

紧凑数据结构专题序言

郑嘉琦, 谢鲲, 杨仝, 黄河, 李福亮, 李猛, 戴海鹏

引用本文

郑嘉琦, 谢鲲, 杨仝, 黄河, 李福亮, 李猛, 戴海鹏. [紧凑数据结构专题序言](#)[J]. 计算机科学, 2024, 51(4): 1-3.

相似文献推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[一种基于部分数据的多级剪枝Obfs4混淆流量识别方法](#)

Multi-level Pruning Obfs4 Obfuscated Traffic Recognition Method Based on Partial Data

计算机科学, 2024, 51(4): 39-47. <https://doi.org/10.11896/jsjcx.231000118>

[分布式网络中连续时间周期的全局top-K频繁流测量](#)

Global Top-K Frequent Flow Measurement for Continuous Periods in Distributed Networks

计算机科学, 2024, 51(4): 28-38. <https://doi.org/10.11896/jsjcx.231000119>

[基于MapReduce的大规模网络社区发现算法](#)

Large-scale Network Community Detection Algorithm Based on MapReduce

计算机科学, 2024, 51(4): 11-18. <https://doi.org/10.11896/jsjcx.231100049>

[学习型过滤器综述](#)

Survey of Learning-based Filters

计算机科学, 2024, 51(1): 41-49. <https://doi.org/10.11896/jsjcx.231000202>

[过滤器数据结构研究综述](#)

Filter Data Structures:A Survey

计算机科学, 2024, 51(1): 35-40. <https://doi.org/10.11896/jsjcx.231000193>

紧凑数据结构专题序言

郑嘉琦¹ 谢 鲲² 杨 仝³ 黄 河⁴ 李福亮⁵ 李 猛¹ 戴海鹏¹

1 南京大学 南京 210023

2 湖南大学 长沙 410082

3 北京大学 北京 100871

4 苏州大学 江苏 苏州 215006

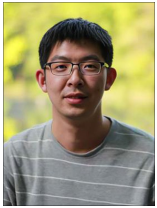
5 东北大学 沈阳 110006

随着现代计算机科学的迅猛发展,紧凑型数据结构逐渐成为研究和应用的热点领域,其核心目标在于尽可能减少内存占用并提升运行效率,以支撑广泛的计算机应用需求。通过精心的设计和优化,这类数据结构能够在限定的内存空间内存储大量数据,并实现高效的数据操作和访问,从而展现出巨大的潜力。具体来说,在嵌入式系统、计算机网络、数据库系统以及数据挖掘等多个领域,紧凑型数据结构不仅能有效解决空间受限和性能低下的问题,而且在推动相关技术的进步和应用扩展方面起到了关键作用。因此,随着科研的推进以及工业界对资源使用效率的持续追求,紧凑型数据结构的研究与应用正吸引着全球学术界和工业界的广泛关注。在此背景下,《计算机科学》推出了紧凑数据结构专题,旨在深入探讨其在计算机科学领域中的重要性和广阔的应用前景。该专题通过邀请和开放投稿的方式,精选了6篇高质量的论文,主要分为两类:一类是针对流式数据应用设计的紧凑数据结构的数据统计方法;另一类是为网络应用设计的紧凑数据结构的流量测量方法。此专题为国内外的专家学者提供了一个交流合作和发布最新研究成果的平台,进一步促进了紧凑数据结构领域的学术交流和科技进步。

为了有效应对计算机应用发展中涌现的海量流式数据,尤其是在对这些数据进行有效统计与挖掘方面,流数据处理技术成为了关键工具。这些技术至关重要,因为它们支持对数据的实时分析和决策制定过程。然而,流式数据的无限性和对处理效率的高要求带来了不小的挑战,导致在数据挖掘和统计方面仍存在诸多问题。为此,福州大学的陈昕杨等提出了一种紧凑数据结构——IntervalSketch,旨在以较低的时空开销进行流式数据中间隔项的 top-K 统计。该方法通过模拟退火技术对数据进行分块,加速统计过程,并利用 Sketch 技术对间隔项进行有效存储,实现了数据流间隔项的高效近似统计。进一步地,基于紧凑数据结构提供的高性能流式处理能力,南京大学的王瀚橙等提出了一种创新的基于 MapReduce 的 k 中心聚类分布式技术。该技术能够有效实现社区发现算法,提高算法的执行效率。此外,陆军工程大学的宋金玉等针对通用型与行业型数据质量的测量框架进行了深入研究和总结。他们提出了一套构建针对特定领域数据质量测量框架的准则,并依据这些准则构建了相应的测量框架。这不仅能使紧凑数据结构在流式数据处理场景中得到更加系统的应用,也为流式数据处理的高效性与准确性提供了有力支持。

随着互联网技术的日益更新,网络流量达到了前所未有的高峰。在巨大的数据压力下,网络流量测量在入侵检测、路由选择、网络质量维护以及网络设备管理等方面都有着深远的意义。此外,新型硬件设备如可编程交换机的出现赋予了网络节点一定的计算和处理能力,使得在节点上进行流量测量成为可能。然而,由于网络设备的特性,网络场景下的数据测量面临着较大的资源限制,使得一些测量方法在实际网络场景中的应用受限。因此,不同场景下的高效网络流量测量方法成为学术界研究的热点。针对分布式网络的场景,苏州大学的毛晨宇等通过在分布式节点部署紧凑的概率数据结构来测量网络流量,在中心节点定时通信,收集信息并进行处理,实现了在分布式网络中对于连续时间中期的 top-K 频繁流测量。在此基础上,利用不同的优化方法减少传输开销,以保证测量方法的高效性。另外,如何对特定的流量进行识别在网络测量领域中也得到了广泛关注。苏州大学的徐宸涵等提出了一种基于部分数据多级剪枝的 Obfs4 混淆流量识别方法。为了识别网络中的突发流量,福州大学的吴艳妮等提出了一种基于紧凑数据结构 Sketch 的测量框架 RBF Radar,并基于 PISA 架构实现了 RBF Radar 以验证其在实际网络场景中的适用性。

近年来,我国在紧凑数据结构领域的研究成果不断涌现,此领域在计算机科学中的应用及其影响日益广泛。希冀本专题能够为紧凑数据结构领域的持续研究与发展注入新活力,为学术界与产业界的协同创新架起桥梁,助力该领域持续向前、蓬勃发展。展望未来,期盼我国科研人员在此领域能够有更多突破,攀登更高峰,并有效地将研究成果转化为实际应用,从而为构建数字化中国提供坚实支撑。



郑嘉琦 南京大学计算机科学与技术系助理教授、特聘研究员、博士生导师, IEEE/CCF/通信学会高级会员, 中国通信标准化协会 TC610 开放网络设备工作组成员, CCF 网络与数据通信专委会委员, CCF 互联网专委会委员, CCF-B 类期刊 Frontiers of Computer Science 青年编委。目前在计算机网络领域 ACM SIGCOMM, ACM CoNEXT, IEEE INFOCOM, IEEE ICNP 等国际刊物上发表 CCF-A/B 类论文 70 余篇, 3 次获得 CCF 推荐国际会议最佳论文奖。承担科技部重点研发计划青年科学家项目, 国家自然科学基金青年、面上项目, 华为、阿里巴巴、蚂蚁金服等企业项目。曾获江苏省科学技术一等奖、华为“难题揭榜”火花奖、ACM SIGCOMM 中国新星奖、“小米青年学者”科技创新奖、CCF 优博奖、江苏省优博奖等荣誉。



谢 鲲 湖南大学二级教授, 国家杰出青年基金获得者, 湖南省杰出青年基金获得者, 湖南省青年骨干教师, 湖南省优秀硕士生导师, 长沙市“巾帼建功”标兵, 湖南大学岳麓学者, 湖南大学科研标兵。目前担任超算与人工智能融合计算教育部重点实验室主任。研究方向为计算机网络、网络安全、大数据和人工智能。先后主持国家自然科学基金重点、杰青、面上、青年项目, 湖南省重点研发计划、湖南省基金重点项目, 以及多项华为技术开发项目。坚持科研一线工作, 授权发明专利 37 项, 在国际主流会议和期刊发表论文 170 余篇, 其中 CCF A 类期刊和会议 46 篇。科研成果应用于多个实际平台。两次获得湖南省科学技术进步奖二等奖, 一次获得吴文俊人工智能科技进步一等奖。



杨 全 北京大学计算机系网络所研究员, 博士生导师, 教育部青年长江学者。主要研究网络、大模型、大数据等方向的新型概率数据结构, 以可控且小的误差换取时间和空间上的显著降低。近年在 SIGCOMM, SIGMOD, SIGKDD 上以第一作者/通信作者发表论文 24 篇。负责 1 项国家自然科学基金重点支持项目, 3 项国家重点研发计划课题/子课题, 国家自然科学基金面上基金、青年基金, 以及华为、中兴、今日头条等项目。研究成果在国家自然科学基金委网站发布 2 次, 在教育部科技发展中心发布 1 次, 多项研究成果在华为海思、华为产品线、今日头条、Redis 数据库中得到了应用部署。



黄 河 苏州大学特聘教授、计算机科学与技术学院副院长(主持工作), IEEE 高级会员, ACM 会员, 中国计算机学会高级会员。2019—2020 年为美国佛罗里达大学访问学者, 2021 年入选教育部青年长江学者。近年来专注网络流量测量、网络资源的优化分配和调度等方向的研究, 在 SIGMOD, MOBICOM, INFOCOM, ICML, NeurIPS, VLDB, ToN, TMC 等国际会议/期刊上共发表论文 150 余篇, 其中 CCF A 类顶级会议/期刊 50 余篇。担任 INFOCOM, ICC, MASS 等多个知名国际会议的程序委员会成员 40 余次。多次担任 INFOCOM, MobiHoc 等国际会议 workshop 的程序委员会主席或宣传主席。担任 IEEE Transactions on Network Science and Engineering (TNSE) 客座主编。曾获得 2022 年度安徽省自然科学一等奖(排名第二)、2014 年度国家教育部自然科学二等奖(排名第五), 以及多个著名国际会议的最佳论文奖。



李福亮 东北大学计算机科学与工程学院教授、博士生导师, 华为中央研究院专家顾问、华为数通产品线智库顾问, CCF 互联网专委会委员、CCF 体系结构专委会委员。主要从事未来网络技术、网络智能运维、网络安全等领域的研究。近年来在各大国际期刊/会议上发表学术论文 50 余篇, 包括 TON, JSAC, TC, TSC, TVT, TCC, NSDI, INFOCOM, ICNP, ICDCS, IWQoS 等。申请发明专利 20 余项, 出版学术专著 6 部。作为项目负责人主持国家级项目 6 项, 包括国家自然科学基金重点、面上、青年项目以及科技部重点研发计划项目/课题等, 主持知名企业研究类项目 7 项, 包括华为公司、阿里巴巴、字节跳动等。获辽宁省百千万人才、辽宁省通信协会优秀科技工作者、沈阳市中青年科技创新人才等荣誉 10 余项。



李 猛 南京大学计算机系助理研究员,2021—2022年在新加坡南洋理工大学担任研究助理。主要研究方向为数据密集的物联网系统,关注设计概率型和学习型索引以提升物联网数据查询性能,这些索引在物联网数据查询方面取得了显著的性能提升,并获得了授权发明专利5项。部分专利技术进行了落地转化,并获得了教育部“互联网+”大赛银奖,相关工作已在 TON, VLDBJ, TKDE, WWW, ICDE 和 VDLB 等国内外优秀期刊和会议上发表。担任 ACM CIKM, IEEE ISPA, IEEE MSN 等国际会议程序委员。



戴海鹏 南京大学计算机科学与技术系副教授、博导,国家级青年人才计划入选者。获 ACM 中国新星奖、IEEE 可扩展计算技术委员会职业中期卓越研究成就奖、中国电子学会优秀科技工作者等荣誉。研究方向为物联网、数据挖掘、移动计算等。在国际著名会议/期刊上发表论文 200 余篇,含 CCF A 类 100 余篇,包括 UbiComp, INFOCOM, SIGMOD, VLDB, ICDE, WWW, EuroSys, ATC, TON, JSAC, TPDS, TMC, VLDBJ, TKDE 等国际一流会议/期刊。曾获 CCF A 类会议 INFOCOM 最佳论文提名奖,CCF A 类会议 UbiComp 杰出论文奖,CCF B 类会议 ICNP 最佳论文奖,CCF B 类会议 SECON 最佳论文奖亚军等。担任国家重点研发计划项目课题负责人,主持和承担国家自然科学基金面上、联合基金、重点等项目十余项。担任 ACM SIGCOMM China 秘书长、中国计算机学会物联网专委会常委、网络与数据通信专委会常委、中文 CCF A 类期刊《电子学报》青年编委等职务,担任 ISPA, HPCC 等会议主席职务十余次。