软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2021,32(3):601−603 [doi: 10.13328/j.cnki.jos.006187]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

支撑人工智能的数据管理与分析技术专刊前言*

陈 雷[1], 王宏志[2], 童咏昕[3], 高 宏[2]

[1](香港科技大学 计算机科学与工程学系,香港 999077)

[2](哈尔滨工业大学计算学部,黑龙江 哈尔滨 150001)

[3](北京航空航天大学 计算机学院,北京 100191)

通讯作者: 王宏志, E-mail: wangzh@hit.edu.cn

中文引用格式: 陈雷,王宏志,童咏昕,高宏.支撑人工智能的数据管理与分析技术专刊前言.软件学报,2021,32(3):601−603. http://www.jos.org.cn/1000-9825/6187.htm

近年来,支撑人工智能的数据管理与分析技术正成为大数据和人工智能领域研究的热点问题之一.利用和发展数据管理与分析理论技术,为提升人工智能系统全生命周期的效率和有效性提供基础性支撑,必将进一步促进基于大数据的人工智能技术发展与其在更大范围的推广应用.本专刊聚焦在数据管理与人工智能融合发展的过程中,数据库技术对人工智能的优化支撑作用,包括两方面:(1) 传统数据管理分析的理论技术对人工智能的数据和计算过程的优化;(2) 传统数据管理系统设计理念对开发通用且易用型人工智能平台的促进作用.因此,需要利用和发展现有数据库理论,构建形成新的技术和系统经验.专刊重点立足于数据库核心技术,探讨数据管理与分析技术对人工智能研究发展推动作用,特别是数据管理分析的理论技术对人工智能在数据和计算密集环节的优化,以及数据管理系统设计理念与开发经验对构建通用型人工智能平台的促进作用,重点关注数据管理与分析技术对人工智能在数据存储、算法优化、模型管理、模型服务、系统构建等方面的支撑作用.

本专刊公开征文,共收到投稿 36 篇.论文均通过了形式审查,内容涉及支撑人工智能的数据管理、分析、系统与应用.特约编辑先后邀请了 60 多位专家参与审稿工作,每篇投稿至少邀请 2 位专家进行评审.稿件经初审、复审、NDBC 2020 会议宣读和终审共 4 个阶段,历时 6 个月,最终有 17 篇论文入选本专刊.根据主题,这些论文可以分为 5 组.

(1) 支撑人工智能的数据管理技术

《支撑机器学习的数据管理技术综述》从数据管理的视角对机器学习训练过程进行解构和建模,从数据选择、数据存储、数据存取、自动优化和系统实现等方面,综述并提出支持机器学习数据管理的若干关键技术挑战.

《数据库内 AI 模型优化》提出一种“预筛选+验证”对 AI 模型推理进行优化的框架,分析探讨了决策树等多个机器学习模型的优化技术,并通过扩展 SQL 支持了决策树训练与推理,所提出的方法能够对“借助决策树模型推理结果对数据进行筛选”的应用场景起到较好的加速效果.

《图嵌入算法的分布式优化与实现》提出一种通用的分布式图嵌入框架,将图嵌入算法中的采样流程和训练流程进行解耦,并设计了一种基于参数服务器的模型切分嵌入策略,从而大幅减少分布式计算中的通信开销.

《时序图节点嵌入策略的研究》提出了一种对时序图节点进行自适应嵌入表达的方法 ATGEB.结合信息在时序图中的传播特征,提出一种自适应方式对其活跃时刻进行聚类,并设计了双向多叉树索引结构和节点采样策略,在时序图中节点间时序可达性检测以及节点分类等问题上取得很好的实验效果.

《面向企业数据孤岛的联邦排序学习》提出了一种面向企业数据孤岛的联邦排序学习框架,并设计了交叉分割的联邦学习策略、基于略图的隐私保护技术和联邦半监督学习方法,进而验证了所提方法的有效性.

《多区间速度约束下的时序数据清洗方法》提出了多区间速度约束下的时间序列数据修复方法,并采用动

收稿时间: 2020-12-10; jos 在线出版时间: 2021-01-21

态规划方法来求解最优修复路径,进而验证所提出方法的可行性和有效性,特别是其可提升人工智能结果质量.

(2) 支撑人工智能的数据分析技术

《基于 Motif 聚集系数与时序划分的高阶链接预测方法》提出了一种基于 Motif 聚集系数与时序划分的高阶链接预测模型,通过同时结合网络中高阶结构的聚集特征与网络结构演变信息,提升预测效果与性能.

《面向时空图建模的图小波卷积神经网络模型》提出了一种新的时空图建模图小波卷积神经网络模型,通过结合图小波卷积层和扩展因果卷积层捕获时空图节点间属性特征的相关性,并设计了利用自适应邻接矩阵从数据中动态学习隐层空间依赖关系的有效方法.

《捕获局部语义结构和实例辨别的无监督哈希》提出了一种基于语义结构保持和实例分辨力的深度无监督哈希学习框架.其对语义结构进行学习的同时也指导哈希编码学习,并被验证可有效提升哈希编码的辨识力.

《用于表格事实检测的图神经网络模型》提出用于表格事实检测的图神经网络模型,利用表格的结构特征结合图注意力网络和图卷积神经网络,设计了以表格的行为单位的 Row-GVM 和以表格的单元格为单位的 Cell-GVM,进而证明所提方法的高效性.

(3) 支撑人工智能的数据库系统

《PandaDB:一种异构数据智能融合管理系统》提出了基于智能属性图模型的分布式数据融合管理系统 PandaDB,该系统实现了结构化/非结构化数据的高效存储管理,并提供了灵活的 AI 算子扩展机制,具备对多元异构数据内在信息的即席查询能力.

《KGDB:统一模型和语言的知识图谱数据库管理系统》研发了统一模型和语言的知识图谱数据库管理系统 KGDB,提出统一的存储方案,解决了无类型三元组的存储问题,并实现了两种不同知识图谱查询语言的互操作,进而验证该系统比 gStore 和 Neo4j 节省 30%的存储空间,查询速度最高可提高 2 个数量级.

《基于 Seq2Seq 模型的 SparQL 查询预测》研究如何利用已有的信息进行知识图谱的查询预测,从而进行数据的预加载与缓存,提高系统的响应效率,提出了将 SparQL 查询提取为序列形式的方法,使用 Seq2Seq 模型对其进行数据分析和预测,并使用真实的数据集对方法进行测试,实验表明所提出的方案具有良好的效果.

(4) 支撑人工智能的数据应用

《LFKT:学习与遗忘融合的深度知识追踪模型》针对学生遗忘行为对其知识掌握程度的影响,提出了融合学习与遗忘的深度知识追踪模型 LFKT.通过结合 4 个影响知识遗忘因素,采用深度神经网络可实时追踪由学生遗忘造成的知识水平变化过程.

《多尺度时序依赖的校园公共区域人流量预测》提出了一种基于深度学习的多尺度时序卷积网络 MSCNN 以对校园公共区域人流量进行预测.通过在真实校园环境测试,所提出模型的预测效果优于其他已有的校园区域人流量数据预测方法,特别在捕获多尺度时序模式方面更具优势.

(5) 赋能人工智能的数据库技术

《基于人工智能方法的数据库智能诊断》研究了 OLTP 数据库在实际运行时可能遇到的异常,分析了这些异常和一系列监控指标之间的影响关系,提出了一种智能的数据库异常诊断框架 AutoMonitor,包括数据库异常监测、异常指标提取和根因分析这 3 个模块,并部署在 PostgreSQL 数据库,实验结果表明该框架对于异常诊断具有较高的精确度,并且不会对系统性能造成太大的影响.

《GPU 数据库核心技术综述》综述了以 GPU 计算为核心的数据库系统(GDBMS)发展历程,深入剖析 GDBMS 的四大核心组件:查询编译器、查询处理器、查询优化器和存储管理器,并展望了其与人工智能、时空数据分析、数据可视化、商务智能等领域的交互应用.

本专刊主要面向数据库、数据挖掘、大数据、机器学习、信息检索等多领域的研究人员和工程人员,反映了我国学者在支撑人工智能的数据管理、分析、系统与应用领域最新的研究进展.感谢《软件学报》编委会和数据库专委会对专刊工作的指导和帮助,感谢专刊全体评审专家及时、耐心、细致的评审工作,感谢踊跃投稿的所有作者.希望本专刊能够对支撑人工智能的数据管理、分析与系统相关领域的研究工作有所促进.

陈雷(1972－),男,博士,香港科技大学讲席教授,博士生导师,IEEE Fellow,ACM 杰出科学家,国家自然科学基金海外杰出青年基金获得者.香港科学技术大学信息技术教育部-微软重点实验室主任,香港科技大学大数据研究所所长.主要研究领域为大数据、数据库、群体智能和机器学习等.《VLDB Journal》联合主编,《IEEE Transaction on Data and Knowledge Engineering》副主编,VLDB 理事会理事.获得“SIGMOD 十年时间检验论文奖”和“VLDB 2014 杰出演示系统奖”等.

王宏志(1978－),男,博士,哈尔滨工业大学教授,博士生导师,英才学院副院长,CCF 杰出会员.主要研究领域为大数据管理与分析、数据库系统、数据治理.主持国家自然科学基金重点项目、国家科技支撑计划课题等 10 余项项目,发表论文 200 余篇.CCF 哈尔滨分部主席,CCF 数据库专委会常务委员,ACM SIGMOD China 秘书长,CCF 大数据专委会委员,CCF 计算机应用专委会委员,ACM 数据科学学科标准编写组专家.获得黑龙江省自然科学一等奖、教育部高等学校科技进步一等奖、黑龙江省青年科技奖等.

童咏昕(1982－),男,博士,北京航空航天大学教授,博士生导师,CCF 高级会员.主要研究领域为大数据,数据库,联邦学习,时空大数据计算与群体智能等.主持国家自然科学基金优秀青年科学基金项目、国家重点研发计划课题等 10 余项,发表论文 80 余篇.曾获得阿里巴巴集团评选的首届“达摩院青橙奖”、“VLDB 2014 杰出演示系统奖”和“KDD Cup 2020 强化学习赛道冠军”等.

高宏(1966－),女,博士,哈尔滨工业大学教授,博士生导师,大数据科学与工程省重点实验室主任.主要研究领域为海量数据计算与分析、社交网络分析、时空序列数据管理与分析、数据质量、物联网感知数据收集、分布式感知数据计算等.承担国家自然科学基金重大项目课题、国家自然科学基金重点项目、科技部重点研发课题等 20 余项.发表学术论文 200 余篇.