

计算科学评论

WWW.COMPUTER.ORG

2018年第1期

计算科学：展望2018



假新闻的经济学 P12
机器人士兵的良知 P80

合作机构



ISSN 1673-5153



9 771673 515122

云计算

人工智能

工控机

制造业

电子
嵌入计算

传感器

互联网

3D 打印

绿色计算

图形图像

虚拟现实



互联网金融

芯片

软件架构

大数据

纳米架构

人机交互

多媒体

普适计算 MEMS



微信名：计算机人 微信号：jisuanren



环球科学 SCIENTIFIC AMERICAN

阿尔·戈尔 (Al Gore)

2007 年获诺贝尔和平奖
1990 年和 1991 年在《科学美国人》
上发表两篇文章，大声疾呼人类应该
尽早行动，抵御全球变暖



比尔·盖茨 (Bill Gates)

美国微软公司创始人
2007 年在《科学美国人》
上发表文章，预言机器人产
业的兴起



斯蒂芬·霍金
(Stephen Hawking)

著名物理学家
2010 年在《科学美国人》
上发表文章，阐述他对物理
学和我们这个世界的思考



1845 年创刊

162 位诺贝尔奖得主撰稿
传承百年的科技媒体品牌
科技精英分享智慧与见解
的全球化平台

罗伯特·布鲁斯·梅里菲尔德
(Robert Bruce Merrifield)

1984 年获诺贝尔化学奖
1968 年在《科学美国人》上发表文章



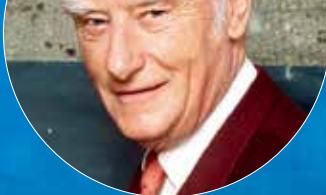
安德烈·K·海姆
(Andre K. Geim)

2010 年获诺贝尔物理学奖
2009 年在《科学美国人》上发表文章

穆罕默德·尤努斯
(Muhammad Yunus)
2006年获诺贝尔和平奖
1999年在《科学美国人》上发表文章，
介绍用小额信贷帮助贫困人口



朱棣文 (Steven Chu)
1997年获诺贝尔物理学奖
1992年在《科学美国人》
上发表文章



弗朗西斯·克里克
(Francis Crick)
1962年获诺贝尔生
理学或医学奖
多次在《科学美国人》
上发表文章

《科学美国人》独家版权内容
全球数千科技精英撰稿
为你报道最新科研趋势与最具
前景的科技成果



《环球科学》是全球著名科技媒体《科学美国人》杂志中文版，创刊170年来，它始终站在科技最前沿，在美、德、法、意、日等15个国家出版同步发行，内容涵盖生物、医学、IT、电子、能源、经济等广泛领域，是科技企业、知识阶层与政府决策者首选的科技指南。

您可采取以下方式订阅：



微店二维码



淘宝二维码

② 邮发代号：80-498

③ 各大电商，书店，报亭均有销售

更多信息，请登录官方网站：
www.huangjukexue.com

订阅热线
010-57458982

Copyright

版权

主管单位 Authorite in Charge

中华人民共和国教育部 Ministry of Education of the People's Republic of China

主办单位 Sponso

中国大学出版社协会 China University Presses Association

出版单位 Publisher

《环球科学》杂志社有限公司 GLOBAL SCIENCE MAGAZINES Co.,Ltd

社址 Address: 北京市朝阳区秀水街1号建外外交公寓4-1-21 Office 4-1-21, Jianguomen Diplomatic Residence Compound, No. 1, Xiu Shui Street, Chaoyang District, Beijing, China. 邮编: 100600

联系电话: 010-85325810 / 85325871

社长 / 总编辑 Editor-in-chief

陈宗周 Chen Zongzhou

副社长 / 副总编辑 Deputy Editor-in-chief

刘芳 Liu Fang

执行出版人 Publisher

管心宇 Xinyu Guan
张岚 Landy Zhang

资深编辑 Senior Editor

马法达 Falda Ma
刘妍 Yan Liu

特约编辑 Contributing Editor

史彦诚 Yancheng Shi
刘大明 Daming Liu
高天羽 Tianyu Gao
费麟 Yong Fei
王璇 Xuan Wang

运营中心 OPERATING DEPARTMENT

运营机构 Publisher
上海灵宸文化传媒有限公司

发行部 Circulation Department

发行总监 Circulation Director
谢磊 Xie Lei 010 - 57439192

市场部 Marketing Department

市场总监 Marketing Director
赵子豪 Zhao Zihao 010 - 85325810 - 807

广告部 Advertising Department

销售总监 Sales Director

范欢 FanHuan 010-85325871-802 010-85325981

读者服务部 Reader Service

杜君 Du Jun 010 - 57458982

印刷 北京博海升彩色印刷有限公司

如发现本刊缺页、装订错误和损坏等质量问题, 请在当月与本刊读者服务部联系调换(请将证书寄回)。

国际标准刊号: ISSN 1673-5153

国内统一刊号: CN11-5480/N

广告经营许可证号: 京朝工商广字第8144号

知识产权声明:

IEEE, IEEE Computer, IEEE中文网站的名称和标识, 属于位于美国纽约的电气电子工程师学会有限责任公司所有的商标, 仅通过授权使用。这些材料的一部分由IEEE Computer英文版翻译而来, 版权归IEEE所有, 并经IEEE授权翻译复制。

IEEE Computer杂志的中文版权归, 由美国电气电子工程师学会有限责任公司授予上海灵宸文化传媒有限公司, 并由本刊独家使用。

本刊发表的所有文章内容由作者负责, 并不代表上海灵宸文化传媒有限公司、美国电气电子工程师学会有限责任公司的立场。

本刊内容未经书面许可, 不得以任何形式转载或使用。

编辑团队

流程编辑

Carrie Clark

colark@computer.org

资深编辑

Chris Nelson

编辑

Lee Garber, Meghan O'Dell
Rebecca Torres, Bonnie Wylie

多媒体编辑

Rebecca Torres

设计与印刷

Carmen Flores-Garvey
Erica Hardison

封面设计

Matthew Cooper

资深广告经理

Debbie Sims

产品与服务总监

Evan Butterfield

会员总监

Eric Berkowitz

出版人

Robin Baldwin

主编

Sumi Helal

Lancaster University,
sumi.helal@computer.org

副主编

Elisa Bertino

Purdue University,
bertino@cs.purdue.edu

副主席, COMPUTING PRACTICES

Rohit Kapur

Synopsys, kapurfamily04@gmail.com

副主席, PERSPECTIVES

Jean-Marc Jézéquel

University of Rennes jean-marc.jezequel@irisa.fr

副主席, SPECIAL ISSUES

George K. Thiruvathukal

Loyola University Chicago,
gkt@cs.luc.edu

2018 IEEE计算机协会主席

Hironori Katashara

Waseda University,
kasahara@waseda.jp

行业编辑

大数据和数据分析

Naren Ramakrishnan

Virginia Tech

Ravi Kumar

Google

云计算

Schahram Dustdar

TU Wien

计算机结构

David H. Albonesi

Cornell University

Greg Byrd

North Carolina State University

Erik DeBenedictis

Sandia National Laboratories

信息物理系统

Oleg Sokolsky

University of Pennsylvania

数字健康

Christopher Nugent

Ulster University

顾问委员会

Doris L. Carver

Louisiana State University (EIC Emeritus)

Carl K. Chang

Iowa State University (EIC Emeritus)

Theresa-Marie Ryne

Consultant

Bill Schilit

Google

Savitha Srinivasan

IBM Almaden Research Center

Ron Vetter

University of North Carolina Wilmington (EIC Emeritus)

Alf Weaver

University of Virginia

领研网

专注科研招聘与学术分享

www.linkresearcher.com



访问领研网获取招聘信息、一手科研资讯，动态追踪学者研究成果



领研网是《科学美国人》中文版《环球科学》旗下科研招聘与学术分享网站，
服务百万学者，为高校、机构与科技企业搭建人才桥梁，
助力学者传播优秀成果，提升学术生涯。

合作请致电: 010 - 85321181, 或邮件 contact@linkresearcher.com



科研求职者可扫描二维码
注册，即可在站内信获得
价值千元就业礼包



科研机构 / 科技企业可扫描二维码
成功注册可免费发布职位信息

计算科学评论



18

导读

展望 2018

2018 年的年度展望专辑突出展示了几个“信息化为知识”的典范，例如，共享内存架构的回归，还有普适计算是怎样最终找到合适的操作系统的。

舒米·埃尔尔 (Sumi Helal)，兰卡斯特大学

2018 年第 1 期

封面故事

24

34

46

从英国脱欧到特朗普：社交媒体在民主中的作用

撰文 温迪·哈勒 (Wendy Hall)、莱明·蒂纳蒂 (Ramine Tinati)、威尔·詹宁斯 (Will Jennings)，南安普敦大学

信息科技前沿： 香农遇上图灵

撰文 沃伊切赫·希潘诺夫斯基 (Wojciech Szpankowski)、安纳斯·格拉玛 (Ananth Grama)，普杜大学信息科技中心

朝着普适操作系统发展：一个软件定义的角度

撰文 梅宏、郭耀，北京大学

2018 年第 1 期

目录



54

再谈可扩展相干共享内存

撰文 C. 戈登·贝尔 (C. Gordon Bell)，软研究院（退休）、艾克·纳西 (Ike Nassi)，TidalScale 公司和加州大学圣克鲁兹分校梶

趋势

12 “假新闻”经济学

撰文 尼尔·科舍特里 (Nir Kshetri) 杰弗里·沃阿斯 (Jeffrey Voas)，北卡罗来纳大学格林斯伯勒分校，IEEE 会士，

访谈

66 自然语言处理为什么这么难？

——专访两位顶级 NLP 大师

撰文 管心宇、林宗辉

竞赛

74 创客革命

撰文 斯库特斯·威利斯 (Scooter Willis)，TechGarage

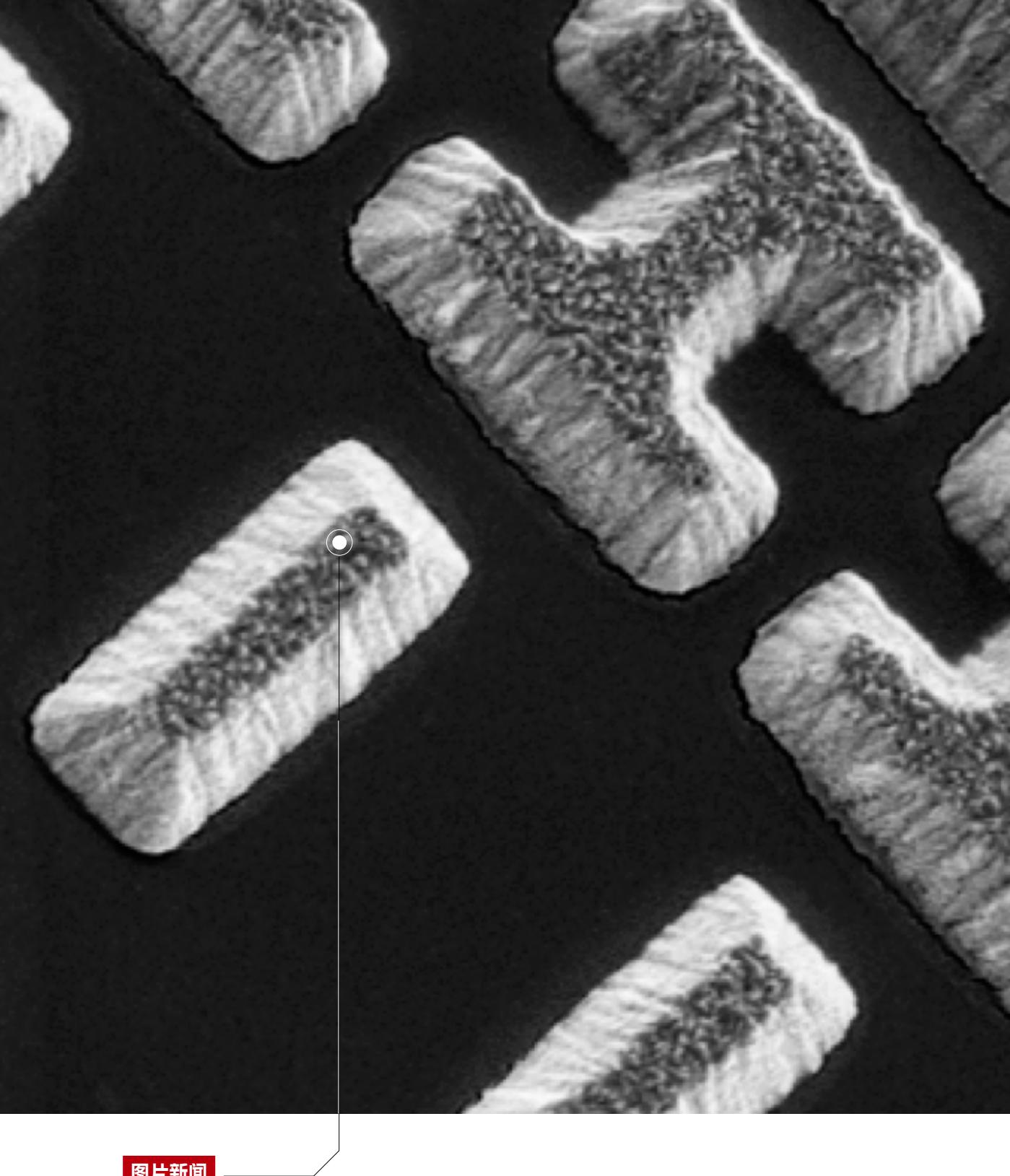
专栏

78 人工智能和攻击 / 防御平衡

撰文 布鲁斯·施奈尔 (Bruce Schneier)，哈佛大学

80 机器人士兵的良知

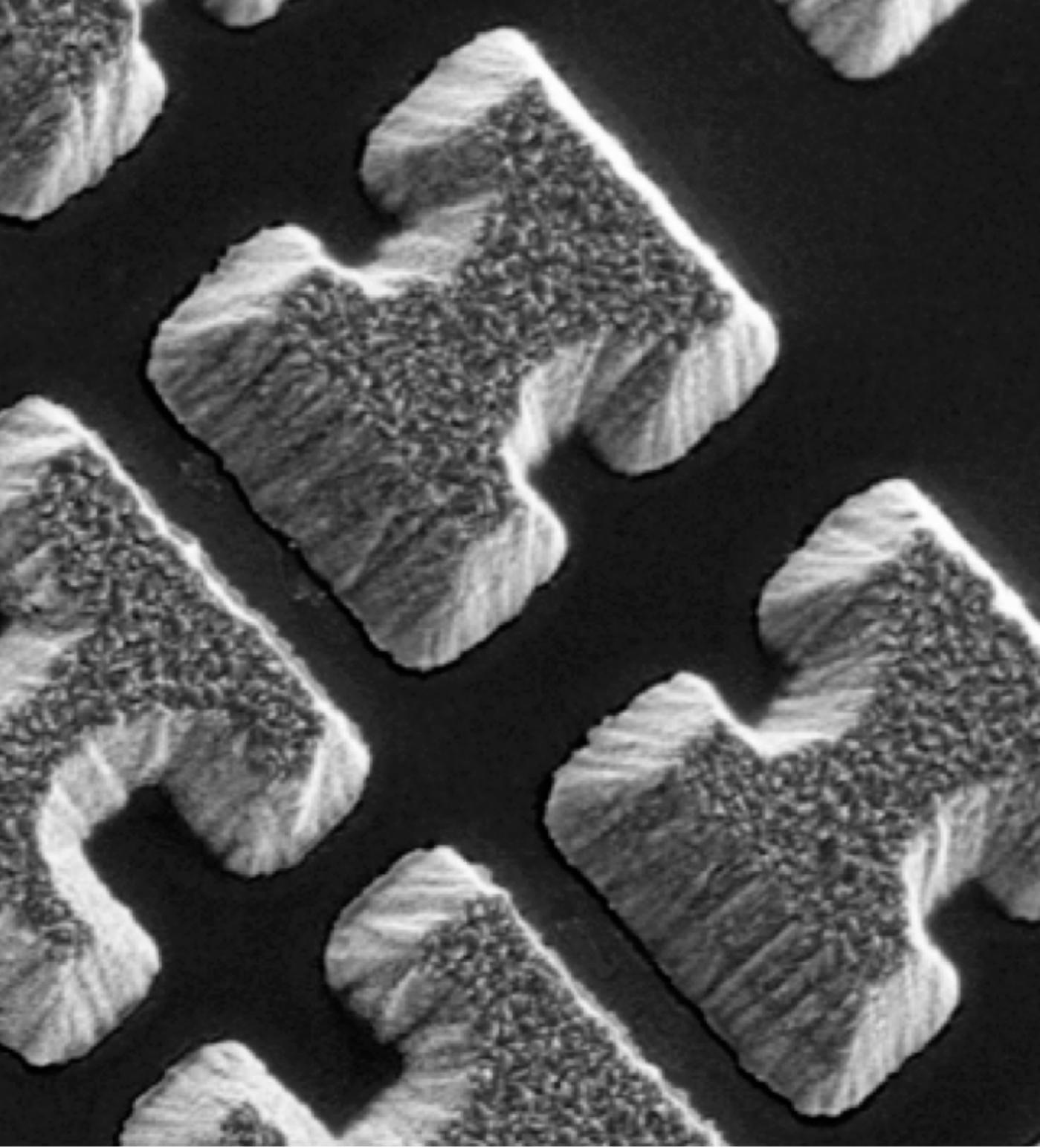
文 | 查尔斯·戴 (Charles Day)，《今日物理》(Physics Today)



图片新闻

成像新方法

麻省理工学院等机构的研究人员共同开发了一种利用中红外波段光谱拍摄图像的新方法，可应用于热成像、生物医学传感和自由空间通信等领域。



上图为扫描电子显微镜图像，显示了精心设计的硫族玻璃，沉积在透明衬底上。这些图形被研究人员称为“元原子 (meta-atoms)”，它们决定了中红外光穿过材料时的折射情况。

来源：哈佛大学

论文：*Ultra-thin high-efficiency mid-infrared transmissive Huygens meta-optics*. *Nature Communications*, 2018; 9 (1) DOI: 10.1038/s41467-018-03831-7

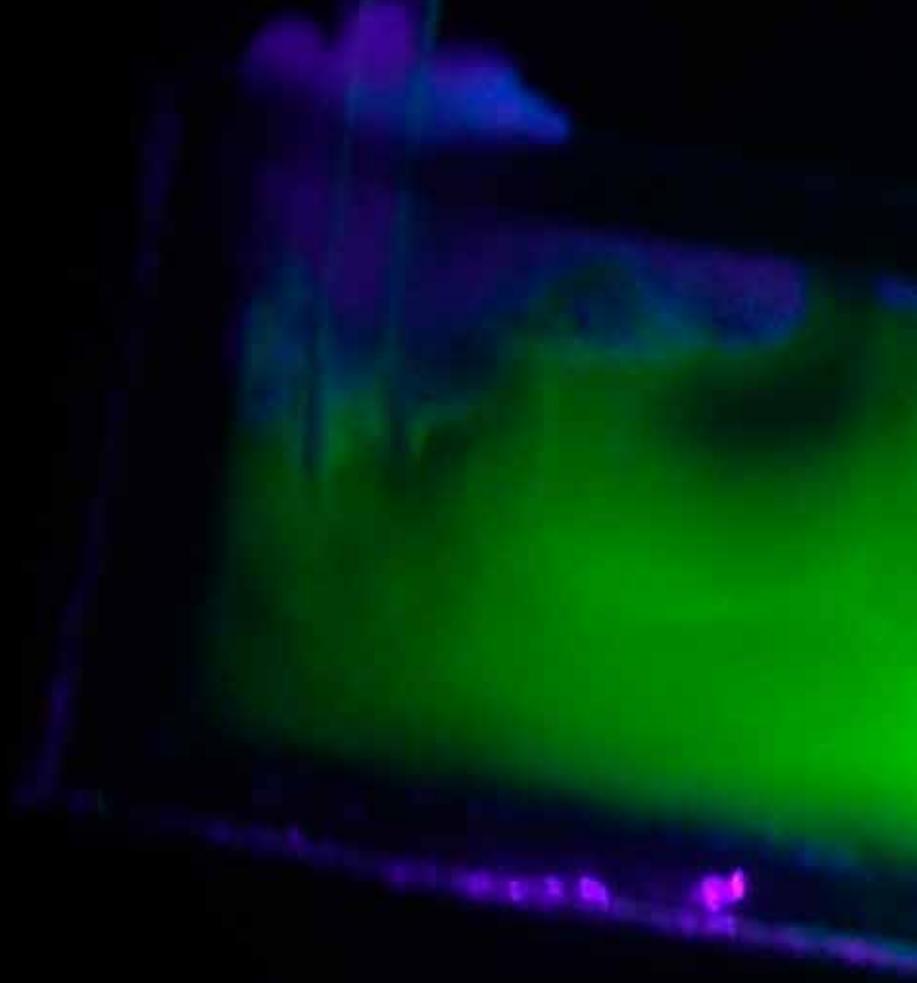
图片新闻

不用马达的机器鱼

虽然已经有很多小型水下机器人和机器鱼被付诸应用，但是发动机的噪音会对水中生物造成影响。加州大学圣迭戈分校的研究人员用人造肌肉设计了大约 10 厘米长的机器鳗鱼，不用发动机也能以 1.9 毫米 / 秒的速度在水中移动。

来源：加州大学圣迭戈分校

论文：*Translucent soft robots driven by frameless fluid electrode dielectric elastomer actuators. Science Robotics, 2018; 3 (17): eaat1893 DOI: 10.1126/scirobotics.aat1893*







“假新闻” 经济学

文 | 尼尔·科舍特里 (Nir Kshetri) 杰弗里·沃阿斯 (Jeffrey Voas)，北卡罗来纳大学格林斯伯勒分校，IEEE 会士，
译 | 郑米兰

今天，复杂的算法已经成功地生成了被知名会议接受的虚假论文。同时，新闻机构已经开始使用人工智能 (AI) 系统来生成合法的文章，报道诸如体育赛事比分这类例行程序式的题材。这些进展都表明，在不久的将来，更高级的算法将自动生成“可信”的错误信息。

虚假信息主要有两种形式：误传和假消息。假消息是为了欺骗而传播的错误信息。误传则仅仅是不正确的信息，例如，“他们告诉我的开会时间是错的，但我知道这不是故意的。”在本文中我们探讨的是“假新闻”——当前的一种（据我们所知）仍然是手动生成的假消息。

关于假新闻

近年来，通过社交媒体传播的新闻一直是人们关注的焦点。一项研究表明，62% 的美国成年人从社交媒体获得新闻²，超过40% 的人从Facebook上获得新闻。

据估计，在2016年美国总统大选的

最后三个月里，Facebook 上阅读量最高的虚假选举新闻比《纽约时报》、《华盛顿邮报》、《赫芬顿邮报》或 NBC 等主要新闻媒体上的最热门报道吸引了更多的点击量。在那段时间里，来自假新闻网站的20则阅读量最高的虚假选举新闻在 Facebook 上共产生了870多万次分享、回应和评论，而19个主要新闻网站的最热门新闻仅为730多万次。⁴据曾运营假新闻网站的人说，广告商在选择刊登广告前并不关心这些网站的质量，只要网站符合最低要求（例如没有色情内容）。大多数广告商的主要要求是流量来自人而非机器人。⁵

大多数与社交网站上的虚假内容相关的欺诈行为并不是需要“超级黑客”技能的高科技犯罪。相反，“攻击者”采用的是社会工程和欺骗手段。请注意，“欺骗”所涉及的是心理过程而非技术过程。其方法是信息的发送者（例如假新闻的制造者）给信息接收者（比如读者）带去“这些新闻是真实的”的错误印象。由此，虚假内容的制

造者操纵了读者的行为（例如说服他们点击网络链接）。

社交网站上的虚假内容会产生经济、政治和社会后果。《华盛顿邮报》于2016年11月24日刊登的一篇文章称，PropOrNot专门监测那些报道俄罗斯政治宣传的网站，其研究人员说，在2016年美国总统选举期间，超过200个网站经常性地推广俄罗斯的宣传。这些网站至少有1500万美国读者。来自这些网站的报道在Facebook上被浏览超过2.13亿次。⁶

其他假新闻制造者的组织性较差，但受金钱激励的驱使。已知他们在格鲁吉亚共和国和马其顿等国家运营这类网站。例如，在2016年美国总统选举前的一年里，马其顿的韦莱斯镇（人口4.5万）的居民打造了140多个美国政治网站。大多数域名看上去都像是美国内网站，例如WorldPoliticus.com、TrumpVision365.com、USConservativeToday.com、DonaldTrumpNews.co 和 USA

DailyPolitics.com.⁷。

分析假新闻的价值

用众所周知的经济分析方法来看，当满足以下公式时，欺诈者会参与假新闻的制造和管理⁸：

$$M_b + P_b > I_c + O_{1c} + P_c + (O_{2c} \pi_{arr} \pi_{con}), \quad (1)$$

其中：

- M_b = 参与利用假新闻的欺诈案带来的金钱利益；
- P_b = 参与利用假新闻的欺诈案带来的心理（非经济）利益；
- I_c = 直接投资成本；
- O_{1c} = 参与制造和管理假新闻的机会成本；
- P_c = 参与利用假新闻的欺诈案的心理成本；
- O_{2c} = 被定罪的金钱机会成本；
- π_{arr} = 被捕的可能性；
- π_{con} = 被定罪的可能性。

公式 1 中最右边的乘积项 $O_{2c}\pi_{arr}\pi_{con}$ 也叫做“预期惩罚效应”。

现在，我们来看看公式中的前五个变量，以更好地理解这个公式：

- M_b 。鉴于前面提到的原因，制造假新闻可以赚到钱。例如，2016年8月至11月期间，一名马其顿少年从两个亲特朗普网站赚取了约1.6万美元⁹；一位主要以Facebook为发布平台的假新闻作者据报道每月从AdSense获得1万美元。¹⁰

- P_b 。格鲁吉亚一名假新闻作者撰写的文章通过流露他个人更喜欢特朗普来吸引特朗普的支持者。在这种情况下，他获得了心理上的好处。
- I_c 。假新闻作者拼凑来自其他网站的故事，所花费的精力、时间和资源都很有限。他们并不需要吸引读者长时间地停留在某个网页上——所有假新闻作者只希望读者点击一次链接。
- P_c 。《纽约时报》最近采访了格鲁吉亚的假新闻作者。受访者说，他们的主要动机是赚钱。因此，他们

如何打破环路

假消息的制造者、消费者和仲裁者有相互强化的作用。这导致了一个假新闻生态系统。图1显示了这个环路中的要素。应注意的是，社会政治和经济因素使得炮制假新闻对世界上部分地区里的众多年轻人产生了吸引力。此外，一般人群对社交媒体上的新闻消费增加了，而其中大部分人都缺乏客观评估社交媒体内容和来源的能力。最后，各类仲裁者的行为鼓励了（或至少没有劝阻）假新闻的制造。

一项研究表明，62%的美国成年人从社交媒体获得新闻，超过40%的人从Facebook上获得新闻。

不会感到内疚。而且，他们相信自己的行为没有对他人造成身体伤害。¹¹

• O_{1c} 。炮制假新闻是一项具有吸引力的经济活动，对于高失业率经济体中的年轻人尤其如此。例如，在2017年第一季度，马其顿的失业率为22.9% ([bit.ly/2fPanJ9](#))，那里的劳动者的平均月收入为371.9美元。

然后是公式1中的预期惩罚效应 ($O_{2c}\pi_{arr}\pi_{con}$)。它是指因伪造新闻而被捕和定罪的可能性。在马其顿和格鲁吉亚这样的国家（大多数假新闻都源自这里），这种可能性小到可以忽略不计。而格鲁吉亚的假新闻制造者没有违反该国任何法律。¹²

消费者

消费者可能缺乏客观评估信息及其来源的准确性和质量的能力。一项研究发现，大约8%的成年人“愿意相信任何听起来似乎合理、且符合他们对政治中英雄和恶棍先入之见的事。”¹²

斯坦福大学教育研究院的研究人员调查了7000名刚从中学毕业即将上大学的学生，看看他们对于自己在互联网上看到的信息判断力如何，结果发现千禧一代“很容易被骗”。这项被《华盛顿时报》引用的研究发现，尽管这些学生在使用手机和电脑等数字技术上非常熟练，但他们大部分人都缺乏辨别内容、区分事实和非事实、评估来源可靠性的能力。许多被调查者不能区分广告和报道、假新闻和基于事实的新闻。大

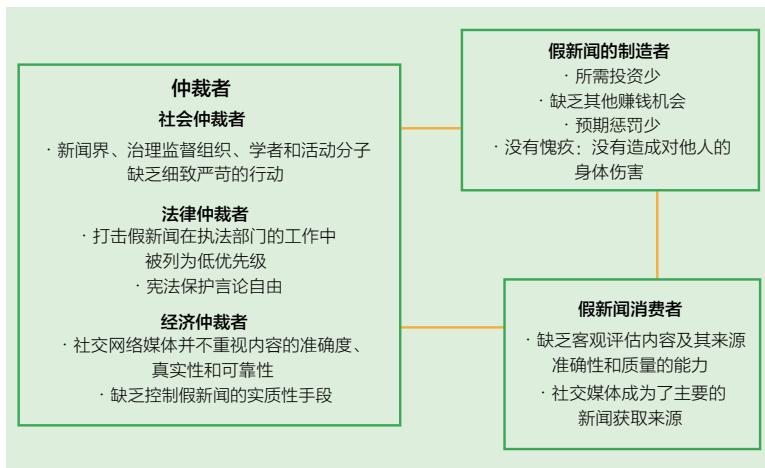


图1. 假新闻制造的恶性循环。在世界上某些地区，社会政治和经济因素使得制造假新闻成为了对年轻人具经济吸引力的选择。一般人群在社交媒体上阅读新闻的数量增加了。各种仲裁者的行为没有阻止假新闻的产生。

多数人不知道或不关心政治偏见。¹³

制造者

如前所述，制造虚假信息所需的投资很小。假新闻的增长也与宪法保护言论自由有关。例如，在美国，第一修正案使得执法机构无法监管自由言论。其他国家也有类似的言论保护措施。这意味着社交媒体用户实际上可以用任何方式表达自我和发表观点。如前所述，马其顿和格鲁吉亚等国的假新闻作者可能并没有违反自己国家的任何法律。

仲裁者

研究人员已经确定了三类“仲裁者”——社会的、法律的和经济的。¹⁴

社会仲裁者包括新闻界、治理监督组织、学者和活动分子。一个关键问题是这些团体缺乏有组织和系统性的应对措施。此外，第三方事实核查机构和媒体等团体尚未做出细致严苛的行动。¹⁵

执法机构和政府机构可以是强制执行规则和法规的合法仲裁者，但打击假新闻并没有被他们列为优先级任务。例如，FBI正投入更多精力和资源来防止与伊斯兰国组织（IS）相关的攻击。他们的一大关注重点是开发一个线民网络，可以向FBI提供有关密谋事件的信息，并将互联网上的激进策划者与为FBI工作的线民连接起来。¹⁶在他们的工作中，与假新闻相关的威胁似乎没有那么紧要。

商业组织可以是经济仲裁者。Facebook的商业模式基于用户的参与度、互动和内容消费。用户阅读、点击、分享和参与的内容越多，它获利就越多。相比之下，准确度、真实性和可靠性并不那么受重视。¹⁷

2014年，60多家新闻媒体、学者和社交网络发起了一个“信任工程”（Trust Project），旨在“恢复新闻媒体在公民生活中受信任的角色”。这个联盟的目标之一是通过建立准则、信任指

标和排名算法来使已知的可信赖新闻来源获得更多重视，从而帮助消费者和科技公司。

诸如Facebook和Twitter之类的商业平台的功能与从前的媒体大不相同。一个关键的问题是，这些网站缺乏严格的编辑人员发贴前审核其质量。¹⁸用户无需经过第三方筛选、事实核查和编辑选择就能相互分享图片、视频、文字信息、新闻和其他内容。一些用户的粉丝和阅读者人数之众堪比《纽约时报》、福克斯新闻和CNN等大媒体。

社交网站是否应为其用户消费的新闻和信息扮演守门人的角色？这似乎更是一个哲学问题而非技术问题。¹⁹正如Facebook首席执行官马克·扎克伯格所言：“一些恶作剧可以被完全揭穿，但更多内容，包括来自主流消息源的内容，常常基本属实，但会弄错或遗漏一些细节。还有更多故事表达的观点会遭到许多人反对，即使符合事实也会被人标识为不正确。”²⁰

打破环路

如果没有适当的措施来打击假新闻，那么这个环路里的元素就会相互加强，导致公众对所有媒体和执法机构的不信任，却对假新闻日益趋之若鹜。那么，我们可以从哪里开始打破这个环路，改变与假新闻相关的成本效益计算？

巨大的社交网络服务可以率先行动来打破这一环路。它们已通过部署先进技术来实施某些举措。2017年8月，Facebook宣布启动一个专案来打击被称之为“伪装”（cloaking）的技术——一种虚假内容制造者用来误导用户的

方法。这些内容制造者会伪装广告或帖子的真实目标页面或目的地页面的真实内容，以规避或绕过Facebook的审查程序。²¹新的举措依赖人工智能并扩展了人员审查流程。公司并表示将屏蔽使用“伪装”技术的广告商和网页。在媒体大量报道了假新闻事件后，谷歌和Facebook已经切断了与假新闻网站的广告业务。²⁰这可能会降低撰写假新闻的财务激励。

乌克兰的一群记者创建了“StopFake News”，专门揭穿假新闻。该组织最早由Kiev Mohyla大学的教授和记者发起，自视为一家提供公共服务新闻的传媒机构。²²到了2017年7月，其记者每周都会做一次有关俄罗斯媒体对乌克兰的虚假和错误报道的回顾。该节目在Hromadske电视台播出，《基辅邮报》刊登了该节目的英文版。

23

此外，客户可以向企业施压，要求它们提供筛查假新闻方面的支持，视之为企业社会责任的一部分。²⁴

了解假新闻中的各种作用因素是很重要的。对于以美国读者为受众的假新闻作者，“醒目”的标题吸引来了即时点击和广告费；营利性社交网络的动机可疑；读者很容易被欺骗，欠缺能力或方法来评估信息及其来源的可靠性。

像StopFake这样的反假新闻组织做出了一些努力。公众意识提升和教育活动可能会有所帮助。但是，我们须意识到，目前信息被推向公众的速度可谓势不可挡。比如，CNN和其他主要媒体都在不间断地发布“突发新闻”——谁

能跟得上这样的节奏？而最重要的一点是：反假新闻的方法还不够。

参考文献

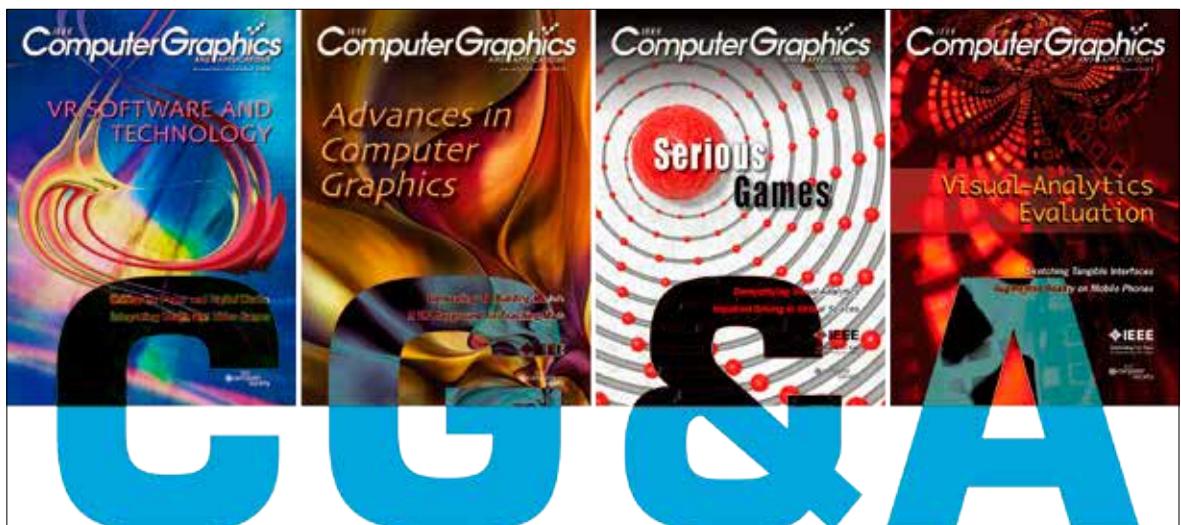
- K. Kakaes, “How Gobbledygook Ended Up in Respected Scientific Journals,” *Slate*, 27 Feb. 2014; slate.me/1cd92Dc.
- J. Gottfried and E. Shearer, *News Use Across Social Media Platforms* 2016, Pew Research Center, 26 May 2016; pewrsr.ch/27TOfhz.
- R. Cellan-Jones, “Facebook, Fake News, and the Meaning of Truth,” *BBC News*, 27 Nov. 2016; bbc.in /2gkuI4j.
- C. Silverman, “This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook,” *Buzzfeed News*, 16 Nov. 2016; bzfd.it/2g9hUPR.
- C. Silverman, J. Singer-Vine, and L.T. Vo, “In Spite of the Crackdown, Fake News Publishers Are Still Earning Money from Major Ad Networks,” *Buzzfeed News*, 4 Apr. 2017; bzfd.it/2hK5Wzy.
- C. Timberg, “Russian Propaganda Effort Helped Spread ‘Fake News’ During Election, Experts Say,” *Washington Post*, 24 Nov. 2016; wapo.st/2grg8Iw.
- C. Silverman and L. Alexander, “How Teens in the Balkans Are Duping Trump Supporters with Fake News,” *Buzzfeed News*, 3 Nov. 2016; bzfd.it/2ynwU7o.
- N. Kshetri, “The Simple Economics of Cybercrimes,” *IEEE Security & Privacy*, vol. 4, no. 1, 2006, pp. 33–39.
- S. Subramanian, “Inside the Macedonian Fake-News Complex,” *Wired*, 15 Feb. 2017; bit.ly /2kphWY3.
- A. Ohlheiser, “This Is How Facebook’s Fake-News Writers Make Money,” *Washington Post*, 18 Nov. 2016; wapo.st/2g8vLHg.
- A. Higgins, M. McIntire, and G.J.X. Dance, “Inside a Fake News Sausage Factory: ‘This Is All About Income,’ ” *New York Times*, 25 Nov. 2016; nyti.ms/2lef2L.
- N. Irwin, “Researchers Created Fake News. Here’s What They Found,” *New York Times*, 18 Jan. 2017; nyti.ms/2jA6cQy.
- S. Fields, “‘Digitally Savvy’ and at the Mercy of Media Fakers,” it is important to understand the various actors in fake news. For a fake-news writer serving the computer.org/ITPro 11
- B.M. Wiesenfeld, K.A. Wurthmann, and D.C. Hambrick, “The Stigmatization and Devaluation of Elites Associated with Corporate Failures: A Process Model,” *Academy of Management Rev.*, vol. 33, no. 1, 2008, pp. 231–251.
- L. Leong, “Fighting Fake News: How Google, Facebook, and Others Are Trying to Stop It,” *TechRadar*,

趋势

- 25 May 2017; bit.ly/2fO52Se.
16. M. Zapotosky, "Why Law Enforcement Can't Get Ahead of Pizzagate and Other Online Conspiracy Theories," Washington Post, 8 Dec. 2016; wapo.st/2xfnXrF.
17. O. Solon, "Facebook Won't Block Fake News Posts because It Has No Incentive, Experts Say," The Guardian, 15 Nov. 2016; bit.ly/2gfOjWn.
- H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," *J. Economic Perspectives*, vol. 31, no. 2, 2017, pp. 211–236.
19. A. Ohlheiser, "Even Mark Zuckerberg Can't Stop the Meme that He Is Running for President," Washington Post, 3 Aug. 2017; wapo.st/2vtNXCr.
20. D. Pogue, "The Ultimate Cure for the Fake News Epidemic Will Be More Skeptical Readers," *Scientific American*, 1 Feb. 2017; bit.ly/2fPx0gp.
21. R. Leathern, "News Feed FYI: Advertising Cloaking so People See More Authentic Posts," Facebook Newsroom, 9 Aug. 2017; bit.ly/2uGntd3.
22. A.E. Kramer, "To Battle Fake News, Ukrainian Show Features Nothing but Lies," *New York Times*, 26 Feb. 2017; nyti.ms/2mvR8m9.
23. P. Stockmans, "Ukrainian Fact-Checkers StopFake: 'Fake News Is a Weapon of War,'" *Mondiaal Nieuws*, 25 July 2017; bit.ly/2uVJR5A.
24. J. Aylon, "How Corporate Citizens Can Stop Fake News and Hate News—and Help Save Quality Journalism in the Process," *The Daily Beast*, 1 May 2017; http://thebea.st/2fJgD1i.

尼尔·科舍特里 (**Nir Kshetri**) 是北卡罗来纳大学格林斯伯勒分校布莱恩商学院的管理学教授。可发邮件至nbkshetr@uncg.edu与他联系。

杰弗里·沃阿斯 (**Jeffrey Voas**) 是Digital公司的创始人、《计算机》杂志的“网络信任”专栏编辑。他是IEEE会士。可发邮件至j.voas@ieee.org与他联系。



《IEEE计算机图形及应用》(IEEE Computer Graphics and Applications, 简称CG&A)把计算机图形学领域的理论和实践联系在一起。《IEEE计算机图形及应用》提供了包括从某个特定算法到全系统实现在内的同行评议的深度报道。它为那些处于计算机图形技术前沿的人们提供了必不可少的资料。无论他们处于商界还是艺术界,这本杂志都能让他们受益。

请点击: www.computer.org/cga

中国中文信息学会、语言与知识计算专业委员会 (CCKS)



CCKS·2018

中文电子病历的命名实体识别

2018.05.01–2018.07.20

清华大学 x 哈工大 x 医渡云



扫码了解详情



2018 展望： 新兴网络和信息科学

文 | 舒米·埃拉尔 (Sumi Helal)，兰卡斯特大学 (Lancaster University)
译 | 本刊编辑部

IEEE《计算科学评论》的年度展望专辑突出展示了几个“信息化为知识”的典范，例如，共享内存架构的回归（这次是由于在规模扩大上取得了成功），还有普适计算是怎样最终找到合适的操作系统的。我们也介绍了杂志编委会和内容的变化。

欢迎来到IEEE《计算科学评论》的年度展望专辑，它像往常一样带你浏览新兴以及快速发展中的计算技术。然而，在我介绍这个专辑的文章之前，我想先简单介绍一下编辑委员

会的变化。

编委会变动

经过数年的志愿服务后，几位编委会成员离任了。其中包括“计算对

话”专栏（2013-2017）的编辑查尔斯·塞弗伦斯（Charles Severance）；“科幻小说原型”专栏（2014-2016）和“未来今日”专栏（2017）的编辑布赖恩·戴维·约翰逊（Brian David Johnson）；“无异于魔术”专栏（2015-2017）的编辑潘蒂·欧拉斯维塔（Antti Oulasvirta）；“云封面”专栏（2013-2017）的编辑萨恩·穆鲁盖桑（San Murugesan）；“标准”专栏（2014-2017）的编辑查伦·瓦尔拉德（Charlene Walrad）；安全与隐私板块（2011-2017）的编辑罗尔夫·奥普利格（Rolf Oppliger）；和软件板块（2012-2017）的编辑勒妮·布赖斯（Renee Bryce）。我们衷心感谢他们对《计算科学评论》卓越服务和奉献，

祝他们一切顺利。

IEEE 标准协会主席唐·赖特 (Don Wright) 将加入我们，接替查伦·瓦尔拉德负责编辑重要的“标准”专栏。我们欢迎唐加入编委会。

泽利科·奥布雷诺维奇 (Zeljko Obrenovic) 是荷兰 Software Improvements Group 公司的首席顾问与软件评测主管，他将会加入《计算科学评论》担任网站编辑。他已经为杂志网站做出了贡献，给杂志的封面、文章和作者开发了一个易于搜索的索引系统。我们欢迎泽利科加入，期待与他一起改善杂志的网站。

新的专栏和栏目

《计算科学评论》将从 2018 年开始推出两个新专栏。“政策角”专栏将由 Synopsys 的高级应用顾问明娜·J·汉娜 (Mina J. Hanna) 负责编辑，她同时也是 IEEE-USA 的人工智能政策委员会主席及研发政策委员会副主席。这一新专栏将介绍与《计算科学评论》读者相关的最新重要技术政策信息。它将让人们更加清楚地认识到世界各地是怎样制订出这些政策的，还有研究人员和科学家如何影响或参与政策相关事务。

“网络 - 物理系统”是另一个每两月推出一篇文章的专栏，由帕特拉斯大学教授迪米特里奥斯·塞尔潘诺斯 (Dimitrios Serpanos) 负责编。该专栏将讨论应用于汽车和航空航天工业，医疗设备和系统，能源系统和智

能制造的网络 - 物理系统的关键安全问题。

《计算科学评论》还将推出一个新的漫画栏目，栏目名为“穿越时间的计算”，漫画由得克萨斯农工大学的可视化（和辅助计算科学）教授埃尔贡阿克莱曼 (Ergun Akleman) 创作。漫画栏目将与埃里克若伊霍尔德 (Erich Neuhold) 编辑的“50 & 25 年前”栏目放在同一版。

我们欢迎明娜、迪米特里奥和埃尔贡加入杂志编委会。

2018 展望

今年，我们会展望信息的未

进和变化，都欢迎提出来。

在本期的封面文章中，前两篇分别通过两个新兴科学（网络科学和信息科学）仔细探讨了信息相关的问题。在第一篇文章中，读者可以领略社交媒体信息的影响力，并与其他来源的信息进行比较。在第二篇文章中，我们可以了解一个激动人心的新科学领域，这一领域拓展了香农和图灵的信息理论和计算概念，让我们能对信息领域的复杂问题提出解决方案，并产生大量实际应用。

在第三篇文章里，我们采用了一种有趣和强大的方法，重新探讨了共享内存的架构问题。这种方法通过一种新的软件定义服务器架构，让我们

我们获取信息的能力并将它传播给世界各地成千上万亲密朋友的能力非常强大——这是一种明显可能遭到滥用的力量，而且难以控制。

来——它的价值与影响力，并去探讨，在充斥着各种噪音的环境中如何构建、解释信息，怎样才能让信息值得信任。从网络攻击到虚假新闻，从网络谣言和数据泄露，到病毒性传播的社交媒体错误信息，我们获取信息的能力并将它传播给世界各地成千上万亲密朋友的能力非常强大——这是一种明显可能遭到滥用的力量，而且难以控制。

我希望你喜欢今年的展望专辑，并期待得到你的建议和评论，无论你希望《计算科学评论》做出怎样的改

使用更简单的编程模型（不过仍然在部署时忽略了一些细节）。第四篇文章研究了网络、云、智能家居和智能汽车等操作系统在普适频谱上的进化。文章再次研究了软件定义概念，并提出一个“软件定义一切”的操作系统改进和设计的框架，让抽象和资源管理的原则继续适用，而且更加灵活。下面是这四篇文章内容的具体介绍。

在《从英国脱欧到特朗普：社交媒体在民主中的作用》一文中，温迪·哈勒 (Wendy Hall)、莱明·蒂纳

蒂（Ramine Tinati）和威尔·詹宁斯（Will Jennings）考察了政治运动对信息的重要利用，即信息在其中产生的影响力和可观测价值。本文使用取自 Twitter 的数据集来分析社交媒体在英国脱欧公投和 2016 年美国总统大选中扮演的角色，试图回答这样一个问题：与通过传统民调和由政治专家主导的主流媒体分析获得的可靠信息相比，为何通过社交媒体渠道流动的信息最终却被证实是预测选举结果的更好指标？虽然作者们承认，由于所获 Twitter 数据集及抽样方式的局限性，他们的研究尚非结论性的，但文章阐明了未来研究的重要领域。如今，社交媒体平台已被广泛运用于各种各样的政治活动，我们需要更深入地理解传输信息流的“管道和渠道”实际上如何影响了信息的价值、我们如何看待接收到的信息或信息如何影响我们，这样的探究越来越重要。

在《信息科学前沿：香农遇到图灵机》一文中，伊切赫·希潘诺夫斯基（Wojciech Szpankowski）和安纳斯格拉玛（Ananth Grama）向我们描绘了新兴“信息科学”的蓝图。与香农着重信号采集处理与通信的信息理论相对，这种信息科学扩大了分析范围，从只研究通信，到囊括处理、推论、聚合分析，并通常来说，还有特定领域的语义与分析。结构、空间、时间和语义这些新式信息，使得此科学围绕更复杂的信息难题得以提出方程式和理论。举一个简单的例子，按时到达的部分信息或许比延迟到达的完整

信息更有价值：究竟是不是呢？延迟，或者时间要素，并不是原有信息理论的关注点。这种时间概括也使我们得以运用嵌入在某些领域信息中的结构，如基因组序列中的对齐，进行强大的信息分析。而这种新式科学方法的核心，则是所谓的观察者功能的强大数学运算；这些数学运算提取并代表了数据库中相关的、可辨识的和可习得的信息。本文展示了此新式科学对其它科学领域和难题的价值，包括数据科学、信息安全和隐私。

在《再访可扩缩连续共享内存》一文中，C. 戈登·贝尔（C. Gordon Bell）和艾克·纳西（Ike Nassi）再次探讨一个著名的问题——扩大共享内存系统的连续性。如果没有严密复杂的程序设计模型，如果没有创造出难以管理的分布式计算机群，这个问题长久以来都无法解决。作者描述了一种创新性解决方法，通过由软件定义软件服务器架构模拟简单的单影共享内存系统，程序编写极大地简化。这种方法成功避免了内存访问冲突，同时亦实现了内存扩缩和简化编程。作者饶有兴致地描述了这种方法的历史，解释它如何诞生，何时诞生，发展过程中的停滞不前，当然还有近来被再次提及的原因。

在《朝着普适操作系统发展：一个软件定义的角度》一文中，梅宏和郭耀也运用软件定义的概念解释几年来操作系统的各种变革走向。从数据处理中心操作系统，到物联网的各组成要素，文中展示了基于网构软件的

操作系统结构，而其中的关键资源和管理员映射到云资源，操作系统调用则映射到网构软件应用程序界面。这个结构运用了应用程序框架概念，与安卓操作系统的概念相似，允许结构通过使用适合的框架、服务和运行时库专攻某一类型。文章简要介绍了三种由作者团队研发的软件定义操作系统原型，展示它们的框架。

我们希望本年度的展望栏目能吸引您，也希望您能提供宝贵的建议和意见。■

苏米·希拉勒是兰卡斯特大学（Lancaster University）数字健康教授和主任。他是 IEEE 会士，也是 IEEE 计算机协会的董事会成员。联系方式：s.helal@lancaster.ac.uk。

微信名：计算人
微信号：jisuanren



中国中文信息学会、语言与知识计算专业委员会 (CCKS)



CCKS·2018

音乐领域的命令理解

2018.05.01–2018.07.20

中科院自动化所 × 云知声



扫码了解详情



图片新闻

预测材料性质的快速算法

哈佛大学的研究人员年发明了一种新算法，可以根据材料晶体中的化学元素预测材料的性质（例如导电性和导热性）。和已有的算法相比，这种新算法简化了发现新材料所需的计



算量，把速度提高了大约 1 万倍。

来源：哈佛大学应用科学与工程学院

论文：*Accelerated Screening of Thermoelectric Materials by First-Principles Computations of Electron-Phonon Scattering*. Advanced Energy Materials, 2018; 1800246 DOI: 10.1002/aenm.201800246



从英国脱欧到特朗普： 社交媒体在民主中的作用

文 | 温迪·哈勒 (Wendy Hall)、莱明·蒂纳蒂 (Ramine Tinati)、威尔·詹宁斯 (Will Jennings),

南安普敦大学

译 | 郑米兰

如今人们可以通过社交网络在志同道合或不同圈子的民众之间分享、获取及连接事实和观点。这促使政客们开始普遍利用 Twitter 和 Facebook 等平台。然而，对于社交网络在民主进程中的影响力，我们的理解还非常有限。本文作者分析了社交媒体在 2016 年美国总统大选和英国脱欧公投中发挥的作用。

基于几项现有研究的框架，我们探究社交媒体在两个引人注目的政治事件中扮演的角色：英国脱欧公投，即英国全体公民投票决定是否继续留在欧盟；以及美国 2016 年大选，其中特朗普和希拉里分别代表共和党和民主党竞选总统。我们考察社交媒体的形式、结构、内容和情绪，以及它们与主流民调及预测的结果相比有何不同。我们的分析揭示出对政治事件的解释如何受到分析方法的影响，而这又突显了需要制定标准来支持针对这一研究领域的通用分析方法。

过去十年中，Twitter 和 Facebook 等社交媒体平台已成为在全球范围内广泛的人际网络中进行社交和政治交流的关键

工具。各种主题的信息在这些平台上以前所未有的规模交换，为用户提供了一个开放的、与特定领域无关的交流场所，同时带来了益处和挑战。

对个人而言，这些平台可被用来推进各种话题，只需简单的几项功能就可实现：关注、点赞、转发。对政客和政党而言，社交媒体可被广泛用于宣传公民投票、参与辩论，以及提供有关全国选举的信息。在政治学中，社交媒体分析已成为理解竞选过程中政治参与特性的关键。¹

有一些研究工作帮助我们提高了对政治极化、社交媒体机器人检测²以及升级竞选活动设计的理解（如<http://truthy.indiana.edu>）。社交媒体在政治话语中的作用不断扩大，促使

我们加深对它们如何影响与政治成功相关的复杂机制的理解³,包括它们如何预测与候选人及其支持者意识形态相关的选举结果。随着包括竞选活动和新闻报道在内的政治活动越来越多地转向在线媒体平台,研究人员正在努力识别Twitter和Facebook用户在人口结构和政治立场方面对广大选民的代表性。^{5,6}

奥巴马在2012年竞选连任获胜的关键原因之一是他在竞选活动中使用的15个社交媒体网站所产生的影响。由于人们越来越多地使用社交媒体来和整个社会及潜在的选民交流,对传统政治民调和新闻报道如何受这些活动影响的兴趣也在增加。在最近几次选举中,传统的投票方法没能准确地预测投票结果。相比之下,社交媒体活动似乎更好地反映了整体情绪和潜在选举结果。然而,尽管有这些最近的趋势,重要的一点是要认识到实际上只有很小比例的人口参与了社交媒体活动。

我们希望通过分析社交媒体数据来更好地理解公民的参与情况、观点及政治倾向⁷,获得对正在进行的讨论的宝贵洞见,了解信息如何被传播或压制,辨别政治事件和运动所运用的总体战略。我们对“点赞”建模,统计分析转发次数,加权混合性话题,7绘制社交网络结构,追踪所分享内容的类型、用户情绪,预测特定事件的结果。这些研究方

法有助于我们了解选民参与度、候选人及其平台的受欢迎程度,以及各种在线竞选活动的效果。⁸

我们分析了社交媒体在英国脱欧公投和2016年美国总统选举中的作用。我们的目标是了解在这些政治事件的前期拉票过程中人们是如何使用社交媒体的。我们也希望开发出一个关键的网络科学视角,通过它来审视传统的政治结果报告及预测方法,因为事实上这些结果与民意测验专家、政治分析家及统计学家的预测并不一致。

我们将在此介绍对这些政治事件的社会技术分析。我们开展这些分析是为提高对世界各地民主选举中社交媒体作用的一致性和多变性的理解。此外,我们还试图深入了解时间动态、网络结构变化和政治极化,确定哪些话题最突出,又有哪些公众情绪与这些话题相关。

相关研究

Twitter的受欢迎程度为其带来了前所未有的日常流量。候选人、当选官员和政府机构现在已把Twitter这样的社交媒体平台用作他们最常见的大众传播模式。因此,已有大量学术研究在探索社交媒体在现代政治运动和活动中扮演的角色。⁹

不同研究已经表明,在政治运动和辩论期间,信息的散播能在覆盖大型个

体网络方面发挥重大的作用。¹⁰因此,成为特定网络一部分的人将主要接收到被持有相同观点和相同人际网络的人所共享的内容。网络结构的动态特征也使得人们对共享内容产生了更深入的了解。¹¹各种研究还分析了社交媒体网络的时间性,这种特性有助于说明:政治运动如何导向了基于政治联盟的社区结构;人群如何转变立场;过程中出现了哪些模式和结构。¹²

研究还表明对话的情绪如何与影响政治运动的各种现实事件密切相关。¹³关于网络的时间因素也探讨了话题情绪,表明网络上的情绪可随时间变化,它与线下活动和事件高度关联(比如在全国电视上播放的公开辩论)。与此密切相关的是使用社交媒体预测选举结果,包括通过网络的结构特性、产生的内容类型,以及网络空间的总体情绪和极化来预测胜方。¹⁴然而,预测建模在预测人类行为方面的能力是有限的。通常,建模是基于社交媒体平台生成的数据和社交媒体用户在代表性上的偏差。

数据收集方法

我们在这两个重大政治事件中收集了两个Twitter数据集。我们通过南安普敦网络天文台(Southampton Web Observatory)这个平台来使用Twitter标准搜索API,¹⁵巡检速率为20分钟。

表1. 在政治结果分析中使用的Twitter数据集

数据集	英国脱欧公投 (2016年6月23日)	2016年美国总统选举 (2016年11月8日)
查询关键字	Brexit	USElection16
数据收集时段	2016年5月1日至7月31日	2016年10月1日至12月31日
推文数量	4,064,217	985,280
转发数量	2,197,300	659,906
提及	437,036	74,484
话题标签	149,758	43,865

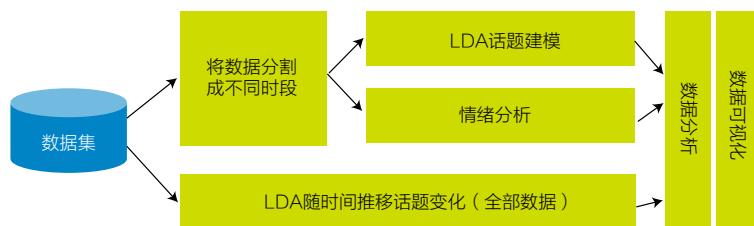


图1. 数据管道和分析一览。LDA 意为潜在狄氏配置 (Latent Dirichlet Allocation)。

对标准搜索API的使用提供了包含特定关键字的推文样本，我们之前已根据对特定主题的热门标签（例如#Brexit、#US Election16）的初始分析对这些关键字做了预定义。尽管该API仅为特定查询提供了推文样本，但我们在这些实验中验证的一个假设是这些数据集为Twitter上的对话提供了具代表性的样本。同样，尽管Twitter并不是社会中不同人口构成的代表性反映，但我们在这项研究中的假设是这些数据能让我们了解积极参与该平台的人群。

表1详细列出了两个数据集，包括用于查询Twitter搜索API的关键词和一组关于数据集的一般描述性统计数据。在这项分析中，我们把数据集内的推文限制在选举前一个月到选举后一个月的时间范围内。比如，英国脱欧公投于

2016年6月进行，我们的数据集包括从2016年5月初至7月底的推文。

如表1所示，两个数据集在数据分布和扩散方面具有相似性。

分析方法

我们使用了现有的分析方法来探索两次重大政治事件中社交媒体的使用情况。在判断一场政治运动的效果时，对话情绪分析可以产生令人印象深刻的结论。与此相辅相成的是，话题建模可以帮助阐明那些最引发热议的话题类型。我们还使用话题建模技术的一项衍生技术来发掘数据的时间特性，使我们能随时间推移对话题建模。这样，我们就有了一套方法来展示特定时间段内某个话题“流行度”的变化。本论文开

展的分析考虑了情绪分析和话题建模中的时间因素，以确定随时间推移人们在社交媒体上对某个政治事件的参与度的变化。令人兴奋的是，这粗略地描绘出社交媒体上大众态度和情绪如何随时间推移而变化（如果真有这种变化的话），以及这可能如何影响了政治事件的总体结果。

图1显示了使用的数据管道，以及如何使用不同的情绪和话题建模来生成有关社交媒体使用情况随时间变化的快照。鉴于我们考察的两个政治事件的二元性质，我们使用了一些预定义的关键词，根据政党（如共和党与民主党）或投票结果（赞成或反对）将数据划分为不同类别。

分析社交媒体在政治事件中的作用

为了理解社交媒体在政治事件中的作用，我们使用表1描述的Twitter数据集，并使用一组分析方法来帮助描述网络结构、内容，以及人与人的互动。我们的分析集中在三个主要领域：网络的时间演变、结构，以及其中的内容主题和情绪。

事件概述

英国脱欧公投，即英国关于是否退出欧盟的全民公决，是2016年欧洲一个

核心政治事件，也可以说是半个世纪以来英国最重要的政治事件。官方（和非官方）活动、政客及公民非常积极地参与到社交媒体平台上的讨论中，交流有关英国脱欧的信息和论点，吸引关注并影响选民。新闻记者和评论员们利用这些平台来报道这场政治运动中的事件。虽然传统的民调仍被用来预测选民意向（它们的预测大多给出了“留欧”的结果），但社交媒体为众多媒体机构提供了一个有关这场运动进展的信息来源，发挥了重要作用。

与英国脱欧这个一次性事件不同的是，2016年美国总统大选是一个历史悠久的政治周期的一部分。在几场竞选活动中，社交媒体一直是讨论的重要场所，也是候选人和新闻媒体参与的焦点。在大选之前，共和党和民主党的两位候选人特朗普和希拉里主导了这一讨论。由于Twitter等社交媒体平台的普及以及它们在政治报道中的使用，它们在全国电视台播出的辩论以及竞选双方的其他活动中有着显著的参与度。在关于民调的讨论中也提到了社交媒体：在全国新闻广播和在线新闻中报道了基于不明方法捕捉的社交媒体对话中的民众情绪。

时间动态

从政治运动中社交平台的互动分布可以了解人们的参与度。这种互动既

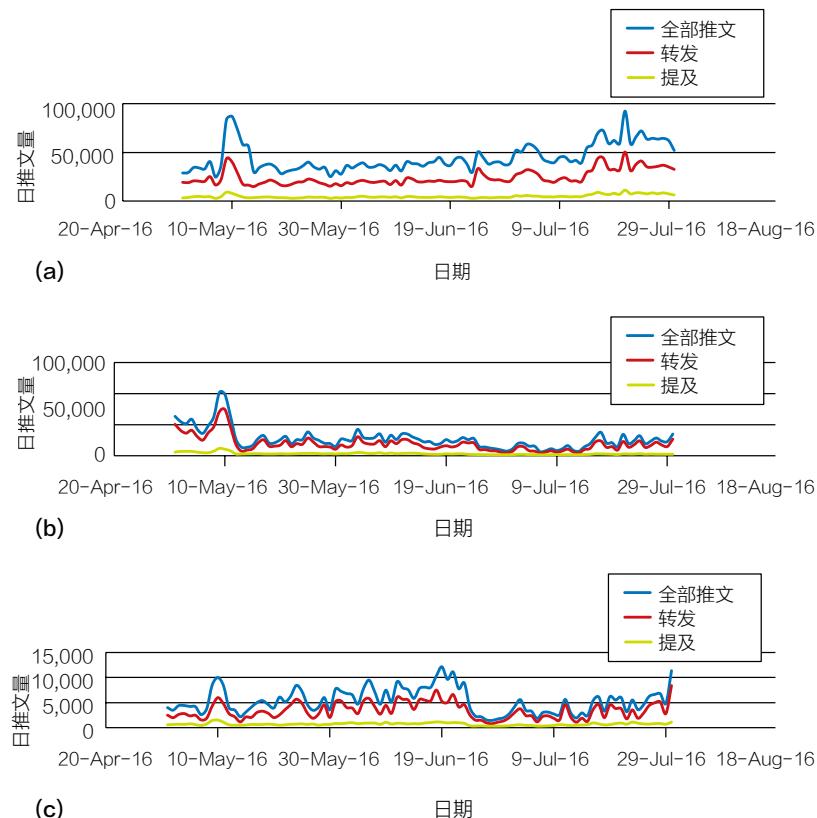


图2. (a) “英国脱欧”推文数量。(b) “留欧”推文数量(每日)。(c) “脱欧”推文数量(每日)。

由运动、党派和候选人官方帐户发起，也包括与普通用户的互动。从有关英国脱欧公投的欧盟推特数据看，这些互动（推文、转发和提及）均匀地分布在竞选前和竞选后时期，而在投票日（6月23日）当天和前后两三天里激增。与其他许多被研究过的政治运动相比，此次互动分布的时间跨度要大得多。如图2a所示（并在图2b和图2c中更详细地展示），尽管在公投前出现了大量讨论，但我们的分析也显示，公投后这种互动还在继续，这无疑是源于公投结果出来后持续展开的政治辩论——主要是围绕该结果对英国的影响。

相比之下，图3a所示的美国总统大选时间表显示，大选之前社交媒体

活动不断增加，并在选举日大增（超过200%）。图3b和3c更详细地说明了这一点，显示选举日前两天出现活动高峰。然而，与英国脱欧公投不同的是，美国总统大选后社交媒体活动减少的速度要快得多。这种差异可以归因于常规性总统选举与一次重大宪法公投相比更短暂的性质，也可能是由于社交媒体讨论和活动转移到了另一个主题标签或在线社区和空间里。

在英国和美国的数据集中，增加的活动量中有超过90%是在投票日当天转发的内容。在两个案例中，在围绕选举日的一周里都包含了全部转发量的70%以上，也包含了最长的转发链（即在特定时间窗口内重复共享单条推文

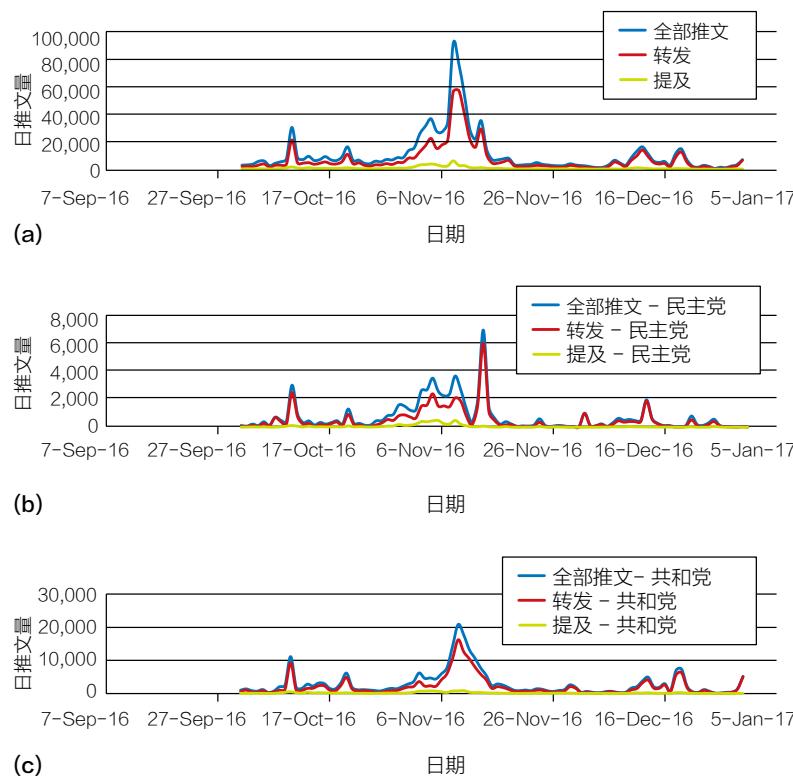


图3. (a) 2016年美国总统大选推文数量(每天)。(b) 民主党推文。(c) 共和党推文。

们可能与话题的多样性密切相关。这意味着政治讨论可能会被少数个人或组织主宰。

在结构上, 2016年美国总统大选的社交媒体活动与英国脱欧公投的网络类似。我们同样观察到网络中的几个主要参与者产生的内容分布不均。由于少数高度活跃参与者的存在, 有几个强连接的参与者集群(比英国脱欧公投中更多)生成了超过60%的网络内容(不包括转发)。对这些参与者进行人工检视后发现, 他们代表了两个对立的政党、全国新闻和媒体资源, 以及有强大政治背景的知名人士。图5a和图5b更深入地探讨了这一点, 显示不同政治立场的两方(民主党或共和党; 脱欧或留欧)发布一定推文数量的用户数对比。

在美国总统选举中, 当我们比较发布超过500条推文的用户百分比时, 发现在共和党中仍存在一定比例的这类用户, 而在民主党中完全没有。如果以网络中的“可见内容”来考察这一点, 这有令网络结构发生偏差的风险, 因为生成大量内容的账号和用户可能会吸引Twitter公众时间表上的可见性, 而盖过那些不够活跃的用户生成的内容。与此相反, 在英国脱欧公投数据集中, 以发布推文的数量来衡量, 倾向留欧和脱欧的用户数是平衡的, 在发布的推文数量从小于2到大于1,000的各个区间里, 两方的用户数量比例是相似的。

而不修改其内容)。因此, 推文和转发之间的比例表明, 随着选举日临近, 社交媒体网络上新发布的内容减少, 而信息的再循环变得越来越普遍。这是一个重要的研究点, 可用于了解线下活动影响在线活动特性的时间表和战略。

网络结构

网络结构指的是活动集群以及这些集群之间的总体连通性。它衡量了网络中参与者之间的互动, 以及参与活动的社区的多样性。图4展示了转发推文的网络结构, 其中的节点代表用户, 线条代表用户之间的内容转发。在政治事件中, 网络结构提供了一种方式来评估不同政党或候选人及其支持者之间的

讨论、网络内子社区的结构, 以及信息的传播方式。

整体网络结构与过去对社交网络政治参与度的研究得出的结果类似: 推文并非均匀地散播, 而是存在几个非常活跃的贡献者, 以及大量只发布几条推文的用户。网络中也有大量转发(这也增加了那些高活跃度贡献者的推文数量), 这体现了不同的、强连通的活动中心之间的互动。这对于社交媒体用户而言意味着什么呢? 它意味着网络中的一般观点可能被忽略, 因为算法偏爱流行的账号(我们将“流行”定义为拥有的朋友/追随者和发布的推文数量都很多)而将他们的帖子推到最前方。此外还有许多重复同样信息的转发长链, 它

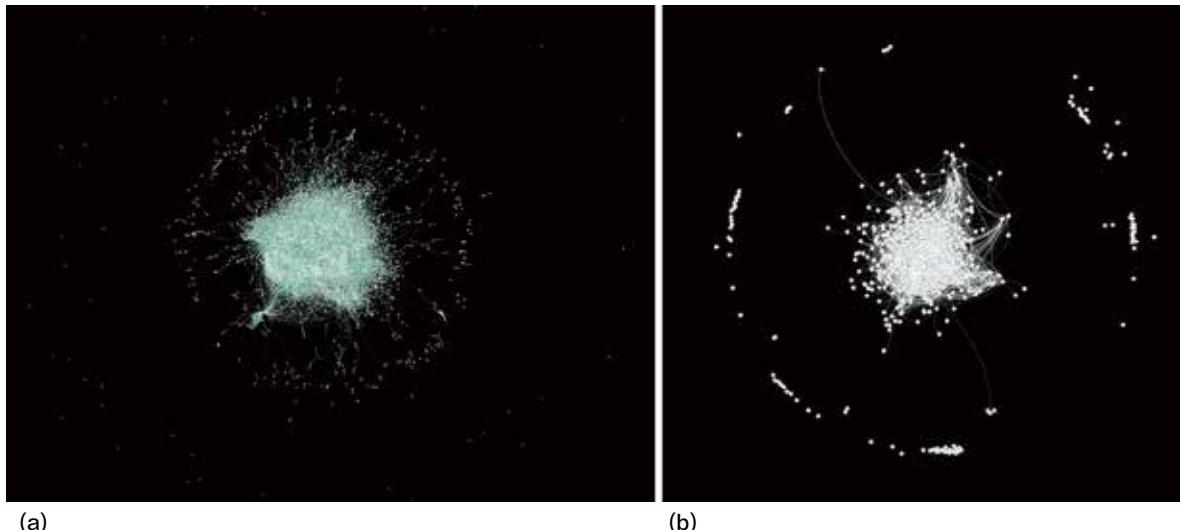


图4. (a)2016年美国总统选举推文转发网络,显示存在单个巨大的核心构成。(b)英国脱欧公投推文转发网络,显示存在一个大型核心构成和多个较小的互不连通的社区。

分析表明,虽然网络中存在许多强大的社区,但低网络模块度显示这些社区之间也在相互作用。也就是说,尽管存在着各自代表不同政党或候选人的强关联网络,但在不同的政治观点(或政治候选人)之间也发生了讨论(比如推文和转发)。

当观测那些令网络模块度下降的参与者(即与几个集群同时交互的参与者)时,我们发现这些人发布的原创内容和转发的内容同样多。就这些人的影响力而言,他们有能力将内容推向不同的社区,而这些社区原本可能已与外界断开联系或者不知道还有其他讨论正在发生。但是,同样也有可能的是,存在机器人或某些参与者负责汇总内容或分享误导性信息。

极化

尽管两极分化可被视为网络结构和网络内容导致的结果,而非明确的度量标准,但在一些研究中,它已被证明

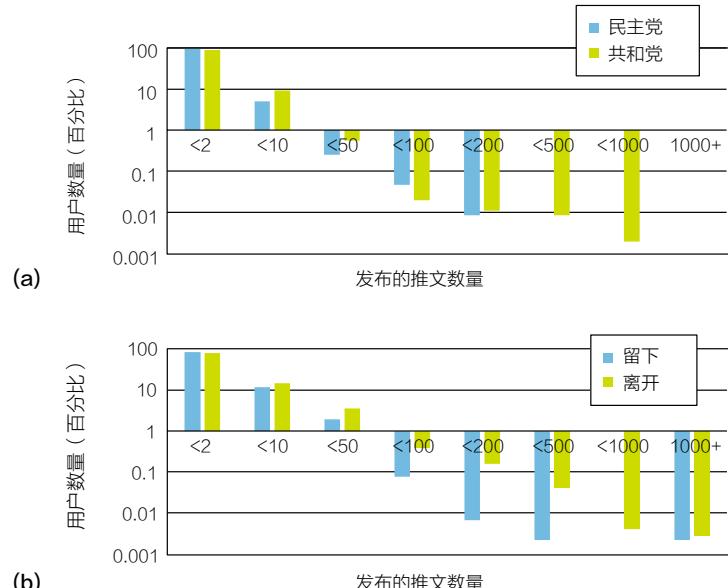


图5. (a)美国总统选举推文分布:民主党 vs. 共和党。(b)英国脱欧公投推文分布:留欧 vs. 脱欧。

是政治运动中特别有用的一项指标。¹⁷一种计算网络中两极分化的常用方法是审视网络结构,然后根据特定的政治隶属关系标记网络中的参与者。测量结果可被用于判断特定标签或类别之间的平衡关系。

在脱欧公投中,我们使用简单的方法将数据划分为“脱欧”和“留欧”两个阵营,我们观察到天平明显向“脱欧”倾斜。超过65%的推文包含了字符串“离开”。随着时间推移,包含“离开”的推文比例不断增加(在投票期间,包含“离

表2. 2016年美国总统选举话题关键词

类别	10月	10/11月	11月	11/12月	12月
民主党人	克林顿 希拉里 投票 总统 历史	克林顿 希拉里 联邦调查局 电子邮件 主导	克林顿 希拉里 选举 早期 删除	克林顿 希拉里 总统候选人 受欢迎	克林顿 共和党 热爱 游戏 选举
共和党人	特朗普 唐纳德 选举 投票 总统	特朗普 唐纳德 投票 总统 受欢迎	特朗普 唐纳德 选举 总统 赢	特朗普 唐纳德 性 重大 面临	特朗普 唐纳德 选举 世界 时间
混合	特朗普 唐纳德 基本上 碾压 事实核查员	特朗普 克林顿 基金会 极大地 唐纳德	克林顿 特朗普 希拉里 唐纳德 选票	克林顿 特朗普 希拉里 唐纳德 民调	克林顿 特朗普 唐纳德 希拉里 俄国
没有关键词	选举 投票 选举的 事实 俄国	选举 新闻 阅读 俄国 结果	选举 投票 糟糕 主要 美国	选举 俄国 已投的 消息 人	选举 时间 人们 民调 para

开”的推文数量发生了变化）。如果在竞选活动期间我们就知道了这一点，那么我们可能已经得出结论：“脱欧”阵营已成功将其理念传达给了民众。

当我们采用相同的分类法来标记与两位美国总统候选人相关的内容，该网络呈现出对共和党候选人特朗普的明显倾向性。这与特朗普在初选和大选期间比希拉里更多地主导了新闻报道的情况是一致的。在我们收集的数据集中，超过80%的推文提到特朗普或共和党。我们还发现，越接近选举日，候选人之间的两极分化越明显。在选举日那一周以前，65%的推文提到共和党或特朗普。接近选举日时，这一比例发生了巨大的变化，占原创和转发内容的85%以上。

在不查看推文的内容、语言及表述目标的情况下，不可能知道它们到底是赞成还是反对某位候选人或某个政党。比如，在特朗普这个案例中，当10月7日他过去在《走进好莱坞》(Access Hollywood)节目中受访的视频被公布后，很有可能有相当大比例的推文都对他持负面态度。同样地，10月28日联邦调查局宣布重启对希拉里担任国务卿

期间使用私人服务器处理机密邮件的调查后，局面对希拉里也一样。尽管如此，我们的方法大致概括了辩论发生期间的关注倾向，无论它们对讨论对象的态度是正面或负面。

话题和情绪

为理解网络语境，主题建模提供了一种方法来学习带有大量文本的常见话题，而无需使用任何关键字或主题的初始列表来引导它。虽然这需要大量调整，但已被证明在Twitter和其他类似的社交网络上表现出色。令政治学研究更感兴趣的可能是话题的时间多样性，它衡量了随时间推移而讨论的话题数量。

如果使用得当，网络情绪是另一个有用的指标。可以对一个数据集内的所有内容进行情绪衡量（例如，同时对所有消息流运用情绪指标），或者在根据主题建模的结果对文档分组后再应用该指标。适用于政治事件的另一种方法是根据极化过程中建立起的类别来应用该指标，这让研究人员能在类别层级上检测到更细微的情绪。表2罗列了2016年美国总统大选数据集包含的那

三个月里被识别出的话题关键字。我们根据以下几种情况来做主题建模：包含与共和党相关的关键词的推文；包含与民主党相关的关键词的推文；同时包含与两个政党相关的关键词的推文；根本不带有上述关键字的推文。这些话题关键词体现了主流媒体中讨论的争议性话题，比如希拉里的电子邮件丑闻和俄罗斯黑客丑闻。

表3罗列了英国脱欧公投的主题关键词。话题多样性（可以在一段时间里衡量）表明，在“脱欧”类别里的话题最初是相当多样化的，内容所带有的情绪往往都是负面的。然而，越接近投票日，话题多样性越减退（也就是说，讨论越来越集中在一个类似的主题上），而社交媒体活动的情绪变得不那么消极了。与此相反，“留欧”类别里的推文主题仍然多样化，情绪也没有改变（保持负面）。这些趋势和一种流行的批评观点一致：“留欧”阵营的宣传活动缺乏一条广泛流传的积极信息，而“脱欧”阵营在投票日前的最后几周里把内容聚焦于“收回控制权”，产生了很好的效果。

同样，我们发现在2016年美国总

表3. 英国脱欧公投话题关键词

分类	主题1	主题2	主题3	主题4	主题5
离开	离开 英国脱欧 投票 英国 欧洲	离开 英国脱欧 运动 民调 投票	离开 投票 英国脱欧 英国 运动	离开 英国脱欧 运动 决定	离开 英国脱欧 投票 运动 英国
留下	留下 英国脱欧 投票 阵营	留下 英国脱欧 运动 投票 企业	留下 英国脱欧 运动 公投 投票	留下 英国脱欧 民调 运动 往前	留下 英国脱欧 投票 团队 运动
混合	离开 留下 民调 往前 放	离开 留下 英国脱欧 投票 选民	离开 留下 英国脱欧 运动 民调	离开 留下 欧洲的 投票 英国脱欧	留下 离开 英国脱欧 辩论 伦敦
无	英国脱欧 投票 人们 英国 移民	英国脱欧 特朗普 贸易 唐纳德	英国脱欧 卡梅隆 大卫 英国 投票	英国脱欧 投票 经济 银行 英国	英国脱欧 del los por para

统选举中，虽然社交媒体上对共和党的讨论情绪起初是负面的，但随着选举日的临近，话题的多样性减少了，也就是说，用户在讨论共和党时越来越多地说着同样的一两件事，而讨论的情绪变得不那么消极了。这与共和党在竞选活动临近尾声时受欢迎程度提高的趋势是一致的，也反映在对国会中期选举的民意调查中。相比之下，围绕民主党的话题情绪保持稳定（趋向负面），话题多样性没有减少。

讨论

在我们的分析和实验中，我们展示了一系列技术，可将它们用于判断在政治活动或事件中社交媒体是如何被使用的，以及生成的对话的特性。通过运用一系列方法，我们有可能发展出对互动结构的高级概述，并深入了解网络内容，了解正在被讨论和分享的内容，以及特定社区的总体情绪。

在解读这些结果时，我们必须考虑到一点：社交媒体上的讨论只是政治运动总体讨论中的一小部分，在整个政

治生态系统中也只占一小部分。然而，正如最近的结果所显示的那样，传统的投票和政治预测似乎并不能正确预测投票结果，而社交媒体平台的分析却越来越显示出它们对投票结果的影响。在英国脱欧公投和美国总统大选中，实际结果公布后，主流的预期结果被完全推翻。而从我们的分析来看，社交媒体上的对话却强烈暗示了实际结果。

这只是巧合吗？或者，在了解普通大众的情绪时，社交媒体对话中的消极情绪的强度是一个更好的指标，程度超出了我们的理性预期？对于这些问题，仅仅是开始探讨它们也需要多得多的实验分析。

众多研究社交媒体的论文的一个主要局限都在于数据访问和数据质量，而非方法应用。例如，Twitter仅提供了讨论的样本，只占其全部“流水”的1%而已，对于生成这些样本所使用的方法也未作披露。因此，无法确定这些样本对于全部内容而言是否真正具有代表性。另一个限制是数据收集方法——它们都基于关键字或提问搜索。而这是有问题的，因为它假设与特定主题相关的所有内容

都至少会带有某个特定的关键字。

社交媒体在政治运动中的使用还处于初级阶段，对其结果数据集的分析就更是如此。尚没有用来运行实验的常用方法、标准或基准。另外，研究人员在实验中使用的数据集都不尽相同。不同团队可能使用同样的社交网络（如Twitter）作为其数据来源，但使用的样本却各不相同，实验设计和分析算法也不一样。我们不能拿本论文描述的成果来声称任何普遍结果。我们只是介绍了我们对从Twitter获得的英国脱欧公投和2016年美国总统选举的特定数据集所做的分析。

我们的分析在我们收集到的数据中广泛地考察了时间动态、网络结构、极化、主题和情绪。在未来的研究中，我们计划更深入地分析数据以考察某些影响因素，比如极端负面情绪的影响力：《走进好莱坞》视频的发布对特朗普竞选的影响；联邦调查局重启希拉里邮件门调查对她竞选的影响。我们也在把本论文中开发的算法应用于其他政治运动。

正如本文已清楚表明的那样，要分

析社交媒体在民主和政治竞选中的作用，有许多不同的方法。然而，各种研究团队正在积极而孤立地工作着，对以不明方式收集的数据集使用不同的算法和方法。如果要在这个领域里创建一个研究机构，我们要能够重复和重现实验结果，并重复使用相同的数据集来运行其他实验。这突显我们需要为了研究目的而实现共享的数据库。网络天文台就是这样一个数据库，它支持了网络科学研究。但在许多其他学科中也有同样的需求。¹⁸我们收集的用于分析的数据集可以在南安普敦网络天文台上找到。^{19,20}

最后，我们要说一说做这类研究的风险。社会学家曼纽尔·卡斯特尔尔（Manuel Castells）有一句名言：“权力不在机构中，甚至不在国家和大公司中。它存在于构建了社会的网络中。”

因此，控制网络的人就拥有权力。而未来，研究人员和商业企业正在开发的这些分析社交媒体的工具可被用来决定民主选举的结果。也就是说，采用最佳算法的竞选活动会获胜。我们已经在最近的选举中看到了这样的迹象。

对于在选举活动中运用社交媒体数据分析，我们的社会很可能需要划定新的伦理界限。否则，人工智能可能会决定我们的下一位领导人是谁。■

参考文献

1. J.E. Carlisle and R.C. Patton, "Is

- Social Media Changing How We Understand Political Engagement? An Analysis of Facebook and the 2008 Presidential Election," *Political Research Quarterly*, vol. 66, no. 4, 2013, pp. 883–895.
2. E. Ferrara, et al., "The Rise of Social Bots," *Comm. ACM*, vol. 59, no. 7, 2016, pp. 96–104.
3. X. Yang, et al., "Social Politics: Agenda Setting and Political Communication on Social Media," *Int'l Conf. Social Informatics*, 2016, pp. 330–344.
4. R. Bond and S. Messing, "Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook," *Am. Political Science Review*, vol. 109, no. 1, 2015, pp. 62–78.
5. J. Mellon and C. Prosser, "Twitter and Facebook Are Not Representative of the General Population: Political Attitudes and Demographics of British Social Media Users," *Research & Politics*, vol. 4, no. 3, 2017; doi: 10.1177/2053168017720008.
6. S. Flaxman, S. Goel, and J.M. Rao, "Filter Bubbles, Echo Chambers, and Online News Consumption," *Public Opinion Quarterly*, vol. 80, no. S1, 2016, pp. 298–320.
7. Y. Wang, et al., "Catching Fire via 'Likes' Inferring Topic Preferences of Trump Followers on Twitter," *Proc. 2016 Int'l AAAI Conf. Web and Social Media (ICWSM 16)*, pp. 719–722.
8. P.T. Metaxas and E. Mustafaraj, "Social Media and the Elections," *Science*, vol. 338.6106, 2012, pp. 472–473.
9. A. Jungherr, "Twitter in Politics: A Comprehensive Literature Review," *SSRN*, 27 Feb. 2014; <https://dx.doi.org/10.2139/ssrn.2402443>.
10. Y. Mejova, P. Srinivasan, and B. Boynton, "GOP Primary Season on Twitter: Popular Political Sentiment in Social Media," *Proc. 6th ACM Int'l Conf. Web Search Data Mining (ICWSM 13)*, 2013.
11. P. Debjyoti et al., Compass: Spatio Temporal Sentiment Analysis of US Election What Twitter Says!, *Proc. 23rd ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD 17)*, 2017, pp. 1585–1594; doi.org/10.1145/3097983.3098053.

关于作者

12. S.A. Myers et al., "Information Network or Social Network?: The Structure of the Twitter Follow Graph," Proc. 23rd ACM Int'l Conf. World Wide Web (WWW 14), 2014, pp. 493–498.
13. C.S. Park, "Does Twitter Motivate Involvement in Politics? Tweeting, Opinion Leadership, and Political Engagement," Computers in Human Behavior, vol. 29, no. 4, 2013, pp. 1641–1648.
14. A. Tumasjan et al., "Predicting Elections with Twitter: What Characteristics Reveal About Political Sentiment," Proc. Int'l Conf. Weblogs Social Media (ICWSM 10), 2010, pp. 178–185.
15. R. Tinati et al., "A Streaming Real-Time Web Observatory Architecture for Monitoring the Health of Social Machines," Proc. 24th ACM Int'l Conf. World Wide Web (WWW 15), 2015, pp. 1149–1154.
16. A.O. Larsson and H. Moe, "Studying Political Microblogging: Twitter Users in the 2010 Swedish Election Campaign," New Media & Society, vol. 14, no. 5, 2012, pp. 729–747.

温迪·哈勒 (Wendy Hall) 是南安普敦大学计算机科学系教授,也是网络科学院 (Web Science Institute) 的执行总监。她和蒂姆·伯纳斯-李 (Tim Berners-Lee) 爵士及奈杰尔·谢德波特 (Nigel Shadbolt) 爵士于2006年共同创办了网络科学研究计划,并且是网络科学协会 (Web Science Trust) 的常务董事。她在2009年英国新年荣誉榜中成为大英帝国爵位二等勋爵士,并且是皇家学会会员。她曾担任ACM主席、皇家工程科学院高级副总裁、英国首相科学技术委员会成员。她还是英国政府AI Review的联合主席。可通过wh@ecs.soton.ac.uk与她联系。

莱明·蒂纳蒂 (Ramine Tinati) 是新加坡微软“全球黑带团队”(Global Black Belt Team)的资深数据科学家。在参与本文所报告的研究时,他是南安普敦大学的研究员。他的研究兴趣包括实时大数据流处理和查询。他是南安普敦大学网络和互联网科学小组的新前沿研究员和高级研究员,在由英国工程和自然科学研究委员会 (EPSRC) 资助的项目SOCIAM中工作,开发方法和分析工具来了解网络的发展和连接性。他在南安普敦大学获得网络科学博士和硕士学位。可通过r.tinati @ soton.ac.uk与他联系。

威尔·詹宁斯 (Will Jennings) 是南安普敦大学政治学和公共政策教授。他的研究探讨有关公共政策和政治行为的议题,特别是与议程设置、舆论、选举、民主创新、政治地理、政策灾难和反政治有关的议题。他是由英国投票委员会和市场研究协会发起的独立调查的成员之一,负责调查2015年英国大选前民调的效果。他在牛津大学获得政治与国际关系博士学位。可通过W.J.Jennings@soton.ac.uk与他联系。

17. M. Conover et al. "Political Polarization on Twitter." Proc. Int'l Conf. Weblogs Social Media (ICWSM 11), 2011, pp. 89–96.
18. R. Tinati et al., "Building a Real-Time Web Observatory," IEEE Internet Computing, vol. 19, no. 6, 2015, pp. 36–45.
19. "2016 UK BREXIT Tweet Dataset," Web Observatory; webobservatory.soton.ac.uk/datasets/F8NC2AxztDeL3csPw.
20. "2016 US Presidential Elections Dataset," Web Observatory; webobservatory.soton.ac.uk/datasets/hNbdddJesSpnJM4Y5.



信息科技前沿： 香农遇上图灵

文 | 沃伊切赫·希潘诺夫斯基 (Wojciech Szpankowski)、安纳斯·格拉玛 (Ananth Grama),
普杜大学信息科技中心
译 | 任筱仪、葛杨贊，浙江大学

作者统一了香农和图灵假设的要素，提出一个信息科技框架，可用于数据分析、数据隐私和数据安全，提供了数据科学亟需的基础理论。

信 息和计算是现代计算机的两大基本概念。克劳德·香农奠定了信息理论基础，演示了通信和压缩如何被准确模拟、公式化及分析。而阿兰·图灵形式化描述了计算概念，即通过算法进行的信息转换。

我们相信将香农和图灵假设的要素结合起来的信息科学可以解决在通过高效计算将数据向信息再向知识转变过程中遇到的重大瓶颈。而且要发展高效计算方法首先需要信息化。在该文章中，我们着眼于数据到信息到知识转变过程中的基本问题，解决与学习极限、取样适切性、数据增值、学习方法最佳性等相关的问题。为实现上述目标，我们将利用信息理论、统计学、算法和分析、组合数学、数值方法和学习理论等。

目前在机器学习、信息理论、信息隐私和信息安全及数据挖掘和分析等广大领域已有了显著进展。技术方法的提升应用于日常生活，产生了巨大变化，如无人驾驶汽车和智能温控装置。同时，这些变化又推动了技术攻关：从学习到数据的瓶颈是什么？系统什么时候会采样不足？学习模型什么时候过度拟合？附加数据的增益有多少？如何验证学习模型？如何实现可证明为最佳的学习方法？

为了解答这些问题，我们需要形式化建模和分析。即便这些问题解决了，还会有更多问题，比如数据隐私，尤其是涉及到数据分析和查询时。在精确性（查询和分析中）和公开性之间必须要有明确权衡，如何权衡需要通过分析量化研究。信息科学提供了一个统一框架，在该框架下这些问题都

可以公式化，用统一理论进行研究。

背景

信息科学凭借其与数据科学、生命科学、通信工程和经济学等诸多学科的相关性及其在诸多领域的应用，成为研究的重要新兴领域（更多信息请看侧边栏）。数据分析重要应用的兴趣域可以用一个系统或过程建模。“数据”的定义是域的采样值。“信息”这个术语几个世纪来在许多文献中被广泛讨论，我们将其定义为区分域的要素。该非正式定义表明对于兴趣域稀疏的采样使我们无法将其与其它域进行区分，因此它几乎不包含信息。相反，对相同域进行过采样不会增加信息，因为区分域的能力没有增强。我们用区分的概念和域空间的相关区分对与数据相关的信息进行特性化。最终模型、相关分析和方法及其在不同应用中的特殊化与信息科学相一致。

我们将“知识”（或“含义”/“观点”）定义为信息和域相关语义的结合。域相关语义的概念可将信息置于上下文中研究。我们将提取数据过程定义为信息通过学习转变为知识的过程。当该过程由机器人执行时，称为“机器学习”。“数据科学”指将原始数据转化为知识的所有途径，包括预处理、清理、分析和解释。统计学家也

信息科学中心

2010 年，美国国家科学基金会建立了信息科学中心 (www.soihub.org)。作为一个科学技术中心，信息科学中心寻求发展支撑

各方面信息的基础理论，并推广其在科学、工程、社会和经济等众多领域的应用。其任务是通过对生物学、物理学、社会学和工程学的信息进行表征、通信及处理等量化研究的新范式发展科学技术。其致力于以信息理论、计算机科学、统计学、数学和物理学作为基本工具，探究延伸信息科学的新领域。

许会认为这种“知识”是假设、结构函数、模型或集合。该文章中最小描述长度 (minimum description length, MDL) 和贝叶斯信息准则 (Bayesian information criterion, BIC) 经常用于选择公式化最优过程的数据模型¹。计算机科学家经常用柯尔莫戈洛夫复杂性 (描述复杂性) 检验这些抽象概念，这是可输出观测数据的最短程序¹。计算学习理论界主张其它方法，如精准学习²、PAC 学习 (probably approximately correct learning)³、VC 理论⁴ 和贝叶斯推理。

香农的研究重点是压缩 (源编码) 和通信 (信道编码) 的数据恢复，信息科学借助他的理论提出了做巨量分析的框架，因为信息不只是通信，还有采样、表征、推测、处理、合计、管理、估值、加密和计算。在这一层面上香农的逆定理尤其具有意义。根据香农对“信息”的经典定义，逆定理是如果信息存储或通信超出了基本限制，随着误差概率的消失没有算法能够实现目标任务 (传输或存储恢复)。

该定理的另一种形式：如果能设计出一种算法实现目标任务，那么基本限制一定满足。

但是，逆定理没有给出实现目标任务的计算复杂度。当计算复杂度超过存储和通信限制，就不得不将其纳入重要考量。该文章中我们提出了信息高效的公式，能同时满足信息的基本限制以及较低的计算复杂度。

众所周知，脱机最优 LZ 压缩算法是一个 NP 难题。但是，数百万使用者每天都需要 LZ 压缩算法，用典型数据序列的线性时间算法实现渐进式最优解。比如 DNA 组装问题就是个 NP 问题 (通过汉密尔顿路径)。但是，在 DNA 组装问题上已经找到满足信息限制的有效方法⁵。更普遍的，因为物理数据不可避免地会损坏或不完整，在数据科学领域一味关注最优解的计算复杂度不一定总是适用。在这种情况下，我们提出信息效率高的更简便算法，可满足信息限制。但这不能说明我们总能找到满足限制的有效办法 (比如 DNA 组装问题⁵ 就不行，还有随机

块模型的群落恢复问题）。因此，必须先对信息限制进行分析，再努力用算法将数据转变为信息。这就是我们的“香农遇上图灵”。

信息科学将框架、空间、时间、联通和语义学等元素纳入香农的信息论中。比如，分析基因组（蛋白组）序列的一个重要工具就是定位。基础假设是被定位子序列的结构和功能信息可以推测。在分子生物中，通过结构基元我们可以知道分子功能。社交网络中，重复的互动模式被认为是信息流的规范机制。尽管不同领域都已找到提取信息的方法，信息科学在复杂的相互作用系统中仍要解答一些基本问题，比如：如何量化、表述和提取多个系统共同使用的抽象信息？信息如何被创造出来？又如何转移？用无数抽象概念表述的信息的价值是什么？从数据库提取信息的基本界限是什么？

信息科学

信息科技的飞速发展和信息系统与服务的广泛应用极大地掩盖了一个事实：虽然为了了解信息已经投入了无数人无数的精力，但信息仍没有从根本上被定义。1953 年，香农写道：“‘信息’这个词已经有了很多意思，至少许多意思在各自领域中都非常实

用，值得深究，并得到认可”⁶。早在 1948 年，他就写道：“交流的基本问题是，在一处产生或准确或大致的信息在另一处被筛选，渐渐地信息有了含义：通过物理或抽象实体和统一系统表达相关含义。这些语义方面的交流与工程问题无关”⁷。

我们采取的对信息的定义可用于区分每一组数据样本。区分的过程用到了“观测函数”，将一组数据标定为观测对象。传统信息理论中的观测函数类似于信道。在更一般设定中观测函数可能是随机函数，比如给出学习算子优化过程的公式。比如，一组给定的数据可以通过分类器（比如逻辑回归函数）观测，在该情况下，两组数据 d_i 和 d_j 如果用同一种学习方法的分类器将会产生相同的信息（见例 2）。注意目标函数值，表明我们在观测函数中使用了极好的办法（比如逻辑回归适用于整个测试）。

更严格地证明， S 表示包含代表每个输入数据组的所有元素的空间，在 S 中定义一个观测函数，正如我们定义信息那样。目标是用 ball B_i 的最小值几乎完全覆盖空间，一个 ball B_i 中的所有元素都可用观测函数 f 区分。这里“几乎”相当于基础概率总空间的 $(1 - \epsilon)$ 。通过观测函数覆盖了 S 中 $(1 - \epsilon)$ 的 ball B_i 的数量等于 $N_\epsilon(f)$ 。然后我们将可用 / 可学习信息定义为 $I(f, S) = \log$

$N_\epsilon(f)$ as $\epsilon \rightarrow 0$

例 1：香农信道容量

我们先举一个例子，来说明信息的一般定义如何体现在稳定传输的传统框架中。

比如一个简单的点对点通信信道，正如香农的定义⁷。我们要通过信道传递一行长度为 n 的二进制字符串，目标是随着 n 增大，保证接受信息时出错概率 P_e （符合观测函数）趋向于 0。对于长度 2^n 的字符串，我们不能保证传送过程中出错率为 0，但是能否保证里面的一部分是对的，也就是 2^{nR} ， $R < 1$ 并且十分接近 1？换句话说，我们能否用 P_e 将“几乎所有”信息区分出 2^{nR} 的子集？对于任意 $R < C$ 都能保证 2^{nR} 正确传输，这个 C 就是香农容量，公式是 $I_n(P_e) = nC$ 。

例 2：学习资源

该例中，我们将演示一般学习中如何将我们的信息公式区分资源。在这个例子中，我们将展示标准化可区分来源的信息在通用学习中的应用。

我们将聚焦于通过一类参数分布源 $M(\Theta)$ 产生的序列 $x^n = x_1 \dots x_n$ 中的可计算可习得信息。 $\{P_\theta : \theta \in \Theta\}$ ，令 $\hat{\theta}(x^n)$ 为最大似然法的评估值，即 $\hat{\theta}(x^n) = \arg \max_{\theta \in \Theta} P_\theta(x^n)$ 。

观察到对于给定的序列 x^n 产生的

模型 θ 和 θ' ，假设这两个参数在某些距离测度下隔得足够远，我们都可以用 $\hat{\theta}(x^n)$ 去选择误差概率最小的模型。如果 θ 和 θ' 这两个模型离得太近，那么他们几乎无法区分且无法带来其它任何有用的信息。因此，序列 x^n 的可习得信息数量为它产生的可区分分布（模型）的总和。

举个例子，当类源 M 是由一族 $(m-1)$ 个未知参数产生的无记忆分布 $I_n(\Theta)$ 。如前所述，我们将使用 MDL 评估法。¹

对于两个分布 / 模型间的距离测度我们将使用 Kullback–Leibler (KL) 散度 $D(\bullet||\bullet)$ 。令 $B_{KL}(\theta_0, \varepsilon) = \{ \theta : D(\theta || \theta_0) \leq \varepsilon \}$ 为 KL 球中围绕 θ_0 的半径 ε 。

模型们的可区分度取决于误差概率，这部分可以用如下方法评估

$$\theta \in \Theta :$$

$$P_\theta(\hat{\theta} \neq \theta) = P_\theta(\theta(X) \notin B_{KL}(\theta, \varepsilon / n)) \sim 1 - O(\varepsilon^{m/2})$$

当 $\varepsilon > 0$ 时，令误差概率为 $O(1/\sqrt{n})$ ，在之前的工作中我们观察到可区分的分布数 $I_n(\Theta)$ 为：

$$I_n(\Theta) = \sum_{k_1 + \dots + k_m = n} \binom{n}{k_1, \dots, k_m} \left(\frac{k_1}{n} \right)^{k_1} \dots \left(\frac{k_m}{n} \right)^{k_m}.$$

上述和是 m 基于 $k^k / k!$ 的一种卷积，因此我们可以用分析组合学的方法找到它的渐近性。通过梅兰 (Mellin) 变换和兰伯特 (LambertW) 函数的奇异值分析，我们发现：

$$I_n(\Theta) = \log C_n(\Theta) = \frac{m-1}{2} \log \left(\frac{n}{2} \right) + \log \left(\frac{\sqrt{\pi}}{\Gamma \left(\frac{m}{2} \right)} \right) + \frac{\Gamma \left(\frac{m}{2} \right)^m}{3 \Gamma \left(\frac{m-1}{2} \right)} \cdot \frac{\sqrt{2}}{\sqrt{n}} + O(1/n),$$

其中 Γ 是欧拉伽马函数，这是可以从一类无记忆源中提取出的可习得信息数。它同时等同于通用源代码的极小化极大冗余度。¹

信息影响因子

现在我们强调一下与观测函数有关的公共因子，因其极大提高了可区分性，从而提升函数提取信息能力。信息的这些方面现已包含在不断扩大的信息科学中。

框架和组织。数据经常含有明确或潜在的与其相关的框架和组织。我们缺少方法定义和量化体现在框架组织内的信息⁹。小弗雷德里克·布鲁克斯 (Frederick P. Brooks Jr.) 写道：¹⁰

“香农和韦弗对信息的定义和测度已经广泛应用，这是他们留给后人的无价之宝。但我们还没有理论能对框架内信息进行测度，

这是信息和计算机科学理论基础的最根本缺陷。愿意把大把时间花在最根本问题上的年轻学者其实不需要这样研究。”

数据框架和组织的具体应用包括纳米编码信息、生物分子、基因检测及蛋白质相互作用网络和社交网络。理论上，这些方法必须说明相关环境，各种观测值（比如同一系统可以测量其相关性、几何性及属性）和时变状态。

滞后。在典型的交互系统中，信号的及时性对其功能发挥必不可少。一般情况下不完整信息的及时传输优于完整信息的滞后传输。但及时性与系统状态（是否高强度运行）和接收者有很大关系。香农信息论对传输或接收的复杂度或滞后没有限制。

空间。在交互系统中，空间定位经常限制数据交换，这是明显的弊端，但也有好处，比如减少干扰，增强系统调节与应变能力。

信息和控制。除了延迟/带宽权衡，系统经常修改根本设计样式（比如网络拓扑结构或配电和网络路由）。简单来说，数据交换需要决定空间和时间，因此传输及时性、可靠性和复杂度组成了基本参数。

动态信息。在复杂的网络中，数据不仅被传输还要被处理甚至产生新的数据。比如应激反应就需要在多个

阶段进行处理——即时反应—接受刺激就产生，更高级的反应则在大脑中产生，而对突发事件的反应往往由多个层面共同产生，从现场急救员到指挥控制中心。这些动态信息的考量必须纳入信息框架。

有限资源。在许多情况下数据都被有限的资源限制（比如计算装置和信号信道带宽）。利用有限的资源可以从数据中提取出多少信息？这与复杂度和信息有关，如果考虑复杂度，不同表达的结果可能会大相径庭。比如传输个百位数，接收端一次只能储存一位，最后也许只能恢复大概的数量级而不是准确的数字。

信息表达。两种不同的数据表达是否能传递相同信息？假设有两个文本，一个英语，一个波兰语，语义完全相同，则两个文本包含完全一样的信息，但用传统的信息处理方法很难确定。

信息和计算。计算信息的目的是发现能被提取的信息。无法提取可能有很多原因：目标信息从计算方面很难被提取；信息可能在物理上被分离很难在本地提取；信息可能用（量子）物理方法编码，无法完全提取。相比于传统信息理论中大多数情况下有明确的量化限制，现代信息论认为计算层面信息很难定量，因为在一般设定中易被提取信息数量的上下限差距可

能是指数级的。信息提取受到计算资源的限制是密码学的基础，对这一点的处理也是现代信息应用的关键（比如电子投票）。这些导致了许多深远的问题，比如：一个量子态包含多少比特的信息？一条信息如何定量为出错？分布式系统中的时间是什么？

合作。子系统经常会产生冲突（比如拜占庭将军问题、服务拒绝和计算机系统的自私攻击）或共谋（比如限价和内线交易）。当数据采样自不同来源，如何量化与识别信息？

信息价值。经济系统和复杂的通信网络有许多共同点：都由许多实体（代理）组成，具有强大且多样的获取、存储、共享和处理信息能力以及对信息进行权威性分级以便于应用。目前对于所有共享通信网络的民众来说，将其与经济网络区分开来的重要特点是代理的目标不满足于可靠通信。在经济系统中信息是有价值的。经济学的一个重大难题是公式化信息价值，特别是包含大量代理的动态经济体。迄今为止“信息价值”尽管已有了几个较为合理的解释，但还没有普遍认可的定义。但我们可以得到一个基本概念：信息价值与信息状态的变化有关，而信息状态由所有代理可获得回报相关的知识决定。一般来说，一个随机相关变量包含的信息价值对于可实现的期望效用影响巨大。

我们利用从数据中提取信息的观测函数，提出了信息的一般概念。一个数据组中的信息量由其可区别性决定，可区别性又依赖于观测函数输出。这样定义信息有许多优点：用数据量化信息，可识别从数据中学习的限制，通过附加数据量化信息增量可以判断采样过疏或过密。

数据科学中的复杂难题往往与以下问题相伴而行——数据范围过大（如大数据，尤其是分布式平台上的大数据）；在多重维度出现的高维数据和稀疏模式；动态结构、时间结构和异质结构；复杂依存关系；噪声和数据缺失；以及计算机语义学。数据科学的基础性综合理论必须特别关注数据中可习得和可辨识的信息，并关注满足如下条件的运算法则的设计和分析：

- › 从数据中有效提取全部可习得信息
 - › 能根据数据中的有效信息，而非仅仅是数据的大小，计算运算成本（运算精度和运行时间），由此形成高效处理信息的计算模式
 - › 具有统计学意义上的稳健性（换句话说，能在噪声和其他人为干扰环境下良好运行）
- 我们的信息结构正是旨在处理这样的问题。如图一所示，我们把

语义学整合进信息结构中，以此提取出数据中可处理的部分。

若我们无法提取出可习得的信息，且不考虑已经用过的算法，我们可以展开如下讨论。

例三：图表中的节点到达顺序

假设在一个动态网络中添加节点（或删除节点）。每个节点都有一个到达的时间戳。我们也许无法获取这些时间戳（例如，Facebook 用户的隐私设置，或蛋白质互作网络的演变过程）。如果给定一个没有时间戳的图表结构，那我们的任务就是把所有节点按从旧到新的顺序依次排列出来。结果表明，这对 Erdős-Rényi 模型和优先链接模型生成的图表无效。¹¹ 更确切地说，没有任何算法能为这些模型还原出一个较为精准的排位。有趣的是，我们可以设计算法来优化似然估计，但这些算法都无法非常精准地得出参考标准。

信息与学习极限

数据分析应用自然而然会产生一个问题，这个问题也非常重要：学习的极限在哪里，以及是否有方法能无限接近这个极限。因此，区分从数据中学习和从模型 / 分布（数据出自模

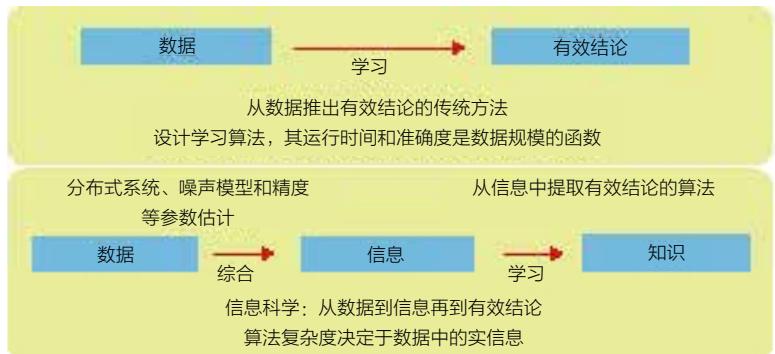


图 1. 从数据中提取有效结论的信息科学方法。

型）中学习是很重要的。大量的分析应用程序能脱离模型运行，也就是说，它们关注的是从数据中学习。相反，典型的信息理论方法关注原始代码 / 来源（分布）分析，是在一类可被允许的分布中证明极限的存在。另一方面，相关的统计学习理论则专注于建立函数，这个函数能在一系列训练数据中将给定的输入和输出放置在合适的位置上。在基于模型的环境中，已知的模型能提取出一个给定的资料组，而我们面临的问题就是从这个资料组中制定出学习的极限。为了说明这个问题和相应的解决方法，我们参考一下例三——在一个动态网络中推导出节点到达的顺序。为了便于讨论，我们假定动态图表是从优先粘贴模型中产生的。在此情况下，一个数据输入对应于一个动态图表快照，而这个观测函数又对应于一个到达顺序和它产生

输入图表的概率。学习问题需要我们识别出最大概率的到达顺序（注意，这只是其中一个公式，这个问题也会有别的对应公式）。这两个输入图表如果得出了一致的到达顺序和概率，那么从观测函数的角度来看，它们是无法被区分的。因此，为了计算出学习的极限，我们可以描述不可区分的输入的小数部分，并给出优先粘贴模型的随机观测函数结果。我们在 node-age 案例中观察到了这一点，不管我们有多强的计算能力，我们还是不能非常精准地推断出节点的到达（顺序）。

学习与数据

与学习和信息相关的一个非常重要的问题是，附加数据的可得性是否会显著地提高信息和联想学习的效率。在分布的情况下，这与采样不足的问题相关。在我们的信息框架中，这个

问题由下列思考构成：附加输入数据是否会增加信息内容和联想学习？如果附加信息增加了信息内容，那么我们认为这个模型确实是采样不足的。相反地，如果附加的信息并没有增加信息的可辨识度，那么这个分布就是采样充分的。总的来说，如之前讨论过的那样，附加信息如果增加了数据

噪声、损耗特性和缺失数据

速率失真理论和损耗压缩是信息理论的重要概念。过度拟合和抽样的理论是统计学建模的核心概念。损失函数和最优化是机器学习的核心概念。这些密切相关的概念形成了一条线索，勾连起从数据到信息到统计模型到学习模型的一条脉络，但目前仍没有理

数据库非常不完整的情况下，就像 Netflix 那样）。

数据相关的一致性模型

为了在数据科学中应用精密的信息论技术（例如极小极大原理和逐点 vs. 均匀收敛原理）¹，我们面临的第一个问题与模型选择相关。追求更丰富的模型类以更好地再现数据的需求，往往会与我们处理这些数据类型的实际数学能力形成矛盾。许多应用程序，尤其是大数据系统中的那些，迫使我们考虑一些模型类，而这些模型类过于复杂，无法与传统的模型不可知保证下的估计量兼容。这些传统的估计量是统一的，它们均匀地聚合在整个模型类上。相比之下，逐点收敛估计量往往难以用于预测，因为它们的收敛性无法被验证。Narayana Santhanam 和他的同事们摆脱了这种二分法，并通过表现一批丰富的、只允许逐点保证的模型类，提出了一个新的分析框架，可以从样本中推断出所有估计精度所需的未知模型的信息。

更正式地说，在信源编码的情况下，存在弱（或逐点）收敛性公式的问题是，我们知道一个度量是对于足够长的序列中未知分布 p 的良好通用序列编码。然而，“足够长”的概念取决于未知的 p ，因为在弱收敛公式中收敛到极限可能不是一致的。

通过不同的集成传感器，智能眼镜能够连续提供相关的食物摄入细节。

空间的覆盖范围，那么它就能增加可习得的信息。

我们用于开发方法的信息与学习系统带来了许多重要成果。尽管许多学习技巧被认为是（关于观测函数的）优化问题，但它们没有量化学习成果的影响。例如，在我们的节点到达顺序问题中，可能有多个到达顺序都是等概率最优的。这表明我们虽然可以（相对容易地）得出最优到达顺序，但同时也有许多其他的顺序有相同的发生概率。因此，对这个问题来说，即使是最优的解决方案也不那么重要。为此，在最优价值上的输入的可区分性也为解决方案的意义提供了精辟的思考。

论能将这些概念整合成一个统一的理论框架。相似的是，丢失数据和抽样这两个概念是密切相关的，但它们对于学习的影响尚未被完全得知。数据点的存在意味着什么？在处理真实世界的数据样本时，我们只能接收到实际存在事物的样本信息（我们无法肯定地说两种蛋白质在实验中不会互相反应，同样地，我们只能接收到电影观众评分后呈现出的选择性信息）。我们所需的信息论特征是有关特定数据点的存在的信息内容，这能帮助我们在未来通过实验直接解决测量新的数据点。而第一个根本性的挑战却常常被忽略，那就是评估是否存在能提取出来的、可习得的信息（尤其是在

Santhanam 的结果显示了通过引入数据驱动的弱收敛概念来使用数据本身，可以对这一点作出澄清。从广义上讲，它的目的是找到一个具有给定准确度的通用估计量 / 编码，使我们能找到一个指标，比预定的序列长度要小。为了描述这样的分布类型，**Santhanam** 和他的同事介绍了欺骗性分布¹²。在较高层次上，如果 p 的邻域的强冗余在极限范围内随着邻域缩小到 0 而发散于 0 以外，则分布 $p \in P$ 是欺骗性的。已经证实，如果一类分布中没有 $p \in P$ 是欺骗性的，那么这一类分布就是数据驱动的。

安全与隐私

随着我们建立起复杂的系统，而其运行很大程度上依赖于从数据中提取的信息，确保用于处理数据的方法的隐私性和完整性也越来越受到人们的重视。我们可以利用信息科学的基础工具来研究信息处理的其他方面，以提供关于安全和隐私的基本看法和整体观点。事实上，我们将安全和隐私看作是从数据变成信息的转化渠道中两个不可或缺的方面。在这里，我们描述了隐私保护这一背景下的我们提出的公式和结果，并向读者介绍 **Jonathan Ponniah** 及其同事的研究¹³，该研究对新的安全保护方法进行了

探索。我们通过可区分性构建起一个信息统一框架下，在这个框架的指导下，我们提出了一个有效的体系，用于从数据库中提取统计信息，同时为隐私提供可信的保证。这个整体思路为我们提供了一个对操纵数据库的算法所保存的统计信息的合理理论测量，并建立起一个保护信息的铜墙铁壁，为隐私提供切实的保证。

对隐私保护数据分析的问题探究源于数个学科的贡献，包括信息理论，理论计算机科学和统计学。缺乏明确的解决方案和对强大隐私保护算法的迫切需要，催生了关于隐私的几个概念，包括 k - 匿名性， l - 多样性和 t - 亲近性。从本质上说，这些概念的目的都是为了在类似项集合中隐藏个人数据，使个人信息无法被识别。但这些模型很大程度上忽视了潜在的分布，并没有考虑对手可能拥有的辅助信息。

受到信息科学视角的启发，我们首先确定数据库的信息内容。这要求我们明确区分不同数据库的基础是什么。当然，可区分性的基础取决于我们要需要给出的是什么查询的答案。查询实质上是数据库的功能。对典型查询产生相同反应的数据库我们自然不会去区分。事实上，从查询界面的角度来看，这些数据库拥有相同的信息。我们注意到，大多数机器学习和统计查询的目的是收集存储在数据库

中的属性之间的相关性。

一个典型的数据库查询的案例是有多少高加索人在患有 2 型糖尿病后开始吃素？在此，查询涉及种族、饮食和健康之间的相关性。数据库的直方图表现了所有属性间的相关性，因此，我们根据它们的直方图来区分数据库。换句话说，数据库的直方图封装了数据库在查询（观测函数）上的所有信息内容。两个具有相同直方图的数据库难以区分。我们注意到，我们的框架概括超出直方图的概念范围。在确定了数据库的信息内容后，我们所面临的问题自然就成了：能否提供这些信息作为回应查询的答案？我们可以允许直方图交易来促进统计信息的交换吗？我们如何防范隐私泄露？

大多数隐私研究着眼于重要的查询类别，如计数和线性查询，并设计了防止攻击者从查询反馈中重建个人条目的响应机制。这种方法的一个缺点是，虽然它可以通过响应少量查询来限制泄露的私人信息量，但如果数据库被任意次数查询，它就无法防范隐私泄露。更具体地说，在当前的典型方法中，泄漏的信息与查询的响应数量呈线性关系。为了保护隐私，大多数当前的数据库被不断查询且很少被销毁。

受到这些实际考虑的启发，我们提出了一种架构，在这个架构中，真

正的数据库通过一种特定机制进行“消毒”，让隐私无法泄露。换句话说，不管被清理的数据库有多少个查询，即使向对方提供了对所有查询的回复，对方也不能重建单个数据库条目。在意识到“消毒”机制必须配套以有力的后续处理时，我们采用差异性隐私（DP）的概念来确定“消毒”机制。Cynthia Dwork 提出了 DP 的概念，其中算法的漏洞通过其输出对单个条目的敏感度来进行量化。¹⁴ 因为 DP 认为有必要对输出进行随机化以保护隐私，所以在数据库 D 上运行的算法输出的特征在于其输出的概率分布 $P_D(\cdot)$ 。算法对个体条目的敏感性可以通过分布密度 $D(\cdot)$ 和 $P^*(\cdot)$ 来获得，其中 D 和 \hat{D} 是恰好在一个个体条目中不同的数据库。在 DP 文献中被称为机制的随机算法是 ϵ -DP，如果对于在单个条目和每个输出 y 中不同的每对数据库 D ，其对数似然比是

$$\left| \log \frac{\mathbb{P}_D(y)}{\mathbb{P}_{\hat{D}}(y)} \right| \leq \epsilon = \ln \theta. \quad (1)$$

较大的 ϵ 对应于对个体条目更敏感的隐私保护，反之亦然。因为 ϵ 量化了对方可以从 ϵ -DP 机制的输出中获取到任何个体的条目数量，所以它必须被视为隐私预算。DP 在计算机组合和后期处理中享有两个理想的恒定性。如果一个人被提供了多个 DP 机

制的响应，那么这个装置就能对对手了解他的条目的范围进行限制。其次，DP 机制的输出不能以任何可能会揭示关于个人条目更多信息的方式进行后续处理。这些特性，特别是后者，使其非常适用于目前遇到的问题。

由此，我们来到了以下结构中。真正的数据库通过 ϵ -DP sanitizing 机制进行“消毒”，其中 ϵ 是指定的隐私预算。原始数据库是安全的，并且有已经“消毒”过的数据库用于查询。我们在给定数据库上构建直方图的情况下说明了这种方法。该框架自然延伸到相互关系的计算。数据库的直方图保留了所有相互关系，并且保留直方图的任何转换的输出都保留了大多数典型数据库查询所需要的信息。这表明，设计直方图来保护转换过程，可能是保护隐私和提供“最佳”准确性的一种自然方法。为了解决后一个问题，我们采取了率失真的方法。在率失真理论中，任何信息源与其（压缩）表示之间的紧密度是由一种类似于使用源与其表示之间的度量来量化的。

科学、工程、农业和商业中还有许多其他领域都可以从统一的信息科学中获益。

相应地，我们定义一对直方图之间的度量来量化信息源（原始数据库的直方图）与其表示（转换数据库的直方图）的接近程度。

例四：保护隐私和真实性

我们提出了一个优化问题（线性规划）的公式，能在最小化不准确度时保留 DP。¹⁵ 假设有一个包含 n 个主题的数据库，每个主题都用一份记录来标识。我们令 $R = \{a_1, \dots, a_k\}$ 表示记录集合。我们也写下 p_k 为主体的概率是 $a_k \in R$ 的记录。令 $r = (r_1, \dots, r_n) \in R^n$ 表示一个（通用的）有 n 条记录的数据库。数据库的直方图定义如下：对于数据库 $r \in R^n$ 和记录 $a_k \in R$ ，我们令 $h(r)_k = \sum_{i=1}^n 1_{\{r_i=a_k\}}$ ，其中 1_A 是 A 的指标函数，用记录 a_k 表示对象的数量，并且 $h(r) = (h(r)_1, \dots, h(r)_n)$ 表示与数据库 $r \in R^n$ 对应的直方图。让 $\mathcal{H}^n = \{(h_1, \dots, h_K) \in \mathbb{Z}^K : h_i \geq 0, \sum_{k=1}^K h_k = n\}$ 表示直方图的集合。为了保护隐私，我们采用 DP 数据库消毒机制 M （DSM）输出随机消毒过的数据

库。更确切地说，我们将 DSM, $M : \mathcal{R}^n \rightarrow \mathcal{R}^n$ 定义为 θ -DP, 其中 $\theta = e^\epsilon$, 如果对于每一对相邻数据库 r, \hat{r} 和每个数据库 $s \in \mathcal{R}^n$, 我们有 $\theta W_M(s|r) \leq W_M(s|\hat{r}) \leq \theta^{-1} W_M(s|r)$, 其中 $W_M(s|r)$ 是将数据库 r 转换为 s 的概率。现在，我们可以制定我们的优化问题，以尽量减少误差，同时保护 θ 隐私。

$$D_n^*(\theta) = \min_{W(\cdot| \cdot)} \sum_{h \in H^n} \sum_{g \in H^n} \binom{n}{h_1 \dots h_K} \left(\prod_{k=1}^K p_k^{h_k} \right) W(g|h) |h - g|$$

受限于：

$$W(g|h) \geq 0, \sum_{g \in H^n} W(g|h) = 1, \text{ and } W(g|h) - \theta W(g|\hat{h}) \geq 0$$

每一对直方图中，其中 $|h|_1$ 是 L_1 的长度。这个线性变成问题可以解决 n 的问题，导出以下：

$$D_K^*(\theta) = 2\theta \left\{ \frac{K-1}{1-\theta} + \frac{S_{K-1}(\theta)}{S_{K-1}(\theta)} \right\}, \text{ where } S_{K-1}(\theta) = \sum_{j=1}^{K-1} \theta^j \left[\binom{K-1}{j} \right]^2 \quad (2)$$

其中 $S'_{K-1}(\theta)$ 是 $S_{K-1}(\theta)$ 的导数，最佳机制由下式给出：

$$W^*(g|h) = \xi_{\rho,f}^{-1}(\theta)(\theta)\theta^{-\frac{|g-h|_1}{2}},$$

其中 $\xi_{\rho,f}(\theta)$ 是与 Ehrhart 序列有关的归一化常数。

在上面的案例中，我们回答了这个问题：给定 ϵ -DP，泄露一个已消毒的数据库时，我们在直方图上的准确度会损失多少？不过，相反的问题更加有趣：当发布真实数据库 D 的“匿名”数据库 D' 时，要量化 D' 发布时的隐私损失有什么好方法么？迄今为止，我们认为这个重要问题没有很好的答案。

科学、工程、农业和商业中还有许多其他领域都可以从统一的信息科学中获益。例如，最近在模型和方法开发方面取得的进展在生命科学等领域表现出了惊人的成果。信息是生命系统的一个重要方面。Paul Nurse 认为，生物学目前站在了一个有趣的关头上，因为我们仍然缺乏对重要的高级生物现象的综合理解。¹⁶ 他进一步认为，系统生物学等学科的成功必须辅之以深入研究生命系统如何聚集、处理、存储和使用信息的问题。在经济学中，信息如何被重视的问题非常重，经济系统中的信息流动以及相关的控制问题也是如此。在社会科学领域，了解信息的来龙去脉，对个人和集体行动的影响以及它对整体社会和经济福祉的影响也是至关重要的。

为了实现这些具有挑战性的目标，我们必须对我们的科学和工程教育方

法进行改进和现代化。人们普遍承认，今天的学生需要一套丰富的信息技能。我们需要教育和多样化，尤其是考虑到学术界、工业界、政府和非盈利组织之间越来越大的差距，这些差距都迫切需要具有不同背景、经验和想法的受训人才来解决。这些学生必须具备信息科学方面的素养，以及解决大数据及其应用中的复杂问题所需的专业能力和专业知识。■

感谢

这项工作得到了美国国家科学基金会信息科学中心 (CSol) Grant CCF-0939370 的支持。我们要感谢信息科学中心的成员为中心的科学计划做出的贡献。特别要感谢 D. Gleich, P. Dineas, A. Magner 和 A. Padakandla 对本文的付出和贡献。

参考文献

1. T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2006.
2. D. Angluin, “Queries and Concept Learning,” *Machine Learning*, vol. 2, no. 4, 1988, pp. 319–342.
3. L.G. Valiant, “A Theory of the Learnable,” *Comm. ACM*, vol. 27, no. 11, 1984, vol. 1134–1142.
4. V. Vapnik and A. Chervonenkis,

关于作者

沃伊切赫·希潘诺夫斯基是普渡大学计算机科学讲席教授。他的研究兴趣包括算法分析、信息理论、生物信息学、解析组合数学、随机结构，以及分布式系统的稳定性问题。他在波兰格但斯克理工大学获得计算机工程博士学位，他还是IEEE会士。联系方式: szpan@purdue.edu。

安纳斯·格拉马是普渡大学的计算机科学教授。他的研究兴趣包括并行计算和分布计算、大规模数据分析及其在生命科学上的应用。他在明尼苏达大学获得计算机科学博士学位。联系方式: ayg@purdue.edu。

“On the Uniform Convergence of Relative Frequencies of Events to their Probabilities,” Theory of Probability & Its Applications, vol. 16, no. 2, 2004, pp. 264–280.

5. S.A. Motahari, G. Bresler, and D. Tse, “Information Theory of DNA Sequencing: A Basic Model,” IEEE Int’l Symp. Information Theory Proc. (ISIT 12), 2012; doi: 10.1109/ISIT.2012.6284020.

6. C. Shannon, “The Lattice Theory of Information,” Trans. IRE Professional Group on Information Theory, vol. 1, no. 1, 1953, pp. 105–107.

7. C. Shannon, “A Mathematical Theory of Communication,” The Bell System Technical J., vol. 27, no. 3, 1948, pp. 379–423.

8. W. Szpankowski, “On Asymptotics of Certain Recurrences Arising in Universal Coding,” Problems of Information Transmission, vol. 34, no. 2, 1998, pp. 142–146.

9. Y. Choi and W. Szpankowski, “Compression of Graphical Structures: Fundamental Limits and Algorithms and Experiment,” IEEE Trans. Information Theory, vol. 58, no. 2, 2012, pp. 620–638.

10. F.P. Brooks, “Three Great Challenges for Half-Century-Old Computer Science,” J. ACM, vol. 50, no. 1, 2003, pp. 25–26.

11. A. Magnier et al., “TIMES: Temporal Information Maximally Extracted from Structures,” to be published in 2018; www.cs.purdue.edu/homes/spa/papers/www2018.pdf.

12. N. Santhanam et al., “Data Dependent Weak Universal Redundancy,” IEEE Int’l Symp. Information Theory, (ISIT 14), 2014; doi: 10.1109/ISIT.2014.6875159.

13. J. Ponniah, Y.-C. Hu, and P.R. Kumar, “A Clean Slate Approach to Secure Ad Hoc Wireless Networking—Open Unsynchronized Networks,” IEEE Trans. Control of Network Systems, vol. 4, no. 1, 2017, pp. 37–48.
C. Dwork, “Differential Privacy,”

14. Proc. 33rd Int’l Colloquium Automata, Languages and Programming (ICALP 06), 2006, Springer Verlag; www.microsoft.com/en-us/research/publication/differential-privacy.

15. A. Padakandla, P.R. Kumar, and W. Szpankowski, “Preserving Privacy and Fidelity via Ehrhart Theory,” preprint, 2017; www.cs.purdue.edu /homes/spa/papers/isit18-privacy.pdf.

16. P. Nurse, “Life, Logic, and Information,” Nature, vol. 454, 2008, pp. 424–426.

中国中文信息学会、语言与知识计算专业委员会 (CCKS)



CCKS·2018

智能客服问句匹配

2018.05.01-2018.07.20

哈工大 x 微众银行



扫码了解详情



朝着普适操作系统发展： 一个软件定义的角度

文 | 梅宏 (Hong Mei)、郭耀 (Yao Guo)，北京大学 (Peking University)

译 | 陈英祁，浙江大学

近年来，操作系统已超出传统计算系统的范畴，朝着云、物联网设备和其它新兴技术范延伸，不久将广泛存在。尽管现有的操作系统各不相同，但它们有一个共同点那就是所谓的“软件定义能力”——即资源虚拟化和函数编程。

在《电脑》(Computer) 2016年1月刊中，Dejan Milojicic 和蒂莫西·罗斯科 (Timothy Roscoe) 根据当今的硬件和应用发展趋势，¹预测10年后的操作系统。他们没有聚焦于传统计算系统中的操作系统，如个人电脑，服务器和嵌入式系统，而是从软件定义的角度讨论操作系统的未来走向。近年来提出并研发了各式各样的操作系统用于大小不一的设备，规模从单个计算机到计算机组不等，既涉及硬件又涉及软件，应用范围从智能家庭到智慧城市不等。虽然这些操作系统看上去天壤之别，但它们都体现传统操作系统的原理和特点，即资源虚拟化和函数编程。

资源虚拟化和函数编程是所谓的“软件定义”系统的核

心，包括软件定义网络 (SDN)，² 软件定义存储 (SDS) 和软

件定义数据处理中心 (SDDC)。就像传统的操作系统运用软件抽象管理硬件系统为应用提供运行支持那样，我们认为未来的操作系统会为新兴技术提供所有软件定义能力。因此，一个软件定义网络就是网络硬件的操作系统，而一个软件定义云就是云的操作系统。我们把这些操作系统成为普适操作系统 (UOS)。

操作系统简史

操作系统是一层在应用和计算机硬件之间的系统软件，管理如处理器、存储器和储存等资源，同时为在其之上运行的应用提供支持。³

最早的计算机没有操作系统，那时的软件应用直接在裸

表1. 传统操作系统的演变

时间表	代表性操作系统	计算机系统	主要特点
1956	GM-NAA I/O	IBM 704	最早实用型操作系统 单道批处理系统 输入 / 输出管理
20世纪 60年代	IBM OS/360 系列	IBM OS/360 系列——大型主机	分时操作 多道批处理 内存管理 虚拟计算机
20世纪 70年代	尤尼斯操作系统	小型计算机/工作站	第一个现代操作系统 用独立于计算机的语言研发 提供标准接口 集成开发环境
20世纪 80年代	Mac操作系统, Windows微软操作系统, Linux操作系统	个人电脑 (PCs)	提供现代图形用户界面 (GUI) 为个人用户提高可用性
21世纪	苹果iOS操作系统, 谷歌安卓, Windows Phone手机操作系统	智能手机	对传统操作系统用户化 为移动设备提高可用性 新型应用程序调运模型 (苹果应用商店、谷歌市场)

机上运行。然而，随着计算机系统越来越复杂，直接在一个应用上管理资源越来越难。因此，越来越多的功能抽象成驱动程序和函数库，创造出一个可在不同应用中共享的系统软件层。这个软件层称为“操作系统”，因为它最初用于抽象系统的操作能力，减轻操作员的负担。然而，如今的操作系统不再着重于系统“操作”，而是着眼于资源管理、应用研发和为特定系统提供运行支持。

表1列出了主要的传统操作系统并对比了它们的主要特点。现代操作系统多半运用基于 Unix 的计算机操作系统，同时为特定类型的系统量身定制个性化功能。例如，微软(Windows)和苹果 mac 的操作系统关注图形用户界面(GUIs)，为台式机用户提供更好的用户体验。而如安卓和苹果手机(iphone)的操作系统包含支持手机程式研发的应用层和为手机用户提供更好使用体验的执行程序。

20世纪 80 年代，随着计算机网络普及，日益需要操作系统提供联网能力，由此产生了联网中间件和许多新生联网操作系统(NOSs)。最早的联网操作系统，如 Novell Netware，着眼于在本地网络中联接计算机。然而，后来随着联网成为许多用户的必需品，这种联网方式中断了，且联网能力多半加进更新版本的台式机操作系统中。

20世纪 90 年代中期掀起了互联网操作系统的讨论热潮。那时正值微软公司(Microsoft)和网景公司(Netscape)互相竞争，推出了一系列面向下一代基于互联网应用的新工具和编程接口。从那时起推出了许多互联网操作系统，包括基于 Java 的 Java 操作系统，还有最近的基于谷歌浏览器的 Chrome 操作系统。

许多联网操作系统和互联网操作系统通过操作系统样式的结构提供联网能力或者加入与互联网相关的数据

管理功能。它们的原件可能在地域分布式的计算机系统甚至是虚拟计算机(VM)上运行，通过互联网提供特定的服务。这些新式“元操作系统”通常在如微软(Windows)或 Linux 的传统操作系统之上运行，为基于互联网的应用和服务提供支持。

软件定义的角度

“软件定义”已经成为学术圈和行业中最热门的流行词之一。它描述了一组技术，包括自防御网络(SDN)、科学数据系统(SDS)、软件定义数据中心(SDDC)，还有有时称为“软件定义的一切事物”(SDX)。在一个软件定义的系统中，硬件资源可虚拟化，由操作系统例程或控制面板管理，用户可通过编写程序获得和管理虚拟化资源提供的服务。⁴

我们认为操作系统提供同样的能

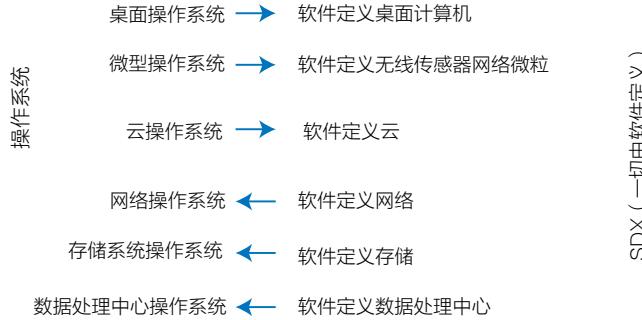


图 1. (a) 操作系统和软件定义系统大多基于相同的原则。

力。例如，传统的操作系统，如 Linux 或 Windows，通过硬件驱动程序提供虚拟化的硬件资源，通过软件开发包（SDKs）和元件库提供应用开发和运行支持。手机操作系统如安卓则以更高层次的抽象为手机应用程序提供额外软件定义层，包括手机数据管理（例如，联系人和位置数据），一组支持应用程序研发的 API 和一组支持应用程序执行的元件库。不管操作系统是在小型设备（如微操作系统 TinyOS5）还是大型集群（如云操作系统）中运行，它们提供虚拟化资源和函数编程能力。换言之，“软件定义”技术是针对于特定技术的操作系统。例如在自防御网络（SDN）中，控制面板提供编写高层次应用程序的能力，管理联网功能，同时控制面板虚拟联网资源。

总而言之，如图 1 所示，操作系统和软件定义系统大体上基于相似的工作原理。

普适操作系统

25 多年前，马克·维瑟（Mark Weiser）设想一个计算机编程广泛存在的世界。⁶那时看来他的预测不切实际。但随着物联网出现，随着日常用品能编

写成程序，如智能灯能根据环境调节亮度，这种预测越来越成为现实。同样地，我们认为操作系统也会广泛存在。

普适操作系统包含一种用于软件定义世界的新型操作系统，能用软件管理生活的方方面面。为了弄清普适操作系统的广泛影响，请看以下例子：

- 网络操作系统。（Web OS）网络操作系统，也称为网络桌面或者网桌，在浏览器内为用户提供像 Linux 那样的环境运行应用程序和管理所有他们的数据和储存。其中也包括为软件开发人员提供的应用程序界面（API），帮助其研发能在浏览器内运行的应用程序。网络操作系统这个例子中就包括了火狐浏览器操作系统，谷歌浏览器操作系统，eye 操作系统，You 操作系统和 G.ho.st 操作系统。
- 机器人操作系统。机器人操作系统（ROS）是一个元操作系统，为复杂强大的机器人应用程序提供研发和运行支持。⁷ 它广泛的开源工具、元件库、抽象和应用程序界面（APIs）可用于众多不同的平台。

，家庭操作系统。它由微软提出，旨在创造“每个智能家庭”。家庭操作系统旨在简化家庭自动化技术的研发和管理。⁸ 它为设备编排提供直观的用户控制和更高水平的抽象。家庭操作系统的科研样机已配置在超过 12 个家庭中。

，城市操作系统。许多倡议倡导研发推进经济发展、促进能源利用和推动环境可持续发展的操作系统。例如，“生活计划信息技术城市操作系统”（living-planit.com）为能源、水资源、废物管理、交通运输、电信和医疗保健系统提供抽象和管理界面和保证不同平台互相操作的应用程序界面（APIs）。

，云操作系统。从概念上说，云操作系统的作品内容与传统的操作系统一样：管理应用程序和硬件。但它的规模扩大到云计算，对象存储代替了文件系统，此外拥有几乎无限的存储能力和输入/输出吞吐量。与在实体计算机上管理处理器不同，云操作系统在虚拟存贮器（VM）上管理任务。更重要的是，它为云应用程序提供各式各样的应用程序界面（API）使用云资源。许多云服务提供商已创建他们自己的云操作系统，包括微软智能云（Microsoft Azure）、亚马逊网络服务（AWS）和华为

FusionSphere云操作系统。还有许多广受欢迎的源代码开放云操作系统,如Open-Stack和Apache Cloud Stack。

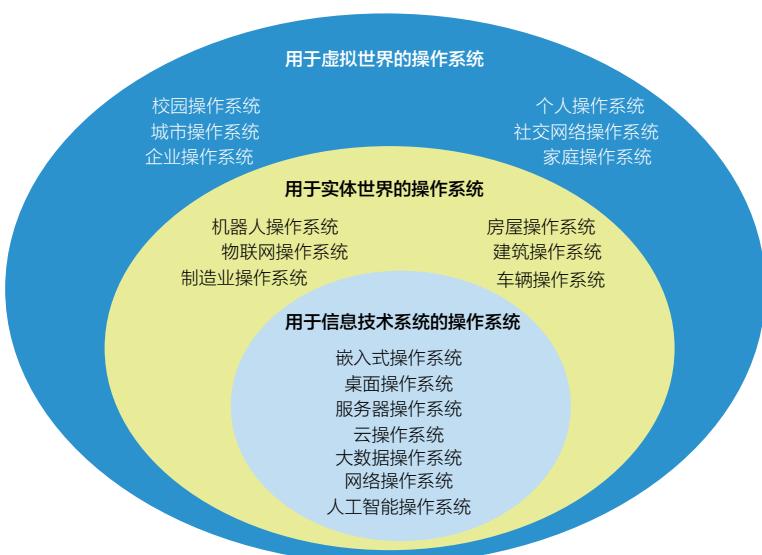
- 物联网操作系统。谷歌的Android Things (Brillo)是个专为低功耗和内存受限的物联网设备设计的嵌入式操作系统平台,它使用安卓应用程序界面和谷歌服务。

图2展示了普适操作系统的一般结构,与传统操作系统的结构类似。普适操作系统体现了与传统操作系统相同的关键概念:资源虚拟化和函数编程能力。但这些概念通常在广泛运用中更清晰:

- 为资源管理抽象。

普适操作系统提供抽象能力管理超出传统计算的各种资源并储存资源。这些功能很像传统操作系统中的驱动程序或者硬件的抽象层,但能更广泛地景象资源虚拟化。此外也为用户和应用程序提供应用程序界面(**API**),获得这些虚拟化资源。例如,用于社交网络的普适操作系统不仅管理用户的信息和关系,还会追踪用户的行为和各用户间的交流。

- 为普适应用程序的研发和运行



支持。普适操作系统像传统操作系统一样，为应用程序提供应用程序界面、程序设计模型、元件库和研发工具。然而，在普适应用程序在普适操作系统之上运行时，这种支持层次更高，而普适操作系统反过来在如Linux和Windows的传统操作系统之上运行。它们之间的区

智能手机和平板电脑等移动设备、传统台式电脑(PC)和笔记本电脑、单机工作站和网络服务器、以及服务器集群和云的操作系统已经出现。我们预计，几乎所有传统系统和下一代系统都将使用普适操作系统，从微型边缘计算设备到跨大洲的巨大分布式计算环境，概莫能外。普适操作系统也可以用于新兴应

我们可以为虚拟世界中的每个实体构建一个普适性操作系统。除实体对象和系统以外，我们也可以为特定应用程序领域中的实体创建操作系统。例如，我们可以为包括家庭、企业、公共机构和政府机构在内的各种类型和规模的组织添加软件定义功能来管理人员、信息、时间表和库存。操作系统将提供抽象来管理资源以及支持新应用程序的开发和执行。

普适操作系统包含一种用于软件定义世界的新型操作系统，能用软件管理生活的方方面面。

别在于，普适操作系统支持第三方研发商的应用程序，而现有的典型非操作系统方法则安装在系统中的专有层。

用程序领域，如大数据和人工智能。

我们可以为实体世界中的每个对象(或对象集合)构建一个普适性操作系统。普适性操作系统的目标是将计算能力扩展到传统信息技术系统之外，以使所有对象变得更加智能。这将最终意味着要使这些实体可受程序控制，而这就需要一个操作系统。我们早已在机器人(甚至乐高机器人)身上安装了操作系统。在智能住宅中，所有电器(包括电视机、洗衣机、冰箱、灯、微波炉和咖啡机)都需要一个程控的操作系统。所有移动物体，包括车辆、无人机、自行车、轮椅、甚至婴儿车，也都需要普适性操作系统。

普适操作系统展望

如图3所示，我们设想了一些可用于不同实体(包括实际和虚拟实体)以及传统信息技术系统的普适系统。

普适操作系统原则

这一设想以三条基本原则为基础。

普适操作系统可以用于任何规模的系统。目前，用于小型嵌入式系统、

普适性操作系统种类

鉴于这些原则，我们有望看到许多不同种类的普适性操作系统，例如：

- 大数据操作系统。目前，人们已经为多个不同领域构建了大数据应用程序。大数据操作系统可为数据抽象和管理、数据访问和应用程序接口管理、以及大数据应用程序的编程模型和语言提供特殊功能。
- 企业操作系统。未来的企业或组织可能需要操作系统来支持业务流程以及人员、资金、机器等资源的高效管理。我们可以在现有的信息管理系统、企业资源规划系统等企业系统的基础上，通过添加编程应用程序接口来创建企业操作系统，以支持企业应用程序的灵活开发。
- 工业/制造业操作系统。许多制

造商已经配备了自动化生产和机器人控制系统。虽然这些系统中有很多已经在通过简单的嵌入式系统进行管理,但新的软件定义抽象和通信功能将进一步提高系统的效率,并使系统更加智能。

› **人-网络-实体操作系统。**计算机行业的一个新趋势是三个先前相互隔绝的领域——人类、网络系统和实体世界——正在融合。这将带来除目前的信息物理系统和物联网以外的许多有趣的应用。然而我们将需要新的软件定义抽象和功能来支持人-网络-实体系统管理、应用程序开发和通信。

› **人工智能操作系统。**我们将需要操作系统来为机器学习或深度学习功能提供抽象,以及为人工智能应用程序提供编程支持。安卓系统的联合创始人安迪·鲁宾(Andy Rubin)近期预测,人工智能将成为操作系统的下一个主要突破。人工智能操作系统将成为新型智能应用成功的关键基础。

技术挑战

普适性操作系统虽拥有广阔前景,但也面临着诸多技术挑战。

› **普适性操作系统模型和架构。**通用的普适性操作系统模型和架构

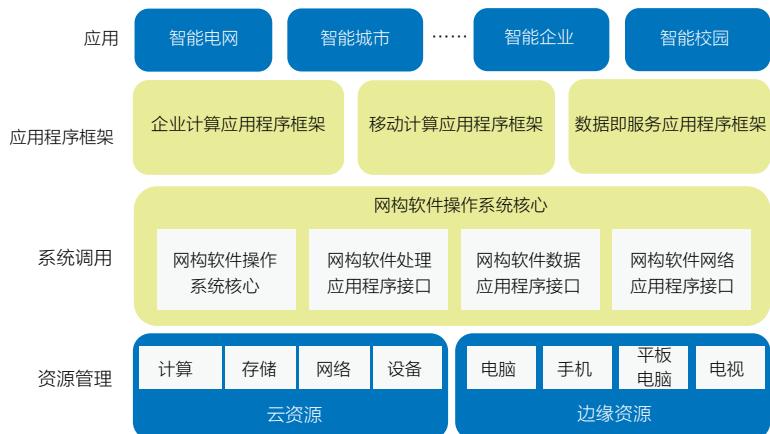


图4. 网构软件操作系统架构。网构软件应用程序运行在云端和边缘设备之上。网构软件操作系统内核提供抽象以管理云和边缘资源,而应用程序框架层则适用于不同领域的应用程序。

可能不适用于所有普适性操作系统。最重要的因素是抽象和编程接口的粒度。较小的粒度可以实现更大的灵活性,但是可能会降低应用程序运行时的性能。在这两者之间作出权衡将成为普适性操作系统架构设计的核心。

› **资源虚拟化。**虚拟化是所有操作系统和软件定义世界的关键技术。借助普适性操作系统,我们将把计算从中央的云挤压到边缘,例如智能手机和物联网设备。因此,我们需要调查轻量级虚拟化技术,以提供高效的操作系统抽象,并支持软件定义的边缘计算。

› **性能优化。**在用于小型计算机系统或计算能力较弱的对象的普适性操作系统中,提高应用程序执

行性能将变得至关重要。随着更多类型的硬件、资源和应用程序出现,提供高效服务将成为一项挑战,特别是在高吞吐量和大规模并行处理的情形下。

› **安全和隐私。**软件与硬件相比在安全威胁面前更加脆弱。在普适性操作系统安装完毕后,软件会成为系统或环境的控制中心,进而成为攻击者的主要目标。此外,若普适性操作系统被用于管理敏感个人数据或重要信息的系统,隐私也将成为首要考虑因素。

› **特定领域的编程语言。**当前的高级编程语言,例如C/C++和Java,都是为计算机设计的。为特定的普适性操作系统如企业操作系统开发更高效的应用程序需要新的特

定领域的语言。

- 实现真正的智能。软件是所有智能应用程序的基础。为实现真正的智能，普适性操作系统和应用程序必须能够“思考”，即智能地管理和执行。

网构软件操作系统：普适性操作系统的原型

网构软件是新型互联网应用的典范，具有自主性、演化性、协同性、多态性和反应性等特征。网构软件由一系列分布在互联网上的自主软件实体和一组连接器组成，通过连接器实现自主软件实体间的各种协作。软件实体感知运行时环境中的动态变化，并通过结构和行为演变不断适应这些动态变化。

我们一直在进行研究，力图为网构软件构建一个拥有一组软件定义特点的操作系统，以抽象出网构软件应用程序的低级资源管理功能。图 4 显示了我们的网构软件操作系统的总体架构，我们认为它是未来基于互联网的应用程序的普适性操作系统的原型。在操作系统内部，网构软件应用程序在包括云和边缘设备在内的现有硬件系统之上运行。网构软件操作系统内核提供抽象以管理云和边缘资源，而应用程序框架层则适用于不同领域的应用程序，例如企业计算、移动计算和数据即服务。

我 们 构建的网构软件操作系 统 实 例 包 括 以 下 内 容：

- YanCloud。YanCloud 是一个用于组织内部私有云计算系统的云操作系统，它支持几乎所有现有的虚拟机技术，包括 Xen、VMware 和 KVM。它具有软件定义功能，能通过基于框架结构、模型驱动的运行时管理机制生成云管理应用程序。众多企业以及联想和方正等主要云服务器厂商都已安装了 YanCloud。

- CampusOS。CampusOS 是支持基于网络的大学校园应用程序的操作系统的原型，它提供抽象来管理师生的个人信息、课程时间表和学校活动等资源，以及软件定义的应用程序接口和软件开发工具包，以支持校园应用程序的开发和执行。

- YanDaaS。最近，我们开发了一个用于数据管理和在不同类型的传统软件系统之间实现共享的网构软件操作系统。顾名思义，YanDaaS 提供数据即服务 (DaaS) 功能。其主要目标是在无需使用原始源代码的情况下，通过自动化的应用程序接口生成和新应用程序的开发，将彼此孤立的传统

软件系统和应用程序连接起来。

YanDaaS 已成功在中国智慧城市项目覆盖的数百个传统工业系统中安家落户。

- 随着物联网快速发展，许多普适性操作系统将会出现，它们将提供软件定义功能，特别是资源虚拟化和功能可编程性功能，以支持新型普适性应用程序的高效安装和管理。但是，在普适性操作系统的架构、系统性能以及安全和隐私方面，仍有若干关键性的技术挑战亟待解决。虽然如此，我们预测，普适性操作系统将会出现在各个计算机和计算机辅助领域，包括机器人学、企业计算、制造、大数据和人工智能。为此，我们未来的工作将包括开发用于无人系统、工业控制和类脑计算等新领域的网构软件操作系统。■

参考文献

1. D. Milojevic' and T. Roscoe, "Outlook on Operating Systems," Computer, vol. 49, no. 1, 2016, pp. 43–51.
2. N. McKeown et al., "OpenFlow: Enabling Innovation in Campus Networks," ACM SIGCOMM Computer Communication Rev., vol. 38, no. 2, 2008, pp. 69–74.

关于作者

3. H. Mei and Y. Guo, "Network-Oriented Operating Systems: Status and Challenges," *Scientia Sinica Informationis*, vol. 43, no. 3, 2013, pp. 303–321 (in Chinese). 梅宏、郭耀. 面向网络的操作系统——现状和挑战[J]. 中国科学(信息科学), 2013, 43(3):303–321

4. H. Mei et al., "Understanding Software-Defined from the Perspectives of Software Researchers," *Comm. China Computer Federation*, vol. 11, no.1, 2015, pp. 68–72 (in Chinese).

5. P. Levis et al., "TinyOS: An Operating System for Sensor Networks," *Ambient intelligence*, W. Weber, J. Rabaey, and E. Aarts, eds., Springer, 2005, pp. 115–148.

6. M. Weiser, "The Computer for the 21st Century," *Scientific Am.*, vol. 265, no. 3, 1991, pp. 94–105.

7. M. Quigley et al., "ROS: An Open-Source Robot Operating System," *ICRA Workshop Open Source Software*, vol. 3,no. 3.2, 2009; www.willowgarage.com/sites/default/files/icraoss09-ROS.pdf.

8. C. Dixon et al. "An Operating System for the Home," *Proc. 9th USENIX Symp. Networked Systems*

梅宏教授, 北京大学高可信软件技术重点实验室教授, 北京理工大学副校长, 研究领域包括系统软件和软件工程, 拥有上海交通大学计算机科学博士学位, 是中国科学院院士、电气和电子工程师协会成员、中国计算机联合会和世界科学院院士。联系方式: meih@pku.edu.cn。

郭耀, 北京大学高可信软件技术重点实验室和信息科学技术学院教授, 研究领域包括移动计算、操作系统和移动应用程序分析, 拥有马萨诸塞州阿默斯特大学计算机工程博士学位, 是电气和电子工程师协会和国际计算机协会成员。联系方式: yaoguo@pku.edu.cn。

Design and Implementation
(NSDI), 2012, pp. 25–25.

9. T. Haselton, "The Man behind Android Says A.I. Is the Next Major Operating System," CNBC, 18 Aug. 2017; www.cnbc.com/2017/08/18/andy-rubin-says-ai-is-next-big-operating-system.html.

10. H. Mei, G. Huang, and T. Xie, "Internetware: A Software Paradigm for Internet Computing," Computer, vol. 45,no. 6, 2012, pp. 26–31.

11. H. Mei and J. Lü, *Internetware: A New Software Paradigm for Internet Computing*, Springer, 2016.

12. H. Mei and Y. Guo, "Development and Present Situation of Internetware Operating Systems," *Science & Technology Rev.*, 2016, vol. 34, no. 14, pp. 33–41 (in Chinese). 梅宏、郭耀. 面向网构

软件的操作系统:发展及现状[J].
科技导报, 2016, 34(14): 33–41

13. X. Chen et al., "Towards Runtime Model Based Integrated Management of Cloud Resources," *Proc. 5th Asia-Pacific Symp. Internetworkware (Internetworkware 13)*, 2013, article no. 1.

14. P. Yuan, Y. Guo, and X. Chen, "Towards an Operating System for the Campus," *Proc. 5th Asia-Pacific Symp. Internetworkware (Internetworkware 13)*, 2013, article no. 24.

15. G. Huang et al., "Programming Situational Mobile Web Applications with Cloud-Mobile Convergence: An Internetware-Oriented Approach," *IEEE Trans. Services Computing*, 2016; doi:10.1109/TSC.2016.2587260.

再谈可扩展 相干共享内存

文 | C. 戈登·贝尔 (C. Gordon Bell)、艾克·纳西 (Ike Nassi)，微软研究院（退休）、

TidalScale 公司和加州大学圣克鲁兹分校

译 | 詹冰冰，浙江大学

可扩展与相干共享内存一直是人们梦寐以求的目标，但这一目标却相当难以实现。与当下流行的分布式计算模型相比，作者提出了一种软件定义纵向扩展式共享内存多处理器，这一服务器架构仍然运用常见的商品横向扩展集群。

在对这一架构探索的初期，我们很少配置有高性能系统，因为这种系统不仅价格昂贵，并且访问也会受到限制。而我们现在的系统则前进很大一步，不仅能够按需提供多处理器和多线程计算平台，而且价格也更加低廉，但同时，这一弹性的计算服务可能会让人产生基础硬件系统正常工作的错觉。在这方面，一些关于亚马逊网络服务 (AWS) 的实验得出了令人意想不到的结果。¹ 在 AWS 四处理器实例上运行的 CPU 与内存密集型数据生成应用程序显示，一个使用 32 个并行线程的程序，其 CPU 标准化利用率为 89%。该应用程序运行了 4,832 CPU 秒（见表 1）。为了进一步探究这一点，我们将并行度降低到了 16 主存，这实际上加快了程序的运行速度。重复这一过程，最佳试验仅用 10% 的线程就使性能提升了 15.79 倍。这一结果表明，除去非

因素 I/O，物理处理器的内存争用是一个不容忽视的存在。为此，我们该如何应对？

背景

高性能计算机 (HPC；也称超级计算机) 从单存储器计算机到多处理机系统的市场转型始于 1987 年，当时桑迪亚国家实验室团队成功地将 1024 台个体计算机（称为节点）组装成单台 Ncube 计算机，该团队也凭借此项成果获得了第一个戈登贝尔并行奖。1993 年，思考机器 (Thinking Machines) 上的一台 1024 节点连接机在性能上超过了所有的传统超级计算机，比如已经达到纵向扩展极限的 Cray YMPs。1994 年，MPI (Message Passing Interface 消息传递接口) 和 Beowulf

标准确立了开始向HPC计算机集群过渡。2018年，所有HPC应用程序都在一组计算机集群中运行，这些计算机紧密连接到一个网络中，数目多达几百到一千万台。

由于在应用级别上，应用程序运行在假定并行化的系统中，因此商业事务处理的情况有很大的不同。随后，显式和隐式并发支持的新编程语言得到了开发。1979年，来自加州理工学院的一个研究团队组建了Teradata，并构建了一个高度可扩展的并行关系数据库，该数据库建立在对磁盘存储的“无共享”访问上。类似地，今天的云网络应用程序相当于一个大型的并行和流水线程序集合，但是这种配置仍然是静态的，并且具有明确的进程间通信。

21世纪中期，随着MapReduce²和Hadoop³的推出，一些程序间转换的实现，逐步建立在访问大量更加多样化资源的基础之上。这些技术都需求巨大的存贮记忆。几年前，机器学习开始进入生产状态。最初，这些应用程序与HPC相类似，反之亦然，两者之间正在发生着一定的融合。总之，由于需要管理和分析不断增加的数据量，因此这些技术必须在分布式架构中取得进展，以简化编程。这些硬件和软件体系要求具备更大的存储量，更广泛的连接和更强的计算能力，其中包括具有各种中央处理器单元和图形处理单元集群、现场可编

表1. 关于亚马逊网络服务（AWS）四处理器实例上运行的CPU与内存密集型数据生成应用程序的实验结果

并行度	CPU秒	CPU标准化利用率 (%)	总运行时间 (秒)
1	259.0	3	257.7
2	335.0	6	165.0
2	333.0	6	164.2
3	306.0	9	102.4
3	375.0	10	122.5
4	484.0	13	118.6
4	482.0	13	117.7
4	475.0	13	116.1
4	356.0	13	86.4
5	563.0	16	110.4
5	590.0	16	115.6
6	670.0	19	109.2
8	993.0	26	121.1
16	2,200.0	50	137.1
32	4,832.0	89	170.1

程门阵列以及TPUx的非结构化体系结构等。由于基础结构处理的复杂性以及所需的软件更新，管理这些分布式架构的成本正在不断增长。

在当前的架构中，48位内存地址的容量限制了非常大的内存。处理器需要物理通道来存储寻址和传输数据，如果内存共享，则需要在处理器之间建立仲裁机制来协调对共享内存的访问。随着

处理器数量(n)的增加，协调量(n^2)也将随之加倍。因此，硬件复杂性转让为软件复杂性。我们构建分布式系统，允许数据和计算分散在大量的服务器上，用于数据划分和计算以及服务器管理操作。但是，尽管数据分区和计算分区看起来似乎很简单，但同时进行这两项操作则要困难得多。此外，如果将这些决策嵌入到软件中，那么当格局发生变

化时，软件就需要进行修改。通常情况下，当数据量或其“形状”发生变化时，数据就必须被重新划分。为此，我们建立了更多的抽象层来解决这一问题，但这往往也会带来意想不到的负面影响。数据中心的操作员和软件工程师每天都面临着这些复杂的挑战。

现在，面对行业提供的廉价、普遍的计算机集群，我们是否可以重新思考这个问题，并建立一种更易处理的不同类型的计算机？与这些横向扩展系统相比，我们定义了一个新的纵向扩展系统。纵向扩展的优势在于拥有更简单的编程模型，但缺点是硬件昂贵，且灵活性较差；横向扩展具有更灵活和更具成本效益的硬件优势，但会因软件的复杂性和数据分区而产生更高的成本。幸运的是，这两种模式并不相互排斥。我们可以利用横向扩展的硬件来构建纵向扩展计算机，从而使横向扩展软件得以运行。

早期，这一问题并未完全得到解决。这里有两个例子：为DARPA和KSR-1⁵构建的Encore Ultramax4 提供了一种易于使用的编程模型：多个处理器共享高速缓存相干内存。许多人指出，与简单地使用更快捷的单处理器相比，尽管微处理器的单流性能在早期取得了一些进展，但却使这些高复杂度的项目更为不理想。而且，微处理器也碰了壁；单流性能受限于功耗和散热而停

止增长。即便单流处理器演变为多核处理器，也不足以满足新兴化需求。为了提高系统利用率，超线程应运而生，虽然给人一种类似处理器的错觉，但它实际上却在争夺通用硬件。分层结构的级别越来越多，包括每个内核的多个超线程、每个处理器的多个内核、每台服务器的多个处理器，以及每台机架的多个服务器等等。反之，多处理器多核服务器、服务器机架行和联网数据中心均由主要云提供商的服务器机架组成。尽管有些问题无法避免，但还是十分有必要考虑是否可以简化其中的一部分。此外，我们是否也可以考虑减少管理中操作系统(OS)的映像数量？

为什么现在重新讨论这个话题？

1984年，我们向DARPA提交了一项研究建议，即寻找一种分布式途径来管理相干共享内存，并使用最近推出的硬件多处理器，这些多处理器连接在总线拓扑结构中。⁴随后，该提案得到了资助，并且，我们在1989年初向DARPA展示了这台机器。

贝尔写到：

对虚拟内存来说，最重要的即是在工作集和硬件管理缓存中所体现出的局部性。虚拟内存和缓存是全

缓存架构用来“缓存”程序的活动部分，并自动利用时间和空间的局部性。

他还写到：

仅缓存（虚拟内存和多处理器缓存的自然扩展），首先允许单个数据项同时存在于多个位置。而一旦将内存页从辅助内存传送到其中一个节点，硬件和软件就会自动移动、复制和控制其他节点上的基于元素的数据流。

然而，由于这种维护相干共享内存的分布式方法十分复杂，成本也更加昂贵，于是，贝尔在1999年发表了一篇论文⁷，文中他对该方法是否会被广泛采用并与简单而直接的集群相竞争进行了一定的质疑。因此，大型程序注定要使用MPI进行强化重新编程。

2012年，我们再次来讨论这个话题，并且决定重新考虑单系统映像系统^{8,9}和分布式共享内存^{8,9}的问题。TidalScale公司成立的宗旨即是构建软件定义的、由多台横向扩展计算机组成的纵向扩展服务器。公司的目标如下：

- 使用现成的、相对便宜的商品服务器。
- 利用支持虚拟化的硬件优势作为

现代计算架构的一流部分。

- 构建能够在一组紧密耦合的网络服务器上运行的虚拟机。该虚拟机建立在一组协作的超内核上，而每个超内核则运行在单个独立的服务器上。我们该虚拟机称之为软件定义服务器（SDS；图1）。这与我们通常所理解的虚拟机¹¹相反，其中许多虚拟机同时在一台服务器上运行。

- 创建生成架构，可以随着成本的增加和时间的推移而线性动态地扩展。

- 在虚拟机中运行任一组原始分布的访客操作系统，无需任何更改（也就是说，它们是比特兼容的）。

- 在操作系统上继续运行任一应用程序，无需任何更改。

- 通过反思和机器学习来优化虚拟机，无需任何人工干预。

- 支持虚拟机尽可能借鉴未来的硬件创新。

在啤酒预算中实现香槟可扩展性

最近，有一篇论文¹²讨论了一个有趣的财务应用程序，该程序利用历史库存信息，在一张包含六百万行的表格中使用了大约6种类型的内存数据，并按

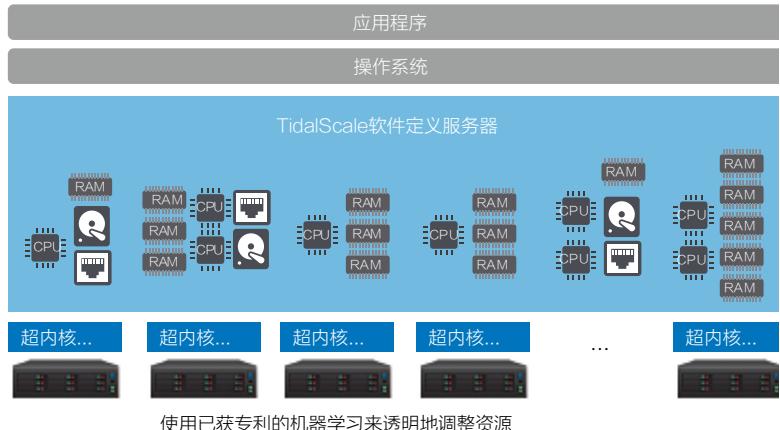


图1. TidalScale 软件定义服务器

其中一栏对行进行排序。这听起来似乎很简单，但由于内存限制，操作起来却并非易事。我们不是在内存中移动数据行，而是创建了一个指向行的指针数组，然后对指针数组进行排序。显然，这是一个算法解决方案，但是却存在着问题。如今，人们可以购买到超大内存的机器，但与集群不同的是，机器的成本与内存并不呈线性变化。得益于大相干分布式共享内存的可用性，我们可选择一个更直接的解决方案。该方案启用可随时间而发展的系统，并且有严格的线性成本配置（即将一个节点添加到一个10节点的系统中，成本只增加10%）

因此，这一方案的成本上是可以负担得起的。此外，它还增加了10%的累计可用内存和PCI带宽。

在过去的30年里，出现了许多可以

被开发利用的硬件和软件创新，包括但不限于：

- 增加内存密度；
- 更大、更多级别的缓存；

X86硬件指令集融合；
大批量成本优化的计算机系统；
高核心密度的多核处理器；
广泛运用支持对称多处理的商品操作系统和应用程序；

- 目前共存于单一主板的多个多核处理器；
- 高性能硬件支持和广泛使用，精确二进制的虚拟化软件；以及
- 基于低成本VLSI切换的低延迟网络，能够实现高带宽利用率。

表2. 三个缓存级别的
Intel Xeon处理器的典型访问率

资源	延迟时间 (纳秒)
寄存器	0
级别1	4
级别2	10
级别3 附近、非共享、未修改	40
级别3 附近、共享、未修改	65
级别3 附近、共享、修改	75
级别3 极大	100
级别3 极小	300

这些进展的一种利用途径是构建大型的可扩展相干内存，这些内存可以被大量处理器轻松利用。然而，说起来容易做起来难。现代操作系统使用具有内存视图的处理器，这些视图具有对称性和强相干性，并且具有一致的访问延迟。虽然操作系统确实具备支持应用程序利用非一致延迟的接口，但这些接口很难有效地使用。此外，随着数据和计算格局的变化，有时难以维护必要的更改。类似地，应用程序也具有相似的虚拟内存视图，在这一视图中，主内存访问被假定具有相同的一致延迟。

计算的一个基本原则是缓存可以提高性能。较大的缓存通常比较小的缓存更好，因为大缓存可以运行更多级别的缓存层次结构。Intel Xeon处理器提供三

种缓存级别，分别为级别1，级别2和级别3。访问没有缓存的主内存并不是特别快。典型的访问比率如表2所示。

多级缓存有助于满足低延迟的错觉。这在一定程度上很有效。基于硬件的缓存具有有限的固定大小，且由物理硅布局约束和成本来定义。此外，硬件需定义相干算法，相比之下，软件则要灵活得多。

对模型进行扩展时，TidalScale超内核将主板上所有的DRAM主内存建模为虚拟机的L4缓存。从某种意义上说，我们已经用分布式的L4缓存替换了所有的物理内存。但在虚拟机中运行的访客操作系统并不知道这一点，就像访客应用程序不知道从L3缓存中访问数据而不是直接访问DRAM一样。有人将此称为无需更改访客操作系统的“全缓存”设计。从上述比例中，我们可以看到，软件缓存具有强大的“魔力”，即在将数据从L4移动到寄存器的短短时间内，就已执行了几百条指令。

鉴于L1，L2和L3的限制，某些应用程序可能会表现不佳。即使是有经验的用户，在现在的“裸机”服务器上也可能会展现出性能非常差的程序。但这并不意味着我们应该抛弃缓存，因为从广泛的实践中，我们认识到，缓存在应用程序中具有一定的优势而且非常有帮助。通过深入研究非一致内存访问延迟，某些应用程序会因此而表现得更好吗？答

案是肯定的。

应用程序通常引用虚拟地址空间中的内存，本文能够呈现应用程序访问内存的过程，但却无法详尽描述目前硬件支持虚拟化的所有方式。我们将在虚拟机上运行的操作系统称为访客操作系统，访客管理的页表称为第一级页表。页表由操作系统维护。当运行应用程序的处理器引用虚拟地址时，处理器通过查询将应用程序的虚拟地址映射到物理地址的页表，将该虚拟地址转换为物理地址。

在虚拟机中，可以有多个级别的虚拟地址到物理地址的转换。就像操作系统获取应用程序的虚拟地址，并将其转换为物理地址一样，虚拟化系统获取访客的物理地址，并将该物理地址转换为一个真实的物理地址。

目前，两个级别的地址转换得到了广泛的支持，但是这一关键概念是可扩展的，因为我们可以很容易地想象到使用相似算法的堆叠虚拟机。

今天，在很大程度上，一个被称为扩容器的软件构造省却了对额外级别的虚拟机的需求。扩容器支持用户对程序和数据进行打包、传输、运行以及将其实例化，以在其他系统上获得一致的结果和类似的性能。在大数据时代，大型扩容器是首选。由于扩容器运行在操作系统上，因此，在SDS中它们无需被更改即可正常工作。

然而，这些虚拟机的机器支持有一个不太明显的优势。由于与硬件相比，软件具有更大的灵活性，管理虚拟机与物理机之间接口的机器是实现许多优化和功能增强的绝佳场所，这些优化和功能增强却不易在硬件中实现。通过这种机制，我们不必修改操作系统，因为这种操作系统可能是专有的，或者是受标准化严格控制的。虚拟化软件可以实现反思、智能、机器学习、资源跟踪、遥测统计、动员、工作集跟踪和I/O虚拟化等操作。通过信息传递，它还可以为SDS行为提供一个增强的全局视图，而无需引入任何共享硬件状态。

我们认为，很有必要引入虚拟主板的概念。在TidalScale，我们提供一个虚拟主板作为SDS的一部分。虚拟主板可以跨越多个单独的硬件服务器（图1）。与物理主板不同，它可以在用户驱动的基础上显式地增长和缩小，或者根据需要进行自动调整。

虚拟通用处理器、虚拟内存、虚拟网络和虚拟磁盘等资源可以进行迁移。虚拟中断可以远程交付。只要不违反操作系统所预期的基本硬件抽象，虚拟机就可以像物理机一样查看操作系统。因为从操作系统的角度来看，虚拟机类似于硬件，所以在兼容性测试中，虚拟机实际上像是一台物理机器。如今，多种操作系统（Centos, Red Hat, Ubuntu和FreeBSD）均已得到支持。Windows

服务器目前正在运行但尚未推出。

在虚拟机的每个节点上运行的超内核都有一个相同的实例。由于物理硬件的限制，物理处理器无法直接处理每条访客的物理地址。当访客的物理地址需要读取或写入时，必须将其转换为处理器能够访问的物理地址。

该转换是通过处理器的第二级页表来处理的。当软件引用访客的物理地址时，如果包含该地址的内存页位于生成该地址的处理器的节点上，则该地址将显示在第二级页表中。然后，自动地址转换硬件将该地址转换为访客物理地址，再转换为真实的物理地址。这种转换通常通过第一级和第二级页面表来实现，但不会降低性能。但是，如果内存地址不在第二级页表中，则处理器将中断，硬件无法将该访客地址完全转换为实际的物理地址。超内核字段中断并分析请求，类似于操作系统在复制非驻留内存页面时的操作，这一页面仅驻留在后备存储器上。该分析可能会要求从不同节点发送该页面，或者将该虚拟处理器迁移到具有该内存页的节点上。

页面读取和页面输入有着不同的处理方式。可读页面可以复制，⁶但是可写页面从其他节点的L4缓存（可能有无效副本）中删除该页面则会产生额外的开销。（实际的一系列步骤比我们在这里概述的复杂得多。）

在迁移虚拟处理器时，超内核使用

一种标准机制来采集处理器的状态快照（大约有6,400字节的数据被写入），并将信息通过专用以太网发送到所选择的目的地，在那里，信息可以恢复到另外一个物理处理器上。目前，处理器支持虚拟化的标准是保存和恢复处理器状态。程序计数器没有跟进，所以指令将被重新启动。虚拟处理器将继续运行，并与页面共存。该指令可能会生成额外的中断以访问不同的非驻留页面，但是其中的机制是相同的。虚拟处理器在迁移的过程中，会随时记录其更新的位置。然而，为了可靠性，我们从不假设完善的位置信息，因为随后处理器可能会重新迁移。

计算机集群上运行的软件定义纵向扩展计算机为什么能够正常工作？

彼得·丹宁¹³在他的一篇关于工作集的开创性论文中断言，需要页面的处理器通常会将这些页面存储在内存中而不是后备存储器上。我们对工作集的概念进行了延伸，不仅包括内存，还包括处理器、I/O、中断、存储等等。如果我们能确保所有的处理器都与其所引用的内存位于同一位置，那么这将会是一项完美的工作。当然，我们不能保证这一目标在一般情况下得以实现，但在统计的基础上，该系统有着良好的运作。同

样重要的是，一旦工作集被建立，它就不会被破坏。此外，由于每个超内核都能独立地掌控计算过程的状态，因此它将继续学习虚拟处理器的内存访问模式，并可以鉴别迁移与页面复制或页面移动。

超内核还可对其正在做的决策进行评级，并为自身提供战略反馈。例如，直接跟踪访客执行时间和摊位数量。如果它们之间的比例很高，则该条计算就会被评定为“好”；否则就会被改进。超内核与虚拟机开始自动适应每个虚拟处理器的内存访问模式。

另外，需使超内核最小化以减少开销。

摊位数量 * 平均延迟时间

每个超内核实例都保持着自己的行为模型，并且通过反思，在同类中共享并调整其行为。这些算法运行得非常频繁，使得虚拟机能够快速地适应所观察到的访问模式。

这就是机器学习的切入点。这些超内核结合在一起，以观察计算的进度，并通过协调，学习处理局部性和非局部性的最佳方法。然后它们会记住所做出的决定，并利用这些决定来帮助做出后续的决定。

性能考虑

我们致力于在操作系统和应用程

序中实现100%的二进制兼容性，从而大幅简化计算机系统格局，同时提供良好的性能和较高的可靠性。为此，我们设计了一个非常通用的系统，它可以处理多级应用程序。我们的目标是，在关于每个流行操作系统的工作方面，我们都能够做到最好。

这已经基本实现了。我们测试过的所有常见应用程序（MySQL, OracleSAP / Hana, R, Python等应用程序）都可以工作。

我们还需要考虑这些系统的性能。可能无法将应用程序的级别概括为SDS的理想替代，但同样需要注意的是，单个应用程序的执行配置文件仅依赖于程序及其数据，认识到这一点很重要。这种访问模式对于单线程应用程序来说是固定的。

实际性能可能会因一些因素而有所不同，如处理器速度、缓存容量、缓存层次级别、内存容量、内存带宽、机上通信争用、分页活动、其他进程或处理器的干扰等等，但访问模式通常是确定的（相比于多线程程序，这对单线程程序来说更是如此）。然而，遗憾的是，同级别的应用程序并不能共享相同的访问模式。

在处理器的设计过程中，设计人员通常不会考虑具体的访问模式，而是针对现有的代表性工作负载进行测试。尽管如此，我们仍然不能保证今天

的架构适合于明天的工作负载。这与虚拟服务器的概念毫无关系；相反，它是根据样本基准的要求来确定的。在TidalScale，我们采取了类似的方法。

为了实现良好的性能，我们需要尽量减少L1-L4的缓存缺失。在L4超内核管理中，如果一个节点上运行着正在请求某个内存页面的内核，而该内存页面又不在该节点上，则开销就会产生。如果是L1-L3，硬件可能会触发缓存失效，缓冲器（TLB）关闭，并且可能不得不从内存中提取新副本。超内核从远程内存而非本地主内存中获取，除此之外，其他情况与L4的情况相同。这将促使节点之间进行净工作交易，类似于处理器在英特尔公司（Intel）的QPI（快速通道互联）和SCI（串行通信接口）或高级微设备公司（Advanced Micro Devices Inc.，简称AMD）的超传输（Hyper Transport）上进行内存访问。例如，L1-L3延迟会因内存争用、套接字数量、内存条数量、内存速度和可用带宽而异。超内核的L4分布式缓存也是如此。正如我们前面所提到的，内存访问次数可能会出乎意料。

关于这类软件虚拟化性能的问题，可以归结为页面访问模式导致缓存“破损”的频率问题。而答案很简单，在L1-L3的情况下，当缓存智能化不足以预测未来访问模式时，会发生缓存丢失。这种情况通常是不可预测的。更

大的L1-L3缓存降低了缓存丢失的可能性。L4也是如此。但是SDS的L4缓存大小相对于L1-L3而言是巨大的。超内核上的L4缓存由主板上的所有内存组成。因此，L4缓存大小可能是256GB, 500GB, 1TB, 或者更多。此外，与硅材料相比，超内核的优势在于可实现更复杂的缓存一致性和缓存管理算法。

人们可能会担心这种系统会需求巨大的带宽，从而给虚拟背板带来巨大的压力，但事实并非如此。如果SDS能够同时定位处理器和内存，那么开销就根本不会产生。与机器学习相结合的智能算法显示，在专用虚拟背板上（例如10千兆以太网），一组模拟应用程序的带宽利用率为5%。

应用：... 的证明

早些时候，我们提出了关于哪些类别的应用程序特别适用于SDS，哪些不适用的问题。遗憾的是，这一问题的答案很难确定。在L1-L4不同层次结构缓存的使用中，应用程序的类别显示不够统一。如果过去的访问模式不能很好地预测未来的访问模式，那么缓存丢失的概率可能很高。

L4缓存命中数据主导性能。观察到两个性能较差的L4缓存的例子：

1. 在计算基因组测序的具体操作

中，当在具有256GB实际内存节点的软件定义服务器上运行时，每次更新都会重新编译大型数组（250GB），从而使预测变得困难。因此，很可能不需要将该部分的数组局部到节点。当内存大小从256GB增加到320GB时，问题即可得到解决。最初的程序一直在尽力节省内存，但需要注意的是，内存的容量是可调的。一旦实现，该程序的新版本就能解决这一问题。

2. 第二个实例还需要做一些难以预测的事情：在一个具有大量多线程（多处理器）的应用程序中，我们观察到，Python运行时库中出现了很多中央共享锁争用，从而导致包含页频繁地在节点之间移动。然而，当在单个主板上访问相同的锁时，不同处理器的多个内核之间便会发生相同的争用模式。在这些处理器的L1-L3缓存中，很难评判哪个缓存是包含该锁的最佳缓存，因此，锁会在缓存之间不断移动，从而导致性能下降。而当修改这一应用程序以减少不必要的共享时，性能就会提高一个数量级。

当我们找到一个用例来攻击“内存悬崖”时，系统就能表现得特别好。该用例即是一个程序及其数据可以从完

全内存驻留到足够的可用内存中受益。所需的内存占用量不需要比可用内存¹⁴大得多，也能够促使系统颠簸以及操作系统开始分页。这是完全正常的，而且经常发生。该应用程序将会继续工作，但表现不佳。同样不尽人意的是，无论后备存储的速度有多快，它都比DRAM（Dynamic Random Access Memor；动态随机存储器）慢几个数量级。即使是市场上最快的SSDs（固态硬盘），通常也比DRAM慢三个数量级。更快的后备存储延迟也远远不及DRAM延迟。

考虑以下的案例研究。¹⁵ 在一个包含100,000,000行和100列的数据库和128 GB内存的服务器上，一个应用程序需花费七个半小时才能完成MySQL三次查询作业。由于所需的内存容量大于服务器的内存容量，从而导致分页频繁产生。而当用两个96GB节点替换运行SDS的192GB物理内存时，完全相同的一组数据查询只需大约7分钟就能完成，从而将性能提升了60倍。并且分页也被完全消除。

在另一项研究中，我们与一家大型金融机构合作，对客户的行为进行了分析模拟。这并不是传统意义上的性能实验：在这之前，客户从未有过如此大规模的模拟，所以没有前后对比。在3.5TB内存的SDS上，该模型在五台物理服务器上透明地运行。由于客户可以在内存中执行模拟，因此能够以1:1的

粒度水平运行模拟，并执行包括粒度分析、灵敏度分析和模型优化在内的三部分分析。实质上，这个实验消除了传统的局限性，并且在不改变模型、工具或操作环境的情况下进行了探索性分析。

兼容性和可靠性

兼容性非常明确。由于超内核类似操作系统的硬件，因此兼容性应为100%。

此外，客户有自身的可靠性特征，例如RAID（磁盘阵列；*Redundant Array of Independent Disks*）。

虽然一般性主题可能超出了本文的范围，但是由于前面提到的新虚拟化级别，SDS提供了提高可靠性的新途径。

当满足以下常见条件时，我们即可获得可靠性：

- › 早期指标表明可能发生硬件故障。
- › 存在机制可在不可恢复的故障发生之前继续准确运行。
- › 在早期指标和不可恢复的故障之间有足够时间来调整行为。

考虑到资源均可移动，因此可以利用先前配置的热备份机器。不久的将

来，如果由于一些原因而引发故障，例如，软纠错代码错误、服务器温度上升或高于正常的网络异常等，则可以通过动态地向群集添加一个额外的热备用节点来处理，并通知所有节点有关待定节点的故障，以防任何虚拟处理器被迁移到此节点，并使故障节点尽早退出虚拟处理器以及活动中使用的内存页。这些机制都是作为资源流动的副产品而存在的。例如，如果处理器无法访问远程节点上的页面，我们通常会将处理器迁移到该节点，或者将页面迁移到发出请求的处理器。如果远程节点出现故障，我们会将该页面迁移到其他带有副本的节点。换句话说，我们的目标不是试图接近单个服务器的可靠性，而是让SDS超过它，因为我们可以在不重启客户系统的情况下实时替换包含资源的节点。

我们提出了一个创新的现代分布式相干共享内存系统，几乎所有的计算机集群均已安装。SDS架构的优势很明显。特别是对于大型扩容器，单系统映像提供了更简单的编程模型和数据中心服务器管理模型。从软件和支持的角度来看，横向扩展系统更为复杂，但是比传统的HPC纵向扩展计算机要便宜。纵向扩展系统的应用程序和操作则更简单。通过使用相同高性价比硬件，可以实现横向扩展系统的硬件成本优势和灵活性效率，以及纵向扩展系统的固有简单性。

我们展示了

- › 横向扩展系统和纵向扩展系统都可以有效地安装SDS；
- › 通过在硬件之上操作系统之下引入一层软件，我们可以提高自动优化的可能性，并创建一个将机器学习引入计算机系统的场所，而无需创建新硬件；
- › 总而言之，可以在不采用新硬件设计的情况下增加内存和I/O带宽；以及
- › 通过调动虚拟处理器和内存资源，我们最大限度地减少了互连带宽的需求。SDSs安装的应用程序可以采用一种更传统、更简单的方式来编写。SDSs通常减少了对显式数据分区和分区管理的需求，这直接转化为快速可靠地构建大型、多样化应用程序的能力。■

致谢

感谢彼得·克里斯蒂、查尔斯·莱文以及本文推荐人对本文早期版本的评论和指导。

参考文献

1. C. Levine, personal communication.
2. J. Dean and S. Ghemawat,

关于作者

“MapReduce: Simplified Data Processing on Large Clusters,” Proc. 6th Symp. Operating Systems Design & Implementation (OSDI 04), 2004, p. 10.

3. D. Cutting, “Welcome to Apache Hadoop!,” Apache Hadoop, 18 Nov. 2017; adoop.apache.org.

4. G. Bell. et al., “The Encore Continuum: A Complete Distributed Work-station-Multiprocessor Computing Environment,” Proc. Nat'l Computer Conf. (NCC 85), 1985, pp. 147–155.

5. KSR-1 Technical Summary, Kendall Square Research, 1992.

6. G. Bell, “Scalable Parallel Computers: Alternatives, Issues, and Challenges,” Int'l J. Parallel Programming, vol. 22, no. 1, 1994, pp. 3–46.

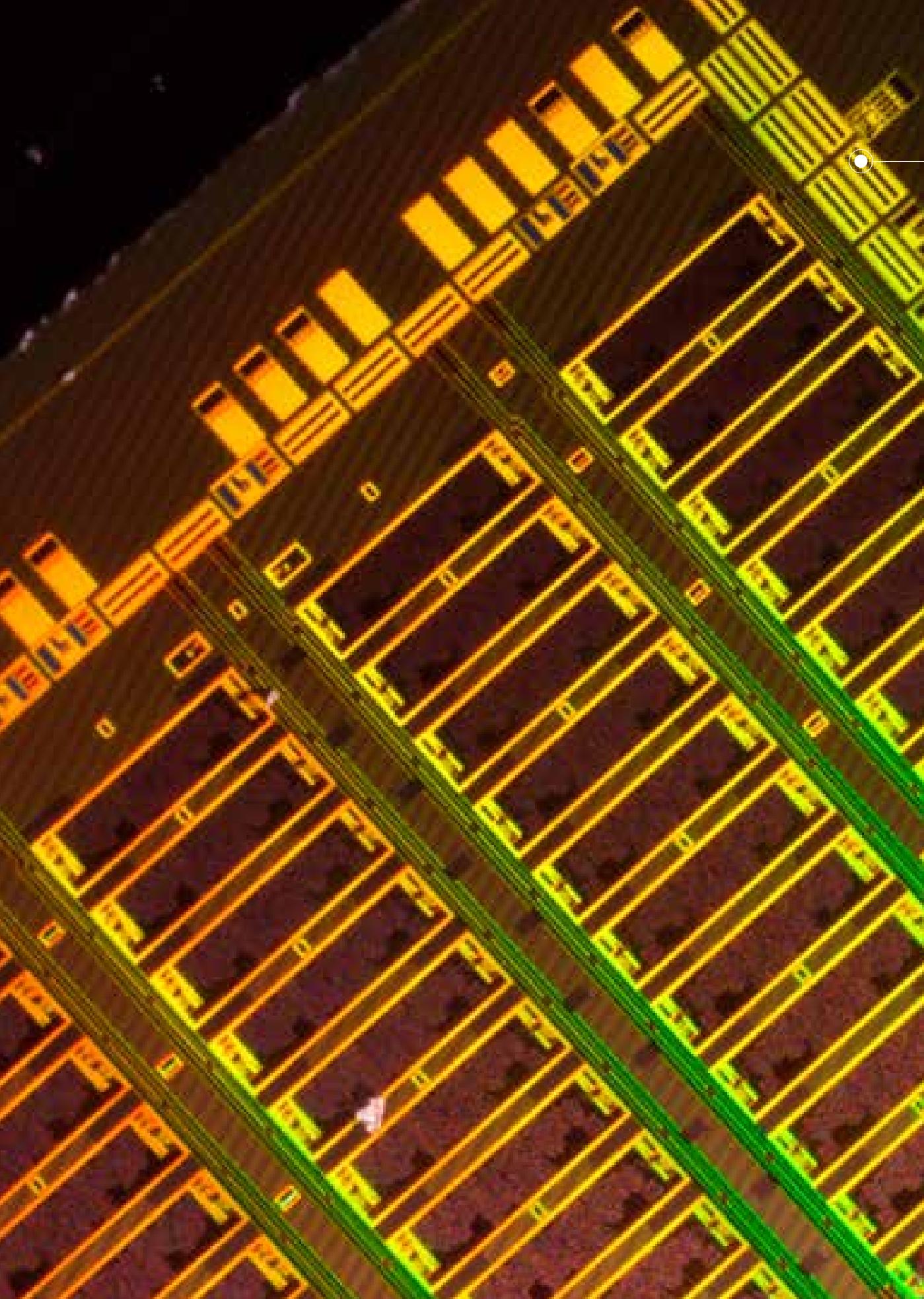
7. G. Bell and C. Van Ingen, “DSM Perspective: Another Point of View,” Proceedings of the IEEE, vol. 87, no 3., 1999, pp. 412–417.

8. R. White, “The Single System Image Feature Delivers Greater Flexibility and Resilience,” IBM Systems Mag., May 2013; [www.ibmsystemsmag.com/mainframe/administrator/Virtualization/ssi_feature_zvm](http://ibmsystemsmag.com/mainframe/administrator/Virtualization/ssi_feature_zvm)

戈登·贝尔是微软公司的一位研究员(退休);曾任数字设备公司研发部主管;AMACAD、IEEE、NAE 和 NAS 的研究员和成员;计算机历史博物馆创始理事。他凭借在计算机、生命历史记录和新企业等方面的贡献获得多项荣誉,其中包括CMU和WPI博士(荣誉),IEEE冯诺伊曼奖章以及1991年国家计算机技术奖。联系方式: gordonbell.azurewebsites.net。

艾克·纳西是TidalScale公司的创始人、董事长和首席技术官;加州大学圣克鲁兹分校计算机科学系副教授;计算机历史博物馆的创始理事;IEEE的资深会员;获得石溪大学计算机科学博士学位。他的研究兴趣包括编程语言、操作系统、系统架构以及计算机科学史和数学史。联系方式: ike.nassi@tidalscale.com。

- feature_zvm. Memory,” Computer, vol. 50, no. 8, 2017, pp 46–53.13.
- 9. R. Buyya, “Architecture Alternatives for Scalable Single System Image Clusters,” Proc. Conf. High Performance Computing on Hewlett-Packard Systems (HiPer 99), 1999; www.buyya.com/papers/ssiArch.html.
- 10. J.K. Ousterhout et al, “The Case for RAMClouds: Scalable High-Performance Storage Entirely in DRAM,” SigOPS, Operating System Rev., vol. 43, no. 4, 2009, pp. 92–105.
- 11. M. Rosenblum, “The Reincarnation of Virtual Machines” ACMQueue, vol. 2, no. 5, 2004, pp. 34–40.
- 12. I. Nassi, “Scaling the Computer to the Problem: Application Programming with Unlimited
- 13. P.J. Denning, “The Working Set Model for Program Behavior,” Comm. ACM, vol. 11, no. 5, 1968, pp. 323–333.
- 14. A. Aho. P.J. Denning, and G. Ullman, “Principles of Optimal Page Replacement,” J. ACM, vol. 18, no. 1, 1971, pp. 80–93.
- 15. I. Nassi, “Advances in Virtualization in Support of In-Memory Big Data Applications,” Proc. Int'l Workshop High Performance Transaction Systems (HPTS 15), 2015; www.hpts.ws/papers/2015/tidalscale.pdf.



图片新闻

电光芯片

这支由麻省理工学院、伯克利大学和波士顿大学的 18 位研究人员组成的科研团队报告了另一个新突破：他们开发了一项新技术，在集成的基础上在芯片上分别装配光学元件和电子元件，这意味着现代的晶体管技术能更好地被加以利用。

来源：波士顿大学工程学院 / 麻省理工学院 / Amir Atabaki

论文：*Electronics and photonics united. Nature, 2018; 556 (7701): 316 DOI: 10.1038/d41586-018-04443-3*

自然语言处理为什么这么难? ——专访两位顶级 NLP 大师

文 | 管心宇、林宗辉

本文原载于《麻省理工科技评论》中文版和 DeepTech 深科技微信公众号。更多信息请关注 DeepTech 深科技微信公众号 (*id: mit-tr*)。

2017 年，人工智能领域迎来了转折之年。在这一年传统的计算机视觉和语音识别都达到了新的高度，也在性能方面趋于饱和。在产业方面，商汤等专注于计算机视觉的公司也获得了长足发展。其中，商汤在 2017 年 11 月新获 15 亿人民币的投资，隐隐有从独角兽变成巨头的趋势。

不过，人工智能另一个相关领域自然语言处理似乎没有达到这种高度。在技术方面，这一领域的技术准确率远远没有达到计算机视觉和语音识别的水平，技术产品（比如个人助手）经常被人讽刺只能用来调戏，缺少实际价值。在创业公司方面，自然语言处理领域也没有产生像商汤、云从这样的小巨头。

然而，这种现状恰恰说明自然语言处理的难度和重要性。人工智能分为两个重要的部分“感知”和“认知”。其中，计算机视觉和语音识别等领域属于感知部分，而自然语言处理属于认知部分的重要内容。对一个“智能”而言，仅仅感知当然不够，理解和消化内容的认知能力才是真正意义上的核心。

所以，人工智能行业的很多从业者认为，自然语言处理是下一个风口和主战场。克服这个领域的难题将对整个人工智能甚至科技领域产生深刻的变革。

微软创始人比尔·盖茨就曾经表示“语言理解是人工智

能领域皇冠上的明珠”。微软全球执行副总裁沈向洋也在 2017 年底的公开演讲时说：“懂语言者得天下”，“下一个十年，人工智能的突破在自然语言的理解”，“人工智能对人类影响最为深刻的就是自然语言方面”。

学术界的看法也大体相同。2014 年，深度学习的开山祖师 Geoff Hinton 说：“我认为下一个五年最值得关注的领域是如何理解视频和文字。” Yann Lecun 在 2015 年 Facebook 巴黎人工智能实验室的启动仪式上时表示：“深度学习的下一个进步点在自然语言理解方面，目标是让机器不仅能理解单个单词，还能理解整个句子和段落。机器学习领域的另一位开创者之一 Michael Jordan 也曾表示：“如果我有 10 亿美元，我会建立一个专门研究自然语言处理的项目。”¹

既然自然语言处理这么重要，为什么像计算机视觉和语音识别那样，产生大量可以媲美人类的应用，也没有产生小型巨头创业公司？这个领域是否有什么独特的技术困难？我们需要如何克服这些技术困难？自然语言处理到什么时候才能在产业界真正落地？

带着这些问题，我们采访了两位自然语言处理领域的领军人物：宾夕法尼亚大学教授 Dan Roth 和微软亚洲研究院副院长周明。对与自然语言处理领域的从业者来说，这两位的名声如雷贯耳。Dan Roth 教授致力于通过机器学习和推理的方法帮助机器理解自然语言，也是 AAAS、ACL、AAAI



如果配备人工智能的无人机把平民错认成士兵而屠杀了他们，是否有人对这一战争罪行负责？他们，是否有人对这一战争罪行负责？

和 ACM 的会士，曾在多个重要会议上担任程序主席一职。而周明博士作为中国自然语言处理最顶尖的学者之一，目前担任微软亚洲研究院副院长、国际计算语言学协会（ACL）候任主席、中国计算机学会理事、中文信息技术专委会主任、术语工作委员会主任、中国中文信息学会常务理事等多个职务，还是哈尔滨工业大学、天津大学、南开大学、山东大学等多所学校博士导师。

前一段时间，周明带领的微软亚洲研究院团队在阅读理解国际评测 SQuAD 比赛中，超过了人类评测员的理解水平。我们的第一个问题就从自然语言处理的评测开始。

SQuAD 的全称是斯坦福问答数据集（Stanford Question Answering Dataset），是由斯坦福大学自然语言处理实验室开发的数据集和比赛。SQuAD 的数据来自 Wikipedia 的文章。数据标注人员去掉了文章里的一些单词，并让参赛队伍利用模型重新填空，借以检测模型对文章的理解程度。

2017 年，微软亚洲研究院、阿里巴巴和哈工大·讯飞联合实验室分别宣布，自己开发的模型对文章的理解已经超过了人类标注员的水平，引起了很大的反响和争议。对此，我们也分别询问了 Dan Roth 和周明的看法。

类似 SQuAD 这样的竞赛是否可以通过一些技巧刷分？类似的竞赛对行业的意义有多大？我们需要什么样的数据集和比赛？

Dan Roth 认为，这种竞赛对于提高技术基础建设会有一定的贡献，但是长期来看，对推动科学的研究和发展方面并没有太多价值。举例来说，如果用相同数据集来进行竞争，持续个一年或两年，比赛本身就会完全失去其意义。主要原因就是，如果人们只是为了竞赛的数据来进行训练，而不是我们所普遍关心的那些真正应该被解决的问题，那么最后我们就不会看到真正的技术进展，而只剩为了拿到比赛名次而发展的各种小技巧。

相较而言，周明对这类比赛的看法要正面的多。他认为，SQuAD 的一些设置可以有效防止刷分。例如，数据集很大，而且测试集也没有公布。总的来说，斯坦福的 SQuAD 可以说是自然发言处理领域一个里程碑式的创新。人们原来做阅读理解，都是泛泛的去做，从来都不知道到底做到什么水平。但是，现在斯坦福做了一个大规模的，不太容易通过微调改进性能（fine tune）的数据集。实际上很有力地来促进这个领域。

SQuAD 确实存在问题。但正确的态度应该是巧妙地设计测试集的新难点，针对这些难点的一条一条地把阅读理解所涉及到的技术难点逐个攻关。久而久之我们整体的阅读理解能力就会循环往复地上升，最后就真的逼近人的平均水平。

例如，SQuAD 没有涉及太多的推理能力，我们就可以

做一个专门测试推理能力的测试集。推理还可以分几级：简单推理可以根据上文就能推理，复杂推理可以根据全文推理，更复杂的推理甚至必须要用到背景和领域知识。如果能把这样一层一层的难度做出来的话，成功就有一半了。

周明认为，未来研究的成功有两个重要的因素，一个是模型，一个是可以用来评测竞赛的数据集。

和一般的机器学习、人工智能领域以及机器视觉这样的方向相比，自然语言处理领域是否有存在属于自己的独特挑战，有什么解决方案？

两位学者都认为，自然语言处理远比计算机视觉复杂。

Dan Roth 认为，计算机视觉基本上就是物体探测。虽然计算机视觉应用很多，但基本上核心算法都离不开物体探测这个方向，背后使用的逻辑也相当一致。此外，由于计算机视觉的技术成熟度已经达到商用化的标准，所以我们可以看到很多不同的公司百花齐放。但自然语言处理的情况完全不同。不同场景、不同语言，甚至不同专业所需要用到的自然语言处理层次都不同，所以自然语言处理远比计算机视觉复杂，且目前的应用还是相当少，要为了这些少数应用而开发自己的算法并不划算。

周明对此做了更进一步地阐述。他表示，语音识别和图像识别都是一输入一输出，问题非常干净。比如输入一个图片，要判断里面有没花或者草，直接判断就行了。这些方向中间没有多轮，不需要交互，一般不太依赖于知识图谱和常识，即使用也被证明没有什么效果。

但自然语言处理有三个重要的区别，让它变得很难：

第一，自然语言是多轮的，一个句子不能孤立地看，要么有上下文，要么有前后轮对话。目前的深度学习技术，在建模多轮和上下文的时候，难度远远超过了一输入一输出的问题。所以语音识别搞的好人和图像识别搞好的人，不一定能搞好自然语言。

第二，自然语言除了多轮特征之外，它还涉及到了背景知识和常识知识，这个也目前大家不知道怎么建模，都没有搞明白。

第三，自然语言处理要面对个性化问题。同样一句话，



不同的人用不同的说法和不同的表达，图像一般没有这么多变化。这种个性化多样化的问题非常难以解决。

周明强调，因为人工智能包括感知智能（比如图像识别、语言识别和手势识别等）和认知智能（主要是语言理解知识和推理），而语言在认知智能起到最核心的作用。所以可以很自信地说，如果我们把这些问题都解决了，人工智能最难的部分就基本上要解决了。

那怎么解决这些问题呢？

周明认为，虽然不保证可以改进技术，但有三个值得尝试的方向：

第一，上下文的建模需要建立大规模的数据集。比如多轮对话和上下文理解。数据标注的时候要注意前后文。没有这样的数据，很难取得突破。

第二，强化学习很重要。我们需要根据用户的反馈倒推模型并做参数修正，使模型更加优化。现在强化学习刚刚开

始用在自然语言领域，性能并不稳定，但在未来很有机会。

第三，要引入常识和专业知识，并把这些知识构建好。这样就能更加精准地回答问题。没有人能证明现在常识知识用在语言问答和搜索中的作用有多大。所以，我们需要一个测试集来检验结果。这个测试集要专门测上下文和常识，可以让我们要不停用新模型（比如强化学习或者知识图谱）去试错，来看系统性能能不能提升。

为什么自然语言处理领域没有产生非常大的创业公司？

周明表示这个问题很值得研究。总的来说，他认为这是因为自然语言处理的技术难度太大，和应用场景太复杂。一

如果配备人工智能的无人机把平民错认成士兵而屠杀了他们，是否有人对这一战争罪行负责？他们，是否有人对这一战争罪行负责？

个公司的成立和发展是由需求驱动。图像识别的需求巨大，例如安防和身份认证的应用场景很多，到处都有摄像头，谁也看不过来。所以，安防领域一直期待着一种技术，只要达到一个阈值，立刻就能用了，恰好这两年深度学习把计算机视觉水平升到了那个阈值。此外，就像上面的回答所说，图像识别问题更干净，再加上有现成且巨大的场景。所以，只要技术有一点突破，场景自然结合，公司一下子就做起来。

特别纯粹的自然语言应用（不包括搜索），主要是就是机器翻译。机器翻译长期有需求，但没有安防和身份认证的需求那么大。而且，机器翻译水平一直不到位。即使到今天，机器也只能翻译简单句子，很难翻译有背景的复杂句子。另外，自然语言处理的应用太依赖于 UI 了。图像识别基本不需要 UI，直接在系统内部集成一些技术就行。包括微软在内的所有公司做翻译软件，如果 UI 做得不行，用户体验不行，根本就没有人用。

技术产业化最重要的是商业模式，也就是怎么让技术挣

钱。图像识别公司的挣钱模式已经成立了，直接就找安防就能收钱。翻译收钱就难多了。所以自然语言是从研究到技术到落地到商业化，面临一系列的挑战。目前的现状是，自然语言处理技术更多的是作为公司内部技术，比如内部的商业情报或人机接口功能。但这不代表我们未来找不到这样的渠道。

Dan Roth 也对自然语言处理的应用场景复杂性进行了补充。他表示，在各种专业应用中，必须要选择正确的自然语言模型，没有任何单一模型可以解决自然语言领域中所遇到的所有问题，自然语言处理没有一个可以解决所有问题的魔术盒子存在，你必须要把所有相关的知识库放进盒子里，选择对的算法，并且针对性的处理特定问题，那么这个盒子最后才有作用。这种现状加大了技术落地的难度，

类似的是，计算机视觉发展到最后已经不是只有单纯识别图像或者是物体，而是要能够做到预测这些物体的本身的一个动作，比如说在桌子上放了瓶水，然后把瓶子往外推，一个先进的计算机视觉系统就能够判断出瓶子最终的动作轨迹可能是掉到桌子下。然而自然语言处理技术达不到这种水平，它无法进行预测。它只能就现有的文字组合、数据库来判断所有文字应该有的意义。计算机视觉的物体识别准确度已经可以达到将近百分之百，而自然语言目前的阅读准确度也不过将近 9 成，而这也是目前自然语言处理商用化的最大阻碍。如果要用到专业领域，那么现有的精准度明显不足。

即使我们不考虑基础研究的困难，就算是现有的自然语言处理的基础研究结果，似乎也没有很好地转化，很多产品在发布会上的效果往往和实际使用的效果完全不同。

周明认为，目前自然语言处理产品出现的问题，很多时候无关技术，而是在产品设计和 UI 方面做得不够好。在做阅读理解和机器翻译研究的时候，我们往往有一个固定的评测集，以及 F- 分数和精确度这样的评测方法。但这些不代表用户的体验，即使在实验中分数达到 100% 也是这样。技术是独立于产品应用方向发展的。做产品的人应用技术的时候要运用之妙存乎一心。他们要考虑，无论是 78% 的技术，还是 88% 或者 98% 的技术，要怎么到产品里，才能让用户体验最好。

如果配备人工智能的无人机把平民错认成士兵而屠杀了他们，是否有人对这一战争罪行负责？他们，是否有人对这一战争罪行负责？

用户体验要考虑什么呢？最重要的是用户界面。因为系统很难达到 100% 的正确，所以要考虑用户怎么操作，怎么容错，让他们接受有缺陷的结果。比如说搜索引擎返回多个搜索结果的设计，其实非常巧妙。因为谁都知道搜索达不到那么好的水准，但当返回多个结果后，用户不抱怨搜索引擎不行，反而认为搜索引擎的结果扩大了他的思路，把好事变成坏事。这种巧妙的用户界面设计和用户体验设计，是我们做自然语言处理的人要好好考虑的。你的系统和研究厉害，不代表能把用户体验做好。要从用户的角度看，如何把你的技术，融入到其他所有的相关的场景中，解决用户的实际问题。

还是以机器翻译为例，在实验室里，所有话都实验了很多遍，也没有什么噪声，效果肯定很好。但做产品的时候要考虑语音、环境噪声、背景噪声、远场识别、专有名词，以及口音等等。如果做不好，会导致翻译结果一塌糊涂。

但是，背景噪声怎么来解决呢？首先好好调整 UI 要好，要解决语音识别的一些问题，然后可能要解决简单的多轮对话的问题，要对用户口音做自动调整，如果用户觉得翻译不好，要有方便的方式和他们互动。这样就能让用户觉得，这个系统虽然没有那么好，但是他也给我解决了很多问题了。这一块就是要考虑设计水平的能力了。

所以这个不是科技要解决的问题，这个是产品设计要解决的问题。

除了这些难点和问题，自然语言处理技术在研究和应用方面，可以在今年或未来几年出现较大的进展？

就这个问题，Dan Roth 和周明一致认为自然语言处理

能在垂直领域很快得到应用。Dan Roth 表示，利用知识库，未来自然语言处理应用会协助企业把专业知识转成特定的自然语言处理模型。利用这些模型，自然语言处理技术就能成为很好的工具，影响更深层次的人类生活。

周明也同意这样的观点。他表示，垂直领域有一定的保护门槛（比如有一些不公开的数据），导致大公司无法直接进入。在这样的领域可以做一些知识图谱的探索，还可以针对本领域特点，做一些特殊的优化和有的放矢的研究，而不是使用通用的自然语言技术。这样就可能会产生一个专业的知识图谱，以及基于专用图谱之上的自然语言理解的技术。最后提升整个领域的生产力。

此外，周明还认为，神经网络机器翻译、阅读理解、聊天对话，和创作辅助这四个应用在今年和明年就会有很多地方普及，相关的应用场景包括搜索引擎、个人助手、语音助手、机器翻译，还有个人制作音乐，个人制作新闻、撰写网络小说、问答系统等等。另外一个重要的应用是机器客服。没人愿意看产品手册，但如果让计算机读一遍产品手

册，你就能问它任何产品的问题，就能在客服、售后服务这些领域产生很好的应用。智能客服可以帮助提高效率，节省人员。系统也可以按照座席收费，有商业模式。

对成熟公司来说，首先搜索引擎还有进步空间。如果搜索引擎有阅读理解的能力，在手机屏幕上返回结果特别精准，会产生很大的竞争优势。第二，现在信息流非常重要。例如今日头条背后的推荐技术需要理解文本，理解用户，然后匹配他们。如果我们的自然语言处理能力提高了以后，推荐水平就提高了。

对创业公司来说，第一个机会是机器翻译，但是要把用户体验和商业模式做好。第二个机会是客服。最后一个开发垂直行业的自然语言处理技术。■

参考文献

1. Computational Linguistics and Deep Learning
Christopher D. Manning https://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00239

万众一心 点石成晶

半导体产业大势论坛

北京·四季酒店

2018年6月2日

技术发展

科技应用

产业趋势

芯片产业驱动的是整体科技产业发展的路径，如果不能自主选择未来的道路，受制于人无疑是必然的结果。DeepTech深科与全球数十位半导体产业精英和创新青年将共同进行一场头脑风暴，我们将提出过去长期被忽视的短板关键，揭示长期坚持自有核心技术将建立的长板可能，探讨如何厚植深耕中国半导体的核心实力，以及如何在世界竞争中取得最具关键影响力的战略位置。

在历史机遇的风口浪尖上，中国半导体需要落地，需要共生，更要能在关键时刻，成为新一代科技产业发展最坚实的后盾。

部分参会嘉宾（已确认）



曾学忠

紫光集团全球执行副总裁
兼紫光展锐CEO

吴雄昂

arm全球执行副总裁
兼大中华区总裁

卢超群

全球半导体联盟(GSA)
前理事长

罗镇球

台积电南京有限公司
总经理

余定陆

应用材料集团副总裁



赵鸿飞

中科创新董事长



彭启煌

明导科技亚太区总裁



潘健成

群联电子董事长



米磊

中科创新创始合伙人
兼首席CEO

彭适辰

中经合北京办公室
董事总经理

周斌

NovuMind中国区总裁



沈亦晨

Lightelligence
联合创始人兼CEO

王星泽

华中科技大学教授兼
合刃科技创始人、CEO

朱明杰

氪信创始人兼CEO



王旭

中国科学院
生态环境研究中心
副研究员

姚颂

深鉴科技
创始人兼CEO

韩璧丞

BrainCo 及 BrainRobotics
创始人兼 CEO

陈成猛

中国科学院
山西煤炭化学研究所
副研究员

黄畅

地平线联合创始人

扫描二维码申请免费门票
(需审核)



图片新闻

神经灰尘

2016 年，伯克利大学的研究人员发明了微小的神经植入设备 StimDust，他们称之为“神经灰尘”。这款设备可植入人体，无需外连电力就能监测、刺激人体内的神经、肌肉和器官。现在，研究人员进一步缩小了设备，目前它只有 6.5 立方毫米那么大。

来源：伯克利大学 /Rikky Muller





创客革命

文 | 斯库特斯·威利斯 (Scooter Willis) , TechGarage
译 | 韩晶晶

3D 打印机和其他创新技术，让我们可以更加迅速、简单地制造多种类型的原型样品和产品，这场创客革命正在显著地影响着科研和工业。



我们正经历着一场被称为创客革命的当代工业革命。有新产品的构想后，为了获得样品，就得经历漫长的设计和制造周期，这样的日子已经一去不复返了。你不再需要与工具和模具制造商见面，也不用在开始生产之前制作大量小部件。

创客革命开始于 2009 年，当时 MakerBot Industries 推出了低成本打印套件。而当你组装并测试了了自己 的 RepRap 3D 打印机，MakerBot 还会鼓励你为朋友打印出另一台 3D 打印机所需的部件。2013 年，商业 3D

打印公司 Stratasys 以 6.04 亿美元收购了 MakerBot。

最近，针对创客革命的快速原型样品制造技术又有了若干进展，在这里我想分享一下我们在 TechGarage 创客空间使用的产品，和我们迫不及待想要购买的产品。

启动 3D 打印

几年前，我们经历了一段黑暗时期，在那段时间内，3D 打印机并没有得到多少应用，因为失败的几率很高。根据 Make:magazine (makezine.

com / comparison / 3dprinters) 的产品评论，我们的 TechGarage 创客空间决定投资购买 Lulzbot Taz 6 3D 打印机。我们现在有六台这种打印机。这几台打印机一直在忙碌工作着，事实证明它们很可靠。我们买的 Prusa i3 打印机效果也不错。

1 月份在拉斯维加斯举办的国际消费电子展 (CES) 上，我见到了 XYZprinting 的达芬奇彩色 3D 打印机 (www.xyzprinting.com/en-US/product/da-vinci-color)，该打印机在印刷过程中会使用无色聚乳酸 (PLA) 丝线和彩色墨水。通常，3D 打印机使用丝线来获得颜色。在 CES 上，XYZprinting 展示了一系列具有全彩打

印功能的 3D 打印机模型。

就像从黑白复印机到彩色复印机一样，我预计未来将出现高度可靠和价格合理的彩色 3D 打印机，可以根据需要打印产品。

人人可用的激光切割机

3D 打印机的缺点之一是生产大型物体所需的打印时间。两年前，我们收到了一笔赠款，让我们得以购买 60 瓦的激光切割机。这是变革性的，有了它，我们可以切割 1/4 英寸厚的丙烯酸树脂或尺寸高达 18" x 24" 的木材。切割速度快，精度高，可快速制造各种零件。而需要赠款才能购买，也表明商用 60W 激光切割机并不便宜；它们通常售价是 15000 美元到 30000 美元。但从我的角度来看，这是我们过去 18 个月来的最佳投资之一，因为我们曾用它来切削 2 000 个不同的零件。

幸运的是，Glowforge 公司通过 Kickstarter 推出了面向本土市场，售价 3995 美元的 40W 激光切割机。它有许多重要的功能，而且易于使用。专业版售价 6995 美元。集成的摄像头可通过检测刚切割过的图案的位置来确定从何处开始切割。摄像头还可识别材料上的条形码，从而自动调节设置，以针对不同材料，例如木材，丙烯酸，皮革，橡胶，纸板，织物等。

打折销售的 5 轴数控机床



图 1. 第二版的 Pocket NC，这种台式计算机数控（CNC）设备可以生产铝、木材、塑料和其他材料的零件。

3D 打印机制造的塑料部件没有铝材料的那种耐用性。虽然有些使用具有金属或木材外观和质地的特殊 3D 打印材料的例子，但用这些材料制成的部件依然不如铝制部件那么耐用。

计算机数字控制（CNC）是制造铝制零件的标准工艺。它从一块坚固的铝块开始，将金属切割成型。这与 3D 打印不同，3D 打印是通过添加材料来制造零件的。业余爱好者水平的数控机床售价大概 5000 美元，而工业制造用的机床需要 50000 美元。

数控机床要使用 G 代码，G 代码是通过刀具路径控制雕刻刀头在三维空间中移动的一系列命令。过去，编写 G 代码是很困难，因为需要高度专业化的软件来生成雕刻 3D 对象所必需的刀具路径。但最近 Autodesk 推出了 Fusion 360，这是一款 3D 计算机辅助设计、制造和工程工具，只需最少量的自定义代码编写就能生成有效的 G 代码。此外，在 Fusion 360 上，

数控机床配置文件可以模拟刀具路径，以便在机床上进行测试之前验证 G 代码。

如果你想要以更少的步骤制造复杂零件，那么你需要一台 5 轴数控机床。但在过去，如果你需要询问这种机器的价格，那你可能买不起。不过，在 2015 年，Pocket NC 发起了一个 Kickstarter 众筹活动，以 5000 美元的价格销售台式 5 轴数控机床，该机床可生产铝、木材、塑料和其他材料的零件。它也适用于 Fusion 360。Pocket NC 众筹到了 35.5 万美元，超过了该公司启动该项目的 70000 美元目标。我们最近购买了 Pocket NC 第二版的机床，如图 1 所示，该机床现在就安置在我们的工作台上，正在等待首次投入使用。

廉价数控机床

最近引起我注意的一个 Kickst-

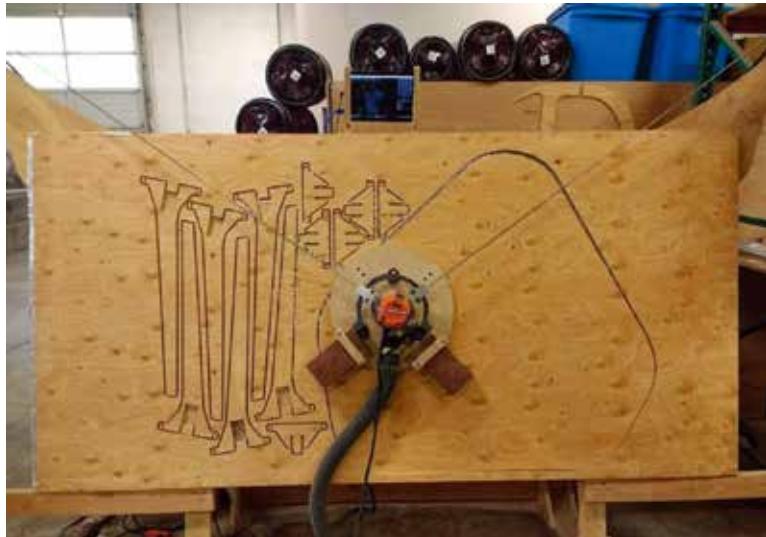


图 2. Maslow CNC，这种计算机数控 (CNC) 设备是一种低成本、大尺寸的产品，用于把胶合板切割为复杂的性状。



图 3. Shaper Origin，这种创新的电动工具可以将胶合板切成各种形状。它使用集成的摄像头通过带有斑点图案的胶带跟踪其在木材上的位置。这有助于操作员沿着预订路径切割。该工具大大降低了用户进行精确切割所需的技术水平。

arter 项目是 Maslow CNC，其目标是开发一个售价 500 美元、4'x8' 大小、由电脑控制的雕刻机床，可以在完整胶合板上切割出错综复杂的图样 (www.kickstarter.com/projects/1830738289/maslow-cnc-a-500-open-source-4-by-8-foot-cnc-machi)。如图 2 所示，该设备在垂直支架的左上角和右上角使用两个电机，通过两个链条连接到木材雕刻机。

/ 1830738289 / maslow -cnc-a-500-open-source-4-by-8-foot -cnc-machi)。如图 2 所示，该设备在垂直支架的左上角和右上角使用两个电机，通过两个链条连接到木材雕刻机。

电机沿着两条链条旋转，可以将雕刻机移动到胶合板的不同部分。使用运行 G 代码的 Arduino 微控制器，用户可以切出精确的形状。

该项目大获成功，募集到了 314547 美元。对于第 2 版 Maslow CNC，创客社区提供了代码修改意见和创新的雕刻机底座设计，可帮助补偿切割时的定位错误。

这是一个很好的例子，从中我们可以看到，如果创客们如果要想降低成本并为自己使用的设备创建新的平台的话，应该做些什么。

电动工具的未来

真正喜欢简单但聪明的解决方案的人们，肯定会爱上 Shaper Origin (shapertools.com)，如图 3 所示。这是一种新型的高科技电动工具，可与传统的雕刻机配合使用，但它也集成了电脑屏幕和摄像头。这大大降低了精确切割所要求的技术水平。

第一步是将多条带有点状图案的胶带粘贴到你要加工的胶合板上。集成在雕刻机上的摄像利用交变跟踪其在电路板上的位置。你可以下载需要切割的图案，雕刻机集成的计算机屏幕会告诉操作人员雕刻机该向哪个方向移动，从而按照虚拟图案进行切割。该工具的调整功能可以左右移动雕刻机刀头，以补偿操作员的追踪虚拟图案时的差错。

他们上传的网络视频展示了这套设备的精确度 (www.youtube.com/

watch?v=DekAjA OIVvQ），其售价为 2399 美元。

办公室可用的桌上水流切割机

我最期待的产品是售价 4995 美元的桌上水流切割机 Wazer，如图 4 所示。它可以精确切割任何材料，例如钢、铝、钛、碳纤维和陶瓷等。预计发货时间是 2018 年 12 月。

该公司的创始人是宾夕法尼亚大学工程与应用科学学院的四位工程学学生，他们在 2012 年在学校结识，意识到市场上对用于零件制造的低成本水流切割机有很大的需求。典型的商用水流切割机至少要花费 10 万美元，而且对于桌面来说太大了。Wazer 团队开始制作原型机，并在 2015 年在 Kickstarter 发起了众筹项目。该项目筹集到了 1331936 美元，表明客户对该产品的需求非常旺盛。

给大孩子玩的乐高

RobotZone 的 Actobotics (www.servocity.com/ / [actobotics](http://actobotics.com)) 是一个提供了多种类型零件的精密系统，用户只需使用一把通用扳手即可搭建出众多机械系统。

Modern Robotics 公司的 goBILDA (modernroboticsinc.com/gobilda) 是一种新的基于度量的系统，可以将机器制造提高到一个新的水平。它提供了构建组件，用户可以将它们



图 4. Wazer，这款台式水流切个装置可精确切割包括钢材在内的多种材料。

组合在一起制成机器人。

我在 TechGarage 开设的课后社区机器人课程中，Actobotics 和 goBILDA 的零件是关键。无需钻削或切割金属，中学生和高中生靠它们就制作出了令人惊讶的机器人。这样更安全并且可以制作出质量更高的机器人。

用户来说，Onshape 是免费的，而专业版则要每月花费 125 美元。

创客经济正如火如荼，对于那些有产品创意，需要开发可用于 Kickstarter 项目的原型样品，来测试客户兴趣的人来说，激动人心的时刻正要到来。随着本地的创客空间充满了最新和最好的原型产品快速制造设备，创客革命将帮助推动这个快速创新的时代。■

云端的 CAD

当使用上述任何方法时，我们首先必须运用计算机辅助设计 (CAD)。这意味着我们也需要 CAD 软件和建模方面的创新。正在改变 CAD 的明星公司是 Onshape，它推出的 Onshape 系统通过“软件即服务”模式工作，并且注重实时协作、参数化建模以及快速引入新功能。它使用 WebGL (一种无需插件就可在浏览器中生成交互式图形的 JavaScript API) 进行 CAD 渲染。对于公开自己的模型和设计的

斯库特斯·威利斯是 TechGarage 的创始人，TechGarage 是佛罗里达州棕榈滩县初高中生的集中式机器人工作空间。威利斯还是阿维拉癌症研究所的计算生物学主任。请通过 willishf@gmail.com 与他联系。

人工智能和攻击 / 防御平衡

文 | 布鲁斯·施奈尔 (Bruce Schneier)，哈佛大学 (Harvard University)
译 | 郑米兰



人工智能技术有可能改变互联网上攻击相对于防御的长期优势。这和以下因素有关：人与计算机的相对优劣势、这些优劣势在互联网安全领域如何相互作用、人工智能技术可能会在哪些方面带来改变。

互联网安全任务可以被分成两个集合：人类擅长的任务、计算机擅长的任务。传统上，计算机在速度、规模和范围上都更胜一筹。它们可在毫秒间发起攻击，感染数百万台计算机。它们可以扫描计算机代码以查找特定类型的漏洞，扫描数据包以识别特定种类的攻击。

人类则擅长思考和推理。他们可以查看数据，区分真正的攻击和虚假警报，理解正在发生的攻击并做出响应。他们能在系统中发现新型漏洞。人具有创造性和适应性，而且能够理解语境。

至少到目前为止，计算机还不擅长人类擅长的事。它们不具有创造性和适应性，无法理解上下文。它们可能因此做出不合理的行为。

人的操作速度慢，厌倦重复性任务。他们在大数据分析上很糟糕；他们使用认知捷径；他们一次只能记住少量信息。因为这些特点，他们也可能表现得不理性。

人工智能将令计算机能接管来自人类的互联网安全任务，然后更快、更大规模地完成这些任务。以下是人工智能可能拥有的能力：

- 发现系统中的新漏洞——更重要的是新类型的漏洞。这既包括通过攻击来利用漏洞，也包括通过防御来修补漏洞，然后自动利用或修补漏洞。
- 就对手的行动做出反应和调整——还是从攻击和防御两个方面。这包括对对手的行为以及这些行为在整个攻击和环境中的含义做出推理。
- 从单个事件中概括出教训，在各种系统和网络中推广，并运用这些教训来提高其他地方的攻击和防御效率。
- 从大型数据集中识别战略和战术趋势，利用这些趋势来调整攻击和防御战术。

这个清单还没有写完。我不认为有谁能够准确预测出人工智能技术的能力。但是，看看人类现在在做什么并想象未来人工智能也会做同样的事（但它们将以计算机的速度、规模和范围去做），并非不合理。

攻击和防御都将从人工智能技术

中受益，但我相信人工智能有能力让天平更多地向防御倾斜。将会有更好的人工智能攻击和防御技术。但是，目前防御处于劣势，而这正是因为人这个组成部分所导致的。当今的攻击就是利用计算机和人的相对优势来打击计算机和人的相对弱点。当计算机进入传统上由人运作的领域，就会改变这个方程式的平衡。

罗伊·阿马拉 (Roy Amara) 有一句名言：我们高估了新技术的短期影响，却低估了它们的长期影响。人工智能很难预测，所以我推测的很多细节可能是错误的，而人工智能很可能带来我们无法预见的新的不对称。但人工智能是我迄今所见最有可能让防御能与进攻势均力敌的技术。对于互联网安全而言，这将改变一切。■

布鲁斯·施奈尔 (Bruce Schneier) 是哈佛大学 Berkman Klein 互联网与社会中心的安全技术专家和研究员。他也是 IBM Resilient 的首席技术官和 IBM Security 的特别顾问。可通过 www.schneier.com 与他联系。

中国中文信息学会、语言与知识计算专业委员会 (CCKS)



CCKS·2018

开放领域的中文问答

2018.05.01–2018.07.20

北京大学计算机技术研究所



扫码了解详情

机器人士兵的良知

文 | 查尔斯·戴 (Charles Day) , 《今日物理》(Physics Today)
译 | 郑米兰



年11月，我参加了芯片制造商英伟达(Nvidia)在华盛顿特区举行的GPU技术大会。它探讨的重点是人工智能。你可能和我一样会奇怪人工智能与图形处理单元（延伸来说也就是视频游戏）有什么关系。答案是：在游戏《光环3》(Halo 3)中追踪从士官长的喷火枪中发射的燃料流所需的计算和运行深度学习（人工智能最流行的变体之一）所需的计算在性质上是相似的。

英伟达的开发者营销副总裁格雷格·埃斯蒂斯(Greg Estes)在开幕主题演讲中列举的人工智能技术令人印象深刻：人工智能驾驶的汽车、人工智能创作的古典音乐、人工智能口译，以及人工智能诊断医学图像等。但在一个一些人可能视之为人工智能顶峰的领域，他着墨甚少。那就是有感知力的机器人。埃斯蒂斯解释了其中一个障碍：“训练一辆车来避开障碍物要比训练一个机器人捡起它头一次看到的东西简单多了。”

虽然技术上的障碍尚未克服，这并不妨碍特斯拉的首席执行官伊隆·马斯克思考人工智能将带来的伦理问题。他是致信联合国特定常规武器公约的人之一。这封信对机器人士兵、无人机及其他高科技武器的危险性提出警

告，因为这些武器配备的人工智能实际上将做出攸关生死的决定。如果配备人工智能的无人机把平民错认成士兵而屠杀了他们，是否有人对这一战争罪行负责？

战事问责的问题经历了长久的历史演变。一战后，一些德国军人因战争罪行受审。被告之一卡尔纽曼(Karl

人员应被追究责任。

如果配备了武器的无人机不是基于程序化的指令而是根据它通过深度学习获得的标准来开火，我认为算法的作者仍然是有罪责的。如果人工智能某天要达到人类的意识水平，那么它大概也将拥有自己的良知和随之而来的道德责任。而在这两者之间的

如果配备人工智能的无人机把平民错认成士兵而屠杀了他们，是否有人对这一战争罪行负责？

Neumann)是一艘U潜艇的指挥官，这艘潜艇发射的鱼雷击沉了英国医疗船“多佛城堡”。德国最高法院以“所有文明人都知道下属执行上级命令这一原则”为由，宣布其无罪。到了二战时，这种辩护被两项附带条件削弱了：首先，如果士兵接到的命令需要他做出道德抉择来执行，那么他仍然有罪。其次，如果命令本身是非法的，那么对它们的执行也是非法的。

我们不难看出士兵接收的命令和今天管制自主无人机的程序之间的相似之处。两者都是指令。事实上，程序更具约束力，因为无人机毫无选择。在这种情况下，当发生导致战争罪行的错误时，程序员和授权执行程序的

某处——比如拥有类似狗的意识水平——那么人工智能和它的人类所有者将需要分担责任。■

查尔斯·戴是《今日物理》(Physics Today)杂志的主编。本文所述观点只代表他个人的观点，不代表该杂志或其出版商美国物理研究所(American Institute of Physics)的观点。



搜索你的工作机会

IEEE Computer Society 招聘可以帮你轻松找到IT、软件开发、计算机工程、研发、编程、架构、云计算、咨询、数据库很多其他计算机相关领域的新工作。

新功能：找出那些建议或要求拥有IEEE CS CSDA或CSDP认证的工作！



点击www.computer.org/jobs，
从全世界的雇主那里搜索技术工作岗位和实习机会。

<http://www.computer.org/jobs>

IEEE  computer society | JOBS

IEEE计算机协会是AIP Career Network的合作伙伴。其他合作伙伴包括《今日物理》杂志(Physics Today)，美国医学物理协会(American Association of Physicists in Medicine)，美国物理教师协会(American Association of Physics Teachers)，美国物理学会(American Physical Society)，AVS科学和技术学会(AVS Science and Technology)，物理学生协会(Society of Physics Students)和Sigma Pi Sigma。



保持联系。

无论你在哪里，都能紧随IEEE计算机协会的脚步。

在Twitter、Facebook、Linkedin和YouTube上关注我们。



@ComputerSociety, @ComputingNow



facebook.com/IEEEComputerSociety
facebook.com/ComputingNow



IEEE Computer Society, Computing Now



youtube.com/ieeecomputersociety