

计算科学评论

环球科学

WWW.COMPUTER.ORG

2018年第3期

金融技术 与应用

卡耐基梅隆大学论文解读 P52

APNet 2018参会记 P60

第二届Byte Cup数据大赛详解 P68

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 学习 科技新闻 竞赛
创业

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

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

本刊内容未经书面许可, 不得以任何形式转载或使用。

编辑团队

流程编辑

Carrie Clark
colark@computer.org

资深编辑

Chris Nelson

编辑

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

多媒体编辑

Rebecca Torres

设计与印刷

Carmen Flores-Garvey
Erica Hardison

封面设计

Matthew Cooper

资深广告经理

Debbie Sims

产品与服务总监

Evan Butterfield

会员总监

Eric Berkowitz

出版人

Robin Baldwin

主编

Sumi Helal
Lancaster University,
sumi.helal@computer.org

副主编

Elisa Bertino
Purdue University,
bertino@cs.purdue.edu

副主席, COMPUTING PRACTICES

Rohit Kapur
Synopsys, kapurfamily04@gmail.com

副主席, PERSPECTIVES

Jean-Marc Jézéquel
University of Rennes jean-marc.jezequel@irisa.fr

副主席, SPECIAL ISSUES

George K. 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

金融科技及其应用

金融科技是推动金融领域颠覆性创新的力量之一。它寻求重塑金融服务业如何构建系统、设置规定和捕捉客户需求。它已经进入到银行、保险、投资管理、证券等金融领域，带来了变革。全球有50%的金融消费者至少使用一种金融科技应用。随着开发人员和金融机构不断寻求利用新技术，他们往往会创造出颠覆性的新商业模式以提供附加值。

2018年第3期

12

利用大数据分析技术，基于公共媒体预测汇率：一项机器学习实验

本文分析了新媒体提供的相关公共信息，以预测美元 / 新台币汇率的变动，并探讨采用机器学习建模的大数据分析是否能有超越随机游走机制和市场效率概念的表现。

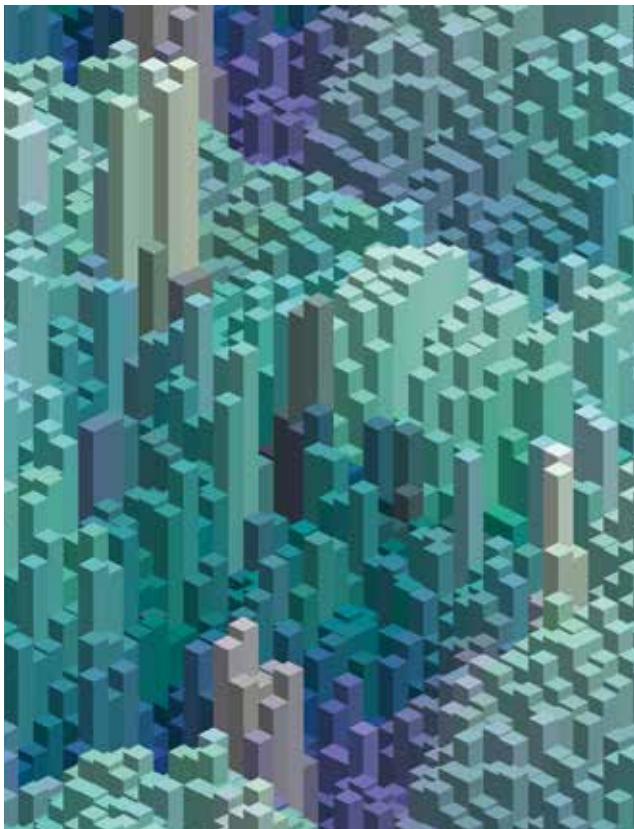
20

如何分析数据雄厚的非上市企业数据

本文阐述了收集和分析风险投资公司等财力雄厚的非上市企业的预处理数据的重要性。作者使用来自三个主要来源的数据集，演示了如何使用 Datastream 和 SQL 来查询和分析数据。

2018年第3期

目录



32

40

不断演变的支付行业：支付系统的 技术革新

成熟的金融机构及其技术合作伙伴正在为支付市场的金融科技创新做出重大贡献。本文全面概述了不断演变的支付领域中的技术创新和挑战。

用不用区块链？ 这是个问题

用不用区块链？这是个问题区块链被认为是一项突破性技术。但你的公司是否需要它？在本文中，作者以保险业为例，讨论了区块链技术的优缺点。这些分析可以被推广运用到其他行业。

竞赛

- 52 第二届 Byte Cup 国际机器学习竞赛详解

趋势

- 56 发展中国家的区块链

专栏

- 60 CMU 论文解读：基于 GAN 和 VAE 的跨模态图像生成

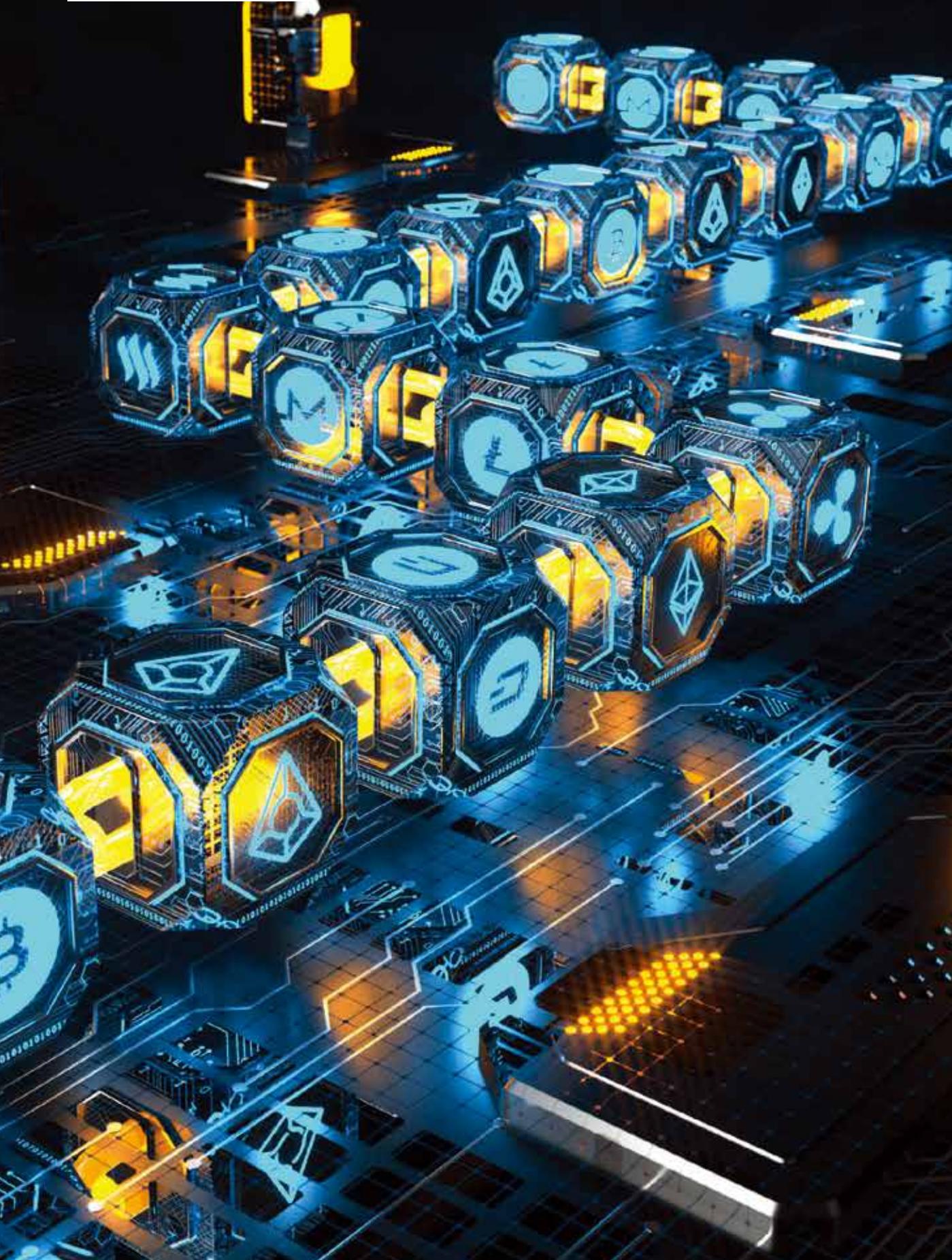
- 64 KDD 18 论文解读：斯坦福大学提出全新网络嵌入方法 GraphWave

会议

- 68 第二届亚太地区网络研讨会 (APNet 2018) 全程纪实

评论

- 72 邦内干戈：剑桥分析的反乌托邦



金融科技及其应用

文 | 李瑞元 (Maria R. Lee) , 实践大学 (Shih Chien University)
 严纪中 (David C. Yen) , 纽约州立大学奥尼昂塔分校 (SUNY at Oneonta)
 乔治·F·赫尔伯特 (George F. Hurlburt) , STEMCorp
 译 | 郑米兰, 浙江大学

本期杂志主题为金融科技及其应用, 本文综合介绍了金融科技的主要技术和商业模式, 以及几篇封面主题文章的内容。

金

金融科技是推动金融领域颠覆性创新的力量之一。总的来说, 它寻求重塑金融服务业如何构建系统、设置规定和捕捉客户需求。¹ 它已经进入到银行、保险、投资管理、证券等金融领域, 带来了变革。² 全球有50%的金融消费者至少使用一种金融科技应用。³ 而中国和印度的这一比例分别为84%和77%, 在全国范围迅速采纳金融科技方面领先全球。⁴ 随着开发人员和金融机构不断寻求利用新技术, 他们往往会创造出颠覆性的新商业模式以提供附加值。图1罗列了部分主要技术和新商业模式, 以及它们所贡献的至关重要的价值。

作为改善金融业的一种手段, 金融科技的早期形式可被追溯至17世纪的欧洲, 那里的金融财团帮助引发了工业革命。19世纪后期, 电报的出现和改进的电信能力帮助建立起全球金融互联。1967年, 简单的计算器和ATM系统为金融科技开创了一个新的数字时代。到了1987年, 计算机系统成为扩大和加速金融部门交易的普遍推动因素。2000年, 互联网和移动设备的出现已经使得金

融科技成为一种全球现象。⁴ 2008年的财政危机进一步推动了金融科技革命。随着技术不断加快流程并扩大金融机构的运营领域, 它挑战了金融界多年发展起来的最佳实践。为了更好理解这些挑战, 我们在此提供一个重要术语的词汇表, 以帮助读者理解在全球范围内不断涌现、种类繁多的金融科技方案。这些方案通常由那些获得大量投资、“咄咄逼人”的创业公司推动。⁵

比特币: 超过600种可用的加密货币中最受欢迎的一种。其曲折的历史和剧烈波动的价值继续导致不确定性。

区块链: 分布式分类账技术(DLT), 是大多数加密货币的技术基础。它已开始在那些需要验证交易的行业中取得合法性, 包括保险、证券和合同。

加密货币: 一种去中央化的数字货币, 使用加密和区块链技术来保护和验证发生在中央权威机构之外的交易。

颠覆性创新: 一种会挑战某个行业的监管或其他运营模式、迫使其发生建设性变革的创新形式。

以太坊: 区块链的一种形式, 其设

计支持去中心化的应用。

金融覆盖: 金融科技解决方案的应用, 覆盖发展中市场尚没有银行服务或银行服务不足的区域。

首次代币发行: 向投资者出售新加密货币的一种方式, 通常使用众筹。

保险科技(InsurTech): 通过专注于颠覆性创新来简化保险业流程及提高效率。

开放式银行业务: 利用应用程序接口将一个网络中的传统金融机构、第三方供应商、终端用户连结在一起。通常通过移动应用程序来实现。

监管科技(RegTech): 利用人工智能或机器学习等先进技术来简化金融合规要求, 主要包括反洗钱和客户识别验证。

机器人顾问: 算法驱动的自动投资顾问, 旨在削减人力投资顾问的角色。

智能合约: 自动化合同执行工具, 通常使用以太坊来编写脚本, 使用区块链来执行。

无银行服务/银行服务不足: 社会中缺少足够银行服务的部分, 可能受益于金融科技应用。

主题文章：金融科技及其应用



现代挑战和机遇

根据世界经济论坛的报告，金融科技涉及通常构成传统金融服务的六大核心功能：支付、保险、存款和贷款、融资、投资管理和市场资讯供应。¹ 其中每一项都具有独特的金融科技吸引力，但也对金融科技的不断涌现提出了重大挑战。

在支付领域，金融科技面临的挑战是如何让支付流程变得更加便捷，为消费者提供更具竞争力的产品。移动支付、简化支付、集成计费以及下一代安全机制是支付领域的关键性颠覆趋势。

聚合器的出现帮助保险客户区分各种传统保险销售渠道，可能有助于巩固保险公司与客户的关系。分销渠道、共享经济的概念、无人驾驶汽车以及第三方资本是进一步促进保险业分解的关键颠覆性创新。要更好地采纳保险科技（InsurTech），挑战最终将在于更智能廉价的传感器、可穿戴设备、物联网，以及其他标准化平台。

2008年的金融危机最终成为了金融科技的关键转折点。这场危机后，传统的金融中介严格审查贷款行为，以更

好地控制信贷风险。这导致信用评级较低的人无法满足贷款人要求而获得传统贷款。替代贷款机制的出现，如P2P贷款、虚拟银行2.0、移动银行、智能合约和开放银行应用程序接口，为客户提供了其他贷款渠道。与此同时，这些颠覆性技术直接影响了客户对贷款服务的期望，从而对传统金融机构构成挑战。

金融机构现在也被迫根据金融科技的发展来重新评估传统的融资方式。创业公司和中型企业目前在运营规划和信用评估流程中享有优势。替代性在线融资平台的出现，如股权众筹，扩大了这些新业务的融资渠道。结果，这些金融科技工具在各行各业的企业家中都变得更加受欢迎。

2008年金融危机之后，许多金融客户再也无法信任人力理财顾问。机器人顾问服务、社交交易平台和创新的零售算法交易提供了先进并且客户友好的理财工具。这些工具通常采用移动应用的形式，正变得比其人类同行更易获得、更便宜，也更可靠。这些技术正在成为在日益动态和快节奏的市场中为投资者服务的一股力量。

支持市场资讯供应的智能机械化

系统有望提高交易的准确性和精确度。为此，可由机器访问的数据、人工智能、机器学习和大数据正在成为算法交易的主要驱动力。从理论上讲，这些工具有望提供更智能、快速的响应。

虽然所有六个领域都面对创新和新技术，但更值得注意的是，随着金融科技不断获得动力，它将不断挑战现有的商业模式。随着时间推移，金融科技可能会显著改变长期建立的商业行为。随之而来的是更复杂、智能和快速的工具。但是，在一个网络攻击日益肆虐的时代，这些工具也可能导致欺诈、掠夺，并严重损害安全和触犯隐私。在一个交易率在很久以前就超过了人类感知新趋势能力的世界里，经认真验证有效性的监管科技（RegTech）将发挥越来越重要的作用。

重要的金融科技技术

金融科技最终旨在通过不断采用现代IT来简化金融行业的流程和提升效率。人工智能、机器学习、大数据、区块链、云计算、网络安全、生物识别、物联网、移动计算和社交网络在新兴的金融科技中扮演着各种不同的角色。此外，人工智能、大数据和区块链被认为是使金融科技能够渗透互联网和移动互联网的三大核心技术。⁶ 然而，这些相对较新的元素确实带来了一些重大的发展

风险。

金融科技技术的主要转变之一是从互联网金融发展成数据金融。数据一直都是金融的核心组成部分。但是,在大数据时代,金融数据必然会对风险定价和风险控制等领域里已加强的监管造成影响。人工智能和大数据一起成为金融数据计算未来的技术基础。金融人工智能应用专注于深度学习、智能分析和智能决策。大数据、云计算、智能硬件和区块链技术进一步增强了人工智能。⁷

区块链技术造就了大幅波动的比特币等加密货币,从中可以一窥这项技术将对创业公司、企业联盟和研究项目带来的巨大影响。⁸在实践中,区块链技术不仅仅只是记录货币交易。金融科技的许多新商业模式都使用基于区块链的技术。它是智能合约的基础,通常与以太坊(区块链的一种形式,其设计支持去中央化的应用)协同工作。举例来说,区块链技术已被用于跟踪医疗服务供应者的现状,以简化医疗保险,以及产生更智能化、更可验证的医疗服务。区块链还辅助保险公司提供新的保险产品,提高欺诈检测和定价的效率,节省管理成本。⁹

为获得竞争优势,金融行业不再关注互联网流量或互联网上的数据流。金融科技正在发生一种真正的进化:从以流量为导向转变为以技术为中心。由此,IT开始推动技术创新的新阶段,这

继而又导致金融行业出现新型商业模式。例如,技术推动了细节改进,增加了模块化,从而提升了金融科技的运作。这些趋势扩大了服务覆盖,加强了全面的反欺诈功能,形成了健全的数据控制机制。重要的是,它们使得风险消减和收购这两个环节实现了的急需的成本缩减。

随着数字化的不断扩大,大数据分析、云计算和金融科技协作很可能成为金融科技开发人员短期的考虑因素。而从长远来看,人工智能和机器学习自动化、密码学、区块链以及量子计算领域的颠覆性创新可能会成为金融科技差异化的最终因素。然而,它们每一项都会带来巨大的道德挑战。这些挑战尚未被充分地认识,更不用说解决了。

创新的商业模式

金融科技常常给现有商业模式带来颠覆性创新。随着P2P传输和加密协议的发展,新的支付类型正在出现。这些发展使得新的流程得以演变进化。不同于传统的集中货币清算机制,去中央化货币、加密货币和移动货币(如非洲的M-Pesa)产生了全新的竞争压力。这些新的支付方式迫使依赖传统高交易费率的既有金融机构重新考虑其战略。此外,Venmo和Dinngo等支付应用提供免费的数字钱包,让用户可与同行和

朋友一起完成共同、拆账付款。这些应用以有趣、廉价、简便又充满人际互动的方式来进行支付。为此,它们倾向于支持对无银行服务或银行服务不足的地区和人群的覆盖。

单个传统金融机构提供的服务可被分解为众多金融科技业务。金融科技往往会拆解金融业务的流程而将其中的组件重新连接起来——通常是通过相关联的金融科技应用用户。因此,金融科技IT正在从中介的角色演变为没有中间商的核心金融服务。其结果是,从更大型的互联的产业链中诞生了自成一体的业务类型。这种金融科技的专注性旨在脱媒,消除对银行、经纪商或其他金融机构等中介机构的需要。当金融科技被采用时,连接资金提供者(如储户或投资者)和资金使用者(如借款人或创业公司)的管道往往会变得直接得多。通过实施彻底改变的商业战略,脱媒似乎正逐渐成为全球金融业自由化的主导影响力。

网络信息聚合平台(聚合器)的出现将简化保险业务。过去,保险公司和客户的关系是基于小型、个性化的商业保险销售渠道。在线保险市场的出现和风险同质化将迫使保险公司的销售策略发生不可抗的变化。¹¹例如,低成本、带有社会责任的机器人顾问现在提供在线的、自动化的投资组合管理服务。这种服务帮助人们了解自己有哪些选项,

并在不需要人工顾问的情况下就做出最佳决策。毫无疑问，将需要更先进的机器人顾问来评估和解决更复杂的财务问题。但是，全球分析技术可能会进一步提高复杂系统的透明度，并使金融市场上的数据更容易获取。这继而可能带来更重要、有效和可行的投资决策。

价值创造

2018年，全球对金融科技公司的投资已超过4,256,202百万美元。此外，预计2022年全球交易额将达到7,971,957百万美元，年增长率为17% (www.statista.com/outlook/295/100/fintech/worldwide)。据高德纳报告，到2025年，区块链带来的商业增值将超过1760亿美元，到2030年将超过3.1万亿美元。¹²

最近金融科技的提升已经开始创造价值，这促进了金融业本身进一步发展。金融科技的关键价值增值之一帮助解决了开发人员自身货币流动的瓶颈。金融科技支付系统加强和推进了金融业的全球扩张和发展战略。

金融科技可以实现更轻松、快速的支付交易，从而绕过不同国家/地区的货币法规。P2P传输技术还可能允许利益相关方绕过传统中介并将资金直接转账到账户。这个过程快速且相对简单，只收取象征性的费用。实际上，P2P承诺以公平和透明的方式提供有价值的服务。

与依赖大型产品设计吸引客户的传统银行不同，金融科技已转向迎合消费者对简化的和模块化的产品及服务的偏好。随着客户越来越依赖并信赖金融科技，IT可能会被用来提供越来越方便的低成本数字化应用。因此，金融科技将改善客户体验，这将成为其最重要的价值创造之一。

本期封面文章介绍

本期IT专刊介绍新兴的金融科技趋势，并探讨充满前景的未来发展、关键的挑战和议题、创新的方法和新颖的解决方案，以及新型金融应用。在经过细致深入的筛选和评审过程后，我们选择了以下三篇文章。

在《利用大数据分析技术，基于公共媒体预测汇率：一项机器学习实验》中，蔡瑞煌、郭炳伸、林子翔、許哲銓验证了一个假设：对包含机器学习建模的大数据分析技术的应用，将能基于新媒体（新闻网站、论坛和社交媒体）中提供的相关公共信息来预测汇率变化。本文提出的大数据分析机制能够产生超过50%的预测正确率，表明它有可能为传统交易规则带来额外的利润。它所提供的金融科技方案有可能在市场效率竞争中脱颖而出。

在《如何分析数据雄厚的非上市企

业数据》一文中，拉塞尔·纽曼（Russell Newman）、张益城、罗伯特·J·沃尔特斯（Robert J. Walters）、加里·威尔斯（Gary Wills）描述了收集和分析非上市公司但资金丰厚的企业（如风险投资公司）的预处理数据所需的技术。他们还演示了如何使用Datastreams和SQL开展查询和分析。

最后，在《不断演变的支付行业：支付系统的技术革新》一文中，菲利普·卡伦（Filip Caron）鉴别了当前支付领域里具有固有的不同特征和风险水平的各个环节。本文将不同的业务需求与不同的支付细分市场联系起来，从未 来支付系统和监管框架的蓝图中提炼出趋势和技术挑战，并识别那些正在挑战传统银行的灵敏的技术公司。

未来的需求

鉴于金融科技的普及和迅速发展（以及从未减轻的网络安全威胁），消费者保护和欺诈保护机制是迫切的需求。在这个金融科技高度发展的时代，在所有市场中，政府的政策法规都必须跟上经济发展，特别是在亚洲和非洲这两个迅速崛起的新兴市场。这些地区具有优势，因为它们没有大量的旧有技术需要被颠覆和淘汰。与此同时，有些社会有责任向全世界传播灵活但却具有潜在风险的移动应用，它们在创造金融科

关于作者

技背后的技术时必须展开深入的道德思考。我们希望读者喜欢本期金融科技特刊, 我们欢迎您的反馈。■

参考文献

1. "The Future of Financial Services," World Economic Forum, report, 2015; www3.weforum.org/docs/WEF_The_future__of_financial_services.pdf
2. "The Future of Fintech: A Paradigm Shift in Small Business Finance," World Economic Forum, report, 2015; www3.weforum.org/docs/IP/2015/FS/GAC15_The_Future_of_FinTech_Paradigm_Shi_ft_Small_Business_Finance_report_2015.pdf.
3. "World Fintech Report 2017," Capgemini, report, 2017; www.capgemini.com/service/introducing-the-world-fintech-report-2017.
4. D.W. Arner, J.N. Barberis, and R.P. Buckley, "The Evolution of Fintech: A New Post- Crisis Paradigm?," Univ. of Hong Kong Faculty of Law Research Paper No. 2015/047, 2015; https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2676553.
5. R. Browne, "Everything You've Always Wanted to Know About Fintech," CNBC, 2 Oct. 2017;

李瑞元是实践大学资讯科技与管理学系教授, 也是苏州大学先进数据分析研究中心的访问教授。她的研究兴趣包括人工智能、大数据分析和电子商务。她在澳大利亚新南威尔斯大学获得电脑科学与工程博士学位。可通过maria.lee@g2.usc.edu.tw与她联系。

严纪虹是纽约州立大学奥尼昂塔分校经济商管学院资讯管理系统教授。他的研究兴趣包括数据通信、电子/移动商务、IT审计/治理, 以及物联网。他拥有内布拉斯加大学林肯分校资讯管理系统博士学位。可通过david.yen@oneonta.edu与他联系。

George F. Hurlburt是STEMCorp的首席科学家。STEMCorp是一家非营利性公司, 在公共部门采用网络科学推动自主技术为人类服务, 促进经济发展。可通过ghurlburt@change-index.com与他联系。

- www.cnbc.com/2017/10/02/fintech-everything-youve-always-wanted-to-know-about-financial-technology.html.
6. M. Swan, *Blockchain: Blueprint for a New Economy*, O'Reilly Media, 2015.
7. S. Chishti and J. Barberis, *The FINTECH Book: The Financial Technology Handbook for Investors, Entrepreneurs and Visionaries*, Wiley, 2016.
8. M.E. Peck and S.K. Moore, "The Blossoming of the Blockchain," *IEEE Spectrum*, vol.54, no. 10, 2017, pp. 24–25.
9. J.T. Lorenz et al., "Blockchain in Insurance—Opportunity or Threat," *report*, McKinsey & Company, 2016; www.mckinsey.com/doc/3627117/forecast-blockchain-business-value-worldwide.
10. "Coming Technology: Fintech Developers Tell You What to Look For and Why the Fintech Revolution Arose," CNBC, 15 Jun. 2017; www.cnbc.com/2017/06/15/fintech-developers-technology-apple-microsoft-xerox-kodak-revolution.html.
11. V. Dhar and R.M. Stein, "Fintech Platforms and Strategy," *Comm. ACM*, vol. 60, no. 10, 2017, pp. 32–35.
12. "Blockchain Business Value, Worldwide, 2017–2030," Gartner, *report*, 2017; www.gartner.com/doc/3627117/forecast-blockchain-business-value-worldwide.



利用大数据分析技术， 基于公共媒体预测汇率： ——一项机器学习实验

文 | 蔡瑞煌，国立政治大学

郭炳伸，国立政治大学

林子翔，国立政治大学

許哲銓，国立政治大学及黎明技术学院

译 | 郑米兰，浙江大学

本文作者分析了新媒体提供的相关公共信息，以预测美元／新台币汇率的变动，并探讨采用机器学习建模的大数据分析是否能有超越随机游走机制和市场效率概念的表现。

近年来，人们越来越依赖新媒体（网站、论坛和社交媒体）访问和共享信息。大多数社交媒体信息被视为公共信息。对于许多人来说，如此广泛的公共信息可以帮助他们做出决策。例如，汇率市场的玩家可以用这些信息来为自己谋取最佳利益，而无需付出什么成本。然而，新媒体提供的数据的种类和数量以及这些数据生成的速度都在不断增加。新媒体数据的真实性也是一个问题。此外，预测汇率的日常波动一直都是一项极其艰巨的任务，大多数计量经济模型在预测短期汇率时的准确度与简

单的随机游走模型无甚差别。¹

在金融经济学领域，市场效率这一重要概念与市场对新闻发布信息的反应有关。当任何新的公共信息进入市场时，交易者立刻根据新信息来调整自己的预期并做出交易。资产价格会立即变化以吸纳新信息，这时我们说市场的效率很高。然而，这种效率也意味着，由于新信息的公共性质，它并不能始终如一地带来利润。

因此，在预测美元 / 新台币汇率的日常变动时，声称大数据分析（BDA）可以超过随机游走机制和市场效率概念的

表现是具有挑战性的。为应对这项挑战，我们提出了一种 BDA 机制，它包含了一种机器学习建模，为处理更复杂的关系提供有效的方法。²

利用台湾期货交易所的真实美元 / 新台币汇率，我们通过排列新闻网站、论坛和社交媒体等信息来源，设置了七种类型的公共信息。我们选用的财经新闻网站是鉅亨网（cnYes）和 moneyDJ 理財网，论坛是“批踢踢實業坊”（PTT）的 ForeignEX 看板，社交媒体平台是 Facebook。我们将文本挖掘技术³应用于所获得的关于汇率的文本信息，以导出相应的数值表达。然后，我们运用机器学习建模来学习导出的数值表达与汇率的日常变动之间的关系。

汇率可以体现概念漂移和异常值。术语“概念漂移”的意思是概念不稳定，会随时间变化。⁴也就是说，随着时间的推移，观察数据中存在的趋势通常会发生变化。异常值是那些大幅偏离从给定的观察子集推导出的拟合函数的结果。⁵当数据受概念漂移影响时，异常值的检测会变得更具挑战性。黃馨瑩和她的同事们提出“包络模块抗拒学习”（resistant learning with envelope module, RLEM）和移动窗口技术来应对这一挑战。因此，我们在所提出的机器学习模型的第 1 阶段采用 RLEM 和移动窗口技术来过滤掉

我们提出的 BDA 机制能够生成超过 50% 的正确预测。这表明利用公共信息可以为传统交易规则带来额外利润。

潜在的异常值。由于期望的输出是二进制的，第 2 阶段采用推理神经网络⁷（RNN）进行训练。为了加快训练速度，我们使用 TensorFlow 和 GPU 来执行机器学习模型。

结果表明，我们提出的包含两种信息类型组合的 BDA 机制能够产生超过 50% 的预测准确率。也就是说，这一结果证明了这种 BDA 机制具有其优势。

汇率

汇率是一国货币与另一国货币的价值之比。⁸大多数主要经济体都采用浮动汇率。在浮动汇率制度下，外国货币和本国货币的市场供需力量决定了汇率。因此，为了稳定汇率，一个国家的政府必须拥有大量储备来控制外国和本国货币的需求和供应。

台湾的外汇市场不断受到台湾央行的干预。其目的是为了在市场比正常情况更不稳定和不规则时稳定市场，以及在有必要时让本国货币贬值以提

高出口行业的竞争力。然而，货币管理部门从未向公众公布在何时、如何或会在多大程度上做出这样的干预，即使是在事后也不曾通报过。金融业和学术界已经意识到，要对本地货币对主要外国货币的汇率做出准确预测是极具挑战性甚至是不可能的。

随着国际贸易的增加，汇率波动会产生巨大的经济影响。这些波动受到诸多因素的影响，比如通货膨胀率、利率和政治稳定性。⁹因此，许多专家和学者致力于研究汇率预测。目前有四种预测方法：技术的、基本的、基于市场的，以及混合的。⁴技术预测使用历史数据来预测未来的汇率趋势。基本预测使用与宏观经济变量相关的基本数据，如国内生产总值（GDP）、失业率、贸易差额和通货膨胀率。基于市场的预测使用即期汇率或远期汇率来预测未来即期汇率。混合预测是对上述预测方法分配权重得到的结果；各种预测的加权平均值提供了一种有用的混合预测法。

推理神经网络 (RNN)

RNN 采用单隐层前馈网络 (SLFN)。它的学习算法属于权重和结构变化这一类别，因为它最初使用一个隐藏节点，并在学习的过程中收集并删除隐藏的节点。RNN 学习目标的实现是有保证的。⁷

对每一个 SLFN， i^{th} 隐藏节点的激活值是 $h(\mathbf{B}_c, \mathbf{X}_i) = \tanh(\theta_i^H + \sum_{j=1}^m w_j^H b_j)$ 。

而每一个输出节点的激活函数是

$$O(\mathbf{B}_c, \mathbf{Y}_i, \mathbf{X}) = \theta_i^O + \sum_{k=1}^p w_k^O h(\mathbf{B}_c, \mathbf{X}_i),$$

上述公式中，

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}; \quad \mathbf{X}_i = (\theta_i^H, \mathbf{w}_i^{H^T})^T$$

$\mathbf{B}_c = (b_{c1}, b_{c2}, \dots, b_{cm})^T$ 是 c^{th} 的输入向量。

$\mathbf{w}_i^H = (w_{i1}^H, w_{i2}^H, \dots, w_{im}^H)^T$ 是所有输入节点和第 i 个隐藏节点的权重向量。 θ_i^H 是第 i 个隐藏节点的阈值。 θ_i^O 是第 i 个输出节点的阈值。

$\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T, \dots, \mathbf{X}_p^T)^T; \quad \mathbf{Y}_i = (\theta_i^O, \mathbf{w}_i^{O^T})^T;$
 $\mathbf{Y} = (\mathbf{Y}_1^T, \mathbf{Y}_2^T, \dots, \mathbf{Y}_q^T)^T; \quad \mathbf{Z}^T = (\mathbf{Y}^T, \mathbf{X}^T); \quad \mathbf{d}_c \in \{-1, 1\}^q$

是第 i 个输出节点的理想输出。 M 是输入变量的数量。 p 是采用的隐藏节点数， q 是输出节点数。上标 H 代表隐藏层的相关量。加粗的字母代表列向量、矩阵或集合，上标 T 代表转置矩阵。

在概念漂移环境中检测异常值

实验设计涉及三个概念：概念漂移、异常值检测、移动窗口。下面所示的 RLEM 被用来有效地检测概念漂移环境中的异常值。在 RLEM 学习过程中，所采用的隐藏节点数量是自适应的。 N 是训练数据的总量， m 表示输入节点的数量， k 表示潜在异常值的百分比，包络宽度是 2ϵ 。

SLFN 的权重 Z 。使用获得的 SLFN 计算所有训练数据的平方残差。然后，采取以下两个方向之一。

(1) 如果获得的 SLFN 的包络包含至少 n 个观测值，则去到步骤 7。

(2) 如果获得的 SLFN 的包络不包含至少 n 个观测值，则设置 $\bar{Z} = Z$ ，并应用增广机制来添加额外的隐藏节点以获得可接受的 SLFN 估计。

步骤 7： 实施删减机制以删除所有可能不相关的隐藏节点。 $n+1 \rightarrow n$ 。转到步骤 2。

步骤 1： 使用训练数据集中的第一个 $m+1$ 参考观测值来设置一个隐藏节点的可接受 SLFN 估计值。设置 $n = m + 2$ 。

步骤 2： 如果 $n > N \times (1 - k)$ ，则终止。

步骤 3.1： 使用获得的 SLFN 计算关于所有 N 训练数据的平方残差。

步骤 3.2： 在所有 N 训练数据的当前平方残差中，提出具有最小 n 平方残差的 n 参考观测值 (\mathbf{B}_c , \mathbf{d}_c)。

步骤 4： 如果所有最小的 n 平方残差都小于 ϵ ，直接去到步骤 7。否则，一个且仅一个平方残差大于 ϵ 。

步骤 5： 设置 $\bar{Z} = Z$ 。

步骤 6： 应用梯度下降算法来调整

大多数移动窗口将最新数据存储在先入先出的数据结构中。这个特点反映了增量学习中的两种机制：基于最新数据更新模型的学习机制，以及丢弃移出窗口的数据的遗忘机制。⁹ 此功能反映在丢弃过时数据和保留最新数据上。

实验设计

该数据集包括 2017 年 1 月 1 日至 6 月 30 日期间从台湾期货交易所下载的真实美元 / 新台币汇率数据。每天 120 个数据项按时间顺序排列，并从 1 到 120 标记索引。我们设定所需（输出）值来表现汇率变动，包括升值、不变或贬值。输出 (1,1) 表示升

值， $(1, -1)$ 和 $(-1, 1)$ 表示不变， $(-1, -1)$ 表示贬值。在我们的训练实例中，所需的升值输出设置为 $(1, 1)$ ，不变设置为 $(1, -1)$ ，贬值为 $(-1, -1)$ 。有 62 个升值（约占 51.7%）、56 个贬值（约 46.7%）和 2 个不变（约 1.6%）。

图 1 展示了实验设计。描述如下：公共信息包括来自鉅亨网和 moneyDJ 理財网的财经新闻、“批踢踢實業坊”论坛的 ForeignEX 看板中的帖子，以及 Facebook 上的相关讨论。这些数据资源的爬网程序在 Python 2.7 中执行。我们采用文本分割技术将文本数据划分为有意义的单元，比如单词和句子。针对繁体和简体中文都有很多工具，比如“中文斷詞系統”（CKIP，<http://ckipsvr.iis.sinica.edu.tw>）、ICTCLAS 汉语分词系统（<http://ictclas.nlpir.org>）和“结巴”中文分词（<https://github.com/fxsjy/jieba>）。我们使用结巴进行文本分割，并使用自然语言工具包（一种用于 Python 的自然语言处理（NLP）工具）来删除与内容无关的停用词，如“a”、“an”和“or”。词性标注系统则用适当的词性标记每个单词。我们还使用结巴标记所有文本数据，以确定每个单词是名词、动词、形容词或其他词性。

情感分析也称意见挖掘，是指使用自然语言处理、文本挖掘和其他方

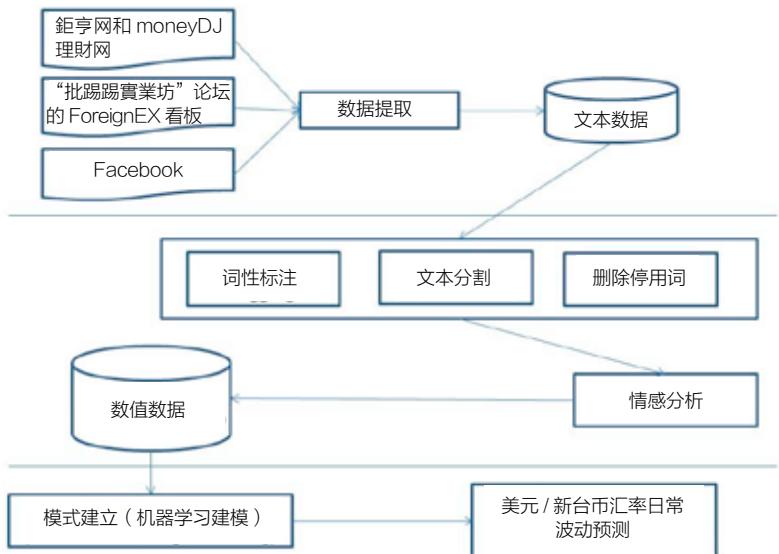


图 1. 实验设计

法系统地识别、提取、量化和研究情感状态和主观信息。我们使用语义字典 NTUSD10 来展开情感分析。该词典包含约 11,100 个具有正面 / 负面情绪的单词。

这里，每种类型的公共信息仅通过统计正面和负面词汇导出 VOLUME（情绪分数）。每天，初始 VOLUME 被设置为零。如果每个取出的单词与正（负）单词匹配，则 VOLUME 增加（减少）1。如果 VOLUME 为正（负），则相关的 SENTIMENT 被设置为 1 (-1)。

因此，对于 cth 数据， $Bc = (\text{社交媒体上的 SENTIMENT}, \text{社交媒体上的 VOLUME}, \text{新闻中的 SENTI-}$

MENT, 新闻中的 VOLUME, 论坛帖子中的 SENTIMENT, 论坛帖子中的 VOLUME)

图2显示了移动窗口技术的执行。实验为训练区块采用 90 个样本实例，为测试区块采用 5 个样本实例。例如，开始时，训练区块由第 1 到第 90 个实例组成，测试区块由第 91 到第 95 个实例组成。随着时间推移，前五个实例被丢弃，训练区块移到第 6 到第 95 个实例。而测试区块同时移到第 96 到第 100 个实例。此过程将一直继续，直至没有更多数据为止。因此，总共有六个窗口。

机器学习建模用于学习派生输入变量和期望输出之间的关系，然后预

表1. 对采样期间输入变量的描述性统计。

	社交媒体上的 SENTIMENT	社交媒体上的 VOLUME	新闻中的 SENTIMENT	新闻中的 VOLUME	论坛帖子中的 SENTIMENT	论坛帖子中的 VOLUME
最小值	-11	-11	-1	-81	-1	-5
最大值	1	41	1	19	1	7
平均值	0.34	3.46	-0.66	-0.93	-0.02	0.19
标准偏差	0.92	8.50	0.72	12.69	0.97	2.56

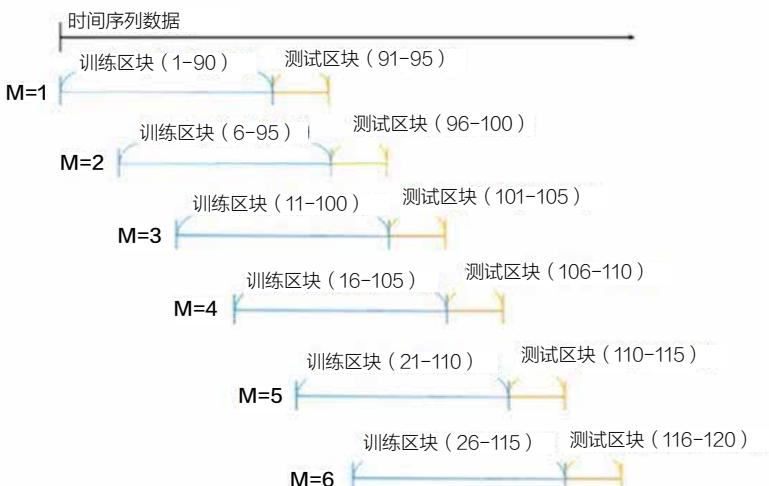


图1. 移动窗口技术的执行

测未来五天的汇率变动。我们提出的机器学习模型有两个学习阶段。在第1阶段，我们采用RLEM和移动窗口技术来滤除潜在的异常值。RLEM在训练结束时包含95%的训练数据，让他5%的数据被归类为潜在的异常值。在第1阶段之后，我们获得经训练的SLFN并移除潜在的异常值。在第2阶段，我们使用在第1阶段获得的RNN

和SLFN来学习余下的数据并获得最终的SLFN。在第2阶段之后，我们将测试数据输送给最终的SLFN。这两个阶段都是通过TensorFlow和GPU执行的。

表2的上半部分显示了每个窗口和每种类型的公共信息的预测率。表2的下半部分显示了每种公共信息的预测率的最小值、最大值、平均值和标

准偏差。

表2展示了以下三个观点。

>有一种组合(如N+S)的某些预测率高于50%，而另两种组合(如F+N和F+S)则表现出较差的预测能力。对于这些类型的信息，第1阶段每个窗口的平均学习耗时为约1小时，第2阶段每个窗口平均为约7小时。

> F+N+S这种公共信息类型的某些预测率高于50%。对于这类信息，第1阶段每个窗口平均学习约1小时，第2阶段每个窗口平均约9小时。

>没有任何单一类型媒体的公共信息显现出良好的预测能力。对于这些类型的信息，第1阶段每个窗口的平均学习时间约为50分钟，第2阶段每个窗口平均约6小时。

结论和意义

表2.七类公共信息的预测率。F:论坛; N: 新闻; S:社交媒体。

	F	N	S	F+N	F+S	N+S	F+N+S
第1到第6个窗口	20%	20%	20%	20%	20%	40%	60%
	0%	20%	40%	20%	40%	60%	40%
	20%	40%	60%	40%	40%	60%	60%
	40%	40%	60%	40%	20%	40%	80%
	40%	60%	40%	20%	40%	60%	40%
	40%	40%	40%	40%	40%	60%	60%
最小值	0	20%	20%	20%	20%	40%	40%
最大值	40%	60%	60%	40%	40%	60%	80%
平均值	26.7%	36.7%	43.4%	30%	33.3%	53.3%	56.7%
标准偏差	14.9%	13.7%	13.7%	10%	9.4%	9.4%	13.7%

在金融经济学领域，市场效率这一重要概念与市场对新发布信息的响应有关。但是，市场效率也意味着新信息因其公共性质并不能总是带来利润。我们的实验结果表明，我们提出的包含两种信息类型组合的 BDA 机制能够生成超过 50% 的正确方向预测。这表明利用公共信息可以为传统交易规则带来额外利润。换言之，我们的金融科技方案可能会有超越随机游走机制和市场效率概念的表现。

我们的研究与先前的研究不同，因为它强调从社交媒体收集的重要定性变量的预测能力。传统预测方法利

用可用的市场数据，包括聚合宏观经济时间序列数据和交易量或价格数据。基于定量数据的实证研究结果尚未达成共识。换句话说，我们完全没法说哪种传统方法主导了预测。由于缺乏公平的基准，把我们的预测结果与其他方法的结果做比较会因为主观判断而变得复杂。即使不做这样的比较，我们的方法应该也能为传统预测方法提供一种补充。

对于金融科技领域的专业人士，它至少有一个实际意义：这种系统地收集、组织和分析公共信息的 BDA 机

制很可能会带来额外的利润。

未来的工作应该包括将采用机器学习建模的 BDA 机制应用于美元和其他货币的汇率预测以及覆盖更长时间的数据。

参考文献

1. R. A. Meese and D. Rogoff, “Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?,” *J. International Economics*, vol. 14, no. 1-2, 1983, pp. 3–24.

关于作者

2. D. Zhang, S. Simoff, and J. Debenham, "Exchange Rate Modelling Using News Articles and Economic Data," *Australasian Joint Conf. on Artificial Intelligence* (AI 2005: Advances in Artificial Intelligence), 2005, pp. 467–476.
3. R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge Univ. Press, 2007.
4. M.G. Masud et al., "Classification and Novel Class Detection in Concept-Drifting Data Streams Under Time Constraints," *IEEE Trans. Knowledge and Data Engineering*, vol. 23, no. 6, 2011, pp. 859–874.
5. R.H. Tsaih and T.C. Cheng, "A Resistant Learning Procedure for Coping with Outliers," *Annals of Mathematics and Artificial Intelligence*, vol. 57, no. 2, 2009, pp. 161–180.
6. S.Y. Huang, J.W. Lin, and R.H. Tsaih, "Outlier Detection in the Concept Drifting Environment," *IEEE 2016 Int'l Joint Conf. Neural Networks (IJCNN)*, 2016, pp. 31–37.

蔡瑞怡是国立政治大学资讯管理学系教授。他的研究兴趣包括人工神经网络、机器学习、基于信息通信技术(ICT)的服务创新及大数据。他拥有加州大学伯克利分校的运筹学博士学位。他是IEEE会士。可通过tsaih@mis.nccu.edu.tw与他联系。

郭炳仲是国立政治大学国际贸易系教授。他的研究兴趣包括将模型平衡应用于估算和预测，重点是模拟汇率变动。他拥有罗切斯特大学经济学博士学位。可通过bsku@nccu.edu.tw与他联系。

林子翔拥有国立政治大学资讯管理学系硕士学位。可通过st01129@gmail.com与他联系。

许哲铨是国立政治大学资讯管理学系博士生，也是黎明技术学院资讯管理系讲师。他的研究兴趣包括知识管理、ICT支持服务、服务能力与供应链管理。可通过chuan1020@gmail.com与他联系。

7. R.R. Tsaih, "An Explanation of Reasoning Neural Networks," *Mathematical and Computer Modelling*, vol. 28, no. 2, 1998, pp. 37–44.
8. M. Levinson, *The Economist Guide to Financial Markets: Why They Exist and How They Work*, Public Affairs, 2014.
9. J. Gama et al., "A Survey on Concept Drift Adaptation," *ACM Computing Surveys*, vol. 46, no. 4, 2014, p. 44.
10. L.-W. Ku and H.-H. Chen, "Mining Opinions from the Web: Beyond Relevance Retrieval," *J. American Society for Information Science and Technology*, vol. 58, no. 12, 2007, pp. 1838–1850.

微信名：计算机
微信号：jisuanren



《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。



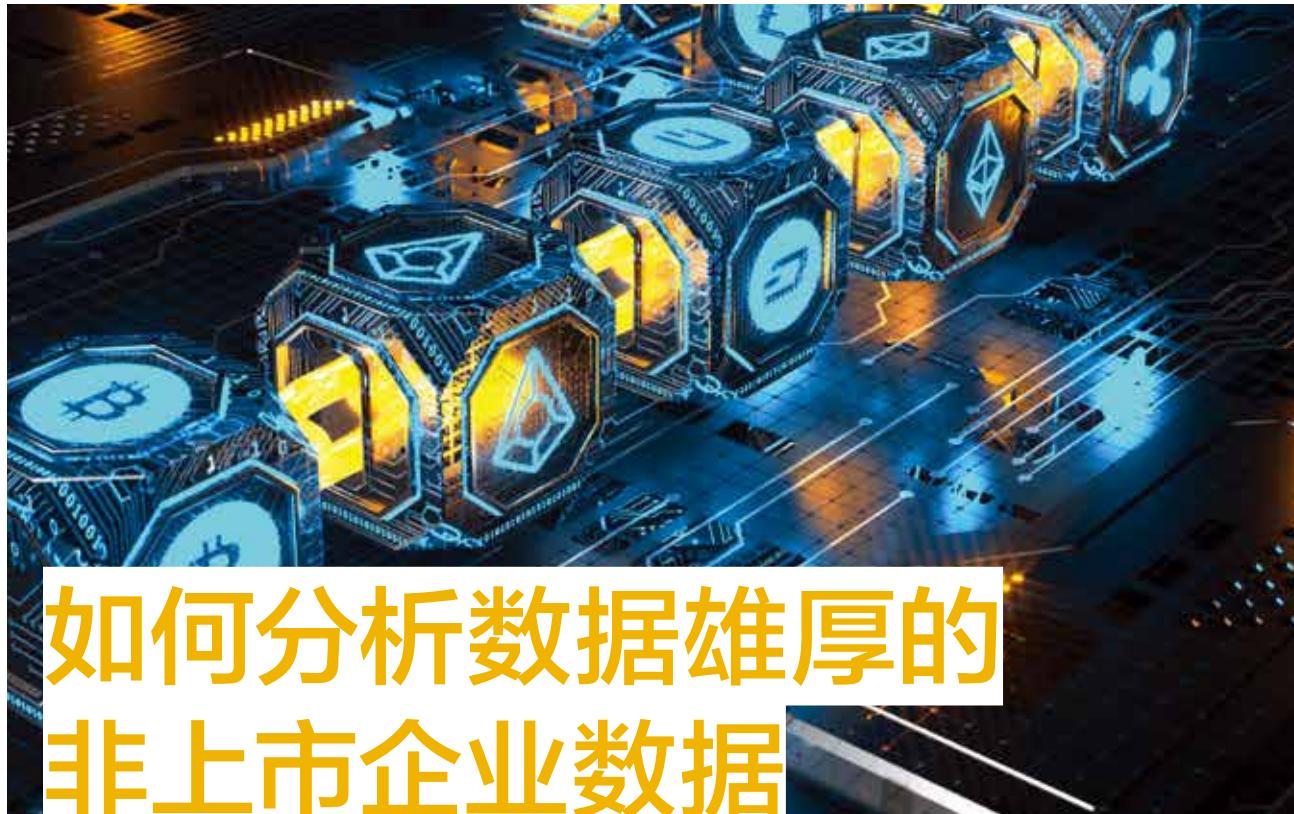
iOS和Android现已同步上线

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



如何分析数据雄厚的 非上市企业数据

文 | 拉塞尔·纽曼（Russell Newman），南安普顿大学（University of Southampton）
张益城（Victor Chang），西交利物浦大学（Xi'an Jiaotong-Liverpool University）
罗伯特·J·沃尔特斯（Robert J. Walters），南安普顿大学（University of Southampton Gary）
加里·威尔斯（Gary Wills），南安普顿大学（University of Southampton）
译 | 吴岑，四川大学

本文阐述了收集和分析风险投资公司等财力雄厚的非上市企业的预处理数据的重要性。作者使用来自三个主要来源的数据集，演示了如何使用 *Datastream* 和 SQL 来查询和分析数据。

在 本文中，我们旨在展示预处理数据的重要性，提供有关用行业主导的解决方案访问重要数据源的见解，并定量分析近年涉足网络行业的公司在多大程度上重现了1999年至2001年间的互联网泡沫。

尽管商业情报服务能够跟踪和呈现某些股票的价格和波动性，¹但有许多公司，特别是风险投资公司（VC），尚未进入股市。因此，需要使用其他类型的金融技术来衡量其商业价值。

我们的重点是研究当前和历史市场数据。为此，我们使用可靠的金融数据集进行计算分析。我们需要使用来自多个来源的数据集分析模型，相关详情将在稍后说明。这就需要开发一种预处理方法来清理、标准化和组合数据集，以便将它们轻松地导入分析软件中。然后我们通过开放式数据库连接（ODBC）运行SQL查询，从而将数据加载到分析工具中。我们发现，预处理是一种灵活的方法，可以使用自行开发的 *Datastream* 服务分析各种维度的数据。

获取 VC 活动数据

我们的概念模型需要以单个公司为单位调用VC活动数据。但是，彭博社没有关于每家公司VC交易的数据，因为许多获得VC支持的企业都是私有的，因而不需要对外发布这些信息。² 在与彭博社的专家讨论后，我们得知这些数据可以通过搜索某些标题和关键字从彭博的新闻报道中收集和提取。

但是，这种方法有四个主要缺点，降低了它的实用性和价值。

1. 并非所有VC交易都是公开宣布的，因此有些交易不会出现在彭博新闻服务中。
 2. 需要编写软件程序来执行从彭博新闻服务中收集VC相关新闻的任务。
 3. 作为一种信息服务而非数据服务，彭博新闻可能包含重复的新闻报道，导致重复的数据。
 4. 从彭博新闻中收集的任何数据都需要经过大量的人工清理和验证。
- 由于无法通过可靠的方式获得每家公司的VC数据，我们发现了三个更高级别的潜在数据来源：

1. NESTA（国家科技艺术基金会）报告：“风投动向”（Venture

由于无法通过可靠渠道获得风险投资机构的数据，我们找到了3个更高级别的潜在数据源。

Capital Now) 和“网络泡沫崩溃后”(After the Dotcom Crash) (英国数据)；

2. 欧洲风险投资协会(EVCA)季度报告(欧洲数据)；
3. 普华永道(PWC)和国家风险投资协会(NVCA)的MoneyTree报告(美国数据)。

NESTA的报告提供了有关英国VC活动的数据，³ 它们是使用Thomson One服务的数据编制的。但这些数据缺乏所需的规模和深度。EVCA每季度发布一份关于VC活动的泛欧报告，并按行业细分。PWC的MoneyTree报告是一项免费服务，使用来自汤森路透的Datastream的数据编制有关美国VC活动的季度报告。⁴ PWC代表NVCA编制这份季度报告。报告中的数据按美国各州、不同市场部门，以及融资的不同阶段进一步分组。

数据精度

年度数据提供了足够高的精度。

但更高的精度能够提升细节。PWC和EVCA通过其季度报告提供了精度最高的数据。

历史数据

经济泡沫本身就为我们设定了应以怎样的时间框架获取数据：理想情况是从互联网泡沫(1999年至2001年)之前的时期开始，一直延续至尽可能接近当前。只有PWC提供了互联网泡沫之前、期间和之后的数据。

数据维度

所有三个来源都提供了按投资额量化的VC数据。NESTA和EVCA进一步提供了获得投资的公司数量。PWC则提供了VC交易的数量。后者是一个更恰当的衡量标准，因为考虑到了一家公司在同一时期获得多次投资。

融资阶段分组

所有数据来源都按融资阶段对数据作了分组，根据VC投资流程中的阶段(种子、扩展等)罗列VC投资额。

主题文章：金融科技及其应用

表1. 分项比较VC数据源。

	新闻中的VOLUMN	论坛帖子中的SENTIMENT	论坛帖子中的VOLUME
数据区域	英国	欧洲	美国
数据精度	年度	季度	季度
最早的数据	2000年一季度	2005年一季度	1995年一季度
最近的数据	2009年四季度	2013年三季度	2013年四季度
数据维度	投资额和公司数量	投资额和公司数量	投资额和交易数量
融资阶段分组	有	有	有
行业划分		自2013年一季度起	有
地理区域分组			有
格式	PDF	PDF	Excel
数据定义	基本	基本	详细

表2.DATASTREAM时间序列查询示例。

参数	输入
序列	@GOOG, @YHOO
数据类型	WC01201, WC08006
开始和终止日期	开始: 31/12/2004 截至: 14/01/2005
频率	2.56

行业分组

只有PWC自1995年第一季以来一直将市场部门对数据分组。这使得我们能够忽略软件/互联网领域之外的VC活动。通过按部门对数据分组，PWC的数据清楚地显现了投资趋势和对某些部门的权重。虽然PWC数据以美国为中心，但在对每个部门的投资做粗略分析

后可以发现，某些部门的权重很大。因此，依赖那些不将数据分解为各部门的数据集将是不准确的。

地理区域分组

NESTA和EVCA的数据都是所覆盖区域的整体数据。PWC数据按美国的州做了分组。但是，在详细的州的层

面展现地理区域与VC活动的关系并非我们这项研究的目标。

格式

只有PWC的数据格式可直接被机器读取。NESTA和EVCA的数据都需要手动输入、修补和验证，才能将数据转换为机器可读格式。

**Datastream 是汤森路透维护的数据
库，用于整合年度和季度公司报告及其
他行业数据源的定量数据。**

数据定义

PWC MoneyTree网站明确定义了用于生成和呈现其数据的分组和类别，解释了他们编制季度报告的方法，包括为何将某些类型的数据排除在研究以外。PWC提供了足够多的细节，可以从基础数据中精确地复制其数据集。NESTA和EVCA未提供有关其数据来源和处理方法的详细信息，因此很难确定这些数据集的来源和可靠性。

选择和理由

虑及以上因素，我们决定使用PWC MoneyTree作为VC活动数据的来源，因为它是可用的最丰富的数据，其详细的数据定义也最适合我们的研究。按行业对VC数据分组是它不同于其他数据集的关键因素，它使得我们可以专注于和我们的研究相关的行业上，而消除了其他噪音。这也和认为我们这项研究将以英国为中心的假设有所不同。其结果是更加一致的分析，其中，数据主体是基于相同的地理区域并受到相同经济条件的影响。

获取公司数据

Datastream是一个由汤森路透填充的数据库，用于整合年度和季度公司报告及其他行业数据源的定量数据。它是一个单一实体，可以通过它查询和比较原本分散和不相干的数据源。⁵

Datastream围绕汤森路透的“世界展望”(Worldscope)中代表各种实体类型的“级别”构建财务数据。它还包含围绕各种主题罗列的公司和证券列表。这让用户可以快速检索一组相关公司或证券的数据。我们使用其中的美国软件公司列表，它列出了在伦敦证券交易所上市的105家公司。

Datastream 查询方法

在访问Datastream时，可获得一个基于助记符的图形查询工具。它被设计成Excel插件，将查询结果直接下载到电子表格中。

Datastream支持各种类型的查询。静态查询将检索可能无法按时间段做比较的数据，例如特定公司的董事名

单。我们的研究则使用时间序列查询来检索根据时间段划分的定量数据。

表2是一个Datastream时间序列查询范例。该序列是这项查询应反馈的有关一个或多个实体（对公司、证券及其他“世界展望”中的“级别”基于助记符的展示）的数据。开始和终止日期是查询应反馈结果涉及的时间段，表现为具体日期或相对日期（例如“年初”）。数据类型是查询应反馈的数据结果的维度，从可搜索列表中选择（对16,432个财务指标的基于助记符的展示）。频率是应返回数据结果的时间频率（每年、每季度、每月或每周）。

我们仔细构建查询，以确保选择有用的数据类型。有效的设计会让查询的实体确实拥有我们所查询的属性。

在指定数据频率时，须注意所查询数据类型的报告周期。假如查询的是年度数据类型（例如WC01001）而指定的是季度频率，则单个年度值将被复制到每个季度的结果中。

查询结果将呈现为Excel电子表格中的一张粗略的交叉表。默认情况下，行表示时间段；列首先表示查询的主体，然后是数据类型。这有效地将所有查询的数据类型包含在同一张表中。

以下摘要显示了一个Datastream查询示例，用于查找2004年12月31日至2005年1月14的三周内谷歌和雅虎的研发支出（助记符号WC01201）和交易

主题文章：金融科技及其应用

表3.DATASTREAM交叉表输出。

名称	谷歌—研发支出	谷歌—交易额	雅虎—研发支出	雅虎—交易额
符号编码	@GOOG (WC0 1201)	@GOOG (WC0 8006)	@YHOO (WC01 201)	@YHOO (WC080 06)
31/12/2004	214289	343760368	368760	153826027
07/01/2005	483978	831408175	547137	178197917
14/01/2005	483978	831408175	547137	178197917

表4.从DATASTREAM中提取的数据类型。

数据类型助记符	名字	报告时间跨度	“世界展望”级别
NOSH	股票数量	每小时	
WC01001	净销售或收入	每年	公司
UP	未经调整价格	每小时	
MV	市值(资金)	每小时	
P	价格 (经调整后-默认)	每小时	
WC01201	研发	每年	公司
W08006	交易额	每年	证券

表5.DATASTREAM查询输出示例。

	@AKAM(W08006)	U:DDD(W08006)	@ADBE(W08006)	@ALTR(W08006)
1990			4895877	363212
1991		13455	6620124	1054756
1992		12331	7975639	1856235

量(助记符号WC08006)。

表3是此项查询的输出，显现查询设计存在问题——一些检索到的数据是重复的。这是因为虽然指定的是周频

率，查询的却是年度数据类型。因此，

须使用年度频率或不同的数据类型重新构建查询。

> Datastream限制返回结果集的

大小。如果查询包含太多维度，则返回的数据可能会被截去超出限制的部分。这在受影响的单元格中

一家公司可能会在所查询的时间期限内停止公开交易，导致在公司被私有化之后难以获得它的数据

显示为“查询错误”代码。

> Datastream不会以真正的交叉表的方式返回数据，这在表3每列标题中的公司名称重复上可以看出来。这使得我们无法在Datastream返回结果后轻松访问和分析Excel中的数据，而必须创建工作流程以将Datastream的输出转换为真正的机器可读格式。

查询 Datastream

表4显示了为本研究目的而从Datastream中提取的数据类型。每个数据类型都从单个查询中取得。

表5至7显示了交易量（助记符W08006）查询的输出样本。选择的四家公司显示出各种类型的输出。公司助记符显示在标题行的括号中。

以上输出数据反映出公司的三种不同情况。

1. @AKAM: 该公司在所查询的时间期限内没有公开交易，因而没有数据。

2. U:DDD:该公司在所查询的时间期限内开始公开交易，因此数据始于1991年。
3. @ADBE和@ALTR: 这些公司在所查询的整个时间期限内都公开交易，因此返回的是完整的数据集。

此外，一家公司可能会在所查询时间期限内停止公开交易，导致在公司被私有化之后数据不可获得。

组织及合并数据

获取数据的过程生成了九个不同的数据集，它们可被合并成一个数据集以实现数据分析。这项工作是在Excel中进行的。

九个数据集中有七个是由类似的Datastream查询生成的，过程中只更改了数据类型。这使得这些数据集可通过识别匹配的公司助记符和每组中的年份来合并。通过将每个数据集逐一合并到目标数据集中，就生成了一个所有数

据类型的组合数据集。

Datastream的七个交叉表输出被转换为数据透视表。我们发现在Excel中的以下工作流程最为有效，并将其应用于所有七个数据集：

1. 隔离公司助记符。使用Excel函数，从标题行单元格中提取公司助记符: = LEFT(A2, FIND("(", A2)-1)
2. 生成数据的多个合并范围数据透视表。这将再现完全相同的表结构，但却是在一个“真正”的数据透视表中。

通过以下步骤3，把要处理的最后一个数据集（编号7）另外展平为线性表：

3. 双击数据透视表的Grand Total单元格。这将创建一个新工作表，其中包含数据透视表中数据的展平线性版本。表6显示这一步骤的样本输出（使用表5中的数据作为输入）。

这项操作生成了一个展平列表，罗列了最后一个数据集中每个公司、年份（参见表6中的行和列）以及数据类型W08006的值。然后，在这张表格上增加包含GETPIVOTDATA函数的列，将

表6.DATASTREAM扁平/线性输出。

行	列	值
1990	@ADBE	4895877
1990	@AKAM	
1990	@ALTR	363212
1990	U:DDD	
1991	@ADBE	6620124
1991	@AKAM	
1991	@ALTR	1054756
1991	U:DDD	13455
1992	@ADBE	7975639
1992	@AKAM	
1992	@ALTR	1856235
1992	U:DDD	12331

表7.DATASTREAM展平/线性输出。

年份	公司	MV	NOSH	UP	P
1990	@AAPL	4430.25	125681	35.250	8.8125
1990	@ADBE	417.23	20604	20.250	1.2496
1990	@ADI	458.25	47000	9.750	1.6252

数据集1到6的数据透视表数据导入此展平列表中。该GETPIVOTDATA函数大致如下：

```
=GETPIVOTDATA(“价格”, ‘股价PIV’! $A$3, “日期”, $A4, “公司”, $C4)
```

这生成了一个单一的平面表，其中

包含代表公司和年份的行，每个数据集一列。表7显示的是包含四个数据集的示例。

生成的二维平面表包含了Data-stream的所有输出，适用于使用SQL数据库进行存储和分析。

VC和在线人口数据则以更有用的

格式提供。这三个数据集都有年度数据，因此从剩余数据集中提取了代表年份的列。

在线人口数据集包含基于每个地区、每年的数据。一些地区是单个国家，另一些则是国家的集合（如欧盟、高收入国家、全球）。从数据集中提取了不

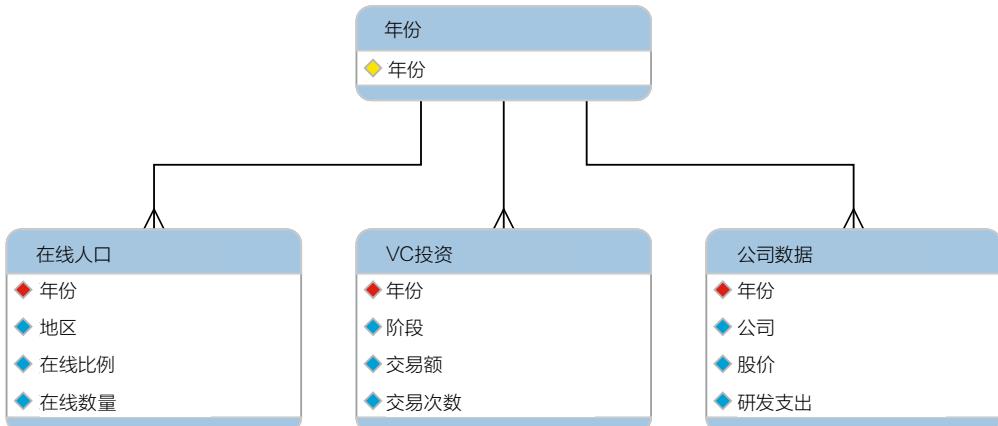


图1. 显示以年份来合并数据集的概念图。

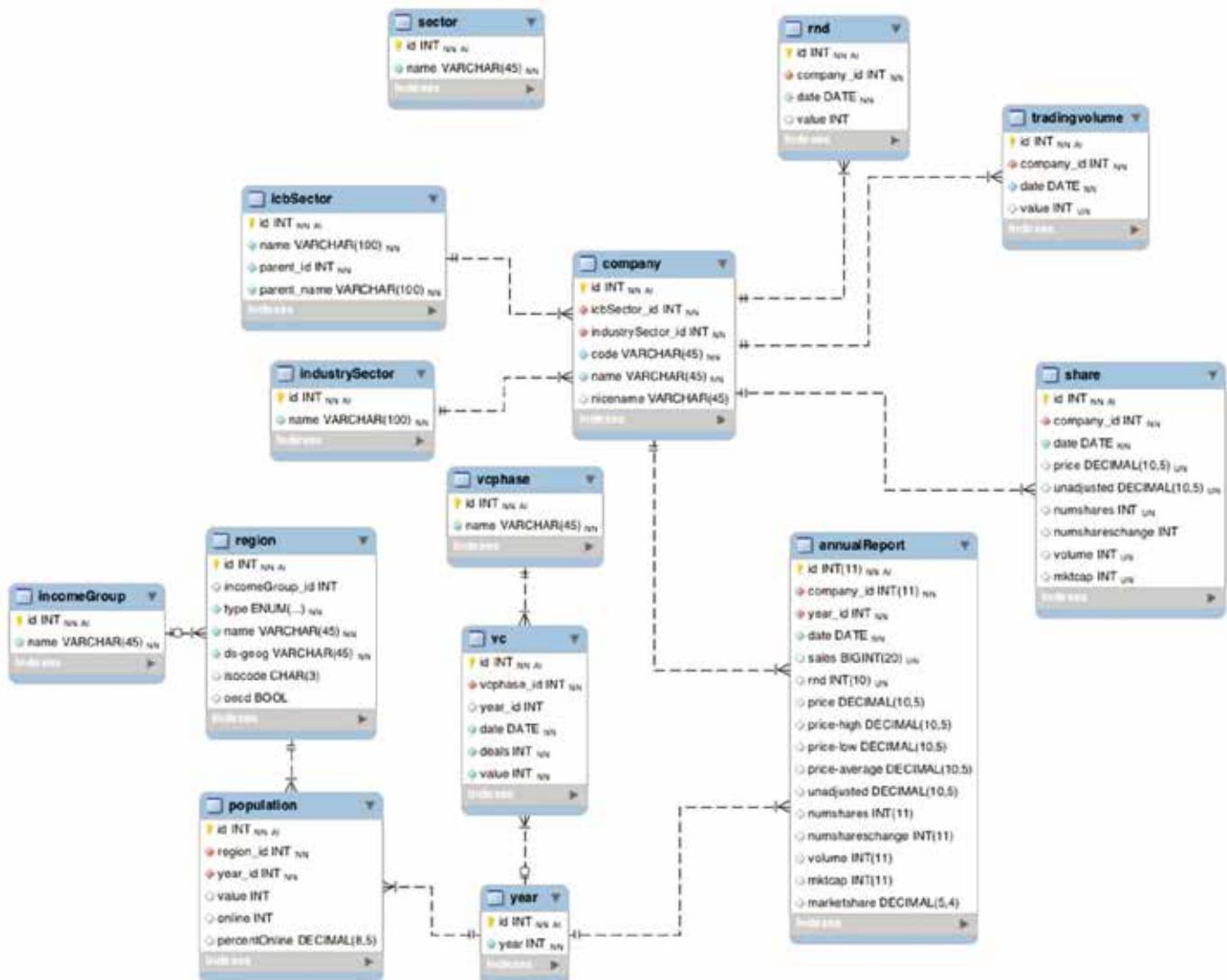


图2. 数据库架构。

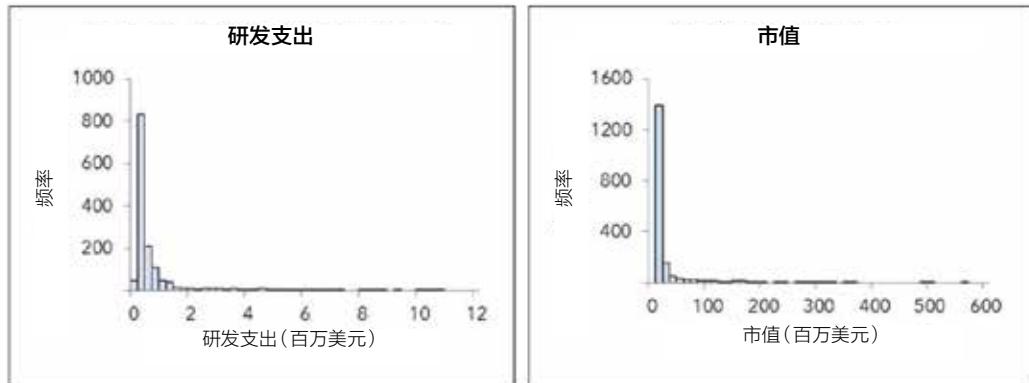


图3. 两个样本数据集的原始值柱状图。

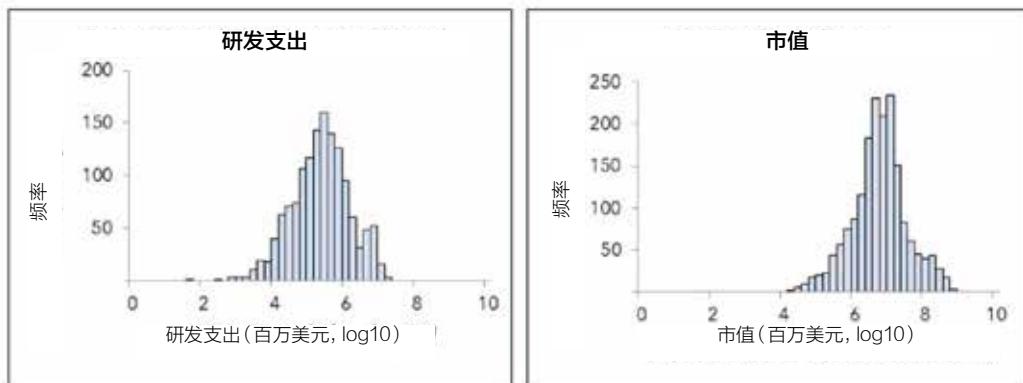


图3. 两个样本数据集的自然对数值柱状图。

同的区域，并根据每行是代表单个国家还是区域进行编码。

VC数据集包含基于每个阶段、每季度的数据。其中“阶段”表示VC流程的阶段（例如，启动/种子、早期阶段、后期阶段和扩张）。

从数据集中提取了四个阶段的列表。数据是按季度提供的，因此添加了

“年”的一列，并从现有数据中填充。

图1是一张概念图，描述了如何用共同的年份字段合并三个数据集。图中仅显示了对列和属性的示例。

检查的过程发现有两个数据集具有多个相同数值。这是源于在查询Datastream时发生的一个令人遗憾的人为失误，导致数据集中出现一页包含

错误数据序列的结果。错误的数据被识别出来，并用Datastream中的正确数据替换。而后，原始数据集和部分克隆的数据集都被仔细查核。

作为进一步的验证点，还应检查数据的分布正态性。为此我们生成一个数据的柱形图，并将其与正态分布曲线进行比较。

关于作者

图3显示了来自两个Datastream数据集的原始数据的柱状图。分布严重偏向横轴的起点，带有长尾，且非正态分布。对这种状态下的数据进行的任何统计分析都可能出现困难，并被认为是不可靠的。

图4显示了在计算了每个数据的自然对数之后，同样这套数据呈现的柱状图。经过这种转换之后，柱状图将数据集显示为正态分布。因此，柱状图中显示的结果对所有Datastream数据集具有代表性。

这一转换被应用于所有Datastream数据集，在所有七套数据中都生成了类似的输出。它也被应用于“VC交易额”的数据列，该列基于每年、每个阶段罗列VC活动涉及的财务额度。在所有情况下，原始数据都不作任何改动，并在视图中计算对数。

分析方法及其在研究议题中的应用

我们查询获取了从1995年开始、经过网络泡沫期公司直到2015年的公司和行业数据。之后对数据做了“协调”处理，让它们可以作为一个连贯的集合被查看和分析。为此，我们根据主题周期对数据集当中的年份进行编码。这使得我们可以基于时间段（包括泡沫前和泡沫后）来分析数据。

Russell Newman是南安普顿大学的访问研究员。他的研究兴趣包括IT和金融。他获得了南安普顿大学的计算机科学博士学位。可通过rn2@ecs.soton.ac.uk与他联系。

张益城是西浦国际商学院副教授、博士及研究硕士项目主任。他的研究兴趣包括大数据、云计算、物联网和安全。他在南安普顿大学获得计算机科学博士学位。他曾多次获奖，包括2015年IEEE杰出服务奖和2017年杰出青年科学奖。可通过ic.victor.chang@gmail.com与他联系。

Robert J.Walters是南安普顿大学的计算机科学讲师。他的研究兴趣包括网格计算、云计算、软件工程和网络科学。他在南安普顿大学获得计算机博士学位。可通过rjwl@ecs.soton.ac.uk与他联系。

Gary Wills是南安普顿大学计算机科学副教授、开普敦大学客座教授以及RLabs的研究教授。他的研究兴趣包括安全系统工程及在工业、医学和教育中的应用。他从南安普顿大学获得了工业超媒体系统博士学位。他是注册工程师、英国工程技术学会（Institute of Engineering Technology）会员、英国高等教育学院（Higher Educational Academy）首席研究员。可通过gbw@ecs.soton.ac.uk与他联系。

我们对协调后的数据集进行了因素分析，以识别出有联动关系的列。这输出了许多派生因子，显示哪些列与每个因素（负面或正面）相关并在多大程度上相关。这是一种降维方式，可用于探索数据列之间的关系，或用于将宽泛的数据集简化为较少数量的列。

我们用这种方法来发现协调后数据集中的任何单独数据列是否彼此相关。如果有相关性，又是什么样的关系。

因素分析显示出数据集中各度量标准之间的关系。它显示出哪些指标对

估值产生了正向贡献，哪些则产生了负面影响。我们也观察了其他指标之间的关联并探索了它们对本研究的潜在价值。

结语

我们将努力开发出更复杂的方法，以识别那些涉及逃税、洗钱、建立空头公司的现金充裕的非上市企业。我们将使用人工智能、大数据和金融智能技术来使我们的研究成果最大化。■

参考文献

1. V. Chang, "The Business Intelligence as a Service in the Cloud," Future Generation Computer Systems, vol. 37, 2014, pp. 512–534.
2. M. Meeker, "Internet Trends 2015—Code Conference," KPCB, blog, 2015; www.kpcb.com/blog/2015-internet-trends.
3. T. Lux and M. Marchesi, "Volatility Clustering in Financial Markets: A Microsimulation of Interacting Agents," Int'l J. Theoretical and Applied Finance, vol. 3, no. 4, 1998; doi.org/10.1142/S0219024900000826.
4. MoneyTree Report: Q4 2014/Full-year 2014, report, PricewaterhouseCoopers and National Venture Capital Association, 2014; www.valuewalk.com/wp-content/uploads/2015/04/pwc-moneytree-q4-full-year-2014-summary.pdf.
5. Y. Bai et al., "Efficient Support for Time Series Queries in Data Stream Management Systems," Stream Data Management: Advances in Database Systems, Springer, 2005.
6. O.S. Ince and R.B. Porter, "Individual Equity Return Data from Thomson Datastream: Handle with Care!" J. Financial Research, vol. 29, no. 4, 2006, pp. 463–479.

The image shows a call for articles for the IEEE Pervasive Computing journal. The background features a close-up photograph of a blue printed circuit board (PCB) with various electronic components like resistors and capacitors. Overlaid on the image is a large, stylized orange text that reads "Call for Articles". Below this, in a white box, is the title "IEEE Pervasive Computing" in bold black font. To the left of the title, there is a section titled "Author guidelines:" with the URL "www.computer.org/mc/pervasive/author.htm". Below that is another section titled "Further details:" with the email "pervasive@computer.org" and the website "www.computer.org/pervasive". To the right of the title, a detailed description of the journal's focus is provided: "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." At the bottom right, there is a logo for "pervasive COMPUTING MOBILE AND UBIQUITOUS SYSTEMS" featuring the IEEE logo and the word "pervasive" in a large, italicized font.

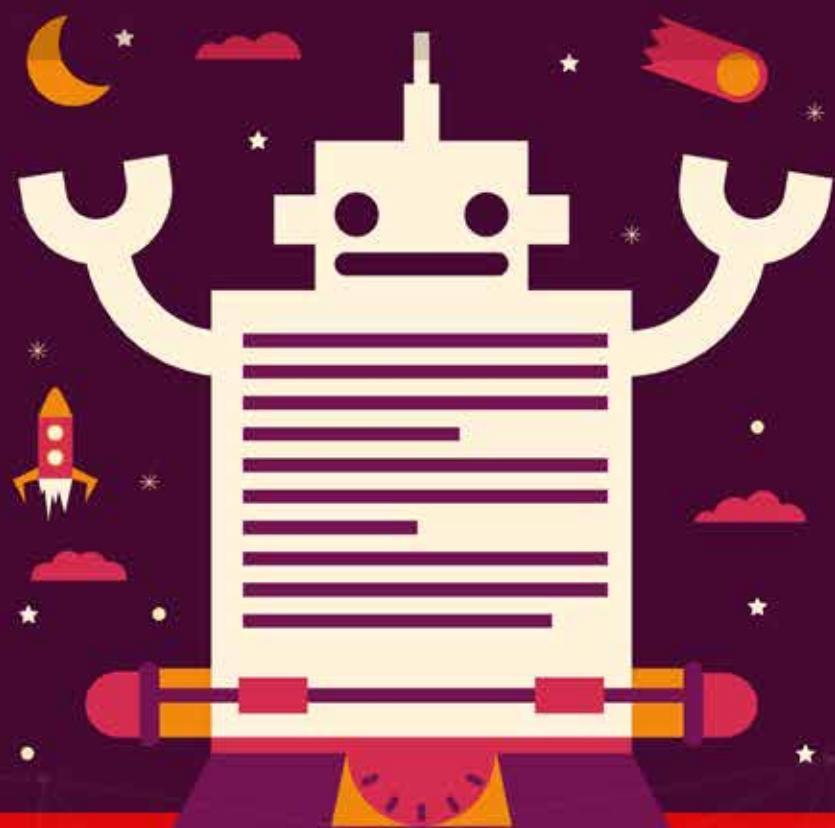
SMP EUPT 2018

SMP & 今日头条

猜猜今日头条的文章作者

是人、写作机器人、翻译机器人还是摘要机器人？

奖金: ￥29000



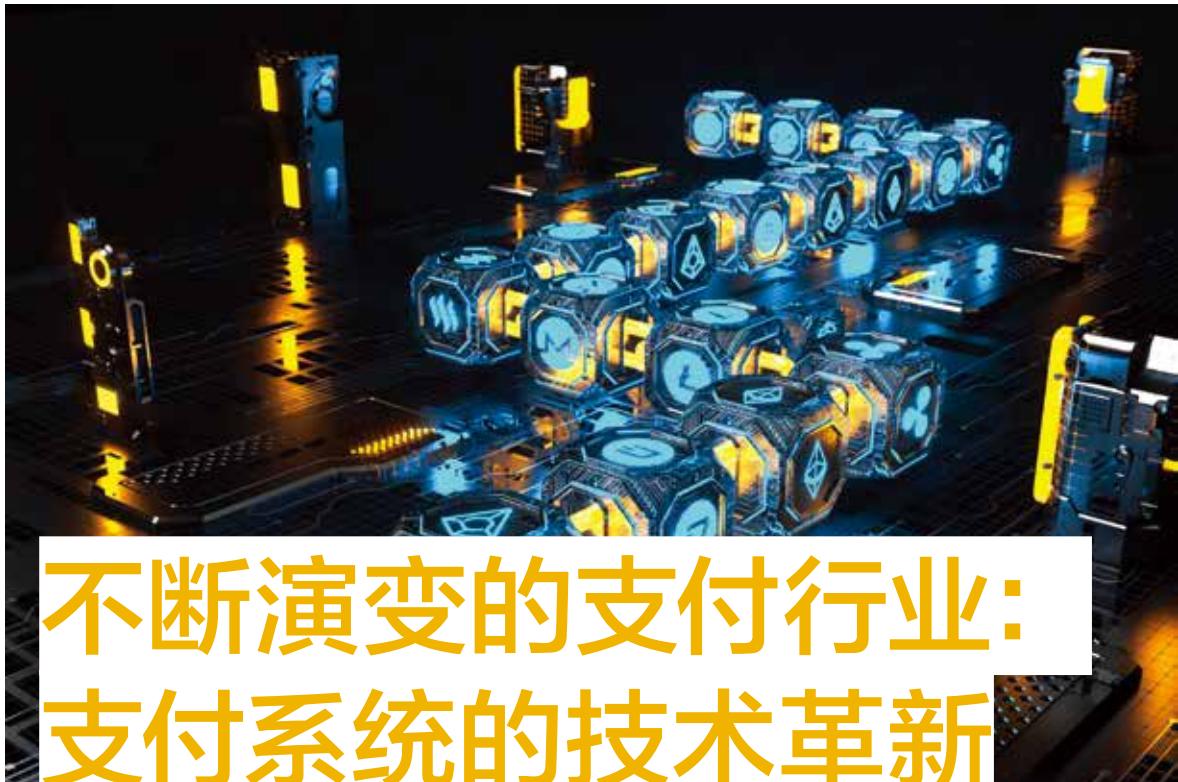
SMP

X

 TOUTIAO AI LAB
今日头条人工智能实验室

报名入口





不断演变的支付行业： 支付系统的技术革新

文 | 菲利普·卡伦 (Filip Caron)，鲁汶大学 (KU Leuven)

译 | 郑米兰，浙江大学

成熟的金融机构及其技术合作伙伴正在为支付市场的金融科技创新做出重大贡献。本文全面概述了不断演变的支付领域中的技术创新和挑战。

成熟的金融机构及其技术合作伙伴正在为支付市场的金融科技创新做出重大贡献。本文全面概述了不断演变的支付领域中的技术创新和挑战。

众多支付服务供应商运营着一系列不同的支付系统，根据正式安排和标准化规则促成这种资金转移。这种多样性源于不同细分市场各自不同的需求和技术挑战。由此，我们在不断变化的支付领域中观察到新动态和创新的解决方案，包括：

- > 零售环节的颠覆性在线支付服务（如支付宝、Venmo、PayPal、Skrill、Dwolla、2Checkout和TransferWise）；
- > 新法规，比如欧洲修订的支付服务指令 (PSD2) 和“沙箱系统”。它们为测试新服务提供了宽松的监管环境；
- > 改善了用户体验并能更快获得资金的新支付系统，包

括澳大利亚的新支付平台 (New Payments Platform) 和英国的快速支付服务 (Faster Payments Service)；

> 新的网络弹性支付系统，用于金融机构之间的实时结算，例如TIPS (TARGET即时支付结算)。

两个关键维度将支付行业整齐划分为不同的集群，每个集群包含具有类似特征和处理需求的付款：

- > 支付参与者的类型。通常分为两类：金融机构和非金融机构¹（政府机构、企业，或私人家庭）；
- > 由付款人和收款人所在地定义的付款地理范围。

批发支付是金融机构之间常见的高额资金转移。保证有序和及时完成批发支付对金融市场的健康运行至关重要。因

此，批发支付系统必须保证在信息安全、运营可靠性和网络弹性方面都达到高度成熟。

零售付款是非金融机构之间的付款。其均值往往比批发支付小几个数量级。相比之下，它们在支付的次数中所占的比例要高得多。这类支付系统的运营商主要致力于改善用户体验、处理及时性、与其他平台的整合，以及确保充分扩展。

通常会在国内和跨境支付之间做出区分。国内付款是同一司法辖区内付款人和收款人之间的交易，而跨境付款的地理范围涵盖多个国家。由于欧元区的支付系统越来越多地与单一系统整合或被单一系统取代，因此该区域内的支付可被视为国内支付。

国内批发支付创新

由支付和市场基础设施委员会（CPMI）等国际标准组织制定的安全要求和效率标准直接反映在下一代批发支付系统的技术蓝图中。此外，这些未来系统的运营商专注于扩大覆盖、扩展功能、促进整合，以及开展端到端风险管理。

网络弹性：加强应对新兴威胁的能力

批发支付处理中信息系统之间的

相互联系显著加深，而网络攻击的复杂性也在增加。不断增加的数据泄露事件引发了对事故发现和响应能力的额外投资。

在实施预防性控制之外，行为分析技术提供了一种补充手段，比如网络入侵探测、安全信息和事件管理解决方案，以及CBEST等基于情报的渗透测试。²

监管机构在恢复节点和恢复时间方面提出了雄心勃勃的目标。当前的网络弹性要求规定在事故发生后两小时内须恢复关键性运作。³下一代批发支付系统的设计蓝图将因病毒、硬件故障、环境事故等常见问题造成的停机时间缩短至接近零。⁴

批发支付系统在多个站点上运行，而且技术多样化的应急平台已经到位。研究表明，在网络攻击使核心系统无法运行的情况下，技术多样性有助于快速恢复运作。

扩大对批发支付系统的访问

一些运营商已在考虑将访问资格扩展到传统金融机构之外。由于这类访问限制仍很常见，合法的非银行支付服务只能通过其主要竞争对手——银行——来访问批发支付系统。支持小型在线支付的非银行支付服务通常必须连接到支付领域的主干基础设施，比如将客户资金转移到收款人的银行账户。

英格兰银行将跟上金融业结构变化的步伐，其中包括由非银行支付服务供应商推动的对金融服务业的拆解。⁴通过扩大批发支付系统的访问资格，运营商可以为不同类型的支付服务供应商创造公平的竞争环境。虽然这可能会增加支付市场的竞争和革新，它可能会给运营商带来重大的技术治理要求，包括第三方风险管理和平端安全策略。

促进支付处理系统之间的互操作性

目前在构建支付消息（比如一条转账命令中包含了发送到某个支付系统的与特定支付相关的所有财务信息）时存在多种标准。在全球各地已经出现了采用ISO 20022所规定的单一开放方法的趋势。ISO 20022是金融机构之间交换电子数据的复合型国际标准（www.iso20022.org）。

采用基于ISO 20022的消息传送标准将带来战略益处。由美联储委托进行的一项商业案例评估强调，标准化的消息格式能够促进自动化并降低成本。⁵英格兰银行寻求基于ISO 20022来促进与零售支付系统的互操作性，这将使用户能在批发支付系统中断时将付款转移到零售支付系统中。⁴最后，标准化将可利用一组通用的数据字典元素来改进监管报告和监控。

批发支付系统的多个运营商已经

表示将采用基于ISO 20022的消息传递标准，包括欧洲央行、美联储和欧洲银行协会清算公司（EBA Clearing）等。

在不断变化的支付环境中增强用户功能

用户对批发支付系统新增的需求主要是优化处理效率和改善与系统的交互。用户越来越多地要求在任何给定时间进行即时支付处理，但这里存在显著的技术障碍。下一代批发系统将需要在服务仍在持续时运行日终批次。此外，系统架构须允许引入并行硬件，以便在服务仍然进行时执行重大的维护和升级。技术上能够实现近全天候运行是英格兰银行下一代批发系统的一项要求。⁴欧洲央行将依赖TIPS系统提供一项服务，在任何时间（全年无休）支持立即处理银行间交易。⁶

下一代批发系统还将为其用户提供面向外部的应用程序接口（API）。这些API允许与用户的系统集成，以及利用系统中可用的支付数据。支付数据分析的例子包括财务风险管理（尤其是流动性管理）以及交易和账户限制。此外，API还可以帮助实现行业正在开发的批发支付追踪功能，其中最引人瞩目的举措是银行拥有的协作系统SWIFT的全球支付创新（GPI）跟踪器（www.swift.com/our-solutions/global-financial-messaging/payments-cash-management/swift-gpi/features）。

financial-messaging/payments-cash-management/swift-gpi/features)。

支付处理中的端到端风险管理

批发支付系统不是在真空中运行，它们与用户和其他第三方的系统高度互联。银行要求更紧密的交织以提高支付处理的效率并建立适当的监控能力，但由于它们本身是一个极广泛的生态系统的一部分，因此网络中的端点数量急剧增加。诸如2016年孟加拉银行信息系统泄漏等网络事件表明，一次攻击可能会在支付处理链的任何组织或步骤中发生，然后影响整个生态系统。风险评估越来越多地覆盖与支付系统相连接的第三方基础设施的相关风险。

运营商参与网络情报获取和共享计划，在其系统中探测已知的攻击媒介。例如，金融服务信息共享和分析中心（FS-ISAC；www.fsisac.com）就是这样一个计划。此外，SWIFT还建立了一个广泛的计划来改善批发支付行业参与者的网络安全意识和举措（www.swift.com/myswift/customer-security-programme-csp/program-description）。该计划要求遵守一套最低限度的被普遍接受的网络安全措施，以及采用实时监控工具来防止欺诈性或不符合政策的支付指令。

国内零售支付创新

与批发支付系统相比，对零售支付的创新主要集中在改善终端用户的体验上。

用户需求的变化：“随时、随地、无论如何”范式

近期数字服务的扩展导致了消费者习惯和期待的重大变化。基于对这些数字服务的预期，消费者要求及时向收款人支付资金，同时又要求很大的灵活性。

零售支付创新强调更快处理支付，但在这方面存在着不同的解决方案设计。基于银行转账的方式正在全球范围内使用，例如澳大利亚的新支付平台、英国的快速支付服务、瑞典的Swish和新加坡的FAST。此外，Visa Direct等支付卡计划、PayPal等电子货币提供商，以及比特币等加密货币正在提供快速支付方案。

这些举措中的绝大多数旨在打造永不中断的系统，实现近乎实时（通常少于10秒）到账或拒付。欧洲零售支付委员会已经批准了一项名为SCT Inst的泛欧即时支付计划的总体设计（www.europeanpaymentscouncil.eu/what-we-do/sepa-instant-credit-transfer）。目前多个符合SCT Inst的系统正在开发中，比如EBA Clearing和

下一代批发支付系统将为其用户提供面向外部的应用程序接口。

Stet。实时处理零售支付的技术挑战之一是要求系统的覆盖面和弹性都大。为保证最大程度的覆盖,欧洲自动清算所协会(EACHA)为符合SCT Inst标准的系统开发了一个互操作性框架。此外,严格的时间限制增加了减轻风险方面的压力,包括分析支付指令以发现欺诈和洗钱行为。

此外,实物正在越来越多地变成能与其环境中的其他事物交互的智能物件(即物联网)。由这些智能物件为服务自动付款是一种经常被提到的交互模式,比如汽车按行驶里程或在每次有人租车时自动支付车险。

人们正在商议智能物件支付领域的不同解决方案。“价值互联网”(Internet of Value)提案使用支持虚拟货币的区块链。此外,现有的支付卡计划提议交换标记化支付卡的数据(www.master-card.us/content/mccom-admin/faq-category-admin/tokenization.html)。

研究表明,现有的加密密钥管理协议和访问管理设计可能都不适用。^{7,8}

实施“价值互联网”需要极大改善区块链技术的可扩展性。苏黎世联邦理工大学的Duplex微支付、比特币的闪电网络(Lightning Network)和以太坊的Raiden微支付分别提出了允许定期重新编码余额的技术架构。

开放零售支付行业

最近,越来越多不同的参与者进入了零售支付市场。新技术加持的数字创新者以及修订升级的监管框架都推动了零售支付服务的发展。而对原本难以触及的消费者的覆盖正在推动需求面的增强。

为促进支付行业的创新、透明度和竞争,欧盟委员会修订了PSD2的监管框架。PSD2的核心是要求银行提供安全访问个人银行账户的受许可服务。欧洲银行管理局已起草监管技术标准,明确了加强客户认证及API标准化方面的信息安全要求。⁹许多参与者从修订后的监管框架中受益,比如SOFORT、iDeal和Trustly这类在电商网站和网上银行平台之间建立软件通道的支付启动

服务。

互联网的普及和信息技术的进步使得点对点支付的崛起成为可能。移动钱包里包含了信用卡的标记化版本,并允许通过近场通信(NFC)芯片和二维码付款。比特币等加密货币连接到基于区块链的点对点支付系统上。这些支付系统的用户面临的主要技术挑战是安全地存储他们所拥有货币的密码,而这些系统的运营商则在寻找可扩展的交易验证算法。

移动网络运营商提供了一些替代性的支付系统,比如像Boku这样的运营商计费系统,或像非洲的M-Pesa这样的运营商存款服务。像M-Pesa这样由移动运营商支持的支付系统已经显著改善了发展中国家的金融服务普及度。在这些地区,对广泛覆盖的移动网络的需求是至关重要的,也是这些系统的运营商面临的一个重大技术挑战。

数字平台带来竞争优势

在大多数司法辖区,人们都预期银行将为新零售系统的发展做出贡献。此外,需要大量投资才能将银行的内部系统与新的支付系统整合。由于这其中涉及巨大的成本,许多计划正在研究灵活而开放的系统设计,使创新服务能够直接与现代化系统相连接。

这些创新服务的供应商从开放的API开发中获益,这类开发不仅仅

是为了达到PSD2或“开放银行标准”(<https://theodi.org/project/open-banking-setting-a-standard-and-enabling-innovation>)的最低要求。额外的API功能将允许银行进一步促进对银行数据的访问并从中获利。例如，澳大利亚的新支付平台构想的服务包括批量提交付款或提供付款凭证资料。

保证信息安全以及不同银行间关键性API的标准化是重要的技术挑战。另一方面，创新服务的供应商必须提供强有力银行账户持有者认证机制，以及与银行建立安全通信的工具。由于尚没有针对额外API功能的监管技术规范，缺乏标准化可能会使得创新服务的供应商难以与一大批银行建立连接。

最近进入零售支付市场的非银行服务供应商也遵循类似的策略。例如，支付宝和WePay为其服务提炼了典型的使用场景，并开始开发广泛的合作伙伴网络。目前，应用场景主要集中在“在线到离线”模式上。这些模式把在线客户获取与服务和产品的实际交付结合起来，比如支付水电费、网上购物、手机钱包充值、购买火车票等。

代理服务简化支付处理

传统上，零售支付系统依赖于付款人知道一部分收款人的个人财务信息，例如账号。这被认为是零售系统中产生阻力的主要原因，特别是在实时或即时

支付的情况下。

支付处理代理服务运作一个集中式数据库，该数据库记录支付服务的用户关系、他们的账户详情，以及一个或多个唯一的代理标识符。付款人可以通过这些唯一的代理标识符来向收款人付款。唯一代理标识符的典型例子是电子邮件地址、社交媒体身份标识和移动电话号码。

支付处理代理服务正被迅速采用，例如泰国的PromptPay、美国的clearXchange和澳大利亚的新支付平台。移动代理论坛(Mobile Proxy Forum)旨在开发一个泛欧标准化代理查询服务，该服务将提供一个互操作性框架，让泛欧支付系统都能访问该服务。

大数据机会

与付款相关的数据包含有关客户的兴趣、活动和行为模式的大量信息。支付服务的运营商可以据此构想多种应用。

首先，对支付数据的深度统计分析

将实现更有效的风险和合规评估。当代大数据解决方案使银行和其他支付服务供应商能将来自不同类型交易(如信用卡、借记卡、支票和电汇)的支付数据汇总到一个中央化平台上，从而能就客户的交易情况生成一个更全面的概览。由此产生的“数据仓库”将使得先进的统计分析技术能够更好地将行为、时间和空间技术结合到对欺诈和洗钱的侦察中。

其次，支付相关数据可用于不断改善用户体验，这会在终端用户越来越苛刻的环境中提供竞争优势。支付服务供应商可以跨多个使用渠道创建单个帐户和交易总览，从而形成对终端用户更好的洞察并能为其提供个人理财应用程序。此外，聚合的数据也可用于基于算法的信用评分。PSD2中建立的监管框架明确认可服务供应商提供的此类账户信息。

最后，支付服务供应商可以尝试通过数据获利。对支付数据进行深度市场分析可以为细分市场提供宝贵的洞见，并协助平台合作伙伴开展奖励活动和

互联网的普及和信息技术的进步使得点对点支付的崛起成为可能。

识别交叉销售机会。

支付相关数据的存储和第三方使用所面临的主要挑战仍然是监管机构规定的信息安全和隐私要求。

跨境支付创新

人们越来越期望跨境支付具有与国内支付相同的特征。收付款双方要求这类支付便宜、及时、可预测且可追踪。当前造成国际交易时间和成本不确定性的主要驱动因素是无处不在的代理银行。代理银行业务是在不同国家的银行之间使用一个或多个中介机构的操作。如果付款人的银行对收款人所在的外国金融市场的访问受限，通常就需要用到代理行，并被要求兑换货币。

由于跨境支付的高度复杂性、货币兑换的风险，以及多个司法辖区的监管差异，这个领域内的技术创新相对受限。

通过脱媒提高效率

数字创新者提出的跨境支付系统越来越多地建立在以金融机构脱媒为核心的商业模式上。加密货币或许可以在没有任何中间人的情况下实现跨境支付。不过，目前可以观察到一系列不同程度的脱媒。

用于跨境交易的点对点支付系统绕过了银行及其代理行，从而避免了这些中介机构收取费用。算法将跨境支付

与相反方向的支付相匹配，反向传送资金流。这类系统的固有挑战是它假设可以在两个方向上找到相同数量的支付交易。**TransferWise**、**Midpoint**、**CurrencyFair**和**WeSwap**提供此类跨境支付系统。

跨境支付中脱媒的尝试不仅来自于金融科技挑战者。银行也在寻找不同的基于新技术的机会来解决跨境交易。银行和技术公司联盟R3分析了**Ripple**的数字资产**XRP**和分布式分类账技术(**DLT**)在解决银行间即时跨境支付(无需在这些银行之间建立联系)上的效用。此外，国际支付框架协会制定了一项标准，使金融机构能够直接与其他司法辖区的国内支付系统相连接。

增强透明度、可预测性和可追溯性

与此同时，银行正在修改其信息系统和业务流程，以改善现有的代理银行业务模式。首当其冲的改革是**SWIFT**的**GPI**计划。**GPI**的核心是在参与该计划的银行之间创建一系列多边服务协议，规定了与及时性、成本透明度以及代理银行业务中丰富的支付数据传输相关的业务规则。

端到端的跟踪解决方案将允许**GPI**银行检查跨境支付的状态——从一项支付的发起直到最终确认的过程。除了提供代理银行间资金转移的透明度之

外，跟踪解决方案还将提供有关银行费用和外汇汇率的概览。**GPI**银行面临的技术挑战包括采用转账唯一识别码(**UETR**)来跟踪其系统中的支付路径，以及使用API基础设施来与跟踪器交互。此外，**GPI**银行可能还会考虑为企业客户提供软件桥梁，让它们能够追踪查看支付状态。

在代理银行链中共享丰富的支付数据进一步提高了代理行的效率。目前，企业客户经常被要求提供额外的支付信息以进行合规检查。此外，支付和发票的核对通常需要人工干预，而当丰富的支付数据与支付一起传送时，就可以免去这种干预。一个解决方案是基于云的信息层，可用支付的**UETR**码来查询。

虽然点对点平台能就支付抵达收款人的时间实现完全的透明度和相当的可预测性，但其中的安全保证被认为还达不到改进后代理银行模式的标准。因此，预计这些点对点平台将主要用于低端零售支付，而非重要的企业支付。

结语

许多运营商正在开发各式各样的支付系统创新。国内批发支付创新主要是为了提高可预测性和安全性，而国内零售支付创新的目标是改善用户体验和成本效益。跨境支付的创新则旨在提高支付的效率、透明度和可追溯性。

关于作者

Filip Caron是鲁汶大学(KU Leuven)决策科学与信息管理系的一名教师，并与鲁汶信息系统研究所合作。他的研究兴趣包括信息管治和金融技术。卡伦获得鲁汶大学管理信息系统博士学位。他撰写了30多篇学术论文，并教授商业分析和流程管理的研究生课程。可通过filip.caron@kuleuven.be与他联系。

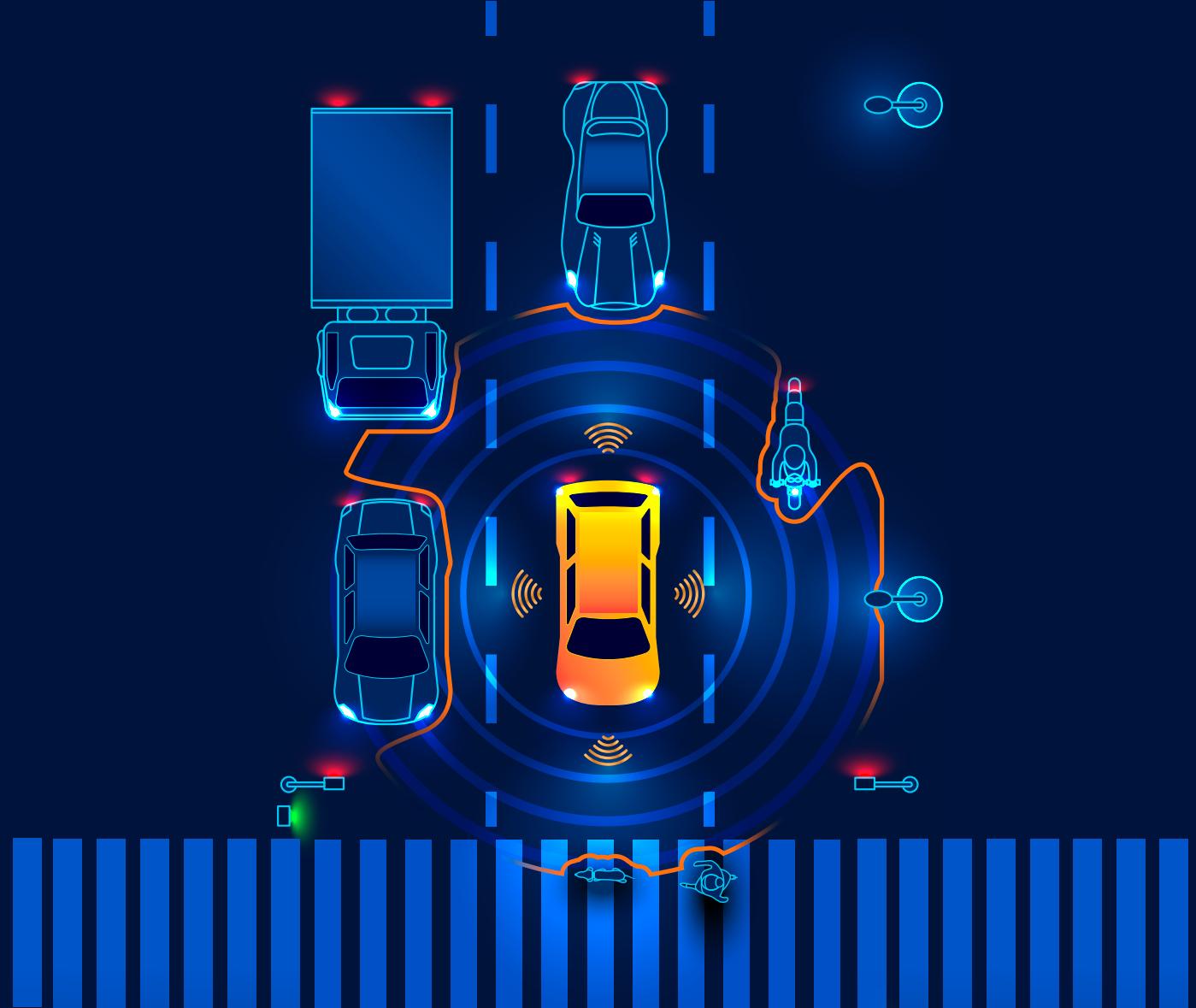
法规和相关要求可能会对支付市场的技术发展产生重大影响。批发支付系统通常面临严格的安全要求和新技术监管方面的不确定性。全球性的零售支付服务必须应对在法规各异的多个司法辖区内运营的挑战。另一方面，新的监管框架(如PSD2)正在为新进者开放零售支付市场。

并非所有支付领域的创新都由灵敏的金融科技创业公司推动。虽然这些新进者通过改善用户体验引领了零售支付的创新，但批发支付系统的改进是在既有企业的内部展开，这些企业在应对安全要求和监管上拥有丰富的经验。

最后，虽然世界经济论坛已将区块链确定为未来金融基础设施的一项关键技术，但下一代批发支付系统的蓝图并未采用区块链。欧洲央行和英格兰银行都宣布它们正在考虑DLT接口的可能性。■

参考文献

1. The Payment System: Payments, Securities, Derivatives, and the Role of the Eurosystem, European Central Bank, 2010; www.ecb.europa.eu/pub/pdf/other/
2. CBEST Intelligence-Led Testing: CBEST Implementation Guide, Bank of England, 2016; www.bankofengland.co.uk/-/media/boe/files/financial-stability/financial-sector-contINUity/cbest-implementation-guide.pdf.
3. Guidance on Cyber Resilience for Financial Market Infrastructures, report, Bank for International Settlements and International Organization of Securities Commissions, 2016; www.bis.org/cpmi/publ/d146.pdf.
4. A Blueprint for a New RTGS Service for the United Kingdom, report, Bank of England, 2016; www.bankofengland.co.uk/-/media/boe/files/payments/a-blueprint-for-a-new-rtgs-service-for-the-uk.pdf.
5. “The Fed’s Resource Center for Adoption of ISO 20022 Payment Messages,” Federal Reserve Banks, 2016; <https://fedpaymentsimprovement.org/payments-efficiency/iso-20022>.
6. Target Instant Payment Settlement User Requirements, report, European Central Bank, 2017; www.ecb.europa.eu/paym/initiatives/shared/docs/tips_user_requirements_document_final_version.pdf.
7. R. Roman et al., “Key Management Systems for Sensor Networks in the Context of the Internet of Things,” Computers & Electrical Eng., report, vol. 37, no. 2, European Banking Authority, 2011, pp. 147–159.
8. S. Sicari et al., “Security, Privacy and Trust in Internet of Things: The Road Ahead,” Computer Networks, 2015, pp. 146–164.
9. Regulatory Technical Standards on Strong Customer Authentication and Common and Secure Communication under Article 98 of Directive 2015/2366 (PSD2), report, European Banking Authority, 2017;



2018·AUTOCAR CONTEST

自动驾驶模拟大赛

奖金: 12,500 元 (颁发证书)

报名入口

线上初赛: 08月15日–10月15日

线下决赛: 10月25日



×



×





用不用区块链？ 这是个问题

文 | 瓦伦蒂娜·加泰斯基 (Valentina Gatteschi)，都灵理工大学 (Politecnico di Torino)

法布里齐奥·兰贝蒂 (Fabrizio Lamberti)，都灵理工大学 (Politecnico di Torino)

克劳迪奥·德马尔蒂尼 (Claudio Demartini)，都灵理工大学 (Politecnico di Torino)

基娅拉·普兰泰达 (Chiara Pranteda)，Reale Group

维克多·圣玛丽亚 (Víctor Santamaría)，Reale Group

译 | 郑米兰，浙江大学

区块链被认为是一项突破性技术。但贵公司是否需要它？在本文中，作者以保险业为例，讨论了区块链技术的优缺点。这些分析可以被推广运用到其他行业。



区块链是分布在某个网络中的公共分类帐，记录网络参与者之间执行的交易（从一个网络节点发送到另一个网络节点的消息）。在添加到区块链之前，每项交易都须由网络节点根据多数共识机制来做验证。记录在分类账中的信息无法被更改或删除。每项交易的历史可以随时更新。

区块链受到的关注日益增长，2016年对它的投资达到约3亿美元。早期采用者认为它是一项突破性技术，可以改变不同应用领域的许多日常活动和业务流程。例如，它可用于记录

选举投票，确保透明度。鉴于其跨国覆盖范围，它可用于追踪有形奢侈品和知识产权等。

根据高德纳顾问公司 (Gartner) 的“炒作周期”报告，区块链目前已达到“期望膨胀峰值”，意味着这项技术的潜力可能已被高估。¹企业可能发现自己难以客观地判断其优势和劣势并评估潜在的机会。

在本文中，我们采用IT专业人员的视角来决定其业务是否可以从采用区块链中受益，并尝试确定达成有效决策所需的信息。

虽然部分讨论将涉及一般商业机会和技术层面（参见“区块链如何运作”部分），但我们将重点关注保险行业这一特定的应用领域。这里有两方面原因。首先，基于保险业内部的技术颠覆炒作周期，² 区块链技术仍处于从“技术触发期”到“期望膨胀期”的曲线的初始部分，表明还存在着创新空间。其次，保险市场的特征（提供的服务、实施的流程、服务的客户等）使区块链的采纳成为一个尤其具有争议的问题，因为尚不清楚它是否值得投资。这方面的探索才刚刚开始：B3i是第一个以区块链为中心的保险业联盟。

尽管我们关注的是保险行业，但我们的观察和结论很容易被推广，以帮助专业人士处理其他应用领域里的问题。

区块链如何运作

区块链技术有时会被描绘为一条DNA长链。与新交易相关的信息会被添加到这根链条上，因而其长度会不断增加。交易信息以区块为单位聚集（“区块链”一名的由来），这些区块按时间顺序排序，每个区块链接到前一个区块上。这条链条由节点网络来维护，这些节点验证交易的有效性，并将它们添加到新的区块上，这个过程称为挖矿。

为更好地理解区块链的工作原理，请参考图1所示范例。艾丽丝想将一定

虽然本文会涉及到一般性的商业机会和技术，但是我们将重点关注保险行业这一特定领域。

数量的加密货币转给约翰（在此，转移的也可以不是加密货币，而是另一种具有数字对等物的资产）。加密货币被存储在数字钱包中，这个钱包由一个地址来标识。为进行转移，艾丽丝明确了要转移的金额和约翰钱包的地址。然后她将这项交易广播到整个网络。这项交易使用存储在钱包中的秘密信息进行数字签名，确保它来自艾丽丝的钱包并且不能被其他人更改。

其他网络节点通过分析数字签名来核查艾丽丝是否真的授权了这项交易。然后，他们通过在区块链的本地副本（存储了网络上的所有交易，包括与艾丽丝的钱包之间的转移）上计算艾丽丝的余额来验证她是否有权花这笔钱。如果交易可以进行，节点会将它插入新的区块中。

新的区块包含了需要被验证的所有交易的列表，并在其标题中记录关于这些交易的摘要（即散列，将一套数据集映射到固定大小的符号序列的数学函数）以及前一个区块的标题的散列。

为了将新创建的区块添加到区块

链中，节点启动挖矿竞赛，它们必须解决复杂的数学问题。这个过程被称为工作证明，它要求节点找到一个随机值，它与交易的散列和前一个区块标题的散列一起，生成一个特定的结果。当一个节点识别出一个可能的解决方案时，它会将结果向其他节点广播，供它们核查。如果大多数节点都认同该方案，则该区块被认为是有效的，并被添加到区块链中，使每个节点都更新其本地副本（获胜者也会获得奖励，比如交易费）。这一挖矿过程的结果是，约翰将在他的钱包里看到艾丽丝转来的金额。

这种复杂的验证机制使得单一节点几乎不可能控制大部分的网络，因为这需要极高的计算能力来创建虚假的区块、抢在其他节点之前解决数学问题，并就刚刚开采的区块获得51%的共识。此外，由于每个经验证有效的区块包含了前一个区块的信息（使用加密方法保护），这防止了对已记录的交易进行恶意篡改。实际上，更改交易也意味着更改包含它的区块以及后面的区块的摘要。

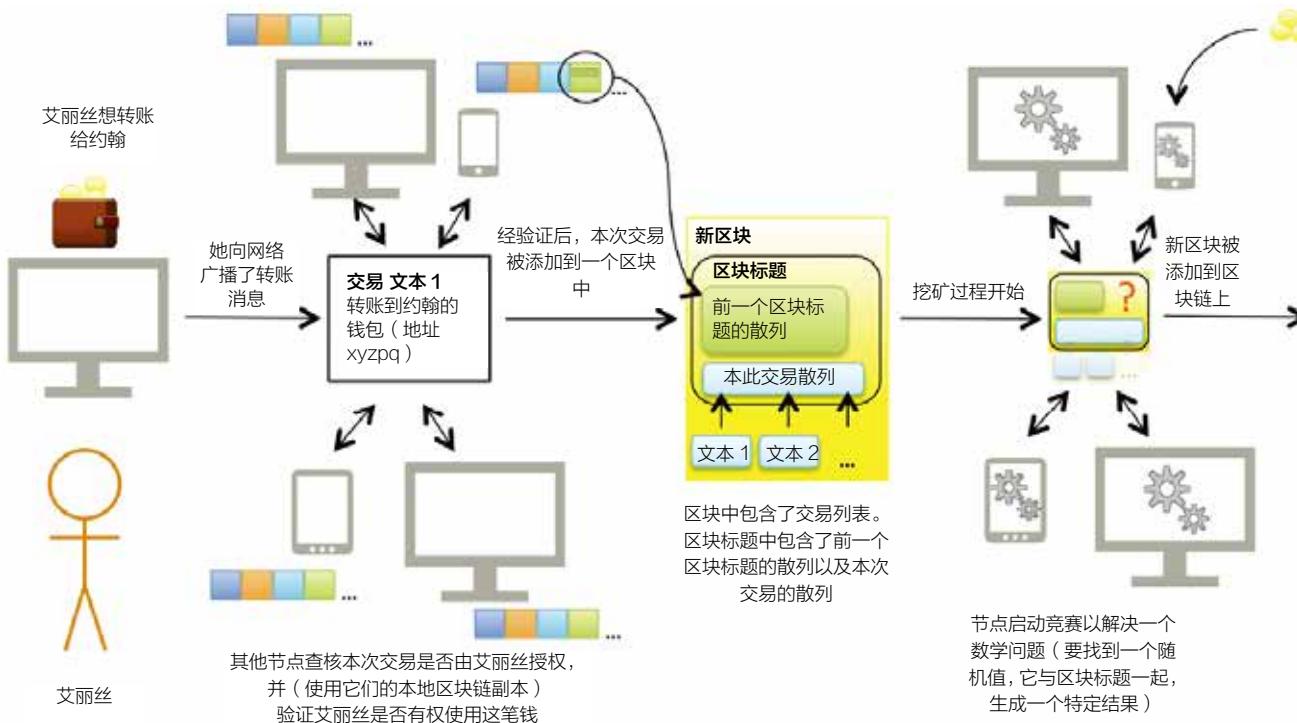


图 1. 区块链交易如何被记录。

区块链的演变和潜在应用

区块链技术构思于2008年，目的是以不可更改和可公开验证的方式记录比特币(<http://bitcoin.org>)交易。比特币是加密货币的第一个原型，它的发明是为了实现各方之间的资金转移而不是依赖中介机构。

随着时间推移，人们发现了新的应用场景（表1罗列了潜在应用）并开发出了许多原型，远远超出了资金转移的用途⁴（参见“现有的基于区块链的应用和原型”部分）。

可以观察到三种不同的区块链演变：区块链1.0、2.0和3.0。⁵ 区块链1.0与比特币等加密货币密切相关。⁶ 自比

特币诞生以来，已经出现了600多种加密货币（通常用作基于区块链的应用的交易代币）。其中最著名的（基于市值数据而言）包括以太坊（比特币一个众所周知的替代品，提供了一个框架来轻松创建基于区块链的应用程序；www.ethereum.org）、Monero（保证交易不可追踪；<http://getmonero.org>）和Ripple（支持即时付款，特别是银行转

帐；<http://ripple.com>）。

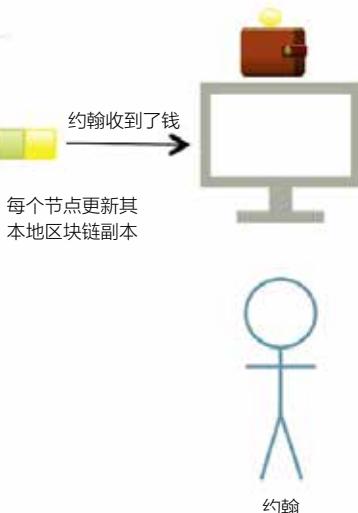
如果说区块链1.0的核心是钱，区块链2.0的核心是登记、确认和转让合同或财产。其应用领域包括使用区块链作为本地数据库（尤其是用作公共记录和证明的数据）的去中心化副本，以及其他更为复杂的应用。

区块链2.0最重要的功能是与智能

合约的集成（最初仅由以太坊提供，目前正在开发针对比特币的应用）。智能合约是存储在区块链上的代码片段，它们被编程，以在满足某些条件时以特定方式运行。它们可在不受第三方管控的情况下自动执行。例如，如果在区块链中编码了一份遗嘱，则智能合约可以在立遗嘱人死亡的情况下，自动将资产转移给受益人。

为了收集他们的激活条件所基于的信息，智能合约依赖于oracles-off-chain服务从现实世界中获取数据并将之推送到区块链。智能合约的应用可用于基于区块链的众筹活动，以在达到目标时自动触发付款；它们可以自动化投注系统，允许人们在事件上下注并向获胜者转账；或者它们可以与物联网设备⁷

赢得比赛的节点
可能获得奖励



一起使用，在付款后自动解锁智能酒店房间锁。

智能合约可以通过编写制定决策和管理人员的规则来创建新型组织，例如去中心化的自治组织(DAO)。

在区块链3.0中，应用领域不再局限于金融和商品交易，而是包括政府、医疗、科学、教育等。

在政府应用中，区块链可用于以不可更改和可公开验证的方式记录选举投票，从而提高选举的透明度。它还可以支持个性化治理系统，让公民只为实际使用了的服务付费。区块链也可用于发布政客的执政纲领，让每个人都可以访问这些内容并验证政客是否履行了承诺。

区块链不可更改的特性也令它有可能在那些执行审查制度的国家发挥优势，因为人们可以在区块链上发表自

己的想法，而不会被任何人删除或更改（不过，内容审查者还是可能找到新的办法，例如创建数千个新帖子来盖过那些他们想要删除的内容）。

区块链还可以帮助人们在其他领域中改善生活，例如医疗和科学。它可用于记录基因组数据（许多国家禁止访问这些数据）并使数据所有者可以访问这些信息。这可以帮助人们改变他们的生活方式，例如当他们发现自己容易患上某种疾病时。此外，研究人员也可以访问在人们检查身体或治病期间以及个人活动追踪器记录的大量健康数据。

在教育领域，智能合约可以用于管理财务奖励。例如，只有当学生成功通过某个学习模块的期末考试时才能拿到汇款。它们还可以记录学生的成绩、确保求职和招聘流程的透明度等。

在工业领域，区块链可以与大数据技术集成，以创建预测-反应系统，收集和存储大量将在稍后处理的信息，并通过结合人工智能和智能合约的力量使其具有执行力。

现有的基于区块链的应用和原型

许多基于区块链的应用已经存在于不同的行业和环境中。有些还处于原型阶段，其他一些已经开放给更广泛的公众使用。在此我们为目前为止在各个

行业出现的解决方案做一番概述。

个人数据管理。通常，这个领域中的应用使用区块链钱包来识别用户身份。它们的优点是用户可以使用唯一的标识符来登录，而不需使用传统的凭证 (<http://bitid.bitcoin.blue>)。此外，用户的数字身份在由经认证的组织验证一次后就可多次使用，从而使他们无需在必须开设银行帐户、购买保险等情况下分享自己的ID和其他个人信息 (<https://kyc-chain.com>)。

知识产权。此类系统通常依赖区块链来存储某个文件的散列及时间戳，以证明其存在和作者身份 (<http://proofofexistence.com>; <http://virtualnotary.org>)。一些平台还允许作者授权使用其作品 (<http://ascribe.io>)，并在其他人访问时通过智能合约接收自动付款 (<http://monegraph.com>; <http://ujomusic.com>)。

金融、交易和投注。一批金融公司正在展开投资，尝试在其应用中包含区块链，并允许用户使用加密货币进行支付。例如，纳斯达克广泛投资区块链技术以降低股票管理成本，并与Chain (<http://chain.com>) 建立合作伙伴关系以开发一项金融网络协议，存储有关已发行或交换的股票信息。几家银行最近采用Ripple来管理实时国际支付。⁹智能合约也大量应用于网络博彩，因为它们可以向赢家确保他们会拿到奖

主题文章：金融科技及其应用

表1.区块链应用。更具所交换的资产来分类。^{5,8}

类型	应用
总体	托管交易、有担保的合约、第三方仲裁、多方签名交易
金融交易	货币、股票、私募/公共股票、众筹、债券、共同基金、衍生品、年金、养老金、保险单、贸易/支出记录、小额信贷、慈善捐赠、航空里程
公共记录	土地/财产所有权、车辆登记、营业执照、企业所有权/合并/解散记录、刑事/法院/政府记录、婚姻/出生/死亡证明、投票ID、健康/安全检查、船舶/卫星登记、建筑/枪支许可、法医证据
半公共记录	学位/证书/学习成绩、人力资源/医疗/交付记录、基因组数据
私人记录	欠条、贷款、合同、投注、签名、遗嘱、信托
身份证明	驾照、身份证、护照、选民登记
认证	保险/所有权证明、公证文件
实物资产的钥匙	家庭/酒店客房、租赁汽车、储物柜、邮箱、物联网设备等的数字钥匙
无形资产	专利、商标、版权、保留权利、域名、数字权利、真实性/作者身份证明、许可证、网络身份
其他记录	文化/历史事件、纪录片、数据（天气、温度、交通、体育赛事比分）

金 (<http://lastis.me>; <http://etherpot.github.io>)。其他系统采用类似的方法，但旨在收集人们对未来事件的二选一预测结果 (<http://augur.net>; <http://gnosis.pm>)。如果用户预测正确，将获得奖励。

软件和互联网。区块链可用于记录系统日志，使攻击者无法删除或更改事件的历史记录 (www.reply.com/en/content/securechain)。其他应用包括使用区块链进行云存储。通常，此类应用会在节点的硬盘驱动器中记录部分文件，并根据借用的空间自动奖励节点 (<http://storj.io>; <http://ipfs.io>; <http://maidsafe.net>)。为避免重

复存储，每次上载新资源时，都会执行与其散列和其中一项已存储资源的对比。域名也被存储在区块链中。在此，目标是用基于区块链的服务器替换DNS服务器，用户可以自动注册域名并用加密货币支付 (<http://namecoin.org>; <http://blockstack.org>)。其他区块链的用途包括通过存储用户产生的内容 (<http://gist.github.com/metacoin/10dea79e15294950c8c3>) 来减少审查，或者根据读者投票来奖励内容生产者 (www.thankscoin.org)。

政府。区块链可用于以透明和可公开验证的方式收集公民投票 (www.reply.com/en/content/cloudchain)，到交易商品、产品和订阅服务，以及更昂贵的商品如汽车 (www.reply.com/en/content/thats-mine)。

ballotchain)。投票可以表现为从选民的钱包转移小额加密货币（或其他等价物）到候选人的钱包。通过这种方式，可以在任何设备（计算机、平板电脑、移动电话或多媒体显示屏）上投票，同时保证匿名性、唯一性和不变性。

商业和供应链。这个领域是受到最大关注和投资的部门之一，因为它可以使用区块链来识别假冒伪造商品，而从中受益。应用范围从交易体育和音乐门票 (www.re-ply.com/en/content/blockchain-ticketing-solution-cloudchain)，到交易商品、产品和订阅服务，以及更昂贵的商品如汽车 (www.reply.com/en/content/thats-mine)。

**智能合约也大量应用于网络博彩，
因为它们能确保赢家会拿到自己的奖金。**

在奢侈品市场中，区块链已被用作钻石及其所有权的全球分类帐 (<http://everledger.io>)，或用于跟踪和定位供应链上的货物 (<http://blockverify.io>)。其他应用使用区块链来改善供应链。比如Provenance (www.provenance.org) 使用区块链来实现食品的源头追溯。最后，像Profeth (<http://profeth.org>) 这样雄心勃勃的项目建议使用区块链和智能合约将智能嵌入供应链，并使得人们可以直接交换商品和服务而无需货币系统。

服务。一些举措建议在现有服务中包含加密货币支付或基于区块链的交易存储。比如La'Zooz (<http://lazooz.org>) 就是一个基于区块链的类优步平台。

物联网。这是区块链技术另一个有前景的领域。在这里，区块链已被用于支持设备之间的互操作性，证明设备收到的消息是由受信任的设备发送的 (www.reply.com/en/content/blokcom; www.re-ply.com/en/content/authentichain)。其他原型

构思的前景是，区块链可以使智能洗衣机在有需要时自动订购洗涤剂，或在发生故障时发出维修要求。¹⁰ 另一个创新的解决方案是Slock.it (<https://slock.it>)，它提出了基于区块链的房屋租赁系统，只允许拥有有效预订的人士打开智能锁。Slock.it团队还参与了另一个名为Blockcharge的项目，该项目依靠区块链来实现电力共享（例如为电动汽车充电）。最后，Lo3Energy (<https://lo3energy.com>) 提出了家庭生产能源P2P交易应用，让家庭可以直接向邻居出售自己生产的电而无需依赖第三方。

医疗保健。在医疗保健领域，基于区块链的解决方案用于收集重要数据和位置信息，并在发生危险时发送警报。依赖区块链技术的优势是确保系统不会停止工作。

区块链的优缺点

表1显示区块链应用的潜力之大，范围之广。然而，一些专家称区块链技术被夸大¹¹、尚未成熟¹²，而且它的某些

应用目标通过其他已掌握的技术就可解决。¹³

那些希望投资这项技术的人应该把力量集中到未解决的问题和新的需求上，并且要意识到，区块链并非先验的最佳解决方案，因为其优势因部门而异，从一个案例到另一个案例各有不同。区块链也存在一些缺点，¹⁴在决定采用它之前应仔细评估。

在这里，我们总结了区块链的主要优缺点。因为我们已经讨论了区块链的潜力，我们将简要地总结其优点，而更详细地讨论其缺点。

优点

- > 实现由对等方共同维护的共享存储库——每个人都可以访问数据并查看交易。此外，在节点上存储信息可防止在意外事件发生时丢失数据。
- > 实现各方之间的信任。数字签名和验证确保每个节点和用户行为正确，而无需中介干预。
- > 可以成为不同参与者访问的全球数据存储库。每个人都可以读/写它。
- > 保证了透明度。每个人不仅可以阅读交易的最终状态，还可以阅读过去交易的历史。
- > 不变性。数据无法被删除或更改。

- > 权力下放。它可以在没有中央部门的情况下运行，不能被控制、审查或关闭。
- > 自动化。通过智能合约，活动可以实现自动化。

缺点

- > 耗电量大。考虑到网络节点消耗的电力，一次比特币交易可能花费6美元。¹⁵
- > 挖矿需要昂贵的硬件，而其中大部分计算能力都被浪费了。挖矿是节点之间的竞赛，唯速度最快者获胜，所以其他的都只是浪费资源。为增加获胜的可能性，节点可以加入挖矿池并与其他节点协作，共享收入。一个减少必须计算能力的解决方案是将挖矿过程从工作证明变成投注证明，节点可以用代币购买挖矿机会，并根据自己拥有的代币数量来相应地扩展挖矿能力。这样，挖矿就变得不那么资源密集，而仅限于代币持有者。
- > 数据的复制需要空间。区块链的本地副本（也就是自创建以来发生的所有交易——比特币为约105 GB，以太币为70 GB；<http://bitinfocharts.com>）被存储在每个网络节点上。因此，其表现尚不能与数据库相比。
- > 添加信息很慢。创建一个比特币

区块需要大约10到60分钟（<http://blockchain.info/charts/avg-confirmation-time>）。以太坊需要15秒，（<http://etherscan.io/chart/blocktime>），虽然相比较而言短了很多，但本身仍然是很显著的耗时。

- > 不可变性和透明度可能会损害用户的隐私和声誉。每个网络节点都会存储区块链的副本，并有可能访问其内容。
- > 智能合约不能依赖外部应用程序接口。每个节点都应能查看以前的交易，并得到与其他节点一致的结果。也就是说，信息必须是不可变更的。因此，智能合约所需的数据须首先被注入到区块链中。实现这种注入需要一个强大的声誉系统或治理机制，它需要和区块链本身一样强大，而不是成为流程中最薄弱的环节。
- > 智能合约可能出错。因为它们的代码是公开可获得的，并且一旦创建它们就成为了自治实体，它们就可能成为黑客钟爱的目标。¹²由于它们被存储在区块链中，智能合约无法被修改。要删除代码错误，开发人员必须创建新的合约，并将所有数据和提示从旧的转移到新的。最著名的智能合约受攻击案例发生在2016年6月，以太坊上约6

千万美元“被盗”。

区块链应用的主要问题

对区块链最常见的批评之一，是大量基于区块链的应用都可以使用现有技术——比如（安全的）中央化数据库——来执行。此外，专家们还确定了以下关键问题¹¹ 供人们确定区块链技术是否是适合他们的解决方案。

- > 是否一定需要一个共享的数据库？
- > 是否有必要让多方写入数据？
- > 潜在的写入者是否不受信任（写入者是否应被禁止修改以前由他人写入的数据）？
- > 是否需要去中介（是否有必要移除验证交易的可信中介）？
- > 是否有必要了解交易之间如何相互关联（不同参与者是否应独立编写有关单个用户的交易）？

如果上述问题中的一个或多个得到否定的答案，那么区块链技术很可能在没有带来真正好处的情况下增加成本。

一个实际案例研究： 保险业

保险行业最近对区块链表现出了浓厚的兴趣。一些大公司已经进行了大

对区块链最常见的批评之一，是大量基于区块链的应用都可以使用中央数据库等现有技术执行。

合约汇集信息，以展开自动和精确的高级计算 (www.reply.com/en/content/insurechain)。

按使用付费的保险

智能合约可以实现按使用付费的保单，依靠物联网来自动购买保险。只有当客户的GPS坐标（由他们的智能手机收集）确认他们在国外时，才能收取旅行保险费。同样，只有在顾客开车时才能支付汽车保险。

量投资，以探索其业务潜力¹⁶，而咨询公司已调查了其对保险业的适用性。¹⁷⁻¹⁹

区块链的应用

在此，我们描述保险行业中的五个应用前景。

庭保险。例如，如果配备有传感器的屋顶受损，则可以触发自动报销。

智能合约的另一个优点是，它们可能会使传统的基于文本的合同中存在的模糊性消失，因为智能合约的所有条款都是硬编码的。这可以提高透明度，降低法律纠纷的频率和影响。

使用智能合约来改善客户体验并降低运营成本

智能合约的自动执行能力可以加快索赔处理（客户甚至会在他们做出索赔申请之前就收到钱，因为智能合约可以在特定事件发生后立即自动触发赔偿）并减少人力。例如，汽车保险智能合约可以被编程设定为只在客户通过受认证的机械师修理汽车时支付赔偿。

在农业保险中，农民会购买保险来应对恶劣天气的风险。智能合约可以读取天气数据，并且在持续不利的情况下授权赔付。

另一个例子是延迟保险，智能合约可以在航班或火车延误时自动退款给旅客 (<http://insureth.mkvd.net>)。

智能合约可与物联网一起用于家

预防欺诈

在全球范围内购买的共享区块链保单¹⁷，加上可能从其他领域获取的数据（医疗报告、警方报告等）可以帮助识别索赔过程中的欺诈行为。

数据录入 / 身份验证

基于区块链的身份验证系统可以在客户购买或更新保单时减少数据输入过载。客户应首先进行身份识别流程，其身份由经认证的中介检查并与其钱包相关联。然后，智能合约可以自动检索身份信息并查核它们。在更复杂的情形下（假设不同参与者广泛采用区块链），区块链可以存储额外的信息（例如健康检查、资产所有权等），并通过智能

P2P 保险

P2P保险并不是什么新想法（几年前就出现了像insPeer、Friendsurance和heyguevara.com这样的服务）。尽管如此，智能合约可以在这一领域提供创新，因为它们可以允许创建去中心化组织，其功能规则是硬编码的。通过这类组织，保险业者团队将能够实现自我管理。

对关键问题的答案

在这里，我们使用前面列出的关键问题来分析五种应用场景中的每一种，以确定是否需要区块链（结果总结在表2中）。

使用智能合约来改善客户体验并降低运营成本

数据收集自多个参与者和来源。对共享数据库的需求取决于单个应用：例

表2.基于五个关键问题分析应用场景 (+: 肯定; -: 否定; +/-: 两种答案都可能适用, 取决于具体情况)。

	共享数据库	多方写入者	不受信任的写入者	去中介化	关联交易
使用智能合约来改善客户体验并降低运营成本	+-	+-	+-	+-	+-
预防欺诈	+	+	+	+	+
数据录入/身份验证	+	+	+	+	+
按使用付费的保险	+-	+-	+-	-	-
P2P保险	+	+	+	+	+

如, 对于汽车维修来说, 可能需要由机械师编写并由公司检查的共享数据库。但是, 我们可以假设公司和经认证的机械师签署了协议并且各方之间存在信任。未经认证的机械师发送的文件将被手动处理。在其他情况下, 公司可依靠外部应用程序接口从在线服务或传感器检索所需信息, 并相应地更新数据库。

脱媒是一个有争议的问题——在索赔处理中, 保险公司实际上充当中间人, 协调政策数据与损害评估。此外, 大量索赔可能无法自动处理, 因为在结算之前, 评估员仍需要对其进行评估。当损害可以自动评估(比如通过物联网技术)或客户不信任公司时, 可能需要非中介化。在发达国家, 保险公司的行为由若干旨在确保信任的法规来规范。因此, 脱媒并不是一种明智的需要。在某

些情况下, 可能会出现必须检索那些链接到某种保单或某个人的所有交易或事件(比如用于客户关系管理)。然而, 这类操作可以通过传统系统来执行。

预防欺诈

由多方(比如医生和警察)编写的共享数据库可能会有所帮助。插入的数据不应被其他方修改, 并应与客户的数字身份相关联。尽管受信任的中间人可以保证记录信息的真实性, 但他们的存在会增加成本。因此, 去中介可能是更好的选择。在这方面似乎可以从区块链受益。尽管如此, 还是应该考量一些问题。首先是需要临界规模, 因为不同的参与者将插入大量数据。其次, 应采取措施确保数据隐私。最后, 由于无法更改区块链数据, 因此应该建立错误管理系统。

数据录入 / 身份验证

与防欺诈类似, 将需要一个共享分类帐来记录客户的文件、所有权证明等, 并将它们与客户的数字身份相关联。这里需要防止写入者修改以前由其他人录入的信息。与防欺诈一样, 去中介可以降低成本。区块链可以是一个很好的解决方案, 它可以实现信息的编写和共享, 只要上面提到的问题得到妥善解决。

按使用付费的保险

区块链可用于证明客户是否或何时激活了保单。多个编写者(公司及其客户)将在一个共享数据库上写入数据, 并且需要保证不能修改保单数据。脱媒似乎不太重要, 因为只涉及两个参与者(客户和保险公司), 而且保险公司已经充当了中间人。同样, 公司已经存储了有关客户先前交易的信息, 使得关联交易不像其他要求那样重要。看来, 在这个场景中, 区块链不会像在之前的两个场景里那样具有破坏性。而且我们应该注意到, 在许多国家, 客户和保险公司之间的信任已经通过监管法规得到保证, 因而无需再依赖可信赖的第三方。但是, 未来大众媒体将越来越多地宣传区块链技术, 因此使用基于区块链的按使用付费保险可能具有竞争优势。如果允许客户使用比特币或其他加密货币支付, 可能会进一步增加这种优势。

P2P 保险

这个场景需要去中介，而智能合约可以管理一起编写共享数据库的欠缺信任的多方之间的交互。它还可以从关联交易中受益，例如在客户身份验证或查询索赔记录上。因此，依靠区块链可能是一个不错的选择。尽管如此，P2P 保险旨在移除中间人。因此，基于去中介化组织的保险服务的传播可能对传统业务构成重大威胁。

讨论

我们的分析表明，部分而不是所有的保险业场景都可以从区块链的采用中受益。如果保险公司采用它，它们可能会开始用它来简化数据输入和客户身份验证。实际上，即使这种情况也需要多个参与者在区块链中插入信息，但涉及参与方的数量将少于其他场景。此外，银行（已经对区块链进行了重要投资）可能对开发或维护共享分类账感兴趣。保险公司和银行可能依赖已有的应用来减少初始投资。

之后，力气可能会被花在开发基于区块链的防欺诈系统上。这将是一项长期的高风险投资，这是因为其成功将需要多方参与和界定标准来存储信息。一开始，保险公司可以在全球范围内存储与客户数字身份相关联的保单和索赔申请，从而减少信息不对称。然后，其

他部门可能会逐渐参与其中。

接下来，保险公司可以投资创建基于区块链的P2P保险基础设施，将潜在威胁转化为商机（比如可以对每份保单收取一笔费用）。在这方面，市场调查发现，众多客户仍然认为与中介机构的个人互动很重要。²⁰ 因此，向P2P保险的转变可能并非迫在眉睫。

如果保险公司希望覆盖新兴市场（信任机制尚未完全建立的地方）或经营微型保险（这在过去因为人力密集的行政环节和针对小额支付的高额费率并不可行，而区块链解决了这些问题），那么使用区块链和智能合约来降低运营成本、改善客户体验和提高透明度可能会成为一个关键选择。尽管如此，保险公司应该考虑到，在某些情况下仍然需要人工来评估和处理索赔。

最后，在按使用付费的保险中，区块链可以提供投保证据。但是，在这方面，有兴趣的投资者应首先核实相关国家规定，例如，是否要求签署文件？如果是，就需要修改国家规则。此外，如果相关保险公司是有信誉的公司，可以使用其他机制来确认投保行为。

结语

尽管区块链可以给许多行业带来创新，而且人们对它也满腔热情，我们不应将它视为“魔力子弹”，而是应根

据公司所在的行业和业务目标来仔细评估是否需要采用它。虽然我们评估了一个特定领域，但我们的分析可以很容易地扩展到其他具有类似应用的场景中，从而帮助专业人员在不同的环境和行业中做出决定。■

参考文献

1. “Gartner: Blockchain and Connected Home are Almost at the Peak of the Hype Cycle,” PR Wire, blog, 2016; <https://prwire.com.au/pr/62010/gartner-blockchain-and-connected-home-are-almost-at-the-peak-of-the-hype-cycle>.
2. S. Gilbert, “The Hype Cycle of Insurance Disruption,” InsuranceThoughtLeadership.com, blog, 2016; <http://insurancethoughtleadership.com/the-hype-cycle-of-insurance-disruption>.
3. S. Higgins, “European Insurance Firms Launch New Blockchain Consortium,” Coindesk, blog, 2016; www.coindesk.com/european-insurance-blockchain-consortium.
4. G. Hurlburt, “Might the Blockchain Outlive Bitcoin?,” IT Professional, vol. 18, no. 2, 2016, pp. 12–16.
5. M. Swan, *Blockchain: Blueprint for a New Economy*, vol. 3, no. 3, O'Reilly Media, 2015, pp. 38–69.

关于作者

Valentina Gatteschi是都灵理工大学的博士后研究助理。她的研究兴趣包括语义处理、智能系统和人机交互。她在都灵理工大学获得计算机工程博士学位。可通过valentina.gatteschi@polito.it与她联系。

Fabrizio Lamberti是都灵理工大学的副教授。他的研究兴趣包括计算智能、语义处理、分布式计算、人机交互、计算机图形和可视化。他拥有都灵理工大学计算机工程博士学位。他是《IEEE计算新兴议题汇刊》(IEEE Transactions on Emerging Topics in Computing) 和《IEEE消费电子品杂志》(IEEE Consumer Electronics Magazine) 的副主编。可通过fabrizio.lamberti@polito.it与他联系。

Claudio Demartini是都灵理工大学的正教授，教授信息系统、创新和产品开发课程。他的研究兴趣包括软件工程、架构、智能系统和教育。他是都灵理工大学控制和计算机工程系的系主任，也是该校学术评议会成员。可通过claudio.demartini@polito.it与他联系。

Chiara Pranteda是意大利保险公司Reale Group创新团队的成员。她拥有都灵大学的物理学学位。可通过chiara.pranteda@realemutua.it与她联系。

Víctor Santamaría是Reale Group下属Reale ITES公司的技术和数字创新主管。他的研究兴趣包括技术和商业创新。他拥有马德里康普顿斯大学的数学学位，并在IE商学院取得了数字创新、IT治理、信息系统管理等多个在职进修硕士学位。可通过victor.santamaria@realeites.com与他联系。

Financial Perspectives, report, vol. 3, no. 3, McKinsey & Company Report, 2015, pp. 38–69.

9. A. Jarrett, “Ripple and R3 Team Up with 12 Banks to Trial XRP for Cross-Border Payments,” Ripple, blog, Ripple, 2016; <https://ripple.com/insights/ripple-and-r3-team-up-with-12-banks-to-trial-xrp-for-cross-border-payments>.

10. S. Higgins, “IBM Reveals Proof of Concept for Blockchain-Powered Internet of Things,” Coindesk, blog, 2015; www.coindesk.com/ibm-reveals-proof-concept-blockchain-powered-internet-things.

11. G. Greenspan, “Avoiding the Pointless Blockchain Project,” MultiChain, blog, 2015; www.multichain.com/blog/2015/11/avoiding-pointless-blockchain-project.

12. F. Zaninotto, “The Blockchain Explained to Web Developers, Part 3: The Truth,” Marmelab, blog, 2016; <https://marmelab.com/blog/2016/06/14/blockchain-for-web-developers-the-truth.html>.

13. A. Cooper, “Does Digital Identity Need Blockchain Technology?,” Gov.UK Verify, blog, 2016;

6. F. Tschorisch and B. Scheuermann, “Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies,” IEEE Communications Surveys Tutorials, vol. 18, no. 3, 2016, pp. 2084–2123.
7. K. Christidis and M. Devetsikiotis, “Blockchains and Smart Contracts for the Internet of Things,” IEEE Access, report Deloitte Report, vol. 4, 2016, pp. 2292–2303.
8. M. Mainelli and M. Smith, “Sharing Ledgers for Sharing Economies: An Exploration of Mutual Distributed Ledgers (aka Blockchain Technology),” J.

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



13. <https://identityassurance.blog.gov.uk/2016/08/15/does-digital-identity-need-blockchain-technology>.
14. M. Peck, "The Blockchain Has a Dark Side," IEEE Spectrum, vol. 53, no. 6, 2016, pp. 12–13.
15. "Let's Quit the Blockchain Magic Talk," ZDNet, blog, 2016; www.zdnet.com/article/lets-quit-the-blockchain-magic-talk.
16. "The March of Financial Services Giants into Bitcoin and Blockchain Startups in One Chart," CBInsights, blog, 2017; www.cbinsights.com/research/financial-services-corporate-blockchain-investments.
17. A. Shelkovnikov, Blockchain Applications in Insurance, report, Deloitte, 2016; www2.deloitte.com/content/dam/Deloitte/ch/Documents/innovation/ch-en-innovation-deloitte-blockchain-app-in-insurance.pdf.
18. Blockchain in Insurance—Opportunity or Threat?, report, McKinsey & Company, 2016; www.mckinsey.com/industries/financial-services/our-insights/blockchain-in-insurance-opportunity-or-threat.
19. Blockchain Technology as a Platform for Digitization—Implications for the Insurance Industry, report, Ernst & Young, 2016; [www.ey.com/Publication/vwLUAssets/EY-blockchain-technology-as-a-platform-for-digitization/\\$FILE/EY-blockchain-technology-as-a-platform-for-digitization.pdf](http://www.ey.com/Publication/vwLUAssets/EY-blockchain-technology-as-a-platform-for-digitization/$FILE/EY-blockchain-technology-as-a-platform-for-digitization.pdf).
20. Voice of the Customer—Time for Insurers to Rethink their Relationships, report, Ernst & Young, 2012; [www.ey.com/Publication/vwLUAssets/Global_Consumer_Insurance_Survey_2012_-_The_Americas/\\$FILE/EY_GIR_AMERICAS_SML.pdf](http://www.ey.com/Publication/vwLUAssets/Global_Consumer_Insurance_Survey_2012_-_The_Americas/$FILE/EY_GIR_AMERICAS_SML.pdf).

微信名：计算机
微信号：jisuanren



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



2018 BYTE CUP

International Machine Learning Competition

国际机器学习竞赛

AAI × ByteDance × IEEE中国

总奖金金额

\$20,000

40名IEEE会员或学生会员资格

第二届 Byte Cup 国际机器学习竞赛详解

文 | 本刊编辑部

2018 Byte Cup 国际机器学习竞赛（以下简称 ByteCup）是一项面向全球的机器学习竞赛，由中国人工智能学会和字节跳动联合主办，数据竞赛平台 biendata.com 承办，旨在促进机器学习的学术研究和具体应用。2016 年，首届 Byte Cup 举办，共吸引全球 1000 多支队伍参赛，共同完成“为头条问答的问题在今日头条专家用户中寻找潜在的答案贡献者”这一竞赛任务。2018 年 ByteCup 发布的新任务是“自动为文章生成标题”。

我们正处于一个信息爆炸的时代。据 IDC 统计，互联网数据量已跃至 ZB 级别（1ZB=240GB），预计 2020 年达到 35ZB。我们每天也能看到大量的信息，包括新闻快讯、社交网络更新、事件评

论，以及专业教程等等。此外，搜索引擎并不能彻底解决信息过载问题，而移动设备的普及和碎片化的阅读方式让这个问题更加严重。因此，如果可以开发出自动提取文本摘要的机器学习模型，就可以帮助消化海量的信息。

文本摘要一直都是机器学习领域一个重要的热点，但是却有很大的难度。例如，给单篇文章起标题/摘要的时候，很难有词频作保证，而是需要模型可以理解内容，甚至做一些推理。在很多地方，摘要生成和机器翻译有类似之

处。然而，和机器翻译不同的是，自动文本摘要的输入和输出很不平衡，此外机器翻译任务的输入输出序的列通常有一些词义层面上的直接对应，这种对应在摘要任务中却没那么明显。

字节跳动于2012年成立，旗下有大量产品可以让用户阅读、消费和创作内容，包括大家熟悉的今日头条和抖音短视频，以及很多海外产品，如TopBuzz 和 Tik tok。此外，字节跳动还在2016年成立了AI实验室，实验室覆盖了很多不同的领域，包括机器学习、自然语言处理、计算机视觉、语音识别、数据挖掘、人机交互等，而自然语言处理领域又有机器写作、机器翻译、NLP 基础、问答、对话以及推荐/搜索等方向。

竞赛任务

Topbuzz是字节跳动为北美和巴西的用户创造的一站式内容消费平台，它利用机器学习算法为用户提供个性化视频、GIF图、本地新闻及重大新闻。目前，Topbuzz每天都会发布大量的文章，利用人工智能为创作者提供更好的标题选择，是Topbuzz优化用户体验的探索方向之一。此次竞赛任务便是为Topbuzz 提供的英文文章自动生成标题。

比赛奖励

前10名的队伍将获得证书和字节跳动招聘面试直通车的机会。周冠军将

获得周冠军证书。

1. 总奖金 2 万美元

一等奖 1 名: 10,000 美金

二等奖 2 名: 每名 3,000 美金

三等奖 4 名: 每名 1,000 美金

2. 周冠军自2018年8月27日起，评

测系统将选出周冠军。周冠军由过

去 7 天内的提交成绩决定。

of Hollywood superstar Tom Cruise and America's sweetheart Katie Holmes...",

"id": 1198440,
"title": "Suri Cruise 2018: Katie Holmes Bonds With Daughter During Dinner Date While Tom Cruise Still MIA"}

文档一共包含三类信息：

1. 文章id (id)：每篇文本对应一个 unique id；

2. 文章内容 (content)：即文章的内容字符串；

3. 文章标题 (title)：文章的标题，参赛选手需要自己生成验证集和测试集的标题。

验证集和测试集格式和训练集相似，但是没有 title，需要参赛选手预测。

说明：训练集用于模型的学习，验证集用于在线实时评估算法效果，测试集用于最终的效果评测。

比赛时间

8月15日：比赛开放提交答案入口

8月15日–11月15日：比赛期（比赛期间可随时报名）

11月16日：发布最终测试集

11月17日：公布测试集排名，排名

前10的队伍一周内提交一篇不超过4页的参赛方法说明 (ACM 双列标准模板，中英文皆可)。

11月30日：公布最终排名。

数据集

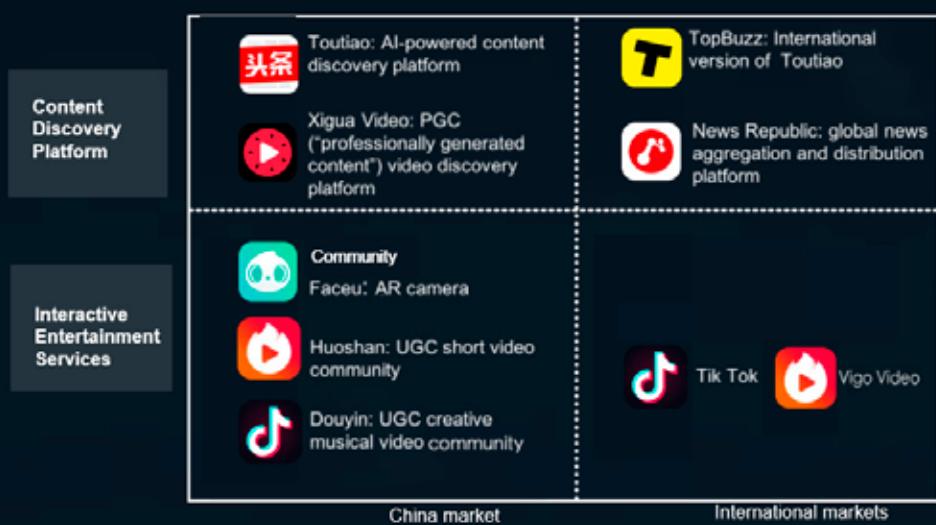
本次竞赛使用的训练集包括了约130 万篇文本的信息，每篇文本都是一个类似json 格式的行，如以下内容：

{"content": "Being the daughter

相关研究

目前，自动文档摘要技术主要分为抽取式 (extractive) 和摘要式 (又叫生成式) (abstractive) 两种。

Products: Whole Spectrum of Content and Geography



抽取式摘要相对较为成熟。这种方法利用如text rank这样的排序算法，对处理后的文章语句进行排序。不过抽取式摘要在语义理解方面考虑较少，无法建立文本段落中的完整的语义信息。

相较而言，生成式技术需要让模型理解文章语义后总结出摘要，更类似人类的做法。不过这种技术需要使用机器学习技术，长期以来并不成熟。转折点出现在2014年。这一年，Bengio 等人发表论文 Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation，正式引入了sequence-to-sequence模型。这一模型通过两个推理神经网络，分别把输入文本转化成向量，再把向量转成输出序

列。这种模型在论文中主要用来完成机器翻译任务，并且后来被应用在谷歌翻译中，但后续在文摘生成任务中也产生了广泛的应用。此后，这种利用深度学习的 sequence-to-sequence方法不断被改进，在一些标准的评测数据集（如 DUC-2004）上，已经超过了传统的抽取式方法。

例如，2016年，Facebook AI 实验室（FAIR）的学者发表论文 A Convolutional Encoder Model for Neural Machine Translation，在编码的时候用CNN取代RNN，获得不错的效果。Salesforce的研究人员 2017年发表的论文 A Deep Reinforced Model for Abstractive Summarization 中，使用了增强学习，在CNN/Daily

Mail数据集上的ROUGE-1分数达到41.16分。同年，又是FAIR发表了论文Convolutional Sequence to Sequence Learning，引入attention 机制，不仅提高了评测分数，还极大地提升了速度。

2016年，来自IBM沃森的研究人员发表论文Abstractive Text Summarization Using Sequence-to-Sequence RNNs and Beyond，和之前的论文不同，这篇论文把机器翻译和文本摘要任务完全分开，专门针对文本摘要提出了更合适的模型，除此之外，文章还发布了两个新的数据集。

中国也有许多学者在从事相关工作，例如北大的万小军老师。他和姚金戈的综述《自动文摘研究进展与趋势》



友情提示，因涉及到数据下载，强烈建议大家登录 PC 页面报名参加~

2018 BYTE CUP

International Machine Learning Competition



竞赛指导委员会

竞赛指导委员会主席：

张钹 清华大学
马维英 字节跳动

指导委员会资深委员：

刘挺 哈尔滨工业大学
孙茂松 清华大学

竞赛程序委员会

竞赛程序委员会主席

杨强 香港科技大学
李航 字节跳动人工智能实验室

竞赛委员会执行主席

唐杰 清华大学
李磊 字节跳动人工智能实验室

竞赛委员会资深委员

Polo Chau 佐治亚理工学院
方飞 卡耐基梅隆大学
黄民烈 清华大学
廖祥文 福州大学
刘知远 清华大学
邱鹏程 复旦大学
孙怡舟 加州大学洛杉矶分校
童行行 亚利桑那州立大学

屠可伟 上海科技大学
王健 东北大学
王威廉 加州大学圣塔芭芭拉分校
熊德章 苏州大学
薛业翔 普渡大学
李锐 谷歌
孙欢 俄亥俄大学

× ByteDance × IEEE中国



把摘要技术框架总结成 4 个步骤：内容表示 → 权重计算 → 内容选择 → 内容组织，并对每个步骤都作了介绍。

自动文摘研究进展与趋势：

<http://qngw2014.bj.bcebos.com/zhuankan/12/%E8%87%A8%E5%8A%A8%E6%96%87%E6%91%98%E7%A0%94%E7%A9%B6%E8%BF%9B%E5%B1%95%E4%B8%8E%E8%B6%8B%E5%8A%BF.pdf>

比赛组委会

主办：中国人工智能学会

联合主办：字节跳动

联合组织：IEEE中国代表处、

Biendata.com

参赛方式

扫描参赛二维码可直达赛事页面，
注册网站 - 下载数据，即可参赛。





发展中国家的 区块链

文 | 尼尔·科西特里 (Nir Kshetri)，北卡大学格林斯堡分校 (University of North Carolina at Greensboro)

杰弗里·沃阿斯 (Jeffrey Voas)，IEEE 会士 (IEEE Fellow)

译 | 吴岑，四川大学

发展中国家的很大一部分人口可以从区块链技术中受益。根据《信息技术事实与数据》(ICT Facts and Figures) 2017年报告，发展中国家已有 42.9%的家庭可以访问互联网。¹随着越来越人买得起智能手机，这一比例还在快速上升。可以说，在许多方面，区块链对发展中国家的价值远高于发达国家。为什么？因为区块链可能会填补有效而正式的制度的匮乏，包括规则、法律法规及其贯彻执行。在本文中，我们将讨论发展中国家制度存在的一些关键问题，评估用区块链解决这些问题的可能性。

产权

据联合国的一份 2011 年报告，治理薄弱导致超过 61 个国家在土地使用和管理中存在腐败——从小规模贿赂，到国家、州和地方政府滥用权力等。

产权的执行能激励投资并提供避免贫困的资源。议定的产权允许企业家

将资产用作抵押，从而增加他们获得资本的机会。但是，很大一部分穷人缺乏产权。

非洲农村约有九成土地未被登记在案或注册产权。同样，在印度，土地所有者的缺乏仍是创业和实现经济发展的障碍之一。³ 根据一项估计，印度有超过 2000 万户农村家庭没有土地，此外还有数百万农民没有合法拥有他们建造了房屋及耕耘劳作的土地。在印度，没有土地可以说是比种姓或文盲更有力的预测贫困的因素。⁴ 此外，据美国国际开发署 (USAID) 统计，只有 14% 的洪都拉斯人合法拥有自己的财产。在合法占用的产权中，只有 30% 登记在案。⁵

政府官员更改已登记财产的拥有者这种事并不罕见，有时他们把变更了所有权的财产分配给自己。据报道，官僚们常常更改那些海滨房产的所有权并登记在自己名下，⁶ 有时他们也接受一些人为拿到财产所有权而提供的贿赂。公民通常难以访问产权记录，而那些可访问的记录可能会提供相互矛盾

的信息。业主往往无法为财产使用权和矿产权被侵犯而自我辩护。⁷

区块链可以减少摩擦和冲突，以及与财产登记相关的成本。可以用智能手机来完成全部或大部分过程。⁸ 在这方面，各种各样的尝试已经展开，令人鼓舞。总部位于美国的房地产登记平台 Bitland 宣布在加纳推出基于区块链的土地登记系统。加纳有 78% 的土地未登记。⁹ 该国法院积压了大量土地纠纷案件。10Bitland 安全地记录交易，配备 GPS 坐标、书面描述和卫星照片。这种方法和其他类似的方式预计将保障产权并减少腐败。截至 2016 年中，加纳有 24 个社区对该项目表达了兴趣。

9Bitland 计划与欧佩克国际开发基金合作，把该系统扩展到尼日利亚。¹¹

比特币公司 BitFury 和格鲁吉亚政府签署了一项协议，共同开发使用区块链登记土地所有权的系统。¹² 目前，在格鲁吉亚买卖土地时，买方和卖方必须求助于公共登记部门。他们将支付 50 至 200 美元——具体价格取决于他们

希望交易被公证的速度。该试点区块链项目将把登记的过程转移到区块链，买家和卖家支付的成本预计在 5 到 10 美分之间。¹³

2017 年，印度的泰伦加纳邦和安得拉邦宣布了将区块链用于土地登记的计划。泰伦加纳邦在首府海得拉巴 (Hyderabad) 开始了试点项目。2017 年 9 月的报道称，该项目将在一年内全面覆盖海得拉巴及附近地区。¹⁴ 2017 年 10 月，安得拉邦政府与瑞典创业公司 ChromaWay 合作，为计划中的城市阿马拉瓦蒂 (Amaravati) 创建一个基于区块链的土地登记系统。

抑制腐败

区块链创建了一个防篡改的数字交易分类账，并共享分类账从而提供透明度。加密设置保证了访问和扩展分类帐过程的安全。更改或删除分类帐上已有的记录即便不是完全不可能，也是极其困难的。鉴于这样的特点，区块链可以减少或消除欺诈和腐败等违规行为，同时也可降低交易成本。

例如，在中国，使用伪造出口发票来掩饰跨境资金转移一度很猖獗。2014 年 4 月至 9 月间发现了价值 100 亿美元的伪造交易。¹⁵ 在青岛港发生了重大欺诈案，当地企业用假收据来给一批金属货品的多笔贷款做抵押。¹⁶ 青岛事件涉及 30 万吨氧化铝、2 万吨铜和 8 万吨铝锭。¹⁶ 之后中国的银行开始收取更高的利率，并减少提供抵押贷款。¹⁷ 区块链可以防止这类丑闻。

区块链还可以生成“智能化”产权

并通过智能合约来控制它。¹⁸ 这类产权可包括实体产权（汽车、房屋等）和非实体产权（公司股份等）。¹⁹ 基于区块链的智能财产仅可根据智能合约中公布的信息发生变更。¹⁸ 如果产权被用作抵押，智能合约可能不会允许所有者将相同的财产拿到另一家银行作抵押或担保。因此，对于财产监管人来说，在贷款之前验证抵押品的过程被大大简化了。²⁰ 这生成了一个针对智能产权的可信赖的交易系统，使信贷变得更易获得也更便宜。¹⁹

弱势群体

区块链也可能帮助难民和流离失所者。目前向这些人群提供的援助存在低效、欺诈和资源严重错配的问题。例如费用和成本占到了援助转账的 3.5%。此外，由于第三方盗窃和管理不善等问题，估计有 30% 的发展援助金无法到达目标受益者手中。²¹

针对这类问题已经出现了各种基于区块链的解决方案。例如，区块链可以确保捐赠到达目标受益人。例如，捐助者可以用比特币为南非的学校购买电力。支持区块链技术的智能电表系统可以让捐助者直接向电表付款，过程中不涉及任何可能重新分配这笔钱的组织或机构。捐助者还可以跟踪学校的用电量并计算自己捐赠的电量。²² 该项目由南非比特币创业公司 Bankymoon 通过一个众筹平台发起。²³

联合国粮食计划署 (WFP) 已经在用区块链帮助难民。援助金被直接支付给商家而非难民。没有银行参与其

中，受益人直接从商家那里收取货品。

²⁴ 2017 年初，WFP 启动了名为“积木”(Building Blocks) 项目的第一阶段，向巴基斯坦信德省的贫困家庭提供食品和现金援助。一台连接互联网的智能手机会验证并记录该署向食品供应商的付款、确保受益人获得预期的帮助、商家获得报酬，而 WFP 可密切关注资金动向。

从 2017 年 5 月开始，WFP 开始向约旦的难民营发放食品券。方法是把逐个单一加密的优惠券分发到参与援助活动的本地超市。超市收银员配备虹膜扫描仪来确定受益人身份并做结算（联合国数据库验证有关难民的生物识别数据）。“积木”项目的分类帐记录了对一个专用版以太坊加密货币的交易情况。WFP 报告说，截至 2017 年 10 月，它已向约旦的 10,500 名叙利亚难民分发了 140 万美元的食品券。²⁵ 该署希望区块链将援助项目的管理成本从 3.5% 降至不到 1%，并加快对偏远或受灾地区的援助（那些没有 ATM 机或银行无法正常运作的地方）。区块链货币甚至可以取代稀缺的本地现金，使援助机构、居民和商家能快速地展开电子交易和转账。

总结

区块链将对发展中国家产生积极的影响：它可以帮助减少欺诈和腐败，增加合法财产权，为世界上最贫困的人群提供创业激励。它还可以令金融交易更快速地展开，并确保援助发放的过程减少盗窃和欺诈。■

参考文献

1. ICT Facts and Figures 2017, report, International Telecommunication Union, 2017; www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf.
2. Corruption Leading to Unequal Access, Use and Distribution of Land--UN Report, report, UN
3. N. Kshetri, "Fostering Startup Ecosystems in India," *Asian Research Policy*, vol. 7, no. 1, 2016, pp. 94–103.
4. T. Hanstand, "The Case for Land Reform in India," *Foreign Affairs*, blog, 2013; www.foreignaffairs.com/articles/india/2013-02-19/united?cid=soc-twitter-in-snapshots-united-022013.
5. USAID Country Profile: Honduras, report, USAID, 2016; [https://usaidlandtenure.net/wp-content/uploads/2016/09/USAID_Land_Tenure_Honduras_Profile_0.pdf](http://usaidlandtenure.net/wp-content/uploads/2016/09/USAID_Land_Tenure_Honduras_Profile_0.pdf).
6. T. Puiu, "How Bitcoin's Blockchain Could Mark an End to Corruption," ZME Science, blog, 2015; www.zmescience.com/research/technology/bitcoin-blockchain-corruption-04232.
7. J. Jeong, "Bitcoin, Blockchain, and Land," *The Global Anticorruption Blog*, blog, 2016; globalanticorruptionblog.com/2016/01/08/bitcoin-blockchain-and-land-reform-can-an-incorruptible-technology-cure-corruption.
8. L. Shin, "Republic of Georgia to Pilot Land Titling on Blockchain with Economist Hernando De Soto, BitFury," *Forbes*, blog, 2016; www.forbes.com/sites/laurashin/2016/04/21/public-of-georgia-to-pilot-land-titling-on-blockchain-with-economist-hernando-de-soto-bitfury.
9. O. Ogundehi, "Land Registry Based on Blockchain for Africa," *IT Web Africa*, blog, 2016; www.itwebafrica.com/enterprise-solutions/505-africa/236272-land-registry-based-on-blockchain-for-africa.
10. A. Jones, "How Blockchain Is Impacting Industry," *International Banker*, blog, 2016; [https://internationalbanker.com/finance/blockchain-impacting-industry](http://internationalbanker.com/finance/blockchain-impacting-industry).
11. "Bitland Partners with CCEDK to Improve Blockchain Land Registry in West Africa," *EconoTimes*, blog, 2016; www.econotimes.com/bitland-partners-with-CCEDK-to-improve-blockchain-land-registry-in-West-Africa-271517.
12. S. Higgins, "Republic of Georgia to Develop Blockchain Land Registry," *CoinDesk*, blog, 2016; www.coindesk.com/bitfury-working-with-georgian-government-on-blockchain-land-registry.
13. S. Higgins, "Survey: Blockchain Capital Markets Spending to Reach \$1 Billion in 2016," *CoinDesk*, blog, 2016; www.coindesk.com/capital-markets-1-billion-2016-blockchain.
14. "Indian State Plans to Store Citizen Data on a Blockchain," *CCN*, blog, 2017; www.ccn.com/indian-state-plans-blockchain-storage-citizen-data.
15. "Leveraging Blockchain for the Real Estate Industry," *Lawfuel*, blog, 2017; www.lawfuel.com/blog/leveraging-blockchain-real-estate-industry.

2017年，WFP开始向约旦难民营发放 加密过的食品优惠券。

News, 2011; [https://news.un.org/en/story/2011/12/397982-corruption-leading-unequal-access-use-and-distribution-land-un-report#.WEMpP33QCWI](http://news.un.org/en/story/2011/12/397982-corruption-leading-unequal-access-use-and-distribution-land-un-report#.WEMpP33QCWI).

3. N. Kshetri, "Fostering Startup Ecosystems in India," *Asian Research Policy*, vol. 7, no. 1, 2016, pp. 94–103.

4. T. Hanstand, "The Case for Land Reform in India," *Foreign Affairs*, blog, 2013; www.foreignaffairs.com/articles/india/2013-02-19/united?cid=soc-twitter-in-snapshots-united-022013.

5. USAID Country Profile: Honduras, report, USAID, 2016; [https://usaidlandtenure.net/wp-content/uploads/2016/09/USAID_Land_Tenure_Honduras_Profile_0.pdf](http://usaidlandtenure.net/wp-content/uploads/2016/09/USAID_Land_Tenure_Honduras_Profile_0.pdf).

Blog, blog, 2016; [https://globalanticorruptionblog.com/2016/01/08/bitcoin-blockchain-and-land-reform-can-an-incorruptible-technology-cure-corruption](http://globalanticorruptionblog.com/2016/01/08/bitcoin-blockchain-and-land-reform-can-an-incorruptible-technology-cure-corruption).

8. L. Shin, "Republic of Georgia to Pilot Land Titling on Blockchain with Economist Hernando De Soto, BitFury," *Forbes*, blog, 2016; www.forbes.com/sites/laurashin/2016/04/21/public-of-georgia-to-pilot-land-titling-on-blockchain-with-economist-hernando-de-soto-bitfury.

9. O. Ogundehi, "Land Registry Based on Blockchain for Africa," *IT Web Africa*, blog, 2016; www.itwebafrica.com/enterprise-solutions/505-africa/236272-land-registry-based-on-blockchain-for-africa.

12. S. Higgins, "Republic of Georgia to Develop Blockchain Land Registry," *CoinDesk*, blog, 2016; www.coindesk.com/bitfury-working-with-georgian-government-on-blockchain-land-registry.

13. S. Higgins, "Survey: Blockchain Capital Markets Spending to Reach \$1 Billion in 2016," *CoinDesk*, blog, 2016; www.coindesk.com/capital-markets-1-billion-2016-blockchain.

14. "Indian State Plans to Store Citizen Data on a Blockchain," *CCN*, blog, 2017; www.ccn.com/indian-state-plans-blockchain-storage-citizen-data.

15. "Leveraging Blockchain for the Real Estate Industry," *Lawfuel*, blog, 2017; www.lawfuel.com/blog/leveraging-blockchain-real-estate-industry.

16. S. Shengxia, "China Uncovers \$10b Worth of Falsified Trade," Global Times, blog, 2014; www.globaltimes.cn/content/883512.shtml.
17. P. Smyth, Blockchain Technology: 7 Ways Blockchain Technology Could Disrupt the Post-Trade Ecosystem, white paper, Kynetix, 2015; www.the-blockchain.com/docs/Seven%20ways%20the%20Blockchain%20can%20change%20to%20trade%20system.pdf.
18. K. Bheemaiah, "Block Chain 2.0: The Renaissance of Money," Wired, blog, 2015; www.wired.com/insights/2015/01/block-chain-2-0.
19. A. Mizrahi, A Blockchain-Based Property Ownership Recording System, ChromaWay, 2016; <https://chromaway.com/papers/A-blockchain-based-property-registry.pdf>.
20. M.A. Calandra Jr. et al., Blockchain Technology, Finance and Securitization, blog, Alston & Bird, 2016; www.alston.com/-/media/files/insights/publications/2016/06/finance-and-financial-services--products-advisory/files/view-advisory-as-pdf/fileattachment/161075-blockchain-technology2.pdf.
21. B. Paynter, "How Blockchain Could Transform the Way International Aid Is Distributed," Fast Company, blog, 2017; www.fastcompany.com/40457354/how-blockchain-could-transform-the-way-international-aid-is-distributed.
22. S. Higgins, "How Bitcoin Brought Electricity to a South African School," Coindesk, blog, 2016; www.coindesk.com/south-african-primary-school-blockchain.
23. G. Mulligan, "5 African Crowdfunding Startups to Watch," Disrupt Africa, blog, 2015; disrupt-africa.com/2015/11/5-african-crowdfunding-startups-to-watch.
24. N. Menezes, "UN Uses Ethereum to Distribute Funds to Jordanians," BTCManager.com, blog, 2017; <https://btcmanger.com/un-uses-ethereum-to-distribute-funds-to-jordanians>.
25. J.I. Wong, "The UN Is Using Ethereum's Technology to Fund Food for Thousands of Refugees," Quartz, blog, 2017; <https://qz.com/1118743/world-food-programmes-ethereum-based-blockchain-for-syrian-refugees-in-jordan>.

Nir Kshetri 是北卡大学格林斯堡分校的布莱恩商学院 (Bryan School of Business and Economics) 的管理学教授。可通过 nbkshetr@uncg.edu 与他联系。

Jeffrey Voas 是 IEEE 会士。可通过 j.voas@ieee.org 与他联系。



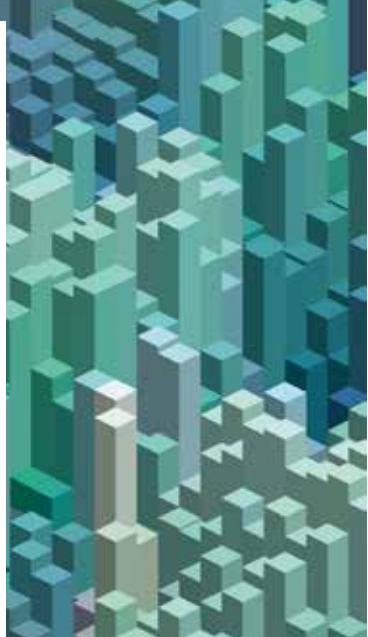
CMU 论文解读： 基于 GAN 和 VAE 的 跨模态图像生成

文 | 武广，合肥工业大学

来源 | PaperWeekly (微信公众号: paperweekly)

本文来自 PaperWeekly (公众号: PaperWeekly)。

PaperWeekly 是一个推荐、解读、讨论、报道人工智能前沿成果的学术平台。



跨 模态之间转换生成在模态间差异大的时候是相对困难的, 将一个模态编码得到的潜在变量作为条件, 在 GAN 的训练下映射到另一个模态的潜在变量是本文的核心。虽然这篇论文只是预印版, 但是文章的这个跨模态潜在变量相互映射的思想是很有启发性的。

论文引入

人类很容易学会将一个领域的知识转移到另一个领域, 人类可以灵活地

学习将他们已经在不同领域学到的知识连接在一起, 这样在一个领域内的条件下, 他们就可以回忆或激活他们从另一个领域学到的知识。

深度生成模型通过将它们映射到潜在空间来编码一个域内的隐式知识是被广泛使用的, 可以控制潜在变量通过条件在学习域内生成特定样本。然而, 与人类相比, 深层生成模型在从一个域到另一个域之间建立新连接方面不够灵活。换句话说, 一旦它学会了从一组域条件生成样本, 使其用于生成以另一组控制为条件的样本通常很难并且可能

需要重新训练模型。

跨域转换对于建立模态之间的联系是重要的, 能够让神经网络更加地智能化, 有一些方法提出来解决跨模态的问题。将条件编码映射到无条件训练的 VAE [1], 以允许它用用户定义的域有条件地生成样本, 并取得了很好的效果。

但是一个限制是那些条件是通过 one-hot 专门定义的。这样做需要特征工程, 并且想要对一些隐含的特征进行条件化时效果较差, 例如使用来自一个场景的图像作为条件来生成学习域中的相关图像。在 Unsupervised Cross-

domain Image Generation [2] 中训练端到端模型, 假设这两个域相关性, 再循环训练。

今天要说的这篇论文, 在更少的模态假设下实现了跨模态的相互生成。总结一下这篇论文的优势:

- > 利用 GAN 实现了跨模态的潜在变量之间的相互映射
- > 在较少的假设下实现跨模态的转换。

实现方法

我们一起来看一下文章实现的模型框架:

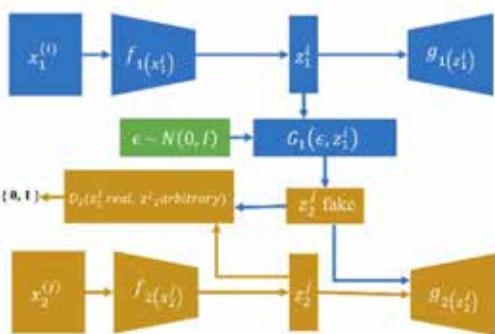


图 1. 上图为本篇论文的模型框架。

下标为 1 的代表着是模态 1 下的数据变换, 相对应的下表为 2 的代表着模态 2 下的数据变换, 上标代表模态下的类别, i 类和 j 类。f 代表着编码器网络, g 代表着解码器网络, G 代表了实现映射的生成器, D 代表了对应的判别器, 噪声 $\epsilon \sim N(0,1)$ 。

我们看到上下两路对应的是 VAE 实现框架, 也就是上路为模态 1 对应的 VAE1 是一个完整的流程, 下路为模态 2 对应的 VAE2 也是一个完整的流程, 当训练 VAE1 和 VAE2 到收敛状态时(训练完成), 此时的编码器已经可以很好的将模态数据编码到隐藏空间了。

我们分析由模态 1 到模态 2 的变换。模态 1 的第 i 类图像 $x_1^{(i)}$ 作为输入到模态 1 的编码器 $f_1(x_1^{(i)})$ 从而得到 i 类的潜在变量 z_1^i , 将 z_1^i 作为条件, ϵ 作为噪声输入到生成器 G1 下从而

实现映射到, 模态 2 第 j 类的图像经过编码得到的潜在空间下, 此时由生成器 G1 得到的映射记为 $z_2^j fake$ 。

判别器 D2 的目的是为了判别真实模态 2 下 j 类图像编码得到的 $z_2^j real$ 和生成的 $z_2^j fake$ 是否是匹配, 如果不匹配就认为是假, 如果匹配了那就说明生成器 G1 成功欺骗了判别器 D2 从而实现相互博弈共同提高。相对应的, 由模态 2 也可以向模态 1 映射, 这个过程和模态 1 到模态 2 是相似的。

分析了框架的实现方法我们再看一下实现的损失函数上的设计, 首先是 VAE 对应的损失优化, 这个大家估计也都熟悉了:

$$ELBO = \max_{\theta} \max_{\phi} E_{p_D(x)}[-KL(q_{\phi}(z|x), p(z)) + E_{q_{\phi}(z|x)} \log p_{\theta}(x|z)]$$

ELBO 下参数的定义和原始 VAE 下是相同的, 我就不重复描述了, 为了达到更好的 VAE 结果, 文章还加了一个像素上的重构误差, 所以最终 VAE 下的损失函数为:

$$\mathcal{L}(\theta, \phi, x^{(i)}) = \frac{1}{N} \sum_{i=1}^N [\lambda_1 C(x^{(i)}, g(f(x^{(i)}))) + \lambda_2 ELBO^{(i)}]$$

这里的 C 就是像素上的重构误差, λ_1, λ_2 为控制的参数, 这个损失将实现 VAE 的整体优化。

对于映射上的 GAN 的损失函数设计, 文章主要采用匹配的方法博弈, 文章匹配有三组。一组是真实匹配 z, z , 对应的损失记为 $\mathcal{L}_{c=1}(z, z) \equiv -\log(D(z, z))$; 一组是生成上的匹配 z, z' , 其中 z' 为生成的隐藏变量, 对应的损失记为 $\mathcal{L}_{c=0}(z, z') \equiv -(1 - \log(D(z, z')))$; 最后一组是噪声对应的匹配 z, z' , 对应的损失记为 $\mathcal{L}_{c=0}(z, \epsilon) \equiv -(1 - \log(D(z, \epsilon)))$ 。则最终的判别器的损失:

$$\mathcal{L}_D = \mathbb{E}_{z \sim p(z|x)} [\mathcal{L}_{c=1}(z, z)] + \mathbb{E}_{z \sim G(z, \epsilon)} [\mathcal{L}_{c=0}(z, z')] + \mathbb{E}_{\epsilon \sim p(\epsilon)} [\mathcal{L}_{c=0}(z, \epsilon)]$$

在生成器中文章引入了正则项 $\frac{1}{n} \|\epsilon - G(\epsilon, z)\|_2^2$, 当生成器将简单噪声移动到映射分布时, 通常会最大化所生成的输出与原始噪声 ϵ 之间的距离。随着这个术语的增加, 最小化损失函数以防 G 和 ϵ 太远。综合上述, 对于 G 的损失函数可以写为:

$$\mathcal{L}_G = \mathbb{E}_{z \sim p(z), \epsilon \sim p(\epsilon)} [\mathcal{L}_{c=1} G(z, \epsilon) + \frac{\lambda_{reg}}{n} \|\epsilon - G(\epsilon, z)\|_2^2]$$

经过对上述损失函数优化, 系统将达到平衡, 此时就可

以实现跨模态的相互生成。输入模态 1 下的 $x_1^{(i)}$ 经过编码得到 z_1^i , 经过 GAN 的映射得到 z_2^{fake} 送入到模态 2 的解码器下得到输出 $x_2^{(j)}$, 从而实现了模态间的转换。

实验

文中主要是在 MNIST 数据集和 SVHN 数据集下进行的实验, MNIST→MNIST 的实现下的模态转换主要是实现 $0 \rightarrow 5, 1 \rightarrow 6, 2 \rightarrow 7, 3 \rightarrow 8, 4 \rightarrow 9$, 其中 0, 1, 2, 3, 4 为模态 1 下的数据, 5, 6, 7, 8, 9 为模态 2 下的数据, 得到的生成的实验结果如下:



图 2. MNIST 到 MNIST 的生成结果。这一模型由全部的训练集训练而来, 并在随机选择的测试集上进行测试。

对于 SVHN 到 MNIST 的转换, 由于 SVHN 下没有 0 这个数字, 所以转换为 $1 \rightarrow 6, 2 \rightarrow 7, 3 \rightarrow 8, 4 \rightarrow 9$, 结果如下:



图 3. SVHN 到 MNIST 的生成结果。这一模型由全部的训练集训练而来, 并在随机选择的测试集上进行测试。

在定量分析上, 文中为生成的数据设计了分类器用来检测生成数据的分类效果, 侧面反应生成的质量, 同时也做了人体辨别实验:

文中还补充了 MNIST 到 Fashion-MNIST 转换的实验,

实验生成上得到了不错的效果:

表1. ACCURACY OF GENERATED IMAGES

Dataset	训练集的大小	准确度(cnn分类器)	accuracy(human)
MNIST→MNIST	500	0.63	0.58
MNIST→MNIST	1000	0.68	0.65
MNIST→MNIST	2000	0.66	0.67
MNIST→MNIST	Full set	0.77	0.75
SVHN→MNIST	500	0.43	0.48
SVHN→MNIST	1000	0.48	0.44
SVHN→MNIST	2000	0.46	0.43
SVHN→MNIST	Full set	0.61	0.58

由于只是预印版, 所以一些详细的对比实验没有加进去, 不过这种方法的启发意义还是蛮大的。

总结

文章利用 GAN 的思想在 VAE 的基础上, 实现了模态间的潜在空间的相互映射, 得到映射的空间可以进一步的解码生成对应于另一个模态的数据从而实现了跨模态的相互生成。这种利用潜在空间的变换实现跨模态生成在很多跨模态之间都可以参考, 虽然文章只是预印版但是很具有启发意义。■

参考文献

1. Jesse Engel, Matthew Hoffman, and Adam Roberts. Latent constraints: Learning to generate conditionally from unconditional generative models. arXiv preprint arXiv:1711.05772, 2017.
2. Yaniv Taigman, Adam Polyak, and Lior Wolf. Unsupervised cross-domain image generation. arXiv preprint arXiv:1611.02200, 2016.



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



KDD 18 论文解读： 斯坦福大学提出全新网络 嵌入方法 – GraphWave

文 | 薛寒生，澳大利亚国立大学

来源 | PaperWeekly (微信公众号: paperweekly)

本文来自 PaperWeekly (公众号: PaperWeekly)。
PaperWeekly 是一个推荐、解读、讨论、报道人工智能
前沿成果的学术平台。



论文动机

驻留在图的不同部分中的节点可能在其本地网络拓扑中具有类似的结构角色。然而学习节点的结构表示是一项具有挑战性的无监督学习任务，其通常涉及为每个节点人工指定和定制拓扑特征。

GraphWave 是一种可扩展的无监督方法，用于基于网络中的结构相似性

来学习节点嵌入。GraphWave 通过将小波视为概率分布，并使用经验特征函数表征分布来开发光谱图小波的新用途。

GraphWave 提供理论保证，具有相似本地网络邻域的节点将具有类似的 GraphWave 嵌入，即使这些节点可能驻留在网络的非常不同的部分中。

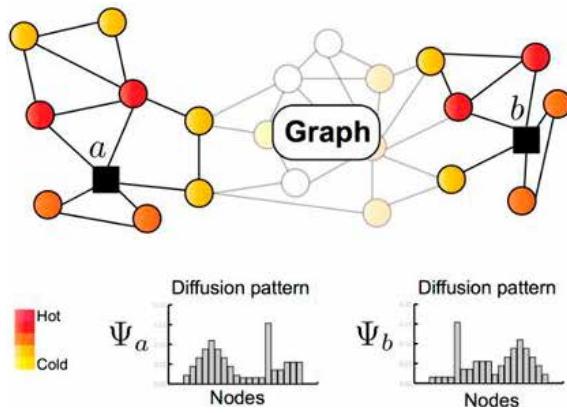
GraphWave 与边数成线性比例，不需要任何人工定制节点的拓扑特征。

GraphWave 模型介绍

GraphWave 基于以该节点为中心的谱图小波的扩散，学习每个节点的结构嵌入。直观地，每个节点在图上传播能量单位，并基于网络对该探测的响应来表征其相邻拓扑。

GraphWave 使用一种新颖的方法将小波视为图上的概率分布。通过这种

方式，结构信息包含在扩散如何通过网络传播而不是传播的位置。为了提供矢量值特征，然后可以将其用作任何机器学习算法的输入，GraphWave 使用经验特征函数嵌入这些小波分布。



在上图中，节点 a 和 b 具有相似的局部结构角色，即使它们在图中很远。虽然 a 和 b 的原始光谱图小波签名/系数 Ψ 可能非常不同，但 GraphWave 将它们视为概率分布，因此可以自动了解系数分布确实相似。Graphwave 利用这些新见解，基于以节点 a/b 为中心的谱图小波的扩散，学习节点 a/b 的结构嵌入。

GraphWave Algorithm

Algorithm 1 Learning structural embeddings in GRAPHWAVE.

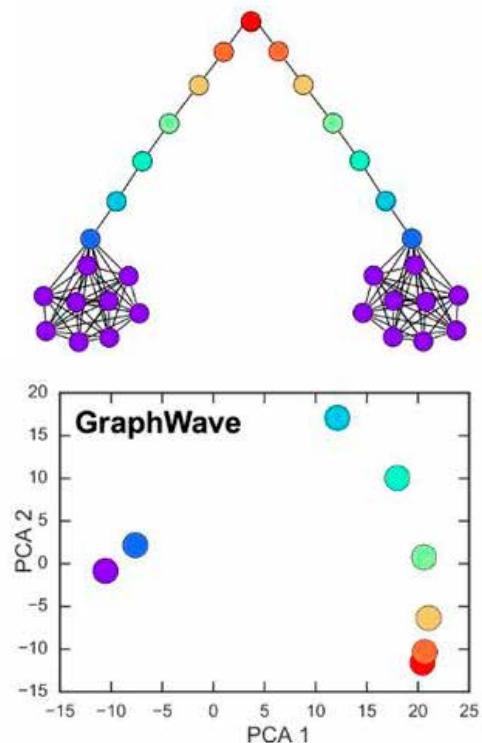
```

1: Input: Graph  $G = (\mathcal{V}, \mathcal{E})$ , scale  $s$ , evenly spaced sampling points  $\{t_1, t_2, \dots, t_d\}$ 
2: Output: Structural embedding  $\chi_a \in \mathbb{R}^{2d}$  for every node  $a \in \mathcal{V}$ 
3: Compute  $\Psi = U g_s(\Lambda) U^T$  (Eq. (1))
4: for  $t \in \{t_1, t_2, \dots, t_d\}$  do
5:   Compute  $\phi(t) = \text{column-wise mean}(e^{it\Psi}) \in \mathbb{R}^N$ 
6:   for  $a \in \mathcal{V}$  do
7:     Append  $\text{Re}(\phi_a(t))$  and  $\text{Im}(\phi_a(t))$  to  $\chi_a$ 

```

学习结构签名的 2D PCA。

从下图中可以看出，该图具有 8 个不同类别的结构等效节点，如颜色（左）所示。结构签名的 2D PCA 投影（右）包含与杠铃图中的节点相同数量的点。这是因为相同的签名具有相同的投影，导致重叠点。



GraphWave 正确地学习了结构等效节点的相同表示，为 GraphWave 的理论保证提供了经验证据。这可以通过图中的结构等效节点（相同颜色的节点）在 PCA 图中具有相同的投影来看出。

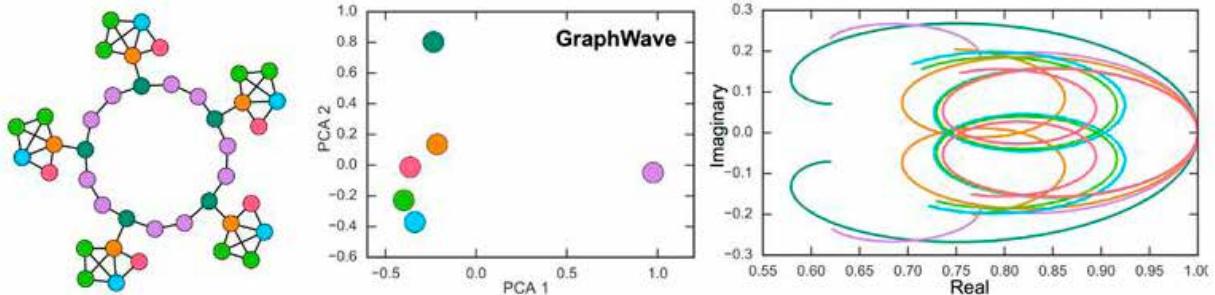
特别是，GraphWave 正确地将 clique 节点（紫色）组合在一起。GraphWave 还正确区分连接杠铃图中两个密集团的节点。它以类似梯度的模式表示那些捕获这些节点的结构角色谱的节点（右）。

A Cycle Graph with Attached House Shapes 在这个例子中，文中考虑一个图形，其中“房屋”形状沿循环图定期放置。和以前一样，我们使用 GraphWave 来学习图中节点的结构签名，然后使用有关结构角色的地面实况信息来评估 GraphWave 的性能。

实验结果

Barbell Graph

在这个例子中，文章考虑一个杠铃图，它由两个由长链连接的密集团组成。我们将 GraphWave 应用于杠铃图并绘制

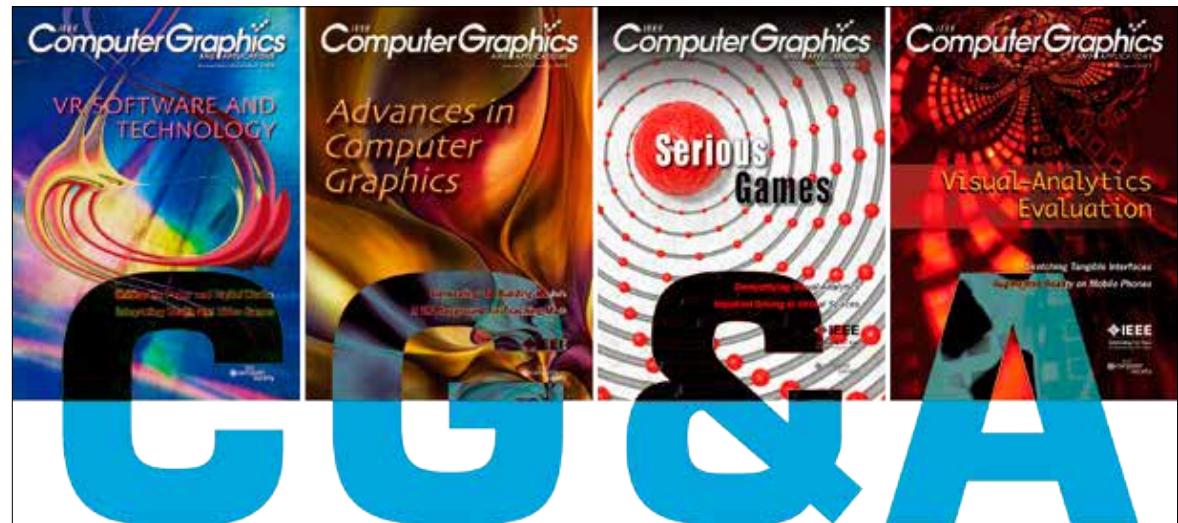


图形在上图(左)中可视化,同时还有 GraphWave 结构签名的 2D PCA 投影(中间)。我们观察到结构等效节点的表示重叠,GraphWave 完美地恢复了 6 种不同的节点类型。

可以看到小波系数分布的最终特征函数(右)。在该图中,不同形状的特征函数捕获不同的结构角色。我们注意到这些曲线所携带的蓝色,浅绿色和红色节点的作用之间的视觉接近度,以及它们与核心深绿色和紫色节点的明显差异。

总结

文中提出了一种全新的 NE 方法 GraphWave,该方法使用谱图小波为每个节点生成结构嵌入,我们通过将小波视为分布并评估结果特征函数来实现。为网络嵌入提供了全新的思路。■



《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

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** 亿万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



ByteDance

字节跳动

x

IEEE中国代表处

HOST: CAAI BYTEDANCE
CO-HOST: IEEE-CHINA



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。

APNet'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SIGCOM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。

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：

网络新技术的部署应用引发广泛

探讨。近年来一些新技术（如RDMA，GPU加速等）不断涌现在网络领域，受到了研究者的广泛关注。本次研讨会上，来自南京大学团队的论文“*Toward Effective and Fair RDMA Resource Sharing*”，针对RDMA在公有云场景中的扩展性部署问题，提出了Avatar模型，可以对不同的RDMA应用进行资源的有效管理，减少资源开销，实现高扩展性和公平性。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。

网络模型的构建和解释正在备受关注。随着网络场景的日益复杂，一些新的网络模型和方法论的研究收到越来越多的关注。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，强调了领域知识对于解释深度学习网络的重要意义。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 Macro*

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

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。

ow 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。

网络协议的优化改进仍然是领域热点。无论是应用层还是底层网络协议的改进和优化，目前仍然是网络领域持续关注的热点。本次研讨会上，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*

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。

此外，本次大会还录用了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。

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，网络与机器学习交叉领域备受关注。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



图 1. 华为的 Yong Ren 在做报告。

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。

同时，本届研讨会邀请了来自字节跳动、微软、腾讯、Intel、阿里巴巴和华为等国内外知名企业的专家进行了专题报告，研讨会首次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学术界和工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。

最后，本届研讨会邀请到了SIGCOM 2018录取论文的相关作者，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。研讨会同样采用座谈形式，在论文报告结束后，由5位SIGCOM论文作者集中分享了自己冲击SIGCOM的历程，并对如何作出SIGCOM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实验室主任郭传雄博士进行了简短交流。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，Fat-Tree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。同时Server-Centric的思想在网络虚拟化中得到广泛应用。因此,郭博士认为,Server-Centric的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领域的各种应用及基础研究,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。

展望

今年的APNet继续展现出多维度、跨领域、创新性的特点,汇聚来自世界各地的优秀学者和工业界,针对网络领域的热点问题展开广泛讨论。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。同时,网络场景的日益复杂也为今后的网络研究提出了更多挑战,希望广大网络研究人员,以本届APNet为契机,继续深入钻研,创造更具学术影响力值和工业实用价值的成果。

APNet 2019将于明年8月继续在北京举办,时间与SIGCOMM 2019临近。届时,一个全新的APNet将为研究者提供更大、更高、更具影响力的展示平台。期待更多有志于网络研究的中国学者的加入! □



图 2. 会议期间,笔者也和会议的组织者之一郭传雄博士进行了简单的交流。

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。



邦内干戈： 剑桥分析的反乌托邦

文 | 哈尔·贝里赫尔 (Hal Berghel)，拉斯维加斯内华达大学
译 | 刘莹，华中科技大学

剑桥分析等党派顾问使用数据分析来控制选民。它们依赖社交网络用户参与自我心理操控。

本

文发表时，剑桥分析

(Cambri-dge Analyti-
ca) 丑闻席卷了媒体头版

头条。这家公司被指干预了2016年的美国总统大选，加之Facebook不负责任地分享用户数据，引发了诸多焦虑愤懑。然而，在媒体难以平复的激动情绪中，存在一个会让我们感到不适的事实：虽然技术的发展已使得党派能以更复杂的方式操纵选民，单单这一点却并非问题的症结。要找到问题真正的根源，我们必须深入审视自己。

新技术，老策略

剑桥分析的故事始于2014年。当时数据科学家亚历山大·科根 (Aleksandr Kogan, 后改名亚历山大·幽灵 [Aleksandr Spectre]) 和其他几个与剑桥大学有不同关系的人创办了一家名叫“全球科学研究所” (Global Science Research) 的公司，专门推销 Facebook 的应用程序“这是你的数字生活” (thisisyourdigitallife)。这款应用从约五万名自认为只是在做某种性

格测试的用户那里收集个人信息，并利用这些信息来推断出其他在政治上有用的情报。¹这些情报之后被转到了秘密政治咨询机构剑桥分析那里，该机构自2012年起与美国共和党合作，并自称在特朗普胜选中起到了举足轻重的作用。²此事被爆料人克里斯托弗·怀利 (Christopher Wylie) 公之于众。

Facebook 在2015年至2018年初对这些信息被利用究竟知情多少仍存在争议，而2016年美国大选究竟受到多少干扰这个问题也一样。有关多方

勾结合谋的指控不断传出，其中涉及剑桥分析、加拿大数据分析公司聚合智商（Aggregate IQ，与几个英国脱欧全民公投运动团体有关联）、俄罗斯特工和“巨魔农场”、特朗普竞选团队以及维基泄密（Wikileaks）等。但在没完没了、令人晕头转向的争论中，各方一次又一次地否认了指控。特别检察官罗伯特·穆勒（Robert Mueller）正在调查部分指控。而如果特朗普政府无法撤销穆勒的调查，那么可以预期指控将进入法庭程序。

要了解这件事的历史背景，我们需要做一番现实考察。用不诚实手段改变公众的选举意向并不是什么新鲜事。在美国，自十九世纪初埃尔布里奇·盖里（Elbridge Gerry）用重新划分选区来操纵选举（“杰利蝾螈”[gerrymandering]这个负面政治术语的来源）至今，这些手段在选举中反复出现。³ 约翰·肯尼迪曾在国会竞选中使出“混淆对手”这一招（为稀释候选人约瑟夫·鲁索斯 [Joseph Russo] 的选票，他游说一个也叫约瑟夫·鲁索斯的清洁工参加了选举）；一封造假的匿名信“加纳克人信”诋毁了埃德蒙·马斯基（Edmund Muskie）；老布什及其竞选经理李·阿特沃特（Lee Atwater）用“威利·霍顿广告”攻击迈克尔·杜卡基斯（Michael Dukakis）；2002年国会参议院竞选中，共和党人用干扰电话阻止新罕布什尔州的民主党选民投票（该州共和党行政主任查尔斯·麦吉 [Charles McGee] 因此被定罪入狱）；小布什及其竞选经理卡尔·罗夫（Carl

Rove）散播了中伤对手的流言蜚语（比如暗示麦凯恩与一名黑人女子生有私生女）；戈尔的支持者对共和党全国委员会网站发起黑客攻击由默茨（Mercer）家族资助的超级政治行动团体“让美国成为第一”发起了“打败狡诈的希拉里”运动（Defeat Crooked Hillary）；尼克松采用“肮脏伎俩”对付几乎所有他不喜欢或不信任的政客。[4-7] 我们只需回顾一下那些候选人赢得了普选却因输给选举人团票的总统大选（1824年、1876

票；
➤ 会强化党派偏见的封闭式初选；
➤ 以各种各样的理由为选民投票制造障碍；
➤ 通过重新划分选区和不分区选举来稀释投票；
➤ 对最具“可塑性”的选民的微定向。

据我所知，上述政治操纵的历史案例与剑桥分析使用战略之间唯一的

在社交媒体平台上进行微定向可以让政治运动利用最强烈的情绪以及我们那些最易被操纵的软弱和恐惧。

年、1888年、2000年、2016年），就可以确信美国总统选举中的党派违规操作并不少见。

一般来说，政党可以通过多种手段来操纵选民：

- 杰利蝾螈；
- 通过控制投票机在选区的分布及投票时间来控制投票等待时间；
- 使用过滤名单（如被控犯有重罪者及非公民）或要求使用选民身份证，往往会影响有色人种、老年人和体弱者参与投票；
- 减少邮寄选票或提前投票的机会，或要求提前注册，往往会妨碍失业者、小时工或移民工人参与投票。

重要区别，是后者使用了数字技术来有效地实现选民微定向，比如将社交媒体用作武器、网络恶意挑衅和僵尸网络攻击等。

剑桥分析使用的战略可以追溯到爱德华·伯内斯（Edward Bernays）的政治宣传理论，该理论基于自一战以来被成功运用的各种方法。即便是现代的在线广告也不过是伯内斯理论的庸常延伸。然而，最近在线微定向技术的改进已经让这种宣传手段达到了新高度，也使剑桥分析成为家喻户晓的名字。该公司从Facebook和其他来源获得数据，而后对这些数据展开一丝不苟的细致分析，并得以在广大选民的层面识别和锁定各个政治热点涉及的敏感人群：仇

外的选民可能会受到与同性婚姻和小规模文化冲突有关的攻击广告的冲击；失业的选民会被美国岗位流失等主题煽动，等等。在社交媒体平台上进行微定向可以让政治运动利用最强烈的情绪以及我们那些最易被操纵的软弱和恐惧。正如最近英国第四频道的一则调查报道所显示，剑桥分析的高管公开声称他们通过操纵三个州的仅仅四万名选民，就令特朗普在选举人团投票中获胜，尽管他在普选中落后了 2,868,636 票（2.09%）(<https://transition.fec.gov/pubrec/fe2016/feedralelections2016.pdf>)。

(Cathy O'Neil) 于 2016 出版的畅销书《数学杀伤武器》(Weapons of Math Destruction) 中得到了强有力的展现。¹⁰ 作为一名数据分析师，奥尼尔担心数据经济会受人类偏见、偏误，以及隐藏在公众视野之外的目的驱使。使用大数据分析客观地从数据中提取模式是一回事，但她也看到了通过操纵数据来强化模式的证据。用她的话说，“在数学杀伤武器中，许多有危害的假设都经过数学的伪装，且大部分未经检验和质疑。”她展示了对数据分析的错误使用和过度信任如何扭曲了一些辖区里的教师评

无需被验证。第三，自动化极大地提升了错误信息的传播效率。两位作者总结说，网络宣传这一新武器的社会效应“才刚刚开始被了解”。两人的计算式宣传研究项目(comprop.ox.ac.uk)探讨了算法、自动化和政治之间的相互作用背后的科学。

在上述两项研究中，数据分析（或其歪曲和滥用）都是问题的核心。无论是像奥尼尔的例举中，通过使用有缺陷的模型来淡化风险或夸大优势，还是像博尔索弗和霍华德观察到的，计算式宣传的数据分析利用社交媒体发布信息的匿名性和不可验证性来操纵选举，偏见、偏误和隐蔽目的被故意渗入了数据分析中，以达到操纵的效果。假如我们对“一人一票”等民主原则深信不疑，那么数据分析可以说正在把我们的政治前景往更糟的方向推进。

假如我们对“一人一票”等民主原则深信不疑，那么数据分析可以说正在把我们的政治前景推向更糟的方向。

pdf）。这样的成就会令伯内斯都嫉妒不已。如果不是这些高管们的公开炫耀被摄录了下来，剑桥分析后来做出的无罪声辩会更令人信服些。⁹

但是，由此得出结论说，单凭剑桥分析的战略就为特朗普赢得了胜利，是错误的。因为操纵选举的力量众多。剑桥分析当然起到了一定的作用，但是否真如媒体所述，是一个主要因素，仍有待证明。

数学杀伤武器

微定向背后的科学在凯西·奥尼尔

估系统、监狱中的累犯建模、过分依赖似是而非的金融工具而导致了 2008 年的经济衰退、影响了数据驱动的大学排名，等等。她称这种过度依赖和滥用数据分析为“大数据的阴暗面”。

与我们目前的讨论更具相关性的 是吉利安·博尔索弗 (Gillian Bolsover) 和菲利普·霍华德 (Philip Howard) 最近有关“计算式宣传”的社论。¹¹ 他们观察到，互联网和社交媒体以三种方式“深刻地改变了政治宣传的风景”。首先，地理障碍的消失使得每一个拥有互联网连接的人都可以自行开展国际宣传。其次，匿名掩盖了消息来源，也使得内容

劝导 VS. 操纵：细细的红线

操纵“可说服的选民”来影响选举（以及其他重要的人类选择）是永恒存在的。这只是另一种形式的滥用——对我们信息空间的滥用。身体虐待、精神虐待、言语虐待、数字滥用（电子邮件、垃圾邮件、网络钓鱼攻击等）都具有相似的马基雅弗利主义根源：想要把自己的意志或信仰强加于人的愿望。在计算式宣传中，这种做法即是使用数字媒体和其他在线资源来促进宣传，过程既高效又不为人所注意（也就是说，人们不会识别出这是在做政治宣传）。通过“深

挖”、贿赂或陷害而从各种秘密来源实现这类宣传并不少见，但更多时候，它们是基于虚假或误导性的信息、谎言或纯粹的鬼话连篇。^{12,13}

在此我们值得做一番短暂的回忆。不到一个世纪前，广告主要强调产品的质量。当然了，广告常常都是虚假或误导的（比如说“这种牙膏比其他主要品牌的牙膏清洁力高 94.7%”，“抽这个牌子的烟比抽其他牌子喉咙更舒服些”，等等），但至少它们还是关于产品本身的。今天，重点已经放在了生活方式适配和自我形象上——或是直接表达（“专业运动员都使用 X 牌肥皂”），或者含蓄展现（让你在广告中看到 10 名豪华车主穿着 Y 牌毛衣）。问题在于，许多消费者对这种转变浑然不觉，以至于产品质量在广告中已经不再重要。例如，亚马逊实则通过它的评论过滤系统强化了这一点：如果有一则评论称 X 商品一直都比 Y 更好，就可能因“存在偏见”而被删除。

更多以地位象征而非品质好坏来评价一个品牌，后果之一是将市场推向了质量更差的产品。我在其他文章中已经把这一现象称为“去背景化”。背景是劣质商品的敌人：如果你不想改进你的产品，那么就确保你潜在的消费者没法把它置于一个有意义的比较视角。这之所以能成一种有效的策略，是因为我们对教育机构错误的信任和信心，以及这些机构的疏忽。毫无疑问，相比学习政客们为何选择某种“州鸟”或“州鱼”，严肃的小学课程应该多讲讲消费主义。

然而这个重要的议题在课程中基本缺

席——对商家而言这无疑是件大好事。

用户都在想什么？

一个更有趣的诈骗方案是“419 骗局”，要求信函和电子邮件的收件人向匿名发件人（通常是外国人）支付一小笔预付款，好在之后换取一大笔钱（数字 419 指的是尼日利亚刑法中欺诈这个类别，该国是许多这类诈骗的源头）。这种欺诈手法可以追溯到 19 世纪，当时被称作“西班牙囚徒”（Spanish Prisoner）。我很难理解那会儿它是怎

想说什么。既然长久以来媒体对网络隐私那么担忧，为何会有五万 Facebook 用户没有先从一些可信来源证实“这是你的数字生活”这款应用是安全的，就被吸引到了“性格测验”中？无论科根的应用声称这项测验会带来什么益处，常识应该令用户们确信，分享自己的数据风险过大，没有足够的理由参与其中。

几十年来，大量书籍和文章都描述了政府和企业利益对网上个人隐私的持续侵害。即使是对在线“免费”服务供应商的商业模式做一番最随意的分析，也会知道它们是通过“用户数据变现”

每个人都应该明白，广告商和营销者——而不是用户——才是免费服务的客户。

么能成功的，现在我就更想不通了。我们可以想象一些人会被“皮尔当人骗局”（Piltdown Man）骗到，这还合理，但“419 骗局”的可信度就和意大利面秘密花园、独角兽以及外星人麦田怪圈差不多。这证明了常识的脆弱。常识必须被不断运用，考验和强化才能发挥作用。数十年来，419 电邮在网络世界中到处可见，受害者少则千百万，多则数十亿。即使到了今天，按说人们应该已经对垃圾邮件这种事了然于心，却仍有足够多的潜在受害者存在，使得肇事者可以继续实施这种骗局。从中可以看出的问题更多在于受害者，而非骗徒。

我想你已经可以看出来接下来我

来获利的否则广告客户为何愿意付钱？现在，每个人都应该明白，广告商和营销商——而不是用户——才是这些免费服务的客户。这在 ProtonMail (<https://protonmail.com/use>) 这类免费服务中已经表现得非常明显。¹⁴ 人们应该知道他们必须放弃一些隐私才能使用免费在线服务，因此他们有义务事先确定自己愿意透露多少。剑桥分析的丑闻爆出后，媒体大量报道了 Facebook、谷歌和 Twitter 等公司追踪用户数据并将它们出售给了第三方^{15,16}。对此我们真的要这么吃惊么？即使最近披露的企业监视用户的程度之深让你吓了一跳，以往媒体对于政府、企业和金融机构黑

客的报道也应足以让你警醒，在网络上更小心谨慎地行事。

自愿无知者

之所以有那么多人愿意参加科根的性格测试，原因简单到令人不安：我们的教育系统让我们失望了！又一次！网上隐私风险本应是小学的标准教育内容，仅仅因为许多孩子在这个年龄段已经在上网了就该如此。

应该被纳入每个核心课程的一条信息是：不要泄露私人信息，除非可信

举产生的影响只是马基雅弗利式政治宣传的又一例证。假如没有一支极强大的军队支持一个强悍的独裁者，那么对有威权倾向的政治人物而言，这类宣传（通过杰利蝾螈和压制投票等手段被增强）就是无需政变也能夺取或维持权力的最有效方式。这种操纵并不新鲜。感兴趣的读者可以在 20 世纪 30 年代的记者费迪南德·伦德伯格 (Ferdinand Lundberg)¹⁷ 和 50 年代的社会学家赖特·米尔斯 (C. Wright Mills) 的著作中

找到类似的有关媒介操纵的论述。¹⁸ 时间和人物会变，但操控大众的恶意是

自利差？理解这种基本概念应该是升上四年级的先决条件——它不需要用到中学代数！同样愚蠢的是，爱荷华州共和党参议员查尔斯·格拉斯利 (Charles Grassley) 声称科技行业有责任保护其用户：“现状不再有效”。²⁰ 对于 Facebook 的股东和广告合作伙伴这些平台背后的驱动力，现状恰恰很有效啊，议员！Facebook 的 20 亿用户并非任何真正意义上的“社区”，他们作为一个整体，是被出售的商品。

许多政策制定者搞不清楚这些简单的概念，这有力地揭示了代议民主制的陷阱。反联邦主义者非常清楚地指出了这种陷阱。他们指责麦迪逊和汉密尔顿“天真”，竟会以为当选代表拥有如此高的道德品行和智慧，以至于能够摒弃偏狭的利益。这些反联邦主义者正确指出，人性决定了这些代表会倾向于重点提升自己的特权和地位而非公共利益，因为他们是从操控性的精英群体而来，受到他们所在的阶层带有的部族主义、仇外心理和偏见的驱使。麦迪逊构想了一个由公民美德驱动的国会，反联邦主义者则预想了一个受自身利益驱动的职业政治家为自身利益服务的国会。我留给读者自行判断哪种展望最接近我们目前的现实。

Facebook 的 20 亿用户并非任何真正意义上的“社区”。他们作为一个整体，是被出售的商品。

赖的人（父母、教育工作者、科学家）已经确认这样做符合你的最佳利益，并且它不会被用来伤害你。显然，Facebook 的受害者从没有得到过这个信息，而这是一个严重的文化问题。另一个缺失的重要课程是“自愿搁置怀疑”这一概念——它几乎是所有隐私泄露的核心。在线钓鱼和 419 等欺诈方案结合了感知管理、社会工程和技术手段（其中最简单也最不重要的因素），以实现“自愿搁置怀疑”。成年人未能识别假新闻和“这是你的数字生活”这类应用采用了与电影电视相同类型的操纵，这个事实既不可理喻，也令人震惊。

剑桥分析对美国、英国及加纳的选

持续存在的。关于这个问题的著述已经太多，所以那些对剑桥分析丑闻大表惊讶的人应该是选择了主动忽视。我必须指出，像剑桥分析这样的“数据末日”事件已被预测了十多年。

我们来看看我说的文化无知程度如何。去年 4 月，犹他州共和党参议员奥林·哈奇 (Orrin Hatch) 在国会听证会上问扎克伯格，免费的网络服务是否“坦承告知外界它们如何从用户那里榨取价值，还是隐瞒了真相？”它们当然隐瞒了真相！宣告自己从侵犯隐私中牟利对这些公司有何好处？这简直是蠢问题的典型——类似于问一个银行总裁，银行是否会直接向客户说明银行的利润来

奥威尔还是赫胥黎

最后，我们该从剑桥分析丑闻中获得什么？问题是自己制造的，任何解决方案也将存在于我们自身。从我们对自己的心理操纵而言，我们是这

桩丑闻的同谋——就和《动物农场》和《1984》中的居民一样。在面对社交网络时,我们都应该从这样一个前提出发来做决策:社交媒体和免费在线服务已经且将继续被与我们利益相违背的方式武装——我们就是产品!我想到了两个操作原则:“买家自慎”(*caveat emptor*),以及问一问“谁会获利?”(*cui bono*)。前者是涉案公司默认的理念,后者则应该是所有潜在顾客的指导原则。

至于剑桥分析的未来,我估计它会在右翼政治势力的帮助下扛过风暴,而这场争议也不会得出什么结果。然而,即便它倒闭了,数据分析这个领域的前景也不会有什么大变化,因为这家公司只是这个市场众多参与者中的一员。科赫(Koch)兄弟资助的i360(i-360.com)和卡尔·罗夫(Karl Rove)创办的DataTrust(thedatatrust.com)从事的业务与它差不多。这两家公司都旨在创建亲共和党的信息群(基于或不基于事实)并分享用户数据。²¹⁻²³尽管民主党人也使用复杂的数据处理技术,例如Narwhal数据库,^{24,25}但这些工具主要是为了解释性的用途。有一件事是肯定的:自1992年以来共和党每次普选都输给民主党(唯2004年例外),他们将继续使用每一种可用的武器——因为数字不站在他们那一边。

我想用富兰克林·罗斯福的话来给本文作结。1935年10月2日,他在加州太平洋国际博览会上发表讲话时警告说:“邦

内干戈’会不时地出现在你面前——他们会提出虚假的议题,歪曲事实,宣扬仇恨,竭力削弱那些为保障人权或精神理想的公众行动的重要性。今天,有些人会播下这些种子,而你们对他们的最好回应,就是拥有关于我们当前现状的简单而显见的事实。”今天,我们都应该听取这段话。■

参考文献

1. L. Ashworth and T. Gillespie, “Who Is Dr Aleksandr Kogan, the Cambridge Academic Accused of Misusing Face-book data?,” Varsity, 17 Mar. 2018; www.varisty.co.uk/news/15192.
2. T.B. Lee, “Facebook’s Cambridge Analytica Scandal, Explained,” Ars Technica, 20 Mar. 2018; <https://arstechnica.com/tech-policy/2018/03/facebook-cambridge-analytica-scandal-explained>.
3. H. Berghel, “Chasing Elbridge’s Ghost: The Digital Gerrymander,” Computer, vol. 49, no. 11, 2016, pp. 91–95.
4. C. Bernstein and B. Woodward, “FBI Finds Nixon Aides Sabotaged Democrats,” The Washington Post, 10 Oct. 1972; www.washingtonpost.com/wp-srv/national/longterm/watergate/articles/101072-1.htm.
5. M. Koncewicz, “How Republican Dirty Tricks Paved the Way for Russian Meddling in 2016,” The Washington Post, 9 Mar. 2018; www.washingtonpost.com/news/made-by-history/wp/2018/03/09/how-republican-dirty-tricks-paved-the-way-for-russian-muddling-in-2016/?utm_term=.ecfaed704065.
6. A. Love and W. Bergstrom, “16 Worst Political Dirty Tricks,” Politico, 6 June 2012; www.politico.com/gallery/16-worst-political-dirty-tricks?slide=11.
7. A.T. Smith, “10 Worst Dirty Tricks in American Politics,” Listverse, 18 Jan. 2014; <https://listverse.com/2014/01/18/10-worst-dirty-tricks-in-american-politics>.
8. E. Bernays, Propaganda, lg Publishing, 2004.
9. “Cambridge Analytica Uncovered,” Channel 4 undercover report, 19 Mar. 2018; www.channel4.com/news/cambridge-analytica-revealed-trumps-election-consultants-filmed-saying-they-use-bribes-and-sex-workers-to-trap-politicians-investigation.
10. C. O’Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Crown Publishing, 2016.
11. G. Bolsover and P. Howard, “Computational Propaganda and

- Political Big Data: Moving toward a More Critical Research Agenda,” Big Data, vol. 5, no. 4, 2017, pp. 273–276.
12. H. Berghel, “Lies, Damn Lies, and Fake News,” Computer, vol. 50, no. 2, 2017, pp. 80–85.
13. H. Berghel, Hal, “Alt-News and Post-Truths in the ‘Fake News’ Era,” Computer, vol. 50, no. 4, 2017, pp. 110–116.
14. A. Yen, “Why Privacy Is under Attack,” blog, 18 Nov. 2014; <https://protonmail.com/blog/privacy-under-attack>.
15. D. Curran, “Are You Ready? Here Is All the Data Facebook and Google Have on You,” The Guardian, 30 Mar. 2018; www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy.
16. H. Day, “It’s Not Just Facebook—Here’s What Data Your Other Favorite Apps Are Tracking,” ShortList, 26 Mar. 2018; www.shortlist.com/tech/facebook-data-apps-tracking-twitter-instagram-snapchat-gmail/351491.
17. F. Lundberg, America’s 60 Families, 5th ed., Lundberg Press, 2007.
18. C.W. Mills, The Power Elite, 2nd ed., Oxford Univ. Press, 2000.
19. A. Rogers, “The Cambridge Analytica Data Apocalypse Was Predicted in 2007,” Wired, 25 Mar. 2018; www.wired.com/story/the-cambridge-analytica-data-apocalypse-was-predicted-in-2007.
20. C. Kang, “Mark Zuckerberg Testimony: Senators Question Facebook’s Commitment to Privacy,” The New York Times, 10 Apr. 2018; www.nytimes.com/2018/04/10/us/politics/mark-zuckerberg-testimony.html.
- M. Allen and K.P. Vogel, “Inside the Koch Data Mine: Meet the Guys Building the Right’s New Machine,” Politico, 8 Dec. 2014; www.politico.com/story/2014/12/koch-brothers-rnc-113359.
22. J. Ward, “The Koch Brothers and the Republican Party Go to War—with Each Other,” Yahoo Politics, 11 June 2015; www.yahoo.com/news/the-koch-brothers-and-the-republican-party-go-to-121193159491.html.
23. M. Gold, “Koch Network Strikes New Deal to Share Voter Data with RNC-Aligned Firm,” The Washington Post, 29 July 2015; www.washingtonpost.com/news/post-politics/wp/2015/07/29/koch-network-strikes-new-deal-to-share-voter-data-with-rnc-aligned-firm/?utm_term=.e089648e8f62.
- A.C. Madrigal, “When the Nerds Go Marching In,” The Atlantic, 16 Nov. 2012; www.theatlantic.com/technology/archive/2012/11/when-the-nerds-go-marching-in/265325. S. Goldmacher, “Hillary Clinton’s ‘Invisible Guiding Hand,’” Politico, 7 Sept. 2016; www.politico.com/magazine/story/2016/09/hillary-clinton-data-campaign-elan-kriegel-214215.

微信名：计算人
微信号：jisuanren





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！

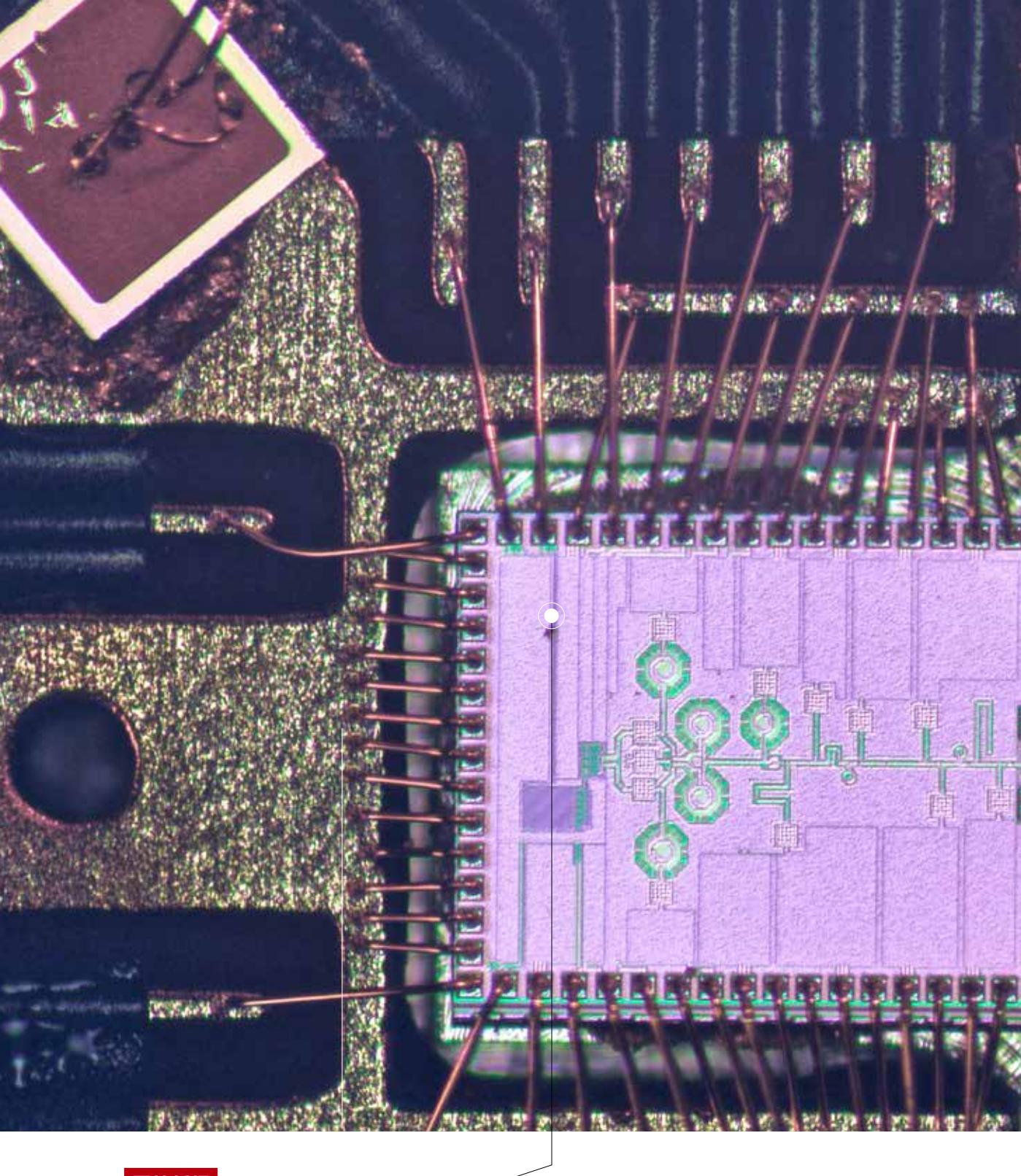


点击www.computer.org/jobs，
从全世界的雇主那里搜索技术工作岗位和实习机会。

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

IEEE  computer society | JOBS

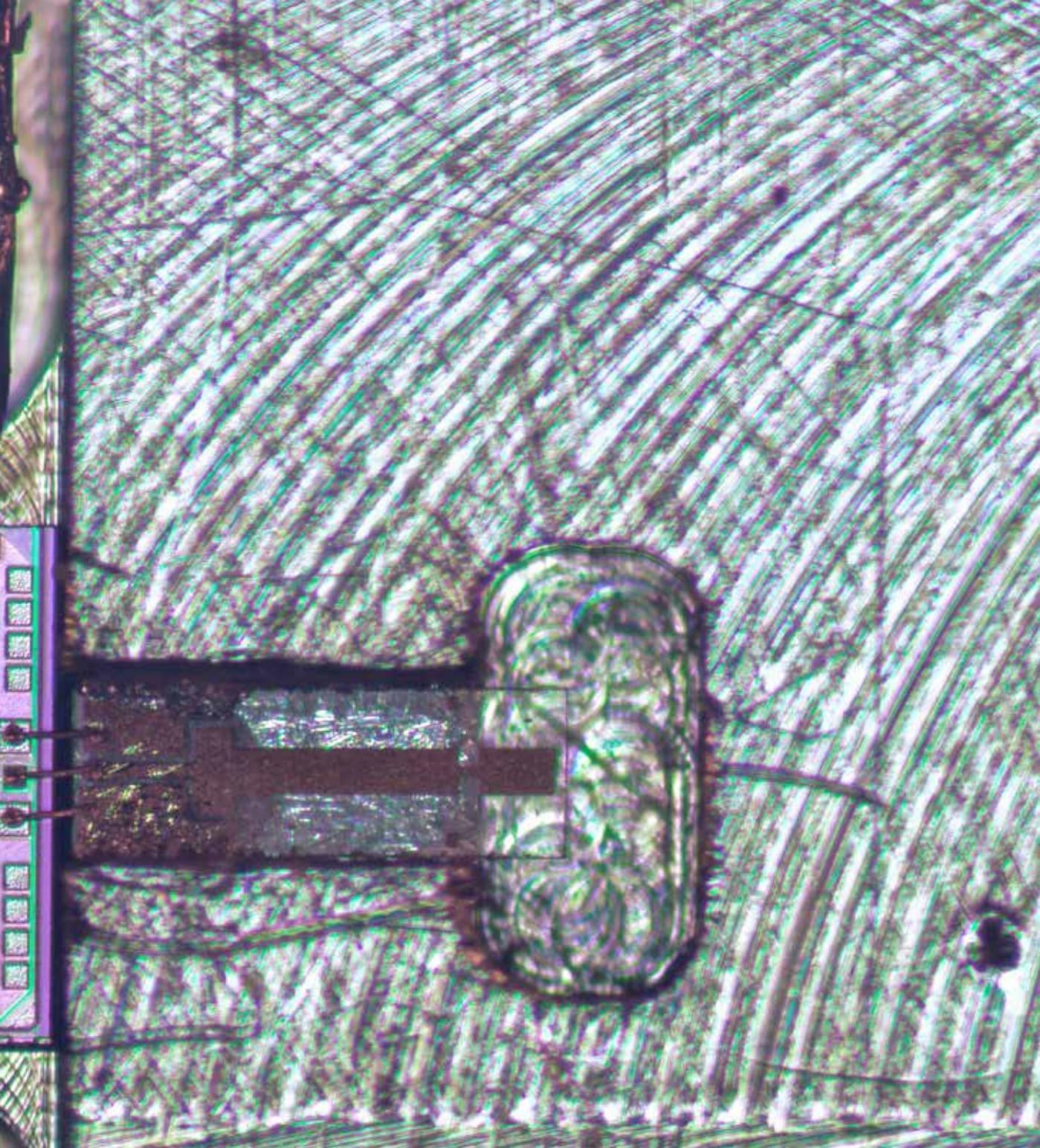
IEEE计算机协会是AIP Career Network的合作伙伴。其他合作伙伴包括《今日物理》杂志(Physics Today)，美国医学物理协会(American Association of Physicists in Medicine)，美国物理教师协会(American Association of Physics Teachers)，美国物理学会(American Physical Society)，AVS科学和技术学会(AVS Science and Technology)，物理学生协会(Society of Physics Students)和Sigma Pi Sigma。



图片新闻

芯片上的分子钟

麻省理工学院的研究人员发明了一个新的芯片，可以利用测量分子稳定地旋转，作为衡量时间的依据。



图中，粉红色的时钟芯片连接到电路板组件上。在实验中，分子钟的平均误差低于每小时 1 微秒，与微型原子钟相当，其稳定程度是采用晶体振荡器的智能手机内置时钟的 1 万倍。研究论文发表在英国《自然·电子学》杂志上。

图片来源：麻省理工学院



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