

计算科学评论

环球科学

WWW.COMPUTER.ORG

2018年第4期

人工智能 开疆拓域

Emoji 表情会昙花一现吗? P46

KDD 2018 参会观感 P70



合作机构



IEEE
computer
society



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

市场 行业分析 职场 竞赛
创业 学习 科技新闻

找工
教育

互联网金融

芯片

大数据

纳米架构

软件架构

人机交互

多媒体

普适计算 MEMS

合作
工作 研究综述
申请
进修

微信名：计算人 微信号：jisuanren



Copyright

版权

主管单位 Authorities 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 Sponsor

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

出版单位 Publication

《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, 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. Thiruvathikal
Loyola University Chicago,
gkt@cs.luc.edu

2018 IEEE计算机协会主席

Hironori Katashara
Waseda University,
katashara@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

06

导读

人工智能开疆拓域

计算有诸多前沿领域，各有其独特的深远影响。在人工智能这个领域发生的变化将提醒我们，人和计算机的竞争环境并不平等。而要让这项技术为我们人类的生活服务，既意味着绝佳的机遇，也意味着棘手的挑战。在本专题中，《计算科学评论》的主编将为大家介绍人工智能领域正在显现的一些新兴变革以及它们对人的影响。

2018年第4期

10

迈向拟人机器学习

未来的智能机器将更便于人类使用，也更人性化，同时它们的处理能力和自动化水平也将大幅提高。这些都将增强我们人类的能力。拟人机器学习 (Anthropomorphic machine learning) 是人工智能 (AI) 和数据科学的新兴方向。这一革命性的转变为下一代机器学习提供了类似人类的能力，更有可能推动技术和日常生活领域里的各种突破。

20

通向人类可理解、可解释的人工智能

近年计算能力的提高加之可获得数据量的迅速增长，让我们重燃了对人工 智能 (AI) 理论及应用的兴趣。然而，如果各行各业以及政府要能信心十足 地推出 AI，则要考虑用户希望通过“可解释 AI”(XAI) 系统提高 AI 的透明 度。本文作者介绍了 XAI 的概念，概述了需要进一步探索的领域，如 2 型模 糊逻辑系统，以确保这些系统能被非专业用户充分理解和分析。

2018 年第 4 期

目录



30

人工智能的情感时代

在科幻小说中，未来的人工智能（AI）技术通常拥有高水平的情商，甚至具有同情心。但我们目前处于什么样的阶段？本文作者深入探讨了人工智能情感智能（AEI），提出 AI 要达到新的情感智能时代所需的三大情感领域——识别、生成和增强。

40

AI 和区块链： 联手改变世界

AI 和区块链是最具颠覆性的技术，它们将深刻重塑我们的生活、工作和人际关系。作者总结了已有的技术，并探讨了两者结合的未来前景，试图回答这个问题：智能、去中心化和安全系统将会如何影响我们的社会。

趋势

- 46 Emoji 表情：普遍通用语
还是一时流行语？

- 52 量子计算机的扩展

大师小传

- 56 万尼瓦尔·布什：虚拟计算与科学管理的大师

教育

- 60 让有趣的学习搞定智能手机一代

会议

- 64 第二届亚太地区网络研讨会
(APNet 2018) 全程纪实

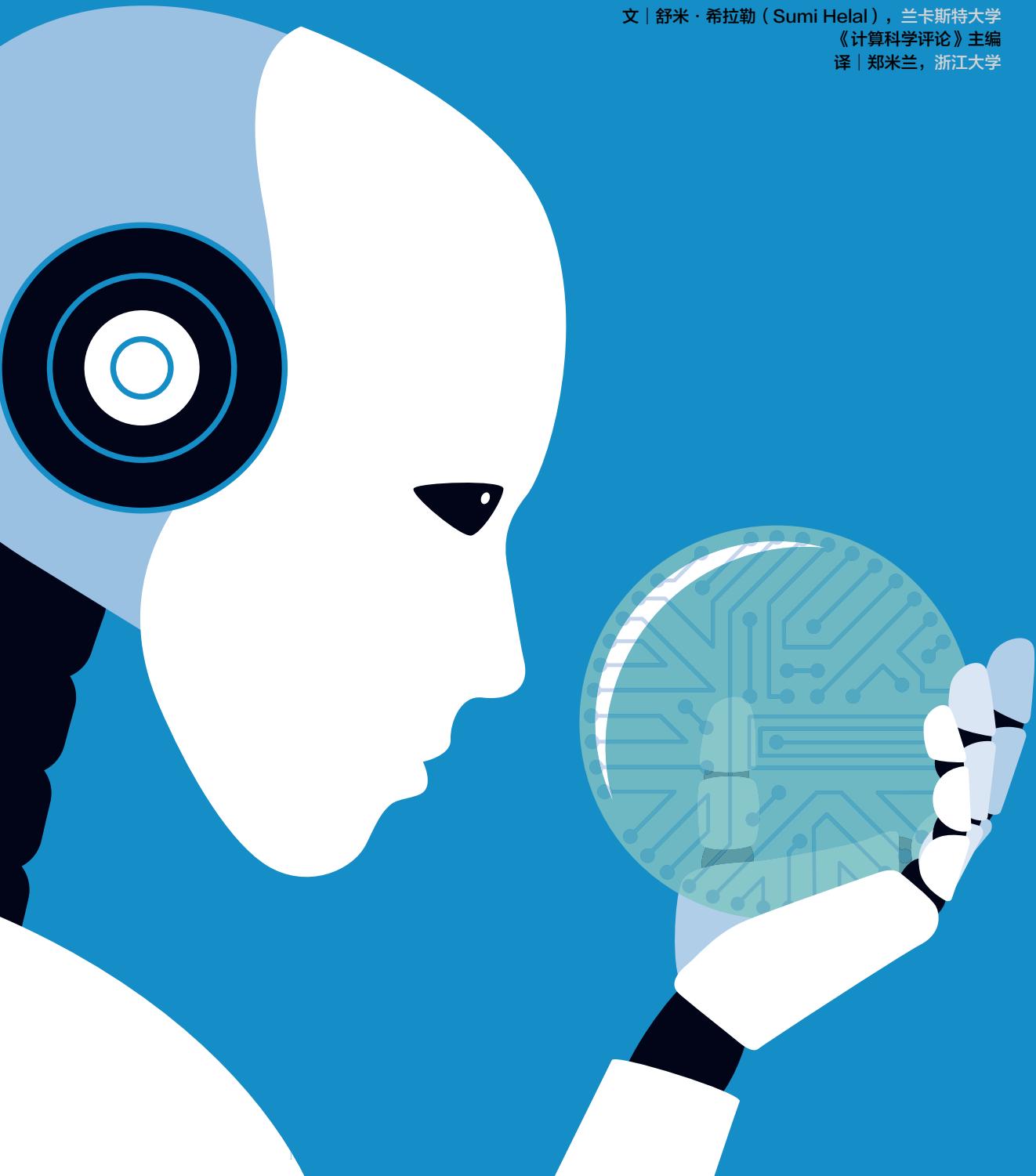
- 70 KDD 2018 参会观感

专栏

- 74 自然语言处理中的语言模型预训练方法

人工智能开疆拓域

文 | 舒米·希拉勒（Sumi Helal），兰卡斯特大学
《计算科学评论》主编
译 | 郑米兰，浙江大学



计算有诸多前沿领域，各有其独特的深远影响。在人工智能这个领域发生的变化将提醒我们，人和计算机的竞争环境并不平等。而要让这项技术为我们人类的生活服务，既意味着绝佳的机遇，也意味着棘手的挑战。在本专题中，《计算科学评论》的主编将为大家介绍人工智能领域正在显现的一些新兴变革以及它们对人的影响。



人工智能这个术语诞生于 60 多年前，但早在这之前很久，人类就已开始设想要让自己喜爱的机器带上我们人的特征或具有人的思维。而尽管一直以来都不乏人工智能如何威胁人类的暗黑想象（从《2001：太空漫游》中的场景，到机器人沃森在《危险边缘》（Jeopardy）节目中赢了人脑这种小得多的威胁），但在 2018 年，我们似乎更急切地要接受和拥抱人工智能技术，因为它们已经清楚展现了帮助我们的能力：智能手机、在线购物、搜索和查找我们喜欢的东西、智能手表、健康、旅行，以及企业如何与客户建立联系和了解客户等等，不胜枚举。

然而，人工智能若要真正载着所有人腾飞，还面临诸多挑战。首先，训练人工智能算法需要海量数据，这是一项昂贵又耗时的苛刻要求。此外，我们为训练机器而生成和处理的数据越多，安全和隐私泄露的风险就越高。然而，最大的挑战或许还在于让各方都了解人工智能如何运作和做出决策，包括非专

业用户、专家、律师、政策制定者及媒体等。未来某天，当两辆自动驾驶汽车发生罕见事故时，可解释的人工智能或许能让地方执法人员和汽车保险公司了解究竟哪一辆车犯错。毕竟，并不是所有的自动驾驶交通工具都会是一模一样和平等的。

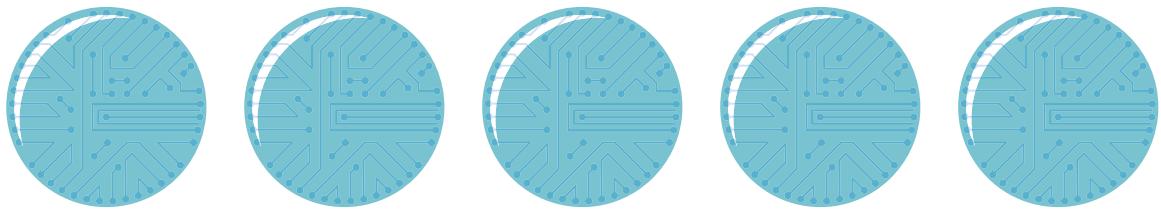
研究人员正在迎击这些挑战。他们要提高我们理解人工智能如何运作的能力、我们对人工智能的信任和信心，并且寻找让人工智能从我们这里接手麻烦问题的机会。为让读者了解人工智能的研究正把我们带向何处，我很高兴地向大家推荐本期专题，一起来听一听计算领域的几位专家对人工智能未来机遇和挑战的见解与探索。

本期内容

在《迈向拟人机器学习》一文中，普拉门·P·安格洛夫（Plamen P. Angelov）（Plamen P. Angelov）和 Xiaowei Gu 介绍了拟人机器学习的概念，这个雄心勃勃的课题是下一代人工

智能的新兴方向。拟人机器学习是指机器可以像人类那样学习。作者介绍了当前主流机器学习技术（如深度学习神经网络）的局限性，包括临时嵌入式结构、严格和数据密集的训练流程、无法在实时和不断演变的系统中应用，以及无法应对各种不确定性。然而目前媒体都在不加选择地炒作人工智能的盛大前景，却无视它面临的严峻挑战。作者提出的拟人机器学习方法基于最近引入的基于深度规则的系统，能从极少的训练数据开始逐步建立起模型，随着系统的发展演变不断学习。这与今天实践的机器学习方法大相径庭。此外，这种学习模型用“原型”来描述或注释学到的观察结果。这是另一个巨大的差异，因为原型可以解释系统为何做出某些决定，以及学习一件事而非另一件事的理由——就像人类在学习时那样。作者用一个基于深度规则的图像分类器演示了这种新方法。

在《通向人类可理解、可解释的人工智能》中，哈尼·哈格拉斯（Hani Hagras）着眼于目前主流机器学习算



法（如深度学习）的决策缺乏透明度的问题。无法解释、不可审计的“黑箱”学习已经引发了政府和消费者的关切，阻碍了它们的广泛采用。作者介绍了现有的三种实现可解释人工智能（XAI）的方法，它们源自美国国防高级研究计划局（DARPA）的相关项目。作者而后提出了第四种方法，它基于模糊逻辑系统，试图通过近似而非精确或量化的连续过程模仿人类思维。具体而言，近期在模糊规则系统（FRBS）上取得的革命性进步能够生成更短也更少的条件（if-then）规则（与输入数量无关），保持了系统清晰度和可解释性，与传统的FRBS形成对比。与《迈向拟人机器学习》一文相似，哈格拉斯强调，为使人工智能被广泛接受，XAI是具有最高优先级的研究课题、一个必须获得满意解决方案的重大挑战。

在《人工情感智能的时代》中，达格玛·舒勒（Dagmar Schuller）和比扬·舒勒（Björn Schuller）阐述了人工智能在情感智力领域的应用。情感智力是识别情绪、产生和适应情绪、应用情绪信息以优化目标达成和解决问题的能力。作者展示了早期人工情感智能（AEI）如何利用机器学习算法（包括隐马尔可夫模型和神经网络等），根据面部表情、语音声学、口语语言和其他形式的变化来识别人类情感。作者提出，深度学习目前要克服传统机器学习技

术在情感识别性能上的巨大差异。作者还展示了目前人工智能在基于规则系统生成情感方面扮演更小的角色，以及在情感增强、继而在规划、推理和目标达成上发挥的更大作用。此外，他们描述了另一种情感增强方式，将情感激励的原则嵌入机器学习算法中。这是一个迷人的概念。作者举例解释了强化学习、认知和抽象学习中的情感增强。文章描绘了AEI如何能实现更自然、高效的人机交互的广阔视野，并指出要实现这一愿景仍缺失的部分和下一步需开展的研究。

在《AI和区块链：联手改变世界》一文中，Thang N. Dinh和My T. Thai关注了AI与新兴区块链技术之间的意想不到的互利关系。正如作者所说，一方面，区块链是一种可以支持AI及其未来发展的技术。作者指出AI由大数据驱动，但由于隐私问题，在某些应用领域（例如，个人健康数据）难以获得此类数据。扩大用户的隐私权和版权（数据所有权）可以用收入鼓励数据共享。此

类方案将要求区块链跟踪个人数据的使用，以实现基于使用的支付方式。作者还指出，AI让消费者（不是大型云计算和云分析公司）能够利用暂时闲置的设备（如移动电话、游戏设备、机顶盒和其他新兴的IoT设备）共同成为区块链（分布式分析平台）来运行AI算法并执行分析。此外，作者（XAI应该是其他东西与AI结合的意思吧，建议不直接翻译出来）还讨论了区块链如何增加AI算法决策的可解释性。另一方面，作者论证了AI如何增强区块链技术的实用性。区块链技术可以让用户控制自己的数据，例如他们在社交网络中的好恶。但仅用区块链产生的匿名化会丧失个性化内容。运行在区块链上的AI则可用安全的方式重新获得个性化。本文预测了AI和区块链之间的这种整合将带来重大创新和革命性的社会效益。

本期杂志的内容既有趣又有启发性。我向所有为内容做出贡献的作者致谢，希望每位读者都能欣赏本期文章，从中受益。■

关于作者

苏米·希拉勒（Sumi Helal）是兰卡斯特大学的数字健康教授。他是IEEE会士，也是IEEE计算机专委会的董事会成员。联系方式：s.helal@lancaster.ac.uk。



2018 开放学术 数据挖掘大赛

奖金: ￥160,000

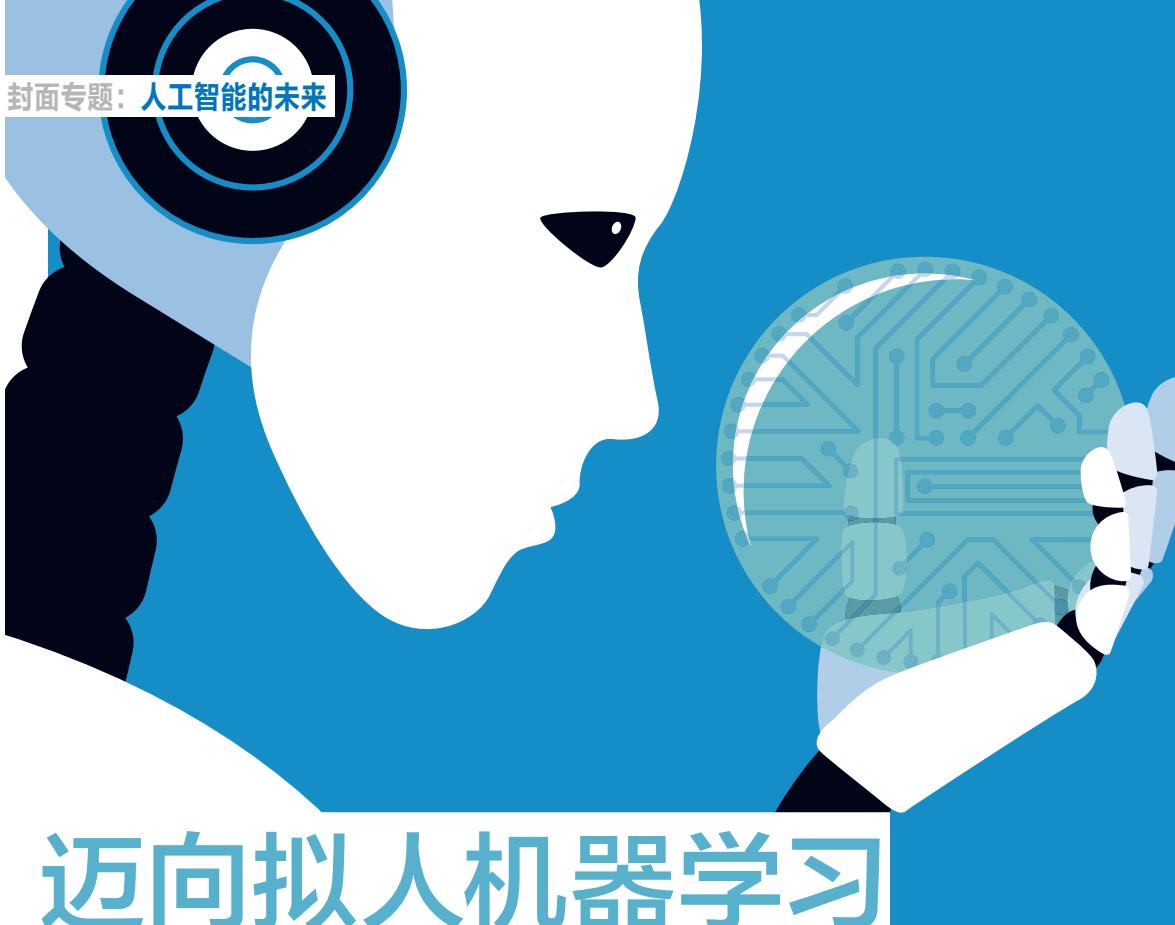
2018.10.29.
—
2018.12.1.

参赛入口



主办

中国人工智能学会、中国工程院知识中心、清华大学 AMiner



迈向拟人机器学习

文 | 普拉门 · P · 安格洛夫 (Plamen P. Angelov) , 兰卡斯特大学

Xiaowei Gu, 兰卡斯特大学

译 | 郑米兰, 浙江大学

未来的智能机器将更便于人类使用, 也更人性化, 同时它们的处理能力和自动化水平也将大幅提高。这些都将增强我们人类的能力。拟人机器学习 (*Anthropomorphic machine learning*) 是人工智能(AI)和数据科学的新兴方向。这一革命性的转变为下一代机器学习提供了类似人类的能力, 更有可能推动技术和日常生活领域里的各种突破。

今

天的技术发展有可能彻底改变我们的行业、社会和日常生活。这些进步主要是基于数据的量、复杂度和来源(数据流)的指数级增长。此外, 由于这些技术极大地影响了我们应对具有丰富数据量的环境的能力, 输入数据日益变成非线性、非固定、多模态/异构流, 混合了各种物理变量和信号以及图像、视频和文字。

现在, 既然摩尔定律在硬件发展中已近失效(因为微型化已达到纳米级这一物理极限), 我们就需要转而在用算法

处理数据这一面上取得突破。我们再次目睹了类似摩尔定律的那种指数式增长, 但现在, 这种增长是有关我们生成和处理的数据量而不是硬件容量和电路集成。人类将不得不应对这种“数字肥胖症”带来的所有积极和消极后果。不过, 我们的文章将侧重于这一革命过程的技术和理论方面。

我们当前的现实与几十年乃至几个世纪前概率论、统计学和统计学习取得基本成果的时代截然不同。它需要采用新方法, 也带来了一些新的特定的挑战, 却也提供了重要的新机遇。

机器学习、统计学习和深度学习

传统的机器学习模型¹的内部结构主要由在学习过程开始时就做出的假设预先确定，比如它们强烈依赖假设的先验数据生成模型。²传统统计机器学习模型的可解释性通常低于所谓的“第一原理”模型，后者本身通常基于代表质量、能量平衡或物理定律近似值的微分方程。然而，传统统计机器学习模型的可解释性水平通常又高于某些计算智能模型，例如人工神经网络（ANN）。而“第一原理”模型的可解释性和透明度更接近于另一种形式的计算智能：基于模糊规则（FRB）的系统。此外，在FRB系统与高斯概率或贝叶斯模型的混合物之间存在一些相似和对立。¹

最成功的分类器的可解释性非常低⁴，比如支持向量机（SVM）³，以及卷积神经网络（CNN）和递归神经网络（RNN）等深度学习神经网络（DLNN）^{4,5}。它们的内部结构对于人类用户而言不清晰也不提供信息，而且自身也不理解正在处理的问题（例如，它们无法清楚地解释做出某些决策的原因，并且不允许探索替代方案）。这对于快速普及的DLNN来说是一个尤其突出的问题。这些DLNN网络能从图像中提取高级抽象，在图像处理中已经取得优异的成绩。

果。但它们缺乏透明度，所用的数亿个权重的价值无法以人类可理解的分析方式与所处理的问题相关联。⁴ DLNN也没有明确的内部模型语义结构，其隐藏层数和许多其他参数都是临时决定的。正如谷歌和Drive.ai在近期所言，不透明和无法在不确定的情境中工作是应用深度学习最具挑战性的两大问题。⁷一项错误的决定可能危及生命。这不是个小问题，需要对DLNN为何做出特定的决策进行分析。若能开发出一个能识别未知情境、从而能应对不确定性的透明的系统，则可以直接帮助解决目前只能通过深度学习来处理的几个已知而未解的问题。

无法应对数据的演变特性（它们具有固定的结构和设置，如类别数量），并且它们也无法“从零开始”学习。

> DLNN容易过度拟合。

> DLNN无法处理不确定性。它们不仅在与训练数据明显不同的推理数据上表现不佳，而且它们自身也意识不到这一点（这意味着实际上不可能分析错误和失败的原因）；它们很容易被“骗”，因此即使是在面对无法识别的图像时也会输出高可信度分数。⁸

机遇和替代方案

深度学习的局限和挑战

尽管主流DLNN获得了成功（报告结果已与人类能力相当或优于人类⁴），引发大量宣传（包括其商业版本），媒体对它们的兴趣也日益增加，但仍存在许多未解决的问题和不足，如下所述。⁶⁻⁸

> DLNN的内部架构缺乏透明度（有许多临时决策和参数，如层数、神经元和参数值）。

> DLNN的训练过程耗费大量数据、时间和计算资源，这使得它们无法展开实时训练和适应，也因而

机器学习和统计学习，以及计算智能和深度学习（深度学习是最先进也最复杂的ANN形式）都与人类从数据（如图像）中学习的方式大不相同。统计学习基于经典的概率论和迭代优化，因其具有坚实的数学基础，一般而言这些学习算法在数据量趋于无穷大时能够保证性能，并且所有数据都来自已知分布。尽管如此，在数据生成模型上强加的假定随机性质和已知分布太过强大而不切实际，无法在真实情境中保持成立。²此外，机器学习算法的预定义参数通常需要对问题具有一定的先验知识，而在现实情境中常常不具备这种知识。

如果我们将人类从数据（包括但不限于图像）中学习的方式与现有的机器学习方法进行比较，我们可以注意到一些基本差异，例如：

- > 人可以轻松地从单个或少数例子中学习；
- > 人投入终身学习，而不是“训练并纠正我们的认识法则”；
- > 我们的认识法则（分类器）是发展演变的（新的规则和/或类别会出现，或者一些规则和/或类别在我们成长时被合并；规则和/或类别的数量并非一经确定就永久固定）⁹；
- > 人可以“实时”添加新的类别和规则（我们不会喊“停”然后重新训练）⁹；
- > 人能解释为何会识别特定的图像（指定某个类别），如发色、耳朵和鼻子（面部识别），或者汽车和房子的细节等等；
- > 人会协作学习（彼此交换聚合数据并从中学习）。⁹

我们看到以拟人方式开发下一代机器学习的巨大机会。在我看来，未来的智能机器和机器智能将更加便于人类使用，也更与人类相像，同时大幅提高其处理量和自动化水平，从而增强了我们人类的能力。

因此，研究所面对的挑战是开发和实现准确、透明、适应性强、自学习和计算效率高的机器学习方法（包括图像分类器、预测器、控制器、异常检测器、推荐系统等），它们要能识别出未知情境。这些下一代机器学习方法不应应用到手工设计、限制性先验假设、用户特定参数或问题特定参数。也就是说，

于单个或少数样本就做出关联、推理、分类或异常检测。因此，从单个或极少数训练样本（即“从零开始”）构建模型的能力就是一种类似人类（拟人化）的特点。

另一个类似人类的特征是能够解释所做的决定。当我们认出一只鸟时，我们永远不会说，“这是一只鹦鹉，但

尽管主流 DLNN 获得了成功，引发大量宣传，媒体对它们的兴趣也日益增加，但仍存在许多未解决的问题和不足。

它们必须高度自治。然而，它们也必须是对人类友好的、透明的、类似人类的（拟人的）。

我不知道为什么。”我们会说，“这只鸟看起来非常像鹦鹉，因为它的头、喙、爪子、翅膀、羽毛的颜色，还有它发出的声音。”然而，DLNN无法做到这一点。如果能够识别并使用与新观察结果相关联的原型，则能够且容易对决策过程做出解释。这是当前的DLNN、SVM和其他最先进的方法无法提供的一种拟人特点。

拟人机器学习

机器学习方法中的拟人化特征让计算机能够“像人类那样学习”。如今，绝大多数机器学习方法，包括主流DLNN，都需要大量的训练数据才能工作。但是，人却可以识别出之前只见过一次的事物。

想象一下悉尼歌剧院或伦敦塔桥——我们需要看过多少张训练图像才能将新图像与之联系起来？人类能基

此外，连续学习的能力也是一种拟人特征，绝大多数现有方法（强化学习法除外）都不具备。但强化学习不能保证收敛。相反，人类却一直在动态地发展他们对现实世界（亦即一种环境模型）的内在感知。我们都能够学习新信息

并更新从前学到的知识。这是大多数主流机器学习方法都缺乏的特点，但数据量和复杂性的急剧增加带来的挑战越来越大，非常需要获得这种特性。⁹

最后，人脑在能耗方面非常高效。它终日协调大量极其复杂的任务，耗电量在12瓦左右，是点亮一个标准60瓦灯泡所需功率的五分之一——这简直是奇迹！在机器学习中使用拟人算法的关键是以精益、高效计算的方式学习。然而，目前最先进的方法（尤其是DLNN）在计算上既昂贵又繁重，需要高性能计算、加速器（即GPU）等。优化硬件的尝试很多（包括用英特尔的Loihi这样的神经形态芯片¹⁰）。但我们认为，巨大的未开发资源是一种从根本上不同的算法处理，它使用递归、非迭代、基于原型的方法。

换言之，我们生活其中的数据丰富的世界急需我们向拟人化的机器学习方法转变。

主流方法的瓶颈

理想情况下，一个拟人系统将能够：

- > 从一个或几个例子中学习；
- > 终生从新观察到的例子中学习；
- > 解释它已知的内容，知其所不知，以及为何发生某个错误；

- > 识别与先前观察明显不同的数据样本，并在必要时形成新的规则和/或类别（自学和自组织）；
- > 以精益、高效计算的方式学习；
- > 与其他（如果有的话）拟人机器学习系统合作。

然而，主流方法最关键的瓶颈是缺乏情境感知能力。此外，以透明的、人类可解释的方式向用户呈现先前所学知识的能力也需要多得多的关注。这也有助于增加系统的可靠性。

DLNN可以从最多样化的场景数据中学习，但它们仍然是“黑盒”，因为它们不能适应新场景。DLNN能够对带有熟悉模式的数据执行高度准确的分类，但当遇到与先前观察不同、不熟悉的场景模式时会完全失败。此外，当发生这种情况时，DLNN自身无法识别它。这种缺陷并非小事，可能导致严重后果，因而进一步阻碍了DLNN在真实世界中的应用。一个众所周知的例子是使用基于DLNN的计算机视觉技术的自动驾驶汽车。^{6,7}它们经过实验研究，甚至在实际道路上测试了好几年。然而，尽管它们在熟悉的场景（高速公路、住宅区、白天或夜晚、雨、雪、车辆少或拥堵等）中表现良好，但无法识别当下所处的情境是已知或未知。这可能导致高风险并造成重大损害，包括人员死亡。因此，能识别未知情境并从中自主学习

对于该技术的安全性和更广泛的推广是至关重要的。⁹

最终，目标必须是建立一种系统，不仅可以识别此前已知的模式，还可以识别“意外”模式。在某种程度上，这个问题可以看作是让系统能够意识到自身局限，能在面对未知和不可预测的情况时启动安全程序，并从中自主学习。

算法的另一个非常重要的特征是它能否在所谓的“单回合”（one-pass）模式下工作。单回合系统在拿到数据样本（如图像）后，用算法处理它，之后就扔掉；样本不在内存中，也不再使用。如果验证数据与训练数据具有相似的特征和模式，则在离线模式下设计的学习模型在拿到验证数据后会有很好的表现。但当它们面对所谓的“移位和漂移”数据时，性能会急剧恶化。

⁹在这种情况下，离线设计的模型需要重新校准或完全重新设计，这不仅在模型设计成本方面非常昂贵，而且还有相关的行业停工成本。在实际问题中这经常发生，因为原材料质量（比如在炼油中）和照明条件（比如图像处理中的白天和夜晚）等因素可能发生显著变化。

单回合系统需要非常高技能的数据处理形式。它们要求算法仅依赖当前数据样本，或许也可能用到一些统计聚合和模型参数（如聚类中心、原型等），但不使用（可能是滑动的）数据样本窗口或所谓的“数据块”。它们在内存使

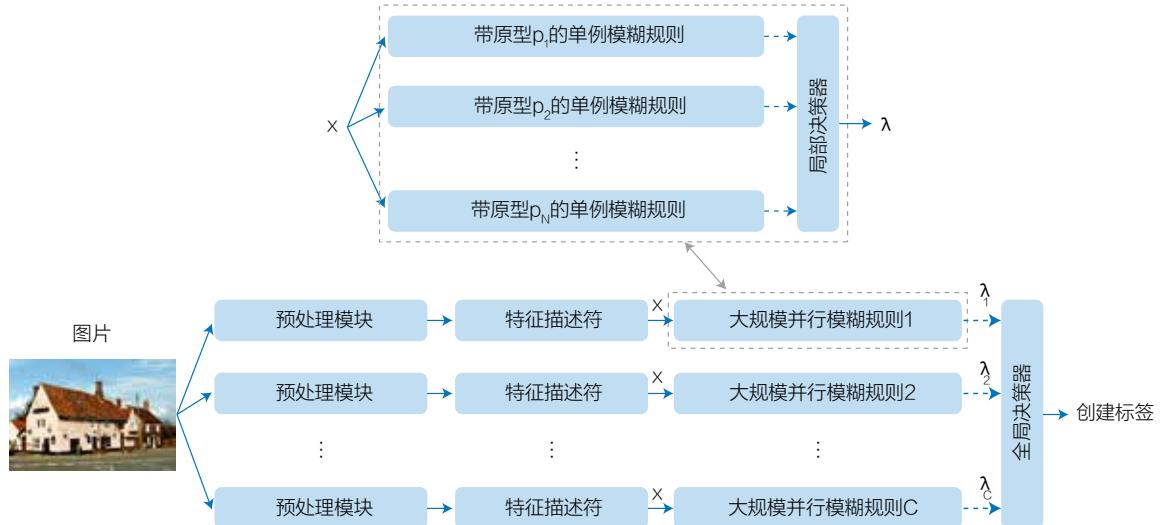


图 1. DRB 系统图片分类器的大概框架。

用和计算资源(搜索和存储)方面使用非常高效的操作模式，并且通常可以不断地自我更新系统结构和参数。差异是非常明显的。例如，在视频处理中，单回合算法仅处理当前图像帧，也不把过去的帧存储在存储器中，而只在内存中存储少量统计聚合和模型参数。⁹

算法要成为单回合模式，关键的一点是它必须是递归而不是迭代的。⁹原因很简单。如果算法涉及迭代搜索，则无法保证它会在下一个数据抵达前就得出结果。递归算法保证了单回合处理能力，因为这种算法在当前时间步更新参数时仅使用了这些参数在前一个时间步的值和一些简单的算术运算，如求

和或乘以数值系数。

迈向拟人机器学习的第一步

作为机器学习领域最新的进展之一，最近推出的基于深度规则(DRB)的系统提供了一种可能的解决方案。^{11,12} DRB系统是一种新的通用方法，它结合了自组织非参数FRB系统和DLNN多层结构的优点。¹¹ DRB系统能够提供自组织、自适应、透明、高度可并行化、基于规则的架构和学习算法，并且具有理论上已被证明的收敛性。须强调的是，这种新方法是一种通用的机器学习

方法，可以通过简单的修改而适用各种分类和预测问题。但在本文中，我们以图像分类为重点，推介它的一般概念和原则。¹¹⁻¹³

用于图像分类的DRB系统的一般结构如图1所示。从图中可以看出，该分类器由以下组件组成。¹¹⁻¹³

1. **预处理块。**它涉及在计算机视觉领域中广泛使用的预处理技术，包括归一化、缩放、旋转和分割。因此，它实际上是由用于各种目的的一些子层组成。
2. **特征描述符。**它将原始图像投影到一个特征空间，使不同类别的图像分离，即 $I \Rightarrow x$ 。
3. **大规模并行模糊规则库。**它是



图 2. 大规模并行模糊规则的图例。(a) 各有一个原型的三个模糊规则。(b) 大规模并行的模糊规则。(c) 通过移动一个原型的位置调整模糊规则。

一种复杂的非线性预测模型，充当系统的“学习引擎”。这个预测模型本身是大规模并行的AnYa类型0阶模糊规则的集合。⁹每个大规模并行模糊规则由从训练集中特定类别的样本中识别的大量（数量非预先设定）原型组成。因此，对于包含C个不同类别的数据样本（例如图像）的训练集，

$$IF \underbrace{\dots OR \dots OR \dots OR \dots}_{number\ of\ prototypes} THEN \dots$$

将识别出C个并行模糊规则（即一个类别一个规则）。

4. 决策器。一个类别带有一个局部/子决策器，给出一个局部建议。根据这些大量并行的局部建议的置信度来决定胜出的类别标签。通常，DRB系统的全局决策器和局部决策器都使用流行的“单个赢家通吃”原则。但是，也可根据具体问题考虑采用其他决策策略（比如“几个赢家通吃”、模糊加权和取平均值等）。

或者，DRB系统也可被视为一种自我演化的ANN。¹¹

具拟人特征的DRB方法运用了数据原型。在实践中，这些原型是从观测数据中自动提取的多模态典型性/数据密度的局部最大值。²原型对DRB方法非常重要，它们因此而与包括DLNN在内的主流方法相区分。^{2, 11, 12}这些原型基于已揭示的集成属性和使用无参数运

算符的数据相互分布，通过完全自主、在线、非参数、非迭代和单回合学习过程被识别。²它们是数据集中最具代表性的样本，直接反映DRB方法从数据中获得的知识。²

基于识别出的原型，DRB从数据中自组织和自我演化一个完全透明且可由人类解释的条件逻辑（if-then）大规模并行FRB系统。¹¹每个大规模并行的模糊规则是围绕（大量）原型确定的。这些原型是从特定类型的训练数据样本中识别出来。局部决策器使用“赢者通吃”原则将这些原型连接起来。公式1显示了如何得出一条模糊规则的例子。¹¹⁻¹³

$$R_i : \begin{array}{l} \text{IF}(x \sim p_{i,1}) \text{ OR } (x \sim p_{i,2}) \text{ OR } \dots \text{ OR } (x \sim p_{i,N_i}) \\ \text{THEN(Class } i) \end{array} \quad (1)$$

其中*i* 表示第*i*个类别的样本；*R_i*表示对应这个样本的模糊规则；“~”表示相似性，也可以看作是模糊的隶属度/满足；*x*是一个特定样本；*p_{i,j}*是这条模糊规则中的第*j*个原型；*N_i*是*i*类别中的原型数量。同样地，*R_i*也可以被显示为带有相同单例后项部分的（大量）更简单的模糊规则罗列。如公式2所示：

$$\begin{aligned} R_{i,1} &: \text{IF}(x \sim p_{i,1}) \text{ THEN(Class } i) \\ R_{i,2} &: \text{IF}(x \sim p_{i,2}) \text{ THEN(Class } i) \\ &\vdots \\ R_{i,N_i} &: \text{IF}(x \sim p_{i,N_i}) \text{ THEN(Class } i) \end{aligned} \quad (2)$$

每个模糊规则的隶属度/满足是基于最接近验证数据样本的原型计算得出的：¹¹

$$\lambda_i = \max_{j=1,2,\dots,N_i} \left(e^{-\|p_{i,j}-x\|^2} \right) \quad (3)$$

大规模并行模糊规则的单例后项部分进一步使得DRB系统不需要针对特定用户和特定问题的参数，也不需要手工构建隶属度函数。

每个大规模并行模糊规则（每个类别一个）可被视为由逻辑运算符“OR”（或）连接的一串更简单的模糊规则（它们每条各带一个原型和相同的单例后项部分）。由于每个模糊规则彼此完全独立，因此它们是并行训练的。此外，可以更新或删除每个规则，而不影响规则库中的其他规则。这就实现了非常高水平的并行化。

至关重要的是，DRB方法不是“黑盒”（不同于之前描述的主流方法3,5）。它最吸引人的部分之一是其基于原型的特质，这提供了系统内部结构的透明度和可解释性。大多数现有机器学习方法都需要大量训练数据，而DRB系统甚至可以从单个例子中学习——也就是“从零开始”。非迭代在线自主学习算法进一步使得DRB系统即使是在完成训练/部署之后也能终身不断地学习新观察到的样本，因此它是“不断演化的”。^{9,12}

DRB 图片分类器示例

举例来说，我们来看看使用来自三个国家（英国、希腊和日本）的各一栋典型房屋的单张图片训练的三个模糊规则，如图2（a）所示。通过提供更多训练样本，DRB系统识别出了更多新原型，每个模糊规则的结构变得大规模并行，整个DRB系统的规模也相应地变大。最重要的是，DRB系统更好地理解了问题。图2（b）中列出了这些规则，从中可以更好地理解从数据中自主提取的建筑风格差异。

DRB方法基于原型的特性带来了透明度和人类可解释性，这对于在现实世界中应用它非常重要。¹¹人们可以在训练过程之后检查汇总的信息，并且可以通过简单地查看系统识别的原型轻松地理解和分析系统学到的内容。这也让用户能在系统出错时识别错误并轻松地解决它们，比如删除被错误指定的原型、添加新原型，等等。

为说明这一点，让我们继续看图2（b）中的示例。人们可能会注意到规则R3的第二个原型p3,2实际上是一栋典型的中式房屋，被错误地标记成了日式房屋。要修改DRB系统，可以简单地从R3中删除p3,2，并创建一个新的规则R4，将p3,2用作其新原型。该过程如图2（c）所示。

DRB系统的另一个拟人特征是其

我们看到了以拟人方式开发下一代机器 学习的巨大机会。

强大的情境感知力，这种能力也与其基于原型的特质密切相关。¹¹⁻¹³对于新观察到的样本，DRB系统会拿它与之前确定的原型相比较。系统可以基于较低的隶属度来轻易地分离出带有不熟悉模式的新样本。这进一步使得系统能够利用基于原型的半监督和主动学习算法，在很少（或没有）人为干预或监督的情况下连续学习。¹²然后，DRB系统可以用实时、自主的方式主动学习新原型和新类别。用户可以选择是否在稍后查核新获得的信息。

协作机器学习

DRB系统透明的、可解释的内部结构及非迭代学习过程还使得不同的机器可以通过交换极少量信息（即原型及与之相应的单例后项部分）从不同的方面协作学习某个问题。¹⁴协作学习是一种拟人化特征，因为人们通常是通过交流来交换聚合信息而非原始数据。这也是一个非常吸引人的功能，因为它通过将问题并行化到多个处理单元而

得以处理超大规模的数据，这不仅加快了整个学习过程，也降低了对参与整个协作学习的每台机器的计算和内存资源的要求。^{9,14}相比之下，主流方法即DLNN的训练过程需要迭代解决方案，因此要实施并行化和扩展会非常昂贵又麻烦。

由于基于原型的特性，DRB系统可以在灵活的架构中协作学习。每个系统不断收集信息并与其它系统交换部分信息，然后汇总信息。通过这种方式，每个DRB系统都拥有所有互连系统学到的关于所处理问题的关键信息，而无需拥有全部观察结果。

拟人决策支持和推荐系统

决策支持系统（DSS）传统上基于FRB系统，后者通常以有人类专家参与的主观方式手工设计。另一方面，推荐系统则通常基于统计分析，但也需要特征选择和许多其他临时决策。两者都经常遇到所谓的“维度诅咒”困境。例

如，一种医疗DSS设计需要计算专家与医学专家合作以确定：

- > 相关特征（如收缩压、体温、血红蛋白水平等）；
- > 属函数的类型——如果使用FRB方法或必须确定一种数据分布模型（如高斯分布）的话；
- > FRB方法中语言术语的数量（如低、中、高、非常低、极低等）；
- > 属函数的参数（或概率密度函数[PDF]）——如果使用统计方法的话）。

很明显，传统的方法（不管是基于FRB系统还是统计方法）都很麻烦。它们涉及临时决策，并且在追求现实世界中的高规模应用时可能遭遇“维度诅咒”困境。此外，一旦构建了这些系统，以后就很难用新数据更新它们，因为它们是固定的、离线的，不会动态演化。

近期提出的替代方案——经验模糊集和系统¹⁵——具有拟人机器学习方法的许多特征。它们的工作方式只是询问人类专家（主观版本）或自动从数据中（客观版本）确定原型（每个类别至少一个）。例如，对于我们上面描述的医疗DSS，必须由医生（或基于运用所谓的“自主数据分区法”¹⁶得出的客观数据）来确定病患和健康人士的原型。类似地，人们可以很容易地想象一个用

关于作者

普拉门·P·安格洛夫 (Plamen P. Angelov)是兰卡斯特大学计算与通信学院智能系统的客座教授；兰卡斯特智能、机器人和自治系统(LIRA; www.lancaster.ac.uk/lira)研究中心主任。他的研究兴趣包括以人类可理解的基于模糊规则的系统和自动机器学习形式开展知识提取的线上演化方法论和算法。他获得保加利亚国家科学院哲学博士和理学博士学位。他是IEEE会士、国际神经网络学会(International Neural Networks Society)副会长、IEEE杰出讲师。他也是《演化系统》期刊(Evolving Systems, Springer出版社)的主编，以及《IEEE 控制论汇刊》(IEEE Transactions on Cybernetics)等几种期刊的副主编。可通过p.angelov@lancaster.ac.uk与他联系。

Xiaowei Gu是兰卡斯特大学高级研究助理。他的研究兴趣包括经验数据分析、自主学习系统和扩频信号处理。他正在攻读兰卡斯特大学计算机科学博士学位。他从杭州电子科技大学获得工科学士学位和硕士学位。他是IEEE的学生会员。可通过x.gu3@lancaster.ac.uk与他联系。

于选择度假屋或买车、租车的推荐系统。然后，经验模糊集方法将自动建立隶属函数（或者也可以设计典型性，它类似于PDF²）。这样，该系统就可用于自动识别未来的患者，或推荐汽车、度假屋等。

这种拟人DSS和推荐系统是：

- > 基于原型的；
- > 高度可解释的（让用户知道决策/推荐是如何做出的；可用于从数据中提取人类可理解的、透明的语言信息；使知识形式化，并通过隶属函数以数学形式表示）；
- > 个性化，而非一刀切的；
- > 计算上高效的；
- > 动态自我演化的。

此外，经验模糊集和系统的拟人特征¹⁵进一步促进了不同地区专家和计算机科学家之间的合作。例如，医生和其他领域的专家通常不愿意参与传统DSS的设计，主要是因为设计过程非常繁琐。而相比要求医生指定几百乃至几千个高斯隶属函数的负担，只提供少数几个病人和健康人的例子（亦即这两个类别的原型）要轻松得多。此外，要保证参数的一贯性及隶属函数（或PDF）的平滑形状通常并不能正确地表示真实数据。

现有的机器学习方法已经被开发了几十年，而与之密切相关的统计学习则基于已有数百年历史的概率论。深度学习这种特定方法借鉴了计算智能和ANN的关键要素，在近年成为了计算机

视觉和机器学习的“黄金标准”。然而，DLNN存在一系列问题（比如它们需要大量数据和计算能力），包括不能适应新场景。

然而，拟人机器学习是AI和数据科学未来发展的新兴方向。这一革命性转变为下一代机器学习提供了近似人类的能力，更有可能推动技术和日常生活领域里的各种突破。毫无疑问，未来将会开发出更多类型的拟人机器学习。■

参考文献

1. C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
2. P.P. Angelov, X. Gu, and J. Principe, “A Generalized Methodology for Data Analysis,” *IEEE Trans. Cybernetics*, 2017; doi: 10.1109/TCYB.2017.2753880.
3. N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge Univ. Press, 2000.
4. V. Mnih et al., “Human-Level Control through Deep Reinforcement Learning,” *Nature*, vol. 518, no. 7540, 2015, pp. 529–533.

5. Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature Methods*, vol. 13, no. 1, 2015, pp. 35–35.
6. A. Davies, "Google's Lame Demo Shows Us How Far Its Robo-Car Has Come," *Wired*, 5 Oct. 2015; www.wired.com/2015/10/googles-lame-demo-shows-us-far-robo-car-come.
7. E. Ackerman, "How Drive.ai Is Mastering Autonomous Driving with Deep Learning," *IEEE Spectrum*, 10 Mar. 2017; <http://spectrum.ieee.org/cars-that-think/transportation/self-driving/how-driveai-is-mastering-autonomous-driving-with-deep-learning>.
8. A. Nguyen, J. Yosinski, and J. Clune, "Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 15)*, 2015, pp. 427–436.
9. P. Angelov, *Autonomous Learning Systems: From Data Streams to Knowledge in Real Time*, John Wiley & Sons, 2012.
10. M. Davies et al., "Loihi: a neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, 2018, pp. 82–99.
11. P.P. Angelov and X. Gu, "Deep Rule-Based Classifier with Human-Level Performance and Characteristics," *Information Sciences*, vol. 463–464, 2018, pp. 196–213.
12. X. Gu and P.P. Angelov, "Semi-Supervised Deep Rule-Based Approach for Image Classification," *Applied Soft Computing*, vol. 68, 2018, pp. 53–68.
13. X. Gu et al., "A Massively Parallel Deep Rule-Based Ensemble Classifier for Remote Sensing Scenes," *IEEE Geoscience and Remote Sensor Lett.*, vol. 15, no. 3, 2018, pp. 345–349.
14. P. Angelov, *Machine Learning (Collaborative Systems)*, US Patent, 8250004, 2012, US Patent and Trade-mark Office.
15. P.P. Angelov and X. Gu, "Empirical Fuzzy Sets," *Int'l J. Intelligent Systems*, vol. 33, no. 2, 2017, pp. 362–395.
16. X. Gu, P.P. Angelov, and J.C. Principe, "A Method For Autonomous Data Partitioning," *Information Sciences*, vol. 460–461, 2018, pp. 65–82.

SUBMIT
TODAY

IEEE TRANSACTIONS ON BIG DATA

► SUBSCRIBE
AND SUBMIT

For more information
on paper submission,
featured articles, calls for
papers, and subscription
links visit:

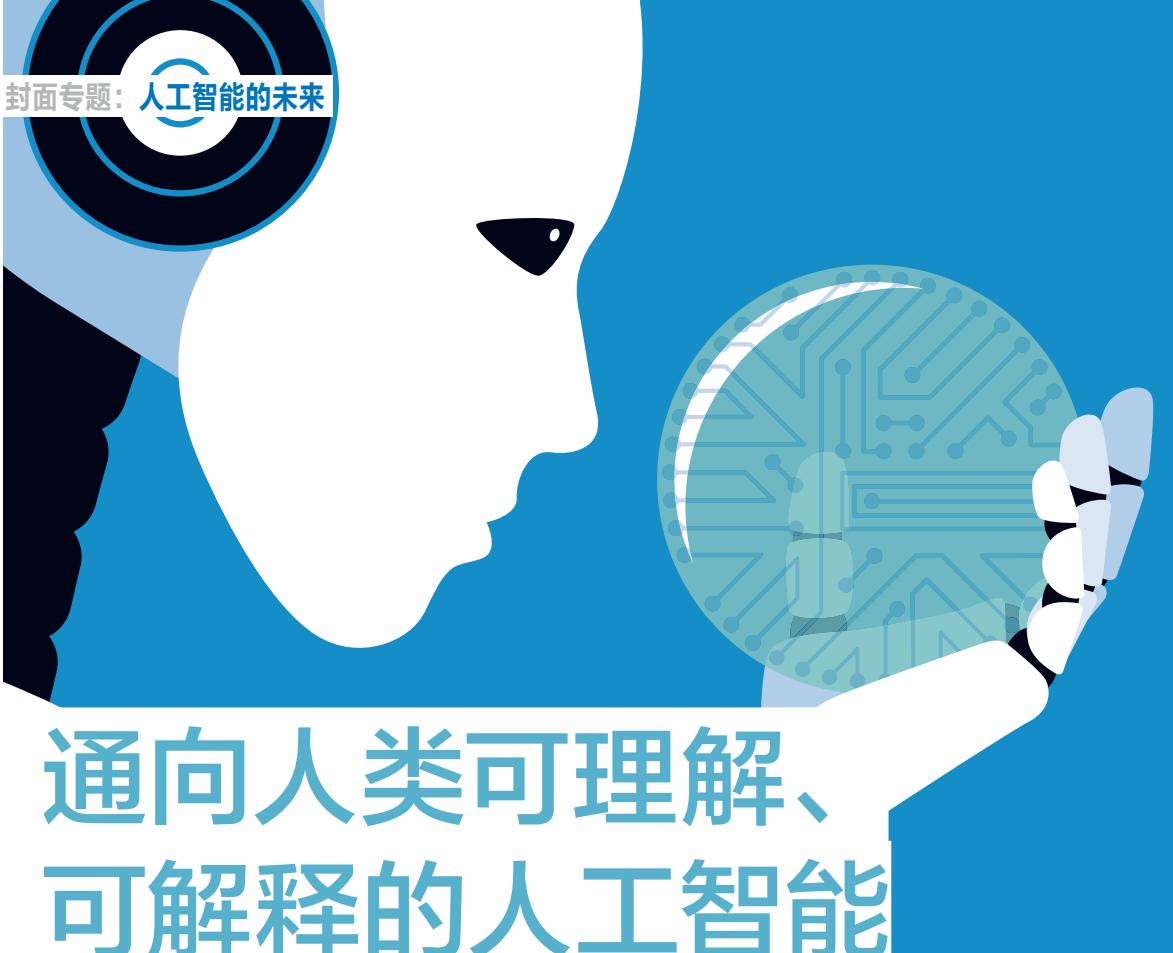
www.computer.org/tbd

TBD is financially cosponsored by IEEE Computer Society, IEEE Communications Society, IEEE Computational Intelligence Society, IEEE Sensors Council, IEEE Consumer Electronics Society, IEEE Signal Processing Society, IEEE Systems, Man & Cybernetics Society, IEEE Systems Council, IEEE Vehicular Technology Society

TBD is technically cosponsored by IEEE Control Systems Society, IEEE Photonics Society, IEEE Engineering in Medicine & Biology Society, IEEE Power & Energy Society, and IEEE Biometrics Council



IEEE
computer
society



通向人类可理解、 可解释的人工智能

文 | 哈尼·哈格拉斯 (Hani Hagras), 英国埃塞克斯大学
译 | 郑米兰, 浙江大学

近年计算能力的提高加之可获得数据量的迅速增长, 让我们重燃了对人工智能 (AI) 理论及应用的兴趣。然而, 如果各行各业以及政府要能信心十足地推出 AI, 则要考虑用户希望通过“可解释 AI”(XAI) 系统提高 AI 的透明度。本文作者介绍了 XAI 的概念, 概述了需要进一步探索的领域, 如 2 型模糊逻辑系统, 以确保这些系统能被非专业用户充分理解和分析。



人工智能 (AI) 旨在令机器能够执行通常需要人类智力的任务。AI 包含所有机器学习 (ML) 技术, 以及搜索、符号和逻辑推理、统计技术、基于行为的方法等其他技术。随着技术的发展以及——更重要的是——我们增进了对人类头脑如何运作及与周围环境互动的理解, AI 的概念已经发生了变化。

随着生成、存储和可用于分析的数字信息极大增加, AI 将发挥越来越重要的作用。构建 AI 系统的一个关键原因是

不仅仅要获得与人类一样的能力, 而是在某些情况下超越人类。这种需求在有成百上千项输入用于决策的情况下是显而易见的: 人类的直觉会集中于一小组输入和一小组交互集, 而难以弄明白大量输入之间的复杂关系和相互作用。将 AI 应用于商业的动力是巨大的, 因为它有机会降低成本、管理风险、加强决策、提高生产力, 以及开发新产品和服务。AI 是一个重大颠覆者, 预期将改变那些正在迅速将它融入各类应用的行业, 包括移动应用、安全系统、语音识别、金融业、物联网

网、智慧城市、汽车技术、生物科学、药剂学，等等。

监管机构和参与者希望，作为技术革命的推动力的AI具有包容性，能泽被每个人而不仅仅是少数人。然而，使用复杂的AI算法，如深度学习、随机森林和支持向量机（SVM），可能导致透明度不足，创建出“黑/不透明盒”模型。¹这个问题并不限于深度学习或复杂模型，也涉及其他分类器，比如内核机器、线性或逻辑回归，或决策树，它们在面对高维输入时也可能变得难以解释。²这样的黑/不透明盒模型不能告诉人们系统为何做出某项决定，它们只是提供一个答案，用户可以采纳也可以忽略。³

监督全球金融体系的金融稳定委员会（Financial Stability Board）称，金融部门对不透明模型（如深度学习技术）的广泛应用可能导致缺乏解释或可审计性，这可能导致宏观层级的风险。⁴该委员会于2017年底发布了一份报告，强调AI的进展必须伴随对算法输出和决策的解释的进一步进展。⁴这不仅是风险管理的重要要求，也是建立公众以及金融服务监管机构更大信任的重要条件。⁴

英国议会下属AI特别委员会在一份2017年报告中写道：

如果AI要成为我们社会中不可或缺又可信赖的工具，那么可理解

AI系统的发展是一个必不可少的基础……它会是以技术透明度还是可解释性的形式，或两者兼而有之，将取决于涉及的背景和利益，但在大多数情况下，我们认为可解释性是对公民和消费者更有用的方法。我们认为，部署任何可能对个人生活产生重大影响的AI系统都是不可接受的，除非它能对将部署的决定给出一个完整且令人满意的解释……在诸如深度神经网络的案例中，目前尚无法对所做决策给出细致全面的解释，这可能意味着要延迟为特定用途而部署这些决策，直至找到替代解决方案。⁵

由此可见，向“可解释AI”（XAI）迈进是非常重要的，这将使负责任又可信赖的AI被广泛采用，从而对全世界的社区和行业产生重大的积极影响。

何为“可解释AI”？

可解释性的概念处于几个活跃的AI研究领域的交叉点，它重点关注以下方面：⁶

> **透明度**: 我们有权要求以我们能理解的术语、格式和语言向我们解释那些将对我们产生影响的决策。⁷

> **因果关系**: 如果我们可以从数据中学习模型，这个模型是否能不仅为我们提供正确的推论，还能为我们提供一些有关背后现象的解释？

> **偏见**: 我们如何能确保AI没有因为训练数据或目标函数的缺陷而学到偏见的世界观？

> **公平性**: 如果决策是基于一个AI系统做出的，我们能否验证这些决策是公平做出的？

> **安全性**: 我们是否能对AI系统的可靠性产生足够的信心，而无需它解释如何得出结论？

XAI（或叫“透明AI”、“可译AI”）是人类可以容易地理解和分析其行为的AI。正如布赖斯·古德曼（Bryce Goodman）和塞思·弗拉克斯曼（Seth Flaxman）的研究所述，XAI可用于实现要求解释的社会权力。⁸因此，预期XAI将提供一份包含与特定预测相关的所有因素和关联关系的可审计的记录，从而提供透明度和合规性。这使企业能够满足合规要求，打消人们对于它可能隐藏信息或不清楚机器如何影响了某项关键决策的顾虑，并能证明算法的决策是公平和合乎道德的。

要获得透明度有其代价：在AI的准确度和透明度之间常常存在权衡取舍。而随着AI系统的内部复杂性增加，

预计需要做出更多取舍。解释复杂AI模型的决策的技术挑战有时被称为保罗·福森(Paul Voosen)口中的“可解释性问题”。根据安德烈亚斯·霍尔津格(Andreas Holzinger)及其同事的说法,¹⁰ XAI的目标应是创建一套能生成更多可解释模型的机器学习技术，与此同时保持机器学习的高性能表现(即高准确度)。此外，XAI模型应能解释自己的原理和优缺点，让人们了解它们未来会如何操作。这些XAI模型可与最先进的人机接口技术相结合，从而将模型译成能让终端用户理解的对话。

生成只能由AI专家分析和理解的格式并不能解决上述问题，因为它无法让利益相关者根据自己的经验来测试和加强所生成的模型。因此，XAI需要生成易于特定领域的普通用户或专家理解和分析的格式和输出。这将使得特定领域的专家能测试系统并轻松地利用自己的专业知识来加强它。这将让用户和利益相关者都能理解AI的认知方式，并使他们能够确定何时该信任或不信任AI。¹⁰ 这将满足前述透明度、因果关系、系统偏见、公平性和安全性的这些问题。

新老研究

XAI是美国国防部高级研究计划局(DARPA)的一个项目，有望实现“第

监管机构和参与者希望，作为技术革命的推动力的AI具有包容性，能泽被每个人而不仅仅是少数人。

三波AI系统”¹¹。这些机器了解它们运行的上下文和环境，并随时间的推移构建起底层解释模型，使它们能够表征现实世界的现象。根据2016年的一则DARPA报告，¹¹ XAI概念解释了单个决策，让人们能够理解系统的整体优劣势，了解系统未来如何运作，以及如何纠正系统的错误。图1a显示了DARPA提供的一份概述，详细说明了现有AI技术的性能与可解释性。其中显示，深度学习等黑盒模型与决策树相比，前者具有最佳预测精度，而后者提供了更高的可解释性。

决策树通过逐步评估数据点来分类，一次一个节点，从根节点开始，以终端节点结束。在每个节点只有两种可能性(左或右)，因此决策树无法学习一些变量关系。尽管决策树通常被认为易于解释，但是要打造决策树，尤其是具有许多分支的大型决策树，是复杂且耗时的。大树不容易被解释，而且还有呈现上的困难。此外，如果分析与某个决策相关的常见原因和特点需要分析决策树的各种路径和子路径，则整个过

程的难度加大，且决策者(特别是非专业用户)可能会被大量信息弄得不堪重负，而使决策力变慢。如果在不同路径上出现相同的子树，则情况还会变得更加复杂。因此，尽管决策树对于具有少量特征的问题而言可以是不错的可解释工具，但在处理具有大量特征的问题时往往不易于阅读、解释和分析(特别是对非专业用户而言)。

如图1a所示，目前有多种方法来实现XAI。第一种方法适用于深度学习和神经网络(在图1a中显示出最高的预测能力¹¹)，被称为深度解释。这种方法试图调整深度学习(或神经网络)技术以学习可解释的结构。这类技术的一些例子可以在格雷瓜尔·蒙塔冯(Grégoire Montavon)及其同事的研究中找到¹²，包括分层相关传播(LRP)技术。¹³

图1a中的第二种XAI方法是可解释模型，这类技术旨在学习更结构化和可解释的随意模型。它们可应用于统计模型(如逻辑回归模型和朴素贝叶斯模型等)、图形模型(如隐马可夫模型等)或随机森林。然而，与深度解释技术一

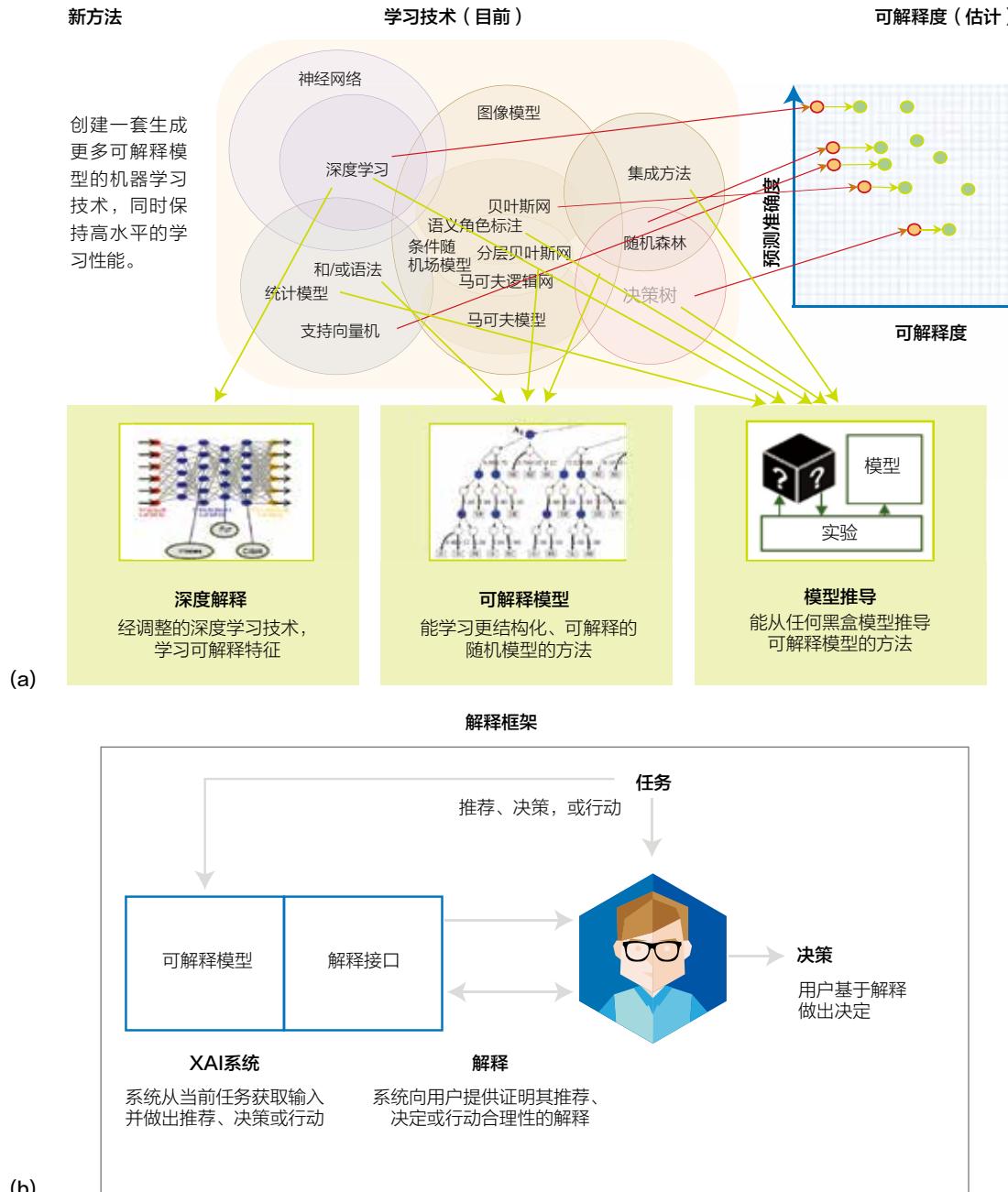


图1. 对可解释AI模型(XAI)的概述。(a)现有AI技术，显示性能与可解释性。(b)据2016年DARPA报告的XAI解释框架。¹¹(图片由DARPA提供。)

样，这类模型的输出只能靠这类技术的专家来分析，非专业用户无法胜任。

第三种XAI方法是所谓的模型归纳法，可用于从任何黑盒模型中推导出可解释模型。¹¹ 马尔科·里贝罗（Marco Ribeiro）及其同事的研究表明，虽然解释通常不可能是完全忠实的——除非它是对模型本身的完全描述，为了使解释有意义，它必须至少局部忠实，也就是必须符合模型在当前预测实例附近区域的行为。¹⁴ 此外，局部保真并不意味着整体保真：对全局重要的特征在局部环境中可能并不重要，反之亦然。虽然有一些模型本身就具有可解释的特性，但像模型归纳这样的解释器应该能够解释任何模型，也就是说，它是“泛模型”的。要让解释能被看懂，则需要使用人类可理解的呈现方式，不论模型本身使用什么特征。里贝罗和他的同事们提出了一种解释预测输出的方法：通过随机均匀地绘制 X 的非零元素来抽样围绕 x' 的实例（以创建新的点 z' ）。该方法而后要生成一个用 z 和 $f(z)$ 来训练的模型。¹⁴ 他们使用稀疏线性解释，但这种方法缺乏对驱动特定决策的各种变量之间关联性的解释。

在另一篇论文中，¹⁵ 里贝罗及其同事描述了稀疏线性模型等解释方法如何能通过只提供对局部的解释而保证高精度和低消耗——即使是面对非常复杂的模型。但是，由于这种解释的

覆盖范围并不明确，可能会导致人为错误。作者举了一个例子：¹⁵ 有一个复杂模型预测出某人的收入低于5万美元。线性解释说明了何以得出这个预测的原因，但不清楚从这一解释中获得的见解能否被应用于其他实例。换句话说，即使解释局部忠实，也不容易知道这个局部覆盖的区域是什么。¹⁵

会预测出不到50K美元的工资水平，而不管其他特点如何。里贝罗及其同事证明，这种方法优于他们在早期研究中提出的基于线性的模型。¹⁴ 不过，这种条件锚定模型¹⁵ 使用了清晰的逻辑，因此可能难以处理没有清晰边界的变量。此外，这种方法无法处理由大量输入生成的模型。再则，用一个条件锚定规则来

模糊逻辑可以模拟和表示不精确和不确定的语言化的人类概念，如低、中、高。

此外，即使在这个局部区域内部，这种线性近似估算何时更准确、何时较不准确也不清楚。因此，研究人员引入了“锚定局部可理解模型通用解释法”（aLIME）——一个用清晰的“如果-那么”（if-then）逻辑规则来解释单个预测的系统，它同样不针对特定模型。¹⁵

这样的“如果-那么”亦即条件规则对于人类来说是直观的，理解和应用起来通常也不会太费劲。¹⁵ 尤其是，aLIME解释法（或锚）是一种充分“锚定”预测的规则，即使改变其余实例也不受影响。例如，上述例子中的锚可能会做出这样的陈述：只要某人未接受过高中以上教育，那么该模型几乎总是

解释预测并不像分类问题那样给出关于做决策的完整图景，而当人们做决定时他们总会在头脑中权衡各种利弊。另一个主要问题是无法理解模型在当下实例的相邻区域的行为，以及如果某些特定特征发生变化时预测可能如何变化，等等。

从上面的讨论来看，为用户提供包含语言标签的条件规则似乎是一种可行的方法，能够解释和分析模型，从而促进模型输出的可解释性，如图1b所示。一种使用条件规则和语言标签的AI技术是模糊逻辑系统（FLS）。然而，FLS并未被当为一种XAI技术来广泛探索，也没有出现在图1a所示的分析中。

其中一个原因可能是FLS与控制问题相关，并且它们并未被广泛视作一种机器学习工具——因为它们需要其他技术的帮助才能从数据中学习自身的参数。以下小节将概述FLS，重点讲述它的优势和对它的误解，并推介2型FLS成为XAI研发中的重要组成部分。

FLS 和可被人类理解的AI

FLS试图模仿人类思维，虽然它并不像神经网络那样试图呈现大脑结构。它的重点是人类的思考方式是近似而非精确的。FLS的一个关键特征是建模和表示不精确和不确定的语言概念，创建一组语言条件(*if-then*)规则，以人类可读的形式描述特定行为。

一个很好的例子是人们在开车时经历的决策过程。我们不会想：“如果和前面那辆车的距离小于2.5米，而路面打滑的程度是10%，那么我要把车速降低25%。”我们会以不精确的语言标签来表达这些数字元素的近似态：“如果和前面的车距离短了，而路有点滑，那我要慢下来。”在不同的司机那里，“短”、“略滑”和“慢下来”这些字眼的数字内涵是不同的。如果询问司机与这些语言标签相关的确切数值，他们很难给出明确的答案。令人惊叹的是，人类仍然能以这些定义不清晰又含糊不清的语言标签交流，他们在讨论时不去探究确切的数值。事实上，这些不确定的概念使人类能够执行非常复杂的任务，比如开车或批核金融申请。

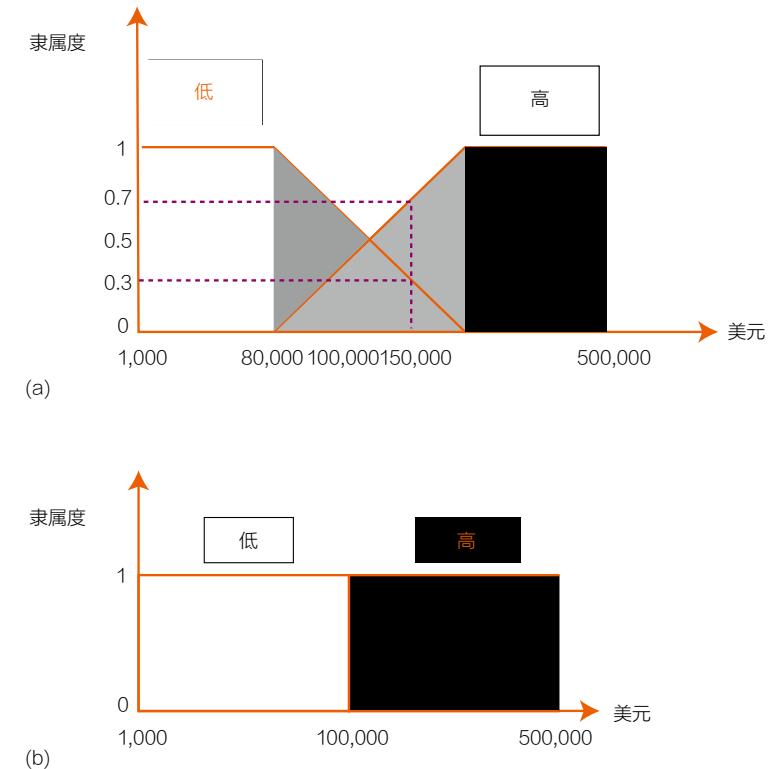


图 2. 使用 (a) 布尔集和 (b) 1型模糊集的低和高收入集表示。

清的语言标签交流，他们在讨论时不会去探究确切的数值。事实上，这些不确定的概念使人类能够执行非常复杂的任务，比如开车或批核金融申请。

模糊逻辑可以模拟和表示不精确和不确定的语言化的人类概念，如低、中、高。如果问一群人他们所说的“低”和“高”的年收入分别指多少钱，并且如图2a所示那样使用布尔逻辑的话，那么我们将不得不选择一个阈值，高于它的收入数额会被认为是高的，低于它

的会被认为是低的。这里存在的第一个难题是要确定大多数人都认同的阈值，因为每个人对这个语言标签的构成看法不同，这是一个挑战。即使达成了一致意见（比如使用10万美元为阈值），这是否意味着100,001美元的收入被认为高，但99,999美元就被认为是低的？显然，布尔数集生硬的分界从人的角度来看并不合理。

而如果用1型模糊集来表示语言标签“低”和“高”，则其中的集合之

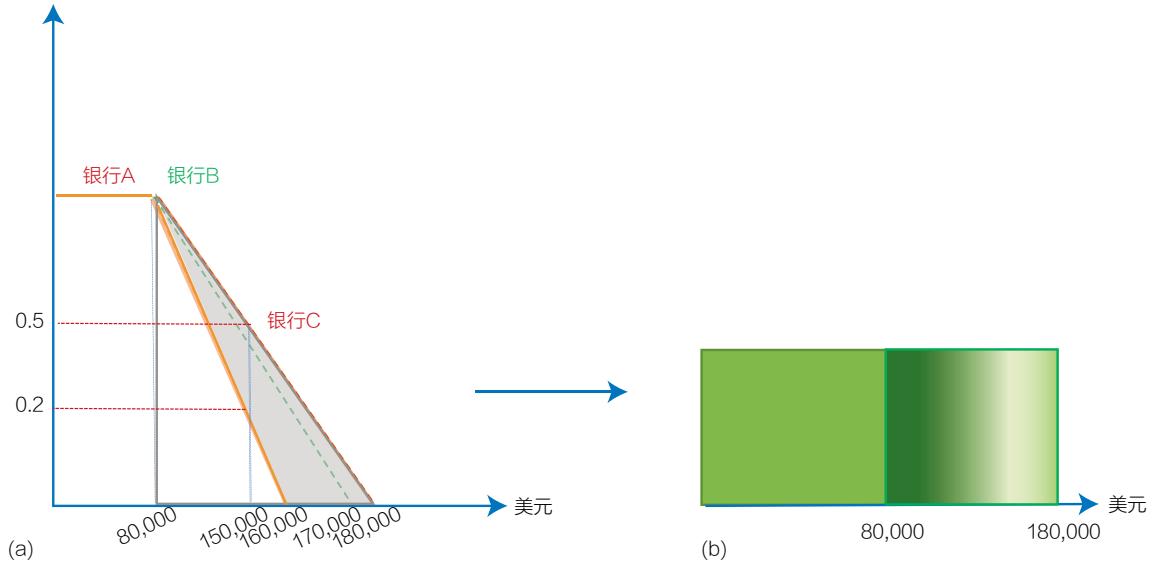


图3. (a)一个2型模糊集，它嵌入了来自三家银行金融专家的“低收入”语言标签1型模糊集。(b)图3a中所示的2型模糊集的图形简化。

间不存在明显的边界，并且x轴上的每个值可以属于具有不同隶属度值的多个模糊集。例如，如果使用布尔逻辑，150,000美元仅属于高集合，隶属度值为1.0(图2a)。在图2b中，使用1型模糊逻辑，150,000美元现在同时属于低和高集合，但程度不同：它在低集合中的隶属度值为0.3，在高集合中为0.7。这可能意味着，当10个人被问及150,000美元算低收入还是高收入，7人会说高(因此隶属度值为7/10即0.7)，3人会说低(隶属度值为3/10即0.3)。如此，模糊集提供了一种计算处于绝对真和绝对假之间的中间值的方法，其结果值介于0.0和1.0之间。也就是说，模糊逻辑可以计算真假之间的灰色度。此外，

模糊集之间的平滑过渡将在面对噪音和不确定性时给出良好的决策响应。再则，FLS采用了语言条件规则，使得信息能够以人类可读的形式呈现，可由非专业用户轻松阅读、解释和分析。

1型模糊集(如图2b所示)清晰、准确，因此只能处理轻微的不确定性。然而，不同的概念对不同的人群意味着不同的事物，而且在不同的情况下，图2b中所示的隶属度值可能在不同国家、不同职业和不同银行的不同人员之间有所不同。如图3a所示，我们邀请了来自三家不同银行(银行A、银行B和银行C)的三位金融专家分享他们对低收入范围的看法。从图3a中可以看出，每个专家可能给出不同的1型模糊集来表示

“低”的语言标签。

表示语言标签的另一种方法是采用如图3a所示的2型模糊集。它将银行A、银行B和银行C的所有1型模糊集嵌入到2型模糊集的不确定性足迹(FoU)中(图3a中的灰色阴影部分)。因此，2型模糊集的特征在于模糊隶属函数。具体来说，该集合中每个元素的隶属度值是[0,1]中的一个模糊集，而不像在1型模糊集中隶属度值是[0,1]中的一个清晰的数字。2型模糊集的隶属函数是三维的，并且包括不确定性足迹(FOU)。这提供了额外的自由度，使得人们可以直接建模和处理不确定性。在图3a中可以看出，150,000美元的低集合隶属度值不再是图2b中的0.3这个清

晰值，它现在是一个模糊函数，在主隶属域中取值为0.3到0.5。有关2型模糊集和系统的更多信息可以在其他研究中找到。^{16,17}

对2型模糊集的一个误解是认为非专业人员难以理解它。然而，当银行专家被问及如何量化语言标签时，情况并

25年。然而，FLS无法处理具有大量输入的系统这一点阻碍了这个领域的发展。这是因为FLS可以生成长规则和巨大的规则基础——这一现象被称为维数诅咒——这使得它们变成不易理解或分析的黑盒。此外，FRBS无法轻易处理不平衡和偏倚的数据（例如欺

将FLS预测的准确性最大化。此前的研究已表明，这种高度可解释系统在准确性方面表现大幅优于C4.5这样的决策树，同时又比决策树更易理解和分析。

最重要的是，与其他白盒技术不同，FRBS使用语言标签生成条件规则，这可以更好地处理信息中的不确定性。举例而言，当银行审查贷款申请时，一条规则可能是：如果收入高且拥有房屋和居住时间很长，则该申请被视为来自优质客户。任何用户或分析师都可以阅读此类规则。更重要的是，这些规则使得数据开始“说”与人类相同的语言。这使我们能够轻松地分析和解释所生成的模型。最重要的是，我们能用采集了专业知识的规则来扩充这些规则库，并可能覆盖数据中的空白（例如，人类经验可以增强历史规则，利用人类专业知识来涵盖未曾发生过的情形）。这让用户能完全信任生成的模型，并实现与透明度、因果关系、偏见、公平性和安全性相关的所有XAI组件。与里贝罗等人提到的锚定规则不同，¹⁵人类并不根据单一规则做出决策，而是在头脑中权衡语言化的利弊规则后做出相应的决定。

因此，回看图1a可以看出，与其他黑盒技术相比，2型FLS和FRBS可能具有最佳可解释性，并能在可解释性和预测精度之间达到良好平衡。此外，2型FLS可用于解释从更复杂的黑盒建模

在不同的司机那里，“短”、“略滑”和“慢下来”这些字眼的数字内涵是不同的。

非如此。他们对核心值很确定（所有专家都对它达成了共识），但是他们很难给出确切的边界。对于一个特定语言标签的始末点存在不确定性。图3b中显示了2型模糊集的简化版本。其中，对于低收入的语言标签，有一个小于80,000美元的核心值（实心绿色阴影）获得了所有专家的认同；在80,000美元到180,000美元之间存在一个灰色区域（绿色渐变），其中隶属度值逐渐变小；而对于超过180,000美元后哪一点不再被认为是低收入存在不确定性。

对FLS总体上的另一个误解是认为它们是控制机制。这也是不正确的，因为从数据生成的基于模糊规则的系统（FRBS）这个领域已经活跃了超过

诈数据和银行默认数据等）。然而，乔斯·桑斯（Jose Sanz）及其同事¹⁸和米凯拉·安东内利（Michela Antonelli）及其同事最近的研究¹⁹使用进化系统，生成了具有短条件规则和少量规则的FRBS，同时还能把预测准确性最大化。由于这创建出了一个不覆盖整体搜索空间的稀疏规则库，他们提出了一种相似性技术来对输入的示例分类，即使它们与生成规则库中的任何模糊规则都不匹配。为此，他们考虑了未覆盖的样本和规则之间的相似性。他们还提出了多目标进化优化方案，不但能提高可解释性（方法是减少每个规则的长度至包括3到6个前提，即使系统有成千上万个输入，并维持小型规则库），还

技术中获得的决策。因此，2型FLS和FRBS在未来可以提供很好的方式来实现可被非专业用户理解、分析和加强的XAI。■

参考文献

1. G. Nott, "Explainable Artificial Intelligence: Cracking Open the Black Box of AI," Computerworld, Retrieved Feb. 2017; www.computerworld.com.au/article/617359/explainable-artificial-intelligence-cracking-open-black-box-ai.
2. Z. Lipton, "The Mythos of Model Interpretability," arXiv, June 2016; <https://arxiv.org/abs/1606.03490>.
3. A. Griffin, "Facebook's AI Creating Its Own Language Is More Normal than People Think, Researchers Say," The Independent, 3 Aug. 2017; www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-own-language-what-does-it-mean-fair-robots-chatbots-a7874576.html
4. "Artificial Intelligence and Machine Learning in Financial Services," report, Financial Stability Board, 1 Nov. 2017; www.fsb.org/wp-content/uploads/P011117.pdf.
5. "AI in the UK: Ready, Willing and Able?," report, UK Parliament (House of Lords) Artificial Intelligence Committee, 16 April 2017; <https://publications.parliament.uk/pa/ld201719/ldselect/l dai/100/10002.htm>.
6. C. Wierzynski, "The Challenges and Opportunities of Explainable AI," Intel.com, 12 Jan. 2018; <https://ai.intel.com/the-challenges-and-opportunities-of-explainable-ai/>.
7. A. Weller, "Challenges for Transparency," Arxiv, 29 July 2017; <https://arxiv.org/pdf/1708.01870.pdf>.
8. B. Goodman and S. Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" AI Magazine, vol. 38,no.3, 2016, pp., 50-57.
9. P. Voosen, "How AI Detectives Are Cracking Open the Black Box of Deep Learning," Science, 6 July 2017; doi: 10.1126/science.aan7059.
10. A. Holzinger et al., "A Glass-Box Interactive Machine Learning Approach for Solving NP-Hard Problems with the Human-in-the-Loop," ArXiv, 3 Aug. 2017; <https://arxiv.org/abs/1708.01104>.
11. D. Gunning, "Explainable Artificial Intelligence," report, Darpa, 11 Aug. 2016; www.darpa.mil/attachments/XAIIndustryDay_Final.pptx.
12. G. Montavon, W. Samek, and K. Müller, "Methods for interpreting and understanding deep neural networks," Digital Signal Processing, vol. 73, 2018, pp. 1-15.
13. S. Bach et al., "On Pixelwise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation," PLoS ONE, vol. 10, no. 7, 2015.
14. M. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," Proc. 2016 ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD 16), 2016.

关于作者

哈尼·哈格拉斯 (Hani Hagras) 是英国埃塞克斯大学计算机科学与电子工程学院计算智能中心的教授、研究主任和主管。他的主要研究兴趣是计算智能，尤其是2型模糊系统、模糊逻辑、神经网络、遗传算法和进化计算。他从埃及亚历山大大学获得电子工程学士和硕士学位，从埃塞克斯大学获得计算机科学博士学位。他是IEEE会士、英国工程技术学会 (IET) 会士、英国高等教育学院 (HEA) 首席研究员。可通过 hani@essex.ac.uk 联系他。

15. M. Ribeiro, S. Singh, and C. Guestrin, "Nothing Else Matters: Model-Agnostic Explanations by Identifying Prediction Invariance," ArXiv e-prints, November 2016; <https://arxiv.org/abs/1611.05817>.
16. H. Hagras, "A Hierarchical Type-2 Fuzzy Logic Control Architecture for Autonomous Mobile Robots," IEEE Trans. Fuzzy Syst., vol. 12, no. 4, 2004, pp. 524-539.
17. J. Mendel, Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions, 2nd Edition, Springer, January 2017.
18. J. Sanz, et al., "A Compact Evolutionary Interval-Valued Fuzzy Rule-Based Classification System for the Modeling and Prediction of Real-World Financial Applications with Imbalanced Data," IEEE Trans. Fuzzy Syst., vol. 23, no. 4, 2015, pp. 973- 990.
19. M. Antonelli, et al., "Multi-Objective Evolutionary Optimization of Type-2 Fuzzy Rule-based Systems for Financial Data Classification," IEEE Trans. Fuzzy Syst., vol. 25, no. 2, 2017, pp. 249-264.

Call for Articles

IEEE Pervasive Computing

seeks accessible, useful papers on the latest peer-reviewed developments in pervasive, mobile, and ubiquitous computing. Topics include hardware technology, software infrastructure, real-world sensing and interaction, human-computer interaction, and systems considerations, including deployment, scalability, security, and privacy.

Author guidelines:
www.computer.org/mc/pervasive/author.htm

Further details:
pervasive@computer.org
www.computer.org/pervasive

IEEE Pervasive Computing
 MOBILE AND UBIQUITOUS SYSTEMS



人工情感智能的时代

文 | 达格玛·舒勒，audEERING 公司

比扬·W·舒勒，伦敦帝国理工学院、奥格斯堡大学、audEERING 公司

译 | 郑米兰，浙江大学

在科幻小说中，未来的人工智能（AI）技术通常拥有高水平的情商，甚至具有同情心。但我们目前处于什么样的阶段？本文作者深入探讨了人工情感智能（AEI），提出AI要达到新的情感智能时代所需的三大发展领域——识别、生成和增强。



类在面对技术和计算系统时可能会经历非常感性的时刻，特别是在遭遇挫折时。1999年的电影《办公空间》（Office Space）中有这样一个关键场景：三名程序员把一台烦扰了他们太久的复印机砸了个稀巴烂。当然了，我们大多数人在使用计算设备时都曾经历过某种情绪化的时刻，比如死机或蓝屏时。但是，我们是赞美还是责怪手中的设备会不会很快将有所不同？人机交互（HCI）中的情感会从单向变成双向吗？

当谈到AI时，情感以及情感智能通常不是人们首先想到的事。我们目前在日常生活中遇到的传播最广的AI应用常与计算机感知和自然语言处理相关。接下来，我们往往会想到知识表示（学习、计划、推理及解决问题）或动作与操纵等能力。这之后，我们又会想到企图获得类似人类甚至超越人类的智慧这类目标。但目前的AI大体上还不具备创造力，也还需要规模化地增加情感和社交智能。

然而，AEI在很久以前就已出现在电影电视中。想想

《2001太空漫游》(1968)中的HAL 9000计算机，它显露出了社交情感智能的最早迹象，还有2013年的《她》中高度情感化的个人AI设备Samantha。另一个早期的例子是《霹雳游侠》(Knight Rider, 1982)，其中装配了AI的智慧汽车KITT觉察到驾驶员正处于“因疲劳引发的轻微烦躁情绪”中。电影中的AI往往还会堕入爱河，比如《电脑梦幻曲》(Electric Dreams, 1984)中的个人电脑和它的主人为同一个女人争风吃醋，而在《机器人总动员》(2008)中，机器人Wall-E爱上了另一个机器人。这样的例子太多了：《霹雳五号》(1986)；《星际迷航7》(1994)；《钢铁巨人》(1999)；《我，机器人》(2004)；《银河系漫游指南》(2005)；《钢铁侠》(2008)；《月球》(2009)；《机械姬》(2014)；《超能陆战队》(2014)……

显然，这些想象中的人机对话与今天发生的人和实际的AI(如Alexa、Cortana、Siri)甚至非消费级别的AI(如IBM的Watson，它非常缺乏情感智能)之间的交互大相径庭。人类的想象力对AEI的迷恋已经催生出了情感运算¹、社交和行为运算²以及情感增强机器学习³等领域。在本文中，我们将探讨以下几个方面：哺乳动物和人工情感的原理、情感智能的建模和历史，以及识别、生成和通过对目前AI系统的“情感

增强”来应用情感原理的现状。我们也会讨论AI是否会真正拥有情感的问题，展望获得情感增强的AI的未来，以及为抵达这样的未来最先要做出哪些努力。

情感和社交智能

情感智力(Emotional intelligence，也称情绪智力、情感智慧等)一词最早于1964年出现在迈克尔·贝尔多克(Michael Beldoch)的一篇论文中。它常见的定义是能识别情感(包括他人的和自己的情感)、产生和适应情感，以及为达成目标和解决问题而应用情感信息的能力。拥有这些能力的前提是能区分不同的情感。

但是，为何情感智力会在哺乳动物的生活中发挥关键作用？是否可以假设，在下一代AI中它也会如此？从今天的计算角度来看，人们可能会倾向于将人类情感视为最佳沟通或规划决策中的“噪音”。然而，根据心理学和神经科学的研究结果，情感是智能行为的重要组成部分，并且在一系列不同的认知、感知和身体进程中发挥着关键作用。^{4,5}

在整个类人的进化过程中，情感帮助人类生存——“战斗或逃跑反应”就是一个例子。情感是我们主要的动机因素，⁵与我们的记忆和学习系统直接相关，并影响我们的联想、抽象、直觉和

推理，帮助我们从探索转向利用。在现代评估理论中，情感被广泛视为适应性反应，⁶有助于我们评估环境、监控自我福祉，以及了解自己所处的状态。因此，情感在强化学习中发挥奖惩作用，引导我们的注意力，帮助我们做出决策。^{4,5}最后，情感在沟通中发挥关键的作用，但在口语对话系统或一般的人与计算机或机器人的交互中仍被普遍忽略。

人类情感

人类如何体验、处理和运用情感？神经影像学研究揭示了有关人类和一般哺乳动物的生理结构如何处理情感的重要信息。情感是一个多面、复杂的现象，但与之一贯相关的一个生理结构是杏仁核。研究表明，杏仁核与大脑内的其他物质高度连通，与大脑边缘系统等部分直接或间接地分享向内和向外投射。³因此，杏仁核(情感)参与多个神经元过程，如加工原始数据(丘脑)、对象层数据(皮层)和背景数据(海马体)。本文把杏仁核当作单一系统来处理，尽管它结构复杂，而且可以影响其他行为，如运动以及通过其他系统的自动响应。

人们认为杏仁核能够评估环境刺激，将注意力转移到情感相关特征上。这些情感关联的形成称为情感学习。杏

仁核很可能负责存储“刺激-情感”的反应模式，并且已被证明参与了陈述性记忆（即可被有意识地记起的记忆）的形成过程，且提供“情感色彩”。此外，情感唤起会引发名为去甲肾上腺素的神经调节剂的分泌，继而增强记忆力和学习效果。有关杏仁核（以及情感）在注意力和记忆中的作用的证据支持了这样的观点：情感对于评估环境以及识别在环境中生存下来所需的重要刺激（如需要进食）是至关重要的。总的来说，研究揭示了大量证据，证明了杏仁核（以及情感）在大脑神经的决策过程中是一个关键的核心。

情感建模

对情感的最佳建模存在于心理学文献中。而在技术应用中，鲜少有模型被广泛使用。在早期，分类法占了上风，例如保罗·埃克曼（Paul Ekman）提出的“六大”独立情感类别（愤怒、厌恶、恐惧、快乐、悲伤和惊讶）。⁷但这样的分类往往过分简化了现实世界里的情感的微妙特质，而且覆盖面太过有限。⁸按维度连续建模日渐流行。最常被应用的维度是唤起、效价和支配。⁷情感类别可以被映射到这些维度所跨越的空间中的区域。举例而言，恐惧的情感以高唤起、负效价和低支配为特征。分类或连续维度的方法在情感识别和生成

中都特别流行。在强化学习中，基于评估的方法⁶得到了广泛应用。此外也存在其他模型和混合模型，例如将类别转换为维度，或标记模型——此时情感可能同时属于超过一个类别。

建模的另一个重要方面是时间演变。这种建模可能具有挑战性，因为不同的模态可能在不同的时间层面上运作。例如，原始音频的采样率（通常高于8kHz）与心电图信号（通常低于0.5kHz）等大多数生理参数不同，也与视频或深度信息帧（通常低于100Hz）不同。类似地，特征采样率在很大程度上也存在差异（音频通常约为100Hz，而基于特征向量的语言分析通常以小于1Hz的频率采样）。因此，在跨模态建模中可能需要做出折衷。时间连续建模中的典型情感更新频率（正日益流行⁷）约为0.1Hz到1Hz。

最后，需要对具体某种情感建模。在计算机对人类情感的识别中，除了自我评估的“感觉到的”内在情感以外，通常还要评估由他人感知的情感。此外还可以对情感的其他方面建模，例如表达、调节或抑制情感的程度；表达、调节或抑制情感的意向程度；情感的典型性程度；情感的特异程度。这也可能影响AI的“内在”和“外在”情感，即AI自己的“感觉”和外界感受到的它的“感觉”。这样一来，（举例而言）AI也可能抑制它的情感。

计算机和情感的历史

“情感计算”一词由麻省理工学院媒体实验室的罗莎琳德·W·皮卡德（Rosalind W. Picard）于1995年在她的同名开创性著作中首次提出。¹然而，它的概念可以被追溯到略早些时候。例如，约翰·D·威廉姆森（John D. Williamson）在1978年提交了自动语音情感识别的第一项专利。1989年，同样来自麻省理工学院媒体实验室的珍妮特·E·卡恩（Janet E. Cahn）撰写了关于“合成语言中情感生成”的文章。¹⁰1992年，小林宏和文雄原发表了对能自动识别前述六种基本面部表情的神经网络的研究。¹¹

在诠释面部情感方面，欧洲的SEMAINE项目（于2010年结束）提供了首个能够识别用户情感并在一条二维视听输入-输出链中生成适应性智能体输出的实时系统，应用于情感智能对话。类似的项目包括南加州大学的SimSensei（于2011年启动）和ARIA-VALUSPA（即“信息助理的人工检索-具语言理解、社交技能和个性化特征的虚拟智能体”，于2015年启动）。

人机情感交互网络（HUMAINE）创建于2004年，2007年转变成一个不受资助的网络，最终于2014年发展为总部位于英国的情感计算进步协会（AAAC，本文作者之一的比

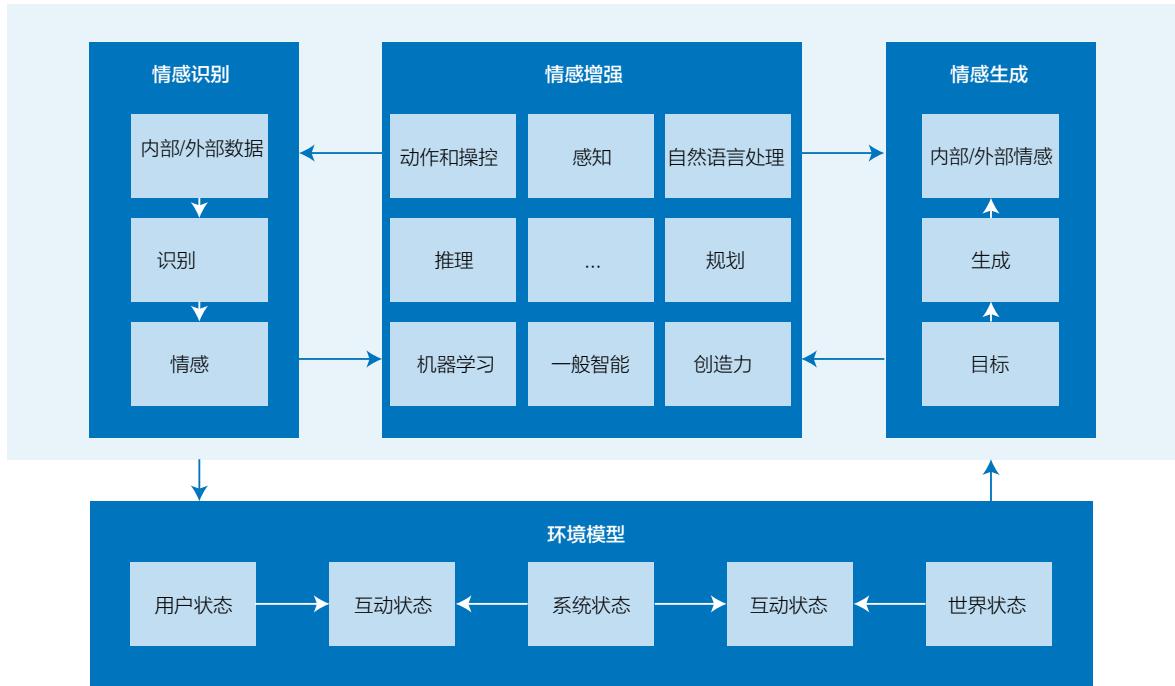


图 1. 将情感智能融入 AI 的主要领域。

扬·舒勒担任第一届主席）。2005年，第一届情感计算与智能互动国际会议（ACII）召开，自此每两年举办一届。2010年，《IEEE情感计算学报》（*IEEE Transactions on Affective Computing*）发刊，至今仍是这一领域里的主要期刊。该领域的第一个公开竞赛活动是“Interspeech 2009情感识别挑战赛”，它基于语音声学和语言学内容分析，由比扬·舒勒发起并与其他人共同组织。舒勒在2011年又举办了首届“音频/视觉情感大赛”（AVEC），丰富了这一概念。同期出现的几项类似

的竞赛包括EmotiW（自然环境下的情感识别大赛）、FERA（面部情感识别和分析竞赛）、MEC（多模态情感计算竞赛）和OMG（一分钟微表情竞赛）。此外还有一些与情绪、情感相关的文本分析竞赛。AVEC在2015年举办了第一项基于生理学的比赛。

专注人类情感识别的商业创业公司包括：创立于2009年的Affectiva，以生理和视频为重点；成立于2012年的audEERING，以音频为重点；成立于2007年RealEyes，以视频为重点。它们如今都已是价值成百上千万美元的

企业。在这个快速增长的市场中，它们的同类也越来越多。然而，消费者产品仍然稀少，且在很大程度上不为公众所知。

AEI 的三个构成模块

迄今为止，对AEI的研究重点主要放在聊天智能体和机器人的自动人类情感识别及情感生成上。在学习算法和对话管理的情感增强上已有所尝试，但规模要小得多。图1显示了情感识别、情感生成和情感增强，它们可被看作AEI

的主要构建模块。

情感识别

计算系统对情感的识别已有二十多年的历史，主要集中在对人类情感的识别上。情感已成为音乐、声音、图像、视频和文本中的一个主流议题。情感分

许从原始数据中识别情感)或浅层建模(例如具最先进性能的跨模态光谱变换)。重复观察到的一个结果是模态的互补性：声音声学携带了有关唤起和支配的信息，面部表情和口语语言与正负效价密切相关。根据上述领域里的基准和竞争状况来看，可媲美人类的水平已

体生成，包括情感驱动的面部表情、身体姿势和运动(如Greta引擎)；情感文本生成。不过，目前的趋势是更多地使用深度学习，例如在情感语音合成中利用WaveNet。这主要是针对“外部”情感，也就是可以从外部观察到的情感，而不是“内在”的AI或系统的情感。

在机器学习中融入情感计算原则具有提高学习效率的巨大潜力。

析的主要模态是声学语音和口头(或书面)语言内容；表情；身体姿势和步态等运动；生理测量，如心率、皮肤电导、甚至大脑活动。触觉交互中的情感也获得了一些研究。

早期的情感识别系统以各种专业制作的特征为标志，例如由openSMILE(音频)或OpenCV(视频)以不同的采样频率提取的特征。⁷早在2008年就有了各种各样的机器学习算法，包括支持向量机和内核机器、隐马氏模型和更通用的图形模型，还有神经网络，比如具有长短期记忆的(深)递归神经网络。

今天的方法越来越多地关注深度端到端学习，例如End2You工具包(允

被部分达到或即将达到，即使是在“自然”条件下。^{7,8}需要注意的是，由于AI的“内部情感”还不是非常复杂，从AI的视角来看，被识别的情感通常是外在的(例如人类和动物表现出来的情感)。

情感生成

类似于情感识别中分析那部分，情感生成中综合的部分也具有更长的历史，特别是对情感语言和面部表情的综合可以追溯到近三十年前。情感生成的其他示例包括文本和触觉反馈。然而，与分析相反，综合的方法在传统上更基于规则而非数据训练。⁷它们主要集中于情感语音合成，例如MARY Text-to-Speech系统(MaryTTS)；视觉智能

情感增强

与情感识别及生成相比，各类研究文献很少尝试讨论AI的情感增强(例如在规划、推理或更宽泛的目标达成中应用情感)。双向情感输入/输出平台SEMAINE和ARIA-VALUSPA(以及其他用于情感增强人机对话的系统)是通过情感原理来丰富对话管理的例子。然而，和情感生成类似的是，这种对话管理目前主要基于规则来操作。此外，大多数情感增强都是单向的情感输入或输出。例如，用户情感已被部分用于某些形式的系统状态调适，比如在视频游戏中。

情感增强的其他例子大多存在于情感增强机器学习(EML)中，⁹旨在把情感作为一种仿生物原则来促进学习效率。在AI中包含杏仁核原理¹²将效法受哺乳动物神经网络启发而发展人工神经网络，或者更具体地说，是效法在哺乳动物的视觉皮层的启发下发展卷积神经网络。实际上，传统机器学习算法的设计通常旨在让误差函数最小化，

而很少将情感作为一种指导因素。

尽管情感计算尚需为我们提供更深刻的发现来指导如何模拟和生成人工情感,但似乎显而易见的是,在机器学习中存在着整合这类原则来提高学习效率的巨大潜力。例如,可通过情感直觉修剪或类似的情感驱动学习指导来加快收敛速度或降低计算成本。与此同时,或许还可以找到更好的整体解决方案,并且学习算法的情感状态同样可以为计算结果的置信度测量提供参考。

把情感原理嵌入机器学习或更广泛的AI的方法可被粗略分为四大类。下面将对这些内容展开讨论和举例,以深入了解目前在AI情感增强方面取得的重要成果。请注意以下每种情况都还可以举出大量案例。

优化。有一点已被反复证明:情感可被嵌入训练机器学习算法的优化过程中,例如用于神经网络。阿德南·凯什曼(Adnan Khashman)的研究就表明,情感可以被有益地融入前馈神经网络中。¹³为此目的,他将焦虑和自信这两种情感形式化,并引入了一种情感反向传播学习算法,这种算法被在隐藏的输出层中发现的情感偏见加强。当新模式出现时,他选择提高焦虑水平,将网络的输出与预期识别错误相关联。焦虑的加剧导致更多新错误被关联,而更高的信心则会强调之前的更新。与传统的反

向传播相比,基于这种情感增强训练的面部识别示例展现出了更高的识别率,执行时间也更短。我们可以将这些原则与如今深度学习中存在的有关注意力的不同观点相联系,清楚地看到在算法执行上存在着的不同见解。在之后的研究中,凯什曼让输入信号中的其他因素或“刺激源”也影响焦虑水平。基于视觉皮层具有背侧(认知)和腹侧(情感)两种相互独立的信息流通路的特点,他引入了处理局部特征的认知神经元和处理全局的情感神经元的“双神经元”构建,进一步改善了算法。

杨易民(Yimin Yang,音译)及其同事开展了与之相关的研究。¹⁴在用人工“感觉”和情感控制从探索到利用的变化的机器学习中,他们规避了陷入“局部最佳”的风险。一个激素系统会进一步提供反馈以激活感觉。这里部署的情感与凯什曼的相似——信心受到调节而焦虑取决于奖励减少。此外,随着焦虑的加剧而出现恐惧的情感,它将控制权衡探索与利用的战略;而根据迭代次数和恐惧的程度,还会出现“温暖”的情感来控制学习终止。该模型还包含了“一种主导情感”的原理。同样地,这个系统受益于情感增强而加快了收敛速度。

其他类似的方法包括EMANN(情感人工神经网络),它基于包含接收和发射“激素”的节点的多层感知器。¹⁵研

究人员的构想是让激素水平控制激活阈值、权重求和,以及最终神经元的输出和特定节点的输出。EMANN能克服学习中的停滞期,从而超越传统的多层次感知器。

强化学习。情感增强的概念已经成功地应用于强化学习中。^{3,16}一个例子是智能体基于评估理论的原则学习旨在让奖励最大化的最优“状态-动作”映射。在这种情况下,评估关联到当前状态以改变奖励。除内在动机(换言之,内部或基于情感的奖励)外,关键因素是考量强化学习中的共同外在动机。实验证明,与传统的智能体激励相比,基于评估的情感整合可以令学习速度加快,因为奖励可被不断提供,而不仅仅是在完成学习之后。如果存在没有情感反应的潜在行为,较长期情感状态(或称“心情”)的原则可以让这种效果进一步受益。³

将情感概念整合到强化学习中的其他形式是基于动态平衡的原理,将心理平衡作为对智能体的一个吸引点来驱动学习。动态平衡即是智能体的动态平衡变量处于平衡状态,这些变量包括主要情感和潜在的次要情感。智能体所经历的驱动将导致影响这些变量的行为,继而形成对奖励的理解。

情感增强也被用于强化学习中的行动选择。³与上述例子类似,效价可被用来引导探索/利用行为。例如,负效

价可以导向考虑更广泛的行动选择，反之亦然。³挫折可以帮助控制驱动策略，例如通过调整行动选择中价值函数的权重。关于该论题的广泛调查可在《强化学习智能体和机器人的情感：一项调查》中找到。¹⁶

AI解剖模型。目前已有受大脑系统启发的整体模型。克里斯蒂安·鲍肯纽斯 (Christian Balkenius) 和扬·莫伦 (Jan Morén)¹²首次提出了这样的模型，其中包括海马体、眶额皮层 (OFC)、杏仁核、丘脑和感觉皮层的模型，通过这些组件的相互依赖性产生情感调节。³模拟的杏仁核用于学习情感关联，模拟的眶额皮质是背景抑制剂，而背景信息是从模拟的海马体中通过将刺激与位置匹配注入的。进一步的例子包括基于大脑情感学习的智能控制器 (BELBIC)，这是一种受大脑边缘系统启发的算法，被成功应用于工程任务的控制器中，显示出良好的泛化性和灵活性，因为该算法有助于适应参数变化和干扰。之后，大脑情感学习 (BEL) 实现了哺乳动物杏仁核的短路径，这种短路径将感知丘脑和那些与刺激额叶皮层交流的长路径连接起来。杏仁核的输出被拿来与输入侧的奖励做比较——如果不存在输入侧奖励仍可以使用这种输出。BEL的表现可以超越多层感知器和模糊干扰。类似地，基于大脑情感学习的模式识别器 (BELPR)¹²

在分类和时间序列预测任务中优于多层感知器。在自适应衰变大脑情感学习 (ADBEL) 中，遗忘过程被添加到杏仁核中。在这类模型中，杏仁核模糊性和眶额皮质变量等额外扩展可以帮助提高系统的性能。最后一个例子是基于边缘的人工情感神经网络 (LiAENN)，它结合了上述几种方案，包含焦虑和信心等情感、短路和长路、杏仁核的遗忘过程，以及通过眶额皮质与杏仁核的互动实现的情感抑制。同样地，这些方法被反复证实超越了多层感知器和其他机器学习方法。³

学习中的认知和抽象。一系列进一步的例子展示了如何在抽象和学习中使用情感增强，例如向由神经网络控制的智能体添加“情感电路”；或者基于情感来控制行动选择，它受情感关联的影响，这种关联基于一个专注模型，其中包括情感生成器、行为系统，以及运动系统等在情感调节机器人智能体 Yuppy 中所见的特征。此外，在学习情报分布智能体 (LIDA) 中，认知行为（以及行动选择）是通过包含注意力、行动选择和动机的情感增强产生的。基于事件评估的情感会引发学习和干扰的变化。读者可以在《情感增强机器学习：一个新兴领域概述》³中找到更多有关这些研究的细节和线索。另一个例子是 Sigma，¹⁷它基于的图形模型受到体现注意力控制欲的评估变量的增

强。类似地，MAMID¹⁸在一个架构中加入情感建模，基于评估过程调节情感状态。这些情感状态会影响处理能力、威胁偏差等。在引入了这类架构的研究中都观察到了优于非情感增强解决方案的表现。

一个天价问题：AI会有情感吗？

在电影《她》中，个人 AI 设备 Samantha 寻思：“这些感觉是真的吗？还是只是程序？”虽然感觉和情感不是一回事，但这仍让我们思考，未来的 AI 会拥有真实的情感还是模拟的情感——尽管后者仍可能达到相同的目的。这在很大程度上是一个哲学问题。不过，皮卡德曾谈到情感驱动的机器，这些情感可能包括好奇心、思想和动机。¹阿兰·图灵在 1950 年表示，“如果一台机器表现得和人一样聪明，那么它就是和人一样聪明。”在面对和人相似的 AEI 时，我们也可以这么想。1956 年，由约翰·麦卡锡 (John McCarthy) 发起的达特茅斯提案中首次提出了“人工智能”一词。麦卡锡在提案中说：“学习的各个方面或任何其他智能特征都可以被精确地描述，以至于机器能够模拟。”麦卡锡似乎确信 AEI 是可以实现的。1976 年，艾伦·纽厄尔 (Allen Newell) 和赫尔伯特·西蒙 (Herbert

Simon) 呼应了他的观点：“一个物理符号系统具有必要和充分手段来实现一般智能行动。”最有趣的是，在确定“强”人工智能的概念时，约翰·R·塞尔(John R. Searle)于1980年提出：“被赋予了正确程序的计算机可说确实具有理解力和其他认知状态。”

有广泛的专家群体认为可在AI系统中实现AEI。但是，对“真实”情感的一种主张是它需要身体以及身体与现

人工智能项目被广泛认为是一种失败”，但他认为AEI的原则可能会改变局面。²³与其他人一样，巴赫支持让情感影响学习、记忆、感知、行动选择和计划。约翰·朱尔斯·Ch·迈耶(JohnJules Ch. Meyer)在《情感智能体的推理》²⁴中还讨论了监控智能体行为的能力。

自纽厄尔的著作《认知的统一理论》于1990年出版以来²⁵，各种文献中都反复强调了将情感纳入构建“深情”

习中——具有提高效率甚至提高准确度的潜力。然而，仍有许多其他领域和AI应用尚未从情感增强中受益。

要克服AEI中的这些“空白”，推导出方法来做以下事情似乎是至关重要的：在AI中包含情感概念；定义基于外部和内部状态评估这种情感随时间而生的变化；定义这种情感变化的后果。理想情况下，要找到这种变化在强化学习、耦合分析和情感综合中的后果。这样，我们不仅可以识别情感，也可以“生成”情感，让我们获得对底层概念的一种关联性理解。

从这个意义上说，AEI需要被更加无缝地融入到未来的AI中，超越目前常常孤立且以用例为导向的现状。因此，我们的观点主要的新颖之处在于我们认为完全嵌入的AEI应该是AI的核心部分而非装饰品或附加物。同时，AEI需要在识别、生成和应用上形成整体性，将目前正考虑把情感原理嵌入计算系统的各种子学科大体上统一起来。

最终，这将导致一系列道德、法律和社会影响，需要从技术的角度加以解决，以创造可审计、可追责、可解释、靠得住、负责任的AEI。这将令我们为不久后完全情感化智能计算系统的到来做好更充分的准备。■

参考文献

1. R.W. Picard, Affective Computing,

可以将情感嵌入到用于训练机器学习算法的优化过程中，比如用于神经网络。

实世界的连接。于尔根森·施米德胡贝(Jürgen Schmidhuber)等人为此提到，疼痛传感器已经存在于机器人技术中。¹⁹正如前文已部分提及的，许多论文作者讨论了AI与情感之间的关系以及它们如何交织在一起。这些构想包括：用情感控制AI的行为选择、行为启用、行为强度或对某些行为的阻止；注意力和感知机制²⁰；在建立对象的功能描述²¹或在向人类学习时作为一种强化来源。²²许多进一步的研究都涉及情感如何影响推理过程。例如，乔舒华·巴赫(Joscha Bach)在十年前就表示“人

机器的蓝图的重要性。这些文献涵盖情感计算、行为计算、认知建模、认知科学、认知系统、适应行为等等，不一而足。AEI已经日益发展成一个成熟的领域，使计算机识别并生成的情感已经具有可被应用于真实世界的性能表现。然而，就完全的双向情感输入/输出而言，视听和文本对话系统目前占主导地位。事实上，目前的AEI主要侧重于在人机交互中应用情感增强，希望以此增加交互自然度，并在受情感增强的检索和监测中加强亲和力。与此同时，人们观察到，AI的情感增强——特别是在机器学

关于作者

达格玛·舒勒 (Dagmar Schuller) 是慕尼黑和柏林的音频智能公司audEERING的联合创始人兼首席执行官。该公司主要研发人工情感智能。她的研究兴趣包括语音情感识别和一般音频智能和情感计算。她拥有维也纳经济大学的经济科学硕士学位。可通过ds@adeering.com与她联系。

比扬·W·舒勒 (Björn W. Schuller) 是伦敦帝国理工学院人工智能全职教授和GLAM(语言音频与音乐小组)负责人、奥格斯堡大学医疗与健康嵌入式智能全职教授、audEERING公司联合创始人和首席科学官、哈尔滨工业大学永久客座教授。他在慕尼黑工业大学获得电气工程/信息技术博士学位。他是IEEE会士、英国情感计算进步协会(AAAC)荣誉会长、国际计算机学会(ACM)高级会员。他撰写或合著了750多篇论文，现任《IEEE情感计算学报》主编。可通过schuller@IEEE.org与他联系。

- MIT Press, 1995.
2. A. Pentland, "Socially Aware Computation and Communication," Computer, vol. 38, no. 3, 2005, pp. 33-40.
3. H. Strömfelt, Y. Zhang, and B.W. Schuller, "Emotion-Augmented Machine Learning: Overview of an Emerging Domain," 7th Int'l Conf. Affective Computing and Intelligent Interaction (ACII 17), 2017, pp. 305-312.
4. C.E. Izard, Human Emotions, Springer, 1977.
5. B.L. Fredrickson, "What Good are Positive Emotions?," Rev. General Psychology, vol. 2, no. 3, 1998, pp. 300-319.
6. K.R. Scherer, "Appraisal Considered as a Process of Multilevel Sequential Checking," Appraisal Processes in Emotion: Theory, Methods, Research, 2001, pp. 92-120.
7. H. Gunes and B. Schuller, "Categorical and Dimensional Affect Analysis in Continuous Input: Current Trends and Future Directions," Image and Vision Computing, vol. 31, no. 2, 2013, pp. 120-136.
8. M. Soleymani et al., "A Survey of Multimodal Sentiment Analysis," Image and Vision Computing, vol. 35, pp. 3-14, 2017.
9. J.D. Williamson, Speech Analyzer for Analyzing Pitch or Frequency Perturbations in Individual Speech Pattern to Determine the Emotional State of the Person, US Patent 4,093,821, 1978.
10. J.E. Cahn, Generation of Affect in Synthesized Speech, MIT Media Lab, 1989; www.media.mit.edu/publications/generation-of-affect-in-synthesized-speech-2.
11. H. Kobayashi and F. Hara, "Recognition of Six Basic Facial Expressions and Their Strength by Neural Network," Proc. IEEE Int'l Workshop Robot and Human Communication, 1992; doi: 10.1109/ROMAN.1992.253857.
12. C. Balkenius and J. Morén, "A Computational Model of Context Processing," Proc. 6th Int'l Conf. Simulation of Adaptive Behaviour, 2000, pp. 256-265.
13. A. Khashman, "A Modified Back-propagation Learning Algorithm with Added Emotional Coefficients," IEEE Transactions on Neural Networks, vol. 19, no. 11, 2008, pp. 1896-1909.
14. Y. Yang et al., "Hybrid Chaos Optimization Algorithm with Artificial Emotion," Applied Mathematics and Computation, vol. 218, no. 11, 2012, pp. 6585-6611.
15. R. Thenius, P. Zahadat, and T. Schmickl, "EMANN—A Model of Emotions in an Artificial Neural Network," The Twelfth European



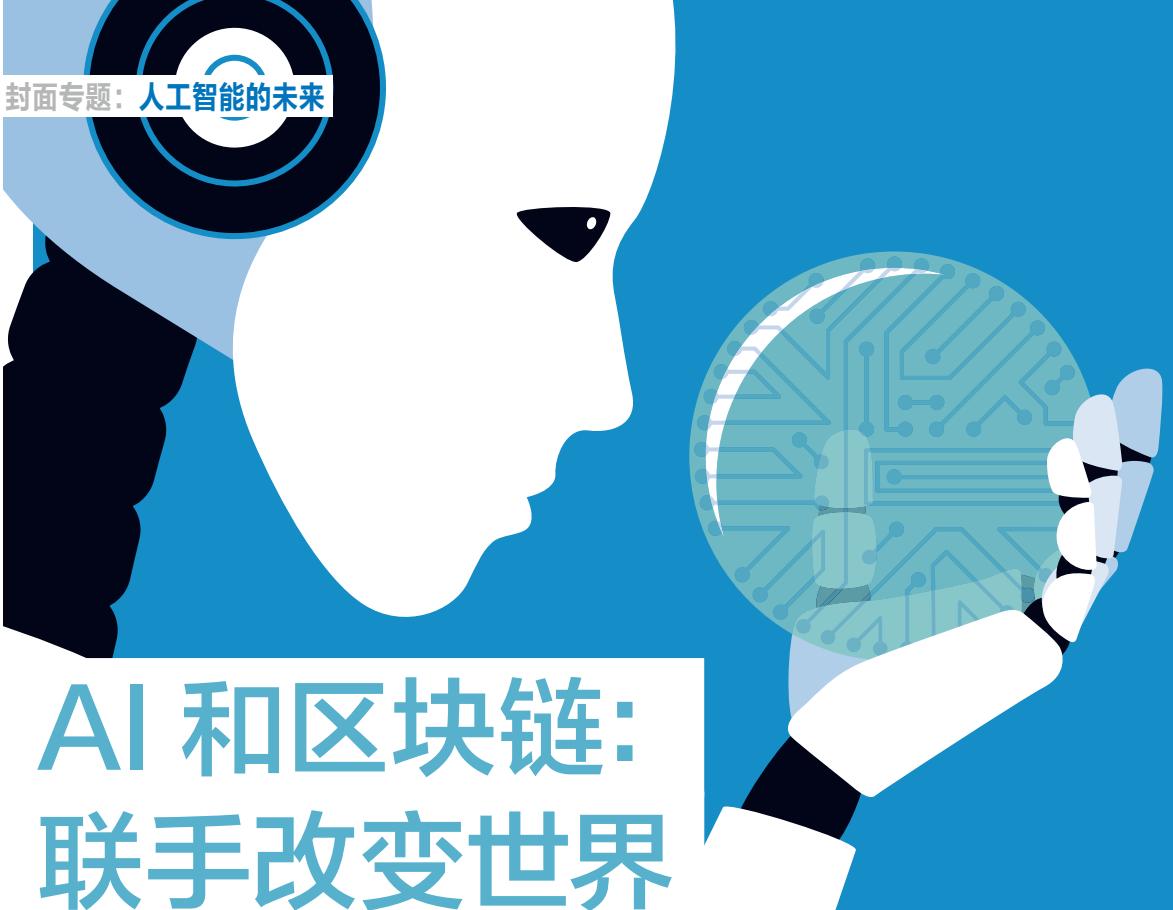
- Conf. Artificial Life (ECAL 13), 2013, pp. 830–837.
16. T.M. Moerland, J. Broekens, and C.M. Jonker, “Emotion in Reinforcement Learning Agents and Robots: A Survey,” *Machine Learning*, vol. 107, no. 2, 2018, pp. 443–480.
17. P.S. Rosenbloom, J. Gratch, and V. Ustun, “Towards Emotion in Sigma: From Appraisal to Attention,” *Int'l Conf. Artificial General Intelligence (AGI 15)*, 2015, pp. 142–151.
18. E. Hudlicka, “Beyond Cognition: Modeling Emotion in Cognitive Architectures,” *Proc. 6th Int'l Conf. Cognitive Modeling*, 2004, pp. 118–123.
19. P. Oltermann, “Jürgen Schmidhuber on the Robot Future: ‘They Will Pay as Much Attention to Us as We Do to Ants,’” *The Guardian*, 18 Aug. 2017; www.theguardian.com/technology/2017/apr/18/robot-man-artificial-intelligence-computer-milky-way.
20. D. Cañamero, “Modeling Motivations and Emotions as a Basis for Intelligent Behavior,” *Proc. First Int'l Conf. Autonomous Agents (AGENTS 97)*, 1997, pp. 148–155.
21. I. Cos and G.M. Hayes, “Behaviour Control Using a Functional and Emotional Model,” *Proc. 7th Int'l Conf. Simulation of Adaptive Behavior (ICSAB 02)*, 2002, pp. 226–227.
22. J. Broekens, “Emotion and Reinforcement: Affective Facial Expressions Facilitate Robot Learning,” T.S. Huang et al., eds., *Artificial Intelligence for Human Computing*, LNCS 4451, Springer, Berlin, Heidelberg, 2007.
23. J. Bach, “Seven Principles of Synthetic Intelligence,” *Proc. 1st Conf. Artificial General Intelligence (AGI 08)*, 2008, pp. 63–74.
24. J.-J. Ch. Meyer, “Reasoning about Emotional Agents,” *Int'l J. Intelligent Systems*, vol. 21, no. 6, 2006, pp. 601–619.
25. A. Newell, “Unified Theories of Cognition,” *Harvard Univ. Press*, 1990.

Call for Articles

IEEE Software seeks practical, readable articles that will appeal to experts and nonexperts alike. The magazine aims to deliver reliable, useful, leading-edge information to software developers, engineers, and managers to help them stay on top of rapid technology change. Topics include requirements, design, construction, tools, project management, process improvement, maintenance, testing, education and training, quality, standards, and more.

Author guidelines:
www.computer.org/software/author
 Further details: software@computer.org
www.computer.org/software





AI 和区块链： 联手改变世界

文 | 升廷 (Thang N. Dinh)，弗吉尼亚联邦大学

梅泰 (My T. Thai)，佛罗里达大学

译 | 颜磊，清华大学

AI 和区块链是最具颠覆性的技术，它们将深刻重塑我们的生活、工作和人际关系。作者总结了已有的技术，并探讨了两者结合的未来前景，试图回答这个问题：智能、去中心化和安全系统将会如何影响我们的社会。

AI 和区块链是当下科技创新的关键驱动力。它们都根本性地改变了我们的生活并将继续为全球经济贡献数万亿美元。未来，自动驾驶汽车和智能助手可以用自然语言和你交谈。新型资讯和经济共享平台的出现将代替“无法让人信服”的Facebook、Yahoo和Equifax们。如果这两种技术结合起来会发生什么呢？我们将分别介绍它们，并分享对这个问题的看法。

AI最早由马文·明斯基 (Marvin Minsky) 和约翰·麦卡锡 (John McCarthy) 定义（他们被誉为“AI之父”），一切由程序或机器运行但却似乎需要智能的行动都可以称作AI。AI系统通常会展现以下与人类智能相关的行为：计划、学习、推理

和解决问题，也包括社交能力和创造力。最近这些年AI的复兴主要归功于机器学习尤其是深度学习领域的突破。同时它受益于数据爆炸，这使机器学习训练更加高效。除了带来自动驾驶汽车和搬运机器人等进步，AI也引起了一些担忧，从制造假新闻（如以假乱真的照片、声音、成人电影）到侵犯个人隐私。另外还有人担心谷歌、微软、亚马逊等巨头会凭借数据、人才和算力优势垄断AI技术。

区块链的发展历程

区块链是公共分类帐，由分布式网络中的所有用户共享

和同意。数据记录（例如交易）与散列值和时间戳一起存储在块中。每个块都连接到前一个块，形成一个链（由此得名）。区块链的一个关键特征是不变性。也就是说，如果没有网络共识，几乎不可能修改任何信息。

根据共识协议，即块的创建方式，区块链技术分为两组。在诸如比特币和以太坊之类的工作量证明（POW）区块链中，用户称为“矿工”——参与挖掘过程以解决创造新块的计算难题。创造了块的“矿工”获得块奖励并收取交易费用。POW协议通常非常耗能。另外，当块奖励减少时，它们受到大多数哈希速率攻击，例如最近BitcoinGold, Verge, ZenCash和其他基于POW的加密货币相关的事件。

新一代协议是利益证明（POS）区块链没有耗能的挖掘过程。取而代之的是由参与者的货币量——也即筹码——决定创造新块的机会。这方面最有名的要数Nxt、Peercoin、delegated EOS、比特币模仿者Ouroboros、iChing和最近诞生于MIT的POS+BFT混合区块链Algorand。

减少当前围绕加密货币的狂热才能释放出区块链技术真正的潜能，它将带来全新的无服务器和分散式的互联网，那时“用户可以控制自己的数据、特征和命运”。区块链也将颠覆医疗系统，人们可以追踪个人数据的用途并

拥有数据版权。到那个时候，我们将有权换掉那些当下很出名却不够好的平台，不论是Facebook、Ebay、Uber、Airbnb还是能源市场。

颠覆性的整合

一方面，区块链存在安全性、可扩展性和效率等弱点。另一方面，人工智能在可信度、可解释性和隐私方面存在相当大的问题。两种技术的结合似乎是理所应当的，它们可以相互补充，彻底改变下一个数据时代。如图1所示，区块链能为AI带来去信任化、隐私和可解释性；反过来，AI可以在块上构建机器学习系统，以实现更好的安全性、可扩展性以及更有效的个性化和管理。

区块链用于AI

区块链可以为AI的各种组件（包括数据、算法和算力）提供分散的市场和协调平台。这将使AI的创新和应用推向前所未有的水平。区块链有望使AI更透明、可解释且值得信赖。区块链上所有的数据都是公开的，因此AI是为用户提供机密性和隐私的关键。

安全的数据共享

当前AI革命的驱动力之一是可用于研究、开发和商业化的数据。数据是

当下数据驱动型经济中的珍贵资源。然而获得这种资源的门槛很高。首先，除非你是Facebook或谷歌这样的顶级玩家，想要获得训练模型所需的有效数据非常困难。这可能会阻碍AI研究者和公司的技术升级之路。

第二，一系列个人数据泄露和滥用事件之后，人们对隐私的担忧越来越强烈。例如最近引起轩然大波的Facebook丑闻中，5000万用户在未许可的情况下被剑桥分析公司（选举第三方机构）定型和定位。这种精确靶向的行动令人担忧，马里兰大学网络安全教授Timothy Summers将这类事件比作“耳语运动”：别有用心的团体通过向每个用户“耳语”分裂美国人。他们抓住用户的恐惧心理，鼓动人们传播这种恐惧，达到某种目的。

区块链能够推动数据共享，因为它通过提供了数据访问、访问时间以及访问者信息建立起透明度和问责制。随着区块链将数据控制权交回用户手中，人们对数据共享更有信心，并且他们知道自己的数据将用于更好的个性化服务或其他有益的事业中。这种潜在的影响是巨大的。例如，医生和研究人员可以访问海量的匿名医疗记录和病例。患有罕见疾病的患者将获得新的希望，因为医生可以从世界各地获取病例。随着一些应用区块链技术的疾控中心和其他健康公司的发展，部分区块链带

来的医疗进步已经变成现实。

数据即财富

除了共享和控制自己的数据之外，区块链技术还可以让你通过智能合约销售数据，这使数据市场无需中间人，因而更加安全和私密。这种市场可以降低小型玩家的门槛，平衡竞争环境，从而促进创新。通过零知识证明等技术，企业和研究人员能够在不知道数据细节或数据所有者身份的情况下搜索相关信息。它带来的影响难以估量，甚至可能会引燃新一轮AI进化。例如，由哈佛大学乔治·丘奇(George Church)共同创办的初创公司Nebula Genomics提供了这样一个市场，将希望基因组测序的人与想要这些数据的公司联系起来。Longenesis是另一个类似的平台，用于共享和货币化生活数据，如医疗记录和健康数据。

出售富余算力

区块链可以通过分散的计算能力市场(即基于区块链的云计算)为AI提供更多的分布式计算能力。AI开发人员现在可以利用游戏玩家的数百万GPU来准备、建模、训练和运行他们的机器学习算法。游戏玩家的GPU经常只使用一小段时间，他们可以以AI智能合约的形式列出用于出售的计算时间，并获得报酬。

如果你认为出售算力的想法并不新鲜，那你是对的。网格计算曾经是非常受欢迎的创意，但后来发现它的应用范围很有限，并不适合大众使用。相比之下，今天AI支持的云计算市场拥有众多应用程序，毫不逊色于AI的应用范围。此外，通过加密令牌支付的激励能够吸引更多用户出售他们的计算资源。

协调不信任设备

人工智能在不受信任的设备上也有一个光明的未来，例如群体机器人、物联网设备或手机。在这些情况下，区块链将作为一个协调平台，攻击者只有占据多数块时才能破坏安全性。从冰箱软件更新到协调战术任务中的群体机器人都有它的用武之地。然而，不利的一面是它可以用于维护僵尸网络和恶意软件程序的命令和控制。目前，安全专家可以通过查找其集中控制中心立即关闭这些程序。如果他们通过区块链取代单点故障，除了逐一清除别无他法。

可解释的AI

尽管深度学习在构建能够感知、学习和自主行动的自治系统方面获得了很大成就，但在实际应用中却很少使用。一个重要的原因是机器学习技术，如深度学习，运行的时候是个难以准确

理解的黑箱。因此，这些系统做出的决策对人类来说无法解释，因而无法证实或信任。医学研究和财务规划领域的忧虑更深，它们对解释能力非常看重，因为这些领域的错误决策可能导致死亡或经济灾难。我们必须有一条不可变的线索来跟踪数据流的发展和AI系统的复杂行为。

好在，区块链可以跟踪数据处理和决策链中的每一个节点。通过跟踪AI系统在不同数据输入和应用场景中的行为，我们可以更好地理解和信任这些系统所做出的决策。更妙的是，人类将有一条清晰的线索来追溯机器决策过程，从而更易于解释这些决策的合理性。此外，它提供了调整这些黑箱的机会，人类专家可以调节其平衡性能和预测准确性与系统的可解释性。如果发生意外，这些基于区块链的线索对于确定人(以及具体是谁)还是机器犯错十分重要。

应用于区块链的AI技术

区块链的设计和操作涉及数千个参数以及安全性、性能、分散性和许多其他参数之间的权衡。AI可以简化这些决策，并且可以优化区块链，使之自动化，以实现更高的性能和更好的管理。此外，由于所有关于区块链的数据

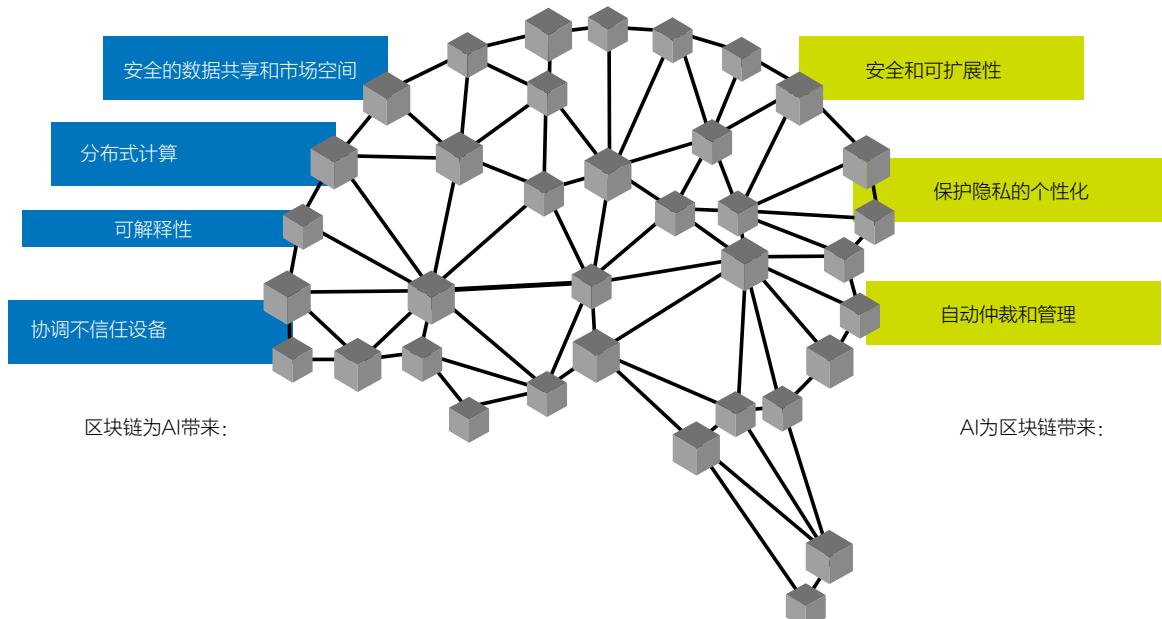


图 1. AI 和区块链融合：区块链在 AI 中的应用和 AI 在区块链中的应用。

都公开，因此AI在为用户提供私密性和隐私方面起着关键作用。

安全性和可扩展性

除非对手拥有多数采矿权，否则区块链几乎不可能被黑客攻击。不幸的是，构建在区块链平台之上的应用程序和功能并不安全。例如，分散的自治组织（DAO）是拥有1.5亿美元加密货币（称为以太币）的最大众筹集团之一，是被盗窃了5000万美元的受害者。黑客利用了写智能合同时犯下的几个错误，这些错误允许重复交易，以取出比基金投入更多的金额。随着机器学习取得了令人难以置信的进步，由智能机器学习算法控制的区块链或许能够检测到攻击的存在并自动调用适当的防御机制。当损失不可避免时，AI至少可以将受到攻击的组件与区块链平台隔离开，使其他组件免受攻击。类似的AI可

用于管理区块链，使其更具可扩展性和可靠性。例如，当交易数量激增时，AI可以增加块创建率，这将以更长的确认时间为代价来增加流通量。

隐私和个性化

如果你担心下一次选举是否会受到损害，或者你的数据在社交网络上是否安全，那么区块链就是你的选择。区块链可以让你重新控制你的个人数据。但它也有代价。在传统的中心化设定中，Facebook、Netflix、YouTube等公司分析收集的用户信息以便为用户提供个性化内容。这就是你为什么可以去Facebook寻找与你最常互动的朋友的帖子，或者登录Netflix获得符合你的品味的电影。收回你的隐私意味着没有人知道你喜欢什么，所以您可能不得不滚动浏览许多页面以查找相关内容，而无法自动定制。但有没有办法可以兼顾隐

私和个人化体验呢？

AI通过新的内容选择模型来解决这个问题。分散的内容提供商（例如区块链上的社交网络）可以在用户侧利用AI来使内容个性化。机器学习程序将在用户的设备上运行，以分析他们的浏览行为和爱好。用户的相关内容将被拉出，而不是推送，并显示给用户。请注意，整个计算是在本地执行的，没有任何个人数据离开用户的设备。此外，用户可以清除个人偏好内容，以防止内容提供者对用户进行分析。因此，这种新的基于牵引的模型将可以同时提供隐私和个人化。

机器的兴起

许多电影都预测了未来机器将会崛起并控制人类生活的方方面面。那样的未来，或者至少是其中的一部分已经存在了。智能合约（一种语言）已经

关于作者

部署在区块链上，用于解决用户之间的合同，决定谁将获得报酬。然而，就目前而言，智能合约并非如此“聪明”，仅限于简单的合约形式。

随着机器学习的快速发展，AI将能够裁判更复杂的情况。用户最终无需前往仲裁室就能解决链内（甚至是脱链）纠纷。在提供证据和文件后，AI可以以无偏见和防篡改的方式执行自动仲裁。所有决策都是数据驱动的，因此更加一致和合理。

尽管发展迅速，人工智能和区块链仍然有很长的路要走。关于AI，最新的发展是Google Duplex的发布，它可以代表你自动拨打电话。这听上去很棒，只不过应用场景仅限于餐厅预订、美发沙龙预约和假期。尽管人工智能近期在算法、算力和数据上都取得了巨大进步，但它依然无法跟上人类世界的无限复杂性。对于区块链，最近使用Etherum、ZCash、BitcoinGold和其他加密货币的安全事件表明，我们还需要一段时间才能享受到实际应用的安全和可扩展区块链。

放眼未来，区块链和AI之间的合作将为我们的社会提供无限的创新和变革，也许只有消亡才能使它们分开。■

参考文献

1. Y. Leviathan, “Google Duplex: An AI System for Accomplishing Real-

升廷 (Thang N. Dinh) 是弗吉尼亚联邦大学计算机科学系的助理教授；区块链跨学科研究实验室Lab-B联合主任；分散式应用程序的生态系统Fractal Platform的共同创始人。他的研究兴趣包括区块链优化、网络科学和安全。Dinh在佛罗里达大学获得计算机工程博士学位。请通过tndinh@vcu.edu与他联系。

梅泰 (My T. Thai)，是佛罗里达大学计算机与信息科学与工程系的UF研究基金会教授。她的研究兴趣包括可扩展算法、大数据分析、网络安全以及网络科学和工程中的优化，包括通信网络、智能电网、社交网络及其相互依赖性。她已经出版了六本专著，并发表了超过140篇顶尖期刊和会议文章，包括2014年IEEE MSN最佳论文奖和2017年IEEE ICDM最佳论文奖。Thai曾担任许多IEEE会议的TPC主席；担任Journal of Combinatorial Optimization (JOCO)、IEEE Transactions on Parallel and Distributed Systems和IEEE Transactions on Network Science and Engineering的副主编；以及Springer Briefs in Optimization的系列编辑。她是Computational Social Networks期刊的创始主编，并获得了许多研究奖项，包括UF Provost的助理教授卓越奖，UFRF教授奖，国防部 (DoD) 青年研究员奖和NSF CAREER奖。Thai在明尼苏达大学获得计算机科学博士学位。请通过mythai@cise.ufl.edu与她联系。

- World Tasks Over the Phone,” 8 May 2018, Google AI Blog; <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>.
2. A. Hern, “AI Used to Face-swap Hollywood Starts into Pornography Films,” 25. Jan 2018, The Guardian; www.theguardian.com/technology/2018/jan/25/ai-face-swap-pornography-emma-watson-scarlett-johansson-taylor-swift-daisy-ridley-sophie-turner-maisie-williams.
3. J. McGrath, “A.I. Will Make Smart Homes Autonomous, But Don’t Expect Perfection or Privacy,” 8 June 2018, Digital Trends; www.digitalerends.com/home/ai-will-make-the-smart-home-autonomous-but-dont-expect-perfection-or-privacy.
4. S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” 2008; <https://bitcoin.org/bitcoin.pdf>.
5. J.I. Wong, “Every Cryptocurrency’s Nightmare Scenario is Happening

- to Bitcoin Gold,” 24 May 2018, Quartz; <https://qz.com/1287701/bitcoin-golds-51-attack-is-every-cryptocurrency-nightmare-scenario>.
6. A. Kiayias et al., “Ouroboros: A Provably Secure Proof-of-Stake Blockchain Protocol,” Proc. 37th Ann. Int’l Cryptology Conf. (CRYPTO 17), 2017, pp. 357–388.
7. J. Chen and S. Micali, “Algorand: Scaling Byzantine Agreements for Cryptocurrencies,” arXiv preprint, 2016, [arXiv:1607.01341](https://arxiv.org/abs/1607.01341).
8. G. Wood, “Web3 Foundation,” 2017; <https://web3.foundation>.
9. T. Summers, “Facebook is Killing Democracy with Its Personality Profiling Data,” 21 Mar. 2018, The Conversation, <https://theconversation.com/facebook-is-killing-democracy-with-its-personality-profiling-data-93611>.
10. M. Mettler, “Blockchain Technology in Healthcare: The Revolution Starts Here,” Proc. IEEE 18th Int’l Conf. e-Health Networking, Applications and Services (Healthcom 16), 2016, doi:10.1109/HealthCom.2016.7749510.
11. A. Maxmen, “AI Researchers Embrace Bitcoin Technology to Share Medical Data,” 9 Mar. 2018 Nature; www.nature.com/articles/d41586-018-02641-7.
12. K. Gammon, “Experimenting with Blockchain: Can One Technology Boost Both Data Integrity and Patients’ Pocketbooks?,” Nature Medicine, vol. 24, 2018, pp. 378–381.
13. A. Pentland and E. Castello Ferrer, “Blockchain: A New Framework for Robotic Swarm Systems,” Media Lab Research MIT; www.media.mit.edu/projects/blockchain-a-new-framework-for-swarm-robotic-systems/overview.
14. N. De, “Hacks, Scams, and Attacks: Blockchain’s 2017 Disasters,” 29 Dec. 2017, coindesk; www.coindesk.com/hacks-scams-attacks-blockchains-biggest-2017-disasters.
15. Basic Attention Token (BAT): Blockchain Based Digital Advertising, white paper, Brave Software, 13 Mar. 2018; www.basicattentiontoken.org/ BasicAttentionTokenWhitePaper-4.pdf.
16. N. Lomas, “What Do AI and Block-chain Mean for the Rule of Law?,” 12 May 2018, Techcrunch, <https://techcrunch.com/2018/05/12/what-do-ai-and-blockchain-mean-for-the-rule-of-law>.
17. L. Fan and H.-S. Zhou, “A Scalable Proof-of-Stake Blockchain in the Open Setting (or, How to Mimic Nakamoto’s Design via Proof-of-Stake,” 2018; <https://eprint.iacr.org/2017/656.pdf>.

微信名：计算人
微信号：jisuanren





Emoji 表情： 普遍的通用语还是 一时的流行语？

文 | 乔治·F·赫尔伯特（George F. Hurlbert），STEMCorp, Inc.

杰弗里·沃阿斯（Jeffrey Voas），国家标准技术研究所（NIST）

译 | 吴岑，四川大学

许

多人对表情符号（以下简称emoji）的创新使用心生敬畏，其他人则嗤之以

鼻，认为正常的英语表达因此而消亡。还有人举双手拥抱这种借助emoji自由表达情绪的方式，同时也有人愤怒于如此简单潦草图像所代表的文化倒退现象。然后，大多数人则摆出事不关己的态度。

Emoji不是什么新鲜事物。简单的Emoji是一种图像符号，用图像的方式表示某一物体，或者是一种表意符号（抽象概念的符号表征），这种符号传承已久。有人还能证明符号的视觉化可追溯回史前岩洞绘画1。

然而，传统的部落规矩大部分都是口口相传。从大约公元前3200年开始，受过教育、精挑细选而出的抄写员在石板上铭刻下古埃及象形文字，描述贵族生平、辉煌战史以及神秘教条。与此同时，出现了苏美尔人的楔形文字，作为一种象形文字，历经数世纪，几经改变，最终成为了一种更具符号性质的表

达形式。汉字起源于公元前1200年左右，也是一种象形文字。大概在同一时候，早期的图像符号早于中美洲的阿兹特克文明出现，它的书写具有独特的说明风格。到了中世纪，接受过教育的僧侣写下泥金装饰手抄本，结合了符号视觉艺术效果和书写文字，在纸上保存下了宗教历史。混合视觉元素的画谜，还加上像emoji一样的插图，配上文字，往往以书谜的形式出现，也越来越受欢迎。

随着时间的推移，符号字母最终推动了印刷品的出现。一经印刷，线性的符号串迅速引领了全面的文化教育。1450年古登堡改革之后，知识变得可复制、可随身携带，也不可或缺。印刷术，这一线性概念得到了牛顿物理学的加持，最终导致了生产线的出现。工业经济紧随其后。终于，广播重新打开了口语空间，而电视重新激发了视觉活力。在相当短的时间内，人们的注意力从大批量生产转向大众媒体，这一切始于1900年，到了1999年伴随着互联网

的活跃达到顶峰。

后现代主义做出了回应，他们将在1900年代后期的消费主义评价为一种狡猾的现象。1963年，黄色笑脸别针一夜之间成为了文化偶像。1982年左右，表情符号，使用字体来表现人类的脸部表情，风靡一时。这些字体组合将情绪传达入枯燥的文本中。受到日本文字的影响，栗田穰崇（Shigetaka Kurita）在1999年第一次创造了emoji。它在网络上迅速传播开。图表1大致描述了人类交流中视觉语言源远流长的历史，一直发展到现代的漫画emoji。

尽管emoji拥有漫长的传承史，他们也很明显是数码时代的产物。今年6月5日公布的统一码Emoji 11.0版本（Unicod Emoji Version 11.0）中，它们大部分被标准化为1644个符号（<http://unicode.org/emoji/charts/full-emoji-list.html>）这样，只要摁下快捷键都能快速打出表情符号，无需动手绘画。使用图像符号和表意符号，可以同时传达思想和情绪。Emoji甚至还



图1. emoji表情出现的概念性时间线

有松散的句法结构和语法规则。这是不是表明人们在某种程度上有能力，可能甚至用emoji文字，通过emoji进行有效地沟通呢？这就引出了下面一个问题：emoji有可能能为新的网络通用语吗？

一种新的表达形式？

有些人可能会赞同emoji正在成为马歇尔·麦克卢汉 (Marshall McLuhan) 口中“地球村”的全新通用语言。比如，牛津大辞典就宣布“笑哭了”这个表情符号是2015年全球年度表情符号。这同时承认了在冷酷无情科技统治的文本世界中，全球已经普遍接受emoji作为表情达意的流行方式，而为emoji设定统一码标准进一步巩固了其普遍性。推特、社交媒体、短信和其他数码媒体对内容量极大的压缩要求更加鼓励了节省字数来表达基本理念。尽管字母提供了有限的文字集来表述观点，而许多文字也只能以规定的

方式组合。144像素18字节大的emoji就很容易取代成本昂贵的文字，更加节省。

广告商们很快抓住了这种流行趋势，他们的目标客户就是那些爱用时髦表情符号发信息的网络用户。这种表情符号“货币化”的程度甚至扩展到服务经济中，雇主鼓励员工展现表情符号一样的笑脸，在一个情感麻木的世界中笑脸相迎顾客。Emoji作为经久不衰的视觉代码似乎保有着“持久力”。以下就是一些常用的的表情符号网站：

- Unicode Emoji Standard V 11.0 <https://unicode.org/emoji/charts/full-emoji-list.html>
- Real-Time Twitter Emoji Usage Tracker <http://emojitracker.com>
- Real-Time IOS Emoji Usage Tracker <http://www.emojistats.org>

- Emoji Encyclopedia <https://emojipedia.org>
- MIT NLP & AI-Based Sentiment Analysis <https://deepemoji.mit.edu>
- Popular Emoji Grams <https://emojisaurus.com>
- Personalized Emojis <https://www.bitmoji.com>
- Moby Dick in Emoji <http://www.emojidick.com>
- Worldwide Use of Emoji <http://nlp.ffzg.hr/data/emoji-atlas>
- Emoji Statistics <https://worldemojiday.com/statistics>

然而，许多emoji的内涵意义很容易被曲解。尽管统一字符标准定义了“核心emoji符号”，更多定义不明的emoji每天都持续出现在世界各地。很快，就会有一套科学的emoji问世。这

有点像巴别塔的情形，因为emoji往往依赖于不同的文化和语境。根据霍夫斯德文化指数(Hofstede Culture Index)定义，具有不同经济描述能力的文化更倾向于用不同的方式使用emoji表情，来表述他们与世界的特殊关系。举例来说，高不确定性规避指数国家的人们往往不喜欢表达积极情绪的emoji。此外，同样的emoji用在不同的文化中，可能会承载不同的涵义。

虽然统一码标准努力强化emoji的涵义，但是至少存在17个基于平台的专有字体，对于相同的统一码emoji表达不同的意义。统一码官方网站 <https://unicode.org/emoji/charts/full-emoji-list.html> 展示了11个标准emoji在不同平台的不同表达。因此，一个给定的标准统一码emoji在iOS平台上表达的内容和安卓平台上的不太一样。当特定的标准emoji跨平台出现时，就会导致其表情达意在统计学上的不同释义。尽管如此，释义的差异也会出现在同一平台内，当然差异程度也比较小。

emoji的释义缺乏相当普遍的共性，这表明emoji其实还不算是普遍的表达形式。如上所述，文化影响、上下文语境和符号变异可能损害内涵意义。更糟的是，emoji似乎不是一种完整的表达形式。

标准化的emoji中并不存在人称代词或者大多数不及物动词。这限制了语言的丰富表现力，但同时却在不同使用者的点击中打开了创造性的大门。实例为证，有志愿者使用亚马逊土耳其机器人(Amazon's Mechanical Turk)

将梅尔维尔(Melville)的整本《白鲸记》(Moby Dick)改编为emoji版，新书名为《Emoji版白鲸记》，书中第一句话“给我打电话，伊势麦尔”(Call me Ishmael)的emoji版本如下：



图2.《emoji版白鲸记》的第一句话

虽然这是一部很明显的时代小说，电话符号的使用(在小说中的那个时代电话还没有发明)产生一种语境讽刺感。同样地，《爱丽丝梦游仙境》(Alice in Wonderland)，作者刘易斯·卡罗尔(Lewis Carroll)的意图是创作一本解谜小说，也被完全翻译成emoji版。然而，在这两个实例中，成功地解开这些表情文本所必须的工作量超过了大部分读者的能力。尽管emoji数据集具有高度创意，但也变得高度主观性，诱导重复性，同时难以语境化。在其他情况下，多个emoji必须创造性地组合起来，来表达常见的事物。比如，“甜心”(sweetheart)可以由棒棒糖符号和心型符号组合起来，几乎是逐字翻译。



最终，emoji是技术导向的。因此，它受技术进步的驱动。当自然语言处理(NLP)和人工智能(AI)合力加强声音互动的效果，emoji可能让位于声音的

音调变化。此外，机器生成的emoji数量可能会压倒人类用户创造的数量，就像邮箱里面压倒性的垃圾邮件。这两种趋势可能标志着emoji推崇者会遇到挫折。

Emoji作为新兴通用语言的概念似乎有所局限。然而，在一个不断要求符号经济和某种程度个性化的世界中，使用emoji作为表达的混合形式以扩充常规文本的可能性似乎会不断增强。配上其他非个人主观色彩的文本，选择性的使用emoji能定下愉悦交流的基调。Emoji使用友好的称呼、问候表情为结语、或者有策略地放置emoji来有意识地强调安抚语气，来消解可能被视为攻击性的信息。

网络时代的emoji

Emoji代表一种网络现象。2018年8月早期对在推特上使用来自网站<http://emojitracker.com>的emoji使用频率快照分析揭示了明显的动力曲线关系。图表3以垂直条形图的形式展示了这种关系。

有意思的是，大部分流行的emoji都表达一些积极正面的情绪，这符合大部分emoji使用研究的结果。其他研究则显示应用网络科学技术要优于最先进的方法，包括用于感情分析的NLP。

如图3所属，印刷术引入了盛行的线性关系，从而迎来工业时代的辉煌，增强了世界经济。大众媒体的来临，尤其是因特网，唤醒了其他感官。emoji作为视觉表达这一流行方法的崛起表明我们回归于长久引以为荣的视觉空间。

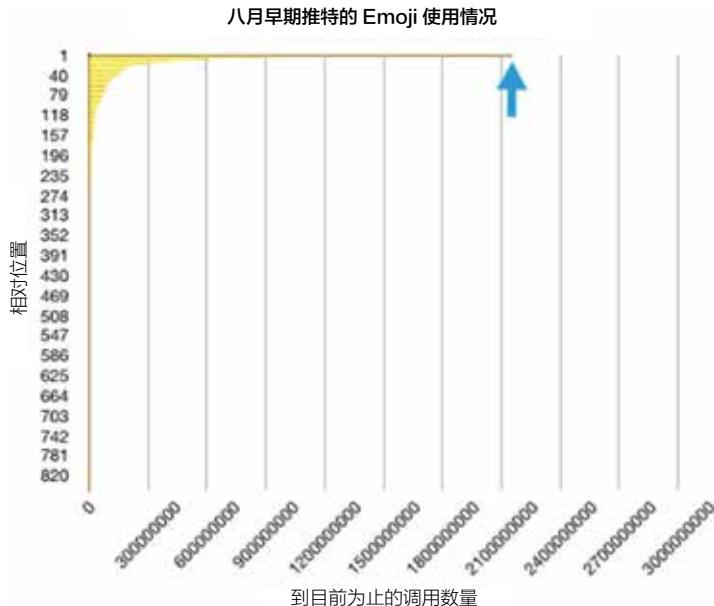


图 3. 2018 年 8 月早期 846 个流行 emoji 在推特上的分布情况



图 4. 2018 年八月推特上排名前十的 emoji

另外，尽管存在使用上的不同文化差异，全世界对 emoji 的普遍接受就是其本身存在的意义。这代表了接受网络全球共享这一事实的广泛趋势。沉浸在旧日线性技术模型中的许多人无法欣赏，甚至害怕这种开放性。对一些人来说，emoji 就是代表了倒退到原始人表达行为的悲剧状态。遵循这些态度观念的进一步社会研究可能会助长许多国家内日益增长的保护主义运动。

随着工业化的推进，注重文化普及的教育变得必不可少。现在，正规教育愈发地寻求创造性的线上教育方式，而传统地以课本识字为基础的教学似乎有些落伍。而且计算机能力不断地获取

人们的信任。更重要的是，正规高等教育不断增加的成本给许多已经负债不少的人设置了超越他们能力的入门门槛。也许是时候该承认，我们要从书本获取的便携式个人知识转向线上网络通用知识。这样的转变可能对未来的教育策略产生深远的影响。在这里，新的数码文化形式成为未来机遇的先决条件。不断增长地持续使用 emoji 可能表明，新的创新研究有助于确定新的教育载体，也许甚至能引申到数学中。

最后，视觉化是大势所趋。举例来说，大部分国家通过不同的形状和色彩线索来制定驾驶行为规范。Emoji 只是代表了数码世界中视觉化复苏的一种

形式。如上所言，增强现实和虚拟现实正在打开全新的感官大门。数据的图像表征也越来越迫切。将各种数据集视觉化的能力越来越成为一种关键技能。这些图像的概念化和构建要求新的数学思路和新的方法来精确且令人信服地描绘它们隐藏的现实。更重要的是，评估并阐释这些视觉表征的能力对于掌握信息的公民来说也同样重要。

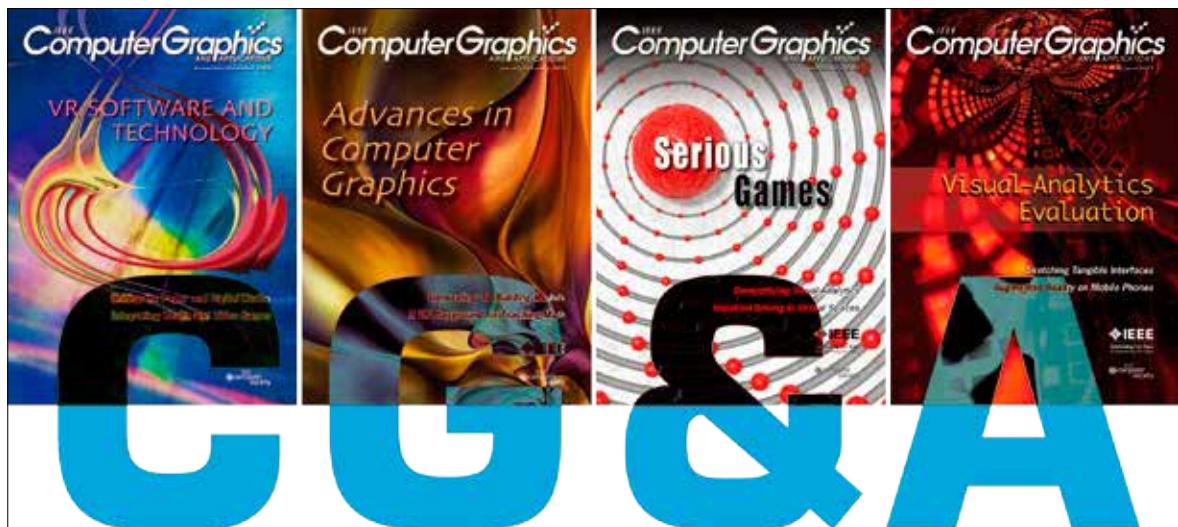
总而言之，掌握视觉能力，包括使用 emoji，成为了一种日益重要的技能，不仅仅是对数据科学家而言，而且涉及几乎所有越发交织在一起的人类知识领域。■

参考文献

1. G. Hurlburt and J. Voas, "Storytelling: From Cave Art to Digital Media," *IT Professional*, vol. 13, no. 7, 2011, pp. 4-7.
2. M. Dansi, *The Semiotics of Emoji*, Bloomsbury Academic, 2017.
3. L. Stark and K. Crawford, "The Conservatism of Emoji: Work, Affect and Communication," *Social Media + Society*, vol. 1, no. 2, 2015; doi.org/10.1177/2056305115604853.
4. X. Lu et al., "Learning from the Ubiquitous Language: an Empirical Analysis of Emoji Usage of Smartphone Users," *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, 2015, pp. 770-780.

5. H. Miller, J. Thebault, and I. Johnson, "'Blissfully Happy' or 'Ready to Fight': Varying Interpretations of Emoji," 10th International Conference on Web and Social Media (ICWSM), 2016, pp. 259–268.
6. W. Radford et al., "'Call me Ishmael': How do you Translate Emojii?", Proceedings of Australasian Language Technology Association Workshop, 2016, pp. 150–154.
7. A. Illendula and R. Yedulla, "Learning Emoji Embedding using Emoji Co-occurrence Network Graph," International Workshop on Emoji Understanding and Applications in Social Media, pending publication, 2018; <https://arxiv.org/abs/1806.07785>.
8. T. McCafery and P.G. Mathews, "An Emoji Is Worth a Thousand Variables," The Mathematics Teacher, vol. 111, no. 2, October 2017, pp. 96–102.

乔治·赫尔伯特(George Hurlburt)是STEMCorp公司的首席科学家，借助网络科学致力于经济发展研究，推动自主技术成为可供人类使用的有用工具。他也参与到基于图像的动态物联网架构研究。赫尔伯特也是《IT专业》(IT professional)杂志的编委会成员，还隶属于南马里兰高等教育部中心(Southern Maryland Higher Education Center)董事会。他的联系邮箱：ghurlburt@change-index.com



《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

《麻省理工科技评论》中英双语APP

《麻省理工科技评论》中英文版APP是其全球范围内的首款APP，包含原版英文杂志上的全部内容，以及更多精选文章，全面整合多方数据源、产业信息、实验室情报，通过综合报道和分析拓展读者的知识纵深。

读者包括关注海外科技信息的产业、创投人士，以及海内外高等院校及研究所的学生及科研人员。



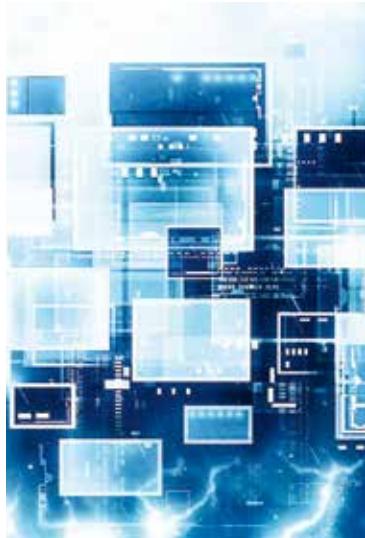
iOS和Android现已同步上线

年度订阅用户福利

- 每日科技英语单词学习卡片
- 每半月科技英语单词复习手册
- 每周科技英语直播讲堂，参与问答与互动
- 不定期更新原版科技知识、论坛独家视频
- 专属科技英语学习社区，不定期举办线下交流活动



扫码下载APP



量子计算机的扩展

文 | 埃里克·P·德贝内迪克特（Erik P. DeBenedictis），桑迪亚国家实验室（Sandia National Laboratories）
 特拉维斯·S·亨布尔（Travis S. Humble），橡树岭国家实验室（Oak Ridge National Laboratory）
 保罗·A·加尔吉尼（Paolo A. Gargini），IEEE 国际设备和系统路线图（IEEE International Roadmap for Devices and Systems）
 译 | 刘建，深圳大学

在晶体管出现几十年后，工程界开发出集成电路，从而利用大量晶体管制造出计算机。一个原初的发明开启了一个产业，其再投资的利润使晶体管的数量增加了一百万倍（这个增长原则现在被称为摩尔定律¹），这些晶体管构建的计算机则改变了世界。

几十年后，物理学家和数学家提出了一种名为量子比特的新型计算设备，似乎像当年的经典计算机一样，量子计算机散发着改变世界的诱人潜力——虽然我们尚不能证实这种潜力真的存在。研究人员已经创造出小规模的量子计算机原型，现在正与IEEE合作，制造标准尺寸的系统。

为了更好地理解当前量子计算机的扩展方向，让我们首先来看看晶体管是如何变成集成电路，而集成电路又如何演变成当今经典计算的庞然大物。在本专栏中，我们将量子计算机基准与半导体当年的发展过程相结合，展示了一种新的前进方向。

制造的扩展

在某些情况下，可扩展性是很容易判断的。图1a是一个平面集成半导体电路，图1b是一个集成量子比特电路。它们看起来几乎相同，所以我们可以假设它们都会根据摩尔定律进行扩展。

我们的假设很可能是真的，但事实远不止于此。图1c显示了一个早期的“飞线”（flying-wire）集成半导体电路，晶体管的集成方式和如今一样，但表面上方连接有金线。制造一个由百万级的晶体管集成的飞线电路，需要放置数百万根细小的金线。

图1d中的金枝形吊灯式构件，是量子计算机研究原型的支撑性结构，图中的量子比特芯片就安装在底部。枝形吊灯将低温冷却的量子计算机组件连接到各个不同层级的控制电子设备，然后连接到顶部的室温电子设备。将量子计算机从7个量子比特扩展到100万比特，需要等比例增加枝形吊灯的复杂性。想象一下，如果将飞线芯片或枝形

吊灯（的复杂度）按比例增加100,000倍，那将是多么巨大的机械结构，不仅很难制造，其中还会包含许多可能在运行过程中损坏的部件。很显然，这些东西无法扩展。然而，平面集成电路和量子比特芯片确实具有一种“整洁”的结构，看起来是可以扩展的；事实上，这些类型的结构还适合进行大规模生产。但是，量子计算机原型是“枝形吊灯”和量子比特芯片的综合体，这一结构因此无法扩展。目前所有的量子计算机的设计都不可扩展，因此我们需要更好的设计。

功能的扩展

制造图1a中的平面集成电路的生产线，与亨利·福特装配大型汽车的生产线类似；然而，还有另一种类型的可扩展性，这成为了计算机对世界产生巨大影响的关键。我们最初假设，每个晶体管的尺寸不会改变。也就是说，具有更复杂功能的芯片将会需要更多数量

的晶体管，这将占用更多的面积，从而导致每个晶圆中的芯片减少，且每个芯片将具有更多的缺陷。我们可以由此得出成本-复杂度曲线，如图2a所示，其中最优点位于每条曲线的底部。

然而，戈登·摩尔(Gordon Moore)1965年的文章¹报道了几代集成电路，这些集成电路的功能一代比一代复杂，但由于晶体管同时变小了，它们占据的芯片面积是相同的。在这种情况下，图2a中的成本-复杂度曲线同时朝着单组件的低成本和单芯片的多组件移动。摩尔根据这些假设预测，到1975年，晶圆的每个裸晶上的晶体管数量将呈指数增长至65,000个，如图2b所示。这种指数级增长现在被称为摩尔定律。

如果量子计算机的规模可以以指数级扩展，原因将会有所不同。量子计算机组件必须要非常精确地支撑量子纠缠，但量子纠缠可以在某些情况下提供指数级的量子加速；随着量子比特数增加，在处理某些特定问题上的性能也能呈指数级增长。如果每一代量子芯片都能比前一代增加一个量子比特，或者如果量子比特能每一代都比前一代更精确，量子计算机就可以指数级地扩展。

材料的影响

双极晶体管主导了最早的集成电路的制造，因为它们比金属氧化物半导体(MOS)晶体管更快、更小；但是，双极晶体管具有固有的扩展性问题。尽管经过了多次尝试，双极芯片中缺陷的数量仍然无法减少到MOS水平。这

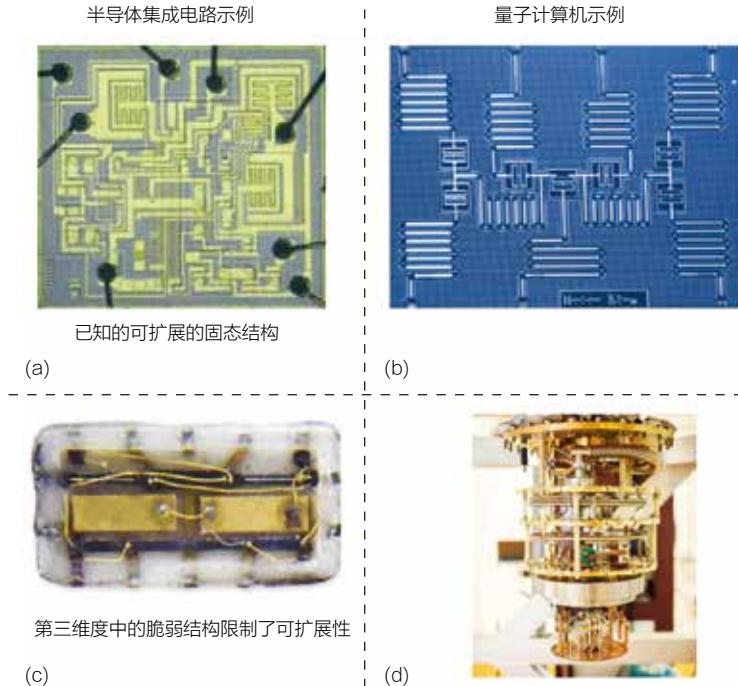


图1. 可扩展(a)半导体电路(约1971年的平面集成电路[555定时器]; https://en.wikipedia.org/wiki/555_timer_IC)和(b)量子比特集成电路(枝形吊灯底部是IBM 7的量子比特芯片); <https://spectrum.ieee.org/tech-talk/computing/hardware/tiny-quantum-computer-simulates-big-molecules>)。然而，量子比特集成电路只是量子计算机的一部分。(c)飞线集成电路(来源：马克·理查兹(Mark Richards)。由计算机历史博物馆提供)和(d)量子计算机枝形吊灯(稀释制冷机冷却阶段); <https://spectrum.ieee.org/computing/hardware/europe-will-spend-1-billion-to-turn-quantum-physics-into-quantum-technology>)。由于第三维度的脆弱结构，这两者都不能在制造中扩展。因此，由于枝形吊灯结构，包括(b)和(d)的量子计算机不可扩展。

为MOS带来了市场优势：到1975年，MOS产品已经征服了芯片市场。

量子比特也有类似的材料问题。与双极和MOS晶体管之间的竞争相似，约瑟夫森结(Josephson junction，即超导传输子，transmon)

量子比特、量子点量子比特和其他类型的量子比特间存在竞争。所有量子比特都可能退相干，当工程化的量子比特与材料中的杂质相互作用时，信息会丢失。当杂质与工程量子比特之间的叠加太强时，杂质会变成额外的寄生量子比特并产生误差。

减少由材料杂质引起的退相干的

需求，类似于减少双极和MOS晶体管中的缺陷的需求。目前没有人清楚任何量子比特中的退相干可以有多大的减少，因此没有人知道哪种类型的量子比特最具长期的可扩展性。

重温提高半导体可扩展性的教训

如果将扩展半导体的过程应用于量子计算机，需要哪些条件？如上所述，该过程需要考虑制造环节、功能和材料。然而，还有其他问题，正如摩尔在一篇短文中详述道：¹

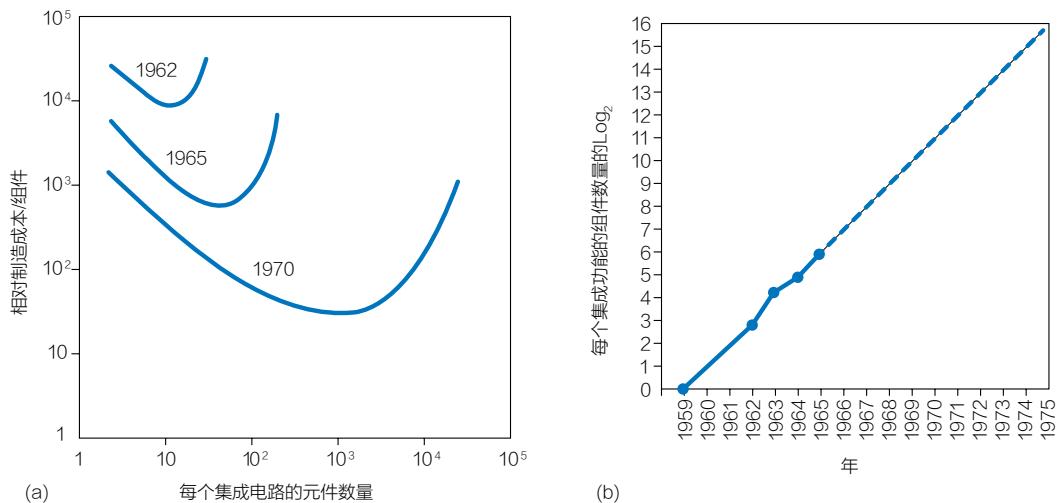


图 2. 来自(a)摩尔的文章¹中的数据,该文章支持(b)对未来几年的最佳芯片尺寸的推断。(a)中的图已经基本在公共语境中被遗忘,(b)中的图表成为了摩尔定律的代表。但是,在量子计算机被认为可扩展之前,需要重新评估这些数据——如果量子计算机注定是可扩展的话。

- > 成本
- > 可靠性
- > 材料(硅)
- > 产量
- > 复杂度
- > 裸晶尺寸
- > 互连空间
- > 热量
- > 速度
- > 单位面积功率
- > 设计自动化
- > 线性和射频

半导体和量子比特的设计目标的差异并不神秘。戴维·迪温琴佐 (David DiVincenzo) 提出的门式量子计算机 (gatetype quantum computers) 的目标,² 被量子计算界公认为以下的“迪温琴佐标准”:

1. 具有良好特性的量子比特的可扩展物理系统。
2. 能够将量子比特的状态初始化

为简单的同一基准态,例如| 000 ...>。
 3. 一套“通用”量子门。
 4. 比门的运算时间长得多的长相关退相干时间。
 5. 对特定量子比特进行测量的能力。

第一个迪温·琴佐 (DiVincenzo) 标准使用了“可扩展的物理系统” (scalable physical system) 这一短语,本质上是简单总结了摩尔标准,以及半导体行业为管理计算机系统的扩展现行而开发的庞大基础设施。我们建议将摩尔的清单与迪温·琴佐 (DiVincenzo) 标准相结合,为可扩展的量子计算机制定合适的设计目标。

量子计算机的基准

量子计算机的扩展工作在几年前已经非正式地开始了,但这却是 IEEE、

新兴的量子计算机行业以及美国政府的代表在2018年8月30-31日的一次更加正式的会议的议题。(会议提出的)那幅平滑的路线图认为,存储密度 (memory density)、时钟频率 (clock rate) 等指标不适合在量子计算机的早期阶段提出。相反,会议参与者为表1中所示的量子计算机基准开发了试验性矩阵,针对量子计算的潜在用途的基准和实现它们所需的工程水平的基准是不同的。当这张表格最终被填上具体的指标(其中一些还有待开发),并被有关各方所共同接受时,由此产生的基准将对量子计算机的扩展带来新的视野。

我们不知道量子计算机是否是为应用而生,是否能达成来自媒体和国会大厅的期望。当前的小规模原型机,与满足应用所需的数百万量子比特级的量子计算机之间存在很大差距。最具前景的方法,是开发和改进现有的量子比特设计,直到一种或多种方法变得可扩

表1.暂定的基准层级

基准水平	门式量子计算机	量子退火机	经典计算机（作为对照）
系统	有待开发	有待开发	时钟周期
门	量子体积，包括量子加速	指标有待开发	门延迟;不相关加速
设备	退相干时间	退相干时间	收益

展。世界上一些最大的公司正在相互竞争，并与资金充足的初创公司合作解决这些问题。然而，对任一单个的公司来说，开发大型量子计算机最终可能都太困难了。我们通过最近的会议认识到，IEEE的社群可以共同努力，提高获得成功的可能性。

致谢

桑迪亚国家实验室是一家多功能实验室，由霍尼韦尔国际股份有限公司 (Honeywell International, Inc.) 的全资子公司桑迪亚国家技术和工程解决方案有限责任公司 (National Technology and Engineering Solutions of Sandia, LLC) 在合同 DE-NA-0003525 的约定下为美国能源部国家核安全管理局 (US Department of Energy's National Nuclear Security Administration) 管理和运营。文章中表达的观点不一定代表美国能源部或美国政府的观点。本文由美国政府承包商根据合同号 DE-AC05-00OR22725 撰写。因此，出于美国政府目的，美国政府保留一份非独家、免版税的许可，可发布或复制本文的已出版形式，或允许其他人这样做。本文件部分总结了2018年8月30日至31日举行的会议，

参会者包括以下人员：保罗·奔克 (Paul Bunyk)，史蒂夫·布什 (Steve Bush)，乔什·柯布斯 (Josh Combes)，汤姆·康特 (Tom Conte)，马科斯·达·席尔瓦 (Marcus da Silva)，埃里克·德贝内迪克特斯 (Erik DeBenedictis)，戴维·弗格森 (David Ferguson)，保罗·加尔吉尼 (Paolo Gargini)，冉-卢克·高地奥特 (Jean-Luc Gaudiot)，李·戈麦斯 (Lee Gomes)，帕特·冈曼 (Pat Gumann)，斯科特·霍尔姆斯 (Scott Holmes)，特拉维斯·亨布尔 (Travis Humble)，阿兰·堪丁 (Alan Kadin)，斯科特·科佐伊尔 (Scott Kozoi)，布鲁斯·克里莫 (Bruce Kraemer)，诺伯特·林科 (Norbert Linke)，阿莱克斯·麦堪斯基 (Alex McCaskey)，凯西·麦吉奥赫 (Cathy McGeoch)，彼得·穆勒 (Peter Mueller)，欧乐戈·姆肯诺夫 (Oleg Mukanov)，保罗·内申 (Paul Nation)，萨泰亚弗鲁·帕特·劳 (Satyavolu Papa Rao)，博佳·佩罗帕鲁 (Borja Peropadre)，劳费尔·鲍塞 (Raphael Pooser)，莫伊奴迪·瓜尔西 (Moinuddin Qureshi)，马克·雷特 (Mark Ritter)，安德鲁·索恩伯格 (Andrew Sornborger)，约翰·斯帕格 (John Spargo)，希曼舒·太皮亚尔 (Himanshu Thapliyal)，比

尔·唐迪 (Bill Tonti)，艾力·特拉克 (Elie Track) 和卡尔·威廉姆斯 (Carl Williams)。

参考文献

1. G.E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, 1965, pp. 114–117.
2. D. DiVincenzo, "The Physical Implementation of Quantum Computation," *Fortschritte der Physik–Progress of Physics*, vol. 48.9, no. 11, 2000, pp. 771–783.

埃里克·P·德贝内迪克特斯是美国桑迪亚国家实验室计算研究中心的技术成员。他是 IEEE 高级会员，也是 ACM 和 APS 的会员。请通过 epdeben@sandia.gov 与他联系。

特拉维斯·S·亨布尔是美国橡树岭国家实验室的杰出科学家。他是 IEEE 高级会员。请通过 humblets@ornl.gov 与他联系。

保罗·A·加尔吉尼是 IEEE 国际设备和系统路线图的主席。请通过 paologargini@gmail.com 与他联系。



万尼瓦尔·布什： 虚拟计算与科学管 理的大师

文 | 乔治·斯特朗 (George Strawn)，美国国家科学院、工程院和医学院
译 | 吴岑，四川大学

在第二次世界大战期间以及战争结束后的一小段时间内，万尼瓦尔·布什 (Vannevar Bush, 1890-1974) 都可以称得上是“科学界的伟大前辈”。他的领域在于虚拟计算，这使他成为本专栏中的第一个非数字计算研究者（虽然本文将介绍他与之前讨论的一些计算大师之间的诸多联系）。职业生涯的成就使他的影响力远远超越了计算和工程领域，跻身科学界的最高领袖。我将简要回顾他的早年生活、他的虚拟计算机、他在战时的领袖作用，以及他在今天仍具影响力的两篇发表于1945年的文章。

如想了解万尼瓦尔·布什的更多生平，请查阅他的维基百科页面 (https://en.m.wikipedia.org/wiki/Vannevar_Bush)，如需更深入的信息，请参阅他的传记¹。

早年生平

布什是马萨诸塞州一位普世主义牧师的儿子，是三个孩子中最小的一

个。他与父亲同样就读于塔夫茨学院 (Tufts College)，是一名优秀且受欢迎的学生。1913年同时获得学士和硕士学位后，他在业界工作了几年，而后回到了研究生院。独立的精神贯穿了他的职业生涯，其表现之一，是他因为不想在论文导师为他指定的领域工作，而拒绝了克拉克大学 (Clark University) 的奖学金。随后他进入麻省理工学院，并于1916年发表了一篇电气工程的论文。

他在麻省理工学院的导师希望得到更多的结果，但布什向上级部门申诉，导师的要求因此被驳回。布什回到塔夫茨学院，开始了他在高等教育界的第一份工作，包括参与早期的无线电研究。当美国进入第一次世界大战时，他加入了美国国家研究委员会 (National Research Council)，这是威尔逊总统于1916年创立的美国国家科学院的“衍生品”。

1919年，布什回到麻省理工学院。他在麻省理工学院做出了诸多成就，其

中包含与人合作的一篇面向电气工程师的介绍性文章，以及他发明的多种电子设备。这其中的一种设备名叫S管，无线电因此不再需要两个电池，而是只要一个就够了。S管的商业化使布什变成了有钱人，生产S管的公司成为了后来的雷神公司 (Raytheon Company)，至今它仍然是一家大型的电子企业和国防承包商。

在另一个凸显布什独立精神的例子中，他与麻省理工学院院长就学校应该允许他做多少“外面的活”产生了争执。布什没有争赢，但因此与院长建立了良好的关系，后者任命他为麻省理工学院副院长兼工程系主任。

虚拟计算

虚拟计算的源头是计算尺 (Slide rule)，而数字计算的源头是算盘。比较年长的读者可能还记得计算尺：如果用长度代表数字的对数，则两个长度 (代表数字) 相加可得到这些数字的乘积。高等数学 (微积分) 定义了函数的微分

和积分运算，这有点类似于数字的乘法和除法运算。包含微分运算的方程称为微分方程，这在科学的研究中很常见。19世纪开始，人们开始尝试制造可以解决微分方程的机械模拟器件。1876年，开尔文勋爵和他的兄弟在英国取得了显著的进展。

1928年至1931年间，布什和麻省理工学院的同事制作了第一个通用的虚拟装置，利用轮盘机制（wheel-and-disc mechanisms）来进行积分，以解决一类重要的微分方程。尽管这基本上是一个机械装置，但这个被布什称做“差分分析仪”（differential analyzer）的东西包含了一些电路。他聘请了一名叫做克劳德·香农（Claude Shannon）²的研究生来操作它，而香农可能从布什的机器里得到了一些使用逻辑来表示电路的灵感。

由于差分分析仪的成功，全美乃至全世界随后涌现了许多类似的机器。宾夕法尼亚大学是制造出差分分析仪的机构之一，他们在随后的20世纪40年代开发出了数字ENIAC³计算机。一些观察家指出，ENIAC更类似于一个数字版本的差分分析仪，而不那么像现代计算机。第二次世界大战期间，布什负责战时科学研究，他在麻省理工学院的同事和知己诺贝特·维纳（Norbert Wiener）⁴建议制造一台数字计算机。许是因为专注于虚拟计算，布什并没有批准这项建议。

组织科研，支援战争

1939年，布什成为华盛顿卡耐基

研究所（Carnegie Institution）的所长，并开始在国家层面影响科学决策和方向。当他与董事会主席和其他领导官员争论并获胜时，他的独立性也立刻显现了出来。不幸的是，布什在社会科学或科学史上看不到什么价值，并且取消了卡耐基资金对前者的资助。大约在同一时间，他被调至国家航空咨询委员会（National Advisory Board for Aeronautics，NASA的前身），并帮助国会批准成立加利福尼亚的艾姆斯研

与发展的中心，是帮助盟军夺取胜利的关键因素之一。

美国参战之前，罗斯福总统创建了科学与研究与发展办公室（Office of Scientific Research and Development, OSRD），其预算基础比国防研究委员会更好，布什被任命为其主任。战争期间，OSRD资助了5亿美元的研发支出。支持的项目包括近炸引信（proximity fuse）的研发。一位历史学家说，“这个项目的规模和复杂程

布什在麻省理工学院与人合作了一篇面向电气工程师的介绍性文章，并发明了多种电子设备。

究中心（Ames Research Center）。但这是他的第二个盲区：布什并不认为导弹和火箭有可观的前景。

随着欧洲爆发战争，布什说服罗斯福总统设立国防研究委员会（National Defense Research Committee, NDRC），并由他自己掌舵。一位官员点出了这个组织（和布什）的重要性：“1940年夏天，如果谁的去世最能让美国陷入灾难，总统是第一个，布什博士可以排第二或第三。”¹但布什经常（带着一丝挫折感地）表示，如果说他为战争作出了什么贡献，那他也只是让陆军和海军告诉了对方彼此正在做什么。在与英国代表团协商后，他在麻省理工学院建立了无线电实验室（Radiation Laboratory，即广为人知的Rad实验室），这里成为了雷达研究

度，足以使其成为战时最杰出三或四个的科学成就之一”。曼哈顿计划也是布什在OSRD亲自重组的，目的是加快项目进度。

作为“伟大的科学前辈”，布什在战后仍然具有影响力，特别是他在1945年7月写的两篇文章（在欧洲战场结束之后但在亚洲战场结束之前）。其中一份报告是以OSRD主任的名义写给杜鲁门总统、题为《科学：无尽的疆界》（Science: The Endless Frontier）的报告。⁵该报告呼吁联邦政府像战时一样，在和平时期继续对科学的资助。

经过五年的政治操作，这一报告的结果之一，是政府立法创建了美国国家科学基金会（NSF），其定位是致力于支持科学的研究的非军事机构（资助对象最终还包括了社会科学、工程和科

学教育)。鉴于美国政府在此后的几年创建或拓展了许多其他的研究资助机构,如AEC(1946年)、DARPA(1958年)和NASA(1958年),布什实现了他的目标。然而,冷战的出现,使得政府的很多此类支持具有国防目的——尽管它们的成果时常以涓滴之势为社会做出了贡献(互联网就是一个很好的例子)。

布什的第二篇据有影响力的文章将他带回了信息技术领域。《我们可能会想到的》(*As We May Think*)⁶发表于《大西洋月刊》(*The Atlantic*),展现了信息在未来的重要性,充满远见卓识。他提出了一种(虚拟的)“记忆”设备,即Memex,从而预见了今天互联网实现的种种功能。他正确地预见了信息过载、以及科学家需要创造新的方法来吸收数百万期刊文章中的信息。包括道格拉斯·恩吉尔巴特(Douglas Engelbart)和泰德·尼尔森(Ted Nelson)在内的计算先驱都受到了这篇文章的启发。在《我们可能会想到的》发表六十年后,比尔·约瑟斯顿(Bill Johnston)和希拉·韦伯(Sheila Webber)回过头来说:⁷

布什的文章像是在描绘信息社会的一个缩影,其边界的设定与当时一位大科学家的兴趣和经验紧密相关,而与21世纪更开放的知识空间无甚关系。第二次世界大战对科学生产和技术应用产生了前所未有的要求,布什由此描绘了一幅“信息爆炸”的图景,指出了信

息对业界和科学界的重要性。他概述了一种图景,即信息科学是科学和技术知识领域的关键学科。他的观点囊括了信息过载的问题,也指出了我们需要设计有效的机制来控制和引导信息,以使信息变得可用。■

theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881.

7. B. Johnston and S. Webber, “As We May Think: Information Literacy as a Discipline for the Information Age,” *Research Strategies*, vol. 20, no. 3, 2005, pp. 108–121.

参考文献

1. G.P. Zachary, *Endless Frontier: Vannevar Bush, Engineer of the American Century*, MIT Press, 1999.
2. G. Strawn, “Claude Shannon: Mastermind of Information Theory,” *IT Professional*, vol. 16, no. 6, 2014, pp. 70–72.
3. G. Strawn, “Masterminds of the Electronic Digital Computer,” *IT Professional*, vol. 16, no. 2, 2014, pp. 10–12.
4. G. Strawn, “Norbert Wiener: Mastermind of Cybernetics,” *IT Professional*, vol. 19, no. 5, 2017, pp. 58–60.
5. V. Bush, *Science: The Endless Frontier: A Report to the President* by Vannevar Bush, Director of the Office of Scientific Research and Development, government report, United States Government Printing Office, 1945; <https://www.nsf.gov/about/history/vbush1945.htm>.
6. V. Bush, “As We May Think,” *The Atlantic*, 1945; <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881>.

乔治·斯特朗(George Strawn)是美国国家科学院、工程学院和医学院研究数据和信息委员会的主任。他是美国国家网络与信息技术研究与发展计划协调办公室(NITRD)的前任主任。他的联系方式是gostrawn@gmail.com。



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。



让有趣的学习 搞定智能手机一代

文 | 斯库特·威利斯 (Scooter Wills), TechGarge
译 | 谭坤, 复旦大学

学生已经准备好拥抱技术驱动的 21 世纪。如果我们让学习变得有趣并富有挑战性——然后让出路来——他们就会准备好学习技能、接替未来。



还记得你最初写的那个计算机程序么？那时候我八年级，身处无线电器材公司的商场里。在卖场的后部有TRS-80 III型计算机和有关基础的编程的书籍，顾客可以自由地摆弄。我着了迷。那时候父母完全不能理解我对计算机的痴迷，我必须承认现在，作为一个十几岁的孩子，我常常觉得很难接受我儿子花大量的时间在智能手机上使用社交媒体。不同之处在于在80年代，如果你想在计算机上做点有趣的事情，不得不自己写代码。现在只能手机应用和电子游戏已经让娱乐的艺术极为完美，

所有闲暇的技术时光都充满乐趣——你没法在智能手机、Xbox 或游戏机上编程。事实上，多数十几岁的孩子都已经不知道什么命令提示符或“IDE”（集成开发环境）是什么意思。但智能手机一代是交流的高手，随时准备和他人分享。

我的世界，这个10年前席卷世界的游戏，2014年被微软以25亿元的价格收购。游戏开发者发现玩家从与他人的互动中，获得了与建造自己的世界同样多的乐趣。我的世界利用红石提供了游戏内的编程体验，玩家可以利用可视化组件进行交流。我的世界还有一个Lua

编程语言插件，可以通过游戏内命令提示符实现与游戏的其他元素的交互。对典型的我的世界的玩家来说，游戏构建才是玩游戏的理由，他们只会在编程上花费很少的时间。在我主持的大型课外机器人工程TechGarage中，最快弄清楚四年级的学生是否会喜欢构建机器人的方法就是询问他们是否喜欢玩我的世界，看他们是否会笑起来。我的世界将会让建筑师的增加，也会增加土木工程师以及机械师的数量，因为在四年级的时候构建虚拟城市让这些年轻的头脑走上热爱建筑的道路。

TechGarage提供了为期四周的综

合性的夏令营项目，我们一直在项目中寻求让学习核心工程技术变得有趣的方法。在帮助学生在年龄足够小的时候找到技术热情的过程中，我们希望通过赋予学习以目来对他们的学校教育带来积极的影响。

五年前，我们引入了机器人大战，一个2001年由IBM开发的，基于Java的坦克游戏来教他们编程（图1）。本游戏由基于Java坦克API（应用程序界面）组成，这些API控制着坦克的所有原件以及拥有各种特点的坦克的集合。这些坦克——它们被冠以疯狂（Crazy）、角落（Corner）、烈火（Fire）和围墙（Walls）之类的名字——总是带着完整的源代码。这促进了学生们的学习，因为他们不得不学习每个机器人的代码。每周我们分配给由十五个学生组成的新队伍的任务就是给他们的坦克写代码，他们的坦克要与游戏里的坦克竞争，更重要的是，要与班里的其他学生竞争。Java里没有任何指示说明——我们只是简单地指导学生进行“电子游戏”，让他们自己研究如何编程。这些年来我们利用机器人大战将几百名学生带进了Java的世界，并一直为当意识到3:00要与同学进行一场对战时，他们能够学习得多么迅速而感到震惊。观看学生间双淘汰制锦标赛的决赛，就像看超级碗比赛的盛会，每个人都会为自己喜欢的队伍加油。如果你让学习变得快乐，剩下所需要做的就只是给他们让路，让乐趣开始。

两 年 以 前 ， 我 开 始 使 用
[CodeCombat.com](#)，这个分级电子游

戏不需要键盘与鼠标，在随着游戏级别升高而变得越来越复杂的过程中，玩家需要用Python或JavaScript写代码来调整角色。这些编程新手面临的挑战是在不存在系统管理员的情况下，安装和编码他们的第一个程序。[CodeCombat.com](#)是基于网络浏览器的，父母试图为他们的儿子或女儿设立账户是唯一的难题。解决的办法就是家长让他们的孩子自己注册账户。[CodeCombat.com](#)提供四种级别的计算机科学、游戏开发和网站设计。它已经成为被大学理事会认可的课程提供者，也是高级计算机科学原理（AP Computer Science Principles, AP CSP）的专业开发者。[CodeCombat.com](#)就集中体现了让学习变得有乐趣而给他们让路的思想。

技术引发了社会各个方面的变革。寻找教育的最优方式、让我们的孩子为技术驱动的世界作好准备最为重要。其中可能让人忧心的一个问题是，如何让来自不同社会经济背景的学生，都能参与到这个越来越被技术推动的世界中并投入工作。我们曾与那些支持高危社区的课后活动与夏令营的社区中心合作，在TechGarage里做有关高危接触模型的工作，去连接这些数字鸿沟。

在2018学年的最后，TechGarage和与其合作的社区中心，举办了三次为期两周的机器人夏令营项目，探索小学生与中学生在基于挑战的学习与机器人学上的表现。在夏令营的之前与之后，我们对那些此前从未参与过TechGarage项目的40名学生进行了有

关STEM(科学、技术、工程与数学) 的调查评估。

由里维埃拉海滩男孩女孩俱乐部（Boys and Girls of Riviera Beach）和德尔雷成就中心（Delray Achievement Center）提供的非正式调查结果显示，在那些自主选择参与TechGarage项目的新人中，90%的学生报告它们喜欢找出有的东西为什么出错了然后将其修复——这就是工程师的核心工作。85%的学生表示他们喜欢电子产品，这很可能是因为他们在很小的时候就接触了智能手机。我们的推測是，智能手机就代表了日常生活中的问题盒子，装满了待解决的问题。在夏令营开始的调查中，30%的参与者说他们对机器的工作原理有兴趣，50%的受访者则回答他们有时候会感兴趣。在两周的机器人夏令营结束后，有70%的学生都回答说对机器是如何工作有兴趣，这个显著的增长的人数来自于那些“有时候感兴趣”的学生，这一组下降到了15%。而“不感兴趣”的比例从20%下降到了12%。在有关焊接线路板调查中我们也看到了类似的改变：在最初的调查中，30%的学生说他们喜欢焊接线路板，在两周的夏令营中，他们利用我们给的电子线路板项目工具箱进行了操作，这个比例在夏令营后飞涨到了62%。这种浸入式的、基于挑战的对技术的更多的接触显著增加了他们进行更多学习的兴趣。

对于关注编程兴趣的调查问题，我们没有期待他们一开始就有强烈的兴趣。当被问及他们是否想编程和开发



图1. 对可解释AI模型(XAI)的概述。(a)现有AI技术,显示性能与可解释性。(b)据2016年DARPA报告的XAI解释框架。11(图片由DARPA提供。)

电子游戏时, 30% 的学生回答“是”。当被他们是否想成为机器人工程师时候, 也有30% 的受访者说了“是”; 在夏令营后调查中, 这一比例上升到了45%。当最初被问及是否想要一份赚大钱的工作时, 95%的人回答“是”, 5%的人回答“也许”。我们还询问学生是否喜欢利用Python编程: 30%的人回答“是”, 尽管对小组的非正式问卷调查显示他们并没有使用Python编程的经验。更重要的是, 在接触了CodeCombat.com两周之后(他们在视频游戏框架中学习了Python), “是”的回应跃升至62%, 这个变化源自于最初回答“不”的人改变了想法——从42%下降到14%。类似地, 在夏令营前62%的受访者对“我喜欢给机器人编程”这个问题做出了肯定的回答; 在调查后, 这个群体增加到了

78%。这些结果在很大程度上支持了我们的假设：学生对技术有浓厚的兴趣，接触技术学习模式会增加他们的热情和兴趣。

以教师为中心的方法强调死记硬背式的学习与记忆，这有让那些在传统学校中表现不佳的学生边缘化，走向低工资的未来的风险。随着2018年卡尔·D·帕金斯职业和技术教育改进法案最近获得国会几乎一致的支持和18亿美元的资助，我们必须认真研究让学生为技术职业做好准备的最佳方式。

在TechGarage，我们相信智能手机一代没有社会经济的隔阂——竞争环境已被有效地填平了。学生们对技术着迷：通过玩电子游戏，他们习惯于在早期和经常失败，通过观看YouTube视频快速学习，并且对直接记住方法没有兴趣。技术教育背景下基于挑战的学习

通过开发急需技能的合作和实践项目，挑战了现实世界。学生们已经为拥抱技术驱动的21世纪做好了准备，而许多学校仍在使用为20世纪的需求而开发的教学模式。如果我们能让学习变得有趣和有挑战性——然后让出路来——学生们会欣然拥抱他们的未来。■

斯库特·威利斯(Scooter Willis)是TechGarage的创始人，TechGarage是佛罗里达州棕榈滩县面向车中和高中生学生的一个集中式机器人工作空间。威利斯也是F1 Oncology公司的R&D生物信息学部分的高级主管。你可以通过willishf@gmail.com联系他。



会议就在你的手中

IEEE计算机协会的会议发布服务（CPS）现在可以提供组织会议的移动应用了！让会议的日程、会议信息和论文列表在你的与会者手中的设备上显示。

会议的移动应用可在安卓设备、iPhone、iPad和Kindle Fire上运行。



欲知更多信息，请联系cps@computer.org



致歉：在上期杂志中，因为编辑流程问题，导致刊登了旧版本。现特刊登正确的修改版本，并向文章作者李丹教授、耿金坤同学致歉。

第二届亚太地区网络研讨会 (APNet 2018) 全程纪实

文 | 耿金坤，清华大学
李丹，清华大学

亚太地区网络研讨会 (Asia-Pacific Workshop on Networking, 简称 APNet)，汇聚工作在业界第一线的网络研究者进行集中研讨，致力于为中国为主的亚太地区网络研究人员提供一个高水平的学术交流平台，为志在冲击顶级学术会议的未成熟的研究工作提供高质量的评审意见和现场反馈，为学术界和工业界越来越密切的联系提供一个高效的互动机会。2017年，第一届 APNet 在中国香港成功举办，并在学术界和工业界引发强烈反响。2018年8月2–3日，第二届 APNet 于中国北京再度举办。

A

PNet' 18历时两天，会议内容包括1篇Keynote主题演讲，10篇口头论文展示，11篇海报论文展示，5篇SIGCOMM论文演讲，以及6篇工业界技术报告。本次大会共有超过170人参会，参会者来自加州伯克利大学、宾夕法尼亚大学、明尼苏达大学、香港科技大学、清华大学、北京大学、中国科学院等国内外顶尖高校，以及微软亚研院、华为、Intel、腾讯、阿里巴巴、字节跳动等国内外知名企业。

主题报告

本届研讨会邀请到了明尼苏达大学双城分校的Zhi-li Zhang教授进行了关于“自运行网络”(Self-Running Network)的主题报告。近年来伴随着人工智能的兴起，网络领域的研究也开始转向与AI/ML技术的结合。张教授立足于网络和人工智能的交叉点，回顾了AI技术在网络领域的典型应用，包括基于AI的网络故障诊断，基于AI的网络搜索性能优化，基于AI的DNS故障分析，基于AI的拥塞控制等等。同时，张教授也指出了网络异构性已成为大势所趋，物联网终将成为未来互联网的发展必然。然而对于AI在网络底层架构方面的应用，目前还基本处于一个“黑盒”状态，仍然需要包含较多的人为因素。针对如何进一步将网络与AI相结合，构造可以自我驱动和自我运行的下一代互联网体系，张教授论述了自己的诸多思考和观点，并和与会者进行了深入交流。

论文研究热点

本届研讨会一共录用了10篇口头报告论文，大致聚焦在以下几个研究热点：

RDMA等新兴传输技术。 RDMA作为近年来网络领域广受关注的新型传输协议，吸引了众多学术界和工业界的研究人员。然而RDMA在大规模部署过程中还存在着诸多问题。硬件资源开销可能成为一个潜在瓶颈，使得单台

物理主机无法支持众多连接。本次研讨会的最佳论文奖授予了来自南京大学的论文“Toward Effective and Fair RDMA Resource Sharing”，该项目针对RDMA在公有云场景中的扩展性部署问题，提出了Avatar模型，一方面对于单台机器上有限的Queue Pair进行共享，从而可以支持更多的RDMA

强调了领域知识对于解释深度学习网络的重要意义。论文以HotNet'16的工作DeepRM为例，重点探讨了DNN在网络场景下的可解释性，即如何利用领域知识将一个黑盒的神经网路转换为白盒模式。该论文阐述了深度学习应用于网络场景中的一系列基本问题，引发了热烈讨论。

近年来，RDMA、GPU加速等新技术不断涌现 在网络领域，受到研究者的广泛关注。

连接，另一方面，设计合适的流量调度方案，保障公平性，避免饿死情况。此外，利用中断和轮询相结合的方式，降低了CPU的开销。Avatar可以对不同的RDMA应用（尤其是在公有云场景中）进行资源的有效管理，实现高扩展性和公平性。

网络与AI的结合。当前人工智能正在如火如荼地发展，并且逐渐蔓延到其他领域。网络与AI的结合在本次研讨会上被广泛地讨论。如本次研讨会的主题报告，由Zhi-li Zhang教授针对网络与AI的结合进行了回顾与展望；SIGCOMM报告中，来自腾讯的陈力工程师也报告了他们利用深度强化学习进行流量优化的工作AuTO。同样地，在本次研讨会录取的10篇论文中，来自复旦大学和芝加哥大学的合作论文“Demystifying Deep Learning in Networking”，初步讨论了深度学习与网络结合过程中的一系列问题和观点，

新型流量模型。流量的调度管理始终是网络领域面临的挑战难题。随着网络应用场景的日益复杂，现有的流量模型已经不足以满足实际需要。于是，一些新式的流量模型被提出，并吸引了众多学者的兴趣。本次研讨会收录了两篇关于新型流量模型的论文。来自香港科技大学和密歇根大学的合作论文“Pas de deux: Shape the Circuits, and Shape the Apps too! ”，重塑了coflow的流量需求矩阵，从而可以匹配回路场景的流量模式，实现高效的流调度。来自南京大学的论文“Using the Macroflow Abstraction to Minimize Machine Slot-time Spent on Networking in Hadoop”，则是针对Hadoop相关场景下的流量特征提出了新型模型Macroflow，通过高效的启发式策略对于工作任务进行调度，从而大大缩短任务完成时间。这些新型流量模型可以更好地近似真实场景中的



图1. 茲位工业界技术专家就关键问题发表见解。

流量特征，从而可以对数据中心中的资源进行更加高效地管理。

其他创新性工作。本次研讨会上其他一些创新性工作也受到了与会者的关注和讨论。来自香港科技大学、微软亚研院和华为的合作论文“Augmenting Proactive Congestion Control with Aeolus”，提出了拥塞控制的新型解决方案Aeolus，能够充分利用前瞻性拥塞控制方案的优良特性，同时消除额外延迟，实现带宽资源的高效利用。来自阿卜杜拉国王科技大学和瑞典皇家理工学院的合作论文“Prelude: Ensuring Inter-Domain Loop-Freedom in SDN-Enabled Networks”，设计了Prelude原型方

案，可以有效检测出SDN中的由于原始路由协议缺陷导致的转发循环，同时避免私有路由信息的泄露。来自清华大学和湘潭大学的合作论文“Dante: Enabling FOV-Aware Adaptive FEC Coding for 360-Degree Video Streaming”，针对360度视频传输场景设计了系统原型Dante，融合冗余编码技术和优先级传输策略，从而可以提高视频服务质量用户体验。来自宾夕法尼亚大学和天普大学的合作论文“Towards Example-Guided Network Synthesis”，设计提出了新型的网络工具Facon，降低了新型领域特定语言(Domain-Specific Language)的使用门槛，可以利用输

入-输出样例快速产生相关应用程序。来自清华大学和普林斯顿大学的合作论文“GEN: A GPU-Accelerated Elastic Framework for NFV”，针对CPU在NFV场景下的低效问题，提出GEN框架，利用GPU加速NFV，实现高性能和弹性扩展。来自清华大学、普林斯顿大学、南洋理工大学和加州大学伯克利分校的合作论文“Preserving Privacy at IXPs”，基于Intel SGX实现路由服务器，可以有效保护信息隐私性，同时显著降低性能损失。

此外，本次大会还录用11篇海报展示论文，内容涉及区块链、机器学习框架、哈希算法等相关领域。相关研究工作的展示人员和与会者热烈讨论交

流，会场气氛热烈。

工业界报告

本届研讨会邀请了来自字节跳动、微软、腾讯、Intel、阿里巴巴、华为等国内外知名企业的专家进行了专题报告。研讨会还设置了工业界专家的Panel环节，采用座谈形式，由六位工业界专家对于业界普遍关注的问题发表了个人观点。问题涉及网络的诸多宏观领域，包括“云计算与边缘计算中的网络挑战”，“网络和AI的作用关系”，“智能网卡的发展”，“网络低时延的解决方案”等等。其中主要观点总结如下：

未来网络领域研究的基本问题。来自阿里巴巴的主任工程师Ming Zhang将网络领域的基本问题划分为四个主要部分：连接（Connectivity），带宽（Bandwidth），延迟（Latency）和安全（Security）。关于连接，在未来10年左右的时间里，网络连接形式可能会出现巨大的变革，对于下一代的网络连接形式研究，是一个值得探索的问题。关于带宽资源，目前对于数据中心中资源的管理，普遍的解决方案是将充足的带宽资源供给终端，然而带宽资源不会无限增长，传统的摩尔定律在不远的将来会面临失效，如何突破带宽的瓶颈，为主机提供甚至超过百G的带宽资源，需要进一步的研究。关于延迟，对于大部分网络服务来说，低延迟是一个普遍的需求，如何降低延迟，提供更加及时的服务，也是网络领域长期存在并

亟待解决的问题。关于安全，如何构建更加安全的网络通信防止嗅探和信息泄露，量子通信是否可以应用到实际的网络安全领域发挥作用？这将是一个宏大而有趣的课题。

可编程交换机和可编程网卡可能成为实现网络低时延的有效解决方案。正如Ming Zhang所述，时延问

具有广阔应用场景，但是问题依然存在。Layong认为，P4语言在硬件层和软件层之间扮演了中间件的角色，使得软件开发人员可以自定义硬件层面的处理逻辑。来自Intel的平台架构师Cunming Liang认为，芯片的智能化和定制化需要将相关的处理逻辑进行完整编译，对比其他脚本语言如Python，

今年的 APNet 会议继续展现出技术前沿、产学研结合、互动性强的特点。

题可以被认为是网络研究领域的一个基本问题。来自微软的软件工程经理Lihua Yuan同样强调了时延问题在云计算和边缘计算中的地位。Lihua认为，云计算和边缘计算中应用的大部分技术，均可以收敛到时延问题的解决方案上。如何有效的解决网络时延问题？来自腾讯的专家工程师Layong Luo认为，目前在大规模网络中的延迟问题大致可以分为端延迟和交换机处延迟。其中端延迟的根本原因在于Hypervisor对于包处理速度过慢，而这一问题恰恰可以由可编程的SmartNIC解决。而对于交换机处造成的网络延迟，根本原因在于网络拥塞和网络资源利用率低下，而利用可编程交换机和可编程网卡能够有效地解决网络拥塞问题，更加精确地进行包的路由转发，相比单纯的ECMP，可以有效降低排队延迟。

以P4语言驱动的可编程性芯片

JavaScript等，P4语言的优势在于成功构建了后端生态，使得开发人员可以根据统一的工业标准，编译部署相关的处理逻辑到硬件层面。然而，P4目前仍然没有做到生态系统的完全统一，不同厂商对于P4功能支持仍有较大出入，P4对于后端生态系统的一致性规范可能还需要较长的发展。

学术环境和工业环境下对网络领域挑战存在不同认知。来自字节跳动的AI实验室主任Chuanxiong Guo认为，学术领域的网络研究往往更为艰深困难，作为学术领域的研究人员，要善于把握领域发展趋势，将网络研究和当前热点紧密结合起来。Lihua认为，好的研究工作在于发现问题和解决问题，而不是局限于尽早发表论文，针对网络领域的问题给出一个比较通用的解决方案是非常有意义的事情。Ming Zhang认为，学术环境下的研究工作具有更高的自由度，而工业背景下的研究

工作则受限于各种实际限制，需要进行多个维度的权衡。作为学术领域的研究人员，最根本的一点要善于发现问题，并去努力解决真正有价值的问题。

SIGCOMM 专题报告

本届研讨会还邀请到了被 SIGCOMM 2018 录取论文的 5 篇中国论文的相关作者，对其各自的论文工作进行了演讲展示。在汇报的这五篇论文中，有三篇论文工作的内容和网络测量紧密相关：来自北京大学的杨全教授和来自中国科学院的黄群教授分别介绍了他们的工作 Elastic Sketch 和 SketchLearn，两者在结合网络流量特征的基础之上设计了高效的 Sketch 方案，以满足网络测量任务的需求。来自华为的李彤研究员介绍了他们在高铁网络场景下多径 TCP 的一些测量研究。此外，来自腾讯的陈力工程师汇报了他们采用深度强化学习进行数据中心流量优化的工作，来自西北大学的彭瑶教授报告了他们在物联网场景下设计的大规模节能方案。研讨会上同样采用座谈形式，在论文报告结束后，由 5 位 SIGCOMM 论文作者集中分享了自己冲击 SIGCOMM 的历程，并对如何作出 SIGCOMM 级别的研究工作给出了宝贵的建议。

专家点评

会议进行过程中，我们对于本次会议的组委会成员、字节跳动 AI 实验室主



图 2. Kai Chen 教授主持 SIGCOMM 作者 Panel。

在未来 10 年左右的时间里，网络连接形式可能 会出现巨大的变革。

任郭传雄博士进行了简短采访。结合本次研讨会的重点内容和网络领域研究的现状，郭博士分享了自己的观点：

关于大规模网络拓扑的应用场景。郭博士指出，目前业界在网络拓扑设计过程中更加关注交换机层面，因此 Fat-Tree 确实是目前最为广泛部署的拓扑结构，但是与此同时，一些经典的 Server-Centric 的拓扑结构，如 DCell, BCube 等，也存在非常大的应用空间，一些大型公司在其下一代网络解决方案中，也将一些 Server-Centric 的拓扑方案纳入进来。因此，郭博士认为，Server-Centric 的拓扑方案在未来依然具有重要的研究价值。

关于 RDMA 技术的辩证看待。郭博士此前一直致力于 RDMA 的推广与创新，RDMA 在高性能网络传输领域受到广泛关注，然而其问题也同样突出，对于 RDMA 的部署风险与性能收益，郭博士认为 RDMA 仍然是未来具有广阔前景的一项技术，虽然目前来看，RDMA 在部署稳定性、可编程性等方面仍然不如 TCP，但是这也正是 RDMA 技术的研究价值所在。Offloading 的思想在网络领域由来已久，RDMA 正在践行这一思想。在未来，一些基于 Offloading 的新兴技术，如 RDMA, FPGA 等，具有很广阔的应用前景。

关于字节跳动的网络领域研究。

郭博士对于字节跳动的网络技术领域进行了简要的介绍，字节跳动致力于AI领用的各种应用研究，大规模的AI应用同样需要高性能、可扩展的网络基础设施支持，因此，网络领域的研究，对于支持并加速AI应用具有重要意义。目前字节跳动的大规模机器学习平台也应用到了RDMA技术，在AI算法之外，AI加速也是具有巨大价值的研究领域。

最后，郭博士对于本届APNet也进行了简要的点评：时值酷暑假期，来自亚太地区高校和企业的众多参会者热情不减，这是对网络研究者的巨大鼓舞，在各位学术界和工业界同行的共同努力下，网络领域研究将会取得更大的

突破。

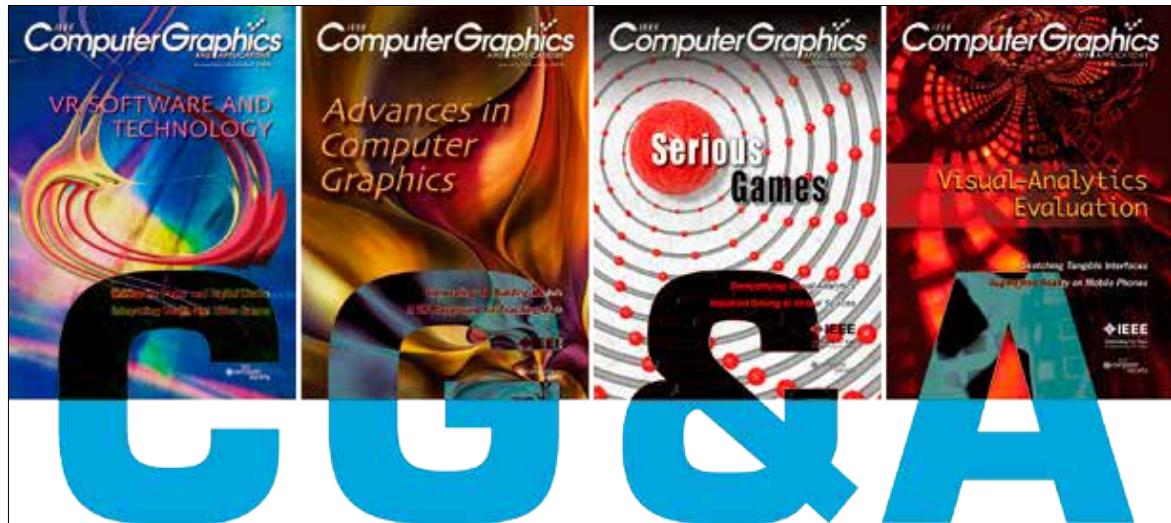
期待更多有志于网络研究的中国学者的加入！

展望

今年的APNet会议继续展现出技术前沿、产学研结合、互动性强的特点，汇聚来自世界各地的优秀学者和业界专家，针对网络领域的热点问题展开了广泛的研讨。时至今日，伴随着人工智能、区块链、软件定义网络、虚拟化等众多领域的兴起，网络研究也正在焕发出新的生机。网络场景的日益复杂也为今后的网络研究提出了更多的挑战。APNet 2019将于明年8月继续在北京举办，时间与SIGCOMM 2019临近。

耿金坤是清华大学计算机系硕士生，中国计算机学会学生会员。主要研究方向为高性能网络协议栈、数据中心网络资源分配、人工智能网络和大规模机器学习。邮箱：steam1994@163.com。

李丹，博士，清华大学计算机科学与技术系副教授，主要致力于计算机网络领域的研究，尤其是互联网和数据中心网络体系结构、系统与协议、AI加速与大规模机器学习等相关技术研究。



《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



KDD 2018 参会观感

文 | 罗宾理，中南大学

2018 年的 KDD 在伦敦举办，本次会议的很多数据都打破了 KDD 的历史纪录。

KDD 简介

ACM SIGKDD 国际会议（简称 KDD）是由 ACM 的数据挖掘及知识发现专委会主办的数据挖掘研究领域的顶级年会。它为学术界、企业界和政府部门的研究人员提供了学术交流和展示研究成果的非常好的平台，

其中涵盖了特邀主题演讲（Keynote Presentations）、论文口头报告（Oral Paper Presentations）、论文展板展示（Oral Paper Presentations）、研讨会（Workshop）、短期课程（Tutorials）、专题研讨会（Panels）、展览（Exhibits）、系统演示（Demonstration）、KDD CUP 赛事以及多个奖

项的颁发等诸多内容。由于 KDD 的交叉学科性和广泛应用性，其影响力越来越大，吸引了来自统计、机器学习、数据库、万维网、生物信息学、多媒体、自然语言处理、人机交互、社会网络计算、高性能计算及大数据挖掘等众多领域的专家、学者。

而 KDD（Knowledge Discovery

万美金；且收到1480篇投稿论文；对学生的奖励而言，KDD 2018的Student Award达15.3万美金。

颁奖情况

大会副主席唐杰主持了KDD 2018的论文投递情况及主要奖项。Research Track Papers共收到投递论文983篇，有104篇oral，录用率10.9%；74篇poster，录用率7.5%。69位Senior PC及512位Regular PC参与了审稿流程。在论文主题上，深度学习仍然占据主流，不过监督及无监督学习和迁移学习等领域呈现新高，论文数量达到了41篇。

根据审稿质量、参与讨论等多重标准，罗切斯特大学的罗杰波获得Senior PC的最佳审稿人奖项。KDD 2018的最佳学生论文，颁发给了Xiaoice Band: A Melody and Arrangement Generation Framework for Pop Music，流行音乐与深度学习的碰撞；而Research Track 最佳论文颁发给了Adversarial Attacks on Neural Networks for Graph Data，图数据是很多高影响力应用的核心，本文探讨了用于分类学习任务的深度学习框架是否真的很容易被欺骗/攻击，其结果的可信程度又在何种水平上。

而在Applied Data Science Track 中，Applied Data Science Track Papers共收到投递论文496篇，创下新高，其中有40篇oral，录用率8.0%；72篇poster，录用率14.5%。29位



and Data Mining, 知识发现与数据挖掘)一词首次出现在1989年8月举行的第11届国际联合人工智能学术会议(IJCAI)上，由Piatetsky-shapiro正式提出；1989-1994年间美国人工智能协会共举办了4届KDD国际专题讨论会(89、91、93、94)。1995年，国际KDD组委把专题讨论会更名为国际会议，并在加拿大蒙特利尔召开了第1届KDD国际学术会议，以后每年召开一次。1998年，ACM成立了KDD特殊兴趣组SIGKDD，于1999年第五届开始组织KDD学术会议。

新纪录

本届KDD在伦敦举办，伦敦也是

欧洲AI之都，其景色和历史文化底蕴让人十分震撼。纯净的天空，闲适的生活节奏，满满的英伦风情。沿着泰晤士河两岸，每一处落眼都是风情。

此次会议是KDD历史上的第24次会议，吸引了大量来自世界各地的学术界和工业界精英参加，并在全球范围内收录了293篇论文。KDD 2018大会主席、IBM的Faisal Farooq介绍了大会概况。本次KDD 2018大会共有4个keynote演讲、10个全天Workshops及10个半天的Workshops、两个panels、5个Hands-on Tutorials及29个Tutorials。Faisal Farooq总结了KDD2018的四个新高。其中共有来自99个国家的3377名学者注册了KDD 2018；而KDD 2018的赞助费用达120

Senior PC及167位Regular PC参与了审稿过程，最佳审稿人奖项由George Valkanas 获得，这一奖项经由SPC提名，审稿质量及相应讨论表现突出；其最佳学生论文和最佳论文分别颁发给了ActiveRemediation: The Search for Lead Pipes in Flint, Michigan、Real-time Personalization using Embeddings for Search Ranking at Airbnb。Keynote Speech。

KDD Cup 2018也是由国际计算机协会数据挖掘及知识发现专委员会(ACM SIGKDD)主办的数据挖掘竞赛，至今已举办21届，是目前数据挖掘领域最有影响力、最高水平的国际顶级赛事。每年都会吸引世界数据挖掘界的顶尖专家、学者、工程师、学生等前来参赛，被外界誉为大数据领域的“奥运会”。今年的KDD Cup基于预测伦敦及北京空气质量问题入手，组委会在比赛中提供中国北京和英国伦敦的数据。比赛选手需要预测未来48小时内PM2.5, PM10 和 O₃ 的浓度(在伦敦只需要预测 PM2.5 和 PM10)。今年比赛共有来自 4180 支团队的 5687 位参赛者参与，覆盖超过 3000 所大学/机构的49个国家。其中有世界著名的哈佛大学、麻省理工学院、剑桥大学和斯坦福大学参加，我国大陆地区也有清华大学、北京大学、复旦大学、中南大学、北京邮电大学等985高校参赛。我国中南大学和北京邮电大学的队伍First floor to eat Latiao 获得本次赛事的第一名。



Keynote

在KDD 2018的Keynote上，不得不提的就是由David Hand教授的“数据科学在金融中的应用”，金融行业与数据科学结合会碰撞出怎样的火花？

现在在国内就有非常火的“金融+数据”的概念，如风控领域，针对客户在金融交易中产生的庞杂数据的分析技术，不仅能够解决业内的模型精度问题，也能很大程度上推动数据挖掘和数据科学概念的发展。在David教授的演讲里，现在更是新模型类型以及新数据来源蓬勃发展的的时候，各种网络结构以及数据类型呈井喷状，这是当代数据科学的重大发展的大机遇。

David教授让我印象最深刻的是，他认为，有效市场假说在庞大的数据里会有些许偏差。“经典经济学中的「有效市场假说」认为金融市场具有不可预测性。这一说法虽然看起来很像真的，但本质上还是错误的。这意味着使用先进

的数据分析方法去发掘传统理论与实际现状间的细微差异成为可能。”其他的 data 科学问题，如数据质量、道德与安全，以及对模型局限性的了解的需求，在金融应用的情境中变得尤其突出。

KDD Cup 和 KDD China Workshop

而今年KDD Workshop的可选择程度非常高，从城市计算到Fintech，以及图对抗网络等非常前沿的选题。就KDD Cup的Workshop而言，先有今年的各个获奖队伍提出自己的Solution。Getmax队伍利用时序上的深度学习模型，已经能够较好地在空气质量预测中取得效果。而我们队伍(First Floor to eat latiao)主要采用了较为创新的特征模型：为了得到时间序列中隐藏的周期信息和波动信息，对空气质量，温度，湿度，气压等时间序列进行傅立叶变换，得到频域统计特征；对于过去

的空气质量，我们又将其分为两类。一类是没有方向的特征，包括压强、温度、湿度；另一类是有方向的特征，污染物浓度和风的扩散都是有方向的。为了建模这一类特征对空气质量的影响，我们将目标站点的方位划分为八个方向，分别考虑八个不同方向区域对目标站点的影响。对于每一个区域，提取其风速和污染物浓度的均值。

此外KDD Cup Workshop中间还穿插了Kaggle排名非常靠前的几位选手的分享。总体而言，这些选手们对整体框架的思考是我们值得学习的地方，对每一种题型都有一种体系化的解决方案，那这是不是也可以成为接下来AutoML的浪潮前驱者呢？

除了KDD Cup Workshop，KDD China Workshop也很精彩。KDD China主要是中国数据科学家成立的组织，每年会在国内组织一些活动，也会在KDD大会上组织一次Workshop。在Workshop上作报告的基本都是华人数据科学家。

字节跳动AI Lab科学家李磊博士报告了关于“机器写作与AI辅助创作”的内容。目前，人工智能在各类APP资讯平台都有交互场景，并且在内容创作及内容交互上有多种潜力，他认为，在未来的10年，新的技术尤其是人工智能技术将更多促进创作和交流，会去连接内容的创作者与消费者。

此外，清华大学副教授、本次KDD会议的副主席唐杰、香港科技大学的杨强、滴滴的叶杰平，以及罗格斯大学的熊辉也作了报告。

字节跳动的Reception

在8.20晚8点，我有幸受邀在Sunborn酒店里，参与了字节跳动的游艇晚宴。为了促进交流，活动现场甚至没有很多座位。所有参会者都可以领取食物和饮料，在场内随意走动和聊天。在活动开始的时候，字节跳动人工智能实验室的李磊简单介绍了公司技术的应用情况。ByteDance立足于前沿的深度学习技术，构建统一的超大规模推荐系统平台，每天从数亿用户的几百亿实时反馈动作中实时分析理解用户的兴趣，为今日头条、抖音、火山、西瓜视频、tiktok、NewsRepublic等国内外产品的用户做出最懂你的精准推荐。公司致力于在深度学习、数据挖掘、CV、NLP、ASR、信息检索等领域做最前沿研究，并落地到产品为用户带来价值。

对于那些热爱机器学习和数据挖掘的同学来说，字节跳动的吸引力在于丰富且异构的数据。例如，今日头条有图片和文本数据，火山、抖音有视频数据，所有这些加上海外的TopBuzz还有来自全世界的用户行为及内容推荐数据。所以，对于想尝试任何人工智能领域的同学，可能都能找到对应合适的产品和工作。

最后，李磊还发布了字节跳动组织的一个新比赛Byte Cup 2018。比赛使用TopBuzz上的文章作为数据，要求参赛选手开发模型，自动根据文章内容生成标题，并与人类编辑起的标题作比较（通过标准的Rouge评测指标计算人类编辑起的标题和自动生成标题之

间的距离）。字节跳动希望通过这种方式，开发一些可供行业使用的数据集，招募人才。

虽然现在比赛不少，但是文本摘要作为一个本身仍然很活跃的前沿学术领域，让Byte Cup和很多纯粹基于业务数据的比赛不同。即使对增加学术背景更感兴趣的同学，也可以通过这次比赛进入文本摘要这一领域，甚至有可能在比赛结束之后发表相关的论文。

现场的许多嘉宾踊跃的提出了业务中的问题，尤其是当外国友人使用了抖音之后，对效果感觉非常震撼，气氛幽默活跃。在晚宴里，我也向数据科学家李磊请教了自己的问题，包括今日头条的广告推荐算法，抖音的人像优化算法，AI-lab的前沿研究等；晚宴后，大家都有同感，字节跳动是一家业务很不错的公司，技术也是走在了业界的前沿。

总结

这次的KDD让我收获颇多，了解了工业界最前沿的应用动态，也透视了学术界目前的发展方向，印象颇深的还有郑宇老师的城市计算，在空气质量方面这是与我们的KDD-CUP的预测主题非常相关的，认识了许多志同道合的朋友，虽然运气较好拿到了今年的KDD CUP，但这绝不是终点，明年KDD再见。■

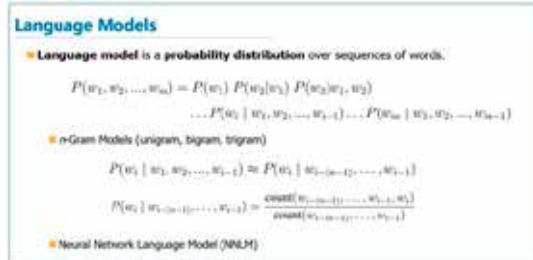
罗宾逊是中南大学软件学院二年级本科生，是本次KDD Cup冠军团队成员之一，曾获得工业AI大赛亚军，IJCAI季军。

自然语言处理中的语言模型预训练方法

文 | 罗凌，大连理工大学信息检索研究室
 译 | PaperWeekly (微信公众号: paperweekly)

在介绍论文之前，我将先简单介绍一些相关背景知识。首先是语言模型 (*Language Model*)，语言模型简单来说就是一串词序列的概率分布。具体来说，语言模型的作用是为一个长度为 m 的文本确定一个概率分布 P ，表示这段文本存在的可能性。

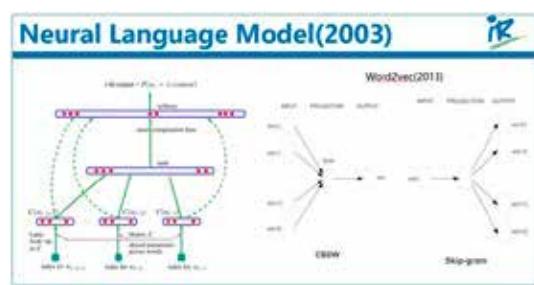
在 实践中，如果文本的长度较长， $P(w_i | w_1, w_2, \dots, w_{i-1})$ 的估算会非常困难。因此，研究者们提出使用一个简化模型： n 元模型 (n -gram model)。在 n 元模型中估算条件概率时，只需要对当前词的前 n 个词进行计算。在 n 元模型中，传统的方法一般采用频率计数的比例来估算 n 元条件概率。当 n 较大时，机会存在数据稀疏问题，导致估算结果不准确。因此，一般在百万词级别的语料中，一般也就用到三元模型。



为了缓解 n 元模型估算概率时遇到的数据稀疏问题，研究者们提出了神经网络语言模型。代表性工作是 Bengio 等人在 2003 年提出的神经网络语言模型，该语言模型使用了一个三层前馈神经网络来进行建模。其中有趣的发现了第一层参数，用做词表示不仅低维紧密，而且能够蕴涵语义，也就

为现在大家都用的词向量（例如 word2vec）打下了基础。

其实，语言模型就是根据上下文去预测下一个词是什么，这不需要人工标注语料，所以语言模型能够从无限制的大规模单语语料中，学习到丰富的语义知识。

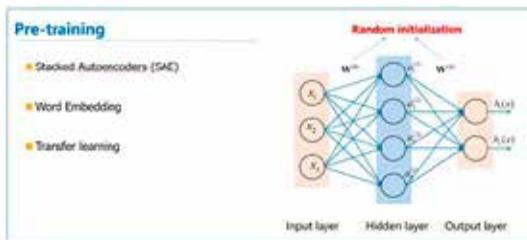


接下来再简单介绍一下预训练的思想。我们知道目前神经网络在进行训练的时候基本都是基于后向传播 (BP) 算法，通过对网络模型参数进行随机初始化，然后通过 BP 算法利用例如 SGD 这样的优化算法去优化模型参数。

那么预训练的思想就是，该模型的参数不再是随机初始化，而是先有一个任务进行训练得到一套模型参数，然后用这套参数对模型进行初始化，再进行训练。

其实早期的使用自编码器栈式搭建深度神经网络就是这个思想。还有词向量也可以看成是第一层 word embedding

进行了预训练，此外在基于神经网络的迁移学习中也大量用到了这个思想。



接下来，我们就具体看一下这几篇用语言模型进行预训练的工作。

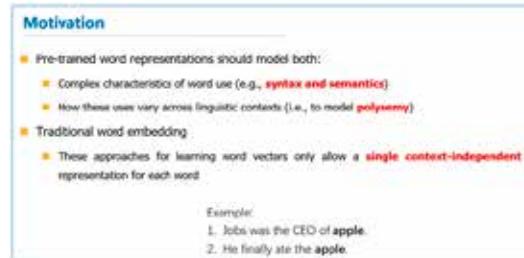
ELMo

引言

Deep Contextualized Word Representations¹这篇论文来自华盛顿大学的工作，最后是发表在今年的NAACL会议上，并获得了最佳论文。

其实这个工作的前身来自同一团队在ACL 2017发表的 Semi-supervised sequence tagging with bidirectional language models⁴，只是在这篇论文里，他们把模型更加通用化了。

首先我们来看看他们工作的动机，他们认为一个预训练的词表示应该能够包含丰富的句法和语义信息，并且能够对多义词进行建模。而传统的词向量（例如 word2vec）是上下文无关的。例如下面“apple”的例子，这两个“apple”根据上下文意思是不同的，但是在 word2vec 中，只有 apple 一个词向量，无法对一词多义进行建模。



所以他们利用语言模型来获得一个上下文相关的预训练表示，称为 ELMo，并在6个NLP任务上获得了提升。

Contribution

- Leveraging Language Modeling to get pre-trained contextualized representation.
- ELMo: Embeddings from Language Models
- Highlight:
 - The higher-level LSTM internal states capture context-dependent aspects of word meaning.
 - These representations can be easily added to existing models and significantly improve the state-of-the-art across six challenging NLP problems.

方法

在 ELMo 中，他们使用的是一个双向的LSTM语言模型，由一个前向和一个后向语言模型构成，目标函数就是取这两个方向语言模型的最大似然。

在预训练好这个语言模型之后，ELMo就是根据下面的公式来用作词表示，其实就是把这个双向语言模型的每一中间层进行一个求和。最简单的也可以使用最高层的表示来作为 ELMo。

Embeddings from Language Models

- ELMo is a task-specific combination of the intermediate layer representations in the bLSTM.
- For k-th token, L-layer Bi-directional Language Models computes $2L+1$ representations

$$h_k = \{x_k^{1,M}, \tilde{x}_k^{1,M}, \tilde{h}_k^{1,M} | j = 1, \dots, L\}$$

$$= \{h_{k,j}^{1,M} | j = 0, \dots, L\},$$
- For a specific downstream task, ELMo would want a weight to combine these representations. (In the simplest case, ELMo just selects the top layer $\tilde{h}(h_k) = h_{k,L}^{1,M}$)

$$\text{ELMo}_k^{\text{task}} = \tilde{h}(h_k; \Theta^{\text{task}}) = \gamma^{\text{task}} \sum_{j=0}^L \alpha_j^{\text{task}} h_{k,j}^{1,M}$$

scale parameter softmax-normalized weights

然后在进行有监督的NLP任务时，可以将 ELMo 直接当做特征拼接到具体任务模型的词向量输入或者是模型的最高层表示上。

总结一下，不像传统的词向量，每一个词只对应一个词向量，ELMo 利用预训练好的双向语言模型，然后根据具体输入从该语言模型中可以得到上下文依赖的当前词表示（对于不同上下文的同一个词的表示是不一样的），再当成特征加入到具体的NLP有监督模型里。

实验

这里我们简单看一下主要的实验，具体实验还需阅读论文。首先是整个模型效果的实验。他们在6个NLP任务上

进行了实验，首先根据目前每个任务搭建了不同的模型作为 **baseline**，然后加入ELMo，可以看到加入ELMo后6个任务都有所提升，平均大约能够提升2个百分点，并且最后的结果都超过了之前的先进结果（SOTA）。

SOTA				
Task	Previous SOTA		Our ELMo + Baseline	Increase (Absolute/ Relative)
	Our Baseline	ELMo + Baseline		
SQuAD	Liu et al. (2017)	84.4	81.1	85.8 4.7 / 24.9%
SNLI	Chen et al. (2017)	88.5	88.0	88.7 ± 0.17 0.7 / 5.8%
SRL	He et al. (2017)	81.7	81.4	84.6 3.2 / 17.2%
Coref	Lee et al. (2017)	67.2	67.2	70.4 3.2 / 9.8%
NER	Petrov et al. (2017)	98.93 ± 0.19	90.15	92.23 ± 0.10 2.06 / 21%
SST-5	McCann et al. (2017)	53.7	51.4	54.7 ± 0.5 2.3 / 6.8%

在下面的分析实验中，我们可以看到使用所有层的效果要比只使用最后一层作为ELMo的效果要好。在输入还是输出上面加ELMo效果好的问题上，并没有定论，不同的任务可能效果不一样。

Analysis				
Alternate layer weighting schemes				
Task	Baseline	Last Only	All layers 3x1	All layers 3x0.001
SQuAD	80.8	84.7	85.0	85.2
SNLI	88.1	89.1	89.3	89.5
SRL	81.0	84.1	84.6	84.8

A: regularize the ELMo weights to the loss.
Large values, such as $\lambda = 1$, effectively reduce the weighting function to a simple average over the layers, while smaller values (e.g., $\lambda = 0.001$) allow the layer weights to vary.

Where to include ELMo?				
Task	Input Only	Input & Output	Output Only	
SQuAD	85.1	85.6	84.8	
SNLI	88.9	89.5	88.7	
SRL	84.7	84.3	80.9	

Including ELMo at both the input and output layers for SNLI and SQuAD improves over just the input layers, but for SRL performance is highest when it is included at just the input layer.

OpenAI GPT

引言

我们来看看第二篇论文 *Improving Language Understanding by Generative Pre-Training*²，这是OpenAI团队前一段时间放出来的预印版论文。他们的目标是学习一个通用的表示，能够在大量任务上进行应用。

这篇论文的亮点主要在于，他们利用了Transformer网络代替了LSTM作为语言模型来更好的捕获长距离语言结构。然后在进行具体任务有监督微调时使用了语言模型作为附属任务训练目标。最后在12个NLP任务上进行了实验，9个任务获得了SOTA。

Contribution

- Their goal is to learn a **universal representation** that transfers with little adaptation to a wide range of tasks.
- Highlight:**
 - Use **transformer networks** instead of LSTM to achieve better capture long-term linguistic structure.
 - Include **auxiliary training objectives** (e.g. language modeling) in addition to the task objective when fine-tuning.
 - Demonstrate the effectiveness of the approach on a wider range of tasks. (significantly improving upon the state of the art in **9 out of the 12 tasks studied**)

方法

首先我们来看一下他们无监督预训练时的语言模型。他们仍然使用的是标准的语言模型目标函数，即通过前k个词预测当前词，但是在语言模型网络上，他们使用了Google团队在Attention is all you need论文中提出的Transformer解码器作为语言模型。

Transformer模型主要是利用自注意力（self-attention）机制的模型，这里我就不多进行介绍，大家可以看论文或者参考我之前的文章自然语言处理中的自注意力机制（Self-Attention Mechanism）。

Unsupervised pre-training

- Use a standard language modeling objective to maximize the following likelihood
$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \dots, u_{i-1}; \Theta)$$
- A multi-layer Transformer decoder for the language model
$$\begin{aligned} h_1 &= UW_e + W_p \\ h_i &= \text{transformer_block}(h_{i-1}) \forall i \in [1, n] \\ P(u) &= \text{softmax}(h_n W_e^T) \end{aligned}$$

然后在具体NLP任务有监督微调时，与ELMo当成特征的做法不同，OpenAI GPT不需要再重新对任务构建新的模型结构，而是直接在Transformer这个语言模型上的最后一层接上softmax作为任务输出层，然后再对这整个模型进行微调。他们还发现，如果使用语言模型作为辅助任务，能够提升有监督模型的泛化能力，并且能够加速收敛。

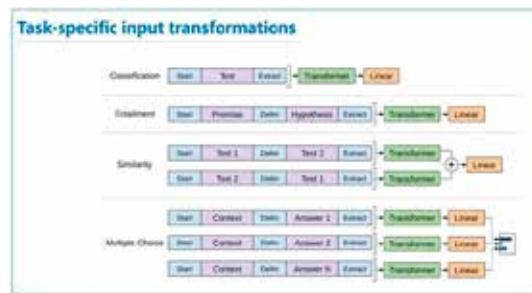
Supervised fine-tuning

- The final transformer block's activation is fed into an added linear output layer
$$P(y|x^1, \dots, x^m) = \text{softmax}(h_m^T W_y)$$
- objective function
$$L_2(\mathcal{C}) = \sum_{(x,y)} \log P(y|x^1, \dots, x^m)$$
- We additionally found that including language modeling as an auxiliary objective to the fine-tuning helped learning by (a) improving generalization of the supervised model, and (b) accelerating convergence.

$$L_3(\mathcal{C}) = L_2(\mathcal{C}) + \lambda * L_1(\mathcal{C})$$

由于不同NLP任务的输入有所不同，在Transformer模型的输入上针对不同NLP任务也有所不同。

具体如下图，对于分类任务直接讲文本输入即可；对于文本蕴含任务，需要将前提和假设用一个Delim分割向量拼接后进行输入；对于文本相似度任务，在两个方向上都使用Delim拼接后，进行输入；对于像问答多选择的任务，就是将每个答案和上下文进行拼接进行输入。



实验

下面我简单的列举了一下不同NLP任务上的实验结果。

语言推理任务：

Natural Language Inference								
Table 2: Experimental results on natural language inference tasks, comparing our model with current state-of-the-art methods. *x indicates an ensemble of 5 models. #x means accuracy as the evaluation metric.								
Method	MNLI-mm	MNLI-mm	SNLJ	SciTail	QNLI	RTE		
ENSM + ELMo [41] (#x)	-	-	89.3	-	-	-		
CAFE [38] (#x)	80.2	78.0*	89.2	-	-	-		
Stochastic Answer Network [21] (#x)	85.8	85.1	-	-	-	-		
CAFE [38]	78.7	77.9	88.5	81.3	-	-		
CoSeNet [40]	70.4	70.3	-	-	82.3	80.2		
Multi-task BiLSTM + Att [64]	72.2	72.1	-	-	82.1	85.7		
Fine-tuned Transformer LM (#ours)	82.8	88.4	89.5	88.3	88.1	86.0		

* RTE, one of the smaller datasets
Given the strong performance of our approach on larger NLP datasets.

问答和常识推理任务：

Question answering and commonsense reasoning				
Table 3: Results on question answering and commonsense reasoning, comparing our model with current state-of-the-art methods. *x means an ensemble of 9 models.				
Method	Story Cloze	RACE-m	RACE-b	RACE
val-LS-skip [55]	76.5	-	-	-
Hidden Cohesive Model [7]	77.6	-	-	-
Dynamic Fusion Net [67] (*x)	-	55.6	49.4	51.2
BiAttention MRU [59] (*x)	-	60.2	51.5	53.3
Fine-tuned Transformer LM (#ours)	86.5	62.9	57.4	59.0

This demonstrates the ability of our model to handle long-range contexts effectively.

语义相似度和分类任务：

Semantic Similarity and Classification

Table 4: Semantic similarity and classification results, comparing our model with current state-of-the-art methods. All task evaluations in this table are done using the GLUE benchmark. (#x = Matthews correlation, acc=accuracy, pcc=Pearson correlation)

Method	Classification		Semantic Similarity		GLUE	
	Cls-A	SST2	MRPC	STS-B	COPA	(F1)
Sparse byte nLSTM [16]	-	93.2	-	-	-	-
TF-ALD [23]	-	-	86.6	-	-	-
ECNU (mixed ensemble) [60]	-	-	-	-	81.0	-
Single-task BiLSTM + ELMo + Att [64]	35.0	90.2	80.2	55.3	66.1	64.8
Multi-task BiLSTM + ELMo + Att [64]	18.9	91.6	83.5	72.8	63.3	68.0
Fine-tuned Transformer LM (#ours)	45.4	91.3	82.3	82.8	78.3	72.8

可以看到在多项任务上，OpenAI GPT的效果要比ELMo的效果更好。从下面的消除实验来看，在去掉预训练部分后，所有任务都大幅下降，平均下降了14.8%，说明预训练很有效；在大数据集上使用语言模型作为附加任务的效果更好，小数据集不然；利用LSTM代替Transformer后，结果平均下降了5.6%，也体现了Transformer的性能。

Analysis

Table 5: Analysis of various model ablations on different tasks. Avg. score is a unweighted average of all the results. (acc=Matthews correlation, acc=accuracy, pcc=Pearson correlation)

Method	Avg. Score	MRPC score	SST2 score	MRPC (F1)	SST2 (F1)	COPA (F1)	STS-B (F1)	MRPC rank	SST2 rank	COPA rank	STS-B rank
Transformer w/o aux. obj.	70.7	43.4	91.3	81.3	10.1	70.9	81.8	48.1	74.0	71.2	71.2
Transformer w/o pre-training	70.8	43.6	91.6	81.4	10.2	71.0	81.9	48.2	74.2	71.3	71.3
Transformer w/o aux. obj. w/o pre-train	58.9	47.9	92.4	84.3	10.5	63.5	82.0	48.3	74.3	71.4	71.4

- Overall, the trend suggests that larger datasets benefit from the auxiliary objective but smaller datasets do not.
- A 5.6 average score drop when using the LSTM instead of the Transformer.
- The lack of pre-training hurts performance across all the tasks, resulting in a 14.8% decrease compared to our full model.

BERT

引言

上周Google放出了他们的语言模型预训练方法，瞬时受到了各界广泛关注，不少媒体公众号也进行了相应报道，那我们来看看这篇论文BERT:Pre-training of Deep Bidirectional Transformers for Language Understanding³。

Motivation

- Pre-trained language representations
 - Feature-based: ELMo
 - Fine-tuning: OpenAI GPT
- Both approaches share the same objective function during pre-training, where they use unidirectional language models to learn general language representations

这篇论文把预训练语言表示方法分为了基于特征的方法（代表ELMo）和基于微调的方法（代表OpenAI GPT）。而目前这两种方法在预训练时都是使用单向的语言模型来学习语言表示。

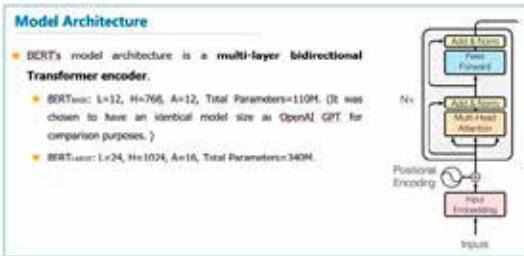
Contributions

- BERT: Bidirectional Encoder Representations from Transformers.
 - We demonstrate the importance of bidirectional pre-training for language representations.
 - Use Transformer encoder as the LM and a new pre-training objective: the “masked language model” (MLM).
 - Introduce a “next sentence prediction” task that jointly pre-trains text-pair representations.
 - They show that pre-trained representations eliminate the needs of many heavily-engineered task-specific architectures. BERT advances the state-of-the-art for eleven NLP tasks.

这篇论文中，作者们证明了使用双向的预训练效果更好。其实这篇论文方法的整体框架和GPT类似，是进一步的发展。具体的，BERT是使用Transformer的编码器来作为语言模型，在语言模型预训练的时候，提出了两个新的目标任务（即遮挡语言模型MLM和预测下一个句子的任务），最后在11个NLP任务上取得了SOTA。

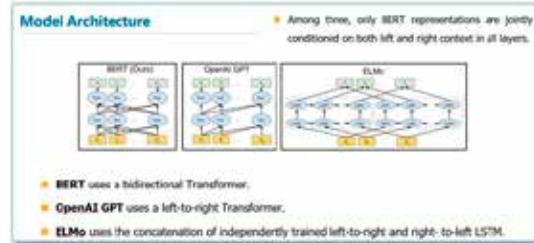
方法

在语言模型上，BERT使用的是Transformer编码器，并且设计了一个小一点的base结构和一个更大的网络结构。

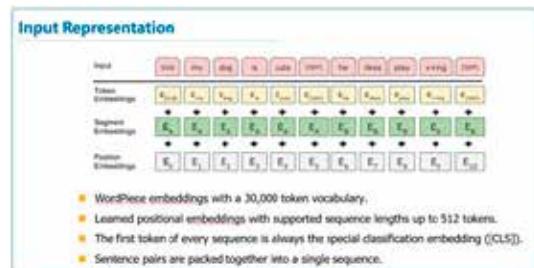


对比一下三种语言模型结构，BERT使用的是Transformer编码器，由于self-attention机制，所以模型上下层直接全部互相连接的。而OpenAI GPT使用的是Transformer编码器，它是一个需要从左到右的受限制的Transformer，而ELMo使用的是双向LSTM，虽然是双向的，但是也只是在两个单向的LSTM的最高层进行简单的拼接。

所以只有BERT是真正在模型所有层中是双向的。



而在模型的输入方面，BERT做了更多的细节，如下图。他们使用了WordPiece embedding作为词向量，并加入了位置向量和句子切分向量。此外，作者还在每一个文本输入前加入了一个CLS向量，后面会有这个向量作为具体的分类向量。



在语言模型预训练上，他们不再使用标准的从左到右预测下一个词作为目标任务，而是提出了两个新的任务。第一个任务他们称为MLM，即在输入的词序列中，随机的挡上15%的词，然后任务就是去预测挡上的这些词，可以看到相比传统的语言模型预测目标函数，MLM可以从任何方向去预测这些挡上的词，而不仅仅是单向的。

但是这样做会带来两个缺点：

1. 预训练用 [MASK] 提出挡住的词后，在微调阶段是没有 [MASK] 这个词的，所以会出现不匹配；
2. 预测 15% 的词而不是预测整个句子，使得预训练的收敛更慢。但是对于第二点，作者们觉得虽然是慢了，但是效果提升比较明显可以弥补。

Contribution

- Their goal is to learn a **universal representation** that transfers with little adaptation to a wide range of tasks.
- Highlight:**
 - Use **transformer networks** instead of LSTM to achieve better capture long-term linguistic structure.
 - Include **auxiliary training objectives** (e.g. language modeling) in addition to the task objective when fine-tuning.
 - Demonstrate the effectiveness of the approach on a wider range of tasks. (significantly improving upon the state of the art in **9 out of the 12 tasks** studied)

对于第一点他们采用了下面的技巧来缓解，即不是总是用 [MASK] 去替换挡住的词，在10%的时间用一个随机词取替换，10% 的时间就用这个词本身。

Pre-training Task1: Masked LM

- The [MASK] token is never seen during fine-tuning.
- To mitigate this, Rather than always replacing the chosen words with [MASK], the data generator will do the following:
 - 80% of the time: Replace the word with the [MASK] token, e.g., my dog is hairy → my dog is [MASK]
 - 10% of the time: Replace the word with a random word, e.g., my dog is hairy → my dog is azole
 - 10% of the time: Keep the word unchanged, e.g., my dog is hairy → my dog is hairy.

而对于传统语言模型，并没有对句子之间的关系进行考虑。为了让模型能够学习到句子之间的关系，作者们提出了第二个目标任务就是预测下一个句子。其实就是一个二元分类问题，50% 的时间，输入一个句子和下一个句子的拼接，分类标签是正例，而另 50% 是输入一个句子和非下一个随机句子的拼接，标签为负例。最后整个预训练的目标函数就是这两个任务的取和求似然。

Pre-training Task2: Next Sentence Prediction

- In order to train a model that understands sentence relationships, they pre-train a binarized next sentence prediction task.
- When choosing the sentences A and B for each pre-training example, 50% of the time B is the actual next sentence that follows A, and 50% of the time it is a random sentence from the corpus.

Input = [CLS] the man went to [MASK] store [SEP] he bought a gallon [MASK] milk [SEP]
Label = 1 (heit)

Input = [CLS] the man [MASK] to the store [SEP] penguin
[MASK] are flight wireless birds [SEP]
Label = 0 (heit)

在微调阶段，不同任务的模型如下图，只是在输入层和输出层有所区别，然后整个模型所有参数进行微调。

Fine-tuning Procedure

- All of the parameters are fine-tuned jointly to maximize the log-probability of the correct label.



实验

下面我们列出一下不同 NLP 上 BERT 的效果。

GLUE结果:

General Language Understanding Evaluation (GLUE) Test results

System	MNLI (exact)	OOP	QNLI	SST-2	CnLs	STS-B	MRPC	RTE	Average
pre-OpenAI SOTA	80.6/80.1	66.1	82.3	94.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.47/6.1	64.8	79.9	90.4	36.0	71.3	84.9	56.3	71.0
OpenAI GPT	82.0/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
BERT _{BASE}	84.6/83.8	71.2	90.1	93.5	52.3	83.8	88.9	66.4	76.6
BERT _{LARGE}	86.7/85.9	72.1	91.1	94.9	49.5	84.5	89.3	70.1	81.9

It is interesting to observe that BERT_{LARGE} significantly outperforms BERT_{BASE} across all tasks, even those with very little training data.

QA结果:

SQuAD results

System	Dev		Test	
	EM	F1	EM	F1
Leaderboard (Oct 9th, 2018)				
Human	-	-	82.3	91.2
#1 Ensemble - sntct	-	-	86.0	91.7
#2 Ensemble - QANet	-	-	84.5	90.5
#1 Single - sntct	-	-	83.5	89.0
#2 Single - QANet	-	-	82.5	89.3
Published				
BiDAF+ELMo (Single)	-	85.8	-	-
R.M. Reader (Single)	78.9	86.3	79.5	86.6
R.M. Reader (Ensemble)	81.2	87.9	82.1	88.5
Ours				
BERT _{BASE} (Single)	80.8	88.5	-	-
BERT _{LARGE} (Single)	84.1	90.9	-	-
BERT _{BASE} (Ensemble)	85.8	91.8	-	-
BERT _{LARGE} (Sgl.+TriviaQA)	84.2	91.1	85.1	91.8
BERT _{LARGE} (Ensl.+TriviaQA)	86.2	92.2	87.4	93.2

实体识别结果:

NER results

System	Dev F1	Test F1
ELMo+BiLSTM+CRF	95.7	92.2
CVT+Multi (Clark et al., 2018)	-	92.6
BERT _{BASE}	96.4	92.4
BERT _{LARGE}	96.6	92.8

Table 3: CoNLL-2003 Named Entity Recognition results. The hyperparameters were selected using the Dev set, and the reported Dev and Test scores are averaged over 5 random restarts using those hyperparameters.

SWAG结果:

Situations With Adversarial Generations (SWAG) results

Given a sentence from a video captioning dataset, the task is to decide among four choices the most plausible continuation. For example:

A squirrel is going across a sort of monkey bars... the squirrel jumps up across the monkey bars.
the squirrel jumps into the bars to grab her tail.
the squirrel runs on the bars and ends up in a monkey's pouch.
the squirrel jumps up and down a bunch of times.

System	Dev	Test
ESIM+GloVe	51.9	52.7
ESIM+ELMo	59.1	59.2
BERT _{BASE}	81.6	-
BERT _{LARGE}	86.6	86.3
Human (expert) [†]	-	85.0
Human (5 annotations) [†]	-	88.0

可以看到在这些所有NLP任务上，BERT都取得了SOTA，而且相比ELMo和GPT的效果提升还是比较大的。

在预训练实验分析上，可以看到本文提出的两个目标任务的作用还是很有效的，特别是在MLM这个目标任务上。

Effect of Pre-training Tasks						
Tasks	Dev Set					
	MNLI-m (Acc)	QNLI (Acc)	MRPC (Acc)	SST-2 (Acc)	SQuAD (F1)	
BERTBASE	84.4	88.8	86.7	92.7	88.5	
No NSP	83.9	84.9	86.5	92.6	87.9	
LR & No NSP	82.1	84.3	77.5	92.1	72.8	
+ BiLSTM	82.1	84.1	75.7	91.6	84.9	

● No NSP: A model which is trained using the "masked LM" (MLM) but without the "next sentence prediction" (NSP) task.
 ● LTR: A model which is trained using a Left-to-Right (LTR) LM.
 ● + BiLSTM: adding a randomly initialized BiLSTM on top of it for fine-tuning.

作者也做了模型规模的实验，大规模的模型效果更好，即使在小数据集上。

Effect of Model Size						
#L	#H	#A	Dev Set Accuracy			
			LM (ppf)	MNLI-m	MRPC	SST-2
3	768	12	5.84	77.9	79.8	88.4
6	768	3	5.24	80.6	82.2	90.7
6	768	12	4.68	81.9	84.8	91.3
12	768	12	3.99	84.4	86.7	92.9
12	1024	16	3.54	85.7	86.9	93.3
24	1024	16	3.23	86.6	87.8	93.7

Larger models lead to a strict accuracy improvement across all four datasets, even for MRPC which only has 3,600 labeled training examples.

此外，作者也做了像ELMo当成特征加入的实验，从下图可以看到，当成特征加入最好效果能达到96.1%和微调的96.4%差不多，说明BERT对于基于特征和基于微调这两种方法都是有效的。

Feature-based Approach with BERT	
Layers	Dev F1
First few AL	96.4
First Layer (Embeddings)	91.0
Second-to-Last Hidden	95.6
Last Hidden	94.9
Sum Last Four Hidden	95.9
Concat Last Four Hidden	96.1
Sum All 12 Layers	95.5

Table 7: Ablation study BERT with a feature-based approach on CoNLL-2003 NER. The activations from the specified layers are combined and fed into a two-layer BiLSTM, without backpropagation to BERT.

● This demonstrates that BERT is effective for both the fine-tuning and feature-based approaches.

总结

最后进行简单的总结，和传统的词向量相比，使用语言模型预训练其实可以看成是一个句子级别的上下文的词表示，它可以充分利用大规模的单语语料，并且可以对一词多义进行建模。

而且从后面两篇论文可以看到，通过大规模语料预训练后，使用统一的模型或者是当成特征直接加到一些简单模型上，对各种NLP任务都能取得不错的效果，说明很大程度上缓解了具体任务对模型结构的依赖。在目前很多评测上也都取得了SOTA，ELMo也提供了官网供大家使用。

但是这些方法在空间和时间复杂度上都比较高，特别是BERT，在论文中他们训练base版本需要在16个TGPU上，large版本需要在64个TPU上训练4天。对于一般条件，一个GPU训练的话，得用上1年。还有就是可以看出这些方法里面都存在很多工程细节，一些细节做得不好的话，效果也会大大折扣。■

参考文献

- Peters, M. E. et al. Deep contextualized word representations. naacl (2018).
- Radford, A. & Salimans, T. Improving Language Understanding by Generative Pre-Training. (2018).
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2018).
- Peters, M. E., Ammar, W., Bhagavatula, C. & Power, R. Semi-supervised sequence tagging with bidirectional language models. Acl (2017).

2018 BYTE CUP

International Machine Learning Competition

国际机器学习竞赛

字节跳动联手中国人工智能学会和IEEE联手举办大数据竞赛

BYTEDANCE, CAAI AND IEEE CHINA JOINTLY ORGANIZE THE DATA CONTEST

为 30 亿篇字节跳动产品文章生成标题

AUTOMATICALLY EXTRACT TITLE FOR 300 THOUSAND ARTICLES FROM BYTEDANCE'S PRODUCTS

总奖金金额

\$ 20,000

40名IEEE会员或学生会员资格

参赛网址 LAB.TOUTIAO.COM

7月31日: 比赛上线, 发布训练数据

10月31日: 发布测试数据

11月1日: 比赛结束, 计算排名

11月15日: 方法说明提交截止

11月20日: 公布排名



x



字节跳动

x

IEEE中国代表处

HOST: CAAI BYTEDANCE
CO-HOST: IEEE-CHINA



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