软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2020,31(3):597−599 [doi: 10.13328/j.cnki.jos.005915]

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

人工智能赋能的数据管理、分析与系统专刊前言*

李战怀[1], 于 戈[2], 杨晓春[2]

[1](西北工业大学 计算机学院,陕西 西安 100191)

[2](东北大学 计算机科学与工程学院,辽宁 沈阳 110169)

通讯作者: 杨晓春, E-mail: yangxc@mail.neu.edu.cn

中文引用格式: 李战怀,于戈,杨晓春.人工智能赋能的数据管理、分析与系统专刊前言.软件学报,2020,31(3):597−598. http://www.jos.org.cn/1000-9825/5915.htm

大数据时代,数据规模庞大,数据管理应用场景复杂,传统数据库和数据管理技术面临很大的挑战.人工智能技术因其强大的学习、推理、规划能力,为数据库系统提供了新的发展机遇.专刊强调数据管理与人工智能的深度融合,研究人工智能赋能的数据库新技术和新型系统,包括两方面:(1) 传统数据管理、数据分析技术及系统与人工智能相结合,将会焕发新的生机;(2) 大数据管理与分析是新一代人工智能技术发展的基石. 因此, 围绕传统数据管理的不同技术层面,需要新的理论和系统经验.专刊重点围绕数据库核心技术,探讨数据管理与人工智能的深度融合,探讨在人工智能大潮下,传统数据管理技术、数据分析技术与数据库系统受到的影响、契机与应对策略,通过数据管理与人工智能融合,重点关注人工智能赋能新技术对传统数据采集、数据存储、索引、查询、统计分析以及数据管理系统的促进和提升.

本专刊公开征文,共收到投稿 37 篇.论文均通过了形式审查,内容涉及人工智能赋能的数据管理、分析与系统.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 2019 会议宣读和终审 4 个阶段,历时 6 个月,最终有 18 篇论文入选本专刊.根据主题,这些论文可以分为 5 组.

(1) 人工智能赋能的数据管理技术

《人工智能赋能的数据管理技术研究》综述人工智能赋能的数据管理新技术的研究进展,总结了现有方法的问题和解决思路,并对未来研究方向进行展望.

《基于中间层的可扩展学习索引技术》提出了基于中间层的可扩展学习型索引模型 Dabble,从而解决索引更新引发的模型重训练问题.

《面向关系数据库的智能索引调优方法》提出一种面向关系数据库系统的智能索引调优技术,利用机器学习方法构造索引的量化模型,可以准确地对索引的查询优化效果进行估计,并设计了一种高效的最优索引选择算法,实现了快速的从候选索引空间中选择满足给定大小约束的最优的索引组合.

《基于时空相关属性模型的公交到站时间预测算法》提出一种基于深度神经网络的公交到站时间预测算法,采用时空组件、属性组件和融合组件预测公交车辆从起点站到终点站的总时长.

(2) 数据处理与优化技术

《面向数据特征的内存跳表优化技术》给出条件社区搜索问题的形式化定义,使用布尔表达式表示搜索条件,提出解决条件社区搜索问题的通用框架及其优化方法,将条件社区搜索分解为多个单项条件社区搜索.

《面向区块链的高效物化视图维护和可信查询》提出一种面向区块链的高效物化视图机制,将视图维护操作与共识过程同时执行,降低该操作对系统性能的影响;使用字典树加快以区块为单位的多物化视图维护进程;以默克尔验证的方式确保物化结果不被恶意篡改,进而确保查询结果可信.

《时间约束的实体解析中记录对排序研究》提出基于二分图上相似性传播的记录匹配可能性计算方法,

收稿时间: 2020-01-09

将记录对、块及其关联关系构建二分图;相似性沿着二分图不断地在记录对结点与块结点之间传播,直到收敛.收敛结果可以通过不动点计算得到.提出了近似的收敛计算方法来降低计算代价,从而保证实体解析的实时召回率

(3) 人工智能赋能的数据分析与推荐

《面向多维稀疏数据仓库的欺诈销售行为挖掘》提出基于分割率的特征提取方法和基于张量重构的挂单行为挖掘算法;设计了基于挂单模式偏序格的特征提取方法,对销售数据集中存在的挂单行为进行分类.

《基于相关性分析的工业时序数据异常检测》提出一种基于序列相关性分析的多维时间序列异常检测方法.对多维时间序列进行分段、标准化计算,得到相关性矩阵,提取量化的相关关系.然后,建立了时序相关图模型,通过在时序相关图上的相关性强度,划分时间序列团,进行时间序列团内、团间以及单维的异常检测.

《基于图神经网络的动态网络异常检测算法》提出了基于图神经网络的异常检测算法,将图结构、属性,以及动态变化的信息引入模型中,以学习进行异常检测的表示向量.

《融合选择提取与子类聚类的快速 Shapelet 发现算法》提出一种快速时间序列 Shapelet 发现算法, 通过对原始训练集采用时间序列聚类,可以得到原始时间序列中没有的 Shapelet,同时在选择性提取算法中加入投票机制,以解决产生 Shapelet 过多的问题.

《基于注意力机制的规范化矩阵分解推荐算法》提出一种基于注意力机制的规范化矩阵分解模型,依据用户信任网络和评分记录构建用户-项目异构网络,并构建用户间的相似关系;引入注意力机制分析用户对项目各个属性特征不同的关注度来获取更准确的用户偏好.

《融合显式反馈与隐式反馈的协同过滤推荐算法》提出一种融合显式反馈与隐式反馈的协同过滤推荐算法.利用加权低秩近似处理隐式反馈数据,训练出隐式用户/物品向量;引入基线评估,将隐式用户/物品向量作为补充,通过显隐式用户/物品向量结合,训练得出用户对物品的预测偏好程度.

(4) 人工智能赋能的数据库系统

《学习式数据库系统:挑战与机遇》提出一种细粒度的分类体系,从数据库架构出发,将现有工作进行了梳理,系统地介绍了学习式数据库各组件的研究动机、基本思路与关键技术,并对学习式数据库系统未来的研究方向进行了展望.

《轩辕:AI 原生数据库系统》提出了原生的支持人工智能的数据库系统,将各种人工智能技术集成到数据库中,以提供自监控、自配置、自优化、自诊断、自愈、自安全和自组装功能,并通过声明性语言让数据库提供人工智能功能,以降低人工智能使用门槛.

(5) 人工智能赋能的数据应用

《基于 PSP_HDP 主题模型的非结构化经济指标挖掘》根据人工构建非结构化经济指标的局限性,以及主题模型在非结构化经济指标挖掘中存在的问题,结合已有经济领域分类标准、词语之间的语义关系和词语对主题的代表性,定义了文档的领域隶属度、词语与主题的语义相关度和词语对主题的贡献度,提出相应的主题模型,提高了经济主题的区分度和辨识度,可以更有效地挖掘与经济有关的经济主题和经济要素词.

《机器学习中的隐私攻击与防御》分析了机器学习模型的训练集在数据采集、模型训练等各个环节中存在的隐私泄露风险为人工智能环境下的数据管理所提出的挑战,指出传统数据管理中的隐私保护方法无法满足机器学习中多个环节、多种场景下的隐私保护要求,总结并展望了机器学习技术中隐私攻击与防御的研究进展和趋势.

《数据集成方法发展与展望》综述数据集成领域从 2001 年开始到现在的相关工作的发展脉络,并展望了未来在数据集成领域的潜在研究方向.

本专刊主要面向数据库、数据挖掘、大数据、机器学习、推荐系统等多领域的研究人员和工程人员,反映了我国学者在人工智能赋能的数据管理、分析与系统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.希望本专刊能够对人工智能赋能的数据管理、分析与系统相关领域的研究工作有所促进.

李战怀(1961－),男,博士,西北工业大学教授,博士生导师,工业和信息化部大数据存储与管理重点实验室主任,CCF 数据库专委会主任委员.获省部级一等奖、二等奖各两项.主要研究领域为大数据管理技术,海量信息存储系统等.

于戈(1962－),男,博士,东北大学教授,博士生导师,CCF 会士,中国电子学会高级会员,美国 ACM 会员和 IEEE 高级会员.主要研究领域为数据库理论与技术,分布与并行式系统,云计算与大数据管理,区块链技术与应用.

杨晓春(1973－),女,博士,东北大学教授,博士生导师,CCF 高级会员.主持国家优秀青年科学基金、国家自然科学基金、973 子课题等 20 项,发表论文 100 余篇,获得国际会议最佳论文奖 3 项,全国会议最佳论文奖 4 项,授权和公示中国发明专利 23 项,其中美国专利 1 项.获省部级奖励 21 项.主要研究领域为大数据管理与知识工程,数据库理论与系统,智能系统推荐,数据质量管理,数据隐私保护.