

科普时报

行业周刊

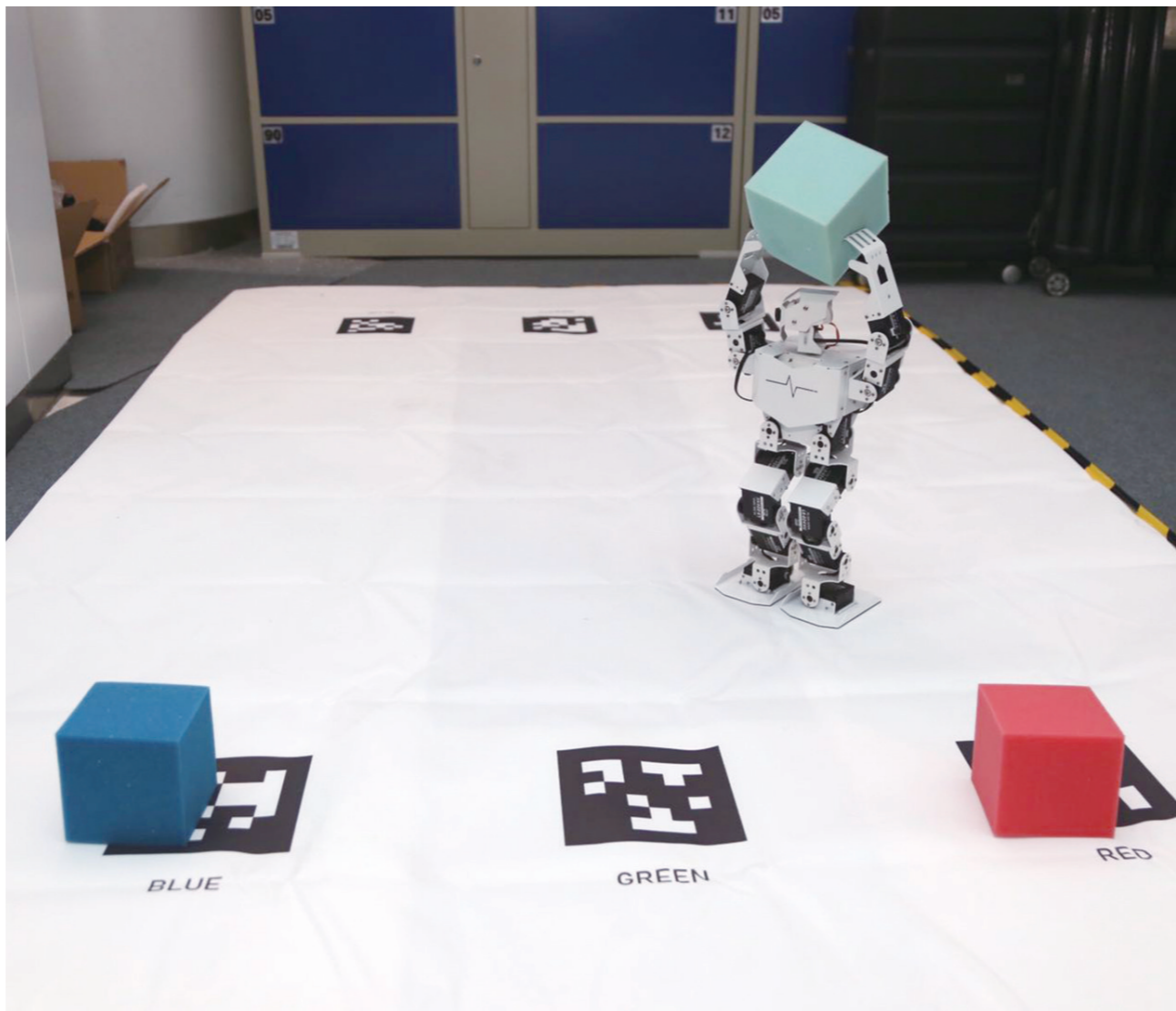
2020年10月20日
星期二
总第271期
主管主办单位：
科技日报社
国内统一刊号：
CN11-0303
社长 尹宏群
总编辑

科技创新、科学普及是实现创新发展的两翼，要把科学普及放在与科技创新同等重要的位置。没有全民科学素质普遍提高，就难以建立起宏大的高素质创新大军，难以实现科技成果快速转化。

——习近平

本期导读

- 携号转网：只是来的有点迟 (2版)
- 智能电视开机广告“扰人”依旧 (2版)
- 森林旅游“百花齐放” (3版)
- 京城赏红叶不必“扎堆儿” (3版)



10月17日-18日，“2020国际自主智能机器人大赛”在北京科学中心举行，竞赛的核心在于“看谁能驯化出更聪明的机器人”，该项比赛是世界大学生在人工智能高级领域的大比拼。图为机器人比赛现场。



北京市科协供图

大数据存储研究“大有可为”

□ 郑邦民

数据存储的意义，用一句话概括就是“记录当下，相约未来”。

数据存储完成了数据在时间维度的传播和跨越，将今天人的智慧、故事、喜怒哀乐、悲欢离合忠实地记录下来，供未来的人体验、学习、借鉴和欣赏。

在计算机产生之前，人们记载这些故事和智慧的手段是纸和笔，能够承载和传递的信息量十分有限，在现代电子计算机诞生这半个多世纪以来，存储介质发生了翻天覆地的变化，从早期的打孔纸带、磁带到磁盘、固态硬盘，再到近些年新涌现的非易失存储等，负载特征和访问模式也早已今非昔比。唯一不变的是计算机存储系统研究在整个计算机领域中的基础地位和重要意义。

从学科建设的角度来看，几乎没有一个应用和数据访问完全没有关系的，所以对于任何一个应用来说，在给定访问模式的前提下，如果希望能够在现有的硬件上充分发挥数据的读写性能，就离不开高效的存储系统。

从时代特点来看，我们已经进入一个大数据时代，全球的数据量积累速度加快，数据来源和种类多样、可信度低、价值密度低，所以在这样一个时代，研究设计新的存储系统大有可为。

从国家需求和战略发展来看，数据是影响国计民生的战略资源，我国在“十三五”期间集中建设数字中国、网络强国，经过近五年的开拓创新和锐意进取，我国数字产业、网络产业蓬勃发展，数字经济、共享经济、人工智能、大数据、云计算等已经飞入寻常百姓家。

例如在这次抗击新冠肺炎疫情的过程中，基于大数据的密切接触者追踪、筛查技术就为迅速控制住疫情作出重大贡献。但从另一方面来说，这也意味着我国经济社会对于数据的依存度在不断提升，尽管我们在计算机领域已经取得长足发展，但是在一些关键技术上依旧和发达国家差距较大，其中就包括大数据存储系统，不断上升的数据依存度意味

着这些技术短板带来潜在国家信息安全风险和发生系统性重大风险挑战的可能性也在不断上升。站在风口浪尖之上，推动我国存储技术的持续发展，成为这一代计算机人责无旁贷的使命与担当。

结合大数据存储的新特点和新挑战，我认为研究工作应主要围绕存储系统扩展性、可靠性和集约性三个方面展开。

存储的可扩展性的主要问题在于，存储系统的建设往往无法一步建设到位，特别是在大数据时代，数据的增长速度往往远远超过预期，因此必须要实现存储容量随着数据量的增加能够逐步动态扩展。

这个问题主要的挑战有两个，一个是设备和网络选择方面，往往扩展时必须选择原来类型的磁盘，新型号的磁盘不行，不同的厂家更不行；二是在线扩展的效率较低，在给用户增加了空的磁盘之后，需要在新旧磁盘之间重新分布数据，从而提高整个系统的读写性能，但是在扩展过程非常慢。

针对第一个问题，我们团队设计了一种可扩展的网络存储系统结构，实现了系统和存储设备的解耦，可以把异构的存储设备、IO网络集成为一个系统中。针对第二个问题，我们团队通过观察发现一种可重排序窗口特性，基于这一特性采用数据批量并行迁移、元数据懒情更新、数据迁移量最小化等方法，显著提高了扩展的执行速度。

存储的可靠性的主要问题在于，在大数据存储系统中，存储设备类多量大，设备出错成为一种常态，可靠性不足造成的损失越来越大。IBM对400家公司的调研表明，由于数据丢失和宕机给这些企业带来的损失高达7000亿美元。

存储的可靠性保证方法有两种类型，一种是容灾，基本思路是做跨数据中心的异地备份，主要用于避免各种因为灾害、战争、恐怖袭击导致的系统总体损毁；另一种是容错，基本思路是通过增加设备冗余、数据恢复等手段，主要用于避免因存储部件失效造成的部分数据丢失。

针对容灾框架一致性保证困难，数据依赖关系复杂的挑战，我们团队提出了一种结构无关的容灾框架，实现了进程状态和用户数据一起备份的全系统保护思想，设计了并行流水的恢复机制。针对容错问题，我们提出了一种轻量级编码方法，设计了一整套防止数据丢失、无需管理员维护的自维护系统，能够用较少的存储空间使得数据不会丢失，系统在生命周期内免人工维护。

存储的集约性的主要问题在于，面对数据产生的速度远快于磁盘生产速度造成的数据产生和存储容量之间的缺口，如何能够去除系统中的重复文件和冗余数据块，尽量降低存储开销。例如通过调研，我们发现兰州大学、清华大学校园网络存储的数据中，相同的文件存储副本数平均达20个，单个文件存储备份最高达1万个，因此利用不同用户之间存在很多内容相同的文件这一特点，我们可以通过用户共享来去掉这些重复的文件，从而节省存储空间。

在此基础之上，我们团队在云存储系统中引入实体化社区，采用面向共享的存储资源自适应组织和数据共享管控等关键技术，实现了数十倍的数据共享率，大幅节省了存储空间和主干网的流量费用。此外，针对现有主存储中存在大量相似数据块的问题，我们提出通过位置敏感哈希识别数据段的主存储内嵌删冗方法，删冗率达到30-70%，基于该方法，我们设计实现了高效删冗的存储系统。

我们的研究工作就是研究并设计可以高效扩展、高可靠、既能够抵御一般性错误也能够抵御各类灾难、同时还高度集约、能够高效存储大量数据的大数据存储系统。近年来，我们的研究成果用于影响国计民生和社会发展的近数千家单位，惠及数千万用户，取得显著经济和社会效益。未来，我们还将继续植大数据存储系统的科学研究，争取做出更多有价值、有分量的成果，经得起时代、历史和人民的检验。

(作者系中国工程院院士)

科普时报 (记者 侯静) 10月17

日-18日，由北京市科学技术协会主办、北京电子学会、清华大学微电子学研究所等单位共同承办的“2020国际自主智能机器人大赛”在北京科学中心举行，竞赛的核心在于“看谁能驯化出更聪明的机器人”。该项比赛是世界大学生在人工智能高级领域的大比拼，共吸引了来自德国汉堡大学、俄罗斯莫斯科国立大学、巴基斯坦国立科技大学以及北京大学、清华大学、武汉大学等国内外知名高校近80支队伍的参与。

经过激烈角逐，最终获得2020国际自主智能机器人大赛一等奖的队伍是：西安电子科技大学Friday队、清华大学深圳研究院Sirius队、哈尔滨工业大学(深圳)HITXGG队。

在本次大赛开幕式上，国际人工智能领域著名专家通过线上线下的方式参与，德国汉堡科学院院士、汉堡大学多模态智能机器人系统研究所所长张建伟教授、韩国科学技术院教授、IEEE life fellow 柳惠俊通过线上方式出席开幕式，并围绕人工智能及机器人领域的科技发展等内容作相关主旨演讲。

为了让全球更多学生参与到大赛中，本次大赛新增了线上虚拟赛，从而克服了因疫情原因导致的交通阻隔、时差、赛道实景等现实困难。多支国际大学生团队以“远程参赛”的方式参与该项赛事。

自主智能机器人是机器人的高级形态，能“自己思考”，因而更加“拟人”。它可以在特定的环境下根据实际条件自主决策。这需要更加强大的智能算法、较强的感知交互能力、姿态控制能力和运动能力，对参赛者的编程能力和机器人本身操控水平都有较高要求。

大赛执行主席、北京电子学会副理事长、清华大学微电子学研究所王志华教授表示，自主智能机械和人机融合是机器人发展的趋势。此次大赛激发了学生的创新意识和创新能力，鼓励了更多年轻人投入到人工智能、芯片设计研究中，可以为国家培养新生代科研力量，为世界智能制造产业的发展输送“顶梁柱”式人才。

大赛主办方北京市科协表示，举办国际自主智能机器人大赛旨在从科技创新、技术交流、人才培养等方面推动科技领域的国际学术交流合作，吸引更多的科研力量，加强原始创新成果的积累，为人工智能、类脑计算和芯片设计制造等前沿领域创新发展注入新动能。本届大赛因疫情防控形势，优化创新赛制，线上线下同向发力，营造良好的学术交流和科技创新氛围，通过打造自主智能机器人领域国际化、专业化、社会化交流合作平台，积极助力具有全球影响力的科技创新中心建设。

谁能称霸机器人脑王

2020国际自主智能机器人大赛在京举行

国产安全电脑“智”造工厂亮相京城

科普时报 (记者 侯静) 10月16日，中国长城(北京)网络安全产业生态示范基地网信产品下线仪式在北京大兴区举行，一座充满数字化的自主安全电脑



及相关产品“智”造工厂亮相京城大地。

中国长城(北京)网络安全产业生态示范基地仅用两个月时间就实现了产线建成投产，该基地以国产整机制造为基础，建设具备产品研发、生产制造、生态适配、解决方案、行业创新应用等功能的网信事业保障体系，面向北京及周边地区，提供基于PKS架构的国产台式机、一体机、笔记本、服务器等产品及服务，一期建成后，基地年产能达30万台(套)。

中国长城董事长宋黎定在致辞中说：通过聚焦新一代网信核心技术攻关突破，中国长城形成了“芯—端—云”完整生态链，孕育出长城“飞腾+麒麟+安全”网信产业生态，形成了体系化的核心优势，完全能支撑起首都网信产业的长远发展，并努力使其打造成为国内网信产业发展的样板高地。

北京大兴区作为“首都新国门”，区位优势明显，产

业发展条件得天独厚，此次中国长城携手大兴区紧抓“新基建”的机遇，深入开展紧密合作、互惠发展。大兴区委副书记、区长王有国认为，此次中国长城(北京)基地网信事业配套保障体系项目自主安全电脑系列产品的正式下线，标志着中国长城在打造网信产业生态、推动网信产业集群发展上取得重大突破，并将对大兴区加快培育壮大新动能、助力高精尖科技产业发展具有重要标志性意义。

据悉，随着北京长城网信项目的全部建成，基于中国架构的中国长城网信产业将具备全面支撑和保障北京市党政机关、在京部委、中央企业网信实践的核心能力，有力支撑大兴区首都高新技术产业基地建设，将成为辐射华北等省市区网信产业自主安全国产化的中坚力量，助推我国京津冀网信产业集群的快速崛起。



微信公众号

科普全媒体平台
敬请关注
欢迎扫码

责编：侯静



头条号