《数据质量》

大作业要求

李振华 长聘副教授、博导

可信网络与系统研究所

2023年12月

lizhenhua1983@tsinghua.edu.cn

组队方式

- 组队不能超过3人
- 3人打分相同

文献研读：40分考察要点

- 阅读提供的2023年CCF A类会议最佳论文及PPT，报告所得与所思

- ✓ 页码齐全5分（底部标注）
- ✓ 时间控制10分（**5分钟**）
- ✓ 知识传达10分（通俗易懂）
- ✓ 重点突出10分
- ✓ 课程关联5分

增加环节：随机选择一名同学提问或评论（提升不报告的同学参与度）

跨学科文献汇报：40分考察要点

□ 阅读提供的2023年Science/Nature期刊论文，制作PPT、报告所得与所思

- ✓ 图文并茂10分（以图带字）
- ✓ 页码齐全5分（底部标注）
- ✓ 时间控制5分（**5分钟**）
- ✓ 知识传达10分（通俗易懂）
- ✓ 重点突出5分
- ✓ 课程关联5分

增加环节：随机选择一名同学提问或评论（提升不报告的同学参与度）

期末会议/期刊研读文献说明

清華大學
Tsinghua University

重要会议数据文献研读

1

重要会议数据文献

根据《中国计算机学会推荐国际学术会议和期刊目录-2022》

从各领域CCF-A类会议中选取Best Paper/ Distinguished Paper Award/ Best Student Paper 共27篇，以2023年为主。

领域	会议简称	文献标题	介绍
计算机体系结构/并行与分布计算/存储系统	FAST	Perseus: A Fail-Slow Detection Framework for Cloud Storage Systems	论文介绍了一种用于存储设备的实用故障缓慢检测框架Perseus，它利用轻量级的基于回归的模型，能够快速定位和分析驱动器级别的故障缓慢故障，从而解决了软硬件中新兴的“故障缓慢”问题。
计算机体系结构/并行与分布计算/存储系统	DAC	Gamora: Graph Learning based Symbolic Reasoning for Large-Scale Boolean Networks	论文提出了一种新颖的符号推理框架，名为Gamora，用于从门级网表等比特爆炸的布尔网络（BNs）中推导高层抽象，这在功能验证、逻辑最小化、数据通路合成、恶意逻辑识别等方面具有显著的益处。
计算机体系结构/并行与分布计算/存储系统	MICRO	Phantom: Exploiting Decoder-detectable Mispredictions	论文发现了一类新的攻击类型PHANTOM，其不同于传统的推测执行攻击，推测执行攻击利用微体系结构层面违反冯·诺依曼顺序处理原则的普遍性，通过在流水线中晚期检测这些违规实现攻击。
计算机网络	SIGCOMM	Memory Management in ActiveRMT: Towards Runtime-programmable Switches	论文提出了一种针对基于RMT的交换硬件的内存管理方法，该方法通过使用活动网络中的胶囊技术来编程RMT设备，使得交换机内存的非破坏性重新分配可以在比P4编译更快的时间尺度上进行，而无需操作员干预。

重要会议数据文献

领域	会议简称	文献标题	介绍
计算机网络	INFOCOM	More than Enough is Too Much: Adaptive Defenses against Gradient Leakage in Production Federated Learning	论文发现，随着对梯度泄露隐私的担忧增加，出现了多种攻击机制，这挑战了联邦学习中隐私保护的主要优势。然而，作者对这些梯度攻击对生产联邦学习系统的实际影响提出了质疑。为此，作者提出了一种新的轻量级防御机制，可以在整个联邦学习过程中提供足够且自适应的保护。
计算机网络	NSDI	DOTE: Rethinking (Predictive) WAN Traffic Engineering	论文探讨了一种新的广域网（WANs）流量工程设计点：通过仅利用有关流量需求的历史数据直接优化WAN上的流量流向。
网络与信息安全	CCS	Victory by KO: Attacking OpenPGP Using Key Overwriting	论文介绍了针对OpenPGP规范及其实现的一系列攻击，这些攻击导致用户私钥被完全破解。
网络与信息安全	S&P	MEGA: Malleable Encryption Goes Awry	论文提出了对MEGA的五种不同攻击，使得用户文件完全泄露。此外，用户数据的完整性受到损害，以至于攻击者可以插入通过客户端所有真实性检查的、其选择的恶意文件。
网络与信息安全	USENIX Security	Don’ t be Dense: Efficient Keyword PIR for Sparse Databases	论文提出了一种可以在稀疏数据库上进行查询的单服务器关键字私有信息检索 (PIR) 结构SparsePIR，其核心基于一种新型编码算法，将稀疏数据库条目编码为线性组合，同时优化兼容包括递归在内的重要PIR。
软件工程/系统软件/程序设计语言	SOSP	TreeSLS: A Tree-structured Microkernel with Efficient Whole-system Persistence on NVM	论文开创性地研究了一种基于树结构的微内核操作系统，旨在提高非易失性内存（NVM）的整体系统持久性和效率。

重要会议数据文献

领域	会议简称	文献标题	介绍
软件工程/系统软件/程序设计语言	OSDI	Triangulating Python Performance Issues with SCALENE	论文提出了Scalene Python分析器，可以同时精确地分析CPU、内存和 GPU 使用情况，而且开销很低。Scalene的CPU和内存分析器通过区分低效Python和高效的本机执行时间和内存使用情况，帮助Python程序员完成他们的优化工作。
数据库 / 数据挖掘 / 内容检索	SIGMOD	Detecting Logic Bugs of Join Optimizations in DBMS	论文介绍了一种名为TQS的新型测试框架，用于检测涉及多表连接的查询导致的逻辑错误。
数据库 / 数据挖掘 / 内容检索	SIGKDD	All in One: Multi-task Prompting for Graph Neural Networks	文章将图形提示和语言提示的格式与提示标记、标记结构和插入模式统一起来，从而将NLP的提示思想无缝地引入图领域。
数据库 / 数据挖掘 / 内容检索	SIGIR	The Information Retrieval Experiment Platform	论文旨在将irdatasets、ir_measures和PyTerrier与信息检索实验平台（TIREx）集成，以推动更加标准化、可重现、可扩展甚至可盲化的检索实验。
数据库 / 数据挖掘 / 内容检索	VLDB	Auto-Tables: Synthesizing Multi-Step Transformations to Relationalize Tables without Using Examples	论文开发了Auto-Tables系统，能自动合成多步骤（使用Python或其他语言）的转换流程，将非关系型表格转化为标准的关系形式，而无需用户手动编程。
计算机科学理论	STOC	The Randomized k-Server Conjecture is False!	论文推翻了长期以来在线算法领域的一个中心开放问题——随机k-服务器猜想。

重要会议数据文献

领域	会议简称	文献标题	介绍
计算机科学理论	SODA	Dynamic Matching with Better-than-2 Approximation in Polylogarithmic Update Time	论文介绍了用于估计图匹配大小的动态算法，这些算法在图的边插入和删除时具有多对数更新时间，并且逼近比优于2。
计算机科学理论	FOCS	Maximum Flow and Minimum-Cost Flow in Almost-Linear Time	论文提出了一种几乎线性时间的算法，用于解决最大流和最小成本流问题。
计算机图形学与多媒体	ACM MM	CATR: Combinatorial-Dependence Audio-Queried Transformer for Audio-Visual Video Segmentation	论文介绍了一种用于音频视觉视频分割的转换器技术，结合了音频查询和视觉处理，提出了一种新的方法来处理音视频数据的结合和分割问题。
计算机图形学与多媒体	SIGGRAPH	Split-Lohmann Multifocal Displays	这项工作描述了一种近眼 3D 显示器，它可以瞬间创建一个虚拟世界，完全支持人眼关注不同距离内容的固有能力。这种能力使观看者能够以以前无法达到的沉浸水平体验 3D 视频和互动游戏。
人工智能	AAAI	DropMessage: Unifying Random Dropping for Graph Neural Networks	论文提出了一个全新的随机Dropping方法，该方法在消息传递的过程中直接对被传递的消息进行dropping操作，统一了图神经网络的随机dropping架构。。
人工智能	CVPR	Visual Programming: Compositional visual reasoning without training	论文提出了VisProg，一种解决复杂组合视觉任务的神经符号方法，不需要任何特定任务的训练。

重要会议数据文献

领域	会议简称	文献标题	介绍
人工智能	CVPR	Planning-oriented Autonomous Driving	论文介绍了统一自动驾驶(UniAD)算法框架，它使用了四个基于Transformer解码器的感知和预测模块，以及最终的规划器。这个框架通过联合优化前置节点，以达到驾驶场景中的最终目标。
人工智能	ICCV	Passive Ultra-Wideband Single-Photon Imaging	论文探讨了如何在极端时间尺度范围内（从秒到皮秒）对动态场景进行成像，同时要求在光线非常稀少的情况下进行被动成像，不依赖于来自光源的任何定时信号。
人工智能	ICCV	Adding Conditional Control to Text-to-Image Diffusion Models	论文提出了ControlNet模型，该模型通过向预训练的扩散模型添加额外的输入，能够控制生成图像的细节。输入可以是多种类型，例如草图、边缘图像、语义分割图像等，允许对生成的图像进行更精细的控制。
人机交互与普适计算	CHI	Changes in Research Ethics, Openness, and Transparency in Empirical Studies between CHI 2017 and CHI 2022	论文评估了人机交互（HCI）社区如何在研究伦理、开放性和透明度方面取得进步，并提出了改进这些实践的方法。
人机交互与普适计算	UbiComp	Selecting the Motion Ground Truth for Loose-fitting Wearables: Benchmarking Optical MoCap Methods	论文提出了一种新的基准测试方法，DrapeMoCapBench (DMCB)，专门设计用于评估光学标记基和无标记MoCap方法在松散服装上的表现。

清華大學
Tsinghua University

重要期刊数据文献研读

2

重要期刊数据文献

从2023年的Science/Nature期刊中选取封面文章或计算机相关的文章，共23篇。

领域	期刊名称	文献标题	介绍
物理学	Science	Ultrafast Mode-Locked Laser in Nanophotonic Lithium Niobate	锁模激光器是精密测量、光谱学等领域的一种重要技术。不过，这些激光器通常很笨重，限制了它们的应用。该论文通过使用物理化学的一系列方法，将锁模激光器缩小到了一个光子芯片的大小。
气象学	Science	Curbing Global Solid Waste Emissions toward Net-Zero Warming Futures	通过全球范围内的分析，作者们发现固体废物对于温室气体排放的重要影响，指出这是导致人为全球变暖的一个重要因素。如果对固体废物采取合适的干预措施，巴黎协定的2050碳中和目标是有望实现的。
水科学	Science	Satellites Reveal Widespread Decline in Global Lake Water Storage	利用卫星观测、气候模型和水文模型，论文展示了过去三十年中，由于人类和气候因素的影响，大型自然湖泊和水库的水量都有超过50%的减少。
遗传学	Science	Dual Domestications and Origin of Traits in Grapevine Evolution	通过分析来自世界各地的约3500种栽培和野生葡萄品种的遗传数据，论文揭示了人类和环境对葡萄驯化的影响，包括气候对历史人口规模的影响，葡萄酒和餐桌葡萄的驯化，以及与葡萄颜色和口感相关的变异。

重要期刊数据文献

领域	期刊名称	文献标题	介绍
地质学	Science	Global Glacier Change in the 21st Century: Every Increase in Temperature Matters	通过计算在全球气温升高1.5°到4°C的情况下，冰川会受到怎样的影响，文章表明，冰川的质量损失和对海平面上升的贡献会比目前人们的估计更大。
地质学	Science	The Magmatic Web beneath Hawaii	作者们利用超过20万个地震事件，绘制了夏威夷岛上著名的几个火山在40公里深度的岩浆供给的关系图。他们发现虽然这些火山在地表相互独立，但在地下却有着不小的联系。
地质学	Science	Dissolution Enables Dolomite Crystal Growth near Ambient Conditions	白云石很难在实验室条件下人工制备。作者们发现了一种在实验室条件下制备白云石的方法，通过循环溶液的过饱和和不饱和状态，可以使白云石的生长速度提高一千万倍，这与自然界中白云石形成的沿海和蒸发环境相一致。
生物学	Science	From Nature to Industry: Harnessing Enzymes for Biocatalysis	本文总结了酶作为自然界的催化剂在各种领域的应用和创新，并展示计算机应用于该领域的潜力和前景。
生物学	Science	Behavioral Responses of Terrestrial Mammals to COVID-19 Lockdowns	利用GPS跟踪数据，作者们记录了2300只哺乳动物在新冠疫情期间的运动模式的变化。他们发现，在实施隔离政策的地区，动物的行走距离更长；在人口密集的地区，哺乳动物的移动频率更低，而且比疫情前更靠近道路。

重要期刊数据文献

领域	期刊名称	文献标题	介绍
计算机体系结构/并行与分布计算/存储系统	Science	Neural Inference at the Frontier of Energy, Space, and Time	该论文设计了一种神经元启发的处理器芯片，在存储访问方面采取了创新的设计，可以实现高性能，高能效和高面积效率，为处理和传输海量数据提供了一种节能的方案。
计算机体系结构/并行与分布计算/存储系统	Nature	All-Analog Photoelectronic Chip for High-Speed Vision Tasks	本文介绍了一种结合电子和光学计算的全模拟芯片，无需模拟—数字转换器，实现了高速，高能效，低延迟，高准确率和高鲁棒性的视觉数据处理。
计算机科学理论	Nature	Evidence for the Utility of Quantum Computing before Fault Tolerance	该论文介绍了一种在噪声环境下实现超越经典计算能力的量子计算的方法，证明了量子计算在强纠缠情况下的优势和应用前景。
人工智能	Nature	A Foundation Model for Generalizable Disease Detection from Retinal Images	本文展示了一种利用无标签视网膜图像进行自监督学习和疾病诊断的人工智能模型，可以在少量标签数据的情况下适应多种疾病检测任务，减轻专家标注负担，实现广泛的临床应用。
人工智能	Nature	Accurate Medium-Range Global Weather Forecasting with 3D Neural Networks	本文介绍了一种基于神经网络进行中期全球天气预报的方法。它利用三维深度网络处理天气数据中的复杂模式，同时也适用于极端天气预报和集合预报。
人工智能	Science	Accurate Proteome-Wide Missense Variant Effect Prediction with AlphaMissense	作者们利用大量的生物序列数据和AlphaFold2蛋白质结构预测工具，开发了一个深度学习模型AlphaMissense，可以预测蛋白质中单个氨基酸变化的影响。

重要期刊数据文献

领域	期刊名称	文献标题	介绍
人工智能	Science	Deploying Synthetic Coevolution and Machine Learning to Engineer Protein-protein Interactions	本文通过深度学习的方法，研究了共进化如何影响蛋白质界面的多样性和特异性，并为蛋白质结构和相互作用的预测提供了新的思路。
人工智能	Nature	Illuminating Protein Space with a Programmable Generative Model	本文提出了一种用于蛋白质设计的生成模型，它可以直接从可能具有功能的蛋白质空间中采样出新颖的蛋白质结构和序列，并可以根据外部约束进行贝叶斯推断，实现对蛋白质的形状，结构，语义的控制。
人工智能	Nature	Transfer Learning Enables Predictions in Network Biology	作者提出了一种迁移学习的模型，通过在一般性的单细胞数据上进行预训练，然后根据下游任务的需求进行微调，为心肌病等疾病的建模和治疗靶点的发现提供了帮助。
人工智能	Nature	Champion-level Drone Racing Using Deep Reinforcement Learning	文章介绍了Swift，一个能够与人类世界冠军进行竞速的自主无人机系统。Swift通过将深度强化学习（RL）模拟与现实世界中收集的数据结合起来，达到了与专业飞行员相媲美的水平。
社会学	Nature	Dopaminergic error signals retune to social feedback during courtship	文章研究了雄性斑胸草雀（Taeniopygia guttata）在面临不同需求（如解渴、唱歌和求偶）时多巴胺信号的变化。
遗传学	Nature	Mexican Biobank Advances Population and Medical Genomics of Diverse Ancestries	研究团队对来自墨西哥所有32个州的6,057人进行了基因分型，特别注意包括该国原住民社区的代表。研究者利用这些数据进行了22种复杂性状的全基因组关联研究，并评估了多基因得分预测疾病风险的有效性。