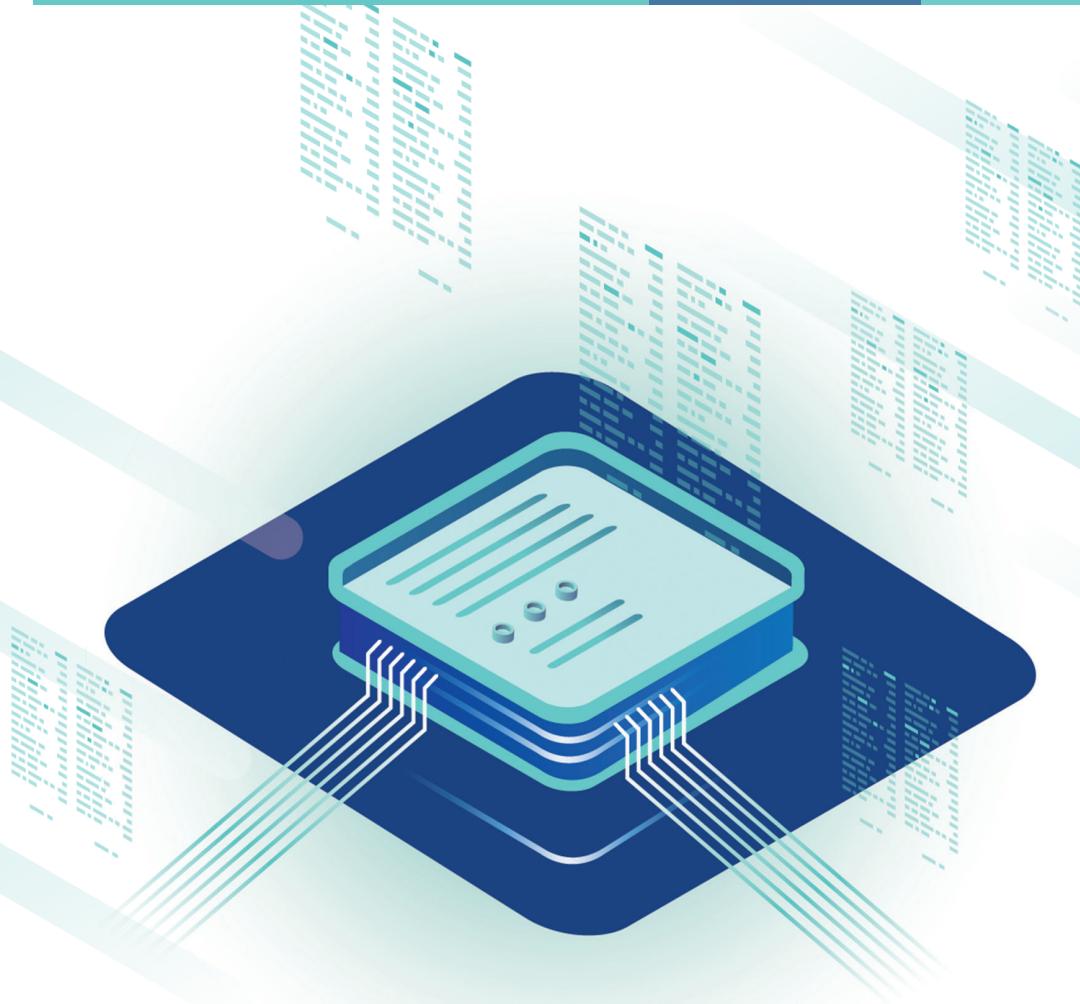


# 中国计算机学会通讯



COMMUNICATIONS OF THE CCF

第15卷 第1期 总第155期 2019年1月



新型存储和内存计算的回顾与展望 P8

从2018年的戈登·贝尔奖说起 P38

科研评价：破“五唯”，立什么？ P74



# ADL 学科前沿讲习班

*The CCF Advanced Disciplines Lectures*

## 站在学科前沿 打开技术之门

**高水平** (资深专家讲授)

**立前沿** (最新和热点技术)

**大剂量** (三整天)

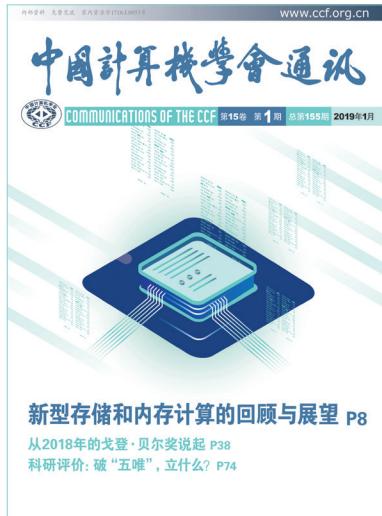
ADL是CCF举办的学科前沿讲习班。目的是使青年学者短期内深入了解计算领域某个学科前沿发展动态，加强学术交流，开拓眼界，提高学术水平，促进职业发展。2009年开始举办，每年10期。

主办单位：中国计算机学会

联系：adl@ccf.org.cn 188 1066 9757



# 中国计算机学会通讯 COMMUNICATIONS OF THE CCF



主办 中国计算机学会  
China Computer Federation  
  
刊名题字 张效祥  
  
编辑 《中国计算机学会通讯》编辑部  
编辑部主任：李梅  
地址：北京市海淀区科学院南路6号  
通信：北京2704信箱 100190  
电话：(010) 6267 0365  
传真：(010) 6252 7485  
http://www.ccf.org.cn  
E-mail: cccf@ccf.org.cn  
封面设计：SEEKLAB

## 声明

《中国计算机学会通讯》(CCCF)刊登的文章，除CCF或CCCF特别署名外，仅代表作者的学术观点。CCCF鼓励与支持学术争鸣。

## 版权声明

中国计算机学会(CCF)拥有《中国计算机学会通讯》所刊登内容的所有版权，未经CCF允许，转载本刊文字及照片会被视为侵权，CCF将追究其法律责任。

编辑单位：中国计算机学会  
印刷单位：北京华联印刷有限公司  
发送对象：中国计算机学会会员  
印刷日期：2019年1月

## 主编

李国杰 CCF名誉理事长，CCF会士，中国工程院院士

## 执行主编

钱德沛 CCF会士，北京航空航天大学教授，中山大学计算机学院院长

## 专题

主编 袁晓如 CCF理事，北京大学研究员  
编委 陈熙霖 CCF会士、理事，中国科学院计算技术研究所研究员  
李向阳 CCF专业会员，中国科技大学教授  
廖小飞 CCF高级会员，华中科技大学教授  
王蕴红 CCF会士、理事，北京航空航天大学教授  
杨珉 CCF专业会员，复旦大学教授  
郑宇 CCF杰出会员，京东集团副总裁

## 专栏

主编 彭思龙 CCF理事，中国科学院自动化研究所研究员  
编委 包云岗 CCF理事，中国科学院计算技术研究所研究员  
郭得科 CCF杰出会员，国防科技大学教授  
徐恪 CCF理事，清华大学教授  
王涛 CCF理事，爱奇艺公司首席科学家  
王长虎 CCF高级会员，字节跳动人工智能实验室总监

## 动态

主编 唐杰 CCF杰出会员，清华大学教授  
编委 鲍捷 CCF专业会员，北京文因互联科技有限公司CEO  
黄萱菁 CCF高级会员，复旦大学教授  
蒋洪波 CCF杰出会员，湖南大学教授  
刘知远 CCF高级会员，清华大学副教授  
宋国杰 CCF高级会员，北京大学副教授  
俞扬 CCF专业会员，南京大学副教授

## 译文

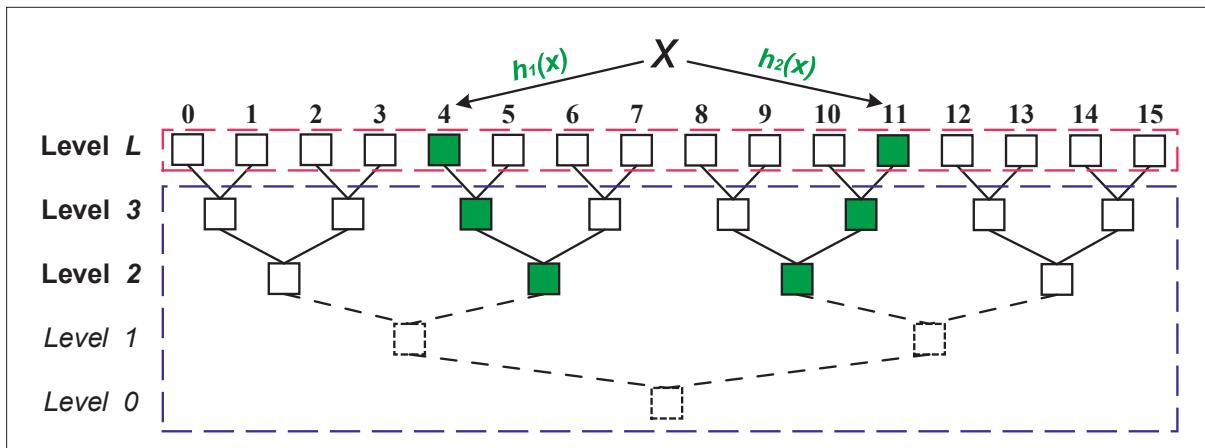
主编 卜佳俊 CCF常务理事，浙江大学教授  
编委 胡春明 CCF理事，北京航空航天大学副教授  
姜波 CCF理事，浙江工商大学教授  
苗启广 CCF理事，西安电子科技大学教授

## 学会论坛

主编 杜子德 CCF秘书长  
编委 胡事民 CCF会士、常务理事，清华大学教授

# CONTENTS 目录

2019年1月 第15卷 第1期 总第155期



## 新型存储和内存计算的回顾与展望

存储和计算是大数据处理的两大核心技术，它们在上层应用和下层硬件之间起着承上启下的作用。针对新型硬件和新应用场景的需求，设计更加高效的存储以及内存计算技术的研究正在蓬勃发展。本期专题以新型存储和内存计算为主题，邀请国内专家撰稿，重点探讨新型存储和内存计算的设计理念以及优化措施，介绍近年来该领域相关的发展趋势。

(P8~37)

## 卷首语

- 7 致读者  
李国杰



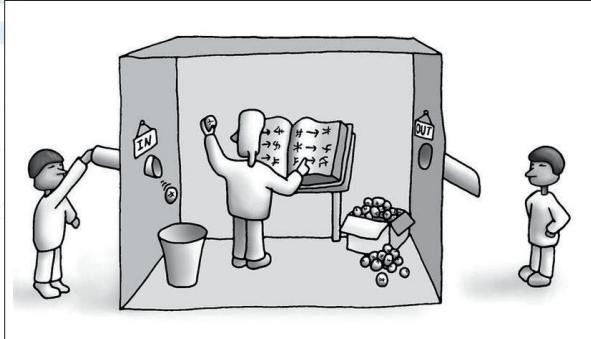
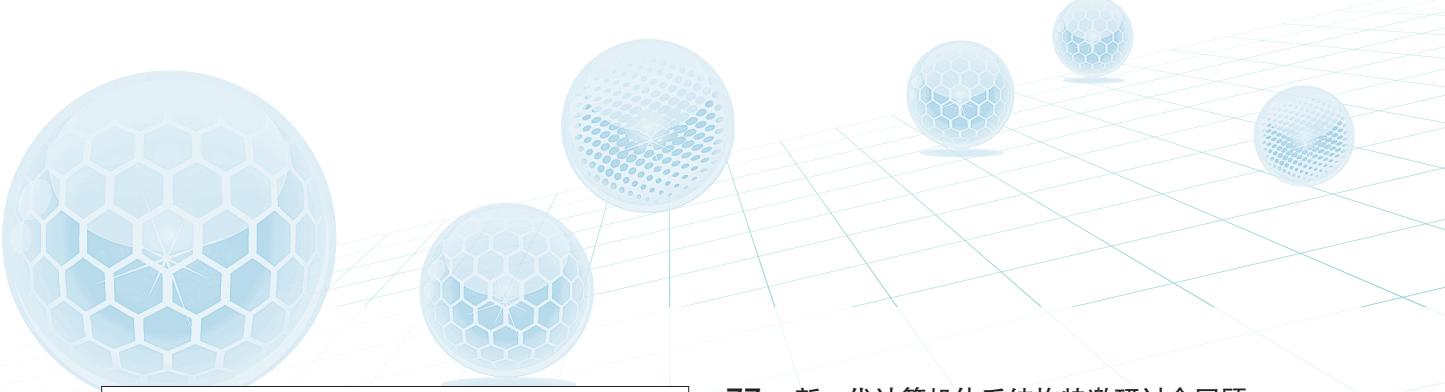
阅读整本

## 敬告读者

欢迎读者提出意见或建议。  
编辑部联系方式：  
电话：(010)6267 0365  
E-mail: cccf@ccf.org.cn  
查阅电子版：  
<http://dl.ccf.org.cn/cccf/list>

## 专题

- 8 新型存储和内存计算的回顾与展望  
特邀编辑：武永卫
- 11 面向非易失内存的高效能数据组织模式  
左鹏飞 华 宇
- 15 持久性内存：从系统软件的角度  
陆游游 舒继武
- 21 面向持久性内存的存储系统设计与优化  
蒋德钧 王 盈 熊 劲
- 27 分布式共享内存与内存计算  
杨 帆 洪 扬 陈 榕 等



中文房间（详见李航专栏文章）

- 32 面向HTAP的内存数据库并发控制技术  
张融荣 蔡 鹏 钱卫宁

## 专栏

- 38 从2018年的戈登·贝尔奖说起  
郑纬民 薛 巍 陈文光 等
- 43 大数据共享及交易中的机遇和挑战  
李向阳 张 兰 韩 风 等
- 52 智能与计算  
特邀专栏作家：李 航
- 57 电脑前传(2): 计算  
黄铁军
- 63 大数据交易市场构建  
郑臻哲 吴 帆 陈贵海
- 69 以“作品文化”取代“帽子文化”  
胡包钢
- 71 **The CS David专栏**  
《AI·未来》  
作者：戴维·阿兰·格里尔(David Alan Grier)  
特邀译者：孙晓明

## 动态

- 74 **CCF YOCSEF视点**  
科研评价：破“五唯”，立什么？

- 77 新一代计算机体系结构特邀研讨会回顾  
钱德沛 陈文光 范东睿 毛 睿
- 80 CIKM 2018最佳论文是怎样炼成的  
张俊祺 刘奕群 张 敏 马少平
- 84 新技术 & 新应用

## 译文

- 86 社交媒体中的多方隐私  
作者：约瑟·萨奇(Jose M. Such)  
娜塔莉亚·卡利亚多(Natalia Criado)  
译者：胡欣宇 岳亚伟

## 学会论坛

- 93 专委发展的历史性进步  
杜子德

## 读编往来

## 信息索引

• CCF ADL	封二
• CCF ADL第98期	6
• CCF推广《计算机专业培养方案编制指南》	10
• 《CCCF优秀文章精选》出版	26
• 第三届计算机视觉及应用创新论坛举办	51
• 第一届中国模式识别与计算机视觉大会召开	70
• CCF推广双导师计划	73
• CCF走进高校 (2018年)	76
• CCF会员活动中心动态 (2018年)	79
• 中日韩在韩国平昌举行CJK会议	83
• 唐杰在KSC2018作特邀报告	83
• 2018年度CCCF“积极评刊奖”评选揭晓	96
• CCF会员续费	封三
• CCF颁奖大会	封底

## Preface

### 7 To Readers

*Li Guojie*

## Features

### 8 Review and Prospect of New Storage and In-Memory Computing

*Guest Editor: Wu Yongwei*

### 11 High-efficient Data Structures for Non-volatile Memory

*Zuo Pengfei and Hua Yu*

Non-volatile memory (NVM) technologies demonstrate salient features of high density and scalability, while suffering from the limited write endurance and asymmetric properties of reads and writes. To deliver high performance, the authors analyze the main challenges of designing efficient index structures for NVM, and present two efficient hashing index structures, i.e., path hashing and level hashing. Path hashing is a cost-efficient write-friendly hashing scheme without extra writes to NVM. Level hashing is a write-optimized and high-performance index scheme with low-overhead consistency guarantee and cost-efficient resizing.

### 15 Persistent Memory: From the System Software Perspective

*Lu Youyou and Shu Jiwu*

Persistent memory is a competitive alternative to DRAM in in-memory storage and computing systems, due to its scalability, DRAM-like performance and data persistency. However, in persistent memory systems, the software overhead against hardware and the volatility-persistency boundary have changed, which led to new challenges in system software. This drives the changes not only in the operating system, including crash consistency, file system, memory management and persistent heap, but also in the distributed storage system.

### 21 Storage System Design and Optimization towards Persistent Memory

*Jiang Dejun, Wang Ying and Xiong Jin*

In this article, the authors investigate recent techniques related to building storage systems on persistent hybrid memory, including index optimization for NVM, low-overhead crash consistency, data path simplification, as well as metadata path optimization. Based on the investigation, the authors discuss the interested research topics for persistent memory based storage system, such as hybrid index design, virtual file system optimization, and hybrid storage medium (including DRAM, NVM, and SSD) management.

### 27 In-Memory Computing and Distributed Shared Memory

*Yang Fan, Hong Yang, Chen Rong and Qi Zhengwei*

In-memory computing promises 1000X faster data access speed, bringing opportunities to boost data processing into a higher level. Distributed Shared Memory (DSM) systems, providing shared memory abstractions for clusters, bring significant benefit to applications in terms of parallelized distributed computation and the ease of programming. Further, the GiantVM design incorporates DSM into the hypervisor and presents an abstraction of virtual shared-memory multiprocessor. In this article, the authors present two prototypes, and introduce their efforts in building DSM systems with emerging hardware for in-memory computing.

### 32 Concurrency Control Techniques in Main-Memory Database Systems for HTAP

*Zhang Rongrong, Cai Peng and Qian Weining*

In this article, the authors introduce different application scenarios and limitations to On-Line Transaction Processing (OLTP) and On-Line Analytical Processing (OLAP) systems. Two kinds of Hybrid Transaction and Analytical Processing (HTAP) applications proposed by Gartner are introduced and existing system architectures for HTAP are reviewed.

## Columns

### 38 Starting with the Gordon Bell Award in 2018

*Zheng Weimin, Xue Wei, Chen Wenguang and Zhang Youhui*

This article summarizes the development of the top level supercomputing systems in the 2018 TOP500 list, and introduces some winning and candidate applications of the 2018 ACM Gordon Bell Award. Finally, two development trends of supercomputing system are put forward.

### 43 The Opportunities and Challenges of Date Trading and Sharing

*Li Xiangyang, Zhang Lan, Han Feng and et al.*

This article introduces the current situation of data sharing and trading market, the relevant policies and regulations, as well as the typical data trading models. Then, along with the process of data transaction, the requirements to different participants, as well as the problems and challenges in each stage of transaction are analyzed. Finally, related works are briefly presented.

### 52 Intelligence and Computing

*Li Hang*

Is human thinking a kind of computing? Can a computer realize human thinking? These are fundamental issues for cognitive science and artificial intelligence. This article gives a brief overview and makes discussion on the topic of computing and thinking (or intelligence).

### 57 The Prequel of eBrain (2): Computing

*Huang Tiejun*

Mathematician Georg Cantor realized that the real number set is uncountable, and proved it with diagonal proof in 1891. Nine years later, David Hilbert posed 23 unsolved problems in 1928, the No.10 was generalized as the Entscheidungsproblem. In 1936, Alan Turing solved the problem by imaging a general computing machine, the conceptual model for all modern computers. The most important conclusion is: even each computable number is computable by finite means, and the computable numbers set is countable, but no algorithm or machine could compute them out one by one.

### 63 A Market Architecture for Data Transaction

*Zheng Zhenzhe, Wu Fan and Chen Guihai*

This article presents the first architecture for data marketplaces. The authors discuss the new features

of data commodity, and conduct an in-depth study of the design problem of data marketplaces. Data acquisition and data pricing are two major components to be build a practical data marketplace. For each component, the authors briefly review the related works, investigate the existing problems, present our preliminary results, and propose some future works.

### 69 Replace “Hat Culture” with “Works Culture”

*Hu Baogang*

This article puts forward the “works culture” after discussing the drawbacks of “hat culture”. “Works culture” pays attention to the works themselves or their connotations, which provides a stage for outstanding young people without eminent reputation. There is still a long way to go for China to move towards “works culture”, and we should make great efforts to cultivate the soil of innovative culture.

### 71 The CS David

#### AI Superpowers

*David Alan Grier (translated by Sun Xiaoming)*

This article is a reading notes of Lee Kaifu’s new book *AI Superpowers: China, Silicon Valley, and the New World Order*. Although its first half parts present an argument very similar to the one found in a thirty-five-year-old book *The Fifth Generation*, its latter parts make a surprising change of focus and talk about how to make the field of computer science more altruistic.

## Advances

### 77 Review of the Invited Seminar on New Generation Computer Architecture

*Qian Depei, Chen Wenguang, Fan Dongrui and Mao Rui*

### 80 A Journey to the CIKM2019 Best Paper

*Zhang Junqi, Liu Yiqun, Zhang Min and et al.*

### 84 New Technologies & New Applications

## Translations

### 86 Multiparty Privacy in Social Media

*Jose M. Such and Natalia Criado (translated by Hu Xinyu and Yue Yawei)*

Online privacy is not just about what you disclose about yourself, it is also about what others disclose about you.



# 深度学习

2019年1月26~27日 成都

学术主任：



张 蕾  
四川大学教授

特邀讲者：



焦李成  
西安电子科技大学教授  
讲座题目：  
影像大数据的深度学习解译  
与智能挖掘



王 亮  
中科院自动化所研究员  
讲座题目：  
人工智能时代的视觉大数据  
的理解



朱 军  
清华大学教授  
讲座题目：  
深度学习的对抗攻击与防御



马尽文  
北京大学教授  
讲座题目：  
基于深度学习的图像语义  
分割与遥感图像处理



于 剑  
北京交通大学教授  
讲座题目：  
深度学习的能和不能



何晓飞  
飞步科技创始人兼CEO  
讲座题目：  
人工智能与未来出行



# 卷首语



CCCF 2019年第1期

## 致读者

从 2015 年第 5 期开始，我每期写一篇主编评语，已经写了 44 篇。每期的主编评语都是千字短文，但写稿时敲下每一个字符我都感到沉重，往往要花一天时间才能写成。我感到紧张是担心由于我思想的局限性误导了 4 万多名学会会员和其他读者。人贵有自知之明，我已经 75 岁，早就不在第一线工作，写出的话很难讲到点子上，该到“谢幕”的时候了。

从本期开始，“主编评语”改成“卷首语”，我希望“卷首语”反映中国计算机学会的集体智慧，成为《中国计算机学会通讯》的“点睛”栏目。不仅计算机学会的领导层、CCCF 各栏目的主编、编委可以写，学会的常务理事、理事和会员都可以投稿。“卷首语”可以对学术研究方向发表独特的观点，对改善科研和产业环境提出鲜明的看法，也可以对取得的重大科技成果做介绍和评述，或对不正之风进行尖锐的批评。总之，只要是观点犀利、给人启迪，有助于科技和产业发展的精品短文，都可能成为“卷首语”。

发动有真心话想说的人写“卷首语”，可能是本刊的一次改革创新。我深信，广大科技人员中蕴藏着巨大的智慧和创造力。中国计算机学会藏龙卧虎、人才济济，给大家一个发表真知灼见的平台，发人深省的“卷首语”一定会源源不断地涌现出来。

### 主编的话

CCCF 2015 年第 5 期



### 人工智能到哪儿了？

人 工智能始终处于不断向前推进的计算机技术的前沿，互联网的普及和大范围应用将人工智能技术推向新的高峰。本期的专题栏目“人工智能到哪儿了？”风格，邀请了我国人工智能领域的几位专家，围绕“人工智能”这个主题，探讨人工智能的发展趋势、应用前景以及面临的挑战。

CCCF 2015 年第 6 期



### 未来互联网向何处去？

未 来互联网应如何构建？这是全世界信息领域科技人员普遍关心的一个重大问题。但我国的互联网主管部门似乎还不着急。今年已经是“十二五”计划的最后一年，遗憾的是，列入“十二五”国家重点基础设施建设计划的“未来网络试验设施”还没有启动。希望本期的专题讨论能对我们的未来网络研究起一点推动作用。

2011 年 7 月 IEEE Communications Magazine 发表的一篇关于未来网络的综述文章，对中国未来网络

主编的话  
Messages from Editor-in-Chief

主编评语  
CCCF 2018 年第 6 期



Messages from Editor-in-Chief

### 卧薪尝胆，发愤图强

在 刚刚召开的两院院士大会上，我聆听了习近平总书记、李克强总理和刘鹤副总理的报告，作为一个科技人员，倍感重任在肩。利用写主编评语的机会，给计算机界的同行们说几句感想和体会。

改革开放以来，特别是近十几年来，我国科技发展取得长足进步，发展速度超出了西方国家的预期，引起了一些西方人士的警惕和恐慌。我国有些媒体，特别是网络自媒体上，有许多盲目乐观甚至自欺欺人的言论，对外宣传上也不善于用国际上可接受的方式表达我们的观点，加深了西方民众对我们的误解。一个国家厉害不厉害是做出来的，不是说出来的。邓小平同志在退休之前曾提出 28 字方针：“冷静观察、稳住阵脚、沉着应付、韬光养晦、善于守拙、决不当头、有所作为”。现在的中国已经比 20 多年前强大，但仍然应多做少说，韬光养晦。

# 新型存储和内存计算的回顾与展望

特邀编辑: 武永卫  
清华大学

关键词 : 新型存储和内存计算

随着信息技术的发展,数据的规模呈爆炸式增长,对于存储效率和计算性能的要求与日俱增。由于大数据在国家发展中起到重要作用,并已渗透在智慧城市、国防安全、日常生活、行业发展等各方面,因而,大数据处理的两大核心技术,存储和计算就尤为重要,在上层应用和下层硬件之间起着承上启下的作用。一方面,它们需要迎合新型硬件新特性的要求,另一方面,又要兼顾上层应用的多样性、海量性及复杂性等变化。因此,为了应对这些挑战,面向新型存储和内存计算的研究层出不穷,目的是构建存储和计算性能更高的大数据处理系统。

在数据量飞速增长的同时,基础硬件的性能也在不断改善。更快速的网络设备、更高效的存储介质等相继问世。这些新型硬件的特性给传统的软件架构带来了不小的挑战。先前的操作系统、文件系统、计算系统等都是依据多层体系架构和缓慢设备进行设计的。如果直接将传统的软件应用于新型的设备,既受到接口的限制,无法充分使用其新特性,又无法充分发挥其性能优势。因此,新型的非易失内存(Non-Volatile Memory, NVM)和远程直接数据存取(Remote Direct Memory Access, RDMA)便成为研究热点。

将数据全部缓存在内存中进行计算,可以显著地避免由于缓慢外存带来的高延迟、低带宽的缺点。然而,虽然单台物理机器的计算物理核数量可以达到数百个(最大主频接近3GHz),但是传统内存的发展却停滞不前,遇到了瓶颈(Memory Wall)。由于

硬件上的限制,传统的动态随机存储器(DRAM)容量难以继续扩展,远远无法满足大数据对于内存的需求。从系统架构上考虑,使用新型的内存设备(纵向扩展)或者分布式共享内存(横向扩展)都可以在一定程度上打破现有单机内存容量的限制。

对于分布式共享内存,网络通信是影响整体性能的主要瓶颈。网络通信最早可以追溯到上个世纪末期,由于当时网络带宽的限制,其只能停留在小规模集群上。InfiniBand<sup>1</sup>网卡和专用网络给这项技术重新赋予了活力。相比于传统的网络传输技术,新型的RDMA技术具有更低延时、更高带宽的特性。同时,传输过程可以旁路远端操作系统,减少操作系统参与通信的开销。通过结合新型RDMA技术,分布式共享内存的性能获得极大提升,将会更适应以互联网应用为代表的新型大数据内存计算、内存存储、云计算应用等等。

在单机内存纵向扩展方面,传统体系结构利用外存作为内存的二级存储,以扩展内存处理超过内存大小的数据以及保证数据掉电不丢失等问题。传统的外存存储介质比如机械硬盘(HDD)、闪存(Flash)、固态硬盘(SSD)等,存储空间虽然很富裕,但是相比于内存具有更高的延迟和更低的传输带宽,并且只能按照块级别进行读写,从而导致存储的接口往往会成为系统的瓶颈。因此,对于内存的更高要求催生了一批类似RamDisk的工作。它们的目的是利用一组外存设备来达到近似于内存的执行速度。

<sup>1</sup> <http://www.infinibandta.org>。

而另一方面，内存 (DRAM) 读写性能很高，读写延时低至 50ns，可以通过字节寻址 (byte addressable)，但是内存的容量相对外存较低，且不能够保证在机器断电或者数据中心发生灾难时数据不会丢失。

NVM 的出现给内存的纵向扩展带来了新的思考。NVM 可以被认为是外存与内存设备的中间态，能够同时兼顾外存存储介质和内存的优点：存储量比内存大比外存小，具有非易失的特点，可以满足字节寻址的需求。这样的新特点促使研究人员开始尝试利用 NVM 来替换传统的内存。与此同时，NVM 相关的电子电路技术也在不断的演变，已经相继出现了以相变存储器 (PCM) 和阻变存储器 (ReRAM) 为代表的技术来实现 NVM。由于 NVM 的中间特性，NVM 被分别用在实现事务系统和存储系统之中。除了这两个特性外，NVM 还具有读写差异性、介于内存和外存存储介质之间的耐久度，因此需要对系统进行重新设计。对于原来以内存为基础的系统而言，在使用了 NVM 之后，需要考虑的是如何减少写操作，从而提高性能且增加器件的使用寿命；另一方面，还需要考虑如何通过合理的算法来保证数据的均衡写，以避免对于单个存储单元的过度写。如其不然，将导致该单元的损坏，最终使得整个存储设备不可用。

由于 NVM 带来的新的硬件特性势必会造成原有软件设计模式的改变。例如，现有系统的 Cache 数据刷回到 DRAM 是由系统控制的，程序无法自动进行缓存的持久化。同时，还会造成机器崩溃时数据丢失或者数据不一致的问题。Intel 的 NVML 提供了针对 NVM 的编程范式，该编程范式通过引入新的编程接口来指导开发者持久化数据。此外，还需要重新针对 NVM 进行部分数据结构的设计（哈希表、B 树、B+ 树等），用来保障数据的持久化、可恢复性，避免写频繁等等。针对哈希表结构的拓展和优化，须有效地减少写比例；针对树型结构，还须解决写频繁的问题。与此同时，传统文件系统在使用 NVM 之后，可以充分利用字节寻址的特性，减轻原有针对多层体系结构的软件执行负担，达到更高吞吐量的目标。

此外，上层应用处理需求的变化也同样推动着

计算和存储技术的发展。例如，同时满足 OLTP 和 OLAP 的新型混合数据库 HTAP，就是在顺应数据增长背景下，处理业务需求的产物。而在全内存模式下，传统的并发控制技术已经成为内存数据库系统的性能瓶颈。针对 HTAP 的应用场景，设计更加高效的并发控制技术，已成为提升处理性能的有效手段。

针对新型硬件和新应用场景的需求，设计更加高效的存储以及内存计算技术的研究正在蓬勃发展，本期专题以新型存储和内存计算为主题邀请国内专家撰稿，重点探讨新型存储和内存计算的设计理念以及优化措施，介绍近年来该领域相关的发展趋势。

华中科技大学的博士生左鹏飞，教授华宇撰写的《面向非易失内存的高效能数据组织模式》一文，介绍了其针对 NVM 的应用场景提出的路径哈希和层次哈希两种高效的哈希索引结构。这两种方法主要关注于降低哈希索引在 NVM 上的写操作频次，同时进一步提升了整体的访问性能。

清华大学计算机系助理教授陆游游，教授舒继武撰写的《持久性内存：从系统软件的角度》一文，从单机的操作系统出发，详细阐述了持久性内存对崩溃一致性、文件系统、内存管理、持久性堆的影响，并介绍了相关的研究工作进展，深度探讨了分布式存储系统在持久性内存上面临的问题和研究现状。

中国科学院计算技术研究所副研究员蒋德钧，研究员熊劲等撰写的《面向持久性内存的存储系统设计与优化》一文，从存储系统的多个维度探讨了持久性内存对其软件架构的影响。他们详细阐述了面向 NVM 的高效索引结构的性能优化方法，如哈希表、B+ 树、LSM-Tree，多种基于软件层面的容机一致性保障技术，以及针对元数据通路和数据通路优化手段。

上海交通大学硕士生杨帆，副教授陈榕，教授戚正伟等撰写的《分布式共享内存与内存计算》一文，详细地阐述了共享内存技术在大数据背景下面临的挑战，深入介绍了新型硬件对内存计算架构设计的影响，并详细介绍了其课题组针对分布式共享内存架构和虚拟化架构提出的两种高效的解决方案 MAGI 和 GiantVM。

华东师范大学教授钱卫宁等撰写的《面向 HTAP 的内存数据库并发控制技术》一文，专注于 OLTP 和 OLAP 混合架构 (HTAP) 下并发控制技术的设计思考，深入探讨了 HTAP 的起源与现有系统架构，详细地分析了传统并发控制技术在 HTAP 架构下面临的问题，介绍了针对 HTAP 架构的内存数据库并发控制技术的研究现状。

新型硬件和应用类型的出现，催动着计算和存储软件设计的变革，萌生出一系列新型的存储和内存计算架构。未来，随着硬件的不断发展、应用场

景的多样化，传统的存储与内存计算系统很难满足发展的需求，诸多软件架构会被重新审视，势必会产生更多理论和应用的发展新趋势。此外，针对现有新型硬件的研究还未成熟，深度优化现有理论会带来更多新机遇。■



武永卫

CCF 理事，CCSP 技术委员会主席。清华大学计算机系副主任、教授。主要研究方向为并行与分布式处理、云计算、存储系统。wuyw@tsinghua.edu.cn

## CCF 推广《计算机科学与技术专业培养方案编制指南》 在河南理工大学举行培养方案评议座谈会

2018 年 12 月 7 日，CCF 教育工委主任、中国人民大学教授杜小勇，CCF 教育工委副主任、北京航空航天大学教授高小鹏，CCF 教育工委委员、河北工业大学教授董永峰，作为 CCF 培养方案评议专家组走进河南理工大学，就“如何制定培养方案”与教师们展开热烈的会谈。河南理工大学计算机学院党委书记王全才、副院长孙君顶、专业负责人朱世松等二十余位教师参与了交流会谈。

王全才详细介绍了计算机专业培养方案的制定情况，CCF 培养方案评议专家组对此提出了从整体到局部的具体整改建议。高小鹏作了题为“复杂工程问题驱动教学设计浅谈”的报告，介绍了以解决复杂工程问题为培养目标的教学体系设计基本思路。董永峰作了题为“工程教育认证的理解与感悟”的报告，指出在教学体系设计中推广认证的理念比通过工程教育认证本身更为重要。孙君顶和朱世松都谈到：“如何响应国家号召，顺应时代变化，在师资力量薄弱的情况下培养一批基础知识扎实并且能适应社会需求的计算机专业人才，是学院面临的难题。CCF 培养方案评议活动如同一场‘及时雨’，是专业整改的重要契机。”

2018 年，CCF 教育工委在参考了工程教育专业认证标准、ACM/IEEE 的 CS2013 等多个标准的基础上，编写出版了《计算机科学与技术专业培养方案编制指南》(以下简称《指南》)。为更好地指导国内高校计算机专业编制培养方案，同时也为《指南》提供更多的案例和反馈，CCF 开展了“教育工委走进高校”系列活动，河南理工大学等 10 所院校是首批试点单位。



# 面向非易失内存的高效能数据组织模式

关键词：非易失内存 哈希索引 写优化 数据一致性

左鹏飞 华 宇  
华中科技大学

传统的动态随机存储器 (Dynamic Random Access Memory, DRAM) 几十年来一直作为计算机系统的主存。但是，由于 DRAM 的可扩展性有限、刷新能耗高，难以满足未来计算机系统的需求。因而，人们研制出新型的非易失存储器 (Non-Volatile Memory, NVM)，例如，相变存储器 (PCM) 和阻变存储器 (ReRAM) 等，因其具有可扩展性高和能耗开销低等优点，被提出作为下一代主存的存储介质来替代或补足 DRAM。NVM 具有的非易失性还使得数据可以持久地存储在主存中来实现瞬时的系统故障恢复。NVM 字节可寻址的特性，以及接近于 DRAM 的访问延迟，使得 NVM 可以直接接在内存总线上通过 CPU 的读写指令进行访问，避免了传统的基于块接口的开销。

## NVM中存在的问题

随着内存特性和架构的重大改变，以及 NVM 的硬件局限性和数据一致性保证的需求，使得传统的面向 DRAM 设计的索引结构在 NVM 中变得低效。现在大量的工作，包括 CDDS B-Tree<sup>[1]</sup>、NV-Tree<sup>[2]</sup>、wB+-Tree<sup>[3]</sup>、FPTree<sup>[4]</sup>、WORT<sup>[5]</sup> 和 FAST&FAIR<sup>[6]</sup>，

已经对基于树的索引结构做了重大改进，使得它们能有效地在 NVM 中使用。基于树的索引结构能够提供  $O(\log(N))$  的平均查询时间复杂度，这里  $N$  是索引结构中的元素数目。与基于树的索引结构不同，哈希索引结构是一个平坦的数据结构，能够实现常数级  $O(1)$  的查询时间复杂度。由于能够提供快速的查询响应，哈希索引结构被广泛使用在主存系统中。例如，哈希索引结构是主存数据库的重要组件，也被广泛用于对 Memcached<sup>1</sup> 和 Redis<sup>2</sup> 这样的键值存储系统的索引。但是，当哈希索引结构被用在 NVM 系统时，又存在许多重要的挑战。

首先，为了处理哈希冲突，现有的哈希索引方法都会造成许多额外的内存写。例如，布谷鸟哈希 (cuckoo hashing) 通过迭代踢出哈希表中的元素来处理哈希冲突，单个的插入操作甚至会造成几十上百的数据移动。这对具有有限写耐久性的 NVM 是非常不友好的，NVM 的写延迟较长，额外的内存写会大大降低哈希索引的性能。另外，哈希索引结构中的扩容操作需要把旧哈希表中的所有元素重新哈希到更大容量的新哈希表中，也会增加大量的内存写操作。

其次，当系统发生故障如掉电或系统崩溃时，

<sup>1</sup> Memcached. <https://memcached.org/>, 2018。

<sup>2</sup> Redis. <https://redis.io/>, 2018。

NVM 上的哈希索引结构需要避免数据的一致性，如数据丢失或部分更新等。NVM 会直接通过内存总线访问这种新的架构，给一致性保证带来巨大的开销。造成这些问题的主要原因是由于处理器和内存控制器经常会打乱内存写的顺序，而且处理器对 NVM 的原子写单元非常小（8字节）。我们需要使用刷新缓存行和内存栅栏指令来保证内存写的顺序，并使用日志或写时复制（copy-on-write）等机制来保证非原子写数据的一致性，这些操作都会带来很大的性能开销。

## 高效能数据组织模式

为了解决这些问题，使哈希索引结构能够有效地在 NVM 中使用，我们提出了一系列面向 NVM 的高效哈希索引结构。

- 路径哈希 (path hashing)<sup>[7]</sup>：一个面向 NVM 的写优化的哈希索引结构，对哈希索引结构中的插入和删除操作不会造成任何额外的 NVM 写，并且能提高访问性能。在路径哈希的基础上，我们进一步提出了一个 cache 优化的路径哈希索引结构 (Cache-Optimized Path Hashing, COPH)<sup>[8]</sup>，通过增强路径哈希中数据存储的局部性来进一步提高路径哈希的访问性能。

- 层次哈希 (level hashing)<sup>[9]</sup>：一个面向 NVM 的保证数据一致性的写优化哈希索引结构，实现了查询、插入、更新和删除操作在最坏情况下常数级时间的复杂度。它通过非常低的开销来保证各种写

操作的一致性，且在大部分情况下不需要使用日志。另外，层次哈希使用了一种原地扩容的技术，在整个扩容过程中，只有不到 1/3 的元素被重新哈希，这大大地提高了扩容的速度且减少了 NVM 写。

## 路径哈希

为了避免 NVM 写延迟长的问题，我们提出了一种面向 NVM 的写优化的哈希索引结构——路径哈希<sup>[7]</sup>，其对哈希索引结构中的插入和删除操作不会造成任何额外的 NVM 写，并且可以提高访问性能。路径哈希使用位置共享（一种不会造成任何额外的内存写的哈希冲突处理技术），同时使用双路径哈希和路径缩减技术来实现高的哈希表负载率和访问性能。

**位置共享**：路径哈希中的存储单元逻辑上组织成一个倒立的完全二叉树，该倒立二叉树的最上面一层，也就是所有的叶子节点，可以被哈希函数寻址，称为可寻址单元。所有的非叶子节点是不能被哈希函数寻址的，被用作叶子节点处理哈希冲突的共享备用位置（如图 1 所示）。在路径哈希中，某个叶子节点  $\ell$  到根节点的一条路径 (Path) 称作为 Path- $\ell$ 。当某个叶子节点  $\ell$  发生哈希冲突时，该叶子节点所对应的路径 Path- $\ell$  上的空的备用位置都可以被用来存储冲突元素。例如，在图 1 中，一个新的元素  $X$  被哈希到叶子节点 4，如果叶子节点 4 已经被占据，元素  $X$  可以存储在 Path-4 的 Level 3 和 Level 2 对应的空位置。

**双路径哈希**：由于在路径哈希中一条路径只

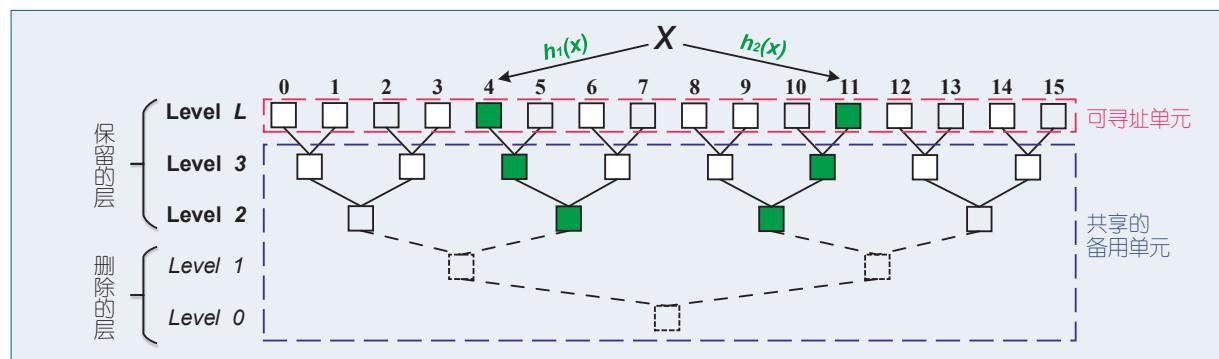


图1 路径哈希索引结构( $L=4$ )

有  $L+1$  个位置，最多只能处理在一个叶子节点上发生的  $L$  个哈希冲突。这样，在一个哈希表处于低负载率时就很容易导致插入失败。为了解决这个问题，路径哈希使用双路径哈希技术，对每个元素使用两个哈希函数，计算两个哈希位置。由于每个元素对应了两条路径，只要这两条路径中任意一条存在空的位置，该元素就可以成功插入。如图 1 所示，对于新的元素  $X$ ，使用两个哈希函数  $h_1(x)$  和  $h_2(x)$  计算得到两个哈希位置 4 和 11。元素  $X$  可以存储在叶子节点 4 和 11 中的任何一个空位置，如果这两个叶子节点都不为空，则路径哈希同时遍历 Path-4 和 Path-11，直到找到一个空位置来插入  $X$ 。使用两个独立哈希函数的随机性和位置共享来处理哈希冲突，使得路径哈希可以达到一个很高的负载率。

**路径缩减：**在路径哈希中，每个查询操作需要遍历两条长度为  $L+1$  的路径，直到找到目标元素。我们发现，在这个倒立二叉树的结构中，位于底部的几层只提供了很少的位置来处理哈希冲突，但是每一层都给读路径长度增加 1。例如，Level 0 层的根节点只包含一个位置来处理哈希冲突，但是增加了读路径长度。为了减少查询操作中读路径的长度，路径哈希进一步提出路径缩减技术，删除位于底部的多层，只保留位于顶部的几层。如图 1 所示，Level 0 和 Level 1 被删除了，只保留了 Level 2、Level 3 和 Level  $L$ 。在删除了底部的几层后，路径哈希依然能够达到很高的负载率。

路径哈希中的插入和删除操作只需要在两条路径中搜索一个空位置或目标元素，不会造成任何的数据移动，因而没有额外的内存写。位置共享和双

路径哈希有效地处理哈希冲突，使得路径哈希能够达到很高的负载率，路径缩减技术也提高了路径哈希的访问性能。

**Cache 优化：**在路径哈希的基础上，我们提出了一个 cache 优化的路径哈希索引结构(COPH)<sup>[8]</sup>。其主要思想是通过增强路径哈希中的数据存储的局部性，来进一步提高路径哈希的访问性能。具体地，COPH 把倒立的二叉树划分成多个小的子树，然后把每个子树的存储单元打包在一起存储在一个连续的空间中。那么，一次内存访问就可以预取一条路径上的多个节点到 cache 中，从而提高 cache 效率减少内存访问次数。例如，COPH 把相邻两层的三个节点打包在一起，一个内存访问总是可以把一条路径上的两个节点预取到 cache 中，如图 2 所示。

## 层次哈希

在 NVM 中，除了减少内存写，保证数据在系统崩溃时的一致性也是非常重要的。我们提出了一种面向 NVM 的保证数据一致性的写优化哈希索引结构——层次哈希<sup>[9]</sup>。层次哈希提供了一个基于共享的两层哈希表结构（层次哈希表），它可以保证查询、插入、更新和删除操作在最坏情况下的常数级时间复杂度，并且产生极少的额外 NVM 写。在这样一个两层哈希表结构中，层次哈希使用了原地扩容技术，只需要重新哈希旧哈希表中  $1/3$  的元素，来提高扩容性能。另外，层次哈希可以通过非常低的开销来保证写操作的一致性，其不需要日志来保证插入、删除和扩容操作的一致性，对于大部分更新操作的一致性保证也不需要日志。

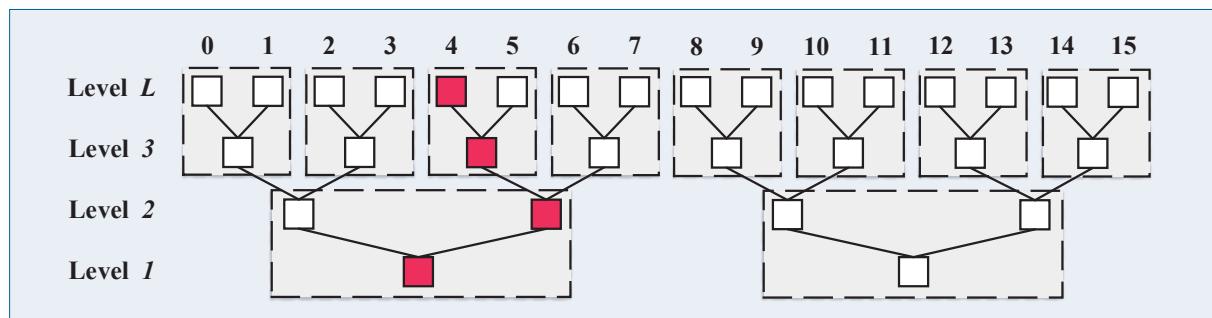


图 2 cache 优化的路径哈希 ( $L=4$ )

**写优化的两层哈希表结构：层次哈希表的结构具有四个特点。**(1) 多槽的桶结构：根据 Facebook 和百度的真实负载，键值存储系统中大部分都是小的键值元素。层次哈希表在每个哈希桶中设置多个槽，每个槽可以存储一个键值元素。那么一个哈希桶就可以存储多个元素。当访问层次哈希表中的一个桶时，该桶中多个元素可以同时被预取到 CPU cache 中，从而提高了 cache 的效率，减少了内存的访问数量。(2) 双哈希函数：如同路径哈希，层次哈希表也使用了两个哈希函数来更好地处理哈希冲突。一个新的元素会对应两个桶，然后被插入到已存储元素更少的桶中。(3) 基于共享的两层结构：层次哈希表中的所有桶被分成两层（顶层和底层），每个底层的桶被两个顶层的桶共享来处理顶层桶发生的哈希冲突。那么，底层桶的数目是顶层的一半。如果顶层的桶发生了哈希冲突并且所有的槽都被占据了，新的元素可以插入到底层中相应的备用桶中。(4) 每次插入至多允许一次移动：插入一个新的元素时，当对应的哈希桶都满了后，层次哈希表至多允许移动一个现有元素，这样可以重新均匀地分布元素来提高哈希表的负载率。具体地，当插入一个新的元素时，如果它在顶层对应的两个哈希桶都满了，层次哈希表将试图移动这两个桶中的一个元素到该元素对应的另一个桶中。如果移动失败，也就是被移动的元素对应的另一个桶也没有空位置，新的元素会被插入到底层。如果新元素在底层对应的两个桶都满了，层次哈希表试图移动底层这两个桶中的一个元素到该元素对应的另一个桶中。如果底层元素的移动依然失败，新的元素才被认为是插入失败。

这样一个哈希表结构具有多个优点。(1) 写优化：大部分元素在插入时不需要移动其他元素，极少的元素在插入时最多移动一个元素。(2) 高性能：查询、删除和更新操作最多只需要探测四个桶，实现了常数级的最坏情况时间复杂度；插入操作至多移动一个元素，时间复杂度也是常数级的。(3) 高负载率：哈希表结构设计上的四个特点使得元素能够均匀地分布在哈希表两层的桶中，从而实现了高的负载率。

## 总结和展望

为了在 NVM 上有效地索引数据，我们提出了路径哈希和层次哈希。在 NVM 上构建哈希索引依然面临着很多待解决的问题，例如在 NVM 上如何设计高并发的哈希索引、如何设计针对高维数据的哈希索引等。另外，一些 NVM 的硬件架构中引入了计算单元，形成了一种 Process-in-Memory(PIM) 的新型架构。哈希索引结构如何利用这种新型的 PIM 架构来实现高的性能也是个很重要的研究问题。■



左鹏飞

CCF 学生会员。华中科技大学博士研究生。主要研究方向为非易失内存架构和系统、数据去重和压缩技术、键值存储系统和存储安全。

pfzuo@hust.edu.cn



华 宇

CCF 杰出会员。华中科技大学教授，博士生导师。主要研究方向为文件系统、云存储系统、非易失性存储器、大数据分析等。

csyhua@hust.edu.cn

## 参考文献

- [1] Venkataraman S , Tolia N , Ranganathan P , et al. Consistent and Durable Data Structures for Non-Volatile Byte-Addressable Memory[C]//Proceeding of the USENIX Conference on File and Storage Technologies (FAST) .2011: 61-76.
- [2] Yang J, Wei Q, Chen C, et al. NV-Tree: reducing consistency cost for NVM-based single level systems[C]// Proceeding of the USENIX Conference on File and Storage Technologies (FAST) .2015: 167-181.
- [3] Chen S, Jin Q. Persistent B+-trees in non-volatile main memory[J]. Proceedings of the VLDB Endowment. 2015, 8(7):786-797.
- [4] Oukid I , Lasperas J , Nica A , et al. FPTree: A Hybrid SCM-DRAM Persistent and Concurrent B-Tree for Storage Class Memory[C]//Proceedings of the International Conference on Management of Data (SIGMOD) .2016: 371-386.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>

# 持久性内存：从系统软件的角度

陆游游 舒继武  
清华大学

关键词：非易失性内存 内存存储 系统软件 文件系统

数据的存储和处理技术推动了计算机技术的快速发展。近二十年来，相比于计算密集型应用，数据密集型应用已经成为主流。高性能计算、大数据应用、物联网以及人工智能等领域均产生了海量数据，同时也对数据进行快速或实时的分析提出了要求。海量的数据存储和高效的数据处理需求对计算机存储系统提出了极大的挑战，内存存储与计算得到了学术界和工业界的广泛关注，比如提出了以 Spark 为代表的大数据处理框架以及以 SAP HANA 为代表的内存数据库等。

非易失性内存 (non-volatile memory) 近年来发展迅速。除了块粒度访问的闪存存储 (flash memory) 已经广泛应用于存储系统之外，字节粒度访问的非易失性内存技术也层出不穷，英特尔与美光公司研制的 3D XPoint 技术即将商用。字节粒度访问的非易失性内存可直接连接于内存总线上，被称为持久性内存 (persistent memory)。持久性内存以接近于动态随机存取存储器 (DRAM) 内存的速度提供数据的持久存储，为数据的实时处理提供了机会。

然而，持久性内存改变了传统易失性内存与持久性外存的结构。这一结构的改变对传统计算机系统软件的设计提出了挑战。如何基于持久性内存构建系统软件是近年来学术界关注的话题。

## 持久性内存的机遇

内存存储与计算既需要数据的大容量存储，也需要数据的高速处理。然而，在当前的计算机存储

层次结构中，主存与外存之间的延迟差异远大于寄存器、CPU 缓存以及主存之间的延迟差异。例如，在一个计算机典型延迟中，寄存器的延迟约为 1 纳秒，CPU 缓存 (L2) 的延迟约为 10 纳秒，DRAM 主存的延迟约为 60 纳秒，然而外存磁盘的延迟约为 5 毫秒。相比于寄存器、CPU 缓存以及 DRAM 主存之间一个数量级的差异，DRAM 与磁盘之间的延迟差异高达 5 个数量级。由于这一差异，内存存储与计算的缓存不命中将极大地影响系统的性能。

然而，DRAM 和磁盘技术的发展均遇到瓶颈。

**DRAM 的容量限制：**DRAM 的容量限制不仅源于晶体管 - 电容式 (1T1C) 单元的密度受限，也源于难以扩展的动态刷新 (refresh) 操作。在 DRAM 单元中，电容需要足够大，以提供可靠读写；晶体管需要足够大以减少漏电，提高数据保存时间。在 DRAM 的芯片上，DRAM 单元需要在限定时间内完成刷新操作。当 DRAM 单元数量越来越多时，刷新操作所占比例越来越高。例如，当刷新一个 64Gb 的 DRAM 时，消耗了约 46% 的性能和 47% 的能耗<sup>[1]</sup>。因而，DRAM 在容量上难以扩展。

**磁盘的性能限制：**磁盘的 I/O 访问中存在盘片旋转和磁头定位两个机械式部件的操作。机械式部件的操作限制了磁盘性能的提升。近 20 年来，磁盘的带宽维持在略高于 100MB/s 的水平，延迟维持在数毫秒量级，几乎没有变化。正是机械部件的存在限制了磁盘的性能提升。

相比之下，持久性内存容量可扩展，且性能接近于 DRAM，是内存存储与计算中存储介质的有力

竞争者。

## 持久性内存的挑战

字节粒度访问的非易失性内存不但自身特点与 DRAM 和磁盘有很大差异，而且改变了传统计算机存储层次结构中易失性 - 持久性的边界，这两方面对构建持久性内存的系统软件均提出了新的挑战。

### 软硬件开销占比的变化

作为持久性存储，持久性内存与磁盘存在很大的差异，包括读写方式（内存总线与 PCI 总线读写）、读写粒度和读写性能等。与读写方式和读写粒度的差异对系统软件的操作方式带来的影响相比，读写性能的差异对持久性内存上的系统软件的效率要求带来的挑战更大。持久性内存的性能与 DRAM 内存接近，远高于磁盘和固态硬盘 (Solid State Drive, SSD) 的性能。加利福尼亚大学圣迭戈分校 (UCSD) 的报告<sup>[2]</sup> 显示，在非易失性存储系统中，软件开销显著高于磁盘存储系统。在磁盘存储系统中，软件开销占比仅为 0.3%。在持久性内存 (DDR NVM) 系统中，软件开销占比高达 94.3%，如图 1 所示。因而，在持久性内存中，软件的高效设计是一个新的挑战。

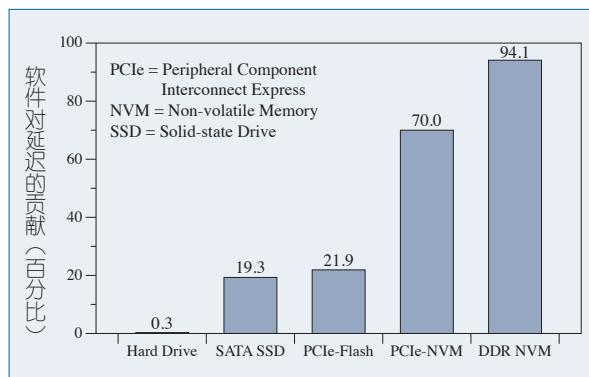


图 1 存储系统中软件开销占比变化<sup>[2]</sup>

持久性内存系统软件的高效设计问题引出了一系列有趣的研究问题。硬件延迟的降低，操作系统自身的延迟不可忽略。在操作系统中，持久

性内存 I/O 操作是否需要经过内核？系统调用和虚拟文件系统 (VFS) 层如何重新设计？文件系统的数据缓存如何管理？如何利用 DRAM 与 NVM 各自的相对优势？

### 易失性-持久性边界的变化

在传统存储层次结构中，寄存器、CPU 缓存和 DRAM 主存均为易失性存储器，而外存为持久性存储器。在持久性内存存储系统中，寄存器和 CPU 缓存较大可能仍使用易失性存储器，而主存和外存采用持久性存储器。易失性 - 持久性边界从主存 - 外存边界上移至 CPU 缓存 - 主存边界<sup>[3]</sup>，如图 2 所示。然而，在主存 - 外存的边界上，操作系统对于 DRAM 主存中的缓存数据管理是白盒管理，可以通过软件形式对数据组织与写回进行灵活控制。在 CPU 缓存 - 主存边界上，CPU 缓存的管理是硬件控制，对于系统软件是黑盒操作。系统软件难以灵活控制，而通过 clflush 等刷写命令对 CPU 缓存效率影响较大。易失性 - 持久性边界的变化对持久性存储系统中高效一致的数据写回也是新的挑战。

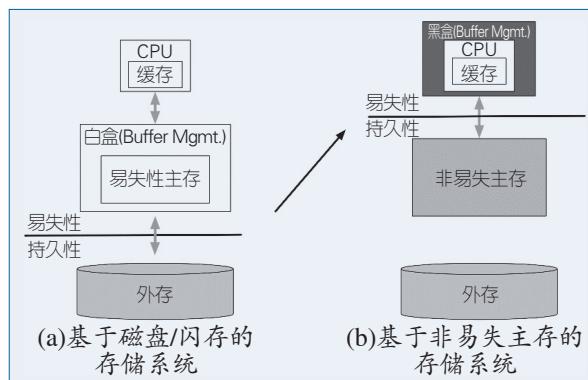


图 2 易失性-持久性边界的变化<sup>[3]</sup>

由于易失性 - 持久性边界的变化，持久性内存系统的崩溃一致性 (crash consistency) 需要重新设计。如何设计高效的机制是新的难点。同时，这一边界的变化，使得程序中的数据可以直接持久性地存储在内存中，为程序中的堆数据提供了持久性保障。“程序级别的数据持久保存与数据共享”这一新的编程模型，也成为持久性内存研究中的新内容。

## 持久性内存对操作系统的影响

针对持久性内存中系统软件的挑战，文件系统和内存管理需要新的设计以满足持久性内存对管理功能以及效率的需求。同时，持久性内存的出现促使了存储系统一致性管理的重新设计，新的编程模型也应运而生。

### 崩溃一致性

存储系统的一致性由并发控制和故障恢复两个部分保证。其中，故障恢复引入的一致性也被称为崩溃一致性。崩溃一致性要求系统遇到意外故障后，系统要么处于新的状态，要么处于旧的状态，不会出现中间状态。在传统的存储系统里面，崩溃一致性通过数据的版本有序写回的方式保证，比如文件系统日志(journaling)机制、数据库写前日志(write-ahead logging)、影子页(shadow paging)等。在持久性内存中，数据缓存在CPU缓存中。由于CPU缓存为硬件控制，因而控制数据版本以及写回顺序性会极大地影响持久性内存系统的性能。针对崩溃一致性的研究可从软/硬件两方面优化持久性与顺序性。

**以硬件方式优化顺序性：**由于CPU缓存为硬件控制，如果在CPU缓存硬件中加入对一致性的支持，将会减少软件的开销。微软研究院早在其设计持久性内存文件系统BPFS时就提出了Epoch的硬件语义<sup>[4]</sup>，即软件程序仅需发送epoch指令给CPU缓存，无须等待数据持久化返回；由CPU缓存硬件保证不同epoch之间的数据按序持久化。清华大学提出了LOC(Loose-Ordering Consistency)技术<sup>[5]</sup>，将预测执行的概念引入数据持久化，放松事务持久化的顺序性要求，从而降低了顺序性带来的开销。密西根大学提出了Strand Persistence技术<sup>[6]</sup>，允许在无依赖的程序片段中并发地执行顺序性约束，提高了并发性能。

**以软件方式优化顺序性：**传统的存储系统也多采用软件方式来弱化顺序性约束，以降低一致性开销，例如采用Checksum、反向指针、计数等方式。在持久性内存存储中，Mnemosyne采用了

TornBit的做法<sup>[7]</sup>，在每个数据块中保留一个比特位torn bit，在写回时被置位，并与数据块原子性写回。由于在初始化时torn bit比特位被清空，可通过torn bit的置位状态判断多个数据块的原子性，而无须在写入路径上维护顺序，从而降低了一致性开销。

**以硬件途径优化持久性：**微软的研究人员提出全系统持久化(Whole System Persistence, WSP)技术<sup>[8]</sup>，CPU缓存采用非易失性存储器的方法，将存储层次结构中除寄存器之外的存储均设计为非易失性存储。寄存器的存储也可通过计算机中的剩余电量保证数据持久写回，以时刻保持系统的运行状态。这种技术虽然能防止意外掉电产生的崩溃一致性问题，但不能处理由程序bug带来的崩溃一致性等问题。而且将CPU缓存完全设计为非易失性存储的做法对于器件要求苛刻，短期内难以实现。Kiln<sup>[9]</sup>仅在CPU末端缓存(last-level cache)上采用非易失性存储，通过末端非易失性缓存的使用降低了数据版本持久化的开销。

**以软件方式优化持久性：**清华大学提出的模糊持久化(BPPM)技术<sup>[3]</sup>，允许未提交数据被写回持久性内存，通过日志数据组织方式使之可检测与可恢复；对于已有数据版本的已提交数据，允许其延迟持久化。通过有条件地放松数据在易失性与持久性状态上的转换，降低了一致性的开销。

近年来，尽管持久性内存的一致性技术涌现，然而尚未出现公认的简洁有效的方法，这一方向仍期待进一步的研究。

### 文件系统

在文件系统中，页缓存(page cache)是持久性内存文件系统中的一个主要性能开销。在持久性内存文件系统中，持久性数据直接放置于内存级别的存储器中，因而页缓存在内存级别的缓存是冗余的。为了避免页缓存的冗余数据拷贝问题，直写(Direct Access, DAX)方式被引入到传统文件系统中以提高其在持久性内存上的使用效率，如Ext4-DAX。

字节访问粒度是持久性内存区别于持久性外存存储的新的访问特点。传统文件系统即使采用

DAX 技术仍然难以利用该特点。微软研究院设计并实现了面向字节粒度非易失性内存的持久性内存文件系统 BPFS<sup>[4]</sup>。BPFS 采用树状结构组织文件的数据与元数据，引入短路影子页 (short-circuit shadow paging) 方法提供高效的数据一致性。短路影子页利用 8 字节的原子更新特性以更新数据指针，从而避免在树结构的整个路径上进行数据拷贝，提高了效率。英特尔公司也设计并实现了持久性内存文件系统 PMFS<sup>[10]</sup>，同样也是为字节粒度访问设计的，从而避免了块粒度 I/O 中的数据拷贝与读改写 (read-modify-write) 开销。

然而，非易失性内存的读写延迟与 DRAM 内存的访问延迟有区别。非易失性内存的读写性能呈现不对称的特点，其读性能与 DRAM 接近，但写性能落后于 DRAM。清华大学的 HiNFS<sup>[11]</sup> 认为，直写方式并不是持久性内存文件系统最合理的方式，而传统缓存模式也存在优化空间。HiNFS 将直写 DAX 方式与传统的缓存方式相结合，对持久性内存文件系统中的读写操作以及不同持久化需求的写操作进行区分，提供了精细化的数据 I/O 控制，有效地提升了文件系统的性能。

NOVA<sup>[12]</sup> 持久性内存文件系统将日志 (log-structured) 结构引入到持久性内存文件系统。NOVA 采用了多日志的方式支持并发的元数据写操作，同时传统日志式结构中数据随机读的性能问题在持久性内存上也得到了有效缓解，从而达到较好的性能。

是否绕开操作系统内核以避免系统调用与虚拟文件系统层开销也是文件系统中一直讨论的问题。Aerie<sup>[13]</sup> 是一个用户态实现的内存文件系统。为了支持文件元数据的管理及数据共享，Aerie 采用一个单独的用户态进程作为可信任的第三方，实现元数据管理以及并发控制。用户态文件系统的缺点在于难以保证文件系统的读写安全，恶意进程可以随意读写文件数据。清华大学提出的 KucoFS，采用了用户态与内核态协同设计的方法，以内核态方式管理元数据，但将元数据操作通过卸载、授权 (lease) 等方式交由用户态辅助执行，既实现了内核态的数据

保护，也提供了用户态灵活高效的数据读写。

## 内存管理

在持久性内存系统中，内存的碎片问题和内存分配一致性问题也显得更为严重。在传统易失性主存中，内存碎片在系统重启后即消失；在持久性内存中，内存数据在系统重启后仍然保存，因而碎片问题日积月累，更为严重。内存分配和释放时的空间管理元数据应能够在系统故障后恢复正确状态，因而需要提供崩溃一致性。

针对内存分配和释放的空间管理元数据一致性问题，Makalu<sup>[14]</sup> 依据数据结构中对象可达性区分空间管理元数据中的关键元数据和辅助元数据，通过延迟辅助元数据的持久化，降低空间分配时的开销。针对内存碎片问题，LSNVMM<sup>[15]</sup> 采用了日志式结构对持久性内存空间进行管理，通过垃圾回收可以减少碎片数量，从而缓解碎片问题。

## 持久性堆

在程序中，堆数据是执行过程中产生的数据。传统程序在进行数据持久化或数据共享时，需要将数据以文件的形式存储于文件系统中。在持久性内存中，倘若需要持久化或共享的数据可以被持久化索引数据结构索引住，这些数据就可以直接以内存数据结构的形式保存在持久性内存中，而无须通过文件系统的读写操作。这一类工作也被称为新型编程模型。

NV-Heaps<sup>[16]</sup> 和 Mnemosyne<sup>[7]</sup> 是其中较早出现的两项研究工作。NV-Heaps 是一个持久性对象存储系统，数据对象直接存储于持久性内存上，同时映射到程序堆空间上。为保证掉电后数据可达，NV-Heaps 检查指针类型，不允许出现持久性内存指针指向易失性内存对象的情形，同时也提供了内存分配的一致性。Mnemosyne 具有类似的功能，将持久性内存空间导出用于持久性对象的分配，并保证其一致性。Heapo<sup>[17]</sup> 基于原生堆的结构提供持久性堆的管理。英特尔公司也提供了 PMDK<sup>[18]</sup> (原称为 NVML) 的用户态程序库，用于导出持久性内存空间，

以方便持久性内存的分配与管理。

尽管在持久性内存上有持久性堆与文件系统等不同系统构建形式的讨论，然而其中有很多相同的技术。持久性堆与传统的程序堆有较大的不同，为了实现数据的持久化存储与数据共享，持久性堆的数据对象需要采用持久化的索引结构，同样需要考虑持久性内存空间的分配与释放，以及内存空间上的数据组织与布局等问题。这些技术与文件系统、内存管理之间的技术是相通的。

## 持久性内存对分布式存储系统的影响

在分布式存储系统中，分布式的软件管理模块通常堆叠于本地存储系统之上，数据访问的路径冗长。因而，在持久性内存上，分布式存储系统的软件效率问题尤为严重。

清华大学将 GlusterFS 分布式文件系统分别部署于两种集群上：一种是以磁盘和千兆以太网构建的传统分布式集群，另一种是以内存和 56Gbps 的 Infiniband 网络构建的高速分布式内存集群。如图 3 所示，在 1KB 同步写的延迟测试中发现，尽管延迟从传统集群的 18 毫秒降低为内存集群的 324 微秒，但是软件开销占比从传统集群的 2% 急剧上升为内

存集群的 99.7%。在 1MB 写的带宽测试中发现，传统集群中软件能够达到硬件裸带宽的 94%，而内存集群中软件仅能达到硬件裸带宽的 6.9%。因而，现有的分布式存储系统并不能有效地发挥出持久性内存等硬件的性能优势。

Octopus<sup>[19]</sup> 是一个面向非易失性内存和 RDMA 重新设计的分布式持久内存文件系统。它通过 RDMA 将分布式持久性内存构建为分布式持久内存池，减少数据在不同软件层次上的内存数据拷贝。针对高速网络与存储 I/O 中的中心化服务节点的处理瓶颈问题，Octopus 提出了客户端主动的 I/O 数据流，以网络代价换取服务器处理效率，将访问负载均摊至客户端，从而提高整体效率。基于 RDMA 的原语特性，Octopus 还提出了新的远程过程调用 (Remote Procedure Call, RPC) 原语以及分布式一致性协议。Octopus 有效发挥了硬件的性能，达到硬件裸带宽的 88% 以上，能大幅提升大数据处理的性能。

HotPot<sup>[20]</sup> 也是新提出的分布式持久内存框架，并在内核级别提供了持久命名机制与空间管理机制，向上层软件提供了透明的内存访问接口。

## 未来研究展望

持久性内存已得到广泛关注，近

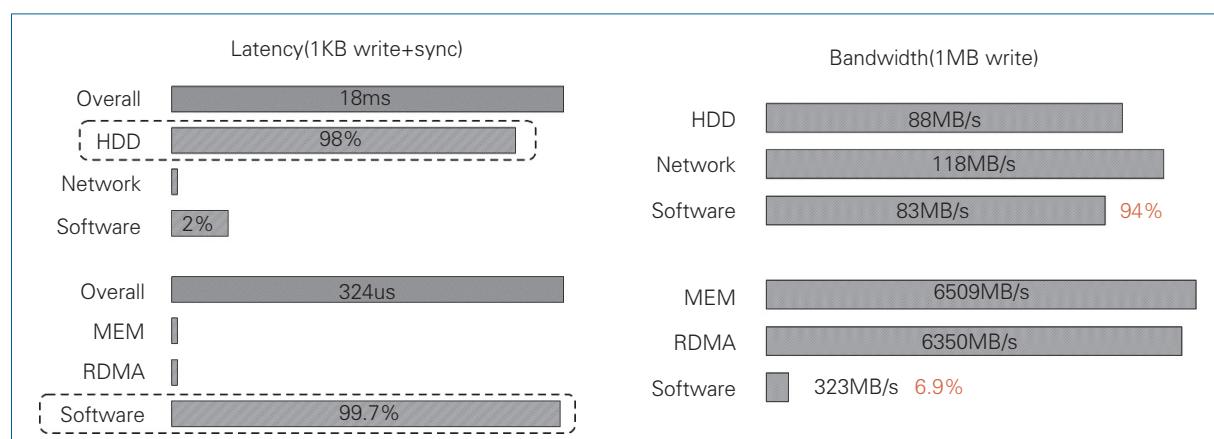


图3 GlusterFS在磁盘/以太网环境与内存/RDMA<sup>1</sup>环境下延迟与带宽的比较<sup>[19]</sup>

<sup>1</sup> RDMA, Remote Direct Memory Access, 远程直接内存访问。

年来相关的研究论文也越来越多，然而以怎样的方式去用持久性内存、如何用好持久性内存等问题仍然需要更多的研究。随着研究的推进，我们认为持久性内存可能将在以下几个方面出现较大的变革。

**1. 软硬件结合更紧密：**与传统外存储相比，持久性内存提供了极高的数据持久化性能。在传统存储系统中，由于外存储较慢，软件处理开销几乎可以忽略；而在持久性内存中，软件处理开销尤为显著。进一步，在持久性内存系统中，CPU 若过多地参与到数据处理中，将显著影响应用性能。近年来 FPGA 以及新型硬件加速器的快速发展对 CPU 数据功能的卸载提供了新的机会。利用硬件的比较优势并实施软硬件协同的设计，将是持久性内存系统设计中一个重要的研究方向。

**2. 系统软件的变革：**持久性内存驱动计算机系统软件的变革，从基础的数据结构到内存与文件系统管理，再到系统调用、操作系统内核设计、以及分布式系统设计，都将出现新的变革。持久性存储介质的以字节粒度访问、读写不对称、磨损等新特性与传统存储介质有较大差异，这些新特性对系统软件设计提出了新要求，并将推动这些新要求被实现。

**3. 新的编程模式与应用：**持久性内存的快速发展一方面是因为硬件技术的发展，另一方面也是因为应用对数据访问苛刻的性能要求。以图计算为代表的内存计算应用呈现较多随机的字节访问模式，这对持久性内存更为友好，同时也需要对持久性内存的可扩展性和数据一致性等方面进一步支持。此外，如何结合持久性内存来优化或设计数据处理框架也有待进一步研究。 ■



陆游游

CCF 专业会员。清华大学助理教授。主要研究方向为存储系统、分布式系统和计算机体系结构等。

luyouyou@tsinghua.edu.cn



舒继武

CCF 会士。清华大学教授。主要研究方向为网络 / 云 / 大数据存储系统、基于非易失存储器件的存储系统与技术、存储安全与可靠性和并行 / 分布式处理技术等。

shujw@tsinghua.edu.cn

## 参考文献

- [1] Liu J, Jaiyen B, Veras R, et al. RAIDR: Retention-aware intelligent DRAM refresh[C]//Proceedings of the 39th Annual International Symposium on Computer Architecture (ISCA). IEEE, 2012: 1-12.
- [2] Swanson S, Caulfield A M. Refactor, reduce, recycle: Restructuring the I/O stack for the future of storage[J]. Computer, 2013, 46(8): 52-59.
- [3] Lu Y, Shu J, Sun L. Blurred persistence: Efficient transactions in persistent memory[J]. ACM Transactions on Storage (TOS), 2016, 12(1):1-29.
- [4] Condit J, Nightingale E B, Frost C, et al. Better I/O through byte-addressable, persistent memory[C]//Proceedings of the 22nd ACM SIGOPS Symposium on Operating Systems Principles (SOSP). ACM, 2009: 133-146.
- [5] Lu Y, Shu J, Sun L, et al. Loose-ordering consistency for persistent memory[C]//Proceedings of the IEEE 32nd International Conference on Computer Design (ICCD). IEEE, 2014.
- [6] Pelley S, Chen P M, Wenisch T F. Memory persistency[C]//Proceedings of the 41st ACM/IEEE International Symposium on Computer Architecture (ISCA). 2014: 265-276.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>



### 封面设计说明

本期主题为新型存储和内存计算。插图的主角是具有未来感的存储器，代表新型存储，周围点块状分布的块面则代表数据流，表示计算。本次采用 2.5D 的风格，整体具有未来感。

设计：SEEEKLAB（设计总监 田力）

# 面向持久性内存的存储系统设计与优化

关键词：持久性内存 存储系统 优化

蒋德钧 王盈 熊劲  
中国科学院计算技术研究所  
中国科学院大学

## 基于持久性（混合）内存的存储系统

新型非易失性存储器件(Non-Volatile Memory, NVM)，例如相变存储器(Phase Change Random-Access Memory, PCRAM)、电阻型存储器(Resistive Random-Access Memory, ReRAM)、磁阻式存储器(Magnetic Random-Access Memory, MRAM)，以及英特尔和美光(Micron)公司于2017年联合推出的3D XPoint，具有字节寻址、高速访问和可持久化保存数据的特性，可以直接连接在内存总线上作为持久性内存使用。但是，NVM较长的写延迟、磨损特性以及容量成本导致使用NVM完全替代动态随机存储(DRAM)作为单一存储，并不是一个高效的生产环境方案。DRAM和NVM组成的持久性混合内存结构是一种有望最早应用于生产环境的存储结构。例如，惠普公司2016年发布的未来计算机系

统原型“The Machine”就采用了混合内存架构，包括GB级的DRAM和TB级的NVM。

近年来，基于持久性（混合）内存的存储系统研究工作受到学术界和工业界的持续关注，其中两大类代表性存储系统为文件系统和键值存储系统。虽然两者提供给应用的操作接口不一样，但是在面向NVM持久性（混合）内存的系统设计上，两者具有相似性。例如，索引结构在文件系统和键值存储系统中都广泛使用，如何基于NVM持久性内存设计高效的索引结构是两类存储系统都面临的挑战。表1列出了近年来面向NVM持久性（混合）内存的文件系统和键值存储系统的代表性关键技术。

## 面临的挑战

首先，索引结构在文件系统和键值存储系统中广泛使用，传统的哈希、B+树和LSM-Tree(Log-

表1 面向持久性（混合）内存的存储系统关键技术

研究内容	典型系统或工作
数据通路优化	BPFS <sup>[1]</sup> , PMFS <sup>[2]</sup> , SCMFS <sup>[3]</sup> , ext4-dax <sup>[4]</sup> , HinFS <sup>[5]</sup> , NOVA <sup>[6]</sup> , NOVA-Fortis <sup>[7]</sup> , Aerie <sup>[8]</sup> , Strata <sup>[9]</sup> , DevFS <sup>[10]</sup>
元数据优化	SPFS <sup>[11]</sup> , ByVFS <sup>[12]</sup> , BPFS <sup>[11]</sup> , PMFS <sup>[2]</sup> , SCMFS <sup>[3]</sup> , NOVA <sup>[6]</sup> , Aerie <sup>[8]</sup> , Strata <sup>[9]</sup> , DevFS <sup>[10]</sup>
索引结构	FPTree <sup>[13]</sup> , Masstree <sup>[16]</sup> , CPHash <sup>[18]</sup> , NV-Tree <sup>[14]</sup> , wB+-Tree <sup>[15]</sup> , PFHT <sup>[17]</sup>
崩溃一致性	顺序写技术 <sup>[2,13~15]</sup> , 减少日志开销 <sup>[19~21]</sup> , 避免写日志 <sup>[22~24]</sup> , 降低保序开销 <sup>[21,25]</sup>

Structured Merge Tree) 等索引结构，均是针对传统存储介质设计的，并没有考虑 NVM 的特性，如读写延迟非对称、写次数有限，直接应用于 NVM 存储系统中会影响系统性能和 NVM 的使用寿命。

其次，无论是文件系统，还是键值存储系统，都需要保证崩溃一致性，包括持久性 (durability) 和原子性 (atomicity)。持久性是指数据持久地保存在存储介质中，在系统断电、宕机等情况下，数据不会丢失。原子性是指事务数据完全被更新或者完全没变，不存在只更新部分数据的情况。传统的面向磁盘和固态硬盘 (Solid State Drive, SSD) 的存储系统，使用 `fsync()` 和 `msync()` 系统调用来保证 DRAM 内存中数据的持久性。然而 `fsync()` 和 `msync()` 针对的是外存块设备，面向持久性内存时，上述命令无法保证 CPU 缓存中数据的持久性。

最后，由于 NVM 的读写延迟接近 DRAM，相比磁盘和 SSD，持久性内存存储系统的硬件访问延迟大大降低，而软件系统的调用与执行开销增加。因此，在设计与优化中需要重点考虑如何降低软件开销。

## 面向NVM的索引结构优化

哈希索引访问随机性强，广泛应用于随机访问性能高的 DRAM 内存。链式哈希索引是一种常用索引，使用一个哈希函数对 Key 进行哈希计算，决定存储数据的哈希桶，哈希冲突通过链表解决。虽然易于实现，但是链式哈希有两个主要缺点：(1) 哈希桶有多个键值对时，链表的顺序查找导致 CPU 缓存局部性差；(2) 插入、删除请求导致频繁的堆分配和释放。

B+ 树索引节点大小可根据存储介质不同而灵

活设置，性能均衡，因此大量应用于内存系统和外存系统。B+ 树在插入和删除数据时，会引发叶子节点的排序、分裂、合并等操作，并会导致大量的写。B+ 树索引查找时需要从根节点开始层层查找，因此查找效率低于哈希索引。但是对于范围查找而言，B+ 树索引找到叶子节点后，会查找该节点包含的数据，并通过叶子节点间指针继续查找下一个叶子节点，直到找到所有需要的数据，因而性能明显高于哈希索引。

LSM-Tree 采用日志结构写，将小粒度随机访问转换成大粒度顺序访问，适合顺序访问性能明显高于随机访问的介质，如机械硬盘 (Hard Disk Drive, HDD) 和 SSD。LevelDB<sup>[26]</sup>、RocksDB<sup>[27]</sup> 和 WiscKey<sup>[28]</sup> 等持久化键值存储系统均采用 LSM-Tree 索引。但是，LSM-Tree 在将小粒度请求聚合成大粒度请求的同时，也带来了读写放大的问题。

表 2 为上述三种经典索引结构在索引结构特点、优势、不足和所适用的存储介质这几个维度上的对比。

近年来，研究者提出了多种面向 NVM 的索引结构优化技术，相关研究主要集中在优化 B+ 树索引，主要技术可以简单分为两方面：

一方面，传统的 B+ 树索引为保证叶子节点有序，在插入和删除时，平均要移动叶子节点中一半的数据，会导致大量的 NVM 写。为了减少叶子节点排序导致的写，文献 [15] 首先提出针对 NVM 内存的叶子节点不排序设计：插入操作直接在叶子节点的尾部添加一个数据项，并增加有效数据项的计数；对于删除操作，使用叶子节点的最后一项覆盖要删除的数据项，并减少有效数据项的计数。因此，插入和删除操作存在两次 NVM 写。为了进一步减少删除操作的 NVM 写，进一步提出在叶子节点中增加一个位图，删除操作只需要更新位图中相应的

表2 经典索引结构对比

索引结构	特点	优势	不足	适用介质
链式哈希	无序	单键值查找效率高	范围查找效率低，缓存局部性不高	DRAM 内存
B+树	有序	范围查找效率高	单键值查找效率低，排序开销大	DRAM 内存及 HDD/SSD 外存
LSM-Tree	有序	范围查找效率高，随机写聚合成顺序写	单键值查找效率低，读写放大	HDD/SSD 外存

数据位，因而只需要一次 NVM 写。

另一方面，上述减少 B+ 树写的优化，由于叶子节点不排序，对于查找操作而言，需要顺序查找叶子节点，因而查找性能低于传统的 B 树索引。为了提高查找效率，NVTTree<sup>[14]</sup> 在叶子节点的头部增加一个计数器，用于记录键值对的数量，通过计数器直接定位到最后一个键值对的位置，并从后向前查找，提高了查找效率。FPTree<sup>[13]</sup> 则在叶子节点头部增加了与 CPU 缓存行匹配的 Key 的指纹 (fingerprints)，通过预取指纹和判断 Key 的指纹是否相等来减少 NVM 访问。

此外，针对 Hash 索引和 NVM 读写延迟非对称的特点（读延迟与 DRAM 相近，但是写延迟明显高于 DRAM），也有研究工作提出改进方法，例如面向 PCM 友好的哈希表设计 PFHT (PCM-Friendly Hash Table)<sup>[17]</sup>，该设计改进传统布谷鸟哈希 (cuckoo hashing) 主表替换次数限制，获得较高的性能。

## 崩溃一致性保证

传统存储系统普遍使用写前日志 (Write-Ahead Log, WAL)、写时复制 (Copy-on-Write, CoW) 技术或者 log-structure 保证原子性。写前日志技术在将易失性缓存中的数据写回真正的持久化位置之前，需要先在持久化存储介质的日志区记日志，写日志完成后才能将数据写回真正的位置。写时复制一般用在树形索引的存储系统。写时复制在执行前先将旧数据复制下来，然后在副本中进行修改，最后原子性修改父节点的指针指向新的副本，使旧数据无效。修改父节点指针的过程会引起迭代修改祖父节点，一直迭代到根节点，造成写放大。Log-structure 顺序地将所有系统的写操作追加到日志中，不需要额外记日志，对写操作友好。由于系统在写操作过程中会不断修改数据，导致文件部分数据追加到日志中，文件的数据随机分布，读操作的局部性差。

为了降低在持久性内存上提供崩溃一致性保证的开销，近年来有不少研究成果，主要工作可以分为以下几类：

第一，基于顺序写的一致性 (write-ordered consistency) 技术。该技术可以避免写前日志技术和写时复制技术导致的两次写问题。顺序写技术要求存在一个原子粒度的标志，该标志可以使用 CPU 原子指令更新，并用于标识一个请求是否完成。顺序写的基本思想是先写数据（追加或者异地更新），再原子地更新该原子标志。顺序写利用了 NVM 作为持久性内存时可以使用原子指令更新数据的特点。多个基于 NVM 的键值存储系统，例如 NVTTree<sup>[14]</sup>、wB+-Tree<sup>[15]</sup> 和 FPTree<sup>[13]</sup>，使用顺序写技术保证索引结构的崩溃一致性。

第二，减少日志开销的技术。写前日志普遍存在两次写（日志和数据）的问题，性能开销较大。特别是文件系统，即使只修改了页中一小部分，写日志时也要写整个页 (4KB)，因此有研究工作提出减少日志写大小的技术方法。例如，Shortcut-JFS<sup>[19]</sup> 采用了两种写策略。在写日志阶段，如果数据块 (4KB) 中改变的字节数超过一半，则在日志区写整个块；否则，在日志区只写改变的字节。在写文件原来位置阶段，如果一个数据块改变的字节超过一半，则原来文件块的指针直接指向日志块的位置，从而减少再写文件的开销；否则，采用传统的方式，将改变的数据写回文件。为了减少大粒度日志的开销，FSMAC<sup>[20]</sup> 同样提出细粒度日志技术，比如以文件系统 inode 大小 (128B) 写日志。类似地，NVWAL<sup>[21]</sup> 同样在日志区只写 B 树节点中改变的字节，以减少日志开销。

第三，利用 NVM 的非易失性来避免额外的写日志。例如，文献 [22] 提出的 UBJ (Union of Buffer cache and Journaling) 技术采用 NVM 作为非易失缓存，原地提交数据，将更新过的缓存块设置为冻结状态，不再修改，即将该缓存块直接作为日志，不需要额外写日志。当需要再次更新该缓存块时，将更新写到新的缓存块中。Kiln<sup>[23]</sup> 则使用非易失性的末级缓存和内存避免写日志。文献 [24] 提出全系统持久化 (Whole System Persistence, WSP)，进一步要求各级存储均非易失，包括寄存器、各级 CPU 缓存和内存。WSP 只在宕机时刷回缓存 (flush-on-fail)，

并原地更新 NVM 内存。

第四，降低写日志完成判别开销的方法。WAL 需要保证写的顺序性，日志写完后才能写数据，因此，传统存储系统在每个日志的后面增加一个提交(commit)标志，以标识该日志是否写完成，并且在日志写完后才能写提交标志。多个 NVM 存储系统<sup>[3,29,30]</sup> 使用保序指令(如 SFENCE、MFENCE)，保证缓存数据写回 NVM 的顺序，该方式引入频繁的保序操作，影响了系统性能。为了降低写日志和写提交之间的顺序性开销，文献[21] 提出在所有日志写完后只调用一次 SFENCE 指令，并进一步提出在提交标志后增加一个校验和(checksum)，不需要保证写日志和提交标志的顺序性。在宕机时，通过校验和判断日志是否写完成。放松顺序性的一致性策略(Loose-Ordering Consistency, LOC)<sup>[25]</sup> 则采用 Eager Commit 协议避免使用提交标志。LOC 为每个事务的日志增加一个元数据块，在元数据中记录日志的数据块个数，并在日志的最后一个数据块记录已经写的数据块个数。宕机恢复时，如果日志元数据块和日志数据块记录的个数相等，说明日志写完成，否则日志没有完成。

最后，分布式日志技术。传统存储系统采用集中式日志，多个并发的操作只能串行地写日志，影响系统性能。文献[31] 提出 NVM 辅助的分布式日志(distributed logging)，每个日志区配置一个 NVM 缓存，日志写到 NVM 后即可提交。当 NVM 缓存满时才写回外存。分布式日志对比了页级划分和事务级划分两种日志空间划分方式。页级划分指的是访问同一个物理页的请求写到同一个日志缓存中，事务级划分指的是同一个事务的请求写到同一个日志缓存。评测显示，事务级划分避免了跨非统一内存访问结构(Non-Uniform Memory Access, NUMA)的访存，因而性能更高。文献[32] 同样提出分布式日志技术 NV-Logging，每个事务有一个独立的写日志。与文献[31] 划分多个日志区不同，NV-Logging 不在物理上划分日志区，所有的日志项组成一个环形日志区。为了提高性能，NV-Logging 提供了两种写日志策略：flush-on-insert 策略在 DRAM 缓存中

的数据变脏后立即写日志；flush-on-commit 策略在后台异步地写日志。NVM 文件系统 NOVA<sup>[6]</sup> 采用了类似的分布式日志设计，为每个 inode 节点分配一个日志区。

在实际的存储系统中，不同请求操作的数据粒度通常不同。例如，在文件系统中，读请求只需要修改文件的访问时间(atime)元数据，因此可以使用原子指令保证原子性，而创建文件请求则需要修改多个 inode 元数据，无法通过原子指令保证原子性。因此，基于持久性内存的存储系统往往需要综合运用上述各项技术，结合原子写、写前日志和写时复制技术，来保证不同操作的原子性。

## 数据通路优化

在基于磁盘的传统文件系统中，数据通路较长，包括虚拟文件系统(VFS)、高速页缓存(page cache)、物理文件系统、通用块层、驱动等。而在基于持久性内存的文件系统中，上述数据通路需要被精简与优化。现有研究工作对数据通路的优化可以简单分为以下两类。

一是在内核中优化，即降低内核中文件系统的操作栈。BPFS<sup>[1]</sup>、PMFS<sup>[2]</sup>、SCMFS<sup>[3]</sup>、ext4-dax<sup>[4]</sup>、NOVA<sup>[6]</sup> 和 NOVA-Fortis<sup>[7]</sup> 认为针对快速的 NVM 存储设备，文件系统的操作栈冗余。特别对于 page cache、IO 调度层和设备层，在持久性内存文件系统中不是必要的。因此现有的持久性内存文件系统通常只保留虚拟文件系统层。文件的读写操作直接在持久性内存文件系统中执行，不需要 DRAM 缓存。这种方式降低了文件读写时不必要的软件开销，提升了文件系统的性能。但是值得注意的是，NVM 相比于 DRAM 有较高的写延迟，因此，文献[5] 认为直接去掉 page cache 使得所有的写操作直接在 NVM 中执行，增加了写文件的开销。然而，NVM 具有和 DRAM 相同的读延迟，保留 page cache 对于文件同步写和读操作是低效的。因此，该文献提出了持久性内存感知的写缓冲机制 HinFS，建立缓冲区缓存不需要立即持久化的数据，对于那些需要立即持久

化的数据(同步写)则直接写到 NVM 中。除此之外, HinFS 支持同时在缓存和 NVM 中读数据。通过这种方式, HinFS 降低了 NVM 高写延迟对文件系统性能的影响, 并且几乎对文件的读性能无影响。

二是在用户态建立文件系统。Aerie<sup>[8]</sup>、Strata<sup>[9]</sup>、DevFS<sup>[10]</sup> 在去掉 page cache、IO 调度层等软件开销的基础上进一步降低了系统调用和内核干预的开销, 应用可以在用户态直接访问文件的数据。例如 Strata 将热数据存放在 NVM 中, 应用可以直接通过 libFS 访问。当读数据时, 先通过 libFS 来访问数据, libFS 中不存在时再到 kernelFS 中查找。只有在 kernelFS 中才需要内核干预。对于写数据而言, 应用通过 libFS 将文件更新直接写在 NVM 中, 不需要内核干预, 只有文件执行合并操作时才需要访问内核。

## 元数据通路优化

在文件系统中, 对于小文件的操作而言, 元数据的开销通常占比较高, 因此, 在基于持久性内存的文件系统中, 元数据通路的优化同样重要。与数据通路优化类似, 现有持久性内存文件系统的元数据优化也可以简单分为以下两类。

一是基于内核文件系统的优化, 即在内核中进行元数据优化, 具体又可以分为两方面工作。一方面, 保留虚拟文件系统, 文件系统对元数据的优化只限定在物理文件系统内部。例如 SCMFS<sup>[3]</sup> 使用页表索引文件数据块, NOVA<sup>[6]</sup> 在 DRAM 中建立文件数据块和目录项索引, 以降低物理文件系统中基于 NVM 的索引结构导致的性能开销。NOVA 预分配空闲空间和 inode 到各个 CPU 核, 并且给每个文件单独记录日志, 以降低元数据操作中多线程的争用, 提升文件系统的并发性。另一方面, 优化或者移除虚拟文件系统, 例如为了降低虚拟文件系统对元数据操作性能的影响, SPFS<sup>[11]</sup> 移除了虚拟文件系统, 所有的操作都在物理文件系统中执行, 且整个过程不使用 DRAM。这种方式全部去掉 DRAM, 所有的元数据增删改操作都需要立即更新到 NVM 中, 降

低了文件元数据操作的性能。ByVFS<sup>[12]</sup> 则建议绕过虚拟文件系统, 使得元数据操作直接在物理文件系统中执行, 提升了元数据操作性能。

二是直接建立用户态文件系统<sup>[8-10]</sup>, 应用可以直接通过库函数访问文件的元数据。用户态文件系统只有一些必要的元数据修改操作需要进入内核执行。例如 Aerie<sup>[8]</sup> 直接通过用户态库访问文件元数据, 同时使用一个可信赖的第三方服务保证文件系统安全。

## 未来研究展望

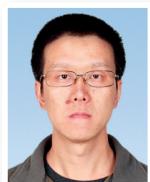
基于笔者前期的研究经验, 在面向持久性(混合)内存的存储系统设计与优化中, 以下几个方面应该获得更多关注。

**索引结构设计与优化。**索引结构是影响存储系统性能的关键因素。持久性内存同时具备字节寻址、高速随机访问和可持久化的特点, 传统面向块接口和慢速设备的存储系统索引结构究竟应该如何面向持久性(混合)内存重新设计或优化, 仍然值得进一步探索。例如, 如何充分组合利用不同类型索引结构的特点组成混合索引支持高效操作, 如何利用硬件技术对索引并发与事务进行高效支撑等等。

**元数据优化。**现有的面向持久性内存的存储系统设计与优化重点关注数据通路优化和降低崩溃一致性开销, 对于元数据操作的优化而言, 还有不少研究问题值得探索与解决。例如, 在持久性内存上构建文件系统时, 虚拟文件系统的收益与开销是否需要重新考虑? 用户态文件系统绕开了虚拟文件系统, 如何保证共享库的访问方式有效支持文件系统的安全性、完整性、正确性和崩溃一致性?

**混合介质管理。**从目前 NVM 器件的研究进展来看, NVM 存储器的容量暂时还无法与传统磁盘或 SSD 相比, 而在实际的生产环境中(例如数据中心应用场景), 数据量往往很大, 使用持久性内存保存所有数据, 就成本而言并不可行, 因此 DRAM、NVM、SSD 和磁盘仍然会组成层次化的存储体系结构。现有研究工作重点关注单一持久性内存的优化,

但是如何基于上述层次化的存储体系结构（尤其是增加了一层持久性内存）设计与优化存储系统，还有很多挑战需要解决。 ■



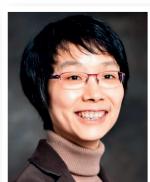
蒋德钧

CCF专业会员。中国科学院计算技术研究所副研究员。主要研究方向为面向新型存储器件的体系结构与系统软件、分布式系统等。[jiangdejun@ict.ac.cn](mailto:jiangdejun@ict.ac.cn)



王 盈

CCF学生会员。中国科学院计算技术研究所博士生。主要研究方向为基于新型存储器件的存储系统。[wangying01@ict.ac.cn](mailto:wangying01@ict.ac.cn)



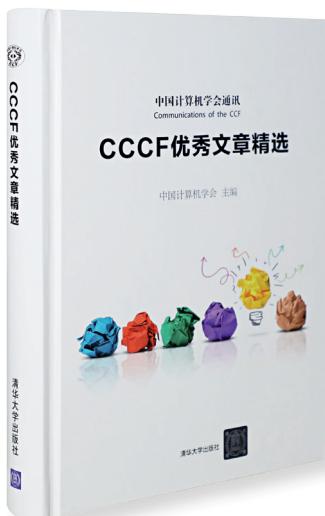
熊 劲

CCF高级会员。中国科学院计算技术研究所研究员。主要研究方向为分布式存储系统、基于SSD和NVM的存储系统、大数据存储系统等。[xiongjin@ict.ac.cn](mailto:xiongjin@ict.ac.cn)

## 参考文献

- [1] Condit J, Nightingale E B, Frost C, et al. Better I/O Through Byte-addressable, Persistent Memory[C]// *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles (SOSP '09)*. ACM Press, 2019:133-146.
- [2] Dulloor S R, Kumar S, Keshavamurthy A, et al. System software for persistent memory[C]// *Proceedings of the Ninth European Conference on Computer Systems(EuroSys '14)*. ACM Press, 2014: Article No. 15.
- [3] Wu Xiao, Reddy A L N. SCMFS: A File System for Storage Class Memory[C]// *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis (SC'11)*. ACM Press, 2011: Article No.39.
- [4] Corbet J. Supporting filesystems in persistent memory[OL]. <https://lwn.net/Articles/610174/>.
- [5] Ou J, Shu J, Lu Y. A High Performance File System for Non-volatile Main Memory[C]// *Proceedings of the Eleventh European Conference on Computer Systems (EuroSys'16)*. ACM Press, 2016: Article No.12.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>



中国计算机学会 主编

# 《CCCF 优秀文章精选》 正式出版

清华大学出版社 出版

书号：978-7-302-51506-7  
各大电商均有销售



清华大学出版社



中国计算机学会通讯



CCF在150期之际，精选出50多篇优秀文章，汇集成书，分“教学篇”“观点篇”“技术篇”“人物篇”等，可窥见计算技术界的心路历程，更可看到对问题的思辨。

# 分布式共享内存与内存计算

关键词：内存计算 分布式共享内存

杨帆 洪扬 陈榕等  
上海交通大学

## 分布式共享内存的必要性

计算和存储是大数据计算的两个主题。随着大数据应用规模的快速扩张，不断推进计算和存储的可扩展性成为提升大数据应用性能的关键。早期的大数据计算主要致力于提高计算的并行度，以基于传统数据库和分布式文件系统的大数据系统为代表。而随着内存成本的降低，数据中心的计算机已经普遍配备了大容量的内存。内存的带宽是传统慢速存储介质的几十倍，而访问延迟可以达到硬盘的千分之一。加之远程直接内存访问 (Remote Direct Memory Access, RDMA) 等新型设备的使用，网络延迟进一步降低，现在 InfiniBand 网卡的性能即将达到 200Gbps 和 0.6μs 延迟（如图 1 所示）。充分利用内存的高带宽和低延迟特性的内存计算，成为了推动大数据计算性能跨越式发展的关键技术。



图1 内存速度和网络延迟变化趋势

内存计算可以应用在各种计算系统中，如基于内存的内存存储系统<sup>[1]</sup>、数据库系统<sup>[2~5]</sup>、图计算系

统<sup>[6,7]</sup>、深度学习系统等。由于单机可扩展性的局限，主流的内存计算系统已经转向分布式和大规模的硬件设施。大多数的大规模内存计算系统都部署在中小规模的集群甚至大规模的数据中心里。管理和协调分布式的内存资源和计算资源是现在内存计算的主要研究方向之一。

分布式共享内存是一种经典的分布式编程模型，在近几年更是借着内存计算的流行而重新获得关注。分布式共享内存为应用提供了一个统一的内存抽象，使分布式系统中的任意节点可以同等地访问位于其他节点上的内存数据，大大简化了编程的难度，优化了逻辑。在近几年，新型加速硬件和存储设备、高速网络等得到广泛普及，100G 甚至 200G 的以太网和一些专用网络都在数据中心得到大量部署，支持 RDMA 特性的网络底层接口为上层应用提供了新的优化维度。另一方面，处理器性能不断提升，集成了多种硬件特性的支持，也为软件的横向扩展提供了新的机遇。在这样的场景下，分布式共享内存作为一种直观的抽象，也被应用于扩展新的软硬件背景下的内存计算应用。近年来，许多基于分布式共享内存的内存计算系统被提出（见表 1），分布式共享内存也将因其简单高效的抽象而得到进一步的发展，在新硬件特性的支持下，成为内存计算的支撑技术之一。同时，基于分布式共享内存的“多虚一”虚拟化架构，也是研究的一个重要方向（关于虚拟化技术的发展介绍可参考《中国计算机学会通讯》2017 年第 6 期专题）。

表1 基于分布式共享内存的内存计算系统

研究领域	典型系统
数据库	FaRM <sup>[8]</sup> , Pilaf <sup>[9]</sup> , DrTM <sup>[10]</sup> , DrTM+R <sup>[11]</sup> , FaSST <sup>[12]</sup> , GAM <sup>[13]</sup>
图计算	Grappa <sup>[14]</sup> , m&m <sup>[15]</sup>
操作系统	LITE <sup>[16]</sup> , Infiniswap <sup>[17]</sup> , LegoOS <sup>[18]</sup>
虚拟化	ScaleMP <sup>1</sup> , TidalScale <sup>2</sup>
存储	Octopus <sup>[19]</sup> , DSPM <sup>[20]</sup>

## 挑战和特点

现阶段对分布式共享内存系统的研究，已与上世纪末其被广泛研究时有了很大的不同，主要可以归纳为硬件性能的变化和新兴应用特点的改变。

## 新兴应用

分布式共享内存系统的广泛研究基本发生在分布式、大数据、内存计算等领域的蓬勃发展之前，当时用于评测分布式共享内存系统的基准（如 SPLASH-2<sup>[21]</sup>）往往是一些用于数学计算等涉及少量内存访问的并行程序。而如今分布式共享内存系统的评测则应该与新兴应用（以互联网应用为代表）相适应，其高并发、低延迟、涉及海量数据等特点，将对系统的设计提出新的要求。显著的应用场景变化往往意味着系统设计的侧重点发生变化，以往的实践经验可能不再成立，技术难点和系统瓶颈发生转移。

## 硬件发展

计算机各部分组件不断进行着高速的创新和发展。二十年前的分布式共享内存系统通常是小规模的集群，节点的硬件也仅为单核或者小规模的多核，内存也较小，网络带宽在100MB左右。而现在的数据中心硬件已经得到了数量级的增强，且不说集群规模已经上千或者上万，单个节点就已经具有数

十个甚至上百个处理核心，数百GB的内存，以及100Gbps的网络互联。但硬件性能再好也不一定能使运行其上的系统软件轻易获得高效的产出。硬件性能的跨越数量级的提升，往往使当时设计软件的一些基本假设发生变化，计算资源和网络资源不对等的发展会带来新的系统瓶颈。

## 软硬件结合的设计

半导体技术的发展逐渐遇到瓶颈，计算机软件已经无法依靠处理器工艺的提升而享受“免费”的性能扩展。以内存计算为代表的大数据应用已经普遍采用多任务并发的设计，充分挖掘多核系统的处理能力，力图实现高效的垂直扩展。另一方面，内存计算的应用也在水平扩展方面不断挖掘新的机遇。而硬件发展也逐渐向专用化、异构化方向发展，不断加入的一些新的特性支持特定的软件模式。这些都意味着研究者需要结合最新的软件模式和硬件特性对分布式共享内存进行全新的设计。

上述挑战和特点并不是孤立的，而是相互交织在一起的。上层应用的特点和底层硬件的变化对于分布式共享内存系统意味着从上到下进行全面的革新和实践探索，工作面广、工作纵深长、工作量大，机遇与挑战并存。我们的团队有幸在其中进行了一部分初步的尝试与探索。

## 分布式共享内存架构

我们提出了一种全新的分布式共享内存架构。该系统通过挖掘现有的RDMA高速网络的特性和应用程序的访存特征，实现了一套高效的机器间内存同步的协议，以及一系列时空局部性感知的优化方案。为了验证该设计的实际有效性，我们实现了一个原型系统（以下称为MAGI）。该系统具有三个创新点：(1)一个分层的分布式共享内存架构，高效地利用了应用程序的访存时空局部性；(2)一套兼容

<sup>1</sup> <https://www.scalemp.com>。

<sup>2</sup> <https://www.tidalscale.com/>。

POSIX(Portable Operating System Interface of UNIX)编程接口的线程库，使原先的单节点应用程序可以透明地运行在分布式的集群上，而不用进行大规模的修改；(3)一套针对非统一内存访问结构(Non-Uniform Memory Access, NUMA)感知应用特征的优化方案，既能提供一个高性能的顺序一致性保证，又能在必要时放松一致性保证，适应NUMA感知应用程序的行为特征。

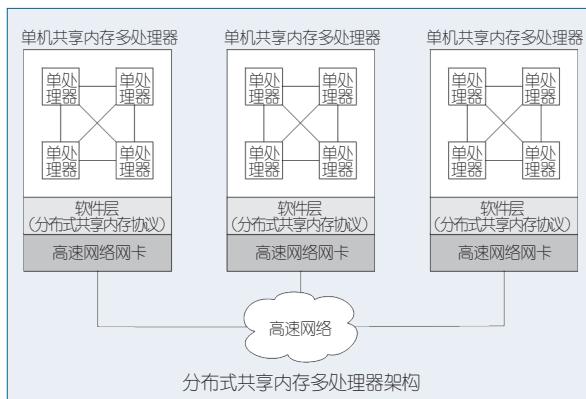


图2 分布式共享内存架构

## 架构组成

MAGI的整体架构如图2所示。该系统运行在一个分布式集群上。该集群由多台多核服务器组成，服务器配备了RDMA网卡，之间通过高速的RDMA网络相连。大数据应用程序运行在每台服务器上，并通过MAGI提供的一致的共享内存抽象访问任意一台机器上的内存资源，使用与单节点程序完全兼容的硬件指令集。

我们采用分层的分布式共享内存架构，即运行在同一台机器上的应用程序线程，通过本地的硬件缓存一致性机制实现对本地内存的共享访问。另一方面，运行在不同机器上的线程，则通过一套分布式共享内存协议来进行内存数据的同步。

运行在该系统上的应用程序使用与POSIX标准兼容的C语言库和线程库进行编程，这一套兼容的

系统由我们实现的软件接口库来提供。这样原本为单机编写的应用程序，几乎不需要修改即可运行在该分布式集群上。

该系统实现了一套NUMA感知的同步原语。传统的线程同步方法往往因为粒度过细以及过多的Cache访问导致在分布式场景下的假共享现象，即虽然两个变量在不同的Cache块中，但是因为分布式共享内存基于缺页的数据同步方式，线程之间往往需要耗费大量的时间处理协议以完成简单的同步任务，甚至这些协议的执行都是多余的。本系统实现的同步原语采用分层的同步办法，首先在节点内进行同步，继而在节点之间进行同步，最后达到减少节点间消息传递和降低同步开销的效果，提升了线程间同步的可扩展性。

## 优化方案

本系统还提出了一系列针对RDMA和当前多核硬件的优化方案，以提升内存计算应用在分布式共享内存场景下的性能。

**预测性缺页机制** 大数据内存计算应用往往需要处理大量的数据，计算过程中也会产生大量数据，更重要的是线程之间也需要对数据进行同步以协作完成计算任务。传统分布式共享内存所依赖的数据共享机制是通过处理器的虚拟内存机制实现的，频繁的缺页异常会导致计算被中断。本工作提出的预测性缺页机制可以基于数据访问局部性的特点，预测性地从页所有者处获取连续的页，减少缺页次数，从而降低协议开销。

**批量刷新TLB机制** TLB<sup>3</sup>刷新是确保所有CPU对虚拟内存翻译的一致性的重要机制。然而，现在的TLB刷新依赖处理器间中断(Inter-Processor Interrupt, IPI)来实现，可扩展性很差。本工作提出的方法是将提升页权限的TLB积攒起来批量刷新，从侧面降低刷新TLB的开销。

**混合一致性模型** 本工作实现的混合内存一

<sup>3</sup> TLB (Translation Lookaside Buffer, 转换检测缓冲区)是一个内存管理单元，用于改进虚拟地址到物理地址转换速度的缓存。

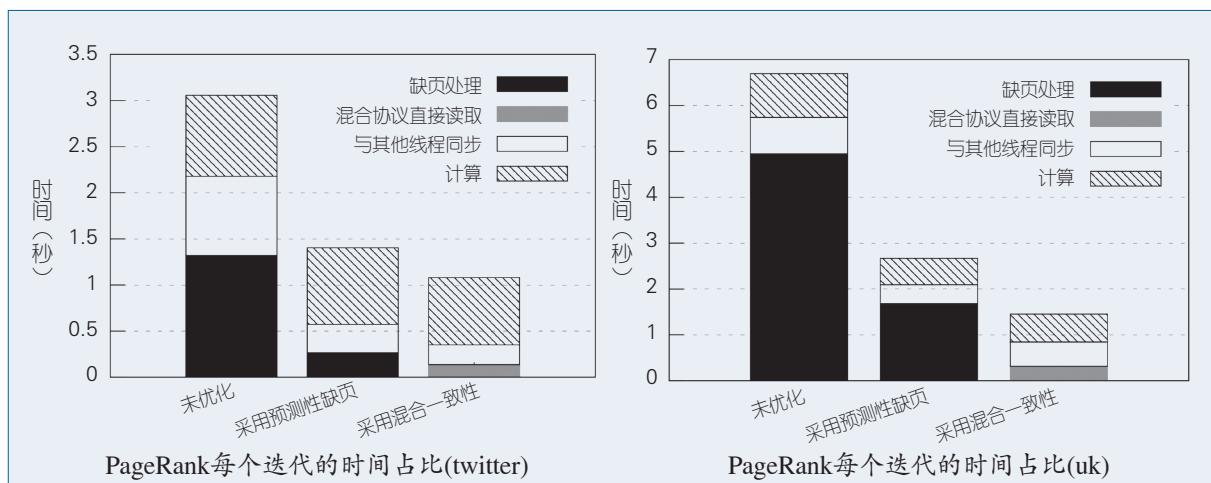


图3 分布式共享内存性能优化

致性模型，默认为程序员提供顺序一致性模型保证，完全兼容单机系统中的内存模型抽象。同时，我们的一致性模型允许程序员手动进行远程内存访问，此时一致性保证将放松以获得更高效的内存传输吞吐和低延时。这种动态可控的模型，为 NUMA 感知的应用程序提供了良好的性能和正确性的平衡点。

在集群上的实验证明，本系统提出的优化方案可以有效地降低图计算引擎执行中由于缺页导致的开销（如图 3 所示），同时也能减少由于缺页数量不均导致的线程执行时间不均，从而减少线程同步所需的时间。

## “多虚一”虚拟化架构

资源的扩展主要有横向扩展 (scale out) 和纵向扩展 (scale up) 两种方式。虚拟化资源横向扩展的优势是弹性分配，资源灵活使用，利用率高，但编程模型复杂；纵向扩展的优势是编程模型简单，避免了由于分布式系统和数据分区产生的软件复杂性，但硬件昂贵，灵活性差。针对内存计算等大规模计算需求，通过虚拟化技术，可以构建由多台横向扩展计算机组成的纵向扩展服务器<sup>[22]</sup>，我们称之为巨型虚拟机。不同于传统的“一虚多”方法，这种“多虚一”的跨物理节点虚拟化架构，可以将计算资源、存储资源和 I/O 资源虚拟化，构建跨节点的虚拟化资源池，对客户操

作系统提供统一的硬件资源视图，并且无须修改客户操作系统（类似于半虚拟化，如果客户操作系统感知到运行于“多虚一”，性能方面可进一步优化）。

目前，典型的虚拟化产品有 ScaleMP 和 TidalScale。其中 TidalScale 提出了一种软件定义服务器，通过超内核来构建虚拟资源池，将主板上所有的 DRAM 内存抽象为虚拟化的 L4 缓存，并且引入虚拟主板提供跨节点的虚拟设备连接；通过虚拟通用处理器、虚拟内存和虚拟网络构建虚拟资源池，并且资源可以迁移，也可灵活使用。通过复杂的缓存一致性算法和缓存管理算法，有效提升了性能。在一个包含 1 亿行、100 列的数据库和 128GB 内存的服务器上，由于内存容量大于服务器容量，导致分页频繁发生，以至于一个应用程序需花费 7 个小时才能完成 MySQL 的三次查询作业<sup>[23]</sup>。这个现象称之为“内存悬崖 (memory cliff)”，即当应用所需内存超过服务器内存时，性能快速下降。若该查询运行在采用两个 96GB 内存节点组成的 TidalScale 服务器上只需要 7 分钟，性能提升 60 倍<sup>[23]</sup>。

基于分布式 QEMU 和 KVM，我们提出了一种开源的“多虚一”巨型虚拟机 GiantVM 架构（如图 4 所示），围绕 RDMA 技术实现虚拟化硬件聚合和抽象，以 Libvirt 为巨型虚拟机上层接口，分布式 QEMU 提供跨节点虚拟机抽象，KVM 为巨型虚拟机下层物理机管理接口，基于 RDMA 提供低延迟

分布式共享内存。

我们分别将 Sv6 和 Linux 作为客户端，运行 PageRank 算法（针对具有 57.6 万个顶点和 510 万条边的图数据，迭代 20 次），与在同样配置的 Linux 集群上运行基于 Spark 的 PageRank 进行对比，实验结果表明，巨型虚拟机 GiantVM 的性能有较大的提升（如图 5 所示）。通过对 RDMA 网络和分布式同步进行优化，性能有望进一步得到提升。

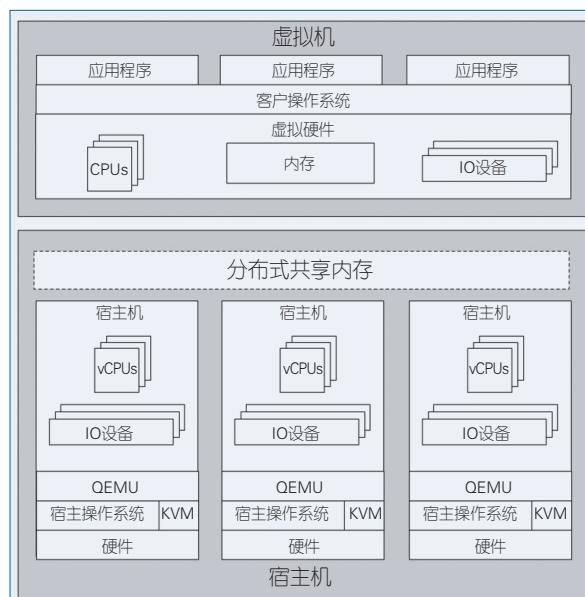


图4 开源的“多虚一”巨型虚拟机架构

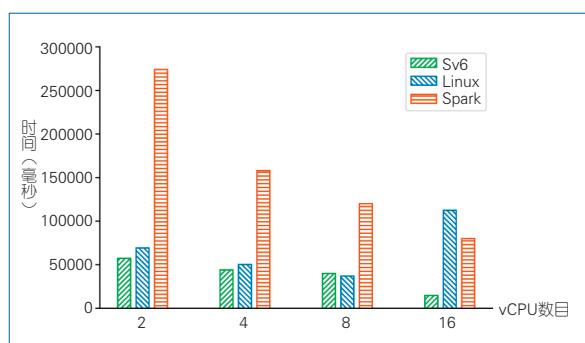
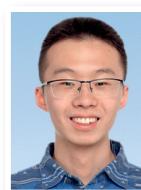


图5 巨型虚拟机性能对比

## 未来研究方向

目前主流的内存计算系统已经转向分布式和大规模的硬件设施，管理和协调分布式的存储资源和

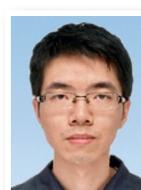
计算资源是现在内存计算的主要研究方向。通过虚拟化技术和分布式共享内存技术，戈登·贝尔(Gordon Bell)等人提出了将纵向和横向扩展相结合的方法，通过使用相同高性价比硬件，降低了横向扩展系统的硬件成本，提高了灵活性，实现了纵向扩展系统的固有简单性。同时，在“资源汇聚(resource aggregation)”和“资源分散(resource disaggregation)”方面，OSDI 2018 最佳论文提出的 LegoOS 操作系统针对应用特征<sup>[18]</sup>，通过高速网络连接，解耦硬件资源，取得了与现有 Linux 服务器相近的性能。与操作系统层面的资源分散相对应，可以在虚拟机监控器层面进行资源汇聚，通过“多虚一”架构，构建异构的虚拟资源池，通过虚拟化实现异构硬件归一化管理，简化编程模型。如何综合利用资源汇聚和分散，也将是今后的一个研究方向。 ■



杨帆

CCF 学生会员。上海交通大学硕士研究生。主要研究方向为分布式共享内存与巨型虚拟机系统。

Fan\_Yang@sjtu.edu.cn



洪扬

上海交通大学博士研究生。主要研究方向为系统虚拟化、并行与分布式计算等。

yang.hong@sjtu.edu.cn



陈榕

CCF 高级会员。上海交通大学副教授。主要研究方向为系统虚拟化、并行与分布式计算等。

rongchen@sjtu.edu.cn

其他作者：戚正伟

## 参考文献

- [1] Ousterhout J, Agrawal P, Erickson D, et al. The case for RAMCloud[J]. ACM Sigops Operating Systems Review, 2009, 54(4):121-130.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>

# 面向 HTAP 的内存数据库并发控制技术

张融荣 蔡 鹏 钱卫宁

华东师范大学

关键词：并发控制技术 数据库系统 HTAP

事务<sup>1</sup>包含一系列的数据库读写操作，其本质是对业务逻辑的抽象，这也是数据库的基本概念之一。在线事务处理(On-line Transaction Processing, OLTP)系统的研发与应用尽管已经有40年的历史，但随着移动互联网和智能手机的普及，大量终端用户依然给OLTP系统造成前所未有的负载压力，这就要求系统能够支持高通量低延迟的事务处理。

传统在线分析处理(On-line Analytical Processing, OLAP)系统的主要目标是分析企业OLTP系统产生的历史数据，基于数据分析结果，辅助市场

与经营决策。如图1所示，在传统IT系统架构下，OLTP系统产生的数据，通过ETL<sup>2</sup>工具，被定期抽取到数据仓库或OLAP系统中，这一过程通常需要数小时或几天时间。OLAP系统针对大规模数据加载和扫描，侧重数据分析与挖掘功能。比如，在销售行业，实时的交易记录往往由OLTP系统产生，而对于促销策略的制定及客户行为分析等则交由OLAP完成。

OLTP和OLAP系统面向不同的应用场景，两者所处理的负载也有很大差异。OLTP系统通常访问少量数据，并且使用索引，支持高效的数据查询、插入、更新、删除操作；而OLAP支持大批量数据加载/更新，以及复杂的SQL查询，这些查询通常需要访问大量的数据。2014年图灵奖获得者迈克尔·斯通布雷克(Michael Stonebraker)曾在2005年提出：通过单一数据库系统支撑各种应用的观念已经过时，即“One Size Does Not Fit All”<sup>[12]</sup>。随后，OLTP系统和OLAP系统分别针对自身的应用场景进行架构设计，在系统实现上充分考虑现代硬件特

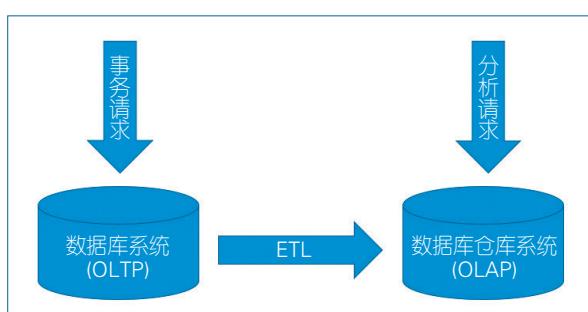


图1 传统OLTP与OLAP系统的架构

<sup>1</sup> 事务(transaction)在计算机术语中，是指访问并可能更新数据库中各种数据项的一个程序执行单元(unit)。事务通常由高级数据库操纵语言或编程语言(如SQL, C++或Java)书写的用户程序的执行所引起，并用形如begin transaction和end transaction语句(或函数调用)来界定。事务由事务开始(begin transaction)和事务结束(end transaction)之间执行的全体操作组成。

<sup>2</sup> ETL是英文Extract-Transform-Load的缩写，用来描述将数据从来源端经过抽取(extract)、交互转换(transform)、加载(load)至目的端的过程。

性。经过重新设计的 OLTP 和 OLAP 系统都已获得了成功应用。在数据分析方面，出现了像 BLU<sup>[1]</sup>，Vertica<sup>[2]</sup>，ParAccel<sup>[19]</sup>，GreenPlumDB<sup>[20]</sup>，Vectorvise<sup>[21]</sup>等一系列基于列存储的数据库。对于批量数据分析而言，列存储是一种相对于行存储更好的数据组织方式。得益于硬件性能的快速提升，也出现了 VoltDB<sup>[3]</sup>，Hekaton<sup>[4]</sup>，MemSQL<sup>[5]</sup> 等内存数据库。这些数据库的出现，极大地提高了事务处理的并发效率，也降低了事务的延迟。

数据的新鲜程度决定了它的商业价值。很多企业已经认识到实时数据分析对提高数据商业价值的重要性。目前，很多企业的在线业务通过新型 OLTP 系统实现了高吞吐低延迟事务处理，满足了互联网场景下的高并发事务请求；并且通过新型 OLAP 能够快速分析海量历史数据，优化业务发展。但是，现有的系统架构还无法实现实时分析最新的事务数据，主要原因是事务处理和分析是通过两套不同的系统完成，事务数据从 OLTP 加载到 OLAP 的过程耗时较长，并且涉及系统之间数据的一致性问题；另外，同时维护两套系统的成本和系统架构的复杂度，进一步阻碍了实时数据分析在企业的实际应用。

## HTAP的起源与现有系统架构

针对实时数据分析的商业前景，Gartner 在 2013 年首次提出了 HTAP(Hybrid Transactional/Analytical Processing) 的新型数据库系统概念<sup>[11]</sup>，旨在用一套系统同时支持 OLTP 和 OLAP 负载（如图 2 所示）。

2017 年，在《支持 HTAP 内存计算技术的市场指南》中，Gartner 定义了两种 HTAP 事务类型，分别是 Point of Decision HTAP (P-HTAP) 和 In-process HTAP (I-HTAP)<sup>[10]</sup>。P-HTAP 描述的应用场景是数据库系统同时承担 OLTP 和 OLAP 两种负载，但是 OLTP 负载中的事务和 OLAP 中的分析型查询来自

不同的应用系统；I-HTAP 是指事务操作和分析型查询可以存在于单个事务内。相比 P-HTAP，I-HTAP 更加具有技术挑战性。目前支持 HTAP 的数据库系统通常针对 P-HTAP 进行设计。

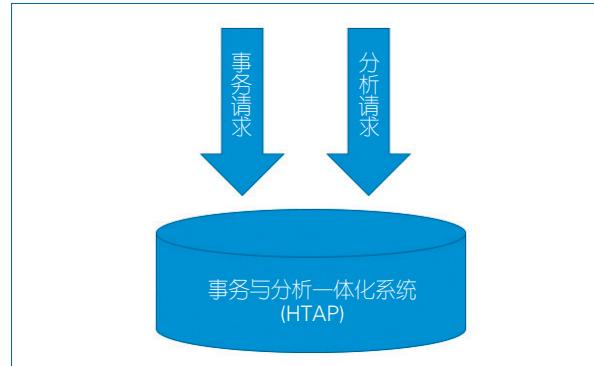


图2 一体化HTAP系统

现有 HTAP 系统的主要架构有三种方式：

- **基于主备复制的系统架构。** 主节点承担 OLTP 负载，备节点用于分析处理。通过数据 / 日志复制技术，将 OLTP 节点和 OLAP 节点进行简单耦合，从而达到同时进行事务处理和数据分析的目的。为了维护这样的混合架构，通常做法是将 OLTP 系统中的操作日志复制到 OLAP 节点。这样的解决方案从某些角度来讲可以被称为 HTAP，但是这样简单耦合的设计存在数据滞后性的问题，分析查询的数据不一定是当前 OLTP 系统中的最新数据，这违背了快速商业决策要求数据新鲜的原则。

- **基于混合存储的系统架构。** 有很多研究抛弃传统关系数据库所采用的单一数据组织方式，采用行列混合的数据存储方式。比如，SAP HANA<sup>3</sup> 在内存中采用行存储，优化在线事务处理性能；大量位于磁盘的历史数据，采用按列格式的方式组织，这样能够快速返回查询引擎所需要的大量数据。在行格式的事务数据逐渐冷却之后，将它们转为列存储格式。

- **两套引擎共享存储。** 还有一些对于 HTAP

<sup>3</sup> SAP HANA 是一款支持企业预置型部署和云部署模式的内存计算平台，提供高性能的数据查询功能，用户可以直接对大量实时业务数据进行查询和分析，而不需要对业务数据进行建模、聚合等。

有效的解决方案是，事务处理和分析处理使用的都是同一存储中的同一种格式的数据。但是为了同时高效地支撑 OLTP 和 OLAP 的应用场景，他们使用不同的引擎分别对两种需求提供支撑。SAP HANA Vora<sup>[9]</sup> 就是这样一个例子，在 HANA Vora 中，事务处理通过 HANA 执行，而分析请求由 Spark SQL 处理，子查询下推到数据库中。

## 可扩展的并发控制技术

普通 PC 服务器的内存容量可以达到 TB 级别，许多应用的数据集可以全部存储在内存中。在内存数据库环境中，传统数据库中的磁盘 I/O 对性能瓶颈的限制已经不复存在。研究发现，在大内存多内核架构的现代服务器中，传统的并发控制技术已经成为内存数据库系统的性能瓶颈。许多研究致力于降低并发控制协议的代价，还有一些优化方案用于减少在内存中数据访问引起的竞争。最近几年发布的数据库系统，如 Hekaton、HANA 和 Hyper 等，已经针对多核多处理器的大内存计算机进行了许多优化。与传统数据库中使用的两阶段锁的并发控制 (Two-Phase Locking, 2PL) 相比，乐观并发控制 (Optimistic Concurrency Control, OCC) 可以使事务执行时避免锁开销，同时在验证阶段花费较小的代价对事务进行冲突检测。在多核大内存环境下，乐观并发控制被人们重新重视并得到广泛应用。

乐观并发控制通常将事务执行划分为读、验证、写三个阶段。在读阶段，事务直接从存储中读数据，并保存到事务执行上下文的读集合中；对于要写入的数据，也是先保存到事务上下文的写集合中。因此，事务在读阶段乐观地执行读写操作而不会阻塞。在验证阶段，乐观并发控制要对读写的数据版本进行验证，以保证事务执行符合可串行化的调度。在写阶段，如果验证通过，证明事务的读写没有被修改，则表明事务可以成功提交。在提交时，事务将本地的写操作数据写入到数据库存储中，此时这些写入的数据将对其他事务可见。

在现代多核的配置环境下，乐观并发控制在保证可串行化隔离级别时，仍然面临着在不同场景中的多核扩展性问题，许多工作针对不同问题提出了他们的解决方法。Silo<sup>[13]</sup> 是一个内存环境下单版本的乐观并发控制协议，在进入验证阶段之前，乐观并发控制通常需要使用原子操作生成事务的提交时间戳，将该时间戳作为事务串行化顺序，而 Silo 采用了一种基于时间的纪元 (epoch) 方法来避免集中式地提交时间戳的分配。对于范围扫描请求，Silo 提出了一种轻量级的幻读检测方法，即通过检查 B 树的叶子节点的父节点版本，判断该分支下是否发生过删除或插入。但是，对于扫描记录的版本验证，Silo 仍然需要在执行阶段将所有扫描的数据保存到读集合，在验证阶段重新读取扫描范围内的数据，与读集合内的数据进行逐个验证。

为了优化乐观并发控制在高冲突场景下回滚率高、性能不佳的问题，MOCC<sup>[14]</sup> 结合锁机制，提出了一种大部分时间使用乐观并发控制而在特殊情况下使用锁的并发控制机制，以避免事务由于热点数据导致的高频率回滚。乐观并发控制如果在验证阶段检测到了与其他事务有冲突，则会自行回滚。BCC<sup>[15]</sup> 通过在验证阶段检测一种事务依赖模式，进一步精确判断该冲突是否违反了可串行化的要求，减少了事务的不必要的回滚，因而在高冲突的环境下提高了事务处理的性能。Wu<sup>[16]</sup> 提出了一种事务的重试机制来代替暴力回滚，在事务发生冲突时，事务通过重新执行有冲突的部分进行自我“治疗”，减少了事务回滚带来的额外开销。

最初的乐观并发控制基于实际并发执行的事务之间的冲突很少，大部分事务可以在短时间内成功提交。然而这样的假设在短更新 (OLTP 负载) 和较长的范围扫描 (OLAP 负载) 相结合的 I-HTAP 应用场景下并不成立。由于乐观并发控制需要在验证阶段对读取的所有数据版本进行验证，因此验证的代价与事务访问的数据数量成线性关系。对于长的扫描请求，可能会需要很长的时间才能够提交事务，从而增加了与其他事务冲突的概率，最终很有可能导致事务回滚。

## 面向HTAP的并发控制

在OLTP和范围查询混合类型的负载下，事务的并发控制通常无法提供高效的可串行化的隔离级别，其主要因素在于对范围查询的请求处理。基于锁的并发控制方法设计了多粒度的锁结构，通过在数据项、页、表等结构上加锁来避免事务读取到的数据被修改，同时防止在读取的范围内有幻读现象发生。而在内存作为主要存储的场景下，乐观的并发控制虽然比悲观的基于锁的并发控制允许更多的事务并发执行，但是对于HTAP的混合负载，仍然不能够有效处理。

### 现有乐观并发控制无法有效处理HTAP负载

在事务验证阶段，乐观并发控制方法主要通过本地读集合验证和全局写集合验证两种方式来验证范围扫描请求的数据是否发生改变。我们设计了试验，并对这两个方法进行了对比。

**本地读集合验证** (Local Readset Validation, LRV) 是在读阶段将读取到的数据保存到读集合中，对扫描过的数据进行重新读取，并逐一与读集合保存的记录版本进行比较。如果全部一致，则证明从事务读的时刻开始到进入验证阶段之前，没有其他事务与该事务有读写冲突，因此事务可以正确提交；否则，事务与其他并发执行的事务有访问冲突，无法判定事务满足可串行化的要求，因此，为了正确性，事务必须回滚。即使在内存作为主存储的环境下，这种方式也需要重新读取读集合中的数据，在有密集的扫描请求的负载下，这种方式效率低下。Silo采用了这种方式进行验证，主要优化在OLTP负载下的性能瓶颈问题，而对范围查询负载没有做更多的优化。

**全局写集合验证** (Global Writeset Validation, GWV) 维护一个全局的写集合，所有已提交的事务的写操作都会添加到这个集合中。事务在读阶段将自己的读操作和扫描操作保存为读谓词；在验证阶段，事务将读取的每个谓词与全局写集合取交集，

如果有与自己并发的事务的写和读谓词有交集，则说明与该事务有读写冲突，需要回滚。如果所有的读谓词都验证成功，则事务可以成功提交。这种方式被 HyPer<sup>[17]</sup> 数据库采用。

对比这两种方法，在读密集型负载下，全局写集合验证只需要访问相对较少的写记录集合，而本地读集合验证需要验证较多的读记录。但是在多核场景下，全局写集合验证存在扩展性问题，并且在OLTP的负载中，全局写集合验证需要维护大量并发的写操作，其性能无法满足OLTP负载的需求。

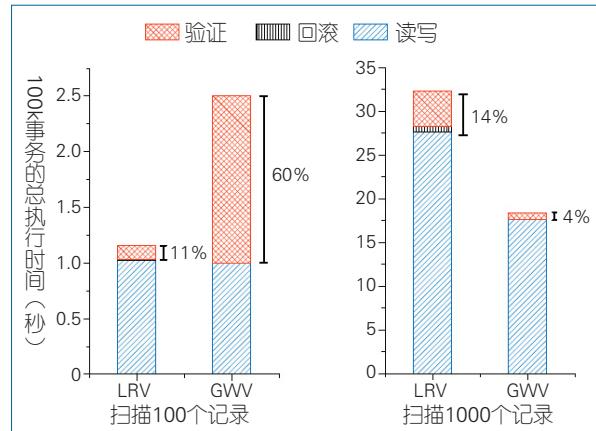


图3 本地读集合验证和全局写集合验证事务执行时间对比

我们设计了一组试验，将全局写集合验证和本地读集合验证在混合YCSB<sup>[22]</sup>负载下进行性能对比，见图3。YCSB表预先导入了10万条记录，实验使用的混合YCSB负载由90%的更新事务和10%的扫描事务组成。每个更新事务包含5个更新操作，每个扫描事务包含4个更新操作和一个扫描请求。左图中的扫描长度为100个记录，右图中的扫描长度为1000个记录。更新操作的主键值和扫描操作第一个记录的主键值是根据同一个Zipfian分布随机生成的，试验生成的混合YCSB负载是低冲突的。在试验中，我们统计了事务执行读写、验证和回滚三阶段的时间。其中读写阶段的时间是所有提交事务的读阶段和写阶段的时间总和，其主要CPU开销是从内存中读写数据，是提交事务必须花费的；验证阶段的时间是所有提交事务的验证时间总和，是并

发控制的额外开销；回滚时间是所有回滚事务的执行时间总和，反映了负载冲突率对并发控制的影响。对于每个试验，我们执行了 10 万个事务，并统计各部分最终的执行时间。

在扫描长度为 100 个记录的试验中，全局写集合验证的验证阶段几乎占据了全部执行时间的 60%（图 3 左图）。这些代价主要是由验证全局写集合中的记录造成的，因此当扫描长度为 100 时，负载中的写操作相对较为密集，这导致扫描事务需要验证大量其他并发事务的写操作，而本地读集合只花费了总执行时间的 11% 来验证。在图 3 右图的试验中，扫描长度变为 1000，负载中的读操作变得密集，全局写集合验证方法的验证代价仅为 4%，而本地读集合验证代价增加到 14%，其原因是，在扫描长度为 1000 时，本地读集合的验证需要在读阶段维护 1000 个扫描记录的版本，而在提交阶段需要重新读取这 1000 个记录进行验证，这消耗了太多时间。此试验说明，在混合类型的负载下，本地读集合的验证和全局写集合验证都存在一定的性能问题。

## 基于键值逻辑范围的乐观并发控制

为了支持可串行化的键值范围负载，微软的 Deuteronomy<sup>[18]</sup> 数据库系统采用键值逻辑范围粒度的并发控制机制。依据数据库中的主键取值范围与数据分布，逻辑范围对象将主键的取值范围切分为一个个连续的数据区间，每个逻辑范围对应一个数据分区，分区之间是连续而不相交的。通常做法是按照等深的方法构造键值逻辑范围，即每个逻辑范围对应的记录数目相同。

- **范围锁** Deuteronomy 数据库采用数据存储和事务处理分离的架构，事务处理节点只有最近提交的部分数据。为了保证事务的可串行化调度，针对包含范围扫描请求的负载，Deuteronomy 采用了一种基于时间戳排序 (timestamp ordering) 的多版本范围并发控制方法 (Multi-Version Range Concurrency Control, MVRCC)，借助逻辑分区对事务的读写冲突检测。

- **范围乐观并发控制 (Range Optimistic**

**Concurrency Control, ROCC)**<sup>[23]</sup> ROCC 的基本思想是将数据库中的数据按照主键划分逻辑范围，在每个逻辑范围内使用一个无锁的循环数组来追踪事务在这个范围内的修改。在验证阶段，事务需要在扫描过的范围内进行冲突检查，验证是否存在其他事务修改、扫描的数据，以及是否在没有扫描的范围内执行插入和删除操作。

与 ROCC 不同的是，MVRCC 的冲突检测等同于在范围粒度上的锁机制。在扫描一个分区之前，MVRCC 将会更新该分区的 Last-Read 时间戳，相当于对该分区添加一个读锁。在修改一个分区内的某条记录时，MVRCC 在该逻辑分区上追加一个新的版本，该版本记录了这个事务的修改意向，等同于意向排他锁。ROCC 也使用逻辑分区的结构，与 MVRCC 不同的是，ROCC 使用了乐观验证的方法。ROCC 有两个优势：第一个是在冲突检测时，对于一个逻辑分区，MVRCC 需要检查的事务数量多于 ROCC。MVRCC 采用按事务开始时间戳排序的提交顺序，然而在每个具体的逻辑分区中，事务并没有按照时间顺序排列。因此，在验证时 MVRCC 需要对事务队列中的全部事务进行检查。而 ROCC 在执行阶段记录扫描的逻辑分区的版本，在验证时重新获取一次逻辑分区的版本，通过这两个版本能够精确找出逻辑分区内的交叉事务，不会验证其他版本的事务。第二个优势是，ROCC 对于扫描边缘逻辑分区内的读写冲突检测更加精确。由于扫描请求范围可以跨越多个逻辑分区，而边缘的分区（即扫描的第一个分区和扫描的最后一个分区）一般不是被完全覆盖的。ROCC 通过谓词记录扫描的开始和结束，在边缘分区精确地描述了读取的数据范围。在验证阶段，ROCC 对于边缘分区内的事务的写集合进行逐一检查，如果有写操作与读谓词冲突，事务才会回滚。因此，ROCC 的读写冲突检测更加精确，减少了事务的错误回滚 (false abort)。

## 未来的研究方向

随着内存、固态硬盘、硬盘驱动器容量的不断

扩大，以及机器配备核数的不断增多，取决于实际应用需求，HTAP 数据库产品既有采用单节点集中式架构，也有采用 share-nothing 的分布式架构。为了实现真正的 HTAP 处理目标，我们认为如下几个方向值得进一步深入研究。

**自适应并发控制。** HTAP 事务中的分析型查询经常需要获取其他系统产生的原始数据，HTAP 系统需要同时处理多种类型的负载。在实践中，同一个数据库系统可能同时处理传统事务、HTAP、报表分析、实时 / 批量 ETL 操作等混合型负载。单一的并发控制方法在现实场景下并不能总是取得最佳性能，HTAP 的事务处理需要结合事务本身包含的操作类型、访问的数据量大小、当前负载的并发度以及读写特征，自动选择代价最低的并发控制策略。

**事务数据可用于实时分析的时间差最小化。**在同一节点进行事务和分析处理势必造成两种负载相互影响，违背了隔离事务与分析处理性能影响的设计目标。高可用的事务处理要求事务日志同步到备节点后，才能在主节点进行事务提交，并响应客户端。一个策略是在备节点进行分析处理。这里的关键问题是如何缩小事务数据在主节点提交后，到该事务数据在备节点可见的时间差，以及分析型查询如何找到拥有最新数据的备节点。

**事务数据与大规模历史数据的高效合并。**事务与分析一体化架构针对事务处理负载和分析处理负载分别采用行存和列存两种存储模式，目标是使得事务和分析处理同时获得最佳性能。在混合存储架构下，内存中的事务热数据（最近几个小时或 1 天产生的事务数据），能够保证事务在主节点提交之后，备节点立即进行实时分析。然而，这些事务型数据需要定期合并到大规模的历史 / 冷数据中（基于列存，位于磁盘 / 固态硬盘），腾出内存空间存放后续的事务数据，这一过程简称为数据合并。整个合并耗时取决于最慢的数据存储节点。过长的合并时间将导致内存中的“旧”事务数据无法及时迁出内存，影响当前事务处理的性能。 ■



张融荣

华东师范大学博士生。主要研究方向为分布式数据库。



蔡 鹏

华东师范大学副教授。主要研究方向为内存事务处理以及基于机器学习技术的自主数据管理系统。

pengcai2010@gmail.com



钱卫宁

CCF 专业会员、CCF 数据库专委常务委员。华东师范大学教授、博士生导师，数据科学与工程学院院长。主要研究方向为面向互联网级应用的数据管理系统、可扩展事务处理等。wnqian@dase.ecnu.edu.cn

## 参考文献

- [1] Attaluri G, Attaluri G, Liu S, et al. DB2 with BLU acceleration: so much more than just a column store[J]. *Proceedings of the Vldb Endowment*, 2013, 6(11):1080-1091.
- [2] Lamb A, Fuller M, Varadarajan R, et al. The Vertica Analytic Database: C-Store 7 Years Later[J]. *Computer Science*, 2012, 5(12):1790-1801.
- [3] Stonebraker M, Weisberg A. The voltDB main memory DBMS [J]. *IEEE Data Eng Bull*, 2013, 36(2): 21-27.
- [4] Diaconu C, Freedman C, Ismert E, et al. Hekaton: SQL Server’s Memory-Optimized OLTP Engine[C]// *ACM Sigmod International Conference on Management of Data*. ACM, 2013:1243-1254.
- [5] MemSQL. <http://www.memsql.com/>, 2018.
- [6] Tu S, Zheng W, Kohler E, et al. Speedy transactions in multicore in-memory databases[C]// *Twenty-Fourth ACM Symposium on Operating Systems Principles*. ACM, 2013:18-32.
- [7] Peter Boncz, Marcin Zukowski, Niels Nes. MonetDB/X100: Hyper-Pipelining Query Execution[C]// *Proceedings of the 2005 CIDR Conference*, 2005: 225-237.
- [8] Färber F, May N, Lehner W and et al. The SAP HANA Database – An Architecture Overview[J]. *IEEE DEBull*, 2012, 35(1):28-33.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>

# 从2018年的戈登·贝尔奖说起

关键词：超级计算机 戈登·贝尔奖

郑纬民 薛巍 陈文光 等  
清华大学

2018年11月16日，第三十届全球超级计算大会(SC18)在美国达拉斯落幕，美国两大国家实验室(美国能源部下属的橡树岭国家实验室和劳伦斯·利弗莫尔国家实验室)的新一代基于图形加速器的异构超级计算机位列TOP500<sup>[1]</sup>前两名，首个每秒百亿亿次(ExaOps)计算能力的基因组学计算应用与最大规模的深度学习应用双双摘得2018年戈登·贝尔奖(ACM Gordon Bell Prize)。

## 2018年顶级超级计算系统发展

表1中列出了2018年排名(2018年11月发布)TOP500榜单中TOP10超级计算机的具体信息，

包括该机器在2017年的排名(2017年11月发布)、峰值性能、LINPACK效率、HPCG排名及其峰值效率、所采用的架构和加速器类型。从表中的数据可以看出：

- 2018年TOP10超级计算机快速更新。与2017年相比，TOP10中增加了4个新系统，分别是Summit<sup>[2]</sup>、Sierra<sup>[2]</sup>、AI Bridge Cloud Infrastructure(简称ABCI)和SuperMUC-NG；同时，3个2017年的TOP10系统得到进一步升级继续保持在TOP10之列。这也反映了目前应用对更高计算能力的迫切需求。
- 新一代基于图形加速器的异构超级计算机(Summit, Sierra, Piz Daint)在LINPACK效率方面较

表1 2018年排名前10的超级计算机

TOP500 2018	TOP500 2017	机器名称	峰值 (PFLOPS)	HPL 效率	HPCG 2018	HPCG 峰值效率(%)	架构	所采用的加速器类型
1		Summit	200.8	71.5%	1	1.5%	加速器异构	NVIDIA Volta GV100
2		Sierra	125.7	75.3%	2	1.4%	加速器异构	NVIDIA Volta GV100
3	1	神威·太湖之光	125.4	74.2%	7	0.4%	异构众核	/
4	2*	天河2A	100.7	61.0%			加速器异构	Matrix-2000
5	3*	Piz Daint	27.2	77.9%	6	1.8%	加速器异构	NVIDIA Tesla P100
6	7*	Trinity	41.5	48.7%	4	1.3%	同构众核	/
7		AI Bridge Cloud Infrastructure	32.6	61.0%	5	1.6%	加速器异构	NVIDIA Tesla V100 SXM2
8		SuperMUC-NG	26.9	72.5%	15	0.8%	同构多核	/
9	4	Titan	27.1	64.9%	13	1.2%	加速器异构	NVIDIA K20x
10	5	Sequoia	20.1	85.6%	12	1.6%	同构多核	/

\*该系统在2018年升级

上一代系统 Titan 有显著提升，LINPACK 效率均超过 70%。当前 TOP10 系统当中，LINPACK 效率能与其媲美的仅有神威·太湖之光和两个同构多核系统 SuperMUC-NG 和 Sequoia。

- 新一代顶级计算系统在稀疏问题求解效率方面有所进步。得益于持续的软硬件创新，TOP500 性能最高的 Summit 和 Sierra 两个系统在 HPCG 测试中的效率并没有因大量使用图形加速硬件而明显下降，与新建的两个同构系统 Trinity 和 SuperMUC-NG 相比，其 HPCG 效率反而更高。最终，这两个新系统也在 HPCG 总性能的排名中取得前两名，超越之前多年在 HPCG 测试中领跑的京系统。这也使得大家对传统科学计算应用有效利用加速器异构架构有了更多信心。

- 基于图形加速器的异构架构在最新 TOP10 系统中已占据 5 席，成为目前构建顶级超算系统的主要方式之一。与前一代图形加速器异构系统 Titan 相比，美国国家实验室采用的新一代异构系统采用了高密度节点（一机多 GPU 卡）的设计，在功耗和节点数量方面有着明显优势，NVLink 技术的引入缓解了节点内多个计算设备间数据传输能力不足的困境。

## 2018 年戈登·贝尔奖入围工作

一年一度颁发的戈登·贝尔奖用于表彰世界范围内高性能计算的杰出成就，尤其是高性能计算应用于科学、工程和大规模数据分析领域的创新工作<sup>[3]</sup>。2018 年入围最终评奖的工作一共有 6 项，分别来自不同领域，既有传统的科学计算应用，也有新兴的深度学习和图计算应用。其中，有 5 项均以 TOP500 最高性能机器 Summit 作为测试和优化平台，仅有 1 项围绕我国的神威·太湖之光展开。

### 持续峰值性能奖：应用超级计算机来应对药物流行病<sup>[4]</sup>

本次获得戈登·贝尔奖持续峰值性能奖的应用来自基因组学计算领域，由橡树岭国家实验室的研

究团队领衔完成，是首个达到每秒钟百亿亿次计算能力的科学计算应用。该工作在橡树岭国家实验室的 Summit 上采用混合精度计算模式，峰值性能达到 2.3 ExaOps。该研究是高性能计算与生物信息学、医学相结合的一个典范。

药物滥用是世界级难题。根据美国疾病控制和预防中心的统计，每天有 115 人死于阿片类（类鸦片）药物过度使用，而该问题还在持续恶化中。如何通过基因组学的分析来寻找致病原因并积极研制治疗方案迫在眉睫。橡树岭国家实验室的研究团队利用超级计算机 Summit 和 Titan，开发了用于大规模上位基因组全关联研究和多效性研究的高性能计算工具 CoMet。研究人员希望利用该工具，基于美国百万退伍军人计划收集的大量基因数据，通过对遗传过程的理解，尝试帮助解决长期慢性疼痛和阿片类药物上瘾问题。

CoMet 的核心操作是特定的相似性计算与相关系数计算，以向量比较操作为核心。该工作通过将向量比较操作转化为负载均衡的大规模分布式稠密线性代数操作，达到最佳利用异构计算资源的目的。同时，针对大规模和采用高密度节点的并行环境，该团队还设计了“计算 - 节点内传输 - 通信重叠”的计算通信模式，最终在 Summit 上取得了 98% 的弱可扩展性测试结果，相关系数计算达到 2.3 ExaOps 的计算性能，与以往工作相比取得了 4~5 个数量级的性能提升，实现了每秒接近 30 亿亿个比较计算。尽管 CoMet 达到了每秒 E 级计算次数，但仍不能算是传统意义上的仅考虑双精度浮点运算性能的 E 级应用，其性能指标计算所采用的总操作数同时包括了整型运算指令、浮点运算指令、位操作指令以及精度转换指令。

### 可扩展性与时效奖：利用高可扩展深度学习方法理解极端天气事件<sup>[5]</sup>

本次获得戈登·贝尔奖可扩展性与时效奖的应用来自气候变化研究领域，由劳伦斯伯克利国家实验室和 NVIDIA 公司的联合研究团队完成，是首个可以有效扩展到 27360 块 GPU 加速卡的深

度学习应用，其半精度计算的峰值性能也达到每秒钟百亿亿次，成功提升了气候变化研究人员在高分辨率气候模拟数据集中有效识别极端天气模式的能力。

科学大数据分析，特别是针对高分辨率海量科学仿真数据的分析，是科学计算与人工智能技术相结合的机会，也是超算的核心应用之一。该研究团队尝试解决大规模深度神经网络训练过程中的共性计算问题，提出了面向异构超算上大规模深度学习训练过程的整体优化方案，重点解决了已有 Tensorflow 系统在 IO、数据载入和通信优化等方面适应异构超算过程中出现的性能问题，以及现有深度学习算法在可扩展性上面临的困难。

利用 Summit 全机系统，该团队借鉴图像识别最新深度学习网络，完成了高分辨率大气模式（全球 25km 水平分辨率）输出结果（数十 TB）的深度神经网络高效训练，半精度计算持续性能接近每秒钟百亿亿次，可有效识别极端天气事件空间结构。值得指出的是，该工作采用数据并行思想，使用更多的计算资源导致更大的批尺寸 (batch size)，大规模并行时可获得的收敛速度和精度成为关键。遗憾的是，可能是由于机时受限，获奖论文仅给出了 1024 节点（每节点 1500 文件）的收敛性分析结果。

## 国内团队入围应用：“神图”图计算框架<sup>[6]</sup>

此次入围戈登·贝尔奖唯一来自中国的应用是名为“神图”的图数据分析编程系统。“神图”所面向的对象不是传统的科学与工程计算应用，而是探索了在超级计算机上如何开展极大规模图数据的高效处理。图数据将数据抽象成点和边的数据形式，是一种典型的非结构化数据。图数据分析是大数据分析中的重要内容，在金融反欺诈、物联网管理、信息安全、网页搜索、社交网络分析、电网分析等领域具有广泛的应用前景。

“神图”基于神威·太湖之光超级计算机，针对极大规模随机通信、图结构的分布不均衡以及异

构结点功能映射等问题，提出了中继消息聚合与路由、分化消息传播技术以及无锁数据分发技术等方法，有效利用了神威·太湖之光全机的处理能力和通信能力，能够高效扩展到全机千万核规模，在国际上首次实现了对包含 4 万亿个结点、70 万亿条边的合成图的快速分析，每一轮 PageRank 算法的时间只需要半分钟。在应用方面，对于搜狗公司提供的 12 万亿条边的真实中文网页图，“神图”完成一轮 PageRank 算法仅需 8.5 秒。与文献中报道的业界最先进系统相比，处理规模增加了一个数量级，同时，处理性能提高了超过一个数量级，实现了图计算节点规模、图数据规模、运行时间上的突破。

“神图”系统并不是一个特定的应用程序，而是一个编程框架，为用户在神威·太湖之光超级计算机上编写多种图计算应用提供了极大的便利。一个基于“神图”的图分析算法通常只需要数十行代码即可完成原先需要编写近万行代码才能实现的图数据处理功能，大大提高了开发效率。

此项工作的研究单位是清华大学、北京费马科技有限公司、卡塔尔计算研究所、数学工程与先进计算国家重点实验室、苏黎世联邦理工学院、国家并行计算机工程技术研究中心、北京搜狗科技发展有限公司和国家超级计算无锡中心等。

## 其他入围应用

另外三个戈登·贝尔奖入围应用也各具特色，分别在量子色动力学模拟、城市震害模拟和电子显微镜数据处理的高效异构计算上取得了突破。

量子色动力学模拟项目<sup>[7]</sup>属于传统的科学计算领域，其性能优化对 Summit 和 Sierra 的高密度节点配置做了深入考虑，主要关注三方面问题：(1) 高密度节点的 CPU 开销最小化。CPU 资源是高密度节点处理能力的短板，因而节点内数据传输和网络通信尽可能少占用 CPU 资源成为提升应用性能的关键之一；(2) 通信策略和通信参数自动调优，高密度节点配置使得节点内不同设备间数据传输存在新的优化空间；(3) 并发任务的最优调度，发掘多任务在同一节点内并发执行的可能性。上述工作均基于已

有软件框架进行拓展实现，与图形加速器异构架构性能优化方面的长期投入与积累密不可分。该工作也从侧面说明了从关注单一应用到关注整个应用工作流的重要性和必要性。

城市震害模拟项目<sup>[8]</sup>关注的是非结构有限元计算中的大规模线性方程组求解问题。与已有工作不同的是，该工作在预处理中对难收敛区域引入局部操作来加速收敛，并引入混合精度计算来充分发挥 GPU 计算资源效能。对于时变计算问题中难收敛区域的识别更是采用了人工神经网络来预测，这也是人工智能技术在传统计算中应用的一个有趣实例。

电子显微镜数据处理项目<sup>[9]</sup>同样关注深度学习的大规模训练问题。其不同之处是采用了类似 AutoML（谷歌在 2017 年创建的一个能够制造神经网络的 AI 系统）的思想，将深度神经网络结构优选和超参调优合并来考虑。由于利用遗传算法来进行网络优选，在传统的数据并行和模型并行的基础上开发了新的并行机会，可以更好地利用超大规模计算系统。该项目也有望推动实现全自动训练，并最终大幅加速训练过程和有效改进训练效果。

## 超级计算机与应用发展趋势

2018 年超算领域在体系结构和应用方面的进步可圈可点，呈现两个重要趋势，一是异构架构在超算系统构建层面被广泛接受，二是人工智能应用有望成为超算的主流应用之一。

### 异构架构在超算系统构建层面被广泛接受

目前的 TOP10 系统中异构超算占据七成。其中，NVIDIA GPU 构建的异构超算系统占 5 席。而且，随着人工智能技术在科学与工程计算中越来越广泛的应用，支持高性能张量计算的图形加速器硬件还可能越来越多地受到超算中心决策者的青睐。特别需要指出的是，该架构最初作为应对功耗墙挑战的一种可行方案起步，其构建的超算系统历经 TSUBAME

系列机、天河 1A、Titan 到如今的 Summit 已经有 10 年时间，从最初在应用移植优化方面饱受质疑到现在已经初步兼具功耗和性能优势，体现了高性能计算社区，特别是美国在这个方向上坚持投入，在应用、算法、软件和硬件等方面持续协同创新的成果<sup>[10]</sup>。从事核能研究为主，关注传统科学计算应用的美国劳伦斯·利佛莫尔国家实验室从上一代同构系统 Sequoia 转而采购图形加速器异构系统 Sierra，也是对这一成果的重要认可。

在 GPU 加速器之外，异构系统天河 2A 选用 Matrix-2000 作为加速计算设备，NEC 的 SX-AURORA TSUBASA<sup>[11]</sup>选用向量处理单元作为加速器，相关技术的未来走向值得关注。

在同一芯片中集成不同计算核心的异构众核架构同样值得期待。该架构已经在我国的神威·太湖之光系统中实现并被证明有效。美国 CORAL 计划(the Collaboration of Oak Ridge, Argonne and Livermore)所支持的第一台 E 级系统 Aurora（预计 2021 年前后建成）也将采用类似架构。

目前看来，异构架构已经成为构建顶级超算系统的大势所趋，加速器异构还是异构众核，争论仍将继续。

### 人工智能应用有望成为超算的主流应用之一

算力一直被认为是人工智能再次起飞的重要基础之一。随着深度神经网络规模的扩大，最新的网络生成和训练往往需要数万 GPU 小时（如 BERT, NASNet 等）甚至更多。具有顶级计算能力的超算系统理应为大规模人工智能应用提供助力，不断拓展后者的技术边界。2018 年的戈登·贝尔奖选择大规模深度学习应用，入围应用中人工智能相关的项目也前所未有地占据了半壁江山，这一切都预示着人工智能与超算的结合将愈来愈紧密。

目前真正具有高可扩展能力的人工智能算法与应用并不多。以应用最为广泛的深度学习为例，增大批尺寸来提升数据并行性有可能导致收敛问题，从而限制可利用的并行资源总量，而模型并行又存

在通信瓶颈。因此，人工智能应用仍需要持续创新以更好地利用未来更大规模的超算系统。而最新出现的进化神经网络方法就是在该方向上努力的成果之一，其相比现有的深度强化学习方法更具扩展性，也更具充分利用超算资源的潜力。美国橡树岭国家实验室基于异构超算系统开发出的多节点深度学习进化神经网络 (Multi-node Evolutionary Neural Networks for Deep Learning, MENNDL) 能够通过遗传算法进行网络拓扑与超参数优化，以自动生成优化神经网络，上文提到入围今年戈登·贝尔奖的电子显微镜数据处理项目<sup>[9]</sup> 就是 MENNDL 的一个应用示例。

可以预见，更好地将人工智能技术与已有科学规律结合，创新科学发现方法和科学计算模型，会为构建未来超算应用和研发高可扩展并行系统创造新的契机。

## 总结与展望

极大规模超级计算机的研制必须要回答的问题是：研制这么大的超级计算机有什么用处？更大规模的计算机是否能够通过更大内存和更强的计算能力解决关键应用挑战，完成从“不能”到“能”的突破，从而为社会提供与其投入相匹配的回报？戈登·贝尔奖的设置为我们回答这个问题提供了一个窗口，可以看到在传统的基于数值模拟的科学计算应用之外，本次入围应用在人工智能、大数据处理等方面探索了新的方向。

我国在过去的三届戈登·贝尔奖评选中，以神威·太湖之光计算机为载体的应用共获得 6 项入围，2 项得奖的好成绩，说明我国在超算系统研制取得突破的基础上，我国的高性能计算算法和应用研究者有能力开展世界最前沿的研究工作。

更进一步，我们认为获奖本身不是目的，我们希望能够超越戈登·贝尔奖，以科技进步和民生服务为目标，推动高性能计算应用和系统研发的广泛使用。例如我国获奖的高精度天气预报算法，能否尽快转化为实用的数值天气预报应用，

提高天气预报的时效性和准确性，为减灾防灾做出切实的贡献？我们在此呼吁，科技部门应对相关领域给予长期稳定的支持，促进应用研发与系统研制的交互引领，形成正反馈机制，促进我国的超级计算领域健康持续发展。 ■



郑纬民

CCF会士、CCF前理事长。清华大学教授。主要研究方向为并行/分布处理、网络存储器等。

zwm-dcs@tsinghua.edu.cn



薛 巍

CCF高级会员、信息存储技术专委会委员。清华大学副教授。主要研究方向为大规模科学计算。

xuewei@tsinghua.edu.cn



陈文光

CCF副秘书长、理事。清华大学教授，兼任青海大学计算机系主任。主要研究方向为并行计算的编程模型、并行化编译和应用分析。

cwg@tsinghua.edu.cn

其他作者：张悠慧

## 参考文献

- [1] <https://www.top500.org>, 2018.11.
- [2] Sudharshan S Vazhkudai, Bronis R de Supinski, Arthur S Bland, et al. The design, deployment, and evaluation of the CORAL pre-exascale systems[C]//The International Conference for High Performance Computing, Networking, Storage, and Analysis. 2018.
- [3] Gordon Bell, David Bailey, Alan H. Karp, et al. A look back on 30 years of the Gordon Bell Prize[J]. International Journal of High Performance Computing Applications, 2017, 31(6): 469-484.
- [4] Joubert W, Weighill D, Kainer D, et al. Attacking the opioid epidemic: Determining the epistatic and pleiotropic genetic architectures for chronic pain and opioid addiction[C]// The International Conference for High Performance Computing, Networking, Storage, and Analysis. 2018.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>

# 大数据共享及交易中的机遇和挑战

关键词：数据交易 数据共享 安全计算 隐私保护

李向阳 张 兰 韩 风 等  
中国科学技术大学

## 大数据交易共享现状

人工智能和大数据科学技术的飞速发展在揭示数据本身的属性和规律的同时，也为自然科学和社会科学提供了新的方法，并将给数据的充分利用带来巨大价值。据统计，2015年全球大数据产业规模达到了1403亿美元，预计到2020年，将达到10270亿美元<sup>[1]</sup>。然而，在看到无限机遇的同时，我们不得不指出，当前开采的只是数据资源的冰山一角，网络空间中绝大部分数据还分散在一座座属于政府、机构、企业的数据孤岛，甚至是从未开采的数据荒岛。正如李克强总理2016年5月在全国推进简政放权放管结合优化服务改革电视电话会议上提到，“目前我国信息数据资源80%以上掌握在各级政府部门手里，‘深藏闺中’是极大的浪费”。由于数据是非独占性资源，复制成本低，且大数据具有一种稀有的属性——协同作用，即多个数据集作为一个整体的价值要大于各个数据集价值的简单相加，因此使这些隔离封闭的数据开放流通、融合应用能极大提升数据资源的利用价值，这也是大数据时代发展的趋势。

由于数据的潜在价值未知、数据所有者的自私性及对数据隐私安全的担忧等，数据所有者大多不愿免费公开/提供自己的数据。为克服上述困难，一种有效途径是将数据作为商品进行交易，数据所

有者通过公开/提供自己拥有的数据获得收益。

## 数据交易共享市场现状

随着数据交易和共享的重要性日益凸显，数据交易和共享平台的建设正在进入井喷期，包括Qlik、CitizenMe、Microsoft Azure Marketplace、DataExchange等国外平台，以及数据堂、数多多、iDataAPI和聚合数据等国内平台。此外，我国还成立了一系列政府指导的大数据交易机构，如贵阳大数据交易所和上海数据交易中心。这些平台上的交易内容包括数据集、网络爬虫、API、分析报告、解决方案等，覆盖的领域包括金融、商业、制造业、地理、交通、天气、电子商务、娱乐、电信、医疗保健、人工智能和各种个人数据（如社交媒体、地点和信用信息），交易形式包括交易已有的数据或订制数据。针对不同的交易内容和交易形式，其定价策略也不同。通常现成的数据集以固定价格出售，而订制数据的价格由卖方和买方协商确定；API的定价策略包括“按次支付”（例如，每次执行0.01元）和“批发”（例如，每千次执行10元）。这些平台上的数据展示根据交易内容而不同，包括元数据、统计信息、文本/视频说明、数据样本和API用法说明。

据不完全统计，2015年我国大数据交易的市场规模为33.85亿元，预计到2020年将达到545亿元<sup>[1]</sup>。虽然现有数据交易市场已具有一定规模，但对数据

交易的探索仍处于初期阶段。现有平台通常要求数据拥有者将数据及其描述和价格提交给平台，由平台代为出售。然而数据作为一种特殊的商品具有数量增长快、易复制、质量价值难衡量、权属难确定、渠道难管控、隐私安全风险高等特点，使得当前的数据市场中还存在数据质量保证、价格管控、数据隐私和版权保护等诸多关键问题亟须深入研究和解决。建立高效、可信、公平、安全的数据交易市场仍面临巨大挑战。

## 数据交易和共享相关政策与法规

通常来说，只要遵守当地的隐私法，各个国家都允许在公司之间进行数据共享/交易。此外，当用户明确同意或者数据来自公开网站时，销售用户数据也不存在法律问题。我国政府明确鼓励数据所有者和数据消费者之间共享数据，以充分发挥大数据的潜在价值，并推动技术创新和经济增长。2015年国务院印发了《促进大数据发展行动纲要》<sup>[2]</sup>，指出到2018年底前要建成国家政府数据统一开放平台，率先在信用、交通、医疗等重要领域实现公共数据资源合理适度向社会开放。2016年底工业和信息化部发布的《大数据产业发展规划（2016—2020年）》进一步分析了数据共享的问题（共享程度低，规范不健全），明确了开放共享的发展原则<sup>[3]</sup>。习近平总书记在2017年12月特别强调数据开放共享和融合作为国家大数据战略一部分的重要性，鼓励政府部门之间的数据共享，以及政府和私营公司之间的数据共享、交易。我国政府还鼓励数据交易，以扩大小数字生态系统。

目前还没有专门针对数据共享和交易行为的法规，但大多数国家都有数据和隐私保护的相关法律和机构来监管公司收集、使用和销售相关消费者数据的行为。面对大数据和人工智能产业对个人数据安全带来的挑战，近年来世界各国正尝试修订或增加法律法规中关于个人信息的保护范围及强化保护力度<sup>[4]</sup>。例如欧盟在2016年出台的《通用数据保护条例》(General Data Protection Regulation, GDPR)，旨在加强公司对用户数据使用的管理。在我国，2017年6月《网络安全法》和“侵犯公民个人信息

罪司法解释”生效，前者提供了一整套数据保护的规定，后者则明确了哪些侵犯个人数据的行为会构成犯罪。对于涉及知识产权的数据，如著作和发明专利等，各国都有相应法律赋予符合条件的著作作者、发明者在一定期限内享有独占权利，并对其所有权和使用权的转让做了详细规定。

## 数据交易模型

### 交易模型

数据交易模型分为两种：数据代理模型和P2P交易模型，如图1所示。

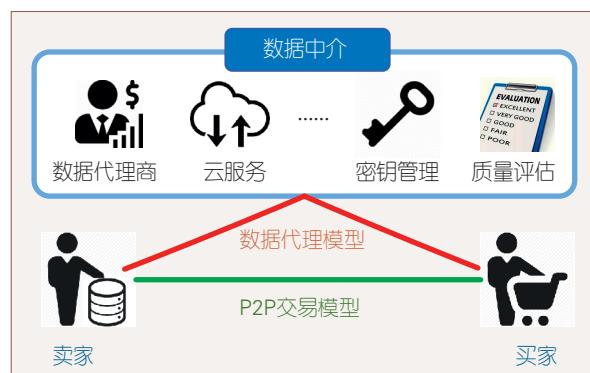


图1 数据交易模型

**数据代理模型：**在该模型中，数据代理商作为中间平台为买家和卖家提供交易数据的市场。交易平台由多个协作但非串通的实体组成，这些实体可以包括管理交易的数据代理商，提供存储服务的云，以及负责密钥管理、数据质量监控、异常检测、执法、税收的实体等。我们将所有这些中间机构作为一个整体称为“数据中介”。

**P2P交易模型：**在该模型中，买卖双方在没有数据代理商的情况下直接交易。其典型示例包括区块链和P2P文件共享网络，如Bit-torrent、eDonkey和Pruna等。

当前大部分数据交易平台都采用数据代理模型，而P2P交易模型由于具有低效率和不透明的特性，尚未成为主流。

## 交易内容

数据交易的内容通常可以分为4种：(1) **数据本身**：买家拥有对数据的永久/指定期限访问权，并可以在数据上执行任意计算以尽可能多地挖掘感兴趣的信息。(2) **数据的直接功能 API**：有时买家只对数据的某项简单功能感兴趣，例如搜索结果、统计信息或使用机器学习模型进行训练等。这种情况下，数据平台可以通过提供API来为买家提供相应功能，并限制其对数据的操作。(3) **数据分析结果**：是指从数据中挖掘出来的更高层次的有用信息。例如一个商家希望基于分析得到什么样的用户最可能是其潜在客户，而对原始数据并不感兴趣。(4) **数据衍生物**：与数据内容无关，而是数据的各种权利许可，例如订阅该数据的相关更新，或持续订阅不断产生的数据流，买断数据的所有权或排他的使用权，甚至一些基于区块链的证书（如基于可信飞行记录的飞行员证书）也可以进行交易。

涉及知识产权的数据交易通常是数据的各种权利许可，如作品的版权或专利的工业产权，根据法律规定其著作者或发明者在一定期限内享有独占权，并在其权利有效期内可以转让约定时间或地域范围内的所有权（购买者拥有独占权）或者使用权（分为排他和非排他两种），甚至其衍生品的商品化权（例如电影、动漫周边商品）。

## 交易形式

当前数据平台有两种主要的交易形式：(1) **交易现有的数据**：由卖家收集、处理数据，并向平台上传展示相关数据信息；买家在平台上检索以选择合适的数据。这种交易形式为卖家节省了订制数据的成本，但买家需要对数据进行进一步处理。(2) **订制数据**：由买家提出对数据内容和格式的要求；卖家根据买家需求整合筛选自己拥有的数据，并为

买家提供符合要求的订制数据。这个过程降低了数据交易的效率，但它提高了买家对数据的满意度。

## 数据定价和支付

交易过程中的关键一环是确定数据的价格，通常分三种情况：(1) **固定价格**：卖家对其数据的供求关系进行市场调查，然后设置固定的出售价格。(2) **变动的价格**：卖家动态决定其出售数据的价格，可以是价格随时间而变化，也可以因买家而异，或因订单顺序而异。(3) **限制出售量和独家买断**：为了增加数据的价值，卖家可以限制出售量，而买家也会独家买断数据以防止其他人获取数据。这种情况通常由买卖双方协商价格。在难以确定数据标价的情况下，拍卖则是一种常见的数据交易形式。

## 数据交易的流程

数据交易的流程可分为交易前、交易中、交易后三个阶段，每个阶段包含不同的操作和问题，如图2所示。

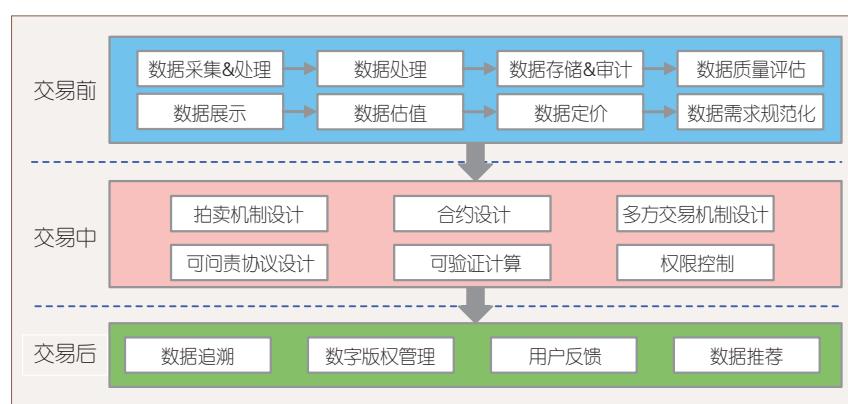


图2 数据交易的三个阶段

## 数据交易前的问题与挑战

在数据商品被交易前，卖家首先需要收集数据并进行清洗、标记、脱敏等处理，然后卖家需要将数据托管至云端以减少存储和通信开销，并委托代理商进行贩售。数据代理商则需要评估上传至平台的数据质量，并向买家以恰当的方式展示数据。对

于委托平台定价的数据商品，代理商还需要对数据进行估值并设计定价机制，给出合适的价格。买家为了更好地检索或定制目标数据，需要生成准确而规范的需求描述。同时买家也有评估数据质量和价值的需求，以确定自己的预算和出价。在数据交易前的准备阶段还存在以下关键问题和挑战。

## 数据审计

许多数据拥有者会使用云服务存储数据，但却对云存储服务不完全信任，因为云服务器可能会丢弃很少被访问的数据以降低维护成本。因此，云必须向用户证明数据是按照要求被正确存储的。这一类问题被称为可恢复性证明 (proof of retrievability)、存储证明 (proof of storage)、数据拥有证明 (provable data possession) 或存储审计 (storage auditing)。近年来也有一系列工作研究相关问题，如文献 [5, 6]。在数据交易场景中，数据代理商需要验证两方面的可恢复性：卖家是否在云端存储了声明的数据，以及数据是否可以实现声明的功能或分析结果。当数据加密存储时，代理商还须证明卖家对数据的拥有权（其可解密该数据）。由于数据代理商也并非完全可信，他无法直接访问云上的明文数据甚至是密文数据，这大大增加了可恢复性和拥有权证明的难度。此外，数据集庞大的体量使得传统的加密算法由于较高的计算和通信开销而变得不适用，研究者必须在准确度和计算复杂度之间进行权衡。

## 数据质量评估

为了提高用户的满意度和平台的声誉，数据代理商和买家需要对出售的数据集进行质量评估。高质量的数据本质上应该具有清晰的表示形式，能被轻松访问，并且适用于买家的任务<sup>[7]</sup>。数据质量包含但不限于以下方面：**(1) 内在质量**是指数据本身在数量、准确性、完整性、及时性、一致性、清洁度、安全性等方面的质量。**(2) 表达质量**侧重于与数据格式（简洁、一致的表示形式）和意义（易于解释）相关的方面。**(3) 可访问性质量**强调买家对数据易于获取或检索的程度，例如访问通信延时等。**(4) 上**

**下文质量**强调数据与应用场景的相关性。前三个指标应作为数据的一般信息由数据代理商评测并向所有用户公开，最后一个指标取决于买家，因此买家应该提供自己的评估目标场景和相应评估函数。

以前的工作提出了一系列从不同的方面和视角定义的数据质量<sup>[8]</sup>，然而面对丰富的数据模态和复杂的数据语义，可定量的数据质量评估还有大量尚未解决的难题。其挑战一方面来自对不同数据形态的各项质量指标的高效准确量化；另一方面来自如何在允许代理商和买家评估数据的同时保护各个主体的隐私，例如应该限制在卖家数据上执行的评估函数的内容和次数，以及保护买家的上下文质量评估函数。目前已有部分工作基于同态加密等方法，为保护隐私的质量评估提供了很好的思想，但距离全面、实用的数据质量评估方案仍有很大距离。

## 数据展示

为了吸引数据消费者并帮助他们更好地选择产品，代理商需要展示数据的信息（如主题、优点和应用范围）。区别于实物商品，数据具有易复制的特点，因此如果将数据完全展示给潜在买家，买家则可以直接通过复制获取数据而不购买。现有部分工作主要通过数据的采样、数据的版本、元数据和数据的摘要四种方式进行数据展示。

现有平台通常采用人工的方式生成元数据和摘要，然而面对海量的数据，数据展示的自动生成是提升效率和准确率的重要方法。由于摘要的生成需要对数据进行语义理解，因此一般比元数据生成要困难很多。已有一系列工作采用机器学习等方法为文本、音频、图像和视频生成摘要。为了更快更好地提供数据摘要，除了需要不断改进摘要模型本身，还需要充分考虑方法在大数据集上的执行效率，以及数据的安全性，即不能泄露太多有价值的信息。此外，摘要个性化（生成符合买家需求的摘要以提升销售量）的需求也提升了摘要生成的难度。

## 数据估值

在交易前，买家、卖家和数据代理商都需要估

算数据商品的价值，以确定自己的购买 / 销售策略。评估无形资产的方法有三种<sup>[9]</sup>。**(1) 基于成本**：数据的价值取决于收集、处理、存储的成本。由于数据通常是作为信息系统的副产品生成的，数据生产成本与其他产品共享，所以基于成本的方法难以估计数据的真实价值。**(2) 基于市场**：数据的价值取决于同一市场上可比数据的市场价格。但“可比性”的定义不明确，而且对大量数据集而言也不存在相似的数据集，故无法准确估值。**(3) 基于收益**：数据的价值取决于买家能从数据中获得的总收益。这种方法是主观的，仅在评估特定应用时有用，因此不同买方的估值可能会有很大差异。

上述三种传统方法都不能准确地评估数据的全部价值。我们也可以从买家、卖家和数据代理商的角度考虑数据的价值。**(1) 买家的估值**：买家需要评估以给定价格购买数据是否值得，或者在拍卖时应该给出什么价格。买家的估值取决于数据质量以及数据将给他带来的利润。**(2) 卖家的估值**：卖家在出售数据时需要设置最低可接受价格，为此，卖方需要根据成本、数据质量以及数据在市场中的稀缺性来估值。**(3) 数据代理商的估值**：代理商需要对数据进行估值以检验卖家设定价格的合理性。一般来说，代理商需要根据数据在市场中的稀缺性和数据质量进行评估。

数据估值的难度来源于三个方面：(1) 买家和数据代理商不能直接访问数据，且质量评估功能受限；(2) 估值依赖于买家的应用场景和需求；(3) 数据的稀缺性和重要性难以衡量。此外，数据功能和分析结果的估值同样重要且难以实现，在限量出售、独家垄断、代理权和所有权转让的场景下的数据估值也有待研究。

## 数据定价

对于数据定价，我们需要回答三个问题，第一个问题是“由谁来设定价格？”，第二个问题是“采用哪种定价策略？”，第三个问题是“什么是合理的价格？”

采用固定价格时，价格取决于供需关系。针对“供应估计”，由于数据是非独占资源，所以理论上它

具有无限的供应；销售数据几乎没有边际成本（卖家只需向买家提供云上数据的密钥）；因为完全竞争市场，Arrow-Debreu 均衡价格几乎为零<sup>[10]</sup>。针对“需求估计”，一种思路是先在没有价格的平台上列出数据，等待客户表达他们的兴趣，然后估计需求。然而，买家可以串通起来隐藏兴趣以降低价格，或者相似类型数据集的卖家可以协商减少销售量以抬高价格。所以我们不能简单地使用供需规律来确定数据的价格。对于变动的价格，如果价格随时间变化，可以使用数学模型预测其变化。如果卖家和数据代理商想要设定针对买家的个性化价格，则需要获知买家的应用场景。一般而言买家不愿透露应用信息，但可以通过买家之前运行的上下文质量评估函数来推测其应用。针对不同的购买顺序，模拟数据价值随购买人数增多而贬值的过程。对于限量出售、独家垄断、代理权和所有权转让等场景，数据是稀缺资源（供小于求），可以进行拍卖以确定其价格。在限量出售时，出售份数的选择是一个问题，同时也存在着卖家进行非法垄断（采用减少销售量的方式收取更高价格）的问题。由于数据具有非常高的固定成本和极低的边际成本，数据定价不仅取决于生产成本和市场竞争，还取决于买家对数据的价值。而如何为不同类型的数据定价，究竟什么是合理的价格仍是尚待探索的问题。

在数据交易前，数据的发现与获取也是一个重要课题。例如基于本体 (ontology) 的数据发现和信息集成，以及基于众包 (crowdsourcing) 的数据获取<sup>[11]</sup>，都能为交易带来更丰富的数据。数据隐私甄别和脱敏也是保障交易安全进行的重要前提。此外，还存在数据的收集处理、存储管理以及买方需求规范化等问题。解决好这一系列难题，才能为高效、准确、公平、安全的数据交易共享做好准备。

## 数据交易中的问题与挑战

在数据交易过程中，“拍卖”是一种重要的交易手段，我们需要为具有不同特性的数据设计合适的拍卖机制，并在此过程中考虑用户的隐私问题。交易

达成时，**合约**作为具有法律约束力的协议，需要被妥善设计以保护各个主体的权利。同时，除单用户购买外，数据集贩卖往往也存在**团队购买和捆绑销售**等情况。此外，对交易过程中不诚实的用户要通过设计**可问责协议** (accountable protocol) 来进行惩罚，以保证良好的数据交易生态环境。

## 拍卖机制设计

当需求多于供应时，可以用拍卖来决定卖给谁和收取多少费用。精心设计的拍卖机制可以将社会福利（每个实体的收益 / 效用的总和）或单边收益最大化。经典的拍卖机制包括英式拍卖、荷兰式拍卖和 Vickrey 拍卖等。数据拍卖的挑战之一是感兴趣的买家可能不会同时表现出他们的兴趣，因此需要很长时间才能将足够多的竞标者聚集在一起进行拍卖。对此，我们可以使用两种可能的拍卖模式：实时决策和延时决策。然而实时决策难以保证出售价格 / 利润最优，延时决策则面临买家可能会放弃等待，同时数据也可能贬值等。另一个挑战来自有时买方希望将出价作为隐私信息进行保护，因此保护隐私的拍卖机制也很重要。对于数据交易的拍卖机制的设计，基于机制设计的理论<sup>[11]</sup>，我们希望其具有**激励兼容 (incentive compatibility)**、**联合预防 (coalition-proofness)**、**个人理性 (individual rationality)**、**预算可行性 (budget feasibility)** 以及**计算效率高**五个属性。为了建立更诚实公平的数据拍卖市场，我们应该针对不同的数据特性和用户需求设计相应拍卖算法，并使其尽可能满足以上性质。

## 合约设计

合约是明确规定了法律强制执行相关方的权利和义务的协议。在数据交易中，合约的目标可以是买家、卖家、数据代理商或他们共同的效用最大化。首先需要回答的两个问题是：如何定义这些实体的效

用？应优先考虑哪个实体的效用最大化？此外，合约设计的最大挑战之一是信息不对称带来的对单方利益的损害，因此需在设计合约时保证激励兼容性，并能够奖励良好的行为，并惩罚坏的行为。另一方面的挑战是我们还需考虑均匀分配风险以使得合约更公平。例如在合约签署后，买家可能重估由数据获得的利润，重估的结果可能高于或低于之前的估值，因此买家或卖家中某一方的利益很可能受到损害。一种解决方案是让买家根据交易后的实际利润而非初始估价向卖家付款。合约设计已经得到了广泛的重视。Bolton 和 Dewatripont 考虑了多种信息不对称的情况，通过将单边或联合效用最大化引入最优合约<sup>[12]</sup>。最近，还有一系列工作利用人工智能和区块链等技术设计合约，例如基于区块链的智能合约进行设计<sup>[13]</sup>，实现在不需要可信第三方的情况下履行合约和追踪交易。

## 多方交易

团购已经成为电子商务中一种流行的交易模式。在团购中，一群买家在发起者的组织下会联合起来与卖家协商并获得低于零售价的折扣价。团购同样适用于数据交易，但需要考虑几个问题，如：(1) 什么是适当的折扣，以使得品牌效应的长期收益大于折扣造成的短期利润损失？(2) 折扣对不同应用需求的买家应该是相同的还是有差异的？(3) 如何保证整个流程的隐私安全？由于不同买家可能对相同数据的估值不同，要求他们为此支付相同金额是不公平的，我们可以通过公平划分理论将团购的总收益分配给买家，例如 Shapley value<sup>1</sup>。

同理，捆绑销售也适用于数据交易，即把来自相同或不同卖家的类似或相关的数据集打包出售给买家。数据交易场景中的捆绑销售存在着以下挑战：(1) 数据集的量大且种类繁多，代理商很难提供适当的捆绑销售方案；(2) 由于数据的协同作用，组合的

<sup>1</sup> 是指所得与自己的贡献相等的一种分配方式。普遍用于经济活动中的利益合理分配等问题。最早由美国洛杉矶加州大学教授罗伊德·夏普利 (Lloyd Shapley) 提出。Shapley 值法的提出给合作博弈在理论上的重要突破及其以后的发展带来了重大影响。

数据可能具有更高的价值，导致数据集的估值和定价变得更加复杂；(3) 如果代理商将多个涉及重合数据对象的数据集捆绑销售，则买家可能会从中推断出较多数据对象的隐私信息；(4) 利用组合拍卖来销售多个商品的问题通常是 NP 难的。除以上两种模式外还有更复杂的交易模式，例如它们的混合模式：一群买家团购捆绑销售的商品再内部分配。

## 可问责协议设计

可问责性<sup>[14]</sup>是一种性质，强调应该指责行为不端的协议参与者。我们可以为数据交易设计可问责协议，以迫使所有参与者诚实地遵循协议。要惩罚行为不端的参与者，首先，我们应该能够检测不良行为的后果，例如一些协议的期望目标没有实现。然后，找出谁应对此负责。最后，根据签署的合约惩罚不端参与者。可问责协议的设计有两个目标：公平性和完整性。公平性要求诚实的参与者不应该受到指责，而完整性要求所有行为不端的参与者都应受到指责。在过去十年中，有一系列工作致力于不同协议的可问责性的研究。2010 年，Küsters 等人<sup>[14]</sup>给出了可问责性的正式定义，并展示了几个可问责协议的设计案例。Jung 等人<sup>[15]</sup>设计了一个简单的可问责数据交易流程。

## 交易后的问题与挑战

在交易完成之后，数据代理商还需进行质量认证和保护数据版权，监督买家和卖家的行为，观察他们是否履行了合约中关于数据质量的承诺和传播的限制（卖家不能卖多，买家不能转卖），但由于数据抄袭难以界定以及离线传播难以追踪，实现数据追溯仍面临着许多挑战。成熟的电子商务市场中，评价系统和推荐算法是促进商品交易进行的重要力量，在数据交易环境下，代理商也需要快速准确地为买家推荐感兴趣的数据集。

## 数据追溯

在一个良好的可持续发展的数据交易市场中，

交易完成后数据传播的可追溯性对整个系统的可靠性至关重要，它决定了用户对系统的满意度和信任度。设计可追溯的数据交易机制是困难的。首先，难以保证数据的可追溯性，因为攻击者可能会采取任意措施来避免数据在传播中被跟踪和识别；其次，抄袭是难以检测的，因为用户可能会修改一小部分数据，然后将其列为市场上的“新”数据集；最后，数据代理商难以检测离线的非法数据交易。

一个想法是引入第三方可信机构监督市场上的所有交易，为每笔交易的数据附加水印，在准许参与者的操作之前先验证数据的现有水印，以检查是否违反了交易政策。然而，串通的参与者和离线数据流通会绕过监视。部分现有的数据交易平台声称已将区块链技术用于数据追溯，例如贵阳大数据交易所和京东万象，利用了区块链分布式数据存储的不可篡改、可追溯、可信任等特性。针对数据抄袭检测，Jung 等人<sup>[15]</sup>设计了相关技术，为了考量数据的原创性，他们定义了各种数据类型的原创性指数，并实验验证了该指标的有效性。还有一些可用于数据篡改检测和数据查重的工具，如默克尔树(Merkle tree)、数字签名和局部敏感哈希(LSH)等。

## 数据版权管理

卖家可以通过数据免费试用或者限期免费退款来帮助买家买到合适的数据，并提高销售额，但如何在试用结束或退款之后确保买家删除数据呢？买家可能不仅不删除数据，而且持有该数据的拷贝或稍加改动的版本，还可能将数据转移至其他存储设备或其他人。由于数据本身易拷贝、易更改、易转移，这些侵权的风险都无法消除。这些风险同样存在于现有的数字商品中，如电子书、音乐、电影。现有的数字版权管理(Digital Rights Management, DRM)通过加密和开发专用的软硬件来保护数字商品的版权，比如只能用特定的软件来看电子书或听音乐并且不允许下载，通过产品密钥、限制软件安装次数、持续在线身份验证等方法自动检测盗版行为。在数据交易中，直接将明文数据发给买家会导致无法恢复的侵权损失，所以我们必须利用 DRM 技术来限

制数据的使用、下载和传播。

现有的 DRM 技术需要做出以下改进以支持数据商品。第一，数据的访问不像流媒体那样是连续的，可能是任意的，所以现有技术对于数据可能难以实现高速的实时在线访问 (streaming)。第二，现有带版权管理的专用软件一般只允许用户浏览（如看视频、听音乐），而在数据交易中，这样的软件不仅要支持浏览，还要允许买家对数据做计算和可视化等。第三，需要禁止截屏功能并利用一些机制（如文献 [16]）阻止买家对屏幕拍照或录像而间接地侵权。第四，不仅要利用产品密钥和持续在线身份验证等机制来防止侵权，还需要检测侵权是否已经发生，例如，在数据被传输时记录发送设备和接收设备等信息，结合数据的交易合约中的版权限制，软件应该自动判断买家是否已侵权，如果是，应该通过扰乱数据或完全禁止使用等方式惩罚买家。

## 数据推荐

许多现有的交易平台一直在使用推荐系统来帮助用户找到他们可能喜欢的新商品。流行的推荐系统可分为三类——基于内容的过滤，协同过滤，以及它们的混合体。数据交易平台同样可以通过推荐系统来推动商品交易。在买家有历史订单记录时，分析他的兴趣是较容易的。对于基于内容的过滤，我们虽难以直接衡量数据集、函数和分析结果之间的相似性，但可以通过机器学习的方式推断它们的相关性。如果买方是新客户，代理商可以应用协同过滤来找到相似用户，然后向他 / 她推荐他们购买过的东西。但此时需要为每个客户建立准确的资料，这可能会带来一些隐私问题，代理商需要通过隐私保护的算法匹配用户，采用安全多方计算或同态加密计算买家的相似性。此外，买家较小的购买频率和不完整的个人信息也加大了代理商做出准确推荐的难度，而数据的稀疏性是研究人员一直在努力克服的问题。

## 基于区块链的数据交易和追溯

近年来区块链技术的飞速发展为数据交易和追

溯提供了新的思路。经典的区块链，即 2009 年中本聪 (Satoshi Nakamoto) 在比特币系统中使用的区块链，融合了非对称加密、数字签名、默克尔树、工作量证明等多种技术，为在无可信中心情况下转账信息的安全、可靠记录，提供了一套完整的解决方案。而后，研究者们对经典区块链进行了不同的修改和补充，以使其适应不同场景下的各项信息分享和记录的要求。2013 年，由 19 岁俄罗斯少年维塔利克 · 布特林 (Vitalik Buterin) 提出的开源具有智能合约功能的公共区块链平台——以太坊 (Ethereum)，在保有之前比特币区块链的支付转账功能基础上，提供了一个开放的、模块化的支持自定义高级应用的平台。以太坊支持用户编辑自己想要的应用，也就是智能合约。合约的调用过程和返回结果被记录在底层区块链中，一样的安全、可靠和不可篡改。

基于区块链及其相关技术，我们可以提出多种可能的数据交易和追溯的解决方案。例如，通过构建一个基本的区块链，即可完成分布式数据交易的基本功能；或者通过使用以太坊平台发行代币的方式，将代币和数字资产绑定，以实现数字资产的证券化和公开化；再或者通过为每一份数据建立唯一的档案合约，将该份数据的相关信息记入与之绑定的档案合约，对该份数据的买卖通过调用档案合约的不同功能来实现，将保有内容标签信息和数据版权的记录上链，利用链上信息的不可篡改性实现数据的防伪和版权确认。数据交易中介方还可以在本地建立合约仓库，将链上的无序合约在链下进行有序组织，通过链上链下结合的方式实现高效服务。

## 总结与展望

从以物易物开始，实体商品的流通及其交易市场的进化已经持续千年，并一步步演变到现今的经济全球化。可以预见数据交易也将不断发展，为国家、企业和个人带来更多的价值。对数据开放共享的急迫需求已经催生了一系列数据交易和

共享平台，但数据交易市场仍处于初级阶段，整个数据交易的流程仍然面临着许多法律及跨学科的难题和挑战，而这也代表着前所未有的机遇。随着相关研究的展开，相信总有一天我们会建立成熟的数据交易生态系统，为社会发展带来一片新的动力和繁荣。



李向阳

CCF 专业会员、CCCF 编委。中国科学技术大学教授，国家千人计划专家，ACM 中国共同主席，IEEE Fellow。主要研究方向为大数据的共享交易和隐私保护等。xiangyang.li@gmail.com



张 兰

CCF 专业会员。CCF 优秀博士学位论文奖获得者，阿里巴巴青橙奖获得者。中国科学技术大学特任教授。主要研究方向为跨域数据的深度理解、隐私保护和数据交易。zhanglan03@gmail.com



韩 风

中国科学技术大学硕士研究生。主要研究方向为数据理解、数据交易、安全隐私。hf1996@mail.ustc.edu.cn

其他作者：薛爽爽 钱建威 郑 翰

## 参考文献

- [1] 2016年中国大数据交易产业白皮书 . [http://www.cbdio.com/BigData/2016-06/02/content\\_4965656\\_all.htm](http://www.cbdio.com/BigData/2016-06/02/content_4965656_all.htm).
- [2] 国务院关于印发促进大数据发展行动纲要的通知 . [http://www.gov.cn/zhengce/content/2015-09/05/content\\_10137.htm](http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm), 2015.
- [3] 工业和信息化部关于印发大数据产业发展规划（2016-2020 年）的通知 . <http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c5464999/content.html>, 2017.
- [4] 国家信息中心大数据研究 . <http://www.sic.gov.cn/Column/551/0.htm>, 2018.
- [5] He K, Chen J, Du R, et al. Deypos: Deduplicatable dynamic proof of storage for multi-user environments[J]. *IEEE Transactions on Computers*, 2016, 65(12):3631-3645.
- [6] Yu J, Ren K, Wang C, et al. Enabling cloud storage auditing with key-exposure resistance[J]. *IEEE Transactions on Information Forensics and Security*, 2015, 10(6):1167-1179.
- [7] Wang R Y, Strong D M. Beyond accuracy: What data quality means to data consumers[J]. *Journal of Management Information Systems*, 1996, 12(4):5-33.
- [8] Batini C, Cappiello C, Francalanci C, and Maurino A. Methodologies for data quality assessment and improvement[J]. *ACM computing surveys (CSUR)*, 2009, 41(3):16.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>



## CCF 计算机视觉专业委员会

### 第三届计算机视觉及应用创新论坛举办

第三届计算机视觉及应用创新论坛 (RACV2018) 于 2018 年 11 月 24 日在广州举办。会议由中国计算机学会 (CCF) 主办，CCF 计算机视觉专业委员会和中山大学联合承办。

论坛由 CCF 计算机视觉专委会副主任、中山大学教授赖剑煌，专委会副主任、爱奇艺资深科学家王涛博士和专委会副秘书长、北京邮电大学副教授马占宇共同主持。专委会副主任、北京大学教授查红彬作了题为“基于三维数据流融合的场景重建与传感器定位技术”的报告；复旦大学教授姜育刚作了题为“面向视频识别的深度学习方法及应用”的报告；叠境数字科技（上海）有限公司创始人、上海科技大学教授虞晶怡作了题为“Learning to Build a New Reality”的报告；Momenta 研发总监任少卿博士作了题为“无人驾驶的技术与挑战”的报告；字节跳动人工智能实验室总监王长虎博士作了题为“抖音背后的那些黑科技”的报告。

# 智能与计算

关键词：脑科学 认知科学 心智计算理论 体验认知理论

李航  
字节跳动科技有限公司  
特邀专栏作家

## 前言

1950年，图灵发表论文《计算机器与智能》(Computing machinery and intelligence)，提出著名的图灵测试。这段时间里，图灵关注的主要问题是，在计算机上是否可以实现人的思考(thinking)<sup>[1]</sup>。他的基本观点是，只要进行适当的编程，计算机可以像人脑一样工作。我们不需要给思考一个严格定义<sup>1</sup>，可以通过图灵测试判断计算机的“思考”能力是否达到了人的水平。

1957年，冯·诺伊曼去世，次年他的遗作《计算机与人脑》(The computer and the brain)出版。该书是他在离世前的两年时间里准备的演讲草稿，讨论他当时最关心的研究课题：计算机和人脑。冯·诺伊曼把计算机和人脑都看作是计算机器(automata)，对两者进行了比较，试图为建立统一的计算机器理论奠定基础。

人的思考是不是计算，是怎样的计算？计算机是否可以实现人的思考？这个问题是认知科学、人工智能的一个核心问题，这一点从计算机领域两位巨人对这个问题的关注程度就可见一斑。

本文对计算与思考（或智能）这个话题进行简单综述与讨论。必须申明，笔者是计算机科学家，对脑科学、认知科学等是外行。因为人工智能的目标是要构建能够“思考”和“行动”的机器，所以

作为人工智能的研究人员又不能不对这些问题进行关注与思考，进而斗胆执笔，写出本篇文章，希望能抛砖引玉，引发大家的思索与辩论。

## 脑科学告诉我们的

人脑是由千亿级的数百种神经元（神经细胞）通过千万亿级的突触连接形成的神经网络，能够实现各种智能性功能，包括感知、认知、语言、情感、创造、意识。脑科学研究虽然取得了一定的成果，但离探明人脑的工作机理还相差甚远<sup>[2]</sup>。

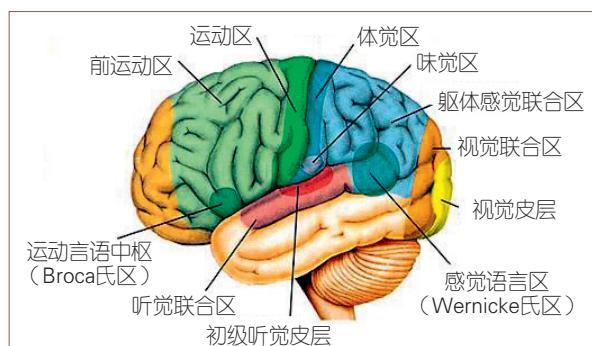


图1 大脑主要区域

在宏观层面，脑科学研究对大脑各个脑区的结构与功能有一定的认识。人脑由大脑、小脑、脑干组成。大脑最重要的部位是大脑皮层，人类与动物的主要区别在于人类拥有极其发达的大脑皮层，可

<sup>1</sup> 这里说的“思考”并没有严格定义，一般包括认知和感知。

所以说大脑皮层造就了人类的智慧。大脑皮层不同区域掌管不同的功能，包括视觉皮层、听觉皮层、味觉皮层、体感皮层、运动皮层、语言区等（见图1）。

在微观层面，脑科学的研究对神经元的信息处理机制有比较清楚的了解<sup>[3]</sup>。神经元通常由一个细胞体、一个轴突和多个树突组成。树突接入信号，轴突接输出信号，神经元与神经元之间由突触连接（见图2）。现神经元从多个前神经元得到输入信号，当输入信号超过一定阈值时被激活，产生输出信号，传递到多个后神经元。神经元之间的信号传递通过突触进行。前神经元在轴突末梢释放化学物质，通过突触传到现神经元的树突，打开现神经元的离子通道（ion channel），促使其细胞内外离子流动，形成现神经元的输入信号。现神经元的输出信号通过轴突以离子流的形式传递到轴突末梢，继续向后神经元传递。

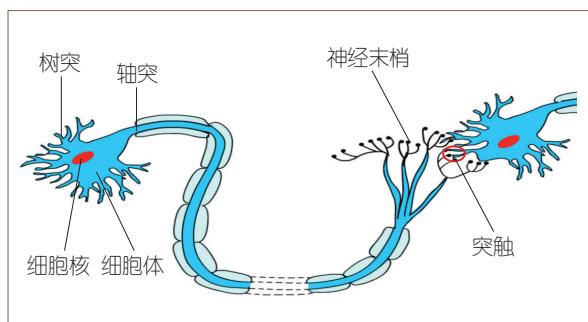


图2 神经元

在介观层面，脑科学的研究对神经环路的信息处理原理有一些认识。神经可塑性是神经网络的重要特点，有所谓的赫伯法则(Hebbian rule)，认为同时被激活的神经元之间的连接被强化，产生新的链路，形成新的记忆(fire together, wire together)。对概念的记忆存储于由密切连接的神经元组成的细胞群中，激活其中的部分神经元可以唤起对整个概念的记忆。

可以看出人脑是一个由庞大复杂网络组成的信息处理系统。它通过神经元之间的信号传递实现信息处理，具有以下特点：处理速度并不很快，进行的是并行处理，计算与存储融合在一起，拥有自学能力。

## 心智的计算理论

心智的计算理论 (computational theory of mind) 认为，人的思考是计算，人脑或心智是计算系统。这里说的计算不是比喻，而是实质上的<sup>[4]</sup>。这个认知科学、脑科学、人工智能等领域的理论，在20世纪60~70年代占据主流地位，代表人物包括认知科学家福多(Jerry Fodor)和平克(Steven Pinker)、脑科学家马尔(David Marr)、哲学家丹奈特(Daniel Dennett)等。

### 计算系统

马尔提出了计算的层次概念，认为无论是计算机还是心智都是计算系统，需要从三个不同且相关的层次理解，包括计算层、表征层、实现层。计算层决定系统的输入与输出，对应计算的功能；表征层决定系统内部的表征与算法，对应计算的软件；实现层决定系统的物理实现，对应计算的硬件。

心智的计算理论把心智看作是图灵式计算机(Turing style machine)，认为人的思考(感知、认知等)是这种机器上的计算。这一点与图灵和冯·诺伊曼的观点一脉相承。有许多理由让人相信这个想法的正确性。给定一个输入，产生一个输出，至少从功能的角度，心智做的是信息处理，可以把心智看作是一种计算系统。神经元对输入的多个信号进行处理，输出一个信号，进而传递信息，从实现的角度，是一种计算器件。

心智的计算理论中，心智的表征理论是重要的一个分支，从表征的角度进一步推进心智是计算系统的想法。

### 心智的表征理论

心智的表征理论 (representational theory of mind) 认为思考是在心智中(图灵式计算机上)的符号操作<sup>[5,6]</sup>。人的思考和行动是基于常识的，由信念或愿望驱动。信念是对事实的描述，愿望是对目标的描述，常识是对世界的描述，而这些描述是通过内心的语言进行的，称为“心智语言”(mentalese)。也就是说，心智中的符号操作基于心智语言。

心智语言同自然语言一样，由符号和语法组成。符号有简单的，也有复杂的，语法规则决定符号的组合方式以及产生的语义。听别人讲一段话，人一般不能复述原话，但可以把内容讲述出来，对这个现象的解释是，人理解自然语言时把它转化成了心智语言。自然语言有歧义（多个语义），但心智语言没有，原因是人能够区别自然语言的歧义，说明人用不同的心智语言表达了不同的语义。

有一些认知学实验支持心智语言存在的假说。比如，让受验者坐在电脑屏幕前，屏幕上瞬间闪出两个英文字母，根据内容快速按下两个按钮中的一个。如果两个英文字母相同，按其中的一个，如果两个字母不相同，按另外一个。有时出现的是同一个字母且大小写相同（如“A A”“a a”），有时出现的是同一个字母但大小写不同（如“A a”“a A”）。结果发现，大小写相同时，受验者按按钮的速度更快，准确率更高。说明在第二种情况，人需要做某种处理把视觉中的符号转换成心智语言中的符号。

## 中文房间

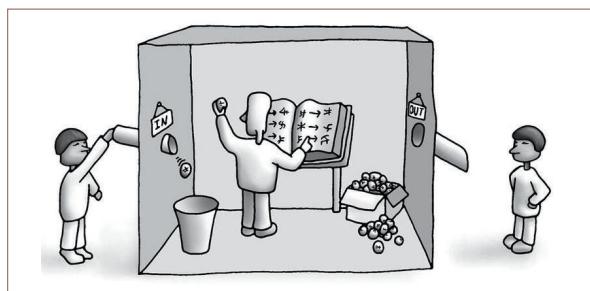


图3 中文房间

哲学家塞尔 (John Searle) 用著名的中文房间 (Chinese room) 思想实验，对“心智是计算系统，思考是符号操作”的想法提出质疑<sup>[7]</sup>。

中文房间思想实验是说，有一个不懂中文的人被放到一个房间里，其他人从房间外塞进写着中文的纸条。房间里有一本书，写着中文会话的规则。他根据书上的规则，对着纸条上的中文符号，找出

相应的中文符号画在纸条上，把纸条塞出房间外（见图3）。从房间外的人看，这个人能够用中文对话，会说中文，但是事实上他完全不会。基于符号操作的计算机器，和中文房间里的人一样，看似在使用语言，其实完全没有理解语言。说明语言理解乃至思考，不是计算和符号操作。

中文房间的论点引起了极大的反响，各种支持和反对的意见接踵而至。比如有一个代表性的反对意见是：确实这个人不会讲中文，但是整个房间会讲中文。因为从功能的角度来说这个房间整体可以完成中文的对话，这个人只是会讲中文的系统的一部分。塞尔对此的反驳是：这个人可以把所有的规则都记住，也可以离开这个房间，但是只要他不能把语义附加到符号上，就不能认为他会讲中文。塞尔的主要论点是符号操作只能代表语法，不能代表语义。

## 体验认知理论

体验认知 (embodied cognition) 理论<sup>2</sup> 是近二十年来兴起的理论，认为生命体（包括人和其他动物）的身体是感知和认知的基础，身体的体验对感知和认知起着决定性的作用<sup>[8]</sup>。代表人物包括认知科学家雷可夫 (George Lakoff)、脑科学家达马西奥 (Antonio Damasio)、哲学家克拉克 (Andy Clark) 等。可以说，体验认知理论对心智的计算理论提出了一定的挑战。

## 脑科学的假说

达马西奥认为，思考是能够在意识中产生表象 (image) 的，在下意识中进行的对神经表征 (neural representation) 的操作<sup>[9]</sup>。神经表征是人脑的神经活动（神经网络中的信号传递）产生的状态。表象是指人的意识中对事物形象的认识，包括视觉、听觉、体感等的表象。比如，提到“黄色的帽子”，我们会在脑海里联想到黄色的帽子，这就是它的视觉表象。

脑和身体是不可分割的有机体（这里说的身体

<sup>2</sup> 也有人译作“具身认知”理论。

指除去脑之外的身体部位)。脑和身体的相互作用，形成一个整体，与外界相互作用，产生人的行为。通过神经系统，外界信号可以从身体器官传到大脑，指令信号也可以从大脑传到身体器官。大脑发出的指令未必都经过思考，有很多属于被动的反应。经过思考的指令，会在意识中产生表象，成为人的主动的命令。达马西奥指出“我们未必是思考机器，其实我们是思考的感觉机器(We are not necessarily thinking machines; we are feeling machines that think)”。

思考也使用单词和符号。单词和符号作为表征被记忆，人在说出或写出一句话之前，单词和符号相关的听觉表象、视觉表象等浮现于意识中。人的逻辑和数学思维也基于表象，而不是符号。一个证据是，许多数学家、物理学家，包括爱因斯坦，都将自己的抽象思维过程描述为表象的操作过程。

这里谈到意识，这也是认知科学、脑科学和哲学关注的一个重要问题，至今仍是一个很大的疑团。因为涉及的内容较多，本文不作讨论。

## 体验模拟假说

体验模拟假说(embodied simulation hypothesis)是关于语言理解的体验认知理论，认为人的语言理解是在心智中进行的，基于自己过去的视觉、听觉、运动等体验的模拟<sup>[10,11]</sup>。

人进行语言理解时既使用语言相关的大脑部位，又使用感知和运动相关的大脑部位。理解语言描述的概念时，会联想到概念相关的图像，这时大脑视觉皮层变得活跃；会联想到概念相关的声音，这时大脑听觉皮层变得活跃；会联想到概念相关的运动，这时大脑运动皮层变得活跃。语言理解的过程就是，唤起大脑各个部位相关体验的记忆，基于这些记忆在心智中生成语言所描述的内容的过程。

语言理解大多发生在下意识，在意识层面，会产生相关的表象。比如，问：“大猩猩有没有鼻子？”

要回答这个问题，我们会在脑里先浮现出大猩猩的视觉表象，然后根据这个表象去回答问题。再比如，听到：“flying pig(飞猪)”，不同的人会根据自己对飞的概念的理解(飞的表象)，以及对猪的概念的理解(猪的表象)组合成不同的新的表象。

如果认为语言理解不是基于符号，而是基于体验模拟，那么中文房间中的人确实没有理解语言，塞尔的观点可能是正确的。语义不是由符号定义出来的，而是从人与外界交互的体验中积累抽象出来的。

有很多认知学实验证明体验模拟假说的正确性。有这样的实验，让受验者先听一句话，然后看一张图片，之后快速按下两个按钮中的一个。如果图片中出现了句子中描述的物体，按其中的一个按钮，否则按另一按钮。例如，句子有“木匠把钉子钉进墙里”(常识中这时钉子的方向是水平的)，“木匠把钉子钉进地板”(常识中这时钉子的方向是垂直的)，图片中显示的物体有水平方向的钉子，也有垂直方向的。结果发现句子中钉子的方向和图片中钉子的方向一致时受验者的反应速度更快，判断准确率更高。更一般地，语言中描述的和图像中显示的同种物体，当方向、形状、颜色相同时<sup>3</sup>，人能更快地判断其同一性。说明人在理解语言时，根据自己的经验在视觉上想象出了对应的场景。

## 比较与评论

### 两个理论

心智的计算理论与体验认知理论在思考即计算问题上有相似的观点，但在思考是怎样的计算问题上观点完全不同<sup>4</sup>。从近年的研究成果来看，体验认知理论对人的感知与认知机制能够给出更好的解释，有很多理由让人相信这个理论的正确性，虽然现在还不能完全否定心智的计算理论。

心智的计算理论以意识为主要对象，基本不考

<sup>3</sup> 形状：“天空中飞翔的老鹰”与“躲在巢中的老鹰”。颜色：“放在橱柜中的牛排”与“放在餐盘上的牛排”。

<sup>4</sup> 其实这两个学派都有不同的学者，他们对具体问题的观点不尽相同。

虑下意识；只关心人脑或心智，而不关心身体，对这个理论来说，身心是可以分开的，智能可以独立于身体而存在。体验认知理论关注的是人脑和身体的统一体，强调下意识对意识的影响，身体对人脑或心智的影响；对这个理论来说，身心是不能分开的，（人的）智能不可能独立于身体而存在。心智的计算理论中的计算是意识中的符号特征的操作。体验认知理论中的计算是下意识中的神经表征的操作，其结果浮现于意识中成为表象。图4给出了两个理论的对比。

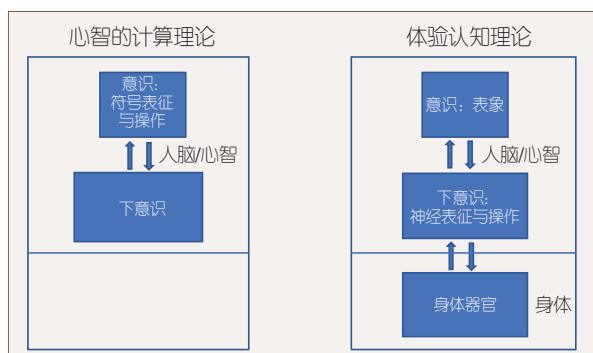


图4 心智的计算理论与体验认知理论的对比

## 人工智能

在人工智能 60 多年的历史中，一直有符号主义 (symbolism) 和连接主义 (connectionism) 之争。前 30 年研究的重点是知识与推理，占主导地位的是符号主义，后 30 年特别是近十年研究的重点是机器学习，特别是深度学习，占主导地位的是连接主义。这与心智的计算理论与体验认知理论的发展在时间上有一定的对应关系。认知科学和人工智能本来就是相互影响的两个学科。

深度学习的重要概念是人工神经网络和神经表征。神经表征将图像、语音和语言的内容都表示为实数向量。人工神经网络是对各种表征进行操作，从而完成各种感知、认知的模式识别任务的机器学习模型。深度学习的“神经表征”与体验感知理论的“神经表征”不尽相同，但也有相通之处。人工神经网络和生物神经网络具有不同的机制，后者借鉴了前者的原理。

## 结语

思考即计算这一命题是认知科学与人工智能的核心问题。图灵和冯·诺伊曼时代以来，其正确性就不断被一些事实所佐证。计算机在数值计算上早已超过人类，近年在智力竞赛、围棋上又完胜人类，在图像分类、语音识别、机器翻译上也接近人类。这些说明对人类来说属于思考的问题，在计算机上都可以实现。由此看来，图像理解、语言理解等现在看来还非常困难的问题，未来将有希望在计算机上实现或部分实现，原因是这些问题的本质也是人的思考。

心智的计算理论和体验认知理论从不同角度对“思考是怎样的计算”给出了答案。前者认为思考是符号计算，后者认为思考是神经计算。虽然现在没有确定性的结论，但体验认知理论似乎对人的思考机制给出了一个令人信服的解释。近年深度学习的巨大成功说明基于神经计算（人工神经网络）的信息处理能更好地实现人的感知与认知能力。希望体验感知理论的研究取得更大的进展，为人工智能提供更多的启发和引导。

如果智能和身体不可分割的假说成立，那么构建像人一样的智能系统就需要从开发智能系统的“身体”入手，让它们在与环境的互动中获得智能，这似乎意味着要走一条非常遥远而艰难的路径。但现实中往往并不需要构建像人一样的智能系统，很多情况下能得到辅助人的智能工具就足矣，所以问题可以被简化，这时体验认知理论仍然具有借鉴意义。 ■



李航

CCF 高级会员，CCCF 特邀专栏作家。字节跳动科技有限公司人工智能实验室总监 (Director of AI Lab)。主要研究方向为自然语言处理、信息检索、机器学习等。  
lihang.lh@bytedance.com

## 参考文献

[1] Proudfoot D.What Turing himself said about the imitation game[J].IEEE Spectrum, 2015, 52(7):42-47.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>

# 电脑前传(2)：计算

关键词：不可数 可计算数 图灵机 判定问题

黄铁军  
北京大学

## 不可数

从结绳记事开始，数和计算就成为人类认识世界、改造世界、创造新世界的有力工具。掰指头数数是最基本的智力活动之一，部分动物也会，这应该是自然数的起源。负数的概念最早出现在公元前三世纪我国的《九章算术》，西方国家直到1637年才由笛卡尔在《几何》中勉强承认负数的地位。0是在公元五世纪左右由印度人发明的（可能源于印度“绝对无”的观念），之后由阿拉伯人列为10个数字之一，广为传播，差不多和负数在同一时期被西方国家接受。小数也是《九章算术》发明的，今天使用的小数计数法则是文艺复兴之后的16~17世纪才确定下来。

“实数”这个词在18世纪才正式登上历史舞台。实数给人踏实的感觉，可以和直线上的点建立一一对应关系。实数集被称为“连续统”。十进制数位的整齐划一加强了实在感：两个整数之间等间隔地插入10个小数，再在两个小数之间更细密地等间隔插入10个小数……模式统一，秩序井然。我国自古以来很多哲学家都深信冥冥之中自有定数，18世纪岳麓书院山长旷敏本总结了这种信念：“是非审之于心，毁誉听之于人，得失安之于数”<sup>1</sup>。

但这种淡定，实为幻觉。

在实数概念出现之前的1683年，自然数就已

经让伽利略隐隐不安了：对自然数做最简单的“数数”动作，只数偶数，每个偶数对应一个自然数（那个偶数的1/2），一一对应，因此偶数和自然数一样多。但明明偶数只是自然数的一半，另一半是迥异的奇数啊，这是怎么回事？

人类在懵懵懂懂中走过100多年，19世纪终于迎来了一位认真“数数”的人——德国数学家、集合论的创始人格奥尔格·康托（Georg Cantor, 1845—1918）。1873年，康托在一封信中提出，他怀疑在自然数和实数之间不能建立让伽利略困惑的那种一一对应关系，即实数不像自然数那样是“可数的”。一年后康托就证明了自己的怀疑。18年后的1891年，康托发表另一种证法——对角线证明法（diagonal proof）<sup>[1]</sup>。这一绝妙思路开辟了一条全新路径，库尔特·哥德尔（Kurt Gödel, 1906—1978）和阿兰·图灵（Alan Turing, 1912—1954）的伟大证明中都会用到它。

对角线法是一种反证法，只用到自然数常识和数学归纳法，不用公式，只借助文字，就能说清楚。

先假定实数像自然数那样是可数的，那么就可以排列成一个阵列：每个实数占一行，各行用小数点对齐，我们只需关心小数点后面的序列，遇到整数或有理数，只把后面的小数位都写成0。实数有无穷多个，因此这个阵列就有无穷行，先后次序无所谓，无论你怎么排，第一行记为第1个实数，第二行记为第2个实数，如此排列下去，直到你相信

<sup>1</sup> 岳麓书院讲堂中一副对联的前三句，由旷敏本撰书。旷敏本（1699—1782年），乾隆十九年（1754年）被聘为岳麓书院山长，即现代意义上的“院长”。

包含了所有实数。

康托的对角线操作是：第1个数，取小数点后第一位；第2个数，取小数点后的第二位；第3个数，取小数点后的第三位……再把取出来的数位按上述自然次序排成一个无限长的数，把小数点放在这个数的最前面，然后让这个数的每位都加1，如果遇到9，就变成0，这样就得到一个新数。

把得到的新数与前面已经排好次序的实数序列逐项对比：第1个数的第一位和这个数的第一位不同，第2个数的第二位和这个数的第二位不同……，第 $N$ 个数的第 $N$ 位和这个数的第 $N$ 位不同……，一直比对下去，如果这个新数和已经排好序列的实数集合中的任何一个数都不同，即这个新数不在实数集合中。这和集合中已经包含了所有实数的假设是矛盾的。“实数不可数”得证！

1895年，为了区分自然数的可数性和实数的不可数性，康托把自然数集合的基数（也称为势）记为 $\aleph_0$ （阿列夫零），称为第一超限数，即可数的无穷大。根据幂集的定义，自然数的幂集的基为 $2^{\aleph_0}$ ，康托证明了这正是实数（连续统）的基数。康托推测，第二超限数 $\aleph_1$ 就是 $2^{\aleph_0}$ ，在 $\aleph_0$ 和 $\aleph_1$ 之间不存在其他超限数，即“连续统假设”。1900年，德国著名数学家大卫·希尔伯特(David Hilbert, 1862—1943)把连续统假设列为20世纪有待解决的23个重要数学问题的1号问题。1963年，美国数学家保罗·科恩(Paul Joseph Cohen, 1934—2007)证明连续统假设无法通过定理证明，这是一条独立于公理的陈述，无法通过公理证明或反驳。

对于第二超限数 $\aleph_1$ ，康托还有惊人发现：首先，实数和它的真子集（例如0到1之间的实数）是等势的，因此考察0到1之间的实数的性质，可以推广到所有实数。其次，连续统（直线）和平面乃至 $N$ 维空间的点之间也能建立一一对应关系，只需要把表示各个维度的数字逐位合并成一个新数就行，因此也是等势的。康托对自己的这个发现震撼得说

不出话来，他在一封信中写道：“我能看到它，但我不相信它。”

无穷对康托不仅是震撼，还有无尽的折磨。他对无穷的研究当时就饱受争议，遭到哲学家、神学家和数学家的抨击，他的老师利奥波德·克罗内克(Leopold Kronecker, 1823—1891, 德国数学家与逻辑学家)甚至斥之为无稽之谈。1884年，康托精神崩溃，他自己归因于对连续统的紧张工作和克罗内克的攻击。

## 不可计算数

从折磨康托的不可数转回具体数字，我们发现的不是秩序，而是更多的诡异。

人们最初认为整数之间的小数排列有序，都可以表示为两个整数之比（分母不为零），也就是后来所谓的“有理数”，这个词反映出人们期望数字能够具有良好的秩序。但这种美好的愿望在公元前六世纪就被打破了，毕达哥拉斯(Pythagoras)的学生希帕索斯(Hippasus)发现，根据毕达哥拉斯定理（勾股定理），直角边均为单位1的直角三角形的斜边无法表示成一个有理数。这一发现让相信数乃万物之本的毕达哥拉斯学派几乎疯掉，他们把希帕索斯扔进了地中海，把这个数称为“无理数”，并一直掩藏这个发现。后来人们逐渐意识到，无理数根本不是个案，相比之下，有理数才是汪洋大海中稀疏的小岛。

代数是数学的一个古老分支，代数方程是解决现实问题强有力的工具。求解整系数方程的整数根在公元前就被人们津津乐道，西方国家称之为丢番图方程<sup>2</sup>，这也是我国《九章算术》第八章的话题。

代数方程的解（根）称为代数数，好奇的人们不禁要问：“所有的实数都是代数数吗？是否存在不是任何代数方程根的实数？”1740年，瑞士数学家莱昂哈德·欧拉(Leonhard Euler, 1707—1783)猜想这样的数是存在的，并称之为“超越数”（因为它们

<sup>2</sup> 丢番图方程(Diophantine equation): 有一个或者几个变量的整系数方程，它们的求解仅仅在整数范围内进行。丢番图是古代希腊人，被誉为代数学的鼻祖，他早在公元3世纪就开始研究不定方程，因此常称不定方程为丢番图方程。

超越了代数)。100年后的1841年,法国数学家约瑟夫·刘维尔(Joseph Liouville)用阶乘构造出了第一个超越数,1882年,德国数学家费迪南德·林德曼(Ferdinand von Lindemann, 1852—1939)证明圆周率 $\pi$ 也是超越数,之后自然常数e(代表欧拉)也被证明是超越数,后来找到的超越数越来越多,但是一直没有一种通用方法证明一个实数是否是超越数。

是否存在能够找到所有代数数的通用方法?这就是希尔伯特1900年提出的23个问题中的第10号问题“丢番图方程可解性的判定”。在1928年数学大会上,希尔伯特又提出三个问题,第三个就是比第10号问题更具一般性的判定问题(Entscheidungsproblem):寻求一种确定的方法,从而能够在有穷步骤内确定某类问题中的任何一个是否具有某一特定的性质。“算法”这个古老词汇从此被赋予了明确的含义。

1936年,美国数学家阿隆佐·邱奇(Alonzo Church, 1903—1995)和图灵分别对判定问题给出了否定回答。两人不约而同地定义了一种“新数”——“可计算数”,只是用词稍有不同:邱奇用的是“Calculable Numbers”,图灵用的是“Computable Numbers”。

图灵在论文开头就给出了定义:“可计算数可以简单描述为其小数表达式可在有限步骤内计算出来的实数。”这里的“有限步骤”(finite means)并不是说确定数位的过程有限(事实上往往是无限的),而是指确定数位的方法是有限的,即算法有限。

有理数显然是可计算的,图灵断言所有代数数都是可计算的,超越数有一部分是可计算的,这其中包括 $\pi$ 和e。

图灵证明了所有可计算数是可数的,而实数是不可数的,因此,实数“绝大多数”是不可计算的。

## 图灵

1930年12月,18岁的图灵第二次参加剑桥大学三一学院的入学考试,仍未被录取。他的第二选择是国王学院。这一次,他决心专攻数学,全心钻研英国大数学家哈代(G. H. Hardy, 1877—1947)的经典著作《纯数学教程》备考。1931年秋,剑桥大

学国王学院迎来了最著名的学生之一。

1932年,图灵研读的是一本新书——《量子力学的数学基础》,这是年轻的匈牙利数学家冯·诺伊曼(John von Neumann, 1903—1957)的著作。当时冯·诺伊曼在大卫·希尔伯特身边研究数学,其所在的哥廷根大学是量子力学的圣地,所以写出这样一部著作也在情理之中。

1933年图灵研读的是英国数学家伯特兰·罗素(Bertrand Russell, 1872—1970)的《数学哲学导论》。这部1919年的作品在末尾写到:“如果有学生因为这本书而迈入数理逻辑的大门,并进行认真的研究,那么这本书就达到写作的初衷了。”图灵显然足以慰藉罗素的衷心。

1934年6月,图灵顺利毕业,凭借优异的成绩,获得了国王学院奖金资助,得以留校从事研究工作,次年4月获聘研究员。

1935年春天,图灵修读“数学基础”课程,授课教师是麦克斯韦·纽曼(Maxwell Herman Alexander Newman, 1897—1984)。这门课涵盖了当时尚未解决的判定问题,纽曼把希尔伯特寻求的“确定的方法”称为“机械过程”:用于解决某个问题的一组明确(但无意识的、非智能的)指令集。

1935年5月,图灵考虑到数理逻辑圣地普林斯顿大学,申请宝洁奖学金未果,但没影响这位年轻研究员的心情。初夏时节,图灵躺在剑桥大学的格兰切斯特草坪上,想到了解决判定问题的思路。第二年春天,图灵把《论可计算数及其在判定问题上的应用》论文草稿交给了纽曼。

就在阅读草稿那几天,纽曼收到了邱奇寄来的短文“判定问题的笔记”(1936年3月发表)<sup>[3]</sup>,基于另一篇已刊出的论文(1936年4月出版)<sup>[4]</sup>,邱奇对判定问题给出了否定回答。

按照惯例,邱奇已经解决了问题,图灵的论文不能再发表了。但纽曼意识到,图灵的方法与邱奇有很大差异,而且更简洁明了,因此他建议伦敦数学学会发表图灵的论文。伦敦数学学会记录的收文时间是5月28日,正式出版于1936年的11月和12月两期,1937年12月又发表了三页的修订。在

论文序言部分，图灵引用了邱奇的两篇论文，声明邱奇“有效可计算性 (effective calculability)” 概念和自己的“可计算性 (computability)” 是等价的。

把论文发给伦敦数学学会后，纽曼很快（5月31日）给邱奇写了一封信，比较了两人证明方法的不同，并直言“我觉得如果可能，明年他应该和你一起研究。”结果是，1936年9月，图灵来到普林斯顿大学就读邱奇的博士。

1936年12月，博士新生图灵在普林斯顿数学俱乐部报告了自己的论文，听者寥寥，不足十人，这让他很郁闷，在家书中写道“只有名人才能吸引听众”。

1938年6月，图灵获得博士学位<sup>[6]</sup>。他的博士论文的要点是：既然存在不可判定的命题，那就以它为真，加入原有系统，从而得到一个新系统，在新系统中，不可判定命题（已经为真）就可判定。当然新系统又会出现新的不可判定命题，解决方法就是再构造新系统，如此迭代，形成分层结构。这篇论文引入的另一个概念是改进的图灵机，它可以中断计算来寻求外部信息。当然，这些内容与图灵的伟大证明相比都算不了什么，但可以换来一个博士学位。

毕业之际，冯·诺伊曼想以1500美元的年薪招图灵为研究助理。图灵婉拒了，回到剑桥大学继续担任研究员，薪水为每学期10英镑。

## 图灵机

为了解决判定问题，图灵想象了一种通用计算机器，也就是我们今天所谓的“图灵机”。图灵机后来的影响超过了判定问题本身，图灵可能也意识到了这一点，所以把论文题目定为《论可计算数及其在判定问题上的应用》。

论文第1节“计算机器 (Computing Machine)”开门见山地给出了定义：“我们可以将一位正在进行实数计算的人比作一台只能处理有限种情况  $q_1, q_2, q_3, \dots, q_R$  的机器，这些情况称为‘m-格局’。”1937年5月，邱奇对图灵的论文发表评论文章：“一位持有铅笔、纸和一串明确指令的人类计算器，可以

被看作是一种图灵机”，这是“图灵机”一词最早见诸文字的地方。

时至今日，已经出现的所有计算机都是图灵机，但不要因此就认为图灵机就是机器。事实上，在计算机出现之前，Computer本来就是指以计算为工作的人（通常是女性）。人在计算时可能会犯错，这也没超出图灵机的定义：一台根本不会正确工作或不会做任何有意义工作的图灵机还是图灵机，图灵称之为“不符合要求的”图灵机。绝大多数数学教育，甚至可以说所有教育，目的就是把你从一个“不符合要求的”图灵机培养成“符合要求的”图灵机。

就像人做演算需要草稿纸，图灵的机器也是如此：一条无限长的可以左右移动的纸带穿过机器，纸带上是排列整齐的一串方格。任何时候都只有一个方格在机器里，机器可以读、写或擦除方格里的字符，就像一个笨拙的，或者说特别认真的，做四则运算的孩子。

实际上这台装置连四则运算都做不了，它的基本动作就是把纸带左移或右移一格以及在当前方格上进行读、写、擦操作，其他什么动作都不会。不过这正是图灵想要的，在论文第2节，他一口气给出了四个定义（原文没编号）：

1. 如果每一阶段的动作完全由格局所决定，我们称这样的机器为自动机。
2. 如果一台自动机打印两种符号，第一类符号完全由0和1组成（其他符号称为第二类符号），那么这样的机器就称为计算机器。
3. 如果给机器装上空白纸带，并且从正确的初始 m- 格局开始运转，那么机器打印出来的第一类符号组成的子序列，就叫做机器计算出的序列。
4. 在这个序列的最前面加上一个十进制小数点，并把它当作一个二进制小数，所得的实数就称为机器计算出的数。

就这么一台简单的机器，就能打印出所有可计算数？图灵的魔法在于“m-格局”，m指的是机器 (machine)，格局 (configuration) 是机器所处的状态，也就是机器所能处理的情况。机器运行就是在不同状态之间切换，机器的功能取决于格局的定义，也

就是会遇到哪些格局？遇到每种格局应该怎么办？

机器如此简单，遇到的情况也简单：目前的方格是空格还是某种字符？能办的事情更简单：左移、右移、写和擦除。简单！这就是图灵机。

在论文第3节“计算机器示例”中，图灵展示了如何定义一组格局，让他的机器打印出一个有理数和一个无理数。

论文第4节“缩略表”定义了一组常用功能，图灵开始称为“骨架表”，后来称为函数，也就是后来程序员都明白的子程序或函数，差别就是程序员是在计算机上写代码，而图灵是在没有计算机的情况下凭空想象。另外他的机器首先关注的不是加减乘除等功能函数，而是在纸带上进行字符串搜寻、拷贝等常用的基本功能。

论文第5节“可计算序列的枚举”。完全格局是指完成一个操作后的纸带快照、读写头的位置和下一个格局。从头开始顺次把完整格局串成行，就是机器完整的历史操作记录，“我们把机器表中这样形式的表达式写下来，并且用分号分隔开来。如此一来，我们就得了机器的完整描述。”

把完整描述进行标准化，再把所有符号替换成阿拉伯数字，就得到一个整数，称为描述数。“一个可计算序列是由计算它的机器所描述。事实上，任何可计算序列都可以通过这样的表描述。”因此，“每个可计算序列至少对应一个描述数，但不存在一个描述数对应多个可计算序列。因此，可计算序列和可计算数是可数的。”简言之，一台机器可以用唯一的一个整数进行编码，它对应一个可计算数，机器是可数的，可计算数也一定是可数的。

论文第6节“通用计算机器”和第7节“通用机的详细描述”是这篇文章的中心，篇幅不长，也相对容易理解。“发明一台计算任何可计算序列的机器是可能的”，这就是图灵的通用机 (Universal Machine)。通用机的输入是“开头写有某台机器 M 的标准描述”的纸带，“可以计算出与 M 相同的计算序列”。

图灵在论文第7节定义了一组骨架表来完成这个任务，图灵第一次用“指令”这个词代替表或函数，这是区分通用机和之前只能打印一个可计算序列的

专用机的关键。用现在的话说，专用机只能完成一个任务，而通用机是可编程的，指令就是纸带上的标准描述。

显然，图灵想象中的机器已经具备了存储（纸带）、软件（描述）和硬件（通用机）等概念，只是他没用我们今天熟悉的词语。15年后的1951年，图灵在曼彻斯特大学做程序员，他是这样定义“编程”的：“一种让数字计算机按照人的意愿工作，并将其正确表达在穿孔纸带上的活动”。

## 可计算数不可计算

简单总结一下：每台专用机能够产生一个可计算序列，对应一个可计算数，专用机的计算过程可以编码成一个描述数，通用机执行这个描述就可以产生专用机同样的可计算序列。我们是否就此可以得出结论：通用机是否可以算出所有的可计算数？

似乎可以回答“是”。前提是能够设计出所有专用机，用今天的话说就是编写出所有可能的软件，并在计算机上执行这些软件。软件数可数但无穷多，因此完成这个任务的前提是编制无穷多个软件，再无穷无尽地执行下去，这实际上是做不到的。

正确答案应该是“否”。尽管每个可计算数都可以算出来，而且所有可计算数是可数的，但是不存在枚举出所有可计算数的算法，只能一个一个地算，不存在找到所有可计算数的通用算法，这就是“可计算数不可计算”的含义。

证明方法是反用康托的对角线法，这就是图灵在论文第8节提到的“对角线法的应用”。既然可计算数是可数的，那就可以把所有可计算数排成队列，用对角线法构造出一个新数。根据对角线法，这个新数与队列中的任何可计算数都不同，因此是不可计算数。然而，构造这个新数的过程是一个典型的可计算过程，因此这个新数是可计算数。这就导致了矛盾。

矛盾的根源在于不可能对可计算数实行对角线法，上述可计算数队列根本没办法构造出来。

自然数可以逐个枚举，因此找出所有可计算数最直接的思路是逐个检查所有自然数，看它是否描

述了一台能够产生一个可计算数的专用图灵机。假定机器 D 能够检查一个整数是否符合要求的描述数，通用机 U 能够按符合要求的描述数执行并产生对应的可计算数，机器 H 按照对角线法调度 U 和 D 来逐位产生新数。

下面这个系统开始运行。

H 从  $i=1$  开始逐个检查所有自然数，用  $r(i)$  来记录已找到的可计算数的个数， $r(1)=0$ ，之后的规则是：如果整数  $i$  不是符合要求的描述数，则  $r(i)=r(i-1)$ ；反之， $r(i) = r(i-1)+1$ ，同时 H 调用 U 计算出  $i$  对应的可计算数的前  $r(i)$  位，并把第  $r(i)$  位添加到新数的第  $r(i)$  位。

三台机器似乎可以联合起来按部就班地运行了。

但机器 H 终究会碰到自己所对应的那个整数，设为  $k$ 。因为 H 的一切行为正常，因此  $k$  是一个符合要求的描述数，D 也应该做出这样的判断，按规则 H 就会把  $k$  转换成标准描述，交给 U 去执行计算任务，也就是产生  $k$  对应的可计算数的  $r(k)$  位。执行描述数  $k$  的机器 U 就等于 H，它会依次产生  $r(1), r(2), \dots, r(k-1)$ ，但要产生  $r(k)$  时却回到了任务本身，陷入无休止的死循环，永远产生不出  $r(k)$ 。

出现上述困境，说明假设错误，因此，不存在能够生成所有可计算数的通用算法，也不存在能够判别任何指定数是否是可计算数的万能机器。这意味着：软件要一个一个地去编，软件中的漏洞也要一个一个地去查，没有万能机器能帮我们完成这个任务。

## 判定问题

从论文第 9 节“可计算数的范畴”开始剩下的十多页，是可计算数在判定问题中的应用，被《图灵传记》<sup>[7]</sup> 的作者安德鲁·霍奇斯 (Andrew Hodges) 誉为“有史以来数学类论文中最不寻常的部分”。其实上一节已经体现了证明的精髓。

“判定问题 (Entscheidungsproblem)”这个德文词是希尔伯特的助手海因里希·贝曼 (Heinrich Behmann, 1891—1970) 创造的。在贝曼的想象中，判定过程是这样的：

这个问题的一个特性至关重要，就是证明过程只允许根据指令进行纯机械式的计算，不允许掺杂任何严格意义上的思考活动。如果愿意，我们可以说机械的或像机器一样地思考（说不定以后我们可以用机器来运行这种过程）。

贝曼这番话是 1921 年 5 月 10 日在哥廷根数学协会关于判定问题的座谈会中讲到的（这个材料近年才公开<sup>[8]</sup>）。

1928 年，已届暮年的希尔伯特在国际数学家大会上将判定问题列为三大问题之一，他梦想得到一个肯定回答。英国数学家哈代对此嗤之以鼻，他指出<sup>[7]</sup>：“当然不存在这样的公理，我们应该感到庆幸，因为如果我们找到了一套机械的规则，为所有数学问题提供解决方案，那么数学家的活动就将结束。”

哈代和希尔伯特各执一词时，图灵还在备考剑桥大学，攻读的正是哈代的《纯数学教程》。四年大学毕业后，他才第一次听到判定问题，躺在剑桥草坪初夏温暖的阳光下，破解了两大顶级数学家争执的世纪难题。

那是 1936 年，图灵 24 岁，一个刚刚大学毕业的翩翩少年，为机械意义上的计算画出了明确边界。

这正是：数可数，非常数

实数不可数，实在在何处？

计算又可数，其他为何物？

其他可想而知不可及

机可及，图灵机

### 黄铁军

CCF 杰出会员。北京大学教授，计算机科学技术系主任、数字媒体研究所所长。主要研究方向为视觉信息处理和类脑计算。[tjhuang@pku.edu.cn](mailto:tjhuang@pku.edu.cn)

### 参考文献

[1] Cantor G. Ueber eine elementare Frage der Mannigfaltigkeitslehre[M]// Jahresbericht der Deutsche Mathematiker-Vereinigung 1890-1891. 1891,1:75-78.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>

# 大数据交易市场构建

关键词：数据交易 众包采集 在线定价

郑臻哲 吴帆 陈贵海  
上海交通大学

## 大数据共享与交易

我们正处于大数据时代，海量数据在各个领域发挥着越来越重要的作用。数据拥有巨大的经济价值，被比喻为新兴石油资源，我国已经把大数据作为国家战略发展方向。根据国际数据公司 (International Data Corporation, IDC)2017 年的预测<sup>1</sup>，当今数据的体量正在以指数级的速度高速增长，预计 2025 年将达到 163 泽字节 (zettabytes)。然而，由于缺少有效的数据流通共享平台，现有的海量数据大都只能被数据拥有者在企业内部分析和使用，缺乏流通、共享，形成大量数据孤岛。这种现象严重地抑制了市场对数据的需求，成为大数据发展的瓶颈。一方面，许多数据拥有者愿意分享他们的数据以获得一定的经济收入，另一方面，数据消费者，如科研人员、数据分析师、应用开发者等，愿意支付一定的费用来获得数据服务。因此需要开放的数据交易平台来促进数据在互联网上的共享和流通，进一步挖掘大数据的经济价值，发现各类数据背后的应用潜力。

由于数据对于市场决策和服务优化的重要性，各个组织机构都想获得有价值的数据资源。数据的流通和交易被认为是一种新兴的商业模式。比如，Xignite<sup>[1]</sup> 公司出售金融行业的数据，Gnip<sup>[2]</sup> 公司出售来自社交网络的数据，Sabre<sup>[3]</sup> 公司则交易旅行用户的订阅和查询信息。为了支持和促进在线数据交

表1 国外代表性数据交易平台

公司名称	数据类型	融资 (百万美元)	创立时间	状态
Factual	通用	62	2008	上线运营
infochimps	通用	5.7	2009	收购 (2013, CSC)
DataMarket	通用	1.2	2010	收购 (2014, Qlik)
Quadrant.io	通用	0.4	2018	上线运营
Azure Marketplace	商业	—	2010	整合重组
Quandl	金融	5.4	2011	上线运营
Benzinga	金融	1.5	2011	上线运营
IOTA	感知数据	—	2017	上线运营
Databroker dao	感知数据	—	2017	上线运营
Terbine	感知数据	—	2013	上线运营
Thingspeak	感知数据	—	2010	上线运营

表2 国内代表性数据交易平台

公司名称	数据类型	融资 (人民币)	创立时间	状态
贵阳大数据交易所	通用	5000万	2015	上线运营
东湖大数据交易中心	通用	6000万	2015	上线运营
华中大数据交易平台	通用	1亿	2015	上线运营
重庆大数据交易市场	通用	1150万	2014	上线运营
上海数据交易中心	通用	2亿	2016	上线运营
数据堂	通用	2.4亿	2011	上线运营
京东万象	通用	60.3亿	2015	上线运营

<sup>1</sup> 数据来源：<https://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>。

易市场的发展，已经有一些国际公司开始提供数据市场的服务，比如微软的 Azure Data Marketplace<sup>[4]</sup>、Infochimps<sup>[5]</sup> 和 Dataexchange<sup>[6]</sup>。国内数据市场的发展也是方兴未艾，贵阳大数据交易所<sup>[7]</sup> 是国内第一家大数据交易所。在此之后，雨后春笋般出现了大小 70 多个数据交易所，如上海数据交易中心<sup>[8]</sup>、武汉东湖大数据交易中心<sup>[9]</sup> 等。数据堂<sup>[10]</sup> 运营了国内第一家大数据电商平台，以电商的形式实现大数据资源的在线交易。京东、百度等公司也纷纷建立了类似的数据共享交易平台。最近，基于区块链的分布式数据交易市场引起了业界的高度关注。IOTA 基金会联合微软等国际公司，利用区块链技术搭建起了物联网数据交易市场<sup>[11]</sup>。类似的公司还有 Databroker Dao<sup>[12]</sup> 和 BAIC<sup>[13]</sup> 等。这些数据市场提供了可用的交易平台，使得数据拥有者能够提交和售卖他们的数据，数据消费者能够寻找和购买他们需要的数据。表 1 和表 2 列举了国外和国内的代表性数据交易平台。

大数据交易市场构建的相关问题近年来引起了业界的热烈讨论。2016 年 5 月，贵阳大数据交易所发布《2016 年中国大数据交易产业白皮书》<sup>[14]</sup>。人民网于 2016 年 7 月梳理了我国大数据的现状和存在的问题<sup>[15]</sup>。2018 年 1 月《光明日报》发表的“2017 年中国智库索引 (CTTI) 系列文章”总结了我国大数据交易产业的七大发展困境，并提供了六大政策建议<sup>[16]</sup>。

然而，大数据交易市场的构建目前仍处于初步探索阶段，还有诸多关键性问题亟待深入研究。

## 数据商品特性

数据商品不同于传统商品和普通电子产品，其具有的新特性为数据交易市场的构建带来了诸多困难。

- **特殊成本构成**：数据一旦生成，就可以被低成本、无损耗地无限复制，一份数据可以同时售卖给多人。数据具有固定的生产（采集）成本，而其边际成本却可以忽略。

- **需求多样、估值困难**：买家对数据的需求

与估值是多样的。由于应用场景差异，不同买家可能需要同一数据的不同子集。比如对于医疗数据，有的买家需要特定病种病人的信息，有的需要某个年龄段患者的情况。数据的价值因应用场景而异，比如 GPS 数据在导航应用中价值较高，在金融征信应用中价值较低，甚至没有价值。数据的价值也与数据的稀疏性有关。对于某些商业金融数据，数据越稀疏，其价值越低。对于政府部门的交通出行数据，涉及到的人数越多，数据价值越高。由于数据应用场景和影响其价值因素的多样化，卖家难以对数据的市场价值进行准确评估，更难以制定合理的数据商品价格。

- **容易伪造、不易验证**：数据是二进制符号（比如数值型传感数据），卖家可以随机地伪造、生成虚假数据，而不是从数据源（传感器）中真实地采集数据，造成数据市场大量的不真实数据。

- **隐私数据敏感**：虽然个人隐私数据能够用来提供个性化服务，但是却不能直接拿来交易。多项实际案例表明，即使不敏感的数据被大量收集后，也会暴露个人隐私。所以在交易隐私数据的过程中需要特别注重隐私保护，需要获得用户的许可才能使用用户数据。在共享交易用户数据的过程中，还要做好数据隐私脱敏和相关性解耦的操作。

- **数据所有权模糊**：个人日常行为所产生的个人数据，所有权毫无疑问属于个人。但数据不同于房子、股票等传统商品，看过即拥有，难以界定清晰的所有权，容易造成盗版数据的盛行，影响数据市场秩序。

- **数据类型多样性**：不同类型的数据具有一些特殊性质。比如对于一些用来决策的数据（商业数据）具有很高的时效性。对于金融数据，具有很强的时间相关性。对于传感器采集的数据，需要考虑其数据质量、精度的不确定性等。

## 数据交易市场关键研究问题

图 1 展示了数据交易市场框架。数据服务提供商将数据采集任务以众包的方式分发给数据贡献

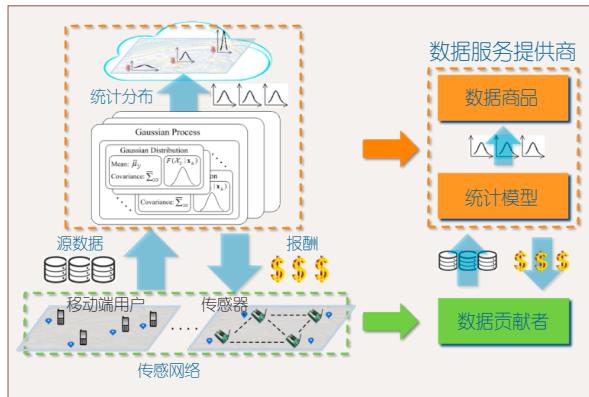


图1 数据交易市场框架

者，以获取大量、优质、可靠的原始数据，再通过质量评估、数据挖掘，并在充分保护数据贡献者隐私的前提下，提炼出有价值的信息，明确数据售卖方式，制定数据价格，最终实现数据无障碍共享流通，发现数据价值，盘活数据资源。

数据交易市场作为新兴研究方向，还有众多待解决的研究问题。我们总结提炼出数据交易市场的三个关键研究问题：数据采购、数据定价、原型系统。希望通过这三个关键问题的研究，为数据市场采集和数据交易定价提供可借鉴的理论依据，为大数据交易提供开放共享的原型系统。

## 数据采购策略

数据是数据交易市场交易商品的原材料。一方面，为了满足数据消费者多样化的数据需求，数据服务提供商需要聚合来自多方数据源的各类数据。另一方面，随着周围环境的变化和时间的推移，固有的数据将失去时效性，变得不准确，甚至产生错误的数据表示。因此数据交易平台需要周期性地向市场补充新鲜的数据，从而提供全面、精准、实时的数据服务。考虑到自身有限的数据采集能力，数据提供商需要利用群体智能的力量，从外部数据源购买数据。众包被认为是采集海量数据行之有效的方法，并且已经部署在实际的数据采集中。数据交易系统需要高效的众包数据采购机制来为数据市场不断补充优质海量的数据资源，从而保证数据交易系统的持续繁荣。

## 研究现状

众包采购平台的首要问题是设计激励机制以吸引足够多的用户参与到众包数据采集中。Lee 和 Hoh 设计了基于动态定价的逆向拍卖机制<sup>[17]</sup>，该机制以众包数据采集平台花费最小化并且保证系统中有足够的数据采集用户为设计目标。然而，该工作并没有考虑数据采集用户在众包平台中可能的操作策略。Yang 等人将用户的策略行为建模成两种不同的博弈模型：以众包平台为中心的模型和以采集用户为中心的模型，并分别设计了基于 Stackelberg 博弈和逆向拍卖的数据采购机制<sup>[18]</sup>。清华大学的杨铮等人考虑了现实中数据采集用户随机出现的情况，并提出了三种在线激励机制<sup>[19]</sup>。

以上数据采集机制的目标主要集中在将社会效益最大化和数据采集酬劳开销最小化两方面，忽略了大数据环境下人工智能、机器学习任务的优化目标。哈佛大学的 Yiling Chen 研究组系统地研究了在策略博弈环境下，针对机器学习任务如何进行数据采集<sup>[20, 21]</sup>。Abernethy 等人为机器学习中的遗憾最小化算法 (regret minimization) 框架设计了真实可信的数据采购机制<sup>[20]</sup>，同时保证了机器学习算法的性能。Waggoner 的博士论文系统地介绍了如何从理性自私的数据采集者中购买、整合信息的理论方法<sup>[21]</sup>。

现有的数据采购方案忽略了数据的时空关联性，只适用于简单的机器学习模型，脱离了数据市场需求。首先，数据之间丰富的关联性使得我们可以通过已有数据推断未知数据，无须采集全部数据，从而降低数据采集成本。其次，新兴的机器学习模型更加复杂，需要与之相适应的数据采购机制。最后，数据交易平台更希望将有限的预算用于采集能够产生更高经济效益的数据集，而不是盲目地采集大量的数据。

## 初步探索

考虑到数据采集者的理性自私策略行为，我们可以将数据采购过程建模成逆向拍卖博弈模型。数据众包采购平台根据应用需求发布数据采集任务，数据采集者提交投标信息来竞争数据采集任务。数

据众包采购平台根据投标信息来分发数据采集任务，并确定数据采集者的酬劳。根据优化目标的不同，可以采用传统的 Vickrey 拍卖酬劳策略（以全局效益最大化为目标）或 Myerson 拍卖酬劳策略（以酬劳最小化为目标）来保证数据采购机制的真实可信。在现实的数据采购过程中，数据需求往往是实时动态变化的，众包平台中的数据采集者通常也是流动的。因此我们需要将静态的逆向拍卖模型进一步拓展为在线拍卖模型，采用在线学习中的竞争比分析 (competitive analysis) 来衡量数据采购机制的性能。在大规模数据市场中，数据众包采购平台往往具有多样的数据采集任务，数据采集者可能同时对多个任务感兴趣。因此，我们还需要将单任务的逆向拍卖模型拓展到多任务的逆向拍卖模型，并采用组合拍卖（多维度机制设计）的思想来保证数据采集者在多维度策略空间上的真实性。

### 待解决问题

上述理论方法虽然在一定程度上能够防止数据采集者的自私策略行为，使数据采购过程有序进行，并能帮助数据众包平台达到一定的优化目标，但都局限于一些理想化的场景，若要投入实际应用，尚有以下四大问题亟待解决。

第一，数据在时间和空间维度具有复杂的关联性，如何挖掘数据的时空关联性，并合理地建模，以指导数据采集？

第二，数据采购目标是以更低的费用采集高质量的数据。数据采集者的数据参差不齐，在没有真实数据的情况下很难对采集的数据进行质量评估。如何将酬劳机制与数据质量衡量有机结合，激励用户真实地贡献高质量数据？

第三，逆向拍卖的目标不仅仅是使社会效益最大化或者酬劳最小化，当前更重要的需求是机器学习训练模型的损失函数 (loss function) 最小化。传统的机器学习模型都假设训练数据是来自可信的第三方，若训练数据是来自理性（自私）的数据采集者，我们应该如何为机器学习模型设计真实可信的数据采购机制？

第四，数据采购和售卖的过程是紧密联系的，

数据交易平台更愿意将有限的预算运用于采集市场需求大、能产生更多经济效益的数据，因此如何设计以市场效益为导向的数据采购机制成为数据交易市场待解决的关键问题。

### 数据定价机制

数据商品具有全新的经济学特性：数据价格取决于数据消费者的数值估值，估值越高，数据价格越高；在不同的应用场景下，不同的数据消费者具有不同的估值。对于数据商品的交易形式和价格制定仍然是经济学领域和计算机领域待解决的基本问题。数据交易平台急需解决的问题是如何确定数据商品的售卖方式，并对这些新型的数据商品进行合理的定价来使交易收益最大化。

### 研究现状

近年来，数据库领域已经涌现出诸多研究关系型数据的定价工作。华盛顿大学 Dan Suciu 教授领导的研究组是这个方向的开拓者，并且在数据交易生态系统项目中研究数据交易市场中的一系列相关工作<sup>[22, 23]</sup>。在他们最早的数据定价文章中<sup>[22]</sup>，Balazinska 等人展望了数据交易市场的前景，并且提炼出数据交易这个方向可能的研究问题。之后 Koutris 等人<sup>[23]</sup>指出工业界中现有数据定价方法的局限性和不灵活性，提出了基于查询的数据定价 (query-based data pricing) 框架，并指出数据定价中两个重要的性质：无套利性 (arbitrage-free) 和无折扣性 (discount-free)。虽然来自数据库领域的数据定价工作关注的大都是关系型数据，但个人数据，包括上网行为数据和移动数据，也已经被众多数据服务提供商采集和分析，并且售卖给其他数据消费者来进行精准营销<sup>[24, 25]</sup>。中国科技大学李向阳教授领导的团队考虑了不可信的数据消费者二次贩卖数据集的问题<sup>[26]</sup>，并将此问题转化成集合相似度的比较问题。他们考虑了多种数据类型的数据交易，包括语音数据<sup>[27]</sup>、视频图像数据<sup>[28]</sup>和图表数据。

数据可以被认为是某种特定类型的信息商品，对信息商品或者是电子商品的定价在经济学和计算机领域也得到了广泛的研究。在文献 [29] 中，作

者提炼出了对于信息服务商品的定价规则，认为有两种有效的策略。其中之一是商品捆绑 (bundling) 销售策略。Bakos 和 Brynjolfsson<sup>[30]</sup> 研究了最优的捆绑策略，并且指出将不相关的信息商品进行合理的捆绑销售能够比分开销售获得更高的利润。另外一个策略是版本划分策略。从文献 [31] 可以看出，很多商家甚至有可能会故意损害他们的商品来实行差异化定价，以此来得到帕累托最优<sup>2</sup>(Pareto efficiency)。Bhargava 和 Choudhary 分析了如何进行版本划分以获得最高利润<sup>[32]</sup>。在数据定价中，在没有买到具体的数据之前，数据消费者无法对数据商品做出有效的估值，该现象被称为非对称信息市场环境。近年来，经济学和理论计算机领域开始关注非对称信息环境下的定价问题，称为信息（结构）设计 (information structure design) 或者信号机制 (signaling)、劝说机制 (persuasion)<sup>[33, 34]</sup>。在博弈环境下，拥有更多信息量的一方通过设计信息结构来引导理性自私玩家向有利于系统总体效益的方向发展。

现有数据市场中的数据定价策略大都是基于经验判断，缺乏相应的理论指导。数据的售卖形式和价格的制定缺乏规范。由于市场信息的非对称性，数据买家对于数据商品很难进行准确估值，难以做出最优数据购买决策；数据卖家也没有相应的机制来学习买家的数据估值，进行准确定价，从而造成数据交易收益的流失，降低了数据卖家和买家参与的积极性。所以需要借鉴微观经济学里的信息设计理论来打破数据交易市场的信息壁垒，并利用机器学习中的在线学习算法预测买家的数据估值，进行合理定价。

## 初步探索

我们从数据售卖方式和数据定价机制两个层面研究非对称信息数据市场下的数据交易策略设计。数据交易首先需要考虑数据以何种方式进行售卖。

在非对称信息数据市场上，数据的交易双方很难对数据商品有准确估值。一方面数据消费者在未购买数据之前无法知道数据内容，因而无法准确估值。另一方面，同样的数据对于不同数据消费者会有完全不同的价值，数据消费者对同种数据也会有不同的质量要求。因此，数据卖家无法知晓数据的市场价值，给数据定价造成困难。然而，数据卖家可以巧妙地设计数据商品的售卖形式来打破这一非对称信息壁垒，通过释放数据商品信号，比如发布免费数据、提供数据展示等方式，让数据消费者了解部分数据信息，辅助其对数据进行准确估值。数据卖家还可以将数据商品划分为不同版本，每个版本拥有不同的质量和价格，比如推出不同数量的 API 查询，并收取不同的费用。数据消费者可以选择适合自己需求的数据版本。数据卖家根据数据消费者选择的版本，可以间接地了解到其数据估值。在确定数据售卖形式之后，需要进一步考虑数据定价问题。经济学领域的定价策略基本都是基于贝叶斯假设，也就是数据卖家根据历史交易信息统计出市场数据估值的概率分布函数，基于估值概率分布函数，计算出最优收益下的价格取值。然而现实中新投入市场的数据商品的定价策略通常无先验分布知识可以借鉴，而只能利用在线学习的思想，在探索 (explore) 和利用 (exploit) 之间做权衡，动态地探索出数据市场估值，并利用该估值信息设定收益最大化的价格。

## 待解决问题

上述方法虽然已经被运用于数据交易市场的数据商品定价，但是大部分数据卖家都是简单地套用，其背后的理论机理还没有被深入探讨。为了能够指导实际数据市场的定价，我们还需要解决如下三大问题。

第一，在确定数据售卖方式的过程中，需要设计出高效的机制来确定需要发布多少免费数据，决

<sup>2</sup> 这个概念以意大利经济学家维弗雷多·帕累托 (Villefredo Pareto) 的名字命名，指资源分配的一种理想状态。假定固有的一群人和可分配的资源，从一种分配状态到另一种状态的变化中，在没有使任何人境况变坏的前提下，使得至少一个人变得更好。帕累托最优状态就是不可能再有更多的帕累托改进的余地，是公平与效率的“理想王国”。

定是否推出数据展示，需要将计算数据划分为多少个版本，决定每个版本的数据内容与质量等。

第二，无论是经济学领域还是计算机领域，现有的定价技术都无法适应动态市场变化下的数据定价问题。数据消费者的数据估值会随着数据时效性而动态波动，如何设计适应市场环境变化的在线学习机制与动态定价机制？

第三，已有的定价技术忽略了数据消费者可能的策略性购买行为，比如套利行为与估值信息谎报行为。我们需要明确数据交易中消费者可能的策略行为，并设计防套利性的数据定价机制。

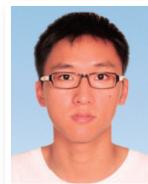
## 开放数据交易平台

虽然目前国内外已存在众多的数据交易平台，但由于商业保密、隐私问题和数据版权等因素，现有的数据市场都是闭源经营，无法为相关的数据交易研究提供可验证性平台。因此，我们急需搭建开源的验证性数据交易原型系统，以实践检验各种理论方法的可行性，发掘系统运行中出现的实际问题，指导理论研究的方向。

目前，我们已经初步实现了基于微信平台的众包数据采集系统，并收集了一定量的环境感知数据和用户行为数据，这为开展采购和定价方面的研究提供了数据基础。同时，我们基于现有的开源区块链数据市场，正在开发科研数据交易系统，并计划开源该系统上的交易细节，以促进数据交易相关研究的开展，用于检验各种理论方法在实际运行中的优势与不足，并从中发掘进一步的研究问题。根据现阶段的初步调研，我们发现基于区块链的数据交易市场能够在一定程度上解决数据隐私保护、数据确权、数据溯源和数据真实性验证等问题。 ■

### 致谢：

本研究工作得到了国家重点研发计划“云计算和大数据”重点专项(2018YFB1004703)、国家自然科学基金(61672353, 61672348)的资助。



郑臻哲

CCF专业会员。上海交通大学计算机科学与工程系博士后。主要研究方向为算法博弈论、云计算、无线网络。  
zhengzhenzhe220@gmail.com



吴帆

CCF专业会员。上海交通大学计算机科学与工程系教授。主要研究方向为网络经济学、无线网络、移动计算、隐私安全。  
fwan@cs.sjtu.edu.cn



陈贵海

CCF会士。上海交通大学计算机科学与工程系教授。主要研究方向为分布式计算、计算机网络、并行计算。

## 参考文献

- [1] Xignite[OL]. <http://www.xignite.com/>.
  - [2] Gnip[OL]. <https://gnip.com/>.
  - [3] Sabre[OL]. <https://www.sabre.com/>.
  - [4] Microsoft Azure Marketplace[OL]. [https://datamarket.azure.com/..](https://datamarket.azure.com/)
  - [5] Infochimps[OL]. <http://www.infochimps.com/>.
  - [6] Dataexchange[OL]. <http://new.thedataexchange.com/>.
  - [7] 贵阳大数据交易所 [OL]. <http://www.gbdex.com/>.
  - [8] 上海数据交易中心 [OL]. <https://www.chinadep.com/>.
  - [9] 武汉东湖大数据交易中心 [OL]. <http://www.chinadatatrading.com/>.
  - [10] 数据堂 [OL]. <http://www.datatang.com/>.
  - [11] IOTA 区块链数据交易市场 [OL]. <https://data.iota.org/>.
  - [12] Databroker Dao[OL]. <https://databrokerdao.com/>.
  - [13] BAIC[OL]. <http://baic.io/>.
  - [14] 2016年中国大数据交易产业白皮书 [OL]. [http://www.cbdio.com/BigData/2016-06/02/content\\_4965656\\_all.htm](http://www.cbdio.com/BigData/2016-06/02/content_4965656_all.htm).
  - [15] 人民网 . 我国大数据交易亟待突破 [OL]. <http://theory.people.com.cn/n1/2016/0705/c367658-28526453.html>.
  - [16] 大数据交易：产业创新与政策回应 [OL]. [http://views.ce.cn/view/ent/201801/25/t20180125\\_27899352.shtml](http://views.ce.cn/view/ent/201801/25/t20180125_27899352.shtml).
- 更多参考文献：<http://dl.ccf.org.cn/cccf/list>

# 以“作品文化”取代“帽子文化”

关键词：作品文化 创新

胡包钢

中国科学院自动化研究所

“帽子文化”终于开始被摈弃了。从2017年9月7日，全国政协第73次双周协商座谈会以及2017年9月16日“CCF YOCSEF论坛‘帽子文化’的利与弊”中各位人员的讨论，到最近国家各部门决定开展清理“唯论文、唯职称、唯学历、唯奖项”（简称“四唯”）的专项行动<sup>1</sup>，反映了文化与制度的双重进步。“帽子”是人才考查后的衡量结果，而问题出在“帽子”形成了一种“评价文化”，成为“通用规则”后引起了诸多弊端。例如，学校与部门围绕“帽子”大做文章，以此来争夺科技资源。而个人则在取得“帽子”后被赋予了永久优势地位，年轻人的价值取向被严重误导。

“帽子文化”本身同时包含了文化与制度层面的问题。考虑到观念永远走在制度前面，破旧立新中应该有替代“帽子文化”的说法。因此，本文提出“作品文化”，并以此展开观念层面上的讨论。

所谓“作品文化”是指围绕作品而形成的思维方式、价值观念、生活习惯等内容。在此，可将“作品”大体解释为具有创作性且以某种形式表现的成品。该说法不仅包含了创新的内涵，而且强调落地时的成品形式。它还可以扩充到“创新文化”“品牌文化”“代表作文化”“精品文化”或“工匠文化”的提法上。从广义上讲，它适用于各行各业，个体或群体，专家或学生。因此这个作品可以以各种形

式出现。作品与帽子之间可以按植物生长规律解释为根叶与花果的关系。无有根叶，焉存花果？“作品文化”的提法是让人明白人才成长的规律。

“作品文化”的说法有利于我们对作品本身或内涵予以关注，而且关注点会自然导向“质”而非“量”。“作品文化”无法避免滥竽充数作品的出现，然而它为那些没有资历的优秀青年人脱颖而出搭设了最好舞台。这反映了中国俗语“不怕不识货，就怕货比货”中的道理。“作品文化”隐含了何为“好”是其关注的研究主题，具体工作中也会以此为重点。我们都应该知道好的作品是需要卓尔不群、精雕细琢的。它能够引导我们更安心地在作品上下功夫。“作品文化”同时为多元化评价提供了空间，讲究的是影响力。具体是论文还是软件没有关系。而“帽子文化”更易诱导人走歪门邪道。那些弄虚作假的名人多因掉入这样的陷阱而身败名裂，历史教训令人警醒。

“作品文化”中的典型人物及具体事例不胜枚举。中国伟大的史学家司马迁，在惨遭宫刑后历经十数年完成《史记》。正是强烈的作品意识中“欲以究天人之际，通古今之变，成一家之言”的理念使他完成这部宏伟作品。我们熟知的许多中外科学家或工程师，都是以他们的作品称道。这个作品可以是概念、公式、定理、原理、方法、仪器、工艺等各种形式。牛顿与爱因斯坦分别因为三大定律和

<sup>1</sup> 2018年10月23日，科技部、教育部、人力资源社会保障部、中国科学院和中国工程院联合发布了《关于开展清理“唯论文、唯职称、唯学历、唯奖项”专项行动的通知》。《通知》指出，根据《中共中央办公厅、国务院办公厅关于深化项目评审、人才评价、机构评估改革的若干意见》和《国务院关于优化科研管理提升科研绩效若干措施的通知》要求，决定开展清理“唯论文、唯职称、唯学历、唯奖项”专项行动，并明确了涉及“四唯”做法的具体清理范围。

相对论而著名。乔布斯留下的可贵精神财富是用具体作品来推动这个世界进步。

中国走向“作品文化”还会有很长的路要走，摈弃“帽子文化”也不会一蹴而就。这种改变不仅是科技界的事，全社会都会涉及。我们曾抱怨中国社会所谓冒牌“大师”横行，但这些“大师”为什么会广受追捧，又总能层出不穷？这有其文化根源。中国先贤的“十年树木，百年树人”，见解十分深刻。一代人才长成二十年足矣。百年之说可能意指改变人的观念、习惯或风俗。对于邻国日本出现那么多位诺贝尔奖获得者，我们要明白其背后的原理与逻辑，这绝非是偶然现象。我们要学习他人之长，如认真办事、注重细节、尊重知识产权等等的文化。中国应该在创新文化的“土壤”方面下功夫，比如在计算机教学中，早就应该加入知识产权、人物历史、计算机伦理等方面的内容。针对中国现今社会发展，笔者认为以下说法当被提出：

“大师时代已然逝去？当下唯有作品为重！”

我们通常理解的“大师时代”是包括古典音乐，

印象派画作，相对论与量子力学这种由数位大师级人物引领的特定时期。上面第一句话用问号结尾为个人自我疑问思考展开了空间。在学术发展背景下，笔者的个人见解是，各时代总有大师级人物需要人们向其学习，但是以其贡献称为时代的说法似已不在。不要期待宗师名人引领，更不要被帽子名气吓倒。第二句话乃为本文核心观点：用作品说话。

“我劝天公重抖擞，不拘一格降人才”。如何培养或评价人才永远是我们需要探讨的话题，其中包括中国创新文化的传承与发展。“作品文化”拟可成为创新文化中的一种见解。对于从事科研工作的我们来讲，需要不断自我思考的问题是：我们的团队或个人的科研成就品牌应该是什么？ ■



胡包钢

CCF专业会员。中国科学院自动化研究所研究员。主要研究方向为机器学习与植物生长建模。

hubg@nlpr.ia.ac.cn

CCF TC

CCF 计算机视觉专业委员会

## 第一届中国模式识别与计算机视觉大会召开

第一届中国模式识别与计算机视觉大会（PRCV2018）于2018年11月23~26日在广州召开。会议由中国人工智能学会、中国计算机学会（CCF）、中国自动化学会、中国图象图形学学会联合主办，注册参会人数超过1600。

CCF计算机视觉专委主任谭铁牛院士出席会议并致辞，中山大学校长罗俊院士致欢迎词。谭铁牛、西安交通大学教授郑南宁院士、北京大学教授查红彬共同担任大会主席。大会邀请美国伊利诺伊大学厄巴纳-香槟分校教授David Forsyth、加拿大约克大学教授Michael S. Brown、美国北卡莱罗纳大学教堂山分校副教授Tamara Berg、腾讯机器人实验室主任张正友博士分别作了题为“The Materials Objects are Fashioned from”“From RAW to sRGB and Back: Modeling the Digital Camera Pipeline”“Words—Pictures”“智能机器人和有情商的人机交互”的主题报告。

大会举办了9个专题论坛、8个专题竞赛、6个讲习班，并设立了顶会顶刊论文交流、应用展示等环节。专题竞赛共有780支队伍参赛，49支队伍获奖，奖金高达107.5万元。

CCF计算机视觉专业委员会工作会议同期举行，CCF常务理事、中国科技大学教授吴枫和CCF专委工委委员、中科院计算所研究员蒋树强代表总部参加了会议。



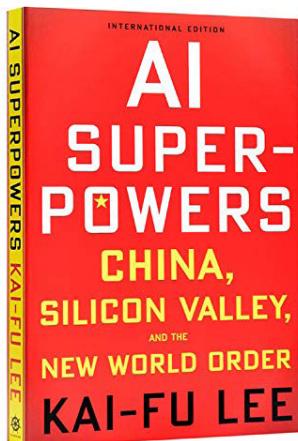
## The CS David专栏

CCCF 2019年第1期

# 《AI·未来》

关键词：人工智能技术 未来

李开复的新书《AI·未来》(*AI Superpowers: China, Silicon Valley, and the New World Order*)从三个截然不同的角度讲述了人工智能最近所取得的成



李开复的新书《AI·未来》

就。就像英语世界里常见的书籍一样，这本书以谈论科技与国家民族的力量开篇。但紧接着话锋一转，开始“聚焦”在李开复本人身上，变成了关于他性格的自白，或者说自传。这本书的最后评论了计算机科学家群体，或者至少是从事人工智能研究的科学家群体。在书中作者发问：计算机科学家实现其目标的方式是否对所有人都有益？

我相信很多读者比我更了解李开复博士。如

果你对他一无所知，我可以做一个简要的介绍：李开复是总部位于北京的一家投资基金——“创新工场”的首席执行官，也是自然语音识别领域的权威专家。早年间他在卡内基梅隆大学(Carnegie Mellon University, CMU)取得博士学位，语音识别这个研究领域最早就起源于卡内基梅隆大学。值得一提的是，李开复还是IEEE Fellow。

这本书是一位朋友向我推荐的，他说书中探讨了国家力量。他告诉我说这本书讲述了中国将如何成为人工智能领域的主导，并将如何在未来支配和控制世界经济。“你会喜欢它的”，他说，“你一向喜欢那些探讨国家如何因科技而强大的书。”

我很快发现其实我的朋友根本没读过这本书，即使读过也仅仅是前面几章。这本书一开始确实如他所言，从一个简单的论断开始，即人工智能将成为未来的重要科技。书中说，中国和中国的研究者将在人工智能领域占据主导地位，作者认为目前中国的研究者已经开始取得和欧美同行一样的地位了。这本书介绍了中国政府如何通过诸如2006年开始的“十五年创新规划”和最近的“中国制造2025”计划资助相关的研究者。作为其论点的佐证，作者强调了中国对创新文化的需求和巨大的国内市场。

如果这就是这本新书的全部，那么它就和

三十五年前出版的费根鲍姆(Feigenbaum)和麦考达克(McCorduck)所著的《第五代计算机》(*The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*)的论断非常相似。那本书聚焦的是三十五年前的日本，并就当时的人工智能技术做出了类似的结论。当时的人工智能技术现在通常称为“专家系统”或者“符号推理系统”。作者声称这些技术在未来极为重要，而日本已经采取了一系列重大举措来推动这些技术，并发展与之配套的高性能计算机。作为结论，作者在书中断言日本很快就会主导世界经济。

《第五代计算机》在美国社会引起了强烈的反响，当时许多美国人对日本日益增长的经济实力忧心忡忡，他们觉得日本经济自我封闭，但却极大地受益于西方的开放式经济。他们非常担心日本政府所主导的计划将使其在一些关键领域超越美国。相关话题的讨论出现在报纸上，以及国会的辩论中。最终国会批准了多项计划以推动计算机的发展，其中最具影响的一项是建立超算研究中心。

《第五代计算机》出版十年之后就不再具有影响力。日本政府资助的计算机发展计划没有取得重大突破，其资助的人工智能研究也未能像预期的那样取得领先业界的丰硕成果。而作为计划的第二部分高性能计算机的研制相较前者要成功一些，但与美国同行相比也不具有太多竞争力。此外，日本经济增长速度放缓，使其失去了在国际市场的竞争力。此后《第五代计算机》逐渐淡出了人们的视野。

有了《第五代计算机》的经验，我们在看待李开复博士的论断时要更加谨慎。费根鲍姆和麦考达克的观点在当时看起来似乎和李开复在书中的观点一样令人信服，《AI·未来》中的观点是否也会和《第五代计算机》的论断遭遇同样的命运？李开复博士似乎也非常清楚这个问题。他提到在过去的12~15年间，人工智能研究者不断地改良原有的技术，但鲜有革命性的突破。在这个过程中他们创造了许多应用广泛且颇具价值的技术，但是这些成功仅仅发生在狭小的知识领域。他们还不能让计算机捕获并学习一般性的知识或常识。从很多方面来看，现在

我们在通用人工智能领域与1983年相比并没有前进太多。

随后，李开复突然将焦点切换到了一个新的话题。他开始思考那些因计算机，特别是人工智能技术的普及而消失的工作岗位。他简要地探讨了这样的观点：科技所创造的就业岗位要多于它所摧毁的岗位。实际上我们并没有掌握充足的数据以作出准确的判断。当一项新技术被引进时，人们通常可以观察到与该技术直接相关的工作岗位的增减，而往往无法了解全局的情况。我们常常会忽略掉包括供应链、消费者市场、处理方式的变更、支撑行业以及培训教育职位的增减等间接因素。

例如上世纪七八十年代，信息化预定系统的引入从根本上改变了航空工业。一方面，它创造了更多的信息技术岗位，每家航空公司都需要聘请通讯专家，所有涉及到顾客服务的人员都需要更多的培训。此外它还创造了一个新的产业，为了提升航班的客座率出现了代理销售折扣机票的公司。另一方面，它摧毁了“旅行代理商”这个职业，他们帮助顾客预订航班，制定出行计划。总的来说，这项技术创造的岗位要多于因它的出现而消失的岗位，它从根本上改变了旅行行业对人力资源需求的结构。当新岗位取代同等薪水的旧岗位时，需要大量额外的教育和培训工作；而替代一个具有同等职业技能的旧岗位时，新岗位通常会降低薪水。

这些关于岗位重构的问题让李开复开始怀疑自己领域所取得的成就，当他在面对危及生命的疾病之时他开始思考这些问题。和之前许多人一样，突如其来变故使他反思自己的价值观，并思考自己是否在做对的事情。他承认自己或许太专注于自己的工作，应该更多地关注家庭，花更多时间陪伴家人。

沿着这一观点，他对计算机领域提出了同样的问题：我们是否过于专注于技术发展而忽略了满足周边人的实际需求。在书中李开复讲述了他在进行精神康复时惊讶地发现其中一个志愿者竟然是一家大型科技公司的首席执行官。这名首席执行官解释说他参与志愿者活动是因为他希望得

到一种不索取任何报酬而服务他人所带来的自我满足感。

在此后的章节中，李开复追问如何才能使计算机科学更加无私？是否可以以更加关注人类福祉的方式来发展科技？是否存在某种方式，使得计算机科学家在服务一些人同时，不会影响另外一些人的生计，不会使他们失去原有的特质和奋斗的目标？这些都是对计算机科学家更高尚情操的呼唤，呼唤人们摒弃私欲而追求公益，克制物欲而拥抱理想，挣脱现实的束缚而追求永恒的价值。不只是对计算机科学家，对于任何群体而言，这都是一个可敬的目标。然而李开复很快就遇到了长久以来一直困扰着全世界先哲的问题：如何构建经济和社会的强制力，以便使人们更少地在意财富和权力？难道必须给人们一个自私的动机才能使他们变得无私吗？

李开复考虑了这些问题并提出了几个方案，但最终认识到他没有答案。他清楚地看到，正在蓬勃发展的计算机和人工智能技术将改变人民和国家的财富和地位。他在书中明确地指出，我们需要更好的方式去发展和应用这些技术，他也希望我们最终可以找到这样的方式。对于计算机科学家而言，李开复的新书发人深省，书中所探讨的问题远远地超越了其标题和开篇章节中所表述的观点。 ■

#### 戴维·阿兰·格里尔 (David Alan Grier)

电气与电子工程师协会计算机学会 (IEEE-CS) 前任主席、IEEE Fellow (会士)。乔治·华盛顿大学名誉教授。他目前是华盛顿特区 Djaghe LLC 公司的技术总监。IEEE Computer 杂志编辑 (2019)。grier@gwu.edu

#### CCCF 特邀译者：

孙晓明 中国科学院计算技术研究所研究员

## CCF 推广双导师计划 已落实首批入选院系

鉴于有些院校计算机专业师资力量不足，影响研究生的培养质量，而有些院校优质师资力量充足但研究生生源不足，CCF于2018年5月开展了“双导师制研究生联合培养计划”(简称CCF双导师计划)，为两方面的导师及研究生搭建资源共享、合作培养的平台，并提供相应的支持。

该计划已征集并遴选了来自国内高校、研究所和企业的100余位优秀导师，并在全国范围内遴选部分试点院系参与。每个院系在项目执行期内（暂定三年），每年组织不少于5名研究生（硕士生或博士生）申报双导师计划，经导师与学生双向选择后，开展双导师联合培养。参加本计划的研究生由原学校颁发毕业证、学位证，CCF颁发结题/结业证。

经专家组讨论审核后，部分院系（见附表）和学生通过遴选，成为CCF双导师计划首批试点单位和成员。

#### 附：入选试点院系

学校	院系
西安理工大学	计算机科学与工程学院
云南大学	信息学院计算机科学与工程系
哈尔滨工程大学	计算机学院
桂林电子科大	计算机与信息安全学院
太原理工大学	大数据学院
华中农业大学	信息学院
安徽大学	计算机学院

学校	院系
武汉理工大学	计算机科学与技术学院
南京邮电大学	通信与信息工程学院
长安大学	信息学院
广东工业大学	计算机学院
航天二院研究生院	706所
湘潭大学	信息工程学院

## YOCSF 视点

# 科研评价：破“五唯”，立什么？

近日，教育部办公厅根据《科技部、教育部、人力资源社会保障部、中国科学院、中国工程院关于开展清理“唯论文、唯职称、唯学历、唯奖项”专项行动的通知》(国科发政〔2018〕210号)的要求，印发了《关于开展清理“唯论文、唯帽子、唯职称、唯学历、唯奖项”专项行动的通知》，决定在各有关高校开展“五唯”清理。此次破“五唯”行动，引发人们高度关注。

“五唯”痼疾近年来屡屡牵动科教界，属社会敏感问题。“五唯”评价标准的特征是简单化、片面化、绝对化，助长了科教界的浮躁风气，给科研和教育带来了诸多负面影响。

然而，从我国科研评价体系发展的历史来看，我国科研评价体系是从“主观评价”逐渐过渡到“客观评价”的过程。论文、学历、帽子等客观评价标准，在打破人情关系、调动科研人员积极性等方面发挥了重要作用。但破除“五唯”，会不会再回到人情关系和学术权威主导的老路？破除“五唯”，我们应该用什么样的标准进行人才和科研评价？甚至，是否存在科学合理、具有普遍适用性的科研评价客观标准？

针对这些问题，中国计算机学会(CCF)联合中国科协第九届常委会人才与继续教育专门委员会，以及中国科协培训和人才服务中心，于2018年12月15日在北京主办了 YOCSF 专题论坛——科研评价：破“五唯”，立什么？论坛邀请了全国人大常委、原中国科协副主席、北京理工大学教授冯长根，CCF 海外杰出贡献奖获得者、美国俄亥俄州立大学教授张晓东，做引导发言。同时邀请 CCF 秘书长杜子德、清华大学教授史元春、浙江大学教授卜佳俊、中科院计算所研究员韩银和作为论坛嘉宾，一起参与讨论互动。CCF 副秘书长、YOCSF 秘书长唐卫清，YOCSF 主席、清华大学教授唐杰出席。

## “五唯”的破与立

冯长根认为，论文引用量水分太多，有争议，而我们的学者们恰恰是被论文引用量绑架了，甚至可以追溯到各位学者的青年学生时期。现在的政策就是要把学者们解放出来。因此，需要新的学术论文评价方法和思路，那就是学术评论句，而学术传承也在其中。学术评论句应该有三个要素：作者姓名、年份、标志词，表明论文学术影响力。倡导学者们在做研究时，学术论文及其研究要有创新点；在撰写论文时，要写学术评论句，要先评大家，才能被大家所评价。所谓“五唯”，就是评优秀的方法，其出现有历史原因。没有“五唯”，青年学者们也很难脱颖而出，但同时像论文引用数影响到科研投入，和科研经费有关，这就是功利。所以要辩证地看问题，才能找到问题的真正根源，排掉根源。但我们要有思想准备，“五唯”可能会以别的形式出现。任何一个评价体系，或者是评价的标准，都有弊端。学术评论句也有弊端，即什么样的评价是合格的评价？是不是指长时间内的评价？对于新兴研究领域的短时间评价是否合适？

张晓东认为，计算机科学的发展要有里程碑的工作。“五唯”，是国家规定的，立也是国家立的，像“唯学历、唯帽子、唯奖项”是国家自上而下立的。国家定义的评价体系也是稳定的：学位需要论文；帽子需要学位、奖项、论文、职称；奖项要学位、帽子、职称等等。本质上说，只有发表论文是在体系评价之内。现在要破“五唯”，是针对评价体系，根子还是在论文，因为发表论文绝对不是目的。发表论文的影响才是学术的重要内容，这可从三个层面来看：第一层，作为相关工作被引用；第二层，是否有创

新和颠覆，是否最后可以进入教科书；第三层，创新的技术是否直接影响国际界。可见，对论文的评价出现了问题，才导致对整个评价体系的质疑。对论文的分级，可以有如下的评判：A+论文，就是工作得到同行认可，或者是很快就可以推进，或者是修改已知的理论；B论文，同行认可，或者是顶级会议接受，但是没有意愿去推进。C论文，自己认同，但是别人没有兴趣，也不太听得懂。不及格论文，自己都半信半疑，别人也不太懂。国内和西方做研究不一样，国内强调手艺，西方强调学问、普适的工作，即所谓的科学精神。“五唯”在国内有深厚的历史、社会和文化的基础，像各种各样的帽子，其实是一种社会身份，是国人所追求的。还有一个潜规则，要想当官一定要拿这些帽子。现存体制异常强大，非常一律化，独立之精神、自由之思想、批判之思维，不能融入我们生活。如爱因斯坦所言：“扼杀个人独立性和多元化的社会，是毫无发展可能的。学校必须以培养能独立行为和思考的个人为目标，而这些个人又把为社会视为他们最高的生活目标。”俄亥俄州立大学评价学者是看成为世界级学者的潜力，这有三项指标：在领域内是否发表最高级别的论文；在业界中的杰出贡献；有量化指标。

**史元春**认为，论文作为评价指标是合理的，不同方向应该区别对待。唯计量论文的成绩，基本不分领域，这才是要去的东西，而不是去论文，一定要有学术自信。至于学历，以清华大学计算机系聘用一位只是中专毕业的老师做教授为例，一个学术单位要有自己的学术判断，这是要“立”的准则之一。

**卜佳俊**的观点是，如果我们没有比“五唯”更好的评价体系，不应该取缔“五唯”。实践证明，“五唯”本身比较公正、比较符合科学规律，只是在执行的过程中有些偏差。例如基础研究，论文是非常重要的，如果没有高水平的论文发表，很难说基础研究水平比别人高。大部分的基础研究，大部分的理论突破，还是在高质量的论文里面。学历和帽子也很重要，总体而言，学历高的人 / 有帽子的人比学历低的人 / 没有帽子的人对社会的贡献更大。在这样的一个体系里面，怎么样才是好的论文，怎么

样才是好的科研成果，怎么样才是真正好的能够传承的成果，非常重要。

**韩银和**则认为，不破不立。破“五唯”，不是不看这五方面，而是“不唯”。我们反“五唯”，暗含着我们会反其他的东西，例如反海外背景、反唯课时论等。如果要“立”，首先要分类。不分类的评价是今天有“唯”的很重要的核心，应该避免行政评价、综合学科评价，可以采用大学科的评价方法，把同行评议做得更公平公正；其次要结合国际化，可以接受一部分的国际同行评论；第三，把一些评价的事情，比如奖励，交由一些更有公信力的组织来评，将行政和评价分开，这样也许会建立一套更合适的体系。

**包云岗**说，美国民间组织设立了很多奖项，整体得到了非常高的认可，甚至是不低于美国政府颁发的国家奖。张晓东认为，社会体制关系很大，很多社会体制是一个自下而上的，根基很深，例如美国国家科学院，是个私人企业，不是政府的，它是一个科学共同体，经常向政府进言帮助政府工作。

## “五唯”对科研的发展造成了阻碍

大家在讨论中认为，(1)各行各业统一标准，对科研工作者发展不利，典型例子是唯论文，对一线的医生们是不公平的；(2)人才引进时，唯论文，甚至唯论文列表，对科研工作者的评价不够公平；(3)没有“五唯”的驱动之前，科研方面我们国家有很多重大突破，有了“五唯”，反而没有了，说明“五唯”评价体系已经不适应于我国科技发展；(4)在“五唯”评价体系下，一度出现科学研究是致富路，教书育人则坐冷板凳；(5)不被“五唯”评价体系看好的科研成果，却成了企业竞相争抢的重要资源……

破“五唯”，破的是“唯”，而不是“五”；量化不一定就是“唯”，主观上如何看待量化指标，是值得探索的问题；“五唯”的正面和负面效应应该有数据支撑，才更有说服力；如爱因斯坦所说，如果不是为了生存而做科学，那么科学是最伟大的，在一个非常纯洁的环境下追求科学，要有一种自由

和独立的精神，这是对科学学者，特别是年轻学者最根本的要求，如果没有这种精神，追的这些东西变成非常的世俗，就不会有里程碑般的科研成果。

## 在实际操作中能否真正落地

(1) 要敢于批评教育部的错误做法，形成共同的价值观和做法，中国的教育和科研是有希望的。(2) 要对我国的行政系统充满信心，不要质疑和怀疑它的执行能力，相信在外界力量的推动下，是可以转变评价观念的。(3) 很多时候，科研工作者不得不利用“五唯”去争取行政力量所分配的资源，但是现在资源主体多元化，得到企业的认可，也可以开展科研。(4) 破“五唯”，根本问题是评价机制可以改进。(5) 形成“五唯”的原因是唯上，其本质是唯权和唯利，如果教育部把评估的权利仍然牢牢掌握在手中，破“五唯”政策就很难落到实处。(6) 破“五唯”被国家提出后，教育部、科技部的学科评估和评奖等，并没有停，这会让科研部门还是按照原有方式进行发展。(7) 人才帽子等带来的科研激励，或者经济收入，在实际上已形成了一种非良态的循环，很难破除。(8) 当下还破不了“五唯”，还有太多问题亟待解决，例如，评价的机制怎么评出来的，去一些什么样的指标，除了帽子和论文之外，还有没有一些指标作为立的依据等。

## 破“五唯”，立什么？

采纳学术评论语是可行的，但是要有数据支撑，这需要进行试验性探索。深化同行评议的同时，还要严防人情问题，这会面临基础保障层面的一些挑战。建立独立培养体系，在当下稳定的培养体系下很难培养杰出的民族科学家，需要一套独立的体系，才有可能实现。分门别类地制定一些标准。不能简单化和片面化。要注重资源分配原则的改变，建设客观评价指标的指挥棒。可选一些高校做迭代优化的试点。学术共同体要有担当，引导行之有效的评价体系。可建立三方面的评价体系：行政管理评价、第三方评价和受益者评价，等等。

无疑，破“五唯”有很强的迫切性。“五唯”一定要破，但在破“五唯”的过程中，又面临着非常多的挑战，现在是行政主导，利益格局已经形成，怎么突破当前的利益格局？我们国家并不是没有同行评议，正是因为我们把同行评议转化成了一个量化机制，才给了年轻人一些发展的机会，而“五唯”的出现，又走向了另一个极端，怎样建立公平公正的评价体系，是一个很重要的值得考虑的问题。未来的评价要更加重视真才实学，重视真正高质量的贡献。在具体评价的过程中，要强化学术共同体在制定学术评价标准方面的主导作用。

（董笑菊 崔鹏 王栋 唐杰）



## CCF 走进高校(2018 年)

序号	演讲人	时间	高 校	演讲题目
678	曹健 谭文安 卢瞰	12月7日	上海电机学院	从协同过滤到跨领域推荐 现代服务工程学科前沿讲座——随需应变的软件服务工程技术 跨社交媒体用户交互关系与行为的分析与理解
684	张加万 肖丽	12月7日	西南科技大学	可视分析及其应用 面向大规模科学与工程计算的可视分析引擎
685	王忠杰 刘譞哲 王尚广	12月8日	河海大学	从图灵奖得主 Tim Berners-Lee 和 Raj Reddy 的近期观点看服务的个人化趋势 AI 时代人机融合的智能软件服务 服务计算那些事儿之服务组合
686	侯宇涛	12月11日	河北大学	深度学习入门——使用开源免费软件 DIGITS 实现手写体数字图片分类

# 新一代计算机体系结构 特邀研讨会回顾

钱德沛<sup>1</sup> 陈文光<sup>2</sup> 范东睿<sup>3</sup> 毛睿<sup>4</sup>

<sup>1</sup> 北京航空航天大学

<sup>2</sup> 清华大学

<sup>3</sup> 中国科学院计算技术研究所

<sup>4</sup> 深圳大学

关键词：新一代计算机体系结构 人工智能与大数据应用

计算机体系结构的发展正面临三个重要的趋势：在底层器件层面，半导体工艺进展的速度已经大大降低，摩尔定律正在走向终结；在需求层面，以人工智能为代表的新型应用对计算机体系结构提出了新的挑战；在工具和生态层面，近年来以 RISC-V 为代表的开源体系结构和以 Chisel 为代表的高级综合语言等技术显著降低了体系结构设计和实验的入门门槛。当前正是计算机系统研究领域发生巨变的时刻。

在此背景下，由钱德沛、陈文光和范东睿等组织的“新一代计算机体系结构特邀研讨会”于 2018 年 6 月 3~4 日在北京召开。会议的宗旨是：在计算机芯片和系统结构技术面临重大变革的关键时刻，通过本次会议碰撞出新的思想，为未来技术发展提出建议。

来自国内主要计算机体系结构研发单位的专家学者参加了本次研讨会，包括中科院计算所李国杰、孙凝晖、徐志伟、陈云霁、范东睿、谭旭、叶笑春，国防科技大学卢锡城、廖湘科，深圳大学陈国良、毛睿，北京大学高文、罗国杰，北京理工大学梅宏，清华大学郑纬民、陈文光、张悠慧、渠鹏，北京航空航天大学钱德沛、刘铁，国家并行机研究中心谢

向辉，华中科技大学金海，美国俄亥俄州立大学张晓东，上海交通大学过敏意、陈海波、陈全等。

会议围绕人工智能及大数据应用、生态和系统软件、芯片研发和设计、程序执行模型、多学科合作研发设计，对计算机系统的挑战与影响展开了讨论。

## 人工智能与大数据应用对计算机体系结构的挑战与机遇

人工智能与大数据是新兴的重要应用类型，其突出特点是展现了海量的并行性。目前的通用 CPU 在处理串行应用时比较高效，但在处理海量并行应用时效率很低，GPU、FPGA 以及各种 AI 加速器虽然能够处理海量并行应用，但是片上内存的限制和现有软件系统（如 Spark）的低效使得现有的异构结构（CPU- 加速器结构）面临处理问题规模有限、CPU 与加速器之间数据传输成为瓶颈的挑战。另一方面，很多人工智能与大数据应用在访存模式上具有只读数据远大于可读写数据的特点，为新的计算机体系结构和系统软件提供了优化机遇。

与会者提出软件层面的改进可以比现有系统

(如 Spark) 的性能提高 100 倍, 而专用 AI 芯片可以比通用处理器快 10000 倍; 实际应用需要“波动小、低熵”的高品质计算机体系结构, 应使用标签化体系结构解决体系结构中由于共享资源而引入的尾延迟问题, 从而显著提高计算资源的利用率。

## 构建包容的计算机系统生态环境

现有软件生态环境以通用 CPU 为中心, GPU 等加速部件被作为外设管理。然而在新型专用处理部件层出不穷的情况下, 现有软件生态环境在编程模型、执行模型和进程管理等方面不匹配的问题不容忽视。模块化和可组合性应该在软件系统的高速演化中发挥根本性作用。

在寻找解决方案时, 可以参考计算机网络领域的经验, 即以 IP 层为“细腰”的计算机网络体系结构构建了一个包容的生态环境, 很好地隔离了底层网络技术与应用层技术。在计算机系统生态环境中, 指令集体系结构 (Instruction Set Architecture, ISA) 本来作为计算机系统的“细腰”隔离了软件和硬件, 但由于其过于复杂而且涉及知识产权等问题而愈发不能胜任。会上提出了一种观点: 能否将一种可执行的数据流图 (executable dataflow graph) 作为面向人工智能的计算机体系结构的“细腰”, 以构建包容的计算机系统生态环境, 有效支撑底层处理器技术和上层编程系统的独立演化?

## 芯片设计中的通用性与效率之间的矛盾

与会专家对未来处理器的发展路线产生了争论。部分专家认为, 计算机走向通用是其成功的基础, 专用处理器如何实现成本的分摊是一个核心问题, 从目前的现状来看, 还是应该优先走通用芯片的道路, 在保证通用性的前提下, 类似于可重构体系结构这样的方案可能是一种比较好的解决办法。另一部分专家则认为, 从技术的角度来看, 未来的处理器会逐渐走向差异化, 应该设法找到一个合适的分

类方法对处理器进行领域分类, 设计研制不同类型的专业处理器, 同时在相关领域找到杀手级的应用。还有部分专家认为, 类似 GPGPU, 在通用和专用之间找到一个合适的折中也是一条可行的出路。

部分与会者提出受数据流架构启发的新一代通用芯片, 利用数据流模型对海量资源进行高效管理和充分调度, 可以有效解决片上处理器核数不断增多导致的管理低效问题; 另一方面, 数据流与人工智能和大数据之间有着几乎天然的联系, 都具有非规则性和细粒度并行的特点。中科院计算所提出的“时敏数据流”为数据流技术加入了实时性保障, 很有创新性。

## 控制流执行模型与数据流执行模型

与会者对程序执行模型 (Program eXecution Model, PXM) 开展了热烈的讨论。部分专家认为现有的控制流执行模型 (冯·诺伊曼模型) 不能胜任未来的人工智能与大数据处理应用, 面临并行墙、存储墙和操作系统或虚拟机锁等问题 (两墙一锁), 需将程序执行模型进行革命性变化, 转变为数据流执行模型。会上对“数据流 (data flow)”和“流数据 (streaming data)”两个概念也做了澄清, 特别是对数据流理论和技术在 60 年代的起源和 70 年代中的定型做了清晰回顾。还有部分专家认为现有模型仍然有潜力可挖, 仅需对冯·诺伊曼执行模型进行优化即可获得较大的性能提升。

类似地, 部分专家认为可设计专用的编程模型和编程框架, 引导编程人员进行符合数据流执行模型的程序编写。还有部分专家认为可保持编程模型和框架不变, 通过编译等方式进行隐式代码翻译转换, 向编程人员隐藏下层程序执行模型, 实现现有代码的无缝迁移。

## 多学科协作与颠覆性计算机体系结构

在摩尔定律已接近尾声的趋势下，基于多学科协作模式研究未来计算机体系结构势在必行，需要融合材料、物理、微电子、生物（计算神经学）与计算机等多学科前沿进展，以新的计算器件、新的存储或存算器件和新的通信器件为基础，新的计算原理、计算模型与应用为牵引，研发突破“冯·诺伊曼”瓶颈的颠覆性计算机体系结构以及相应的软硬件系统，包括基于忆阻器的存算一体化架构、超导计算机、光子计算机、类脑(brain-inspired)计算

机等。

新型器件或工艺的成熟程度以及计算机各层次的“端到端”设计等因素是决定颠覆性计算机体系结构创新成败与否的关键。比如，新型非易失性存储器件的成熟工艺是实现存算一体化架构的核心需求；又如，光计算过程是一个物理过程，而不是一个状态，适用于对精度要求不高的运算场合，因此必须从适合的应用需求出发，做到扬长避短，发挥新器件优势。■



**钱德沛**

CCF 会士，CCCF 执行主编。北京航空航天大学教授，中山大学计算机学院院长。主要研究方向为计算机网络新技术、高性能计算机体系结构、网格计算。  
depeiq@buaa.edu.cn



**范东睿**

CCF 高级会员，CCF 体系结构等专委委员。中科院计算所高通量计算机研究中心主任，中科院特聘研究员。主要研究方向为众核处理器设计、高通量处理器与系统设计、数据流计算。fandr@ict.ac.cn



**陈文光**

CCF 副秘书长、理事。清华大学教授，兼任青海大学计算机系主任。主要研究方向为并行计算的编程模型、并行化编译和应用分析。  
cwg@tsinghua.edu.cn



**毛 睿**

CCF 高级会员。深圳大学教授、计算机与软件学院副院长。主要研究方向为通用大数据管理分析方法和高性能计算。  
mao@szu.edu.cn

## CCF 会员活动中心动态 (2018 年)

**CCF 上海** 12月8日，CCF上海在上海交通大学成功举办“区块链与数字金融”论坛。来自高等院校、科研院所、金融机构及国内外知名企业180余人参加了本次论坛。

12月27日，CCF上海联合上海高级金融学院、上海计算机软件技术开发中心，于上海高级金融学院举办了“行业数据资产研讨会”。来自高等院校、科研院所、金融机构及国内外知名企业的专家学者、行业精英等人员参加了本次论坛。

**CCF 武汉** 6月7日，CCF武汉“大数据与工业互联网最新技术应用学术研讨会”在武汉工商学院举行。CCF武汉副主席、武汉大学计算机学院副院长吴黎兵，CCF武汉秘书长宋伟，武汉大学计算机学院教授应时，武昌工学院信息工程学院院长龚义建等高校院系领导参加了此次会议。研讨会由信息工程学院副院长彭敏主持。

**CCF 绵阳** 12月22日，CCF绵阳举行了2018年年会，内容包括CCF绵阳工作总结暨CCF绵阳颁奖仪式等。120余名会员参加了本次年会。

**CCF 合肥** 10月12~21日，CCF合肥联合安徽省人工智能学会、类脑智能技术及应用国家工程实验室、合肥中科类脑智能技术有限公司等共同举办第一期“深度学习技术讲习班”。

## CCF 推荐 B 类国际学术会议介绍

# CIKM 2018最佳论文是怎样炼成的

关键词 : CIKM 多模态 信息检索

张俊祺 刘奕群 张 敏 马少平  
清华大学

**题记 :**2018 年 10 月 22~26 日, 清华大学计算机系信息检索课题组师生一行 5 人参加了在意大利都灵召开的第 27 届国际计算机学会信息与知识管理会议 (The 27th ACM International Conference on Information and Knowledge Management, CIKM 2018), 并在会议上作了 3 篇长文口头报告和 1 场专题报告, 其中论文 “Relevance Estimation with Multiple Information Sources on Search Engine Result Pages” 获得了本次会议唯一的最佳论文奖, 作者为清华大学博士生张俊祺、刘奕群副教授、马少平教授和田奇教授。

## 会议简介

CIKM 是中国计算机学会 (CCF) 推荐 B 类国际学术会议, 在信息检索、数据挖掘领域享有较高学术声誉。CIKM 2018 的主题是 “From Big Data and Big Information to Big Knowledge”, 定位于知识、信息和数据的交叉研究。这次会议共收到 862 篇长文投稿, 录用 147 篇, 录用率为 17%。本届大会邀请了三位来自学术界和工业界的知名学者作大会主题报告, 分别是荷兰阿姆斯特丹大学教授 Maarten de Rijke 的报告 “Shifting Information Interactions”, DeepMind 高级研究员 Edward Grefenstette 的报告

“Teaching Artificial Agents to Understand Language by Modelling Reward”, 以及亚马逊研究副总裁 Yoelle Maarek 的报告 “Alexa and her Shopping Journey”。

除了最佳论文, 本届大会还评选出时间检验奖、最佳演示奖、最佳短文奖等奖项, 见表 1。

## 最佳论文介绍

### 一个想法的诞生

随着互联网的发展, 搜索引擎中查询结果的表现形式不断发生变化。几十年前搜索引擎刚诞生时,

表1 CIKM 2018奖项设置

奖项	作者	论文
时间检验奖	David Milne, Ian H. Witten	Learning to link with Wikipedia (发表于CIKM2008)
	Tim Finin, Yannis Labrou, James Mayfield	KQML as an agent communication language (发表于CIKM1994)
最佳演示奖	Edward Abel等人	SOURCERY: User Driven Multi-Criteria Source Selection
最佳短文奖	Alfan Farizki Wicaksono, Alistair Moffat	Empirical Evidence for Search Effectiveness Models
最佳论文奖	张俊祺、刘奕群、马少平、田奇	Relevance Estimation with Multiple Information Sources on Search Engine Result Pages



### CIKM2018最佳论文奖

查询结果的展现形式非常单一，均为一条蓝色的标题链接配以简单的文本摘要。用户检索时，根据每条查询结果的文本信息来判断结果是否相关。这样一个纯文本的检索过程对用户来说极为枯燥和不友好。这一方面是由于互联网刚刚诞生，网页均为纯文本内容。另一方面，搜索引擎主要关注于排序效果，不重视用户与搜索引擎的交互过程优化。

随着多媒体数据的涌现，网页已经不仅仅由纯文本构成，大量的图片、视频等其他资源充斥着网页。垂直搜索在信息资源整合中变得越来越重要。搜索引擎结果页面上，单纯文本已经不能够充分反映该结果的内容，图片、视频等也被加入到了查询结果当中。除此之外，为了直接在结果页面为用户提供所需信息，而不需要用户点击进去，特定类型的垂直结果对信息资源加以整合，提供在结果页面，例如知识图谱、天气查询结果等。查询结果中也开始嵌入一些应用，例如计算器、汇率计算、音乐播放器。除了文本、图片等静态信息，还有输入框、按钮、可缩放地图等动态交互插件，使得用户与搜索引擎结果页面的交互行为更加多样。查询结果展现形式的差异化，可以满足用户多样的信息和服务需求，降低用户信息检索的难度。以最高效的方式满足用户需求是搜索引擎的终极目标，因此，结果页面内容的丰富化和异质化是搜索引擎的发展趋势。

然而，目前搜索引擎的排序算法主要关注的是查询结果所对应的原始网页内容，例如文本匹配特

征、链接关系特征以及用户的点击交互特征等。结果页面中的多模态异质化的信息并没有被考虑到排序算法当中。基于这样的观察，我们开始考虑如何从查询结果页面入手，探索这些信息对用户判断结果相关性以及与搜索引擎交互行为的影响，最终改进搜索引擎的排序。

### 曲折的投稿经历

我们首先需要确定查询结果的哪些信息对于用户判断相关性最为重要。大量的研究表明，视觉显著性对于用户的浏览行为会产生极大的影响。用户得到搜索引擎返回的结果列表时，结果的视觉展示形式使用户感知到的信息最直接。例如，用户想要查找关于猫的图片，那么包含一整行图片的垂直结果最容易吸引用户的关注。展示形式的差异化降低了用户信息检索的难度。因此，考虑一条查询结果的视觉信息显得极为重要。除此之外，要想对查询结果所包含的内容有更加具体的了解，标题和摘要的文本也必不可少。通过文本的精确描述，用户可以获得对该结果的基本了解，判断其是否符合自己的信息需求。

因此，可以将整条查询结果的截图作为视觉信息的输入，标题和摘要作为文本语义信息的输入。由于没有可供使用的数据集，我们根据搜狗搜索引擎所提供的用户查询日志筛选了排序算法最主要考虑的一部分中频查询词，然后抓取了每个查询词第一页结果所对应的标题、摘要和截图，同时保存了每条结果的HTML源码。

数据集中，每条查询结果可以匹配到用户搜索日志中的点击信息。因此，我们尝试利用视觉和文本信息，预测每个查询词前五条结果的点击概率分布。相关的工作投稿到了CVPR 2018 (IEEE 国际计算机视觉与模式识别会议, IEEE Conference on Computer Vision and Pattern Recognition)。这项工作主要的关注点在于结合查询结果不同模态的信息，用于用户点击行为建模这一信息检索领域传统的任务，主要的贡献在于考虑查询结果展示在结果页面的内容对于用户点击行为的影响，将视觉方法引入

到信息检索领域。由于这项工作对于计算机视觉领域的贡献不足，缺乏充分的对比实验和深入的结果分析，论文最终没有被录用。

除了利用查询结果的视觉和文本信息来预测用户点击之外，另一个思路是预测每条结果的相关性，用于结果列表的重排序。我们通过众包平台对数据集中结果的相关性进行了标注，用 HTML 树表示每一条结果的结构。除了查询结果截图的视觉信息、标题和摘要的文本语义信息之外，在预测结果相关性时，我们又结合了 HTML 树状结构信息。最终，每条结果所包含的信息源共有三大类（视觉、文本、结构）四种（截图、标题、摘要、HTML 树）。这项工作投稿到了 IJCAI 2018（国际人工智能联合会议，International Joint Conference on Artificial Intelligence）。因为 IJCAI 文章篇幅较短，只能粗略地介绍数据集，部分分析结果也没能充分地展示。这项工作角度较为新颖，没有采用类似数据或方法去做相关性预测和排序的工作，因此只能和搜索引擎原始的排序做比较。最终，这篇文章也没能被录用。

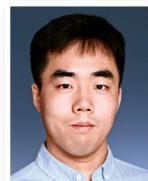
经历了两次投稿失败之后，我们对这项工作的思路和方法产生了自我怀疑，但一直没有放弃。根据 IJCAI 评委的建议，我们对模型和一些相关方法进行了对比。通过对实验结果进行更加深入的分析，对模型的表现效果有了更透彻的认识。之后对实验结果做了补充，留出更多时间对论文撰写进行改进，最终的论文投稿到了 CIKM 2018。因为 CIKM 论文的篇幅较长，可以对数据集的构建和收集做更为详细的介绍，使评委对数据和论文思路有了较深刻的理解。最终，这篇论文被录用，并且获得了唯一的最佳论文奖，给这项工作画上了一个圆满的句号。

## 心得

这篇论文能够被 CIKM 2018 录用并且获得最佳论文奖，一方面是由于异质资源的整合是现代搜索引擎的一个发展趋势，如何根据异质多模态资源做查询结果的聚合排序，是搜索引擎已经面临并且日渐重要的一个问题。这篇论文提出了一个新的思路去解决该问题，也证明了这种方法的优异效果。另

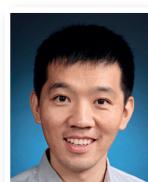
一方面，通过多次投稿到顶级国际会议，收获了很多专家学者关于实验和写作方面的建议。这些建议对我们不断改进工作思路提供了很大的帮助，使我们了解到一项好的工作的要求和标准。一次次被拒稿也使得自己不断提高自我要求，因为工作具有充分的价值，才能得到别人的认可，也才能够对整个研究领域产生一些贡献，对他人的研究工作带来一些启发。这是每一项科学研究的价值所在，也是每一个科研工作者的追求。

正如这届大会的主题“From Big Data and Big Information to Big Knowledge”，互联网的发展带来了大规模、差异化的数据和信息，不同种类的信息源以不同的结构蕴含着人类所需要的知识，这些知识有着更深层次的联系，如何挖掘和整合这些不同形式的数据和信息来建立一个全面的知识体系，或许是该领域的一个发展趋势。 ■



**张俊祺**

CCF 学生会员。清华大学博士研究生。  
主要研究方向为多模态信息检索。  
zhangjq17@mails.tsinghua.edu.cn



**刘奕群**

CCF 高级会员。清华大学计算机系长聘副教授。主要研究方向为信息检索、互联网搜索与自然语言处理。  
yiqunliu@mails.tsinghua.edu.cn



**张 敏**

CCF 高级会员。清华大学计算机系长聘副教授。主要研究方向为互联网搜索与推荐、用户建模、数据挖掘。  
z-m@mails.tsinghua.edu.cn



**马少平**

清华大学计算机系教授，博士生导师。中国人工智能学会会士、副理事长，中国中文信息学会副理事长。主要研究方向为智能信息处理和信息检索。  
msp@mails.tsinghua.edu.cn

## 构建中日韩学会合作联盟

在韩国平昌举行 CJK 会议

2018 年 12 月 20 日，在 2018 韩国软件大会 (KSC2018) 期间，韩国信息科学学会 (KIISE) 代表与应邀出席 KSC2018 的中国计算机学会 (CCF) 代表和日本信息处理学会 (IPSJ) 代表举行了 CJK 三方会议。KIISE 理事长 Young Ik Eom、2019 年理事长 Jongwon Choe、2020 年理事长 Doo-Hyun Kim、KIISE 顶级会议列表委员会主席 Kyuseok Shim，CCF 秘书长杜子德、CCF 杰出会员 /KSC2018 特邀讲者 / 清华大学教授唐杰、CCF 办公室主任朱征瑜，IPSJ 候任理事长 /KSC 2018 特邀讲者 Yasuo Okabe、秘书长 Taizo Kinoshita 参加会议。

会议重点讨论了 2019 年各方重要学术会议的准备情况和后续落实问题，包括 IPSJ 将于 2019 年 3 月举办的第 81 次全国大会，KIISE 将于 6 月举办的学术会议，以及 CCF 将于 10 月举办的中国计算机大会 (CNCC2019)，落实了各方参会人员以及需要落实的特邀讲者的要求。

韩方特别介绍了他们刚刚完成的顶级会议列表相关的组织、信息采集和结果等情况，CCF 希望对方能够给出更加详细的关于列表的材料和支撑数据。CCF 正在进行“CCF 推荐国际会议和期刊列表”的修订工作，希望双方能够在未来携手共同讨论相关问题。

CCF-IPSJ-KIISE (简称 CJK) 会议由 CCF 于 2015 年发起，每年分别在中、日、韩举行 1 次，旨在加强中、日、韩三国计算机领域的合作，共同推动三国计算机专业的发展。

## 唐杰在 KSC2018 作特邀报告



应韩国信息科学学会 (KIISE) 邀请，2018 年 12 月 20 日，CCF 杰出会员、清华大学教授唐杰在 2018 韩国软件大会 (2018 Korea Software Congress, KSC2018) 作了题为“Empower MOOCs with AI”的报告。

近年来，MOOCs 蓬勃发展，在全球范围吸引了数百万的用户，学堂网 (XuetangX.com) 向逾千万注册用户提供数千门课程，但是，完成率总是很低。唐杰在报告中介绍了他们利用用户的内隐反馈(如用户点击)来帮助提高学习效率的方式和结果。

## 新技术 & 新应用

### “超级芯片”取得突破进展

英特尔公司和加州大学伯克利分校的研究人员近日在 *Nature* 上发表了一篇题为 “Scalable energy-efficient magnetoelectric spin-orbit logic”的论文，介绍了其正在研究的一种高集成的自旋电子逻辑器件，称为 MagnetoElectric Spin-Orbit(MESO)。器件的制作结合了多铁性材料和拓扑材料。多铁性材料本身具有磁性和铁电性，可以通过改变施加在多铁性材料的电场方向来改变其内部的磁场方向，并通过自旋轨道耦合效应进一步改变材料中电子自旋的方向。另一方面，拓扑材料可以根据电子的自旋

选择性通过具有特定自旋方向的电子。基于这两种材料的独特性质，MESO 最终可以通过改变电压以及材料中电子的自旋方向，进而以拓扑材料充当电子的传输通道，限定特定自选方向的电子通过。

由于 MESO 使用量子材料，与传统的 CMOS 晶体管相比，能耗可降低至原来的 1/10~1/30，开关电压也可降低至原来的 1/5。比传统的 CMOS 集成微处理器，能效提升约 10~100 倍。因此，MESO 有望在未来取代目前广泛使用的 CMOS 晶体管。

### 量子计算机控制系统研制成功

2018 年 12 月 6 日，中国科学技术大学郭光灿团队发布了我国首款具有完全自主知识产权的量子计算机控制系统——本源量子测控一体机，实现了对量子芯片的操控并发挥其性能优势。

该系统最基本的功能是提供量子芯片运行所需的关键信号，同时负责对量子芯片传回信息的处理，并执行对量子计算机程序的编译；采用模块化设计，当前发布的版本包含四种功能模块，总计 40 个功能

通道，功能通道之间通过 PCIE 高速扩展总线互连实现一体化，可直接支持 8 位超导量子芯片或者 2 位半导体量子芯片。在使用过程中，还可以将其升级至最多 200 通道，以支持更高性能的量子芯片。



### DeepMind 推出 AlphaFold 预测蛋白质的 3D 形状

2018 年 12 月 2 日，DeepMind 推出了其最新的人工智能系统 AlphaFold，该系统在全球蛋白质结构预测竞赛 (CASP) 中名列榜首，针对任务 “蛋白质折叠问题”，可准确地从 43 种蛋白质中预测出 25 种蛋白质的结构。

蛋白质可以在氨基酸之间扭曲、折叠，因此一种含有数百个氨基酸的蛋白质有可能呈现出数量惊人的结构类型。不同的折叠或纠缠会导致糖尿病、帕金森和阿茨海默症等疾病。因此，如果我们可以根据蛋白质的化学构成来预测其形状，那么我们就

可以通过设计新的蛋白质来治疗疾病。为了构建 AlphaFold，DeepMind 在数千种已知的蛋白质上训练了一个神经网络，直到它可以仅凭氨基酸预测蛋白质的 3D 结构。给定一种新的蛋白质，AlphaFold 可以预测氨基酸对之间的距离，连接它们的化学键之间的角度，接着调整初步结构以找到能效最高的排列。基于这两个属性，继续训练一个神经网络以预测蛋白质成对残基 (residues) 之间距离的独立分布以及一个独立的神经网络，该网络使用集群中的所有距离来估计预测的结构与实际结构之间的差距。

## 清华大学发布《人工智能芯片技术白皮书》

2018年12月11日，清华大学—北京未来芯片技术高精尖创新中心结合学术界和工业界的最新实践，联合斯坦福大学、加州大学、圣母大学等领域资深专家，共同发布了《人工智能芯片技术白皮书》。

《人工智能芯片技术白皮书》主要从研究背景、技术趋势、基准测试、技术路线和未来发展等九个方面阐述了人工智能芯片的战略意义。白皮书指出，作为人工智能应用实现的物理基础和关键支撑，芯

片已成为人工智能领域的研究和创业热点。白皮书的发布，一方面希望与全球学术和工业界分享AI芯片领域的创新成果，另一方面则希望通过AI芯片的技术认知和需求的深入洞察，帮助相关人士更加清晰地了解AI芯片所处的产业地位、发展机遇和需求现状，通过对AI芯片产业现状及各种技术路线的梳理，增进对未来风险的预判。在充满信心、怀抱希望的同时，保持冷静和客观，推动AI芯片产业可持续发展。

## 腾讯医疗AI提出器官神经网络辅助头颈放疗规划

腾讯医疗AI实验室与美国加州大学合作，在*Medical Physics*上发表了其最新的研究成果：器官神经网络(AnatomyNet)。

AnatomyNet 基于常用的三维 U 网络(U-net)架构进行扩展，使用在整幅 CT 图像上进行自动分割的新编码方式，在编码层中加入三维 Squeeze-and-Excitation 残差结构来进行更好的特征表示学习，同时提出一种新的结合 Dice 损失和 Focal 损失的损失函数，用来更好地训练该神经网络。从结果上看，与之前 2015 年 MICCAI 竞赛中最好的方法相比，AnatomyNet 将 Dice 相似系数(DSC)平均提升 3.3%，

仅仅需要大概 0.12 秒就可以完全分割一幅尺寸为  $178 \times 302 \times 225$  的头颈部 CT 图像，极大地缩短了之前方法所用的时间。同时，该模型可以一次性完成一幅完整的 CT 图像的处理，而几乎不需要预处理或后处理。

该研究成果使得人工智能系统在头颈等重要器官的放射治疗规划中发挥精准规划作用，最大限度地将放射剂量集中在靶区内，使周围正常组织或器官少受或免受不必要的伤害。与单纯依靠人类医生相比，既可以提升诊疗规划效率，降低勾勒时长，还能提升勾画准确率。

## 美 CSIS 报告：人工智能与国家安全

2018年11月5日，美国战略和国际问题研究中心(CSIS)网站发布题为《人工智能与国家安全——人工智能生态系统的重要性》的报告，阐释人工智能生态系统的组成、当前人工智能投资情况、人工智能在国家安全领域中的应用等，并在分析人工智能生态系统构建必要性的基础上，为美国打造强健的人工智能生态系统提出具体建议。

报告指出：人工智能生态系统包括熟练的劳动力和在行的管理；捕获、处理和利用数据的数字能

力；信任、安全和可靠的技术基础；人工智能蓬勃发展所需的投资环境和政策框架。政府应继续发挥作用，追求那些不能为私营部门带来快速投资回报的更难的技术领域；为关键的政府和国家安全应用开发建立人工智能可靠性所需的工具；以及发展和加强人工智能生态系统。 ■

(本条内容转自中科院成都文献情报中心)

(本栏目内容由动态栏目编委鲍捷提供整理)

# 社交媒体中的多方隐私 \*

作者: 约瑟·萨奇 (Jose M. Such)

娜塔莉亚·卡利亚多 (Natalia Criado)

译者: 胡欣宇 岳亚伟

关键词: 社交媒体 多方隐私

在线隐私不仅与你自己分享的个人内容有关，也与别人分享的关于你的内容相关。

超过 20 亿的用户在使用社交媒体来构建并参与在线社交网络 (Online Social Networks, OSNs)，上传分享数以亿计的数据项<sup>[15]</sup>。在线社交网络规模巨大，预计在未来几年，用户数量和用户上传及共享的数据规模都将持续增长。大多数情况下，社交媒体的大部分数据是由用户生成，与个人有关。因此，需要采取适当的隐私保护机制，允许用户享受社交媒体便利的同时，对其个人信息进行充分保护。保护用户隐私不仅是对《世界人权宣言》的必要尊重，也是减少由社交媒体隐私数据泄露造成的网络犯罪现象及其他非法活动的第一道防线，如社交网络钓鱼、身份窃取以及网络欺凌。

自社交媒体发展初期，就有许多工作致力于研究社交媒体中的隐私以及如何保护用户个人隐私，如格罗斯 (Gross) 和阿奎斯蒂 (Acquisti) 的工作<sup>[10]</sup>。但其中大部分工作是从个体角度来刻画隐私问题。例如，学界<sup>[9]</sup> 和工业界<sup>[16]</sup> 努力通过在社交媒体中普遍存在的二元朋友关系模型之外建立不同的关系和社交圈模型，以帮助个人用户更好地寻找目标受众。虽然这确实提升了内容共享的水平，但与此同时，共享内容对多个用户隐私的影响却鲜有关注。

隐私不仅仅与你所说或分享的关于自身的内容有关，也与别人所说或分享的关于你的内容相关。有证据表明，在关系、家庭、团体或组织中，隐私边界是由其中的个体共同管理的<sup>[22]</sup>。然而，随着社交媒体的大规模增长，群体共有隐私边界的维护变得极富挑战，数以亿计上传的数据项中的大多数由多名用户共有<sup>[14,15]</sup>，但主流社交媒体仅允许共有数据项的上传者对共有数据项进行隐私设置，这往往会导致冲突和严重的隐私侵犯<sup>[33,35]</sup>。

多方隐私 (Multiparty Privacy, MP) 旨在促进在线数据项的所有共享个体进行共有隐私边界的协调，因为每个共享个体的隐私都与共享数据项的其他共享个体息息相关<sup>1</sup>。MP 主要关注多方隐私冲突 (Multiparty Privacy Conflicts, MPCs) 的检测和解决方案，特别是当共享同一数据项的多个个体的隐私偏好出现冲突时。举一个简单形象的 MPC 实例：爱丽丝 (Alice) 拍摄了她和鲍勃 (Bob) 的照片。主流社交媒体只允许爱丽丝 (假设由她上传照片) 来对照片进行隐私设置，但如果鲍勃并不想把照片分享给爱丽丝想要分享的部分朋友呢？MP 不仅关注照片，也关注其他社交媒体内容，如帖子、视频、评论或活动。除社交媒

\* 本文译自 *Communications of the ACM*, “Multiparty Privacy in Social Media”, 2018, 61(8):74~81 一文，有删节。

<sup>1</sup> MP 与其他只关注单个个体保护的协同方法不同。<sup>[4,6,21]</sup>

体之外，MP 还可以在其他诸如协同软件（如基于云的协同文档），内部 / 外部维基页面，博客，群体智能，众包等社会计算领域发挥作用。在这些领域中，信息由多个用户共同创建并共同所有，以便所有用户都对向谁共享这些信息具有发言权。

设计 MP 工具是一项复杂而艰巨的任务，因为用户对隐私持有不同的观点和偏好，他们通过多种途径进行在线社交，并且他们共享不同数量不同类型的内容。

## 社交媒体对MP的支持

主流社交媒体网站支持的 MP 主要通过两种机制实现：标记 / 取消标记以及报告不适当内容。

标签通常用于标记照片中出现的人物，并带有指向其个人主页的链接。然而照片中被标记的人可以将自己从照片中取消标记。在一些社交媒体中（如 Facebook），你可以选择接收你被标注的照片的通知，以便在标签生效之前对其进行验证。标记 / 取消标记就表示了某种 MP，但它主要存在三个限制。第一，即便你在任何人看到照片之前将自己从照片中取消标记，也不表示你的朋友不会看到照片。例如，爱丽丝向 Facebook 上传了一张爱丽丝和鲍勃的照片，其中对鲍勃进行了标记。鲍勃收到一个提醒，他查看了照片并且决定取消了该标记，因为照片使他觉得难堪。关键是，即使没有明确标记鲍勃，照片也会根据爱丽丝的选择进行分享。也就是说，如果爱丽丝决定与朋友进行分享，并且爱丽丝与鲍勃有一些共同好友，那么这些共同好友无论如何都将在爱丽丝的主页上看到这张图片。第二，标记 / 取消标记操作只支持照片，不支持帖子、评论和活动等其他条目。帖子和评论虽然通常可以选择“提及”操作（使用特殊符号，如“@”），但这些“提及”操作只能由帖子 / 评论的创建者控制，而用户只可以删除对他们的帖子 / 照片的评论。第三，许多用户表示将自己从照片中取消标记让他们感到非常不舒

服，因为这可能会冒犯（从社会角度来看）在照片中对他们进行标记的人<sup>[2]</sup>。

关于报告机制，大多数社交网站允许用户对他们觉得不适当的内容进行报告。该机制主要用于处理不适当（甚至非法）内容，如裸露、敌视言论、暴力和其他严重违法行为。报告后，由提供商单方面决定如何对内容进行处理（删除与否）。虽然这种机制对于打击严重违法行为至关重要，但它并不适用于所有 MP 场景，因为在许多情况下，即便没有违法行为，隐私侵权行为依然有可能发生，例如当你不愿意与其他人分享某些信息或希望信息对同事不可见的时候。此外，需要强调的是，报告只是一种反应机制，只有在内容发布并且有人将其标记为不适当才会响应。但是，当内容被标记时，可能为时已晚，隐私侵犯可能已经发生，导致的损失已经无法挽回，或者其他用户已经下载内容并转发到其他渠道。

最近 Facebook 隐私控制的更新证明 MP 问题已经开始被主流社交媒体所认可<sup>2</sup>。特别是 Facebook 的隐私保护机制最近增加了与你不喜欢的照片的所有者联系的选项。该机制工作方式如下：如果用户在照片中被标注，并且他 / 她不喜欢这张照片，那么可以将照片标记为不喜欢，然后系统会打开一个消息窗口，收信人为照片上传者，以便不喜欢照片的用户告知照片上传者将照片删除，并可以附上要求删除的原因。对于公认的 MP 问题，这种做法已经有了进步，但依然存在不少问题：(1)一旦照片已上传，相关进程就已经启动，因此潜在的隐私泄露可能已经发生。(2)从网站上撤下已发布的图片需要时间。例如，Lian 等人<sup>[19]</sup>计算了照片从社交媒体上删除到照片链接不可访问所经历的时间，结果表明 Instagram 需要 3 天，Facebook 需要 7 天，Flickr 需要 14 天，MySpace 和 Tumblr 需要至少 30 天。(3)该机制不支持群体协商，因为照片可能牵涉到其他人，而不仅仅是照片发布者和对该照片不满者。(4)一切操作都需要手动完成，这会给拥有大量线上

<sup>2</sup> <https://www.facebook.com/about/basics/howothers-interact-with-you/>。

好友的用户带来无法忍受的负担。(5)这个机制仅仅实现在照片中，并没有对其他类型的内容进行适配，如帖子、评论和活动。

## 用户对MP的应对策略

如前所述，当前社交媒体的基础架构中缺少内建(built-in)功能，以方便用户间通过主动协商达成一致<sup>[40]</sup>。用户被迫在社交媒体之外进行沟通，并采取一些应对策略以克服技术支持的不足。基本上，这些应对策略中的大多数包含旨在防止线上发生MPC的一系列线下世界的行动或行为。本文通过研究这些应对策略的几个实例，揭示了对MP工具的强烈需求。接下来讨论一些应对策略的实例及其缺点(表1中做了概述)。

表1 应对策略的实例

策略	主要缺陷
对他人带来的影响进行预测 <sup>[18]</sup>	不可能总能预测到隐私相关问题
发布前征求许可 <sup>[18]</sup>	对上传内容的用户造成过多负担
使用内部语言并隐藏 <sup>[3]</sup>	可扩展性差，对某些内容不适用。
使用替代媒体 <sup>[2]</sup>	MPC在其他媒体同样可能发生，同时，用户无法对其他用户使用的分享媒体进行控制。
遇到摄像头时改变线下行为 <sup>[2,18]</sup>	由于手机和可穿戴设备的广泛出现，很难实现。
同其他用户协商共享策略 <sup>[2,18,40]</sup>	由于共有内容数量较多，容易成为用户负担。

信息发布之前对该信息的敏感性进行预估，是人们采用的一种线下MPC避免策略<sup>[18]</sup>。例如，如果爱丽丝和鲍勃同时出现在照片中，但照片中鲍勃看起来显然已经醉了，那么爱丽丝可能会考虑这个情况，要么不上传照片，要么只与有限数目的朋友分享。然而，这种做法并不总是奏效，因为有时信息发布者并不能事先预见对他人造成的影响。Lampinen等<sup>[18]</sup>给出了一个例子：某人的一个朋友在评论中祝贺其获得硕士入学资格，但该人不得不迅速删除评论，因为他还没有将这个消息告诉雇主，

并且雇主也是其在线好友。请注意，即便此人快速删除了该评论，其雇主仍有可能在评论删除之前注意到该评论。

用户有时在分享信息之前会请求其他共同所有者应允<sup>[18]</sup>。此策略的问题在于它是在完全线下、没有任何技术支持的前提下完成的。也就是说，人们需要在线下向其所上传信息的每个涉众请求应允。并且，如果有人不同意时，他们需要对解决方案进行协商(例如，减少初始受众或者决定不上传)。这很快就会成为用户难以承受的负担。

另外一个观察到的现象是，青少年会对他们的信息进行隐藏，使用他们的圈内语言分享照片<sup>[3]</sup>。例如博伊德(Boyd)和马威克(Marwick)<sup>[3]</sup>提到了一个女孩在Facebook上发帖的例子，她知道这篇帖子里的事情只有她和她的密友明白，因为她想阻止其他朋友知道她的真实想法。这种策略的缺点是无法扩展，并且不可能适用于人们想要共享的所有照片或其他类型的信息。例如，你前往毛里求斯的照片就难以只分享给一部分人而不分享给其他人。

由于已经证实社交媒体在MP场景下对隐私管理的不足，一些用户切换到使用其他技术的媒体(如基于云的文件共享，即时消息或电子邮件附件)进行信息共享<sup>[2]</sup>。这样做的好处是不仅可以保护自己的内容，也减少了泄露他人隐私的风险。然而，这种做法有三个主要的缺点。首先，这种做法适用于照片、视频等内容，但不适用于其他类型，如活动或评论。其次，用户无法决定他的朋友使用哪种技术；也就是说，他们的朋友依然可以使用社交媒体上传照片，而用户无法对其进行任何操作。第三，这些技术也可能导致MPC，例如，一个用户可能在WhatsApp群组内分享一段视频，其中有一些人是视频中其他用户不愿与之分享的。

用户也证实，如果没有更好的办法来应对MP状况，他们实际上会改变并严格控制他们的线下行为。例如，当人们发现周边的镜头后，会有意地改变行为<sup>[2,18]</sup>。如果你知道朋友喜欢拍照并经常上传，你可能会决定不和她一起外出，以免任何不希望公开的照片被上传。这凸显了人们无法参与MP决策

的无力感。然而这些策略的有效性也非常有限，由于智能手机和可穿戴设备的普及，时刻保持警觉并不断改变你的线下行为是不现实的。

最有趣的策略之一可能是用户集体协商并形成线下协议，对发布的内容及共享受众达成妥协<sup>[2,18,40]</sup>。例如，一群朋友会同意将他们在旅途中拍摄的照片在他们之间或他们的密友之间分享。有趣的是，事实证明，用户总是非常豁达地考虑并尽可能迎合他人的偏好<sup>[18,40]</sup>。此外，研究表明，用户不希望对他们朋友造成任何有意的伤害，他们通常会听取其合理的反对意见，这也是重申和回报关系的一种方式<sup>[40]</sup>。和目前提到的其他策略一样，这一策略的主要问题在于它不能扩展。用户不可能在没有技术支持的情况下，经常和数以百计的朋友沟通数以百计的照片问题。

## MP工具研究

前文的案例已经很明显地反映出，用户针对MP场景下缺乏技术支持的问题，在积极寻求解决方案。然而，由于这些策略的缺陷，使得这些策略的效果显得特别有限。这激发了研究人员设计新的、相比现有被动应对策略有所改进的用户接口和算法，使得用户能够以更有效且高效的方式进行MP

的共同管理。尽管该领域的研究尚处于初级阶段，但已经有了一些研究思路，我们将其大致分为5类（如表2所示）。由于篇幅和参考文献的限制，我们仅选取了每种方法中我们认为最具代表性的工作进行讨论。

**手动方法。**第一个支持MP的研究方向是帮助用户确定MPC在哪里可能或已经发生<sup>[2,39]</sup>。例如，威舍特(Wishart)<sup>[39]</sup>等人提出了一种对强弱共享偏好进行标识的方法，通过对这些偏好进行检查来发现隐私冲突。此外，Besmer等人<sup>[2]</sup>提出了一个系统，其中照片中被标记的用户可以联系照片的上传者，要求其删除照片或者限制照片可见人群，Facebook不久之后推出了类似功能<sup>[7]</sup>。这些方法作为MP研究领域的垫脚石，意识到了MP问题并给出了部分解决方案，但对于检测到的MP冲突的协商过程是没有任何特别的技术支持的。也就是说，用户必须以手动方式处理每个潜在的MPC，考虑到上传的内容庞大以及用户在社交媒体朋友的数量，这可能会造成难以承受的负担。

**基于竞价的方法。**另外一个研究思路提出使用竞价机制解决潜在的MPC<sup>[30]</sup>。用户对他们最喜欢的共享决策进行竞标，中标者决定某个特定数据项的共享策略。这些方法是第一个采用半自动方法帮助用户共同确定共享策略的方法，例如，竞价的结果根据用

表2 MP方法及样例引用总结

方法	简要描述	主要缺点
手动 <sup>[2,39]</sup>	为用户提供检测MPC的途径，用户对检测到的MPC进行手动解决。	由于不提供冲突的自动解决方案，容易对用户造成负担。
基于竞价的方法 <sup>[3]</sup>	用户获得虚拟货币，可以对最期待的共有条目共享策略进行竞价。	用户可能对竞价过程的理解和管理存在困难。
基于聚合的方法 <sup>[5,12,36]</sup>	所有用户的个人隐私偏好使用某一规则或规则集进行聚合以产生一个共同共享决策。	个人隐私偏好以同一方式聚合或者由上传者选择聚合方式。
自适应方法 <sup>[32]</sup>	基于一系列因素对不同情况进行建模，并根据不同情况选择不同的共享决策。	很难对决定一个情况的所有可能因素以及实现最优共享决策的最佳方法进行建模。
博弈论方法 <sup>[13,17,25,31]</sup>	用户或自动软件代理根据既定协议对某个特定情况进行协商。协议和协商策略均使用博弈论概念进行分析。	用户在社交媒体中的行为似乎并不是完全理性的，因为许多社会特质会在MP中产生影响。
细粒度方法 <sup>[14,36]</sup>	用户对照片内出现的个人身份标识对象单独定义访问控制策略，例如，用户决定他们的面部是否被模糊。	照片中的模糊对象（如人脸）可能并不是信息共享的效用和/或保护用户隐私的最佳方案。

户给定的出价自动计算。但是，用户可能难以理解该机制，在竞价中难以给出合理的出价，并且用户需要为与他人共同拥有的每个数据项进行投标。

**基于聚合的方法。**这些方法通过对所有用户的个人隐私偏好进行聚合形成一个 MPC 解决方案。它们可以抽象地概念化为投票机制，共享同一数据项的每个用户的偏好计为对共享 / 不共享的一次投票（有时是加权）。投票规则描述了个体偏好如何聚合。例如，在多数投票规则中<sup>[5]</sup>，多数用户的偏好被作为最终的共享策略。另外一个例子是否决投票<sup>[35]</sup>，如果存在一个涉及到的用户反对共享，那么该内容就不会被共享。这些方法的主要问题是它们总以相同的方式对偏好进行聚合。例如，在多数投票规则下，即使内容非常敏感并对某个用户隐私导致侵害，如果大多数用户同意，该内容依然会被共享。反之，一直使用否决投票则可能过于严格，并影响用户从社交媒体分享中获得已知权益<sup>[29]</sup>。后续工作<sup>[12]</sup>认识到了这个问题，并考虑使用多种方式对用户偏好进行聚合。但是，这种方法需要由上传者选择使用哪种聚合方式，并需要上传者预估上传信息对他人的影响，前面已经讨论过，这可能是一个非常复杂的任务，并且它可能并不总能给出最佳解决方案。

**自适应方法。**这类方法根据特定情况自动推测解决 MPC 的最佳方法<sup>[32]</sup>。这些方法根据每个用户的个人偏好、内容敏感性以及和潜在观众的关系对场景进行建模。一个特定的场景会形成特定的让步实例，比如用户们对数据共享规则进行线下协商并在达成一致时形成一个特定让步<sup>[2,18,40]</sup>。这类方法自动适应当前的场景，如果场景需要（例如，共享数据条目非常敏感），这些方法也能和反对投票方法一样具有限制性，或者建议在其他情况下共享（例如，某些人有兴趣共享而其他人并不关心）。虽然这类方法捕获了在线下协商期间何时让步的已知场景，但很难对所有可能的场景进行建模，并且也可能无法捕获在潜在未知场景下可能出现的机会性让步或协议。

**博弈论方法。**另一种定义协商协议的方法，是通过规定参与者之间的交互方式来规范参与者协商 MPC 问题解决方案的通信流程。这些协议由用户手

动或由软件代理自动制定，以便对特定数据条目的共享策略进行协商。参与者在制定协商协议时可以遵循不同的策略，并使用众所周知的博弈理论（如纳什均衡）来对这些策略进行分析。这样就能通过系统分析来确定哪些是参与者可选的最佳策略以及平衡策略（任何参与者都无法通过单方改变自己的策略获益）。虽然这些方法在形式上提供了理想的框架，并且建立在成熟的分析工具上，但是在实践中使用效果可能并不好<sup>[13]</sup>。这是因为实际情况下用户的行为并不完全理性（如方法中假设的那样），即使有些方法考虑到了其他因素，如互惠<sup>[17]</sup>和社会压力<sup>[25]</sup>，但他们仍远未考虑到影响 MP 的诸多社会特征<sup>[18,40]</sup>。

**细粒度方法。**最后一个研究思路侧重于通过让照片中出现的每个用户各自决定照片中的个人标识对象是显示还是模糊，以避免 MPC<sup>[14,36]</sup>。其中一种早期方法允许用户自己决定他们的面部是显示或模糊化<sup>[14]</sup>。过程如下：通过人脸识别算法如 Facebook 的 DeepFace 算法对照片中出现的用户进行识别<sup>[34]</sup>，被识别出的用户会收到通知，并给出一个允许访问该照片的好友列表；当某个用户访问照片时，她只会看到已授予她访问权限的用户的面孔，而照片中其他面孔将被模糊化。然而，模糊的人脸（或照片中的其他对象）可能会影响共享照片的效用，进而影响人们通过社交媒体分享的体验<sup>[29]</sup>，另外即使照片中的人脸（或照片中其他物体）已经被模糊处理，依然存在某个人被认出来的风险<sup>[23]</sup>。因此，如果能通过协商达成一个 MPC 解决方案，该方案可能比单独给出访问权限更可取。在这种方式下照片不会失去任何效用（没有任何对象模糊），并且通过协商可以确定照片的受众以避免任何不必要的人对照片的访问。

## 对MP工具的要求

**基于实际经验数据的设计。**现有方法都没有深刻理解 MPC 问题，也没有在实际中形成最佳的解决方案。一部分是由于目前没有足够的 MPC 经验数据。这些经验数据对克服现有文献中 MP 工具设计的局限性至关重要。如上所述，研究人员已经阐明为

了解解决社交网络中缺乏 MP 支持的问题，用户是如何被迫采取在线应对策略的<sup>[2,3,18,40]</sup>，并且有证据表明用户是如何在线下协商共有隐私边界的。虽然此前的研究已经提供了一个非常好的基础，但需要进一步研究以更好地了解 MPC 的发生时机及频率，更重要的是，当它们发生问题或导致潜在隐私侵犯时的实际解决方案。可以对用户遇到的一些 MPC 特定实例进行进一步研究，以了解在使用应对策略后它们是否还会发生，用户如何针对这种 MPC 提出最佳解决方案，以及在这个过程中发挥作用的因素。最近的一些研究在朝着这个方向进行<sup>[33]</sup>，并贡献了 MPC 的第一个经验性的公开数据集。这个 MPC 经验数据集可以为深入理解 MPC 以及影响 MPC 的各种细微因素提供支撑，并作为基础针对不同类型的用户、社交群体、关系等设计 MP 工具，提供技术支持并推荐 MPC 问题的最佳解决方案。近期在隐私工程方面的成果应该被用来简化从经验数据到隐私设计的艰巨任务<sup>[11]</sup>。

**以用户为中心的 MP 控制。**这里面临的主要挑战是如何利用前面的经验基础开发可用的 MP 工具，这样用户可以用最小的代价对 MP 进行有效管理。但是，MP 工具应该以可用性为目标，而不是以完全自动化的解决方案为目标，因为面对社交媒体隐私时，自动化的解决方案可能无法取得令人满意的效果。相反，近期研究表明，如果用户向 MP 工具输入一些数据（如首选隐私策略的原因），MP 工具将会反馈建议，这样可以更好地提出针对某个 MP 冲突的最佳解决方案。但是，如果用户必须参与表达他们的个人隐私偏好，为每个共享数据项和潜在冲突选择或拒绝系统推荐的方案，这难道不会成为用户的负担吗？我们如何在用户参与度与自动化之间找到合适的权衡？之前关于社交媒体中的个人隐私研究可能会提供一些帮助：AudienceView<sup>[20]</sup>可以用来显示和/或修改推荐方案或表达个人偏好；与 Fang 等人<sup>[7]</sup>类似的方法可以用于研究用户对 MP 的响应随时间的变化方式；Waston 等<sup>[38]</sup>的方案可以用来创建合适的 MP 默认设置。

**可扩展可比较的评估。**之前提出的 MP 方法要

么没有同用户进行经验评估<sup>[5, 17, 25, 30, 31, 39]</sup>，要么只是进行至多 50 人的小规模的用户研究<sup>[2, 12-14, 32, 36]</sup>。这些问题一部分是由于缺乏系统性和可重复性的 MP 评估方法及对比方法。为了使得评估更具信服力和推广能力，MP 工具的评估应该考虑更加广泛更加多元的人群。此外，评估协议的制定，应遵循“生态效率”最大化原则，这在此领域极具挑战。首先，用户研究的参与者似乎总是不愿意与研究人员共享敏感信息<sup>[37]</sup>（例如，他们感到尴尬并且不愿在线共享的照片），这会使所有评估偏向于非敏感事务，遗漏了对 MP 工具性能至关重要的场景。另外一个问题是在评估时遇到的假数据/场景，由于隐私态度和隐私行为之间的巨大差异，参与者所说的行为可能与实际行为不符。其次，在“自然环境”下进行 MP 评估是非常困难的，因为需要对受特定内容影响的所有用户一起进行研究，以理解冲突并评估冲突解决方案是否是最佳的。可能的改进是采用基于生活实验室的方法，这些方法将研究对不断变化的现实生活场景进行整合与验证。

**隐私增强的群体识别。**给定特定的上传数据条目，MP 工具应该自动推出受该数据条目影响的用户。例如，某用户上传照片，并在其中标注出现在照片中的其他用户，MP 工具可以直接使用这些信息来了解照片会影响到哪些用户。但是，用户经常未将照片中所有清晰可见的人全部标识，或是错误标识在照片中未出现的人。人脸识别软件可应用在这里，例如 Facebook 研究人员开发的 DeepFace<sup>[34]</sup>，拥有 97.35% 的准确率。这里存在的问题是人脸识别软件的使用是否存在过分的隐私侵入，也就是说社交媒体提供商能够识别照片中出现的任何一个人，即使是社交软件之外的照片，或者是个人被错误标识并与不相关的条目关联（虽然软件准确率很高且误报率很低，但用户数量及数据条目数量基数很大）。有趣的是，与广为人知的隐私效用权衡相比，这似乎开启了一种全新的令人兴奋的隐私相关权衡方式，它将在多方隐私与个人隐私之间做权衡。注意，如果 MP 工具使用了隐私保护人脸识别算法<sup>[27]</sup>，那么就不需要进行多方隐私与个人隐私

之间的权衡。除了照片之外，其他类型内容如活动（用户被邀请或明确提及）中的群体识别会更加容易，而某些类型如文本帖子，由于受影响的用户可能经常不被明确标记，其中的群体识别更具挑战。

**支持可推理隐私。**之前在 MP 场景下未被考虑的另外一个问题是可推理隐私。它不仅包括你的朋友在网上发布的关于你的内容，也包括从这些内容中可推出什么信息，不论内容是什么类型。例如，Sarigol 等<sup>[27]</sup> 使用来自单个在线社交网络的 300 多万个账户的数据，证明了对用户和非用户进行性取向的影子画像可行性。注意，可推理隐私情况下的谈判或者协商可能会更加复杂，因为内容不能发布的原因可能不是由于内容自身，而是由于可能从中推出的信息带来的后果。因此，用户可能更加难以理解这种类型的 MPC 的解决方案，这也会给 MP 工具的可用性和可理解性带来挑战。此外，我们注意到没有任何社交网站为用户提供任何类型的可推理隐私控制机制，也没有同时考虑 MP 和可推理隐私的相关研究。

**隐私保护保证。**最后但同样重要的是，MP 工具应该提供某种个人隐私保证。当无法达成多方协议时，这一点尤为重要。例如，用户故意发布中伤其他用户的内容，这种情况下可以强制施行个人隐私偏好设置。如伊利亚 (Ilia) 等人<sup>[14]</sup> 提出的关于照片的解决方案，允许用户控制其脸部在某个图片中被显示或者被模糊处理。当 MP 冲突出现并且受影响用户未达成一致时，这似乎是一个合理的解决方案。与“赢者决定”的规则不同，此方案能够在一定程度上保证所有受影响用户的个人隐私。然而，正如伊利亚<sup>[4]</sup> 所说的那样，这并没有完全消除被识别的风险，因为她的脸部即使被模糊，但仍然可能被识别出来，虽然还没有算法可以在移除一个人的全身之后将图像重建，但是已经存在识别用户身体/动作的方法<sup>[28]</sup>。

## 结论

MP 是社交媒体中的一个重要问题，也可以扩

展到社交计算中存在共享信息的其他领域，如博客、群体智慧 (collective intelligence)、维基页面、基于云的文件共享<sup>[24]</sup> 以及协作文档，与社交媒体相比，这些领域受到的关注较少。正如本文所强调的那样，主流社交媒体并没有为 MP 提供足够的支持，导致用户不得不使用非理想的各种应对策略。因此，需要开发新的隐私增强技术和机制来帮助用户管理 MP。我们还有很长的路要走，才能使这些机制成为现实，并将它们嵌入到最终用户可以使用的高可用工具中，部分原因在于 MP 和社会行为的复杂性，这需要对 MP 进行跨学科的处理。 ■

作 者：

约瑟·萨奇 (Jose M. Such)

英国伦敦国王学院自然与数学科学系信息学系副教授。  
jose.such@kcl.ac.uk

娜塔莉亚·卡利亚多 (Natalia Criado)

英国伦敦国王学院信息学系、自然和数学科学系助理教授。  
natalia.criado@kcl.ac.uk

译 者：

胡欣宇

CCF 专业会员。山西云时代技术有限公司高级工程师。主要研究方向为物联网和人工智能。  
huxinyu109@126.com



岳亚伟

CCF 专业会员。山西农业大学软件学院讲师。主要研究方向为图像处理、机器视觉。  
yue123161@sxau.edu.cn

校 对：

周若宸 浙江大学智能系统安全实验室

王家乐 浙江工商大学

(本期译文责任编辑：姜波)

## 参考文献

- [1] Acquisti A, Gross R. Imagined communities: Awareness, information sharing, and privacy on the Facebook[C]// PET 2006: Privacy Enhancing Technologies. 2006:36-58.

更多参考文献：<http://dl.ccf.org.cn/cccf/list>



# 学会论坛

CCCF 2019年第1期

杜子德

## 专委发展的历史性进步

2018年中国计算机学会(CCF)各专业委员会对CCF总部财政贡献统计结果出笼,专委上缴总部财政总数是142万元。这是历史性的突破,因为此前,专委基本未向总部上缴过结余,当然,学会也未要求过。

2018年1月,CCF常务理事会通过决议,决议称,CCF专业委员会及其他分支机构承办学会的活动,应按照预算的10%上缴总部。通过这个决议实在不容易,而让最具活力和经营能力的专委接受此决议更不容易。对活动(会议)经营能力强的专委而言,如果上缴财政收入的一部分,就意味着要多开源10%,而对于经营能力较差的专委压力就更大了。尽管如此,各专委对这项决议基本坦然接受(当然也只能接受)。不过,考虑到历史原因和各专委发展的不均衡性,常务理事会给出的是“应”而不是“须”,换句话说,有多就多缴,有少就少缴,没有就不缴。那么,这样岂不乱套!?如果都不缴呢?其实不会!实践证明也不是这样。在过渡期,要给专委机会调整,首先是理念的调整,其次是运作方式的改变,习惯后就好了,就如同我们要交个人所得税一样。

专委向总部缴钱(overhead)是国际各个学术组织的惯例,IEEE、ACM均如此。不同组织规定的比例不一,最高收取预算的20%,最后的结余还要和总部分成。CCF只收10%,且还是“应”,可见力度是不大的。就绝对数而言,这142万元也只占CCF年收入的2%略强,显得微不足道,但意义重大。

首先,专委不仅从组织上是从属于学会的一部分,其经营也是学会的一部分。如果把学会看成一个公司,那么每个专委就是这个CCF公司的一个个车间。如

果每个车间生产了产品自行上市销售,把钱放在自己口袋,这个公司还怎么生存?所以,上缴钱的意义在于总部和分支机构的互动,让分支机构理解自身的义务和责任以及学会和分支机构的主从关系。

其次,通过上缴经费和算经济账让专委负责人有经营的概念(包括定价和销售)。传统上,专委喜欢声称收支平衡、没有结余,认为没有结余表明组织者清廉。但现在不行了,各专委负责人不仅学术活动过硬,还要经营过硬,活动不仅要结余,还要上缴。从上缴的多少看你经营能力的强弱。

第三,学会通过掌握第一手财务数据,规范各专委的财务管理,了解各专委的经营过程和存在的问题。

此前不少专委办会叫苦连天,主要是缺钱,CCF也给各专委撒钱资助过,但“狼多肉少”,不解决根本问题。究其原因,就是专委没有产品经营概念,没有很好的商业模式,把会议就当会议了,于是总是缺钱。如果把专委的会议等活动(events)都看成为软产品,让专委来经营,经营好就鼓励,经营不好就关门,恐怕这样就简单多了。

学会收钱听起来似乎不雅,学会怎么一天到晚盘算钱呢?一个学术组织固然有崇高的使命,但没有经费的保证恐怕一天都过不下去,而政府部门也会让你关门歇业的。

因此,做好学会首要的是先打掉不屑看钱的假清高,大胆谈钱。教授也要懂得商业模式,否则不但科研经费没有着落,自己的工资也拿不到,这样“孔乙己”式的“学者”也只能是大家笑柄而已。

为专委的发展叫好!

# 读编往来

## ■ 2018年第12期主编评语《融合的力量》

李国杰院士提到的刘积仁教授在CNCC报告中所指出的“现有的从事大数据的人对制约技术创造价值的因素了解的太少。”我对此深有体会。目前信息技术的发展中，工业软件、工业信息化和智能化、工业互联网、智能制造是一个趋势，而对我国而言，由

于种种原因，最基础的工业软件等领域目前还处于非常初级的阶段，原因并不是我国缺乏优秀的软件开发人才，而是真正懂行业、具有经验知识的工程师太少。因此，产学研用融合，技术与市场融合，技术与行业融合，都成为迫切需要和必然趋势。

## ■ 2018年第12期专题

### 《数字对象与互联网》

随着国内计算机科学的快速发展，网络设施的日臻完善，特别是5G网络的部署，新一代互联网必将在国内茁壮成长、壮大，与世界先进各国并驾齐驱。DO数字对象，即使现在用不到，将来可能也会用到。这种面向未来，服务社会的专业精神，值得我们学习。

### 《发展数字经济值得深思的几个问题》

◆ 发展数字经济，笔者认为人工智能、区块链、云计算和大数据这四种技术不可或缺，其中大数据在数字经济中是生产资料，人工智能是生产力，云计算是生产工具，而联系它们的纽带就是扮演生产关系的区块链，这些技术将构成数字经济的技术基石。当然，数字经济是一个巨大的范畴，目前没有明确和权威的定义，它不仅仅涉及到技术，还涉及到隐私、安全以及法律法规等方方面面，还需要政府、研究院所、社会组织以及各类专家、学者等给予更多的关注和推动。

◆ 特别高兴地看到我国在大数据与人工智能基础层的几项技术突破，也衷心希望企业界能够介入这些方面的工作，使其商业化。从现代社会的发展来看，一个行业服务于人类社会，取得良好的社会回报，是需要一大批企业来承担技术的产品化、商品化和社会化部署、维护及升级换代。

### 《当大数据遇到商业模式》

◆ 很欣赏东软集团刘积仁董事长的这篇报告，言简意赅，表意清楚，体现了产业界高效做事风格。大数据时代，谁拥有且利用好了数据，就是财富的拥有者。真实有效的数据加上高效率的信息处理技术必不可少，而区块链+人工智能将大数据的价值发挥得更好。商业模式方面，笔者认为一定要考虑到参与商业生态的各个对象，要让数据的提供者也享受数据确权后所产生的红利。

◆ 文章提出的“要了解制约技术创造价值的因素”这一观点值得重视。长期以来，国人（尤其是政府和高校）存在“创新仅仅是科学或技术的创新”这样的认识误区。但是，国内外的实践已经表明，仅仅有科学技术创新还远远不够，这就是文章中提出的超越技术思维（即了解社会、了解生态、了解其他行业）。东软集团在中国27年的大数据实践经验揭示了中国数字化转型需要重视的三要素：洞察社会痛点，领悟政策精神，拥抱数据技术。

◆ 文章为各领域大数据的研究、发展及价值的产生指明了大的方向。笔者联想到，在我国的中医治病领域，如果能够把各位名医大师利用望闻问切理论为病人诊断开方治病的经验构成一个大数据系统并加以应用，不仅能够提高中医治病效果，或许还能够使我国的中医学在世界范围得到认可，从而获得更大的价值。当然，实现这些还有很大难度，

需要在大数据应用中不断总结经验并提出解决方案加以实现。

### 《如何为未来的人工智能做好准备》

认同做人工智能方面的研究不能太注重算法而不注重解决实际问题的观点。文章中关于人工智能

未来7大趋势对改进我们未来的教育很有启发，尤其是要注意4C培养的建议。此外，中国社会和经济环境有其特殊性，因此，未来中国的人工智能发展一定要与中国具体国情相结合，也需要进行相应的“供给侧结构性改革”。

## 2018年第12期专栏

### 《工业技术进步没有捷径》

改革开放四十年来，我国工业不断发展，取得了各种各样的成绩。然而要进一步保证经济和人口平稳健康，就必须实现新的产业升级，自然会引出一些新的竞争关系。所幸的是科学技术能够借助于互联网得以有效传播，我国人口存量也能提供丰富的劳动力。但是不同于互联网和房地产，工业相关的技术、人才、配套商业、社会认同等都需要平凡的积累，踏踏实实地完成。

### 《电脑前传(1)：信息》

◆ 关于信息的本体论和认识论，使我想起另一个典故。一次王阳明与友人同游，友人指着岩中花树问：“天下无心外之物，如此花树在深山中自开自落，于我心亦何相关？”王阳明答：“你未看此花时，此花与汝同归于寂；你既来看此花，则此花颜色一时明白起来，便知此花不在你心外。”这一故事中，友人说的是本体论，岩中花树不因人看到与否而存在，而王阳明说的是认识论，岩中花树因人看而“颜色一时明白起来”。

◆ 黄铁军老师采用传记写法，生动形象地介绍了信息的相关知识。对于每个方向的研究者而言，了解一些计算机大历史大背景下的人文和科学知识都是非常有意义的。美中不足的是，有一个公式采用了插图，导致显示模糊。

◆ 笔者认为信息还应该包含更高、更深的实体存在，比如事物的本质。实际上，世界的本质，就是最终的信息。针对信息的数字化，计算机专业的学者可以有重点地提炼可描述的数字特征，用于科学的研究和产品开发。针对更广义的真理阐述，还是交给更专业的智慧人，这是智慧的选择。

### 《教育系统》

认同“幻想性和挑战性揭示了教育计算机化的有趣本质”的观点。就个人经历而言（1981年参加全国高考），恢复高考前，笔者的小学和初中教育更具有幻想性而缺乏挑战性，而恢复高考之后的学校更具挑战性而缺乏幻想性。数字经济时代，中国学校教育如何与时俱进，适应教育计算机化的幻想性和挑战性的有趣本质值得进一步探索和实践。

## 2018年第12期动态

### 《AI作恶，是世界末日还是杞人忧天？》

当前既然人们已经意识到了AI具有作恶的可能性，那么就应该在世界范围内尽快设立对AI研发的管理规范，从伦理、道德、法律、技术等方面对AI产品的研发进行管控，以备恶劣事件的发生，防患于未然。从另一方面看，由什么级别的人、成立什么样的组织来管理和控制AI作恶的问题，确定什么样的AI研发属于作恶范围，怎样严惩研发作恶AI

的人等等一系列问题，都是当前迫切需要研讨和解决的问题。在信息技术快速发展的当今，或许应该设立管控新技术发展过程中出现严重问题的相关研究领域及相关专业。

### 《ECCV2018——计算机视觉领域的学术盛会》

ECCV作为顶级会议，背景介绍可以简单一点，希望对研究领域和前沿做更多、更深的讲解。

## 2018年第12期译文 《ACM道德规范与专业行为准则》

这篇文章使笔者对 ACM 的敬意油然而生。在高度发达的资本主义社会，技术社团能够始终如一地支持公众利益，值得我们借鉴和学习。

## 2018年第12期学会论坛 《如何主持会议》

虽然文章提到的是学会的会议，在科研和产品开发中，也有对应的会议形式。技术骨干、项目责任人往往是会议的主持人。有效管理会议时间，会前沟通，处理突发事件，都是主持人要缜密考虑的。工作会议会要求一些细化的讨论和决策，为此，需要主持人对会议内容、议题汇报各方有充分的把握。

## 其他

◆ 建议开辟论文撰写指导专栏。提供高水平论文撰写方面的案例及解析，为青年教师的成长提供快速通道，让专家学者的经验得以传承。

**编辑部回复：**谢谢建议。本期动态栏目即刊登了此方面的文章《CIKM 2018 最佳论文是怎样炼成的》。以后我们尽量多邀请论文撰写指导方面的文章。

◆ 内文排版非常棒。建议将封面刊名亮度提高，否则与背景的暗色调混在一起，不够抢眼。建议封面和封底的配色最好相一致，或者相类似，以免造成突兀之感。

**编辑部回复：**谢谢建议。以后我们注意改进。

(2018年第12期参与评刊的有：李振华、李珍妮、廖勇、刘宇擎、吕腾、时成阁、万江平、易小琳、周果)

CCCF《读编往来》向广大读者开放  
欢迎分享观点、提出建议

- 登录学会网站在线阅读、评论
- 关注微博“CCF 通讯”

联系：cccf@ccf.org.cn 来信请注明：姓名、单位

## 2018年度CCCF “积极评刊奖”评选揭晓

CCCF 编辑部设立的“积极评刊奖”，旨在奖励积极、认真地为刊物撰稿的评刊员，每年评选一次。根据 2018 年评刊员的评刊次数和质量，编辑部评选出突出贡献奖 11 名，积极参与奖 12 名，并为其颁发证书和 CCF 定制的奖品。获奖名单如下(按姓氏拼音排序)。

### 突出贡献奖

#### 一等奖

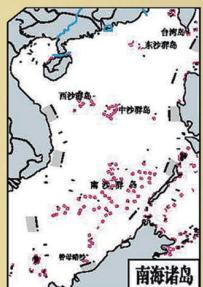
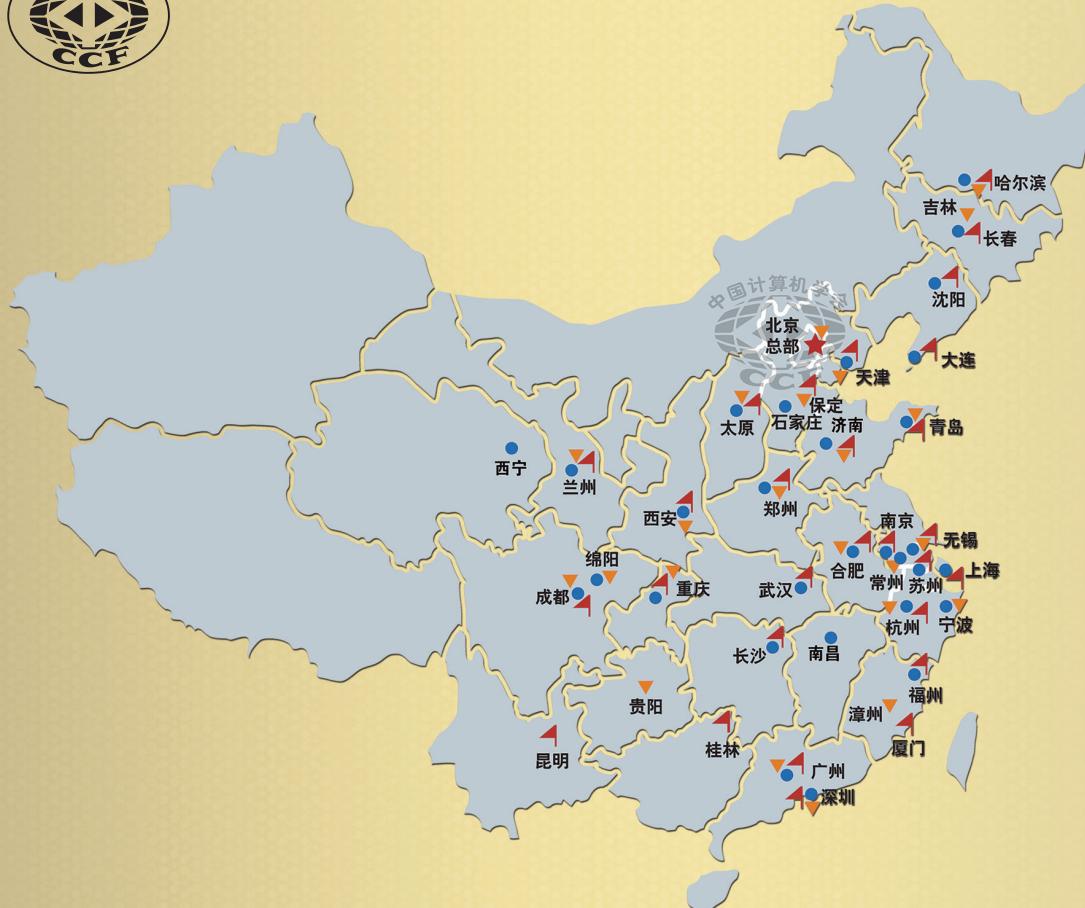
廖勇 CCF 高级会员，重庆大学副研究员  
时成阁 CCF 专业会员  
发码行公司信息技术中心总监  
万江平 CCF 高级会员，华南理工大学教授

#### 二等奖

陈盈 CCF 专业会员，台州学院副教授  
李振华 CCF 专业会员  
浙江商业职业技术学院副研究员  
刘恒业 郑州轻工业学院助理工程师  
刘宇擎 CCF 学生会员，大连理工大学研究生  
吕腾 CCF 高级会员，安徽新华学院教授  
易小琳 CCF 高级会员  
北京工业大学高级工程师  
张福生 CCF 专业会员，哈尔滨学院高级工程师  
周果 中国政法大学讲师

### 积极参与奖

程飞 杜晓舟 范天龙 何剑虹 李睿  
李挺 刘宇航 秦董洪 王波 杨健  
周珂 周元欣



- ▲ YOCSEF
- CCF会员活动中心
- ▼ CCF学生分会

# 只有结成群体 才好发展专业

加入CCF/CCF会员资格延续

专业会员/高级会员/杰出会员/会士:200元/年(一次可交纳5年)

学生会员:50元/年

欢迎 微信支付

## 其他缴费方式

在线缴费 [www.ccf.org.cn](http://www.ccf.org.cn)

银行转账

开户行: 北京银行北京大学支行

户 名: 中国计算机学会

账 号: 0109 0519 5001 2010 9702 028



微 信 支 付



# 2018中国计算机学会颁奖大会

CCF Awarding Ceremony 2018

2019年1月19日（农历腊月十四）

北京金隅喜来登大酒店

CCF终身成就奖

CCF杰出贡献奖

CCF夏培肃奖

CCF卓越服务奖

CCF杰出教育奖

CCF计算机企业家奖

CCF优秀博士学位论文奖

CCF杰出工程师奖

协办：

Sponsors



Microsoft



中科曙光



ByteDance  
字节跳动



Alibaba Group  
阿里巴巴集团