

前 言

随着新一轮科技革命与产业变革的深入发展,以互联网为代表的新型信息技术正在深刻地改变社会各领域,数据作为关键生产要素的价值日益凸显,信息化被赋予新的内涵.为加速传统产业向智能化、绿色化、融合化方向转型升级,亟需利用新型信息技术提供数字化服务,数据是促成这一转型新的动力能源(New Power).数据是人们对客观世界的认知在计算机世界中的表示,是人类社会继“蒸汽能”“电能”后的第三次新能源——数据能.

近年来,国内外学者在数据管理的基础理论与关键技术方面取得了较好进展,数据赋能多个行业领域实现数字化转型,以BAT为代表的数字化原生企业为数字化转型树立了榜样,这也是我国“互联网+”行动计划的底层逻辑.支持互联网应用的平台和App也为未来的企业信息系统发展指明了方向.当前,我们正处在数字化转型的关键时期,支撑各行业信息系统精准服务的数据需求日趋多元.如何在数据搜集与使用的过程中突破屏障,挖掘出海量数据中蕴含的价值,支持信息系统有效利用数据进行决策、创新?这给数据驱动的信息系统服务带来了新的研究机遇与挑战.为了推动我国学者在数据驱动的信息系统领域相关研究,及时报道最新的研究成果,我们组织了本期“数据驱动的信息系统服务”专题,旨在推动我国学者聚焦数据驱动的信息系统在提供精准服务过程中产生的新方法和新技术,围绕信息系统的两个重要组成:数据与服务,为相关研究者和实践者提供集中发表最新成果的平台.本专题于2023年7月征稿结束,经过一轮专题编委会评审、两轮通讯评审以及WISA2023会议宣读和终审环节,最终有11篇论文入选本专题.

在数据驱动信息系统服务的安全与隐私方面,申远等人的论文“支持访问行为身份追踪的跨域密文共享方案”提出了基于密文策略属性基加密机制的支持访问行为身份追踪的跨域密文共享方案,并引入倒排索引结构优化访问行为身份追踪效率,通过聚合签名和隐私集合交集思想实现索引查询的隐私保护.王吉宏等人的论文“基于信息瓶颈理论的鲁棒少标签虚假信息检测”针对虚假信息检测问题,提出一种基于信息瓶颈理论的鲁棒少标签虚假信息检测方法,通过互信息最大化技术融合无标注样本信息,克服虚假信息检测对标签的过分依赖问题,并通过对抗训练的策略模拟虚假信息传播者的恶意操纵行为.傅培旺等人的论文“基于本地差分隐私的分布式图统计采集算法”针对图结构数据的多统计指标安全高效采集,提出基于本地差分隐私的分布式图统计采集算法,同时实现度分布、三角计数序列和聚类系数3个不同统计指标采集,并适应不同有效性和隐私保护的需求.程渝栋等人的论文“基于半监督学习的未知异常检测方法”针对现有的半监督异常检测方法在识别未知异常方面的不足以及表格数据的特性,提出了一种基于半监督学习的未知异常检测方法,旨在同时识别已知和未知异常,该方法利用闭集分类器对已知异常和正常分类,利用未知异常检测器检测未知异常.此外,还考虑了异常场景中异常和正常极端不平衡的情况,设计了有效的数据增强方法来扩充异常样本的数量.

在数据驱动信息系统服务的推荐与决策方面,高云帆等人的论文“大语言模型驱动的选址推荐系统”提出了一个大语言模型驱动的选址推荐系统.对当前推荐系统的应用场景进行

了拓展,展示了如何通过外部工具来让 LLM 在地理空间任务中理解具体位置同时避免虚构地名地址.将大语言整合到选址推荐系统中,为用户提供更加智能、灵活和全面的选址解决方案.王贝伦等人的论文“面向信息系统推荐与决策的高阶张量分析方法”提出了稀疏低秩张量回归模型方法,该方法直接对问题施加结构约束,同时还提出了两层并行解决方案来有效地求解模型,能够有效支撑信息系统服务的高效推荐与决策.黄玲等人的论文“图卷积宽度跨域推荐系统”针对信息系统中的跨域推荐问题,提出了一个基于图卷积宽度跨域推荐系统,通过图卷积网络学习用户与项之间的高阶关系,以聚合领域信息,并提出了多个优化方法提高系统整体效率.赵容梅等人的论文“基于对比学习的多兴趣感知序列推荐系统”提出了一个基于对比学习的多兴趣感知序列推荐模型,从序列内的局部偏好和序列间的全局偏好出发,通过对比学习机制更精准地获取用户的偏好信息,进而进行更精准的推荐.

在数据驱动信息系统服务的交互技术方面,王晨泽等人的论文“基于定位数据的全景超分辨率图像交互可视化框架”提出了一个基于定位数据的全景超分辨率图像交互可视化框架,旨在普通计算机上实现大规模定位数据的快速交互可视化.该框架基于随机采样策略构建定位数据的多分辨率层级结构,以交互方式可视化多尺度全景超分辨率图像;同时采用分块策略和多线程并行策略,分批次处理大规模定位数据,既防止内存溢出又加快了处理速度.

在面向信息系统服务的数据处理方面,潘青峰等人的论文“基于查询编译的 SQL 执行技术研究进展”从信息系统中的数据管理核心技术切入,系统地综述了编译执行技术的研究进展.论文首先概述了查询编译的基本概念,随后从代码生成、中间表示、机器码生成与运行 3 个角度对查询编译现有工作进行了总结分析,论文还总结了该领域的前沿研究方向,同时展望了查询编译未来可能的发展方向.

在数据驱动信息系统服务的领域应用方面,巢成等人的论文“基于空间位置关系的轨迹数据高效降维和查询算法”针对信息系统服务的重要应用领域基于位置的服务中的降维算法和查询技术,提出了基于网格划分的均匀网格降维算法和非均匀网格降维算法,基于降维后得到的数据,采取先过滤筛选后遍历查询的方法,设计实现了实际应用场景中常用的范围查询与最近邻查询算法.

本专题汇集了数据驱动的信息系统服务研究中涉及的理论基础和关键技术,反映了我国学者在该领域的最新研究进展.由于专题征集时间和篇幅有限,无法全面覆盖领域的近期研究进展,不足之处敬请各位学者谅解和批评指正.感谢《计算机研究与发展》编委会和 CCF 信息系统专业委员会对专题工作的指导和帮助,感谢专题全体评审专家及时、耐心、细致的评审工作,感谢踊跃投稿的所有作者.希望本专题能够对数据驱动的信息系统服务相关领域的研究与实践工作有所促进.

周傲英 教授 (华东师范大学)

杨世宇 教授 (广州大学)

张岩峰 教授 (东北大学)

毛嘉莉 教授 (华东师范大学)

2024 年 6 月