>1</sup>, TAI Wenxin<sup>1</sup>, ZHOU Fan<sup>1</sup>, WANG Yong<sup>2,3</sup>, CHEN Kai<sup>2</sup>

1. School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China  
2. Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong 999077, China  
3. Zhengzhou Aiwen Tech Co., Ltd., Zhengzhou 450047, China

**Abstract:** IP geolocation plays a critical role in determining the locations of cyberspace entities and is widely applied in important fields such as communication system optimization, financial risk prevention, and network topology mapping. The research papers in the field of IP geolocation over the past 20 years were reviewed systematically, the characteristics, advantages, and limitations of various approaches were analyzed, while exploring practical application scenarios and potential future development directions. The first comprehensive framework for IP geolocation technologies was also established in this review, different methods and paradigms were taxonomized, while the current challenges in accurate, trustworthy, and robust IP geolocations were analyzed. The goal is to provide theoretical and methodological support for future research, thereby promoting the further development and innovative applications of IP geolocation technologies.

**Keywords:** cyberspace entity, IP geolocation, data mining, anchor point analysis, network delay measurement, machine learning, deep learning

### 0 引言

互联网协议 (IP, Internet protocol) 地址是互联网主机的唯一标识符, 用于虚拟位置寻址。IP地理

定位<sup>[1]</sup>旨在确定IP地址对应的现实地理位置, 通常以经纬度表示, 如图1所示。精确定位技术在网络安全、内容分发、数字广告及网络犯罪追踪等领域

收稿日期: 2024-12-26; 修回日期: 2025-03-06

通信作者: 周帆, fan.zhou@uestc.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62176043, No.62072077); 四川省自然科学基金资助项目 (No.2022NSFSC0505)

**Foundation Items:** The National Natural Science Foundation of China (No.62176043, No.62072077), The Natural Science Foundation of Sichuan Province (No.2022NSFSC0505)

具有重要应用价值。然而，IP地理定位面临诸多挑战。许多方法依赖全球定位系统（GPS, global positioning system）<sup>[2]</sup>、北斗卫星导航（BDS, Beidou navigation satellite）系统<sup>[3]</sup>或移动热点（Wi-Fi, wireless fidelity）<sup>[4]</sup>等数据源，但存在依赖客户端参与、普适性差、无GPS以及蜂窝网络设备的定位准确性受限等局限性。

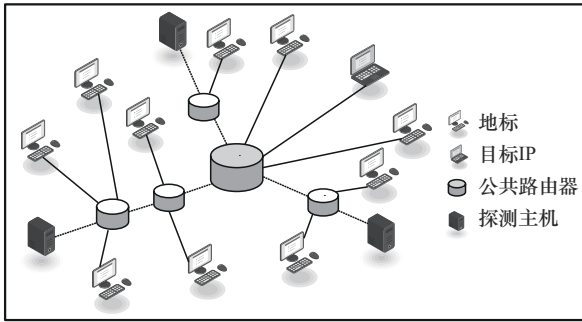


图1 IP地理定位场景示意

IP地理定位是一个重要课题。随着点对点（P2P, peer-to-peer）网络的发展，选择物理距离较近的节点以缩短时延成为关键问题<sup>[5]</sup>，推动了大量研究取得进展。通过IP地理定位可实现节点选择的优化，提高传输效率。传统方法通过维护IP地址与地理位置的映射表实现定位，依赖公共或自有数据资源（如邮政地址<sup>[6]</sup>），并需要定期更新。然而，这种方法成本高，受数据质量和更新频率影响，对信息不完整的IP地址效果较差。因此，未来IP地理定位需要更加灵活和普适，以克服上述局限。

基于网络时延测量的IP地理定位方法常用工具包括Ping<sup>[7]</sup>和traceroute<sup>[8]</sup>，通过时延数据绘制圆并以交点区域估算位置。虽然这种方法较有效，但受网络复杂性、信号衰减和设备障碍等因素影响，时延测量准确性较低。同时，节点间未知空间关系及网络拓扑变化增加了距离估算难度，降低了定位准确性与可靠性。基于机器学习与深度学习的IP地理定位算法快速发展，一部分将其建模为分类任务以确定IP区域<sup>[9]</sup>，另一部分采用回归方法预测经纬度<sup>[10]</sup>。这些方法借助大数据和复杂模型显著提升了定位准确性与可靠性，然而，其在异构网络中的鲁棒性不足，数据集间性能差异限制泛化能力，且模型低可解释性使验证难度增加。优化这些问题是提升算法实用性和可靠性的关键。

随着移动客户端普及，轻量探测模块便于获取

经纬度，降低高准确性定位数据库构建难度。但数据库需准确更新，否则易滞后失真，且隐私问题限制其应用。对无模块或数据缺失设备，可借助数据挖掘与基准点分析、网络时延测量、机器学习与深度学习等方法推断IP位置。本文通过总结这些方法，为优化研究提供了有益参考。

IP地理定位领域的文献数量统计如图2所示，2005年前的文献合并至2005—2006年展示。研究经历了从数据挖掘与基准点分析、网络时延测量到机器学习与深度学习的演变。总体来看，国外在IP地理定位领域的研究数量比国内多，但从综述性文献的角度来看，国内外的综述文献数量差不多，但现有综述普遍存在一些问题，例如，分类模糊、方法交叉重叠等，导致对整个IP地理定位技术框架的理解不清晰。

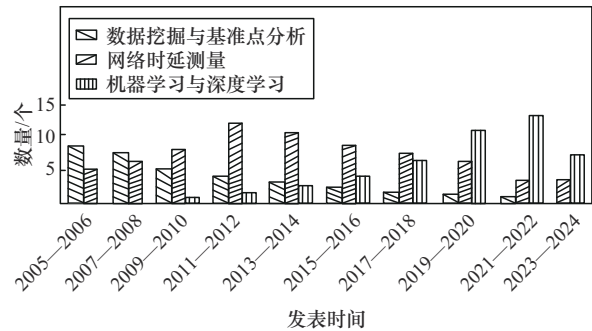


图2 IP地理定位领域的文献数量统计

国外的综述文献在一定程度上推动了该领域的发展。文献[1]提出3种用于推断互联网主机地理位置的技术，即GeoTrack、GeoPing和GeoCluster，并对这些技术的性能进行全面评估。文献[11]讨论IP地理定位数据库的可靠性，揭示这些数据库的局限性，并通过实际数据验证了其准确性问题。文献[12]则对约10万个IP地址进行分组，评估IP地理定位数据库的准确性，揭示现有数据库的优缺点，并指出在复杂网络环境中的异常情况。文献[13]进一步探讨IP地理定位面临的挑战，特别是安全性、云中的数据地理定位以及如何应对通过代理和VPN隐藏地理位置的问题。

国内的综述文献也对IP地理定位技术的发展进行了梳理。例如，早期文献[14-15]概述IP地理定位的概念、应用与分类。文献[14]将算法分为独立与基于客户端两类，讨论隐私保护技术，但未反映2014年以后的发展趋势。文献[15]探讨IP地理定位技术与其他定位技术的关系，但缺乏技术实现与

效果分析。文献[16-18]扩展了应用与发展方向，文献[16]分析溯源与定位技术，对 IP 地理定位技术在网络犯罪追踪中的应用进行展望，但对算法实际效果和局限性讨论不足。文献[17]介绍定位原理和方法，提出攻击与防御模型，但对复杂环境中的定位准确性探讨较少。文献[18]细分了主动、被动及结合方法，总结研究进展与面临的挑战，提出提高准确性与适用性的建议。

本文提出结构化的 IP 地理定位分类方法，涵盖数据挖掘与基准点分析、网络时延测量、机器学习与深度学习等技术，解决现有文献分类模糊和方法重叠问题，构建更清晰的技术框架。从应用角度分析各方法的优缺点及适用性，弥补文献[18]的不足，提供前沿技术进展，为后续研究提供全面视角。本文展望了 IP 地理定位在多源数据融合、实时监控等领域的应用潜力，提出前瞻性研究方向，填补综述空白，整体研究结构如图 3 所示。

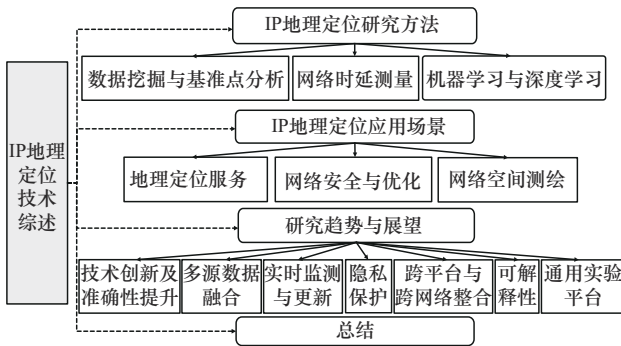


图 3 整体研究结构

### 1 IP 地理定位研究方法

本文系统阐述了数据挖掘与基准点分析、网络

时延测量及机器与深度学习 3 种 IP 地理定位技术，如图 4 所示。其中，数据挖掘与基准点分析技术通过整合 Whois<sup>[19]</sup>和 CAIDA 数据库<sup>[20]</sup>等提取地理特征。网络时延测量技术基于探测主机与目标 IP 之间的时延及拓扑关系推断位置。机器学习与深度学习技术通过构建 IP 地理位置映射模型实现精准定位。

#### 1.1 数据挖掘与基准点分析

数据挖掘与基准点分析作为 IP 地理定位的重要方法之一，通过公开或自有资源提取地址、端口信息、traceroute<sup>[8]</sup>、Whois<sup>[19]</sup>和 DNS<sup>[21]</sup>等特征。这些特征数据经过预处理后，将按特定规则划分为不同的 IP 地址组，以便进一步分析。这些处理后的信息构成数据集中的 IP Block 字段，每个 IP Block 包含最小 IP 地址和最大 IP 地址、覆盖区域、域名的 IP 数量等统计特征。通过对 IP Block 字段进行分析，实现精确定位。

基于数据挖掘的 IP 地理定位技术包括以下方法。首先，通过对现有地理数据库（如 CAIDA<sup>[20]</sup>和 PlanetLab<sup>[22]</sup>）进行性能评估和对比，提升 IP 地理定位的准确性。Poese 等<sup>[11]</sup>证实商业 IP 数据库可显著提升城市级定位的准确性，尤其是在网络拓扑稀疏的区域，为不同区域定位准确性对比提供了有力证据。其次，一些研究通过解析域名和主机名中嵌入的地理信息辅助 IP 地理定位，这些域名和主机名通常包含地理位置线索。例如，DRoP<sup>[23]</sup>通过收集宿主域内的地理提示，将这些提示与实际物理位置进行关联。此外，还有研究通过分析用户行为数据实现 IP 地理定位。例如，Checkin-Geo<sup>[6]</sup>利用用户在位置共享服务中分享的位置数据和用户在计

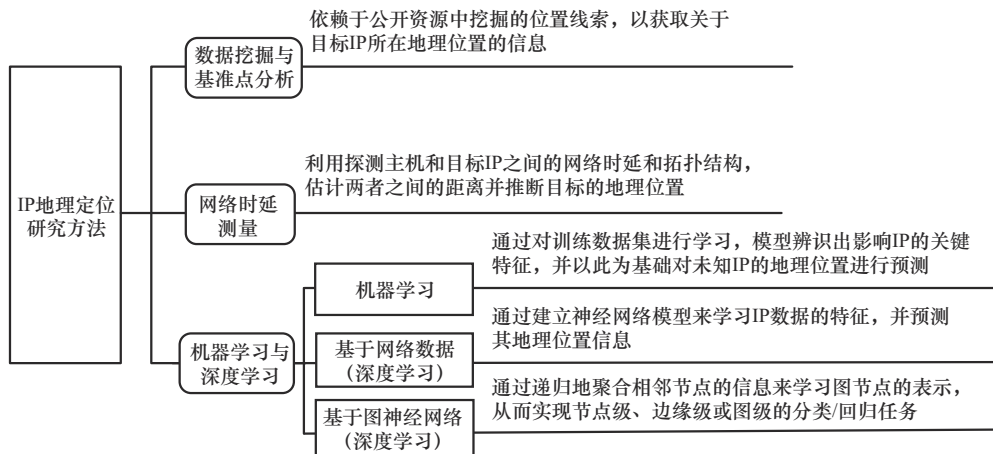


图 4 IP 地理定位研究方法分类

计算机上的登录日志, 准确地进行实时定位。Dan 等<sup>[24]</sup>利用跟踪路由路径上的时延差异特征, 改进 IP 地理定位模型。

基准点分析法利用已知位置基准点(如特定服务器、设备等)推断未知 IP 位置。Lee 等<sup>[25]</sup>结合众包宽带性能数据和用户地理位置数据, 实现精准定位。另一些研究则通过分析 IP 的开放端口及其对应的服务类型, 获取与 IP 地址所属组织或服务提供者相关的线索, 从而辅助定位。Structon<sup>[26]</sup>是一种零探测定位方法, 通过一系列信息聚类、虚假信息过滤、错误纠正和位置推断算法实现从 IP 地址到地理位置的映射。Alidade 方法<sup>[27]</sup>通过可用的约束条件计算所有地址的联合解, 实现高效的 IP 地理定位。Sommers<sup>[28]</sup>从网页客户端的角度出发, 通过网页响应头和未加密 Cookie 提取地理信息, 进一步提升定位准确性。

表 1 从数据集/数据库构成、地理位置资源及定位级别 3 个方面进行对比分析。基于数据挖掘与基准点分析的方法在数据获取方面存在显著困难, 这类方法依赖于网页爬虫和搜索引擎日志等网络数据资源, 难以收集到充足数据以进行有效分析。由于这一限制条件, 这类方法在实际应用中面临较大挑战, 难以提供精确且可靠的 IP 地理位置信息。

## 1.2 网络时延测量

通过测量探测主机与目标 IP 地址之间的网络时延, 结合已知的网络拓扑结构信息推断两者之间的距离关系。然后, 利用地理位置与距离之间的对应关系, 通过三角测量、多次测量、多点定位等方式推断目标的 IP 位置<sup>[29]</sup>。多主机时延数据结合已知位置提供近似定位, 通过多次测量与多点定位, 可有效提升定位准确性与可靠性。

一些研究通过发送 Ping 请求来测量往返时间, 从而推测地理位置。GeoPing<sup>[1]</sup>采用控制报文协议包匹配最近的地标服务器, 并将该服务器的地理位置分配给目标 IP。Ziviani 等<sup>[30]</sup>通过分析时延模式来推断主机位置。Dong 等<sup>[31]</sup>通过建立时延与距离关系的模型来优化位置估计。Arif 等<sup>[32]</sup>通过测量时延数据来提升定位的准确性。Gill 等<sup>[33]</sup>利用多个位置已知的主机对待定位主机的时延测量数据来估计目标位置。Posit<sup>[34]</sup>仅需少量时延测量数据, 结合高效的统计嵌入技术即可实现定位。Chen 等<sup>[35]</sup>构建了一个时延-距离模型, 用于估计目标 IP 的位置以及每个地标的真实偏差。NCRGeo<sup>[36]</sup>通过多时延法计算路由器和地标之间的距离, 并根据此距离估计地理位置。Zhao 等<sup>[37]</sup>通过对比时延分布, 在复杂网络环境中的准确性较高, 对比分析结果如表 2 所示。

另外一些研究使用 traceroute 工具确定数据包从源到目的地之间经过的路由器和节点, 识别 IP 的网络拓扑位置。GeoTrack<sup>[1]</sup>结合 traceroute 与主机名地理信息, 其准确性受限于主机名信息的可用性。Guo 等<sup>[6]</sup>结合网页挖掘技术与 IP traceroute, 显著提升定位准确性。Wang 等<sup>[38]</sup>利用拓扑发现网页服务器地标, 增强稀疏区域定位能力。Shavitt 等<sup>[12]</sup>通过结构与时延分析, 将约 10 万个入网点(PoP, point of presence)分组, 提供高置信度信息。Tian 等<sup>[39]</sup>提出启发式接口聚类方法, 基于我国互联网主要服务提供商(ISP, Internet service provider)的层次结构实现拓扑映射和定位。文献[40]通过路径探测技术追踪数据包传输路径, 利用获取的网络结构与节点位置进行 IP 地理定位。NNC 方法<sup>[41]</sup>利用网络拓扑聚类技术与投票规则提升定位准确性, 减少误差。文献[42]提出基于时延-距离相关性和

表 1 基于数据挖掘与基准点分析的 IP 地理定位方法研究比较

文献	数据集/数据库构成	地理位置资源	定位级别
文献[6]	位置共享服务数据和用户登录日志	位置信息	街道级
文献[11]	多个商业 IP 地理位置数据库	IP 地址和地理位置数据	城市级
文献[23]	DNS 解析数据和路由器位置数据库	地理位置提示	城市级
文献[24]	搜索引擎点击日志和路径时延数据	地理信息	城市级
文献[25]	众包宽带性能数据和用户反馈数据	地理标签	城市级
文献[26]	网页爬虫数据、搜索引擎日志	地理标签信息、网页内容和域名信息	城市级
文献[27]	网络拓扑信息、已知地标数据库	基于约束条件计算地理位置	城市级
文献[28]	网页客户端数据和未加密网页 Cookie	地理信息	城市级

表2 基于Ping时延测量的IP地理定位方法研究比较

文献	数据集/数据库构成	地理位置资源	定位级别
文献[1]	IP地址和地理位置	多分布地标服务器	城市级
文献[30]	测量数据和现有地理位置数据库	全球地标服务器	城市级
文献[31]	时延数据和优化模型	全球地标服务器	城市级
文献[32]	时延测量数据和最大似然估计	大城市地标节点	城市级
文献[33]	时延测量数据	已知地标服务器	城市级
文献[34]	轻量级时延测量数据	地标服务器	城市级
文献[35]	时延测量数据	已知地标节点	城市级
文献[36]	网络测量数据和路由信息	公共路由器	城市级
文献[37]	时延测量数据	路由器节点	街道级

多层路由器的定位方法,结合Wi-Fi地标采集增强准确性。GUI方法<sup>[43]</sup>针对不可达IP,通过相同子网和公共地标提升定位准确性。表3对上述研究的数据集/数据库构成、地理位置资源及定位级别进行对比分析。

一些研究结合Ping和traceroute进行IP地理定位。TBG<sup>[29]</sup>通过结合网络拓扑和时延测量约束主机位置,减少结构化网络误差。SLG<sup>[44]</sup>通过结合网络相对时延与地理相对距离之间的非线性对应关系,突破传统方法在城市细粒度定位上的局限性。CBG<sup>[45]</sup>通过多边定位分配置信区域,支持位置感知应用评估。Spotter<sup>[46]</sup>提出统一概率模型来定位目标主机。GeoGet<sup>[47]</sup>采用多步探测法将目标客户端映射到具有最短时延地标的地理位置。Hu等<sup>[48]</sup>研究如何扩展现有基于测量的地理定位算法,并在保持准确性的同时最小化测量流量。GeoCET<sup>[49]</sup>只需少量单向时延数据,通过结合椭圆轨迹约束和最大似然估计技术进行定位。Jiang

等<sup>[50]</sup>通过比较拓扑图、ISP数据和Whois<sup>[19]</sup>信息进行验证,以评估推理规则的有效性。RNBG<sup>[51]</sup>利用复杂网络的无标度特性来寻找网络中少数重要且稳定的节点,以此对不同区域的IP进行地理定位。通过使用Ping和traceroute,研究者能够在全球范围内通过时延数据估算城市级位置,虽然在复杂网络环境下时延的影响因素较多,但该方法仍显示出高效性和实用性。上述文献的汇总结果如表4所示。

通过上述分析表明,该类方法需要设计精准规则,或利用可靠的拓扑结构信息来准确估计时延和距离<sup>[52]</sup>。在此过程中,需考虑网络环境的动态性、不确定性以及可能存在的干扰因素。

### 1.3 机器学习与深度学习

机器学习与深度学习在IP地理定位领域表现优异,通过表征学习建模IP与位置之间的复杂关系,自动发掘隐藏模式,显著提升了定位准确性<sup>[53]</sup>。其并行化与可扩展性优势使其能够高效处

表3 基于traceroute时延测量的IP地理定位方法研究比较

文献	数据集/数据库构成	地理位置资源	定位级别
文献[1]	主机名数据库	主机名中的地理位置信息和路由路径信息	城市级
文献[6]	网络爬虫获取的网页数据	网页中的地理信息(如地址、坐标)	城市级
文献[12]	IP地址和地理位置数据库	PoP级别的地理位置信息	城市级
文献[38]	中间路由器的拓扑发现数据	拓扑结构和中间路由器位置	城市级
文献[39]	中国互联网拓扑数据	ISP级拓扑信息	城市级
文献[40]	蜂窝数据网络的路径测量数据	基于路径的地理信息	城市级
文献[41]	网络节点聚类数据	基于网络节点的地理聚类信息	城市级
文献[42]	多层普通路由器的测量数据	基于路由器的时延距离相关性	街道级
文献[43]	不可到达IP的地理定位数据	相同子网地标和公共路由地标	城市级

**表 4 基于 Ping 与 traceroute 时延测量的 IP 地理定位方法研究比较**

文献	数据集/数据库构成	地理位置资源	定位级别
文献[29]	时延和拓扑测量数据	网络时延与拓扑信息	城市级
文献[44]	地标服务器和时延测量数据	地标服务器位置、路径时延	街道级
文献[45]	多时延测量数据	地标服务器时延数据, 约束圆	城市级
文献[46]	模型化的路径时延测量数据	概率密度模型	城市级
文献[47]	中等连接互联网区域的数据	网页服务器作为被动地标, 时延测量	城市级
文献[48]	海量 IP 地址的测量数据	拓展测量算法	城市级
文献[49]	单向时延数据	椭圆轨迹约束、最大似然估计	街道级
文献[50]	ISP 的拓扑数据	网络测量数据	城市级
文献[51]	复杂网络节点数据	无标度网络节点	城市级

理大规模数据。本节将分别介绍机器学习与深度学习应用在 IP 地理定位中的应用。其中, 深度学习根据处理的数据类型和结构, 可分为基于网格数据的深度学习方法和基于图神经网络 (GNN, graph neural network) 的深度学习方法。

机器学习通过分析 IP 与地理位置的关联性, 提取关键特征以预测未知位置<sup>[9]</sup>, 如图 5 所示。基于这些特征构建预测模型, 输入 IP 的特征向量即可输出对应的地理位置信息。例如, GeoCop<sup>[54]</sup>结合网络测量、路由策略和机器学习方法优化定位数据库, 在测量数据不完善和路由不规则的情况下保持较高的鲁棒性。Paule 等<sup>[55]</sup>采用排名和投票算法推断未标记地理位置信息的推文。LBG<sup>[9]</sup>利用轻量级测量与概率密度分类方法, 将目标定位至最可能的区域。文献<sup>[56]</sup>通过地标挖掘与聚类算法实现定位, 并采用回归模型进行预测。Hong 等<sup>[57]</sup>结合服务器时延数据与地理位置信息, 进一步提升了定位

准确性。GeoCAM<sup>[58]</sup>利用摄像头数据和地理数据库, 实现了基于网络特征的街道级定位, 展现了机器学习在大规模数据处理和复杂模式识别中的优势。Liu 等<sup>[59]</sup>通过域名筛选与 IPv4-IPv6 连接挖掘 IPv6 地标, 并关联 2 种地址的地理位置。Light-GBM<sup>[60]</sup>采用高效的梯度提升树算法实现地理定位。Li 等<sup>[61]</sup>提出一种基于聚类与子网定位的移动 IP 地理定位方案, 无须用户或 ISP 参与。表 5 对这些文献进行了对比分析。

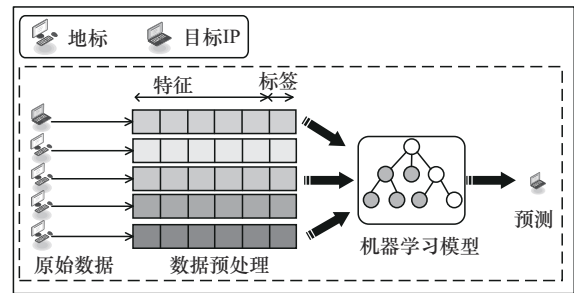


图 5 基于机器学习的 IP 地理定位方法示例

**表 5 基于机器学习的 IP 地理定位方法研究比较**

文献	模型	数据集/数据库构成	定位级别
文献[9]	朴素贝叶斯分类器	数据 (iPlane、Traceroutes) 和 MaxMind 数据库	城市级
文献[54]	决策树	地理定位数据库 (IP138、QQWry 和 IPcn)	城市级
文献[55]	排名方法结合多数投票	地理标记推文 (芝加哥和纽约)	街道级
文献[56]	K-means 聚类	网络地标数据	街道级
文献[57]	支持向量机、随机森林和 K-近邻算法	时延测量数据、地理位置数据和网络拓扑数据	城市级
文献[58]	支持向量机	摄像头数据、地理数据库和网页特征数据	街道级
文献[59]	随机森林	域名和 IPv4-IPv6 连接数据集	城市级
文献[60]	梯度提升决策树	ImageNet、MSLR-WEB10K	—
文献[61]	两阶段聚类	志愿者通过移动设备提交的 12.6 万个地理打卡点	街道级

基于网格数据的深度学习方法利用神经网络自动提取IP数据的高维特征,降低了对人工特征工程的依赖。该方法需要基于大规模数据进行训练,以确保模型能够有效提取IP地址特征并准确预测其地理位置。基于深度学习的自动特征学习机制不仅能提升定位准确性,还能有效处理复杂的非线性关系,从而实现更优的预测性能<sup>[62]</sup>。IP2vec<sup>[53]</sup>通过节点连接与时延分析实现IP节点的向量化。PlaNet<sup>[63]</sup>采用卷积神经网络(CNN, convolutional neural network)解析照片与元数据,预测地理位置。Ebrahimi等<sup>[64]</sup>利用双向长短期记忆(BiLSTM, bi-directional long short-term memory)网络融合推文内容、网络特征和元数据,预测用户位置。Thomas等<sup>[65]</sup>结合CNN和长短期记忆(LSTM, long short-term memory)网络分析推文及其上下文元数据,以预测用户位置。Zhou等<sup>[66]</sup>结合决策树、支持向量机和自动化特征交互模型AutoInt(automatic feature interaction),优化复杂场景下的预测效果。NN-Geo<sup>[10]</sup>通过多层感知机(MLP, multi-layer perceptron)结合地标与时延数据实现定位。MLP-Geo<sup>[67]</sup>引入路径编码和时延特征,有效解决时延相似性无法反映隐蔽通信实体真实地理位置的问题。Zu等<sup>[68]</sup>融合贝叶斯网络与MLP,从路由器数据中学习地理位置预测模型。TabNet<sup>[69]</sup>利用自注意力机制处理表格特征交互,显著提升了定位性能。表6对上述文献进行了系统汇总。

基于GNN<sup>[70]</sup>的深度学习方法是一种图结构数据表示方法,近年来在多领域得到广泛应用。通过GNN强大的结构表示学习能力,显著提升了定位

的准确性和效率。该方法的核心思想是通过递归聚合相邻节点信息,学习和表示图中的节点,进而实现节点级、边缘级或图级的分类/回归任务,如图6所示。GraphNEI<sup>[71]</sup>通过图结构嵌入解决特征缺失或欺骗问题,提高定位准确性。GNN-Geo<sup>[72]</sup>将IP地理定位建模为属性图节点回归任务,利用先验知识加速模型收敛,并展现细粒度定位潜力。GraphGeo<sup>[73]</sup>结合地标与目标IP的路由时延数据,聚焦网络拓扑与IP主机的语义相似性。RIPGeo<sup>[74]</sup>和TrustGeo<sup>[75]</sup>均基于公共路由器连接地标与目标IP,其中TrustGeo<sup>[75]</sup>进一步利用邻近IP的关联性降低预测不确定性,性能更优。GraphCyber<sup>[76]</sup>结合拓扑与属性相似性,通过GNN学习图数据以识别IP场景。Zhou等<sup>[77]</sup>提出层次图神经网络(HGNN, hierarchical graph neural network),通过分层聚类预测地理位置,性能显著优于传统方法。表7汇总了上述文献的模型架构、数据集与定位级别等关键信息。

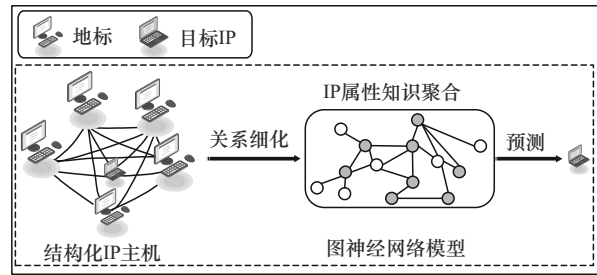


图6 基于图神经网络的深度学习方法的示意图

综上所述,机器学习与深度学习在IP地理定位领域各具优势。机器学习方法依赖人工特征提取

表6 基于网格数据的深度学习研究方法研究比较

文献	模型	数据集/数据库构成	定位级别
文献[10]	MLP	地标数据和时延数据	城市级
文献[53]	IP2vec	节点时延和路径信息(纽约、北京、香港和郑州)	街道级
文献[63]	CNN	含有Exif地理标签的照片	街道级
文献[64]	BiLSTM	推文文本、用户的社交网络信息和其他元数据	城市级
文献[65]	CNN和LSTM	推文文本、用户描述、用户位置、用户时区和其他元数据	城市级
文献[66]	AutoInt	互联网流量数据、网络日志数据(北京、上海、四川和伊利诺伊州)	城市级
文献[67]	MLP	路径编码和时延数据	城市级
文献[68]	MLP	网络路径测量数据、路由器误差信息和地理位置数据	城市级
文献[69]	TabNet	分类(KDD Cup 1999、Forest CoverType)、回归(YearPredictionMSD)	—

**表7** 基于图神经网络的深度学习研究方法比较

文献	模型	数据集/数据库构成	定位级别
文献[71]	GNN	上海、纽约、东京	街道级
文献[72]	GNN	IPv4网络(纽约、香港、上海、北京、东京)、IPv6网络(东京、柏林、慕尼黑)	街道级
文献[73]	GNN	上海、纽约、洛杉矶	街道级
文献[74]	GNN	上海、纽约、洛杉矶	街道级
文献[75]	GNN	上海、纽约、洛杉矶	街道级
文献[76]	GNN	上海、纽约、洛杉矶	街道级
文献[77]	HGNN	GeoText、Twitter-US、Twitter-World	街道级

机制,适用于对可靠性与透明度要求较高的场景。深度学习方法通过自动学习特征机制,能够有效捕捉复杂非线性关系<sup>[78]</sup>。然而,深度学习方法由于参数复杂、可解释性差等特征,在一定程度上限制了其应用范围。未来研究应结合两者的优势,开发兼具可解释性与高效特征提取能力的模型,从而进一步提升IP地理定位的准确性与计算效率。表8系统比较了基于2类方法的不同任务类型,表9则汇总了常用模型的性能指标及相关文献。

**表8** 基于机器学习与深度学习的IP地理定位方法比较

方法	任务类型	文献
机器学习	分类	文献[54-55]
	回归	文献[9,56-57,61]
	分类, 回归	文献[58-60]
深度学习	分类	文献[53,63-66,71,76]
	回归	文献[10,67,72-75]
	分类, 回归	文献[68-69,77]

**表9** 常用模型指标及对应文献汇总

评价标准	公式	文献	描述
地理位置覆盖率	$Coverage = \frac{TP + TN}{N_{total}}$	文献[11-12,16,23,27,29-30,40,42-43,46-48,50]	衡量IP地理定位系统对整体网络的覆盖情况,应用于大规模地理覆盖能力评估
准确率	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	文献[6,9,11-12,23-24,26-30,32-37,40-43,45-49,51,54,76]	衡量IP地理定位算法中正确分类的比例,应用于对定位准确性有严格要求的场景
精确率	$Precision = \frac{TP}{TP + FP}$	文献[53,57,62,66,69,71]	衡量定位为某特定区域的IP地址中实际正确的比例,用于减少误报的场景
召回率	$Recall = \frac{TP}{TP + FN}$	文献[41,64-66,70-71,77]	衡量所有实际应定位的IP地址中成功定位的比例,用于需要尽可能减少漏定位的场景
F1分数	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	文献[57,59,64,66,69,71,76-77]	用于在精确率和召回率之间取得平衡,适合在数据类别不平衡的场景中使用
平均误差	$ME = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$	文献[61,63,77]	衡量所有预测结果与真实值之间的平均差异,用于对总体误差趋势进行分析的场景
中值误差	$median =  y_i - \hat{y}_i $	文献[10,67,72-75]	计算预测误差的中位数,适合在数据中存在显著离群值的情况下使用
平均绝对误差	$MAE = \frac{1}{n} \sum_{i=1}^n  y_i - \hat{y}_i $	文献[69,72-75]	衡量预测结果与真实值之间差异的平均绝对值,用于需要对误差进行直观分析和优化的应用场景
均方根误差	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	文献[69,72-75]	衡量预测误差的平方根均值,能够有效反映模型在极端情况下的表现

## 2 IP地理定位应用场景

基于IP地理定位技术的特性和多样化应用场景，本文将其主要应用分为地理定位服务、网络安全与优化、网络空间测绘3类，如图7所示。基于IP地理定位技术的应用的实际效能主要取决于数据质量、模型准确性和算法效率。首先，IP地理定位技术通过精确获取用户或设备的地理位置信息，显著提升了基于地理位置的定位准确性和应用价值，优化网络资源分配和服务部署，改善用户体验。其次，IP地理定位技术通过识别潜在威胁源，增强网络安全性。此外，IP地理定位技术与地理空间数据挖掘技术的结合应用，能够有效发现地理空间数据中隐含的分布模式和关联规律。这种技术融合不仅扩展了IP地理定位的应用范围，还为地理信息系统开发了新的分析功能，从而促进相关技术的进步与发展。

### 2.1 地理定位服务

地理定位服务在广告定位、个性化推荐、内容本地化和交通管理等应用场景中发挥着关键作用。这些服务通过获取用户的地理位置信息，显著提升了服务精准度和运营效率。

在广告定位领域，地理位置信息被用于展示用户地区相关的广告，有效提升广告效果和用户参与度。一些研究表明，通过精确投放广告可以有效提

升传播效果和受众转化率。传统广告定位方法主要包括用户提供的地理位置信息、基于浏览器Cookie的追踪技术和GPS定位系统3类。文献[79]探讨了“志愿地理信息”概念，指出其具备实时数据和大规模地理信息收集优势，但精确度和可靠性受用户主观因素的影响较大。Turow等<sup>[80]</sup>研究了用户对定制广告和数据收集方法（如使用Cookie追踪行为）的隐私顾虑，同时指出该方法存在精确度低、数据持久性不足等缺陷，主要原因是用户可清除或关闭。Umate等<sup>[81]</sup>回顾了基于GPS的实时定向广告技术，展示了其在提升广告精准度和用户参与度方面的优势。然而，GPS在室内或高楼密集区域的信号覆盖能力不足，影响其适用性。相比之下，IP地理定位技术通过用户IP地址与地理位置信息的映射关系，实现了更精准高效的广告定位。Callejo等<sup>[53]</sup>的研究表明，采用准确性高的IP地理定位数据库可显著提升广告投放的精准度和效果。值得注意的是，IP地理定位技术在提升广告效果的同时，其配套的透明化工具有助于平衡隐私保护与广告定制之间的关系，这为未来广告定位策略的优化提供了重要参考。

在个性化推荐领域，地理信息被广泛应用于周边商户和服务推荐，以满足个性化需求。传统方法主要包括基于浏览器的地理位置API、注册信息和

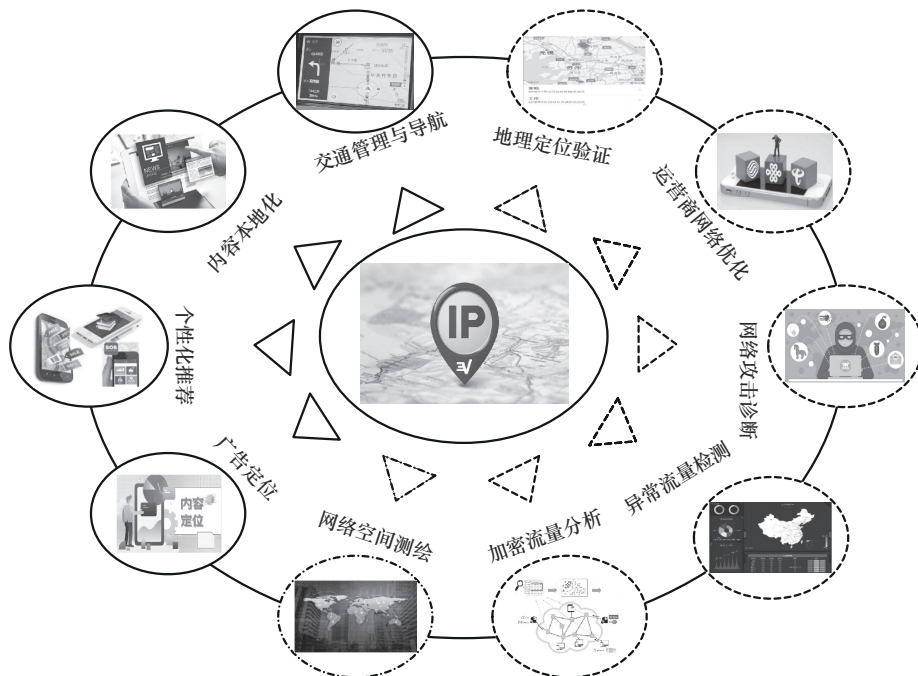


图7 IP地理定位方法的应用分类

历史行为分析。Kliman 等<sup>[82]</sup>发现, 地理位置数据对搜索引擎的个性化推荐效果有显著影响, 提升了搜索结果的相关性。然而, 该类方法依赖地理位置 API, 面临隐私保护方面的挑战。文献<sup>[83]</sup>结合用户个性、人口统计和行为数据提出基于多维度特征的推荐系统, 虽然推荐精准较高, 但系统实现复杂度较高。Kesorn 等<sup>[84]</sup>利用游客签到数据构建推荐模型。IP 地理定位技术通过识别用户 IP 地址确定其地理位置 (精确到城市级别), 再结合其他用户数据进行地域分布和行为分析, 从而提供个性化服务。研究表明, 准确性高的 IP 地理定位技术不仅能保证准确性和扩展性, 还能有效区分用户来源和提供个性化服务, 在保障网络安全的同时提升推荐系统的准确性与用户满意度。

内容本地化根据地区法律和文化特点调整内容, 以提升合规性和本地化体验。Leiva 等<sup>[85]</sup>提出的基于浏览器语言设置的动态翻译方法, 能够实时翻译用户界面文本, 降低国际化和本地化的时间与成本。然而, 当用户因系统限制被迫使用非首选语言 (如界面语言强制绑定地理位置), 或处于多语言混合区域时, 该方法可能无法精准匹配其实际需求。另一种方法是允许用户自定义地区或语言选项, 虽然提高了灵活性, 但依赖手动操作可能导致错误, 且无法为新用户或未登录用户提供个性化服务。相比之下, IP 地理定位技术通过查询 IP 数据库获取用户地理位置信息, 动态调整网站内容 (如语言、货币和地区促销等), 无须用户手动选择即可实现内容本地化。该技术能够提供精准的地理位置信息, 提升用户体验的流畅性和个性化程度, 适用于所有访问用户。

交通管理与导航通过监测交通状况提供实时导航服务, 帮助用户避开拥堵并规划最佳路径。现有技术包括 GPS 定位、交通摄像头和 IP 地理定位等方法。文献<sup>[86]</sup>提出实时监控系统, 该系统使用航空相机采集交通数据, 虽然准确较高, 但成本高昂且可能存在监测盲区。Jeong 等<sup>[87]</sup>研究了基于 IP 的车载网络架构, 指出该架构在智能交通系统中发挥的重要作用, 能够提升交通管理效率、预防事故并改善用户体验。

## 2.2 网络安全与优化

网络安全与优化领域包括地理位置验证、网络优化、攻击诊断、异常流量检测和加密流量分析等

技术。地理位置验证可以提升身份验证和访问控制的安全性; 网络优化通过位置数据优化网络覆盖能力和服务质量; 攻击诊断技术通过流量分析识别威胁; 异常流量检测可以识别潜在威胁, 加密流量分析支持安全评估。这些技术在提升网络安全和性能方面具有关键作用。

地理位置验证广泛应用于金融交易和身份验证等场景, 以提高安全性和防范欺诈事件。传统技术包括 GPS 和基站定位, 其中, 基于 GPS 的定位技术需要设备支持 GPS 功能, 并接收卫星信号。Scott 等<sup>[88]</sup>提出通过收集设备的 GPS 信号特征生成唯一的 GPS 签名来验证位置。然而, GPS 在高楼、封闭空间或恶劣天气下容易受到多路径传播和信号遮挡的影响, 导致定位准确性降低。基站定位技术通过手机与基站通信确定用户位置。Capkun 等<sup>[89]</sup>通过隐藏和移动基站进行安全位置验证, 无须目标设备知晓基站位置。这种方法虽然增强了安全性, 但基站定位仍存在信号干扰和基站密度不均等问题, 导致定位准确性较低。IP 地理定位方法通过查询 IP 地址数据库或调用 IP 地理定位服务提供商的 API 接口, 获取用户地理位置信息。Koch 等<sup>[90]</sup>研究了 IPv6 地址的地理定位和验证方法, 结合数据库查询和网络测量技术, 通过路径特征和时延数据提高定位准确性。尽管 IP 地理定位技术具备自动化和广泛适用性, 但受 IP 地址分配和代理服务影响, 存在一定误差。为提高地理位置验证的准确性, 可采用多源信息融合方法, 将 IP 地理定位与其他定位技术 (如 GPS 或基站定位) 结合, 并引入反欺诈算法和地理位置验证策略, 以提升验证准确性和可靠性, 从而有效降低地理位置验证的误差率和不确定性, 为网络安全提供可靠保障。

运营商网络优化通过技术手段评估和改进网络覆盖能力, 主要方法包括信号覆盖分析、容量规划和干扰管理。Amaldi 等<sup>[91]</sup>提出基于混合整数线性规划的 3G 蜂窝网络优化方法, 通过数学模型提升网络覆盖能力和容量, 有效减少盲区并提高用户满意度。然而, 覆盖分析虽然能够反映网络状态, 但对用户需求和负载情况的考虑不够全面。文献<sup>[92]</sup>结合深度确定性策略梯度和贝叶斯优化算法, 通过模拟优化网络覆盖能力和容量, 显著提升了网络性能与用户体验。Wang 等<sup>[93]</sup>利用大数据和机器学习技术实时调整 5G 网络参数, 在提升用户体验质量

(QoE, quality of experience) 和个性化服务的同时优化资源利用效率,但干扰源的复杂性增加了干预难度。运营商利用IP地理定位技术可以分析用户地理位置信息,评估网络覆盖能力。文献[94]对Geofeeds数据在IP地理定位中的有效性和可靠性进行研究,揭示了其优势与局限性。IP地理定位技术的高自动化和实时性优势,为运营商提供实时优化决策支持。此外,IP地理定位技术还具有成本低、适用范围广的特点,但其准确性受IP地址分配机制和代理服务器的影响。运营商需要将该技术与其他优化技术相结合并进行校准,以提高优化的准确性和效果。

网络攻击旨在窃取信息、破坏系统功能或侵犯隐私,威胁互联网安全。常见的攻击诊断方法包括行为分析、黑名单和入侵检测系统(IDS, intrusion detection system)。行为分析方法通过监控行为模式识别异常活动,Fayssal等<sup>[95]</sup>提出基于异常行为分析的无线网络安全方法,检测率达99%以上。黑名单技术通过屏蔽已知恶意来源提升安全性,Torabi等<sup>[96]</sup>研究了使用被动DNS数据检测滥用行为的系统。IDS则通过分析网络流量和日志发现异常,文献[97]对IDS及其应用进行了全面综述。虽然这些技术能有效应对威胁,但仍存在误报、漏报和对新型威胁反应不足等局限性。IP地理定位技术通过分析网络流量中的IP地址信息帮助追踪潜在攻击者和异常活动。DeCastro等<sup>[98]</sup>利用地理和时间特征,结合机器学习技术对IP地址进行恶意分类,识别潜在网络威胁,显著提高了恶意IP的识别准确率,为早期防御网络攻击提供了有效方法。

异常流量检测通过监测与正常网络流量模式不符的数据传输行为来识别潜在威胁,常用的检测技术包括基于规则的检测、基于特征的检测和机器学习方法。Duffield等<sup>[99]</sup>提出基于规则的检测方法,通过分析IP流量数据来识别网络异常,虽然提高了安全性,但对新型威胁的检测能力有限。文献[100]提出一种自动化特征选择方法,减少了手工操作,提高了对高维数据的处理能力。Shon等<sup>[101]</sup>结合多种机器学习算法提出混合检测方法,在检测准确性和响应速度方面表现优异,能够有效识别和应对网络威胁。IP地理定位技术在异常流量检测中的应用主要包括解析用户的IP地址以确定其地理位置,并基于该位置信息识别异常流量。Lison等<sup>[102]</sup>通过

被动DNS数据和神经声誉模型识别恶意域名和IP地址,提升了恶意活动检测能力。文献[103]采用同源检测方法分析IP地址特征与行为模式,有效识别了欺诈性IP地址。

加密流量分析旨在识别和解码加密数据传输中的信息内容。文献[104]评估了通过分析元数据检测恶意活动的技术。常见方法包括流量特征分析、元数据分析和机器学习算法。Muehlstein等<sup>[105]</sup>通过统计方法和机器学习方法分析网页加密流量,根据分组大小和时间模式推断用户操作系统、浏览器和应用程序信息,但在高加密强度或频繁变化的通信模式下准确性受限。Huoh等<sup>[106]</sup>利用GNN对加密网络流量进行分类,基于流量特征构建图并进行分类,能够准确识别和分类不同类型的加密流量,尤其在处理复杂网络流量模式时表现优异。IP地理定位技术通过解析IP地址确定地理位置,有助于监测和分析加密通信过程。该技术能够提供精准的地理位置信息,其自动化特性无须用户干预,提高了分析效率和用户体验。此外,IP地理定位技术适用于所有访问者,不受设备类型或用户身份限制,具有广泛的适用性。因此,IP地理定位技术在加密流量分析中的应用能够更精准地监控通信内容,提升网络安全性和数据保护水平。

### 2.3 网络空间测绘

网络空间测绘利用地理信息系统、大数据分析和空间建模技术来研究网络空间的地理特征和结构信息<sup>[107]</sup>。文献[108]探讨了网络拓扑与地理位置之间的关联性,发现拓扑分析在实际地理位置映射方面存在局限性,且大规模分析需要消耗大量时间和计算资源。相比之下,IP地理定位技术通过解析IP地址来确定网络资源和实体的地理位置,为地理信息采集与分析提供了有效手段。首先,通过查询IP数据库获取地理位置信息。然后,结合网络拓扑分析、流量分析和数据挖掘技术,进一步分析网络数据和通信行为,构建网络空间的地理和拓扑结构。最后,通过整合这些分析结果识别潜在的威胁和安全漏洞,为安全决策提供数据支持。IP地理定位技术凭借其较强的地理定位能力和自动化信息获取能力,不仅显著降低了数据分析的复杂度,还能适应多样化的网络环境和设备类型,具备广阔的应用前景和可靠性。

### 3 研究趋势与展望

尽管 IP 地理定位技术已取得显著进展,但仍面临很多挑战,主要包括定位准确性提升、数据隐私保护和网络攻击防御等方面。因此,针对这些挑战的深入研究不仅能推动相关学术理论和技术创新,还能促进 IP 地理定位技术的进一步发展与应用落地。

#### 3.1 技术创新及准确性提升

随着技术不断创新和发展,IP 地理定位技术正朝着准确性更高的方向演进。通过优化算法和改进定位技术,当前 IP 地理定位的准确性和可靠性已显著提升,一些先进技术甚至能够实现街道级别的精确定位,这为网络安全分析和流量分析提供了更精细的数据支持。未来研究可进一步改进基础设施,增加定位基站的部署密度,同时提升 GPS 接收器的性能,这将有效增强 IP 地理定位系统的可靠性和稳定性。此外,探索基于区块链的位置验证技术,利用区块链的不可篡改性和去中心化特性,确保 IP 地理定位的准确性和安全性。这些研究不仅为 IP 地理定位技术的发展提供了新思路,也将更好地满足日益增长的网络安全和流量分析需求。

#### 3.2 多源数据融合

通过整合搜索引擎查询数据、社交媒体数据等多种数据,可实现对网络行为和用户信息的全面分析,从而显著提升 IP 地理定位准确性。这种多源数据融合方法不仅优化了 IP 地理定位的准确性,还为广告定位和个性化推荐等应用提供技术支持。未来研究应聚焦于多源数据的整合和分析,利用先进的数据融合算法提升定位的准确性和稳定性,并深入研究数据源之间的关联性和可信度。此外,随着物联网终端设备的普及和智慧城市建设的推进,研究如何有效整合智能设备和传感器数据,将有助于提高定位的精准度,为智慧交通和智能医疗等应用场景提供精准的服务支持。

#### 3.3 实时监测与更新

随着网络环境和 IP 地址的动态变化,实时监测和动态更新对提高定位精准度越来越重要。传统的静态定位模型难以应对网络拓扑和 IP 分配的变化,通过机器学习与深度学习方法可实现自动化监测和更新,确保定位的准确性和时效性。未来研究应聚焦开发智能监测系统,利用机器学习与深度学习算法进行实时分析,设计自适应更新策略动态调

整定位模型。同时,结合传感器网络和物联网技术,通过传感器节点实时获取网络流量和信号数据,以更精准地推断 IP 地址的位置和活动状态。这将显著提高定位准确性,并为网络管理和安全防护提供有力支持。

#### 3.4 隐私保护

随着 IP 地理定位技术的广泛应用,如何加强用户隐私保护已成为亟待解决的关键问题,需要在隐私保护和定位准确性之间取得平衡,优化隐私保护算法,在保持定位准确性的前提下,降低计算和存储开销。此外,还需研究个性化的隐私保护方案,考虑用户的隐私偏好和风险承受能力,将隐私保护技术与 IP 地理定位技术结合,设计可配置的隐私保护模型,以满足不同用户群体的需求,并研究隐私保护与数据利用之间的平衡机制。随着人工智能和数据挖掘技术的发展,还应探索利用这些技术来增强隐私保护的效果和性能。

#### 3.5 跨平台与跨网络整合

随着网络技术的发展,未来互联网将变得更加多样化和异构化,这对 IP 地理定位技术提出了更高要求。相关研究应聚焦于开发适应不同网络环境的智能 IP 地理定位技术,设计自适应算法进行动态调整和优化,以满足多样化需求。同时,需要构建跨平台、跨网络的整合框架,实现信息共享,提升技术的智能化水平。随着物联网、智能城市和车联网等技术的快速发展,IP 地理定位技术将在智慧交通、智能家居等领域发挥更重要的作用。相关研究应探讨其与新兴技术的融合应用,并推动跨域合作和标准制定,促进 IP 地理定位技术的广泛应用和互联网生态系统的持续发展。

#### 3.6 可解释性

随着 IP 地理定位技术的持续发展,其可解释性问题逐渐成为研究的焦点。现有研究主要聚焦于提升 IP 地理定位技术的可解释性,使用户和决策者能够理解和信任定位结果。未来研究应侧重于设计和开发更加透明和可解释的 IP 地理定位模型和算法,以解释 IP 地理定位技术的决策流程和输出结果。此外,需要探索如何将专业领域知识和先验信息融合到 IP 地理定位模型中,以增强其可解释性和可信度。研究者还需要关注解释性与定位精确度之间的平衡,确保提升可解释性不影响定位准确性和性能。

### 3.7 通用实验平台

当前IP地理定位技术缺乏通用且公平的实验平台,这导致算法评估难以进行有效对比和验证。针对这一现状,未来研究应创建通用的实验平台,提供标准化数据集和开源代码,确保算法评估结果的可重复性与可比性。该平台应包括:1) 标准化数据集,涵盖多种网络环境和地理区域的数据,确保算法在不同场景下进行测试和评估;2) 开源代码实现,提供多个现有IP地理定位算法的开源代码,方便研究人员复现和改进;3) 基准测试,设计详细的评估指标和测试方案,确保不同算法在统一标准下进行比较;4) GitHub仓库,将部分代码和实验结果公开,提升研究透明度,并鼓励社区广泛参与和贡献。通过这些措施,不仅能建立IP地理定位算法的统一评估体系,还能推动领域创新和进步,形成合作研究生态系统,加速算法改进和实际部署。

## 4 结束语

本文系统综述了IP地理定位技术的研究与应用进展,涵盖数据挖掘与基准点分析、网络时延测量、机器学习与深度学习等领域。重点讨论了IP地理定位在地理定位服务、网络安全与优化、网络空间测绘中的应用。最后,文章对IP地理定位技术的未来发展方向进行展望,为相关研究与应用提供了重要参考。

### 参考文献:

- [1] PADMANABHAN V N, SUBRAMANIAN L. An investigation of geographic mapping techniques for Internet hosts[C]//Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. New York: ACM Press, 2001: 173-185.
- [2] SAXON J, FEAMSTER N. GPS-based geolocation of consumer IP addresses[C]//Passive and Active Measurement. Berlin: Springer, 2022: 122-151.
- [3] JIAO G Q, SONG S L, JIAO W H. Improving BDS-2 and BDS-3 joint precise point positioning with time delay bias estimation[J]. Measurement Science and Technology, 2020, 31(2): 025001.
- [4] SAPIEZYNSKI P, STOPCZYNSKI A, GATEJ R, et al. Tracking human mobility using Wi-Fi signals[J]. PLoS One, 2015, 10(7): e0130824.
- [5] SEN S, WANG J. Analyzing peer-to-peer traffic across large networks[C]//Proceedings of the IEEE/ACM Transactions on Networking. Piscataway: IEEE Press, 2004: 219-232.
- [6] GUO C, LIU Y, SHEN W, et al. Mining the web and the Internet for accurate IP address geolocations[C]//Proceedings of the IEEE INFOCOM 2009. Piscataway: IEEE Press, 2009: 2841-2845.
- [7] YIHUNIE F, ABDELFAHATTAH E, ODEH A. Analysis of Ping of death DoS and DDoS attacks[C]//Proceedings of the 2018 IEEE Long Island Systems, Applications and Technology Conference (LISAT). Piscataway: IEEE Press, 2018: 1-4.
- [8] LUCKIE M, HYUN Y, HUFFAKER B, et al. Traceroute probe method and forward IP path inference[C]//Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2008: 311-324.
- [9] ERIKSSON B, BARFORD P, SOMMERS J, et al. A learning-based approach for IP geolocation[C]//Passive and Active Measurement. Berlin: Springer, 2010: 171-180.
- [10] JIANG H, LIU Y Q, MATTHEWS J N. IP geolocation estimation using neural networks with stable landmarks[C]//Proceedings of the 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Piscataway: IEEE Press, 2016: 170-175.
- [11] POESE I, UHLIG S, ALI KAAFAR M, et al. IP geolocation databases: unreliable? [J]. ACM SIGCOMM Computer Communication Review, 2011, 41(2): 53-56.
- [12] SHAVITT Y, ZILBERMAN N. A geolocation databases study[J]. IEEE Journal on Selected Areas in Communications, 2011, 29(10): 2044-2056.
- [13] ZILBERMAN A, OFFER A, PINCU B, et al. A survey on geolocation on the Internet[J]. IEEE Communications Surveys & Tutorials, 2024, 72(3): 1235-1245.
- [14] 王占丰, 冯径, 邢长友, 等. IP定位技术的研究[J]. 软件学报, 2014, 25(7): 1527-1540.  
WANG Z F, FENG J, XING C Y, et al. Research on the IP geolocation technology[J]. Journal of Software, 2014, 25(7): 1527-1540.
- [15] 潘晓楠. IP定位技术研究分析[J]. 应用能源技术, 2015(6): 47-49.  
PAN X N. Analysis of IP positioning technology[J]. Applied Energy Technology, 2015(6): 47-49.
- [16] 奚源. IP溯源与定位技术综述[J]. 科技资讯, 2018, 16(31): 4-5.  
XI Y. Overview of IP traceability and location technology[J]. Science & Technology Information, 2018, 16(31): 4-5.
- [17] 王志豪, 张卫东, 文辉, 等. IP定位技术研究[J]. 信息安全学报, 2019, 4(3): 34-47.  
WANG Z H, ZHANG W D, WEN H, et al. A comprehensive survey of IP geolocation and evasion[J]. Journal of Cyber Security, 2019, 4(3): 34-47.
- [18] 林金磊, 李城龙, 宋光磊, 等. 客户端独立的IP地理定位研究综述[J]. 软件学报, 2025, 36(1): 321-340.  
LIN J L, LI C L, SONG G L, et al. Survey on client-independent IP geolocation[J]. Journal of Software, 2025, 36(1): 321-340.
- [19] ELLIOTT K. The who, what, where, when, and why of WHOIS: privacy and accuracy concerns of the WHOIS database[J]. SMU Science and Technology Law Review, 2008, 12: 141-172.
- [20] GHARIB A, SHARAFALDIN I, LASHKARI A H, et al. An evaluation framework for intrusion detection dataset[C]//Proceedings of the 2016 International Conference on Information Science and Security (ICISS). Piscataway: IEEE Press, 2016: 1-6.
- [21] YUAN F X, LIU F L, HUANG D H, et al. A high completeness PoP partition algorithm for IP geolocation[J]. IEEE Access, 2019, 7: 28340-28355.
- [22] KOMOSNY D, MRDOVIC S, ILKO P, et al. Testing Internet applications and services using PlanetLab[J]. Computer Standards & Inter-

- faces, 2017, 53: 33-38.
- [23] HUFFAKER B, FOMENKOV M, CLAFFY K. DRoP: DNS-based router positioning[J]. *ACM SIGCOMM Computer Communication Review*, 2014, 44(3): 5-13.
- [24] DAN O, PARIKH V, DAVISON B D. IP geolocation through geographic clicks[J]. *ACM Transactions on Spatial Algorithms and Systems*, 2022, 8(1): 1-22.
- [25] LEE Y, PARK H, LEE Y. IP geolocation with a crowd-sourcing broadband performance tool[J]. *ACM SIGCOMM Computer Communication Review*, 2016, 46(1): 12-20.
- [26] CHEN C, GUO C X, LIU Y X, et al. Mining the web for IP address geolocations[R]. 2007.
- [27] CHANDRASEKARAN B, BAI M, SCHOENFIELD M, et al. Alidade: IP geolocation without active probing[R]. 2015.
- [28] SOMMERS J. A web client perspective on IP geolocation accuracy[C]// *Proceedings of the 2020 International Symposium on Networks, Computers and Communications (ISNCC)*. Piscataway: IEEE Press, 2020: 1-8.
- [29] KATZ-BASSETT E, JOHN J P, KRISHNAMURTHY A, et al. Towards IP geolocation using delay and topology measurements[C]// *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*. New York: ACM Press, 2006: 71-84.
- [30] ZIVIANI A, FDIDA S, DE R J F, et al. Improving the accuracy of measurement-based geographic location of Internet hosts[J]. *Computer Networks*, 2005, 47(4): 503-523.
- [31] DONG Z Q, PERERA R D W, CHANDRAMOULI R, et al. Network measurement based modeling and optimization for IP geolocation[J]. *Computer Networks*, 2012, 56(1): 85-98.
- [32] ARIF M J, KARUNASEKERA S, KULKARNI S, et al. Internet host geolocation using maximum likelihood estimation technique[C]// *Proceedings of the 2010 24th IEEE International Conference on Advanced Information Networking and Applications*. New York: ACM Press, 2010: 422-429.
- [33] GILL P, GANJALI Y, WONG B. Dude, where's that IP? circumventing measurement-based IP geolocation[C]// *19th USENIX Security Symposium (USENIX Security 10)*. Berkeley: USENIX Association, 2010: 1-16.
- [34] ERIKSSON B, BARFORD P, MAGGS B, et al. Posit: a lightweight approach for IP geolocation[J]. *ACM SIGMETRICS Performance Evaluation Review*, 2012, 40(2): 2-11.
- [35] CHEN J N, LIU F L, LUO X Y, et al. A landmark calibration based IP geolocation approach[C]// *Proceedings of the 2015 10th International Conference on Availability, Reliability and Security*. Piscataway: IEEE Press, 2015: 411-416.
- [36] CHEN J N, LIU F L, SHI Y F, et al. Towards IP location estimation using the nearest common router[J]. *Journal of Internet Technology*, 2018, 19(7): 2097-2110.
- [37] ZHAO F, LUO X Y, GAN Y, et al. IP Geolocation based on identification routers and local delay distribution similarity[J]. *Concurrency and Computation: Practice and Experience*, 2019, 31(22): e4722.
- [38] WANG Z H, LI H, LI Q, et al. Towards IP geolocation with intermediate routers based on topology discovery[J]. *Cybersecurity*, 2019, 2(1): 13.
- [39] TIAN Y, DEY R, LIU Y, et al. China's Internet: topology mapping and geolocating[C]// *2012 Proceedings IEEE INFOCOM*. Piscataway: IEEE Press, 2012: 2531-2535.
- [40] TRIUKOSE S, ARDON S, MAHANTI A, et al. Geolocating IP addresses in cellular data networks[C]// *Passive and Active Measurement*. Berlin: Springer, 2012: 158-167.
- [41] 李明月, 罗向阳, 柴理想, 等. 基于网络节点聚类的目标IP城市级定位方法[J]. *计算机研究与发展*, 2019, 56(3): 467-479.
- LI M Y, LUO X Y, CHAI L X, et al. City-level IP geolocation method based on network node clustering[J]. *Journal of Computer Research and Development*, 2019, 56(3): 467-479.
- [42] DING S C, ZHAO F, LUO X Y. A street-level IP geolocation method based on delay-distance correlation and multilayered common routers[J]. *Security and Communication Networks*, 2021, 2021(1): 6658642.
- [43] ZU S D, LUO X Y, WANG L, et al. GUI: a geolocation method for unreachable IP[J]. *The Computer Journal*, 2024, 67(5): 1963-1978.
- [44] WANG Y, BURGNER D, FLORES M, et al. Towards street-level client-independent IP geolocation[C]// *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation*. New York: ACM Press, 2011: 365-379.
- [45] GUEYE B, ZIVIANI A, CROVELLA M, et al. Constraint-based geolocation of Internet hosts[C]// *Proceedings of the IEEE/ACM Transactions on Networking*. Piscataway: IEEE Press, 2006: 1219-1232.
- [46] LAKI S, MÁTRAY P, HÁGA P, et al. Spotter: a model based active geolocation service[C]// *2011 Proceedings IEEE INFOCOM*. Piscataway: IEEE Press, 2011: 3173-3181.
- [47] LI D, CHEN J, GUO C X, et al. IP-geolocation mapping for moderately connected Internet regions[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2013, 24(2): 381-391.
- [48] HU Z, HEIDEMANN J, PRADKIN Y, et al. Towards geolocation of millions of IP addresses[C]// *Proceedings of the 2012 Internet Measurement Conference*. New York: ACM Press, 2012: 123-130.
- [49] DU F, BAO X G, ZHANG Y Z, et al. GeoCET: accurate IP geolocation via constraint-based elliptical trajectories[C]// *Collaborative Computing: Networking, Applications and Worksharing*. Berlin: Springer, 2019: 603-622.
- [50] JIANG Y, FANG B X, HU M Z, et al. Techniques for determining the geographic location of IP addresses in ISP topology measurement[J]. *Journal of Computer Science and Technology*, 2005, 20(5): 689-701.
- [51] LIU C, LUO X Y, YUAN F X, et al. RNBG: a ranking nodes based IP geolocation method[C]// *Proceedings of the IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. Piscataway: IEEE Press, 2020: 80-84.
- [52] CALLEJO P, GRAMAGLIA M, CUEVAS R, et al. A deep dive into the accuracy of IP geolocation databases and its impact on online advertising[J]. *IEEE Transactions on Mobile Computing*, 2023, 22(8): 4359-4373.
- [53] ZHANG F, YIN M J, LIU F L, et al. IP2vec: an IP node representation model for IP geolocation[J]. *Frontiers of Computer Science*, 2023, 18(6): 186506.
- [54] WANG T, XU K, SONG J D, et al. An optimization method for the geolocation databases of Internet hosts based on machine learning[J]. *Mathematical Problems in Engineering*, 2015, 2015(1): 972642.
- [55] PAULE J D G, SUN Y R, MOSHFEGHI Y. On fine-grained geolocation of tweets and real-time traffic incident detection[J]. *Information Processing & Management*, 2019, 56(3): 1119-1132.
- [56] LI R X, XU R, MA Y Y, et al. LandmarkMiner: street-level network landmarks mining method for IP geolocation[J]. *ACM Transactions on Internet of Things*, 2021, 2(3): 1-22.

- [57] HONG A, LI Y H, ZHANG H, et al. A cheap and accurate delay-based IP geolocation method using machine learning and looking glass[C]// Proceedings of the 2023 IFIP Networking Conference (IFIP Networking). Piscataway: IEEE Press, 2023: 1-9.
- [58] WANG Z H, LI Q, SONG J K, et al. Towards IP-based geolocation via fine-grained and stable webcam landmarks[C]// Proceedings of the Web Conference 2020. New York: ACM Press, 2020: 1422-1432.
- [59] LIU M H, SUN Y C, LIU C, et al. IPv6 landmark mining based on domain name screening and IPv4-IPv6 connection[C]// Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering. New York: ACM Press, 2021: 338-343.
- [60] KE G L, MENG Q, FINLEY T, et al. LightGBM[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM Press, 2017: 3149-3157.
- [61] LI R X, YUAN X Y, YIN M J, et al. Mobile IP geolocation based on district anchor without cooperation of users or Internet service providers[J]. IEEE/ACM Transactions on Networking, 2024, 32(6): 5507-5523.
- [62] PAAß G, HECKER D. Deep learning can recognize complex relationships[C]// Artificial Intelligence. Berlin: Springer, 2023: 75-112.
- [63] WEYAND T, KOSTRIKOV I, PHILBIN J. PlaNet-photo geolocation with convolutional neural networks[C]// Computer Vision - ECCV 2016. Berlin: Springer, 2016: 37-55.
- [64] EBRAHIMI M, SHAFIEIBAVANI E, WONG R, et al. A unified neural network model for geolocating[C]// Proceedings of the 22nd Conference on Computational Natural Language Learning. Stroudsburg: ACL Press, 2018: 42-53.
- [65] THOMAS P, HENNIG L. Twitter geolocation prediction using neural networks[C]// Language Technologies for the Challenges of the Digital Age. Berlin: Springer, 2018: 248-255.
- [66] ZHOU F, ZHANG W F, WANG Y, et al. Identifying IP usage scenarios: problems, data, and benchmarks[J]. IEEE Network, 2022, 36(3): 152-158.
- [67] ZHANG F, LIU F L, LUO X Y. Geolocation of covert communication entity on the Internet for post-steganalysis[J]. EURASIP Journal on Image and Video Processing, 2020(1): 15-25.
- [68] ZU S D, LUO X Y, ZHANG F. IP-geolocator: a more reliable IP geolocation algorithm based on router error training[J]. Frontiers of Computer Science, 2021, 16(1): 161504.
- [69] ARIK S Ö, PFISTER T. TabNet: attentive interpretable tabular learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(8): 6679-6687.
- [70] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.
- [71] MA Z R, ZHANG S C, LI N, et al. GraphNEI: a GNN-based network entity identification method for IP geolocation[J]. Computer Networks, 2023, 235: 109946.
- [72] DING S C, LUO X Y, WANG J W, et al. GNN-geo: a graph neural network-based fine-grained IP geolocation framework[J]. IEEE Transactions on Network Science and Engineering, 2023, 10(6): 3543-3560.
- [73] WANG Z Y, ZHOU F, ZENG W X, et al. Connecting the hosts: street-level IP geolocation with graph neural networks[C]// Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2022: 4121-4131.
- [74] TAI W X, CHEN B, ZHONG T, et al. RIPGeo: robust street-level IP geolocation[C]// Proceedings of the 2023 24th IEEE International Conference on Mobile Data Management (MDM). Piscataway: IEEE Press, 2023: 138-147.
- [75] TAI W X, CHEN B, ZHOU F, et al. TrustGeo: uncertainty-aware dynamic graph learning for trustworthy IP geolocation[C]// Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2023: 4862-4871.
- [76] LIU L, HE Y H, ZHANG L, et al. GraphCyber: identifying IP usage scenarios for cyberspace mapping[C]// Proceedings of the ICC 2024 - IEEE International Conference on Communications. Piscataway: IEEE Press, 2024: 3310-3315.
- [77] ZHOU F, WANG T L, ZHONG T, et al. Identifying user geolocation with hierarchical graph neural networks and explainable fusion[J]. Information Fusion, 2022, 81: 1-13.
- [78] ALZUBAIDI L, ZHANG J L, HUMAIDI A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions[J]. Journal of Big Data, 2021, 8(1): 1-74.
- [79] GOODCHILD M F. Citizens as sensors: the world of volunteered geography[J]. GeoJournal, 2007, 69(4): 211-221.
- [80] TUROW J, KING J, HOOFNAGLE C J, et al. Americans reject tailored advertising and three activities that enable it[R]. 2009.
- [81] UMATE N B, BHUJADE V G. A real time technique for targeted advertising using location-based services: for GPS enabled device: a review[C]// Proceedings of the 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA). Piscataway: IEEE Press, 2017: 689-693.
- [82] KLIMAN-SILVER C, HANNAK A, LAZER D, et al. Location, location, location: the impact of geolocation on web search personalization[C]// Proceedings of the 2015 Internet Measurement Conference. New York: ACM Press, 2015: 121-127.
- [83] KHODABANDEHLOU S, HASHEMI G S A, ZIVARI R M. An effective recommender system based on personality traits, demographics and behavior of customers in time context[J]. Data Technologies and Applications, 2020, 55(1): 149-174.
- [84] KESORN K, JURAPHANTHONG W, SALAIWARAKUL A. Personalized attraction recommendation system for tourists through check-in data[J]. IEEE Access, 2017, 5: 26703-26721.
- [85] LEIVA L A, ALABAU V. Automatic internationalization for just in time localization of web-based user interfaces[J]. ACM Transactions on Computer-Human Interaction, 2015, 22(3): 1-32.
- [86] ROSENBAUM D, KURZ F, THOMAS U, et al. Towards automatic near real-time traffic monitoring with an airborne wide angle camera system[J]. European Transport Research Review, 2009, 1(1): 11-21.
- [87] JEONG J, SHEN Y W, OH T, et al. A comprehensive survey on vehicular networks for smart roads: a focus on IP-based approaches[J]. Vehicular Communications, 2021, 29: 100334.
- [88] SCOTT L. Proving location using GPS location signatures: why it is needed and a way to do it[C]// Proceedings of the 26th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2013). Nashville: Institute of Navigation, 2013: 2880-2892.
- [89] CAPKUN S, RASMUSSEN K, CAGALJ M, et al. Secure location verification with hidden and mobile base stations[J]. IEEE Transactions on Mobile Computing, 2008, 7(4): 470-483.
- [90] KOCH R, GOLLING M, RODOSEK G D. Geolocation and verifica-

- tion of IP-addresses with specific focus on IPv6[C]//Cyberspace Safety and Security. Berlin: Springer, 2013: 151-170.
- [91] AMALDI E, CAPONE A, MALUCELLI F. Radio planning and coverage optimization of 3G cellular networks[J]. *Wireless Networks*, 2008, 14(4): 435-447.
- [92] DREIFUERST R M, DAULTON S, QIAN Y C, et al. Optimizing coverage and capacity in cellular networks using machine learning[C]//Proceedings of the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2021: 8138-8142.
- [93] WANG Y, LI P L, JIAO L, et al. A data-driven architecture for personalized QoE management in 5G wireless networks[J]. *IEEE Wireless Communications*, 2017, 24(1): 102-110.
- [94] LIVADARIU I, VERMEULEN K, MOUCHET M, et al. Geofeeds: revolutionizing IP geolocation or illusionary promises?[J]. *Proceedings of the ACM on Networking*, 2024, 2(3): 1-21.
- [95] FAYSSAL S, HARIRI S, AL-NASHIF Y. Anomaly-based behavior analysis of wireless network security[C]//Proceedings of the 2007 Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services (MobiQuitous). Piscataway: IEEE Press, 2007: 1-8.
- [96] TORABI S, BOUKHTOUTA A, ASSI C, et al. Detecting Internet abuse by analyzing passive DNS traffic: a survey of implemented systems[J]. *IEEE Communications Surveys & Tutorials*, 2018, 20(4): 3389-3415.
- [97] ALKASASSBEH M, AL-HAJ B S. Intrusion detection systems: a state-of-the-art taxonomy and survey[J]. *Arabian Journal for Science and Engineering*, 2023, 48(8): 10021-10064.
- [98] DECASTRO-GARCÍA N, ESCUDERO G D, CARRIEGOS M V. A mathematical analysis about the geo-temporal characterization of the multi-class maliciousness of an IP address[J]. *Wireless Networks*, 2024, 30(6): 5033-5048.
- [99] DUFFIELD N, HAFFNER P, KRISHNAMURTHY B, et al. Rule-based anomaly detection on IP flows[C]//Proceedings of the IEEE INFOCOM 2009. Piscataway: IEEE Press, 2009: 424-432.
- [100] NAKASHIMA M, SIM A, KIM Y, et al. Automated feature selection for anomaly detection in network traffic data[J]. *ACM Transactions on Management Information Systems*, 2021, 12(3): 1-28.
- [101] SHON T, MOON J. A hybrid machine learning approach to network anomaly detection[J]. *Information Sciences*, 2007, 177(18): 3799-3821.
- [102] LISON P, MAVROEIDIS V. Neural reputation models learned from passive DNS data[C]//Proceedings of the 2017 IEEE International Conference on Big Data (Big Data). Piscataway: IEEE Press, 2017: 3662-3671.
- [103] MA Y Y, CHEN A, HOU C Z, et al. DC-FIPD: fraudulent IP identification method based on homology detection[J]. *Computers, Materials & Continua*, 2024, 81(2): 3301-3323.
- [104] PAPADOGIANNAKI E, IOANNIDIS S. A survey on encrypted network traffic analysis applications, techniques, and countermeasures[J]. *ACM Computing Surveys*, 2021, 54(6): 1-35.
- [105] MUEHLSTEIN J, ZION Y, BAHUMI M, et al. Analyzing HTTPS encrypted traffic to identify user's operating system, browser and application[C]//Proceedings of the 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC). Piscataway: IEEE Press, 2017: 1-6.
- [106] HUOH T L, LUO Y, LI P L, et al. Flow-based encrypted network traffic classification with graph neural networks[J]. *IEEE Transactions on Network and Service Management*, 2023, 20(2): 1224-1237.
- [107] ZHANG L, WANG G X, YOU X, et al. Research on the cyberspace map and its conceptual model[J]. *ISPRS International Journal of Geo-Information*, 2023, 12(353): 1-16.
- [108] GRANDIN A. Cyberspace geography and cyber terrain: challenges producing a universal map of cyberspace[J]. *European Conference on Cyber Warfare and Security*, 2023, 22(1): 207-213.

## [作者简介]



刘学婷 (1992-), 女, 山西大同人, 电子科技大学博士生, 主要研究方向为 IP 地理定位和行人重识别。



台文鑫 (1997-), 男, 甘肃兰州人, 电子科技大学博士生, 主要研究方向为可控生成概率模型和可信人工智能。



周帆 (1981-), 男, 四川眉山人, 博士, 电子科技大学教授, 主要研究方向为机器学习、神经网络、时空数据分析、图学习、推荐系统、社交网络数据挖掘。



王永 (1981-), 男, 河南郑州人, 博士, 香港科技大学在站博士后, 主要研究方向为数据挖掘、数据安全、机器学习。



陈凯 (1980-), 男, 安徽马鞍山人, 博士, 香港科技大学教授, 主要研究方向为数据中心网络、机器学习系统和隐私保护计算。