

中国计算机学会通讯



COMMUNICATIONS OF THE CCF

第15卷 第3期 总第157期 2019年3月



数据可视化的新疆界 P8

计算机问题求解的三类方法 P42

电脑前传(3): 逻辑 P47



站在学科前沿 打开技术之门

高水平 (资深专家讲授)

立前沿 (最新和热点技术)

大剂量 (三整天)



ADL 是 CCF 举办的学科前沿讲习班，目的是使青年学者短期内深入了解计算领域某个学科前沿发展动态，开拓眼界，为以后的科学的研究打下基础。2009 年开始举办，每年 10 期。

ADL 学科前沿讲习班

The CCF Advanced Disciplines Lectures

联系: adl@ccf.org.cn 188 1066 9757



中国计算机学会通讯 COMMUNICATIONS OF THE CCF



主办 中国计算机学会
China Computer Federation

刊名题字 张效祥

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

声 明

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

版权声明

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

编辑单位：中国计算机学会
印刷单位：北京华联印刷有限公司
发送对象：中国计算机学会会员
印刷日期：2019 年 3 月

主 编

李国杰 CCF 名誉理事长，CCF 会士，中国工程院院士

执行主编

钱德沛 CCF 会士，北京航空航天大学教授，中山大学计算机学院院长

专 题

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

专 栏

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

动 态

主 编 唐 杰 CCF 杰出会员，清华大学教授
编 委 鲍 捷 CCF 专业会员，北京文因互联科技有限公司 CEO
黄萱菁 CCF 高级会员，复旦大学教授
蒋洪波 CCF 杰出会员，湖南大学教授
刘知远 CCF 高级会员，清华大学副教授
宋国杰 CCF 高级会员，北京大学副教授
俞 扬 CCF 专业会员，南京大学副教授

译 文

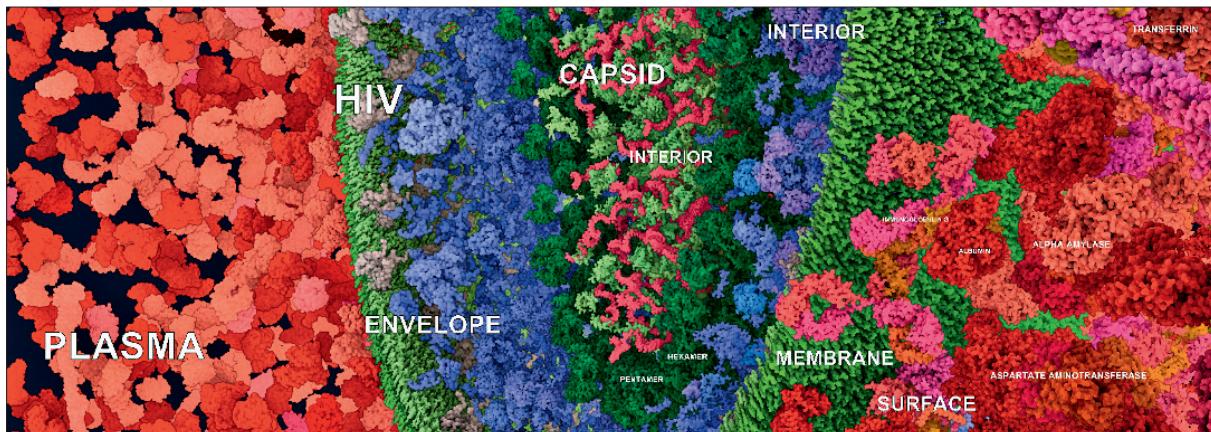
主 编 卜佳俊 CCF 常务理事，浙江大学教授
编 委 胡春明 CCF 理事，北京航空航天大学副教授
姜 波 CCF 理事，浙江工商大学教授
苗启广 CCF 理事，西安电子科技大学教授

学会论坛

主 编 杜子德 CCF 秘书长
编 委 胡事民 CCF 会士、常务理事，清华大学教授

CONTENTS 目录

2019年3月 第15卷 第3期 总第157期



数据可视化的新疆界

数据可视化是关于数据视觉表现形式的科学技术研究，是一个处于不断演变之中的概念，其边界在不断扩大。随着数据科学的发展，可视化在不同的方向扮演着越来越多的角色。本期专题邀请了国内外的专家学者撰文，讲述生物医学、基础生命科学、流场、深度学习可解释性、隐私保护等方面的可视化，读来令人开阔眼界。相信随着社会的进步，我们能看到更多解释新发现的数据可视化新进展。

(P8~40)

卷首语

- 7 科学研究和工程技术的结合
彭思龙

专题

- 8 数据可视化的新疆界
特邀编辑：袁晓如
- 10 信息可视化及可视分析在智慧医疗领域的应用
曹楠
- 17 从细胞到分子——生物信息可视化
吴湘筠 纪海朝 伊万·维奥拉
- 23 从流线到流面：流场可视化新进展与展望
陶钧
- 29 基于可视分析的可解释深度学习
姜流 刘世霞 雷娜



阅读整本



CCCF 微博

敬告读者

欢迎读者提出意见或建议。

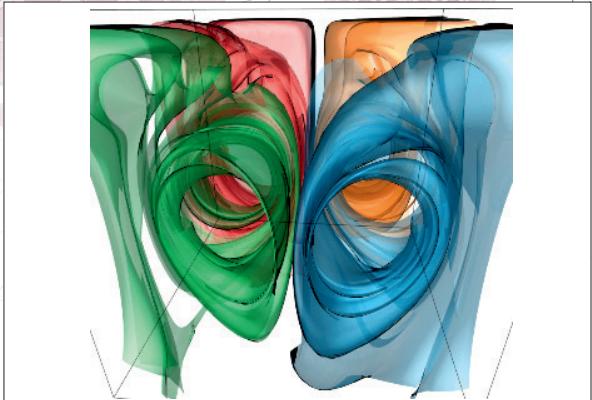
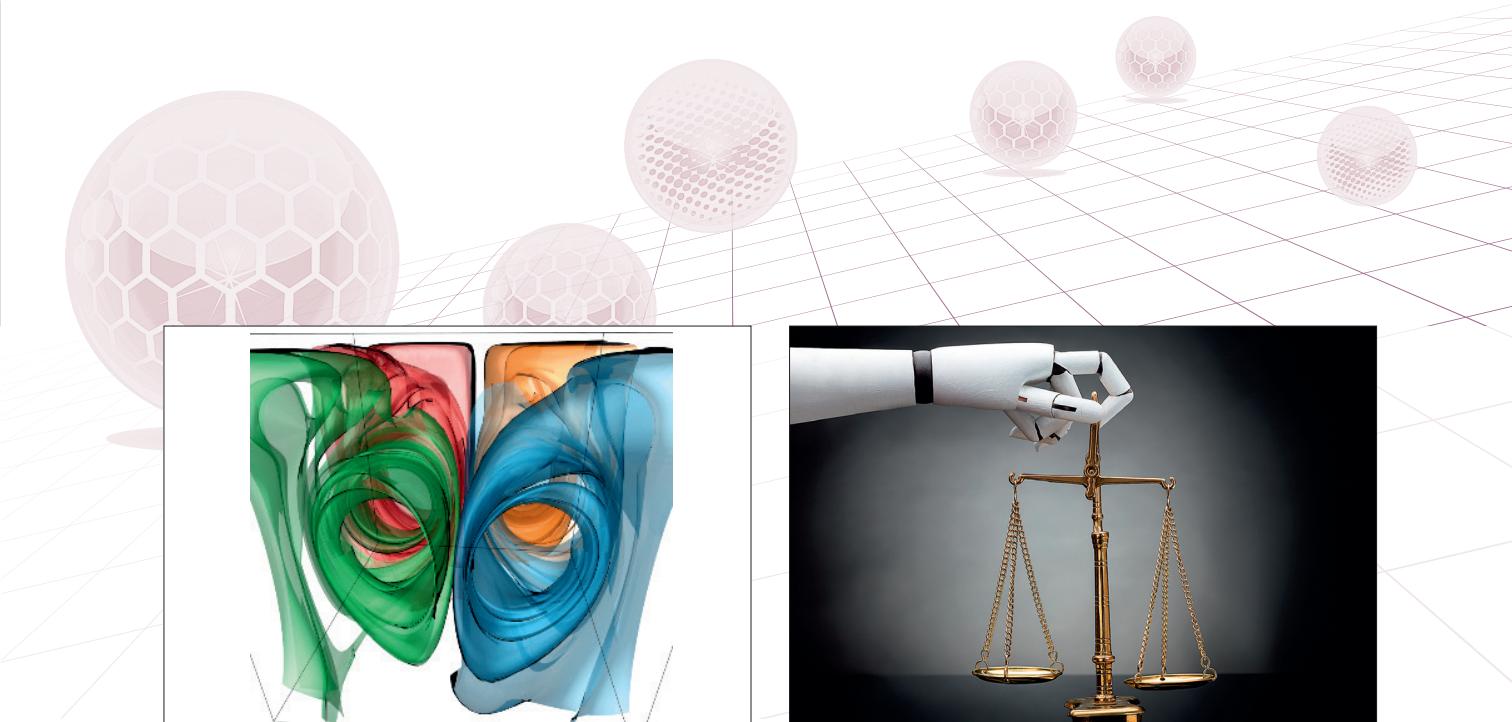
编辑部联系方式：

电话：(010)6267 0365

E-mail: cccf@ccf.org.cn

查阅电子版：

<http://dl.ccf.org.cn/cccf/list>



two swirls 数据的代表性流面
(详见陶钧专题文章)



人工智能法官和陪审团
(详见译文)

36 面向隐私保护的可视分析

王叙萌 陈为

专栏

42 计算机问题求解的三类方法

裘宗燕

47 电脑前传(3): 逻辑

黄铁军

54 我国公民基因数据安全风险与应对

穆琳 李维杰 陈海强

60 用代码可以构建人工心智吗?

宋睿华

65 让机器学会触景生情吟诗作赋

刘蓓 傅建龙

69 大规模视觉智能探索与实践

华先胜

74 The CS David 专栏

下一代网络将是什么样子?

作者: 戴维·阿兰·格里尔 (David Alan Grier)

特邀译者: 孙晓明

视点

76 P4 与可编程数据平面: 回顾与展望

毕军

动态

82 AutoML: 回顾与展望

涂威威

88 新技术 & 新应用

译文

90 人工智能法官和陪审团

作者: 洛根·库格勒 (Logan Kugler)

译者: 魏书寒 汪方野 卢瞰

94 读编往来

信息索引

• CCF ADL	封二
• 2019 CCF 青年精英大会	6
• CCCF 专题选题暨特邀编辑征集	9
• 《计算机科学技术名词(第三版)》出版	28
• CCCF 2018 年最受关注文章	41
• CCF 规章修订组工作启动	46
• 《CCCF 优秀文章精选》出版	53
• CCF 走进高校	59
• 2019 CCF CCD 计算机课程改革导教班	80
• CCF 专委活动计划(2019.4~7)	81
• CCF 会员活动中心动态	87
• CCF 会员续费	封三
• CNCC 2019	封底

Contents

Vol.15 No.3 2019/3

Preface

- 7 **Integration of Scientific Research and Engineering Technology**
Peng Silong

Features

- 8 **New Boundary of Data Visualization**
Guest Editor: Yuan Xiaoru
- 10 **Application of Information Visualization and Visual Analytics in Intelligent Medical**
Cao Nan

Data visualization has long played an important role in the medical field. In this paper, we review the relevant research results in the field of visualization from the aspects of mining and presenting potential disease laws, analyzing similar patients, and presenting medical knowledge atlas, and summarize the opportunities and challenges in this field.

- 17 **From Cell to Molecular: Biological Data Visualization**

Hsiang-Yun Wu, Haichao Miao and Ivan Viola
In this article, we introduce several visualization techniques to support scientists to explore biological data across multiple scales, including mesoscale and nanoscale, and across multiple information spaces. The vision is that we infer that scientists in the future will be able to model, visualize, analyze and communicate cell biology in an engaging 3D visual environment.

- 23 **From Streamline to Stream Surface: Recent Advances and Prospects of Flow Visualization**

Tao Jun
In this article, we review surface-based methods addressing three major challenges: construction (how to build high quality surfaces), placement and selection (where to build informative surfaces), and rendering (how to convey the information). We

further discuss the merits and drawbacks of these methods and highlight the research trends in flow visualization.

- 29 **Visual Analytics for Explainable Deep Learning**

Jiang Liu, Liu Shixia and Lei Na

Deep learning has achieved unprecedented advancement in many fields of academia and industry. Even with such advancement, oftentimes, deep learning is criticized due to the lack of explanation regarding their decisions made by models and absence of control over their internal processes. In this article, we review the explainable deep learning techniques in visual analytics community, and discuss the potential challenges and research opportunities.

- 36 **Visual Analytics for Privacy Preservation**

Wang Xumeng and Chen Wei

Privacy leaks caused by data disclosure are derivative issues that come with the development of the data industry. In this article, we introduce how visual analytics can assist in privacy preservation with state-of-the-art studies.

Columns

- 42 **Three Approaches to Solve Problems by Computer**

Qiu Zongyan

This article summarizes three different approaches to solve problems by computer depending on the characteristics and our understandings of the problems. For each approach, we analyze its application conditions, advantages/disadvantages, and other related topics. By making clear of these points, we may learn more about the potential of computer in solving problems, and explore its power better. We use AlphaGo as examples in the discussion.

- 47 **The Prequel of eBrain (3): Logic**

Huang Tiejun

Logic, originated in the Axis Age, aims to establish a solid framework for human reasoning. In the 18th

century, George Boole represented logic with algebra. Starting from the 19th century, Gottlob Frege, Bertrand Russell and David Hilbert tried to lay a solid foundation for mathematics with logic. In 1929 and 1930, Kurt Gödel proved that Frege's first-order logic was consistent and complete, but a stronger logic system (once including axiomatic system of natural numbers) could not be complete and consistent. Logic entered a new era of open exploration.

54 Risk and Countermeasures of Citizen Genetic Data Security in China

Mu Lin, Li Weijie and Chen Haiqiang

Citizen Genetic Data (CGD) contains all of the citizen's genetic information. In order to reduce the security risk of CGD in China, it is important to strengthen the protection of national CGD resources, build up the counterattack ability to gene deterrence, and participate in the international cooperation to break down the technology barriers.

60 Is it Possible to Use Codes to Build Artificial Mind?

Song Ruihua

In this article, we start with clarifying two different concepts, intelligence and mind. Then we use what XiaoIce team has done on chitchat and AI based creation as examples to introduce a different direction of making AI. Next we introduce some recent research on dialogue models, controlling AI's opinions, cross-modality poem generation from an image, and a melody and arrangement generation framework for pop music. We also discuss some promising trends of building artificial mind.

65 The Poet in the Machine: Poetry Generation from Images

Liu Bei and Fu Jianlong

In this research, we study the generation of poetic language with an image as inspiration and propose a methodology for automatic poetry creation by integrating a poetic embedding model and a poem generation model, which are optimized by two joint discriminators. We evaluate the proposed approach by extensive objective measurements, and a Turing test which includes more than 500 subjects.

69 Exploration and Practice of Large-Scale Visual Intelligence

Hua Xiansheng

In this article, we analyze current challenges of

visual intelligence in the real world and summarize several key points, which help us develop and apply visual intelligence technologies to solve real-world problems successfully. In particular, we introduce some successful applications, such as "City Brain" and "Luban", from the problem definition, technology development, product design, to business values realization.

74 The CS David

What Will Be the Next Network?

David Alan Grier(translated by Sun Xiaoming)

This article is a reading notes of Designing an internet by David Clark. The book contains a lot of valuable information about the state of the network and the way that engineers and computer scientists are thinking about how to advance networking technology. It also gives us a hint that the next generation of the internet may be a sudden change, a "paradigm shift" to support new applications.

Points

76 P4 and Programmable Data Plane: Review and Prospect

Bi Jun

Advances

82 AutoML: Review and Prospect

Tu Weiwei

88 New Technologies & New Applications

Translations

90 AI Judges and Juries

Logan Kugler (translated by Wei Shuhan, Wang Fangye and Lu Tun)

Artificial intelligence is changing the legal industry. The case for using AI-based systems to assist in the legal process hinges on the perceived ability of machines to be more impartial than humans, which would open up the law to many more people. While machines might have superior predictive power, humans will issue the final verdict on their use.



—· 中国 · 成都 · —

2019 CCF 青年精英大会

2019年5月24~25日

主办：CCF

YEF (CCF青年精英大会) 创建于2011年，每年一届。YEF旨在为计算机领域的青年精英提供深入交流和提升的机会，促进青年领军人物的成长，提升他们的领导力，促进优秀青年创新、创业以及相互之间的合作。

2019.5.23

下午： YOCSEF工作会议
YOCSEF换届

晚上： YOCSEF CLUB

2019.5.24

上午：开幕式
思想秀
科技创业秀

下午：专题论坛
晚上：颁奖晚宴

2019.5.25

上午：特邀报告
大会论坛

下午：专题论坛



<http://yef2019.ccf.org.cn/>



010-62670236转17



yocsef@ccf.org.cn

科学的研究和工程技术的结合



彭思龙

CCF 高级会员、本刊编委。
中国科学院自动化研究所
研究员。
主要研究方向为图像处理、
信号处理的算法和系统。
silong.peng@ia.ac.cn

很

长一段时间以来，在中国的科研界，科学技术研究和工程技术甚至产业界脱节是个普遍现象，表现为做科研的人员发表了大量论文，但是产业界很多具体问题无人问津。这种科学的研究和产业脱节的问题一方面反映出工程研发人员的科学修养有待提高，另一方面反映出科学的研究人员不善于从实际应用中发掘出好的科学问题。实际上，科学的研究从一开始就没有真正脱离过实际应用。最早期的欧几里得几何算是当时“高大上”的科学的研究领域，其实是为了丈量土地；即便是天文观测这样看似比较虚的研究，也与定时和定位技术紧密相关，并没有脱离应用的范畴。

在中国出现这样的现象有一定的历史必然性。中国大规模的科学的研究的历史并不长，仅仅几十年时间，所以在过去大多数时间里，中国所谓的科学的研究大多数还是一种科学练习，处于不断学习科学的研究的过程。这就导致中国的科研人员都去研究所谓的世界热点问题，而这些热点问题基本上都是国外根据自己的需求提出来的。也就是说，我国的大多数科研是替别人解决问题，而不是根据自己的需要解决问题。另一方面，在大多数工程领域，中国处于追赶状态，很多问题都有可以参考的对象，这就导致工程技术人员热衷于模仿先进的对象而不愿意从新的角度思考问题，也就提不出像样的科学问题。

随着中国的工程能力不断提升，到了要与国外的产品或者技术进行竞争的时候，一味的模仿已经不能带来有效的利益，这就迫使从事工程技术研发的人员不得不思考新的技术路线，也就有可能发现新的科学问题。尤其是新兴领域，国外没有成型的技术可用，失去了参考对象，我们不得不自己摸索，这时候产生科学问题的机会就会大大增加。另外，科学的研究需要相当多的资源，尤其是实验和各种环境都需要大量的资金和人力投入，中国的企业没有高利润的产品为基础资源进行资本积累，就没有实力进行奢侈的研究，自然也就与所谓的主流科研脱节。随着中国企业的竞争力不断增强，产品的利润不断上升，企业就有更好的能力进行科学的研究的投资。比如在人工智能领域，中国很多大公司都投入了巨资进行科研，其中很多科研成果达到了世界领先水平，这在过去是不可想象的。

尽管过去一直存在科研和工程的脱节，但是放眼未来，中国的科研必然会逐渐落地，落实到解决中国需要解决的问题；另一方面，企业对于科研的投入越来越大，工程和科研结合的技术也就越来越多。期待我国的科研能够在这种结合中结出硕果，不断产生世界级的好成果。这种局面似乎正在到来。 ■

数据可视化的新疆界

特邀编辑：袁晓如
北京大学

关键词：数据可视化

在一些人的眼中，可能可视化只是把计算分析的结果图形化。实则不然，可视化是人类认识、分析复杂数据的重要途径。随着数据科学的发展，可视化在不同的方向扮演了越来越多的角色。本期我们将讨论一些新的可视化工作方向。

从世界范围来看，可视化这一学科方向的形成和发展是在 20 世纪 80 年代末，其重要的推动力来自科学计算和观测技术的进步，特别是医学图像领域中 CT、MRI 等成像技术的出现和发展。如果回顾 90 年代的可视化研究工作，会发现相当一部分研究集中于三维医学成像数据的可视化，即体可视化 (volume visualization)。

在经历了 30 多年的发展之后，可视化和生物医学之间的联系依然紧密，但是研究范围逐步扩大，转移到了更为复杂的对象上。在医疗领域，随着信息技术的发展，医疗过程除了常规性地产生和收集各种影像数据以外，就诊者各项生理指标的测量数据也能被系统地收集并和其他就诊记录一同被纳入电子病历的数据管理之中。对于复杂的病历，一方面需要将其以合理先进的方法提供给病患个体，用于医患沟通；另一方面，挖掘分析其中的内在规律、发现问题也是医疗工作者的迫切需要，可视化与可视分析可以提供相当有效的手段。这方面早期的工作可以追溯到 1994 年耶鲁大学统计学家爱德华·塔夫提 (Edward Tufte) 在《柳叶刀》上发表的论文 “Graphical Summary of Patient Status” 和 1996 年马里兰大学本·施耐德曼 (Ben Shneiderman) 与同事在国际人机交互会议 CHI 上发表的 “LifeLines: Visualizing

Personal Histories”。经过多年的发展，如今不仅可以通过可视化理解分析个人的电子病历记录，相关研究还拓展到大规模人群的群体相似性分析和基于医疗知识图谱的可视分析技术上。本期同济大学曹楠教授撰写的《信息可视化及可视分析在智慧医疗领域的应用》一文将深入介绍上述方向的工作。

可视化和基础生命科学同样密切相关。好莱坞科幻电影的镜头中，经常出现从人体开始，推进放大到血液中细胞乃至分子的运动变化等科幻场景。这些场景并不只是出现在电影的蒙太奇手法之中，科学家本身也非常关心大规模跨尺度、跨模态的成像理解。最近我国支持建设的一系列大科学装置中，就有“多模态跨尺度生物医学成像”国家重大科技基础设施项目，希望提供革命性的研究手段，对生命体结构与功能进行跨尺度可视化描绘与精确测量，进而破解生命与疾病的奥秘。因此本期我们邀请了来自奥地利维也纳科技大学和沙特阿卜杜拉国王科技大学的伊万·维奥拉 (Ivan Viola) 及其同事撰文《从细胞到分子——生物信息可视化》，讨论如何利用最新的可视化和图形学技术，组合结构元素创建完整生物体的结构模型。值得注意的是这些成果是由计算机科学家与来自美国斯克里普斯研究所 (The Scripps Research Institute, USA) 综合结构和计算生物学系 (Department of Integrative Structural and Computational Biology) 的科学家共同合作完成的。

纵观欧美发达国家的科研力量的配置，我们可以看到和生物医学相关的研究占据了相当大的部分。美国国立卫生研究院 (NIH) 的经费力度要

远超过美国国家科学基金会(NSF)。从上面的例子，我们可以看到更多计算机科学与生命科学合作的机会。

回顾可视化初期历史，和研究的医学图像数据相对应，还有一类是模拟计算特别是流场数据。对于流场的图形化描绘可以回溯到达芬奇的笔记，在其中已经可以看到对水流的线形描绘。类似的抽象绘画形式也可以在中国传统的山水画中找到踪迹。然而，对于模拟流场数据忠实、高精度的科学呈现还是近几十年来的工作。这方面的研究，对于提高航天航空飞行器设计水平、理解大气海洋行为等都有重要意义。而从可视化方法来看，流线相关的方法已经发展得较为成熟，我们邀请了来自中山大学的陶钧教授撰写《从流线到流面：流场可视化新进展与展望》一文，讨论流场可视化面临的挑战及机遇。

除了以上提到的几个传统应用的新扩展外，在计算机科学内部，可视化和诸多最新发展的前沿领域有密切的融合。随着深度学习近年来在各个应用领域的成功，以机器学习为代表的人工智能技术迅速扩展到社会的各个角落。使用者在享受全自动人工智能模型便利的同时，也因为无法理解模型内部的工作机理，遇到了巨大的困扰。提高工作机理复杂的深度学习模型的可解释性，对于技术发展本身和关键决策过程应用都有重要意义，也因此受到世界各国重视，成为当前学科的一个热点。清华大学的刘世霞教授是可视分析领域中较早对深度学习的可解释性进行交互式可视化探索和分析的学者。我们邀请她撰写《基于可视分析的可解释深度学习》一文，从模型理解、模型诊断和模型改进方面讨论利用可视分析对学习模型开展的研究。

在智能和大数据时代，另一个引起关注的问题是数据的隐私保护。如果把整个可视化看作一个信息的传播系统，我们可以看到在信道的不同阶段，通过合理控制信息的转化，有可能在数据通过可视化传递给用户的同时，提供相应的隐私保护。本期专题我们邀请了浙江大学的陈为教授撰写《面向隐私保护的可视分析》一文讨论这一新兴的议题。

人类对于理解世界的需求是无止境的。可视化通过视觉这一通道，和交互结合，为人类用户提供高效的数据分析理解能力。人作为婴儿来到这个世界的时候，通过眼睛看到母亲，看到缤纷的世界；在数据时代，人通过可视化看到变幻万千的数据空间，相信随着社会的进步和发展，我们将看到更多的拓展和尝试。 ■



袁晓如

CCF 理事、CCCF 专题主编。北京大学研究员。
主要研究方向为可视化与可视分析等。
xiaoer.yuan@pku.edu.cn

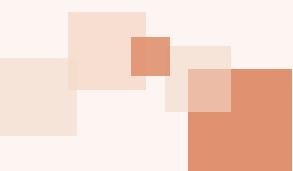
CCCF专题选题 暨特邀编辑 征集

专题是《中国计算机学会通讯》(CCCF)最主要、最受读者欢迎的特色栏目，通过业界顶级专家撰稿，对前沿技术论题全面而深入的阐释，使读者了解计算机领域前沿动态。

专题采取邀稿、投稿相结合的组稿方式。CCCF 编辑部持续向业内专业人士广泛征集专题选题暨特邀编辑。如果您是计算机及相关领域的专家，对本领域的科学技术有深入的研究和实践积累，并能组织本领域专家撰写一组高水平的专题稿件，希望您能应征提出选题计划，担当此选题特邀编辑的职责，为计算机科学技术的传播和发展尽一份力量。

读者也可提出您感兴趣的选题建议。

联系 : cccf@ccf.org.cn



信息可视化及可视分析在智慧医疗领域的应用

曹楠
同济大学

关键词：信息可视化 可视分析 智慧医疗

医疗健康是与每一个人直接密切相关的重要科学领域。科学家在探索生命奥秘和疾病产生机理的过程中，一直重视对跨学科技的运用。从基于虚拟现实技术的仿真手术到手术机器人，从医学成像技术到医学图像处理，从大数据分析到人工智能，越来越多的新兴科技被应用到医疗领域，也提高了患者在就诊治疗过程中的安全保障。

长期以来，科学可视化 (scientific visualization) 技术¹ 在医疗领域一直扮演重要角色。无论是平面 X 光扫描，还是三维 CT 影像，都应用了科学可视化的相关技术。然而这些技术仍然局限于对具象数据（例如人体骨骼、器官组织结构等）的展现。随着互联网的普及和可穿戴设备的广泛应用，越来越多与医疗相关的抽象数据被采集了起来，对信息可视化 (information visualization) 技术² 提出新的需求，主要包括：(1) 展现用户的个人健康信息，例如心跳、血压等状态；(2) 汇总并展现公众健康信息，例如禽流感的扩散趋势、不同地区的人民健康状况等；(3) 分析并展现临床电子病历记录 (Electronic Health Record, EHR) 中的规律与模式，例如疾病的演变过

程以及诊疗方案的疗效等。前两类可视化应用一般面向不具备医疗知识的普通用户，因此往往采用传统直观的信息可视化形式，如柱状图、折线图等，便于用户理解与阅读。第三类应用主要面向医生等具有专业背景，需要对数据进行深入调查并做出职业判断的用户，因此更具针对性，其可视化及相关分析技术的设计也更具挑战性。

在过去近十年，研究人员针对第三类可视化应用设计并开发了一系列与智慧医疗相关的可视化技术，主要用于：(1) 挖掘、展现并预测电子病历记录数据中潜在的规律及风险；(2) 帮助医生针对病患特征进行病人群体的相似性分析 (cohort analysis)；(3) 展现大规模医疗知识图谱 (knowledge graph)。

针对电子病历记录的可视化及可视分析

电子病历记录了患者就诊以及用药治疗的完整过程。它从患者和医生两个角度分别刻画了疾病在不同人群中演变发展的过程以及治疗方案在不同人

¹ 科学可视化是一个跨学科研究与应用领域，主要利用计算机图形学来创建视觉图像，帮助人们理解科学技术概念或结果的那些错综复杂而又往往规模庞大的数字表现形式。

² 信息可视化是一个跨学科领域，通过利用图形图像方面的技术与方法，帮助人们理解和分析数据。与科学可视化相比，信息可视化则侧重于抽象数据集，如非结构化文本或者高维空间当中的点。

群中所带来的不同疗效。因此，对于电子病历数据的分析与可视化具有重大临床意义。

电子病历数据及其带来的挑战

电子病历数据往往以文本或者表格的形式存储于计算机系统中。它可以被进一步抽象并转化成由医疗事件构成的事件序列数据。序列中的每个项记录了电子病历中的一次“就诊”“诊断”“用药”“化验”“手术”的记录，或者“入院/出院”这样的行为，或者“康复/死亡”这样的结果，以及这些事件发生的时间。针对这样的数据类型，可视化的设计空间可以由时间和事件两个主要的信息维度构成。

- **时间信息维度**：包括事件发生的时间，先后顺序，事件之间的时间间隔，周期性规律等可供设计的主要元素。

- **事件信息维度**：包括单个事件的类型，出现的频率，对应参数属性的大小（例如化验结果、用药剂量等），以及多个事件同时出现（或共同发生）的规律（co-occurrence）等设计元素。

任何针对事件序列数据的可视化都可看作是在这两个信息维度上挑选不同的设计元素，采用不同的设计方案及编码方式而构成的。然而，这个看似简单的任务在实际应用中却面临着诸多挑战。

首先，就时间信息维度而言，同样的疾病，有可能因为患者的自身差异、不同的时间安排、医生采用不同的诊疗手段等诸多因素，导致电子病历所记录的事件序列具有极大的差异，很可能同一组事件在不同患者身上，发生的具体时间、先后顺序、持续的周期、对应的属性取值均不相同，为事件序列数据的汇总与可视化带来了困难。

其次，从事件信息维度而言，考虑到药物、化验以及治疗手段的多样性，真实的电子病历数据中可能包含有数以万计的事件类型（例如，服用不同的药物均可看作是不同的医疗事件），从而导致数据具有高维异构性。不仅如此，真实的电子病历数据往往规模庞大，包含数以万计的病人和历时数十年的记录。这些都为可视化的设计带来了挑战。

大规模电子病历数据的可视化汇总

对大规模电子病历数据（即事件序列）进行汇总，是挖掘数据中潜在模式的必要手段。现有基于数据挖掘算法的技术实现了对原始事件序列数据的高度概括。通过分析，能够直接获得频繁子序列等事件序列中的相关模式^[1]，但同时也丢失了数据中的细节及上下文信息。而现有的可视化技术，例如EventFlow^[2]和CareFlow^[3]，多是对原始数据的直接展现，无法对较大规模的事件序列进行汇总。因此需要一种既能够汇总展现大规模事件序列的数据信息，又能够展现足够上下文细节的可视化技术。

为此，研究人员提出了EventThread^[4]，一种面向大规模事件序列数据的可视化汇总方案。如图1所示，该技术通过一系列数据处理及分析步骤，将原始的事件序列汇总转换成用于可视化的“潜在序列(latent threads)”。在具体应用中，潜在序列可被视为一种潜在的“治疗方案”。该技术首先剔除了数据中的小概率事件，以减小数据噪声（图1(a)）。接着，将事件序列靠左对齐（图1(b)），并将对齐后的序列按照定长的时间窗口切割成若干个阶段（图1(c)），用于代表离散化的时间维度。基于上述处理，该技术进一步将数据转换成一个包含病患、

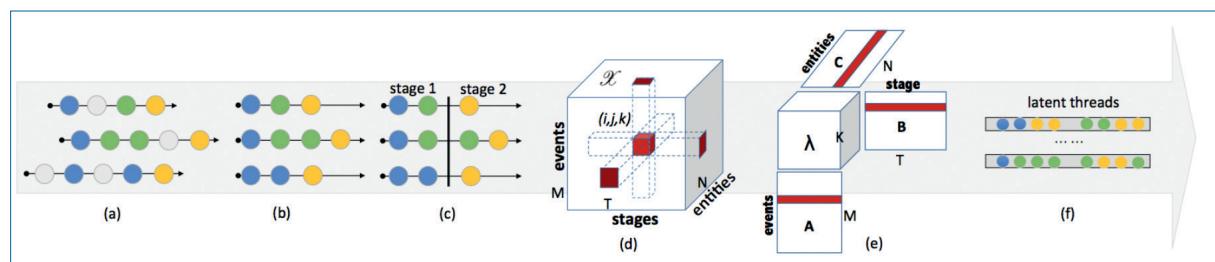


图1 EventThread 对事件序列的处理及建模流程

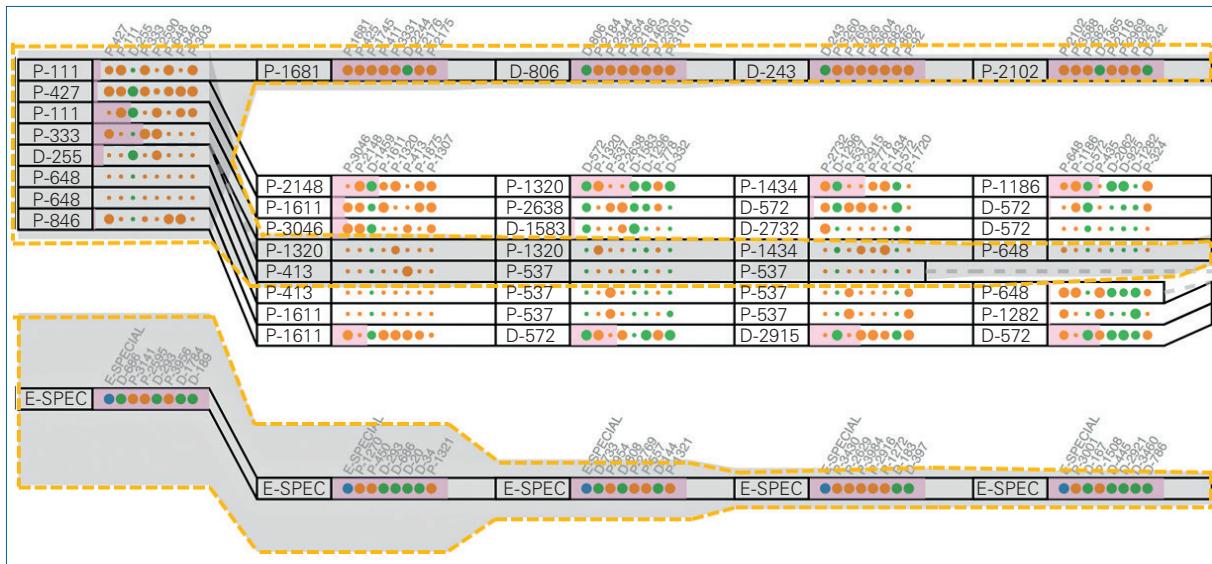


图2 EventThread 可视化设计。每一个“潜在序列”被画成平行的一行，相类似序列之间的距离也较为接近。序列中的小点代表对应阶段中的关键事件

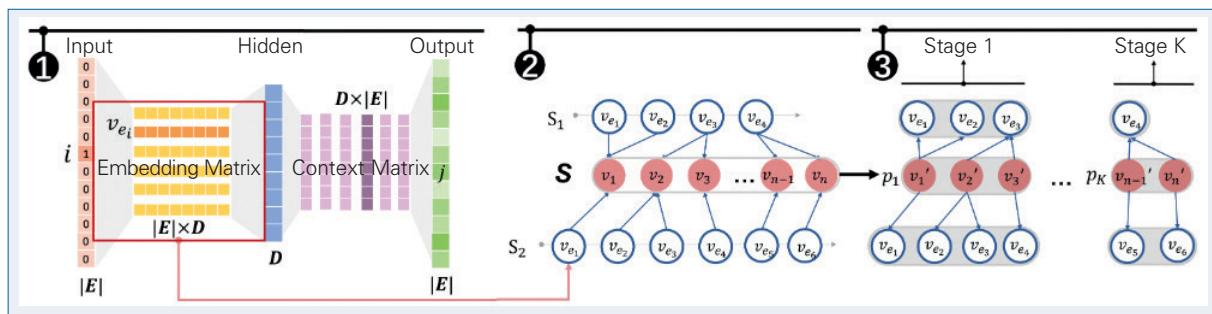


图3 EventThread-2 中针对事件序列数据的阶段性分析

阶段、事件三个维度的张量（图 1(d)），并将其进一步分解为三个因素矩阵 (A, B, C)。其中， B 和 C 分别代表了潜在序列（治疗方案）在病人群体中以及在时间维度之上的分布，而 A 则代表了事件在“潜在方案”上的分布，即该方案是由哪些具体医疗事件构成的。结合 A 和 B 中的信息，便能够得到“具体有哪些病人，在什么阶段与哪一个‘方案’相关”这样的上下文信息。

图 2 展现了我们对上述数据处理结果的可视化展现方案。该图汇总显示了 2 万名“慢阻肺”患者前后 8 年的电子病历数据。通过直接展现“潜在序列”，我们能够很好地汇总大规模电子病历数据。通过结合因素矩阵中的相关信息，我们能够为每一个

序列提供足够的细节及上下文信息，以帮助用户理解每一个序列的具体含义。图 2 中的可视化设计清晰地揭示了“慢阻肺”采用的三种不同的诊疗方案，其中中间的一组是较为主流的方案，由多根并列的序列共同展示。

针对疾病演变的阶段性分析及可视化

EventThread 方案虽然能够汇总并显示大规模数据，但存在两点缺陷：(1) 未能解决序列多样性的的问题，简单地将序列靠左对齐，并不是一个合理的方法；(2) 将序列分割成阶段时，使用了固定长度的事件窗口，这样的分割有可能把本来连续发生的事件序列（例如一次住院过程中所产生的所有事件），分

割成两个不同的阶段，从而导致分析错误。为了解决这两个问题，研究人员对 EventThread 方案做了进一步修改，提出了 EventThread-2^[5]。该方案采用新的数据分析方法及可视化设计方案，针对事件序列数据进行阶段性分析，以更加清晰的视角展现疾病的演变过程或诊疗方案的阶段性进展。

图 3 展现了 EventThread-2 所提出的阶段性分析算法。该算法首先通过神经网络将事件序列数据中的每一个事件，根据其相互之间的相关程度映射到高维空间，实现事件到向量的转变（图 3 ①）。接着，通过动态时间规整(dynamic time wrapping)算法，将长短不一、顺序各异的事件序列根据相关事件之间的相似度进行对齐，并将对齐后的结果汇总在一个虚拟序列 S 之上（图 3 ②）。最后，通过对 S 进行切分，划分出序列发展的不同阶段。在此过程中，算法确保被划分在同一段中的事件具有较大的相似度，而不同段之间的事件在向量空间中具有较大的差异性，将切分结果进行拆分，可得到针对每一条

序列的阶段性分析结果。该技术很好地解决了事件序列长短、顺序不一致等问题。

图 4 展示了利用 EventThread-2 技术对一组患有心脏病的 ICU 患者的电子病历数据及相应事件序列进行阶段性分析的结果，清晰地再现了患者从入院到化验，再到药物治疗、手术、及最终出院的全部治疗过程。可以看出，该技术能够准确地对医疗事件进行切分与汇总。

病人群体的相似度可视分析

病人之间的相似性分析在医疗领域也具有重大意义。首先，当遇到疑难杂症时，在病例库中查询相似病人及相应的诊疗方案对救治当前患者具有一定的参考意义。除此之外，将病人按照相似度进行归类，有助于医生根据不同的病患特征，制定不同的诊疗方案。现有的基于机器学习的相似度分析往往缺少可解释性，在给出病人之间相似度数值的同

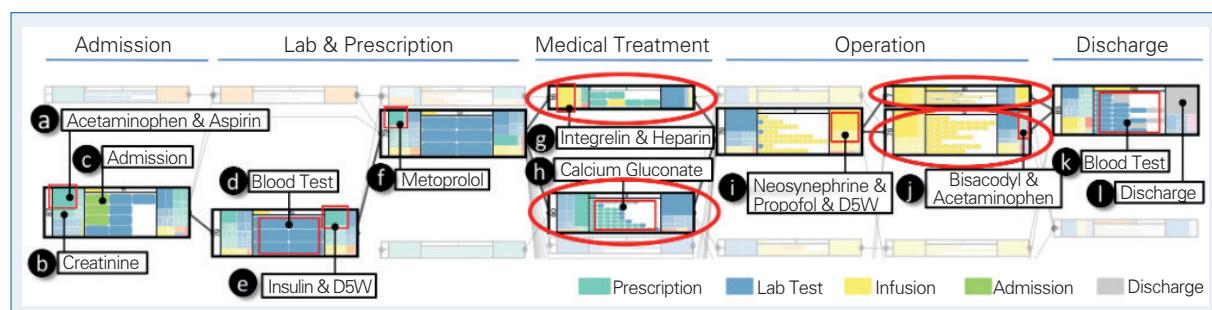


图 4 针对 ICU 患者入院序列的阶段性分析及可视化展示

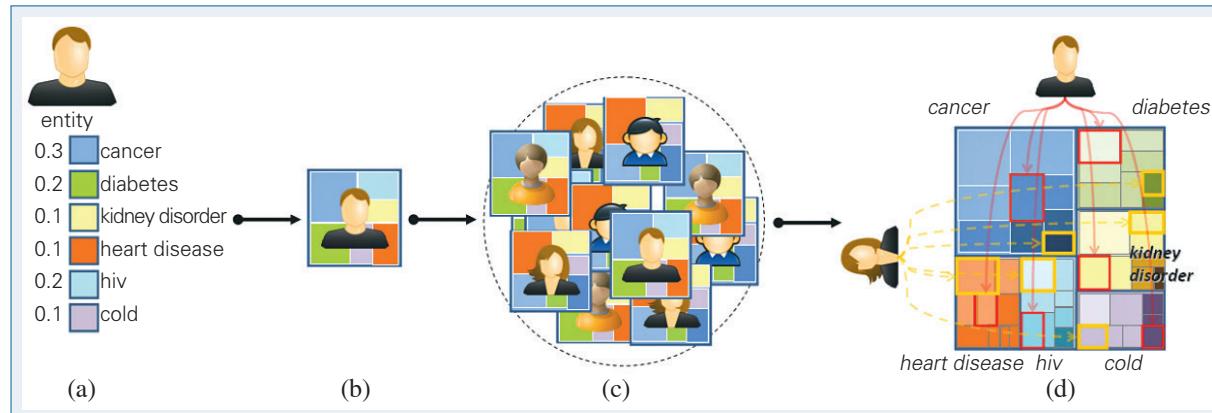


图 5 利用动态图标技术展示病人间的相似程度

时，不能够清晰直观地为缺少技术背景的医生解释病人为什么相似，在哪里相似。因此可视分析技术在该研究方向上的应用也非常重要。

为了提供直观的相似度分析工具，研究人员开发了动态图标技术 DICON^[6]，用于直观展现病人的多维度特征，方便不同病患之间的相似度对比。该技术仍然基于对电子病历数据的分析，但在设计 DICON 时，我们忽略事件维度的信息，仅统计病人因特定疾病去就诊的次数，并以此作为病人的特征。如图 5 所示，根据电子病历记录，每一位病人都可能患有多种疾病，每一种疾病都被展示成一个矩形，疾病的种类用不同的颜色表示（图 5(a)）。病人因病就诊的次数被归一化处理后可以用矩形的大小表示，因此每一位病人都可以通过封装代表疾病的小方块，而被展示成一个方形的图标（图 5(b)）。面对一个拥有众多患者的病人群体，这样的设计依旧有效。通过拆分重组，我们把群体中不同病人的相同疾病封装在一起，从而构成了整个群体的图标（图 5(d)）。

这种可视化设计充分利用了图标技术的高可比性，让具有相似疾病的患者有相似的表示，不相似的患者有所区分。同时，我们对病患个体以及群体采用类似的设计方案以及完全一样的可视化编码方式，从而降低了可视化在理解上的难度。动态图标技术还可以与其他可视化图表一起使用。如图 6 展

示了把动态图标技术应用在散点图中，用来展示多维度上下文信息的场景。

医疗知识图谱的构建及可视化

医疗知识图谱涵盖了医疗领域的相关概念（例如疾病、药物、症状等），主要用于构建智慧医疗体系中的自动问答系统。与其他知识图谱类似，医疗知识图谱往往是异构的，并且包涵大量节点与链接。知识图谱的构建需要对大规模的文本文件进行处理，以提取其中的相关概念，并构建这些概念之间的联系以构成知识本体。此外还需要直观展现知识图谱中所包含的复杂关系，提供高效的查询机制，从而提高分析人员浏览知识图谱的效率。

基于上述需求，研究人员设计开发了一系列针对文本数据的知识图谱构建及可视化系统。如图 7 所示，研究人员首先对疾病文档进行分析处理，将每一个记录疾病信息的文本文件按照疾病不同方面的描述（如症状、治疗方案、病因等）切割成多个信息层面，并从不同的信息层中提取实体关键词（如疾病或症状的名称），去重后构成实体关键词集合。基于该集合，实体之间分别以“内在关联”和“外在关联”两种形式建立相互联系。“内在关联”是实体在同一信息层面出现的关联（如两种互为并发



图6 在散点图中使用动态图标技术。横轴及纵轴分别代表肾功能失调（蓝色）及糖尿病（橙色）两种不同疾病，其他颜色代表其类型的疾病

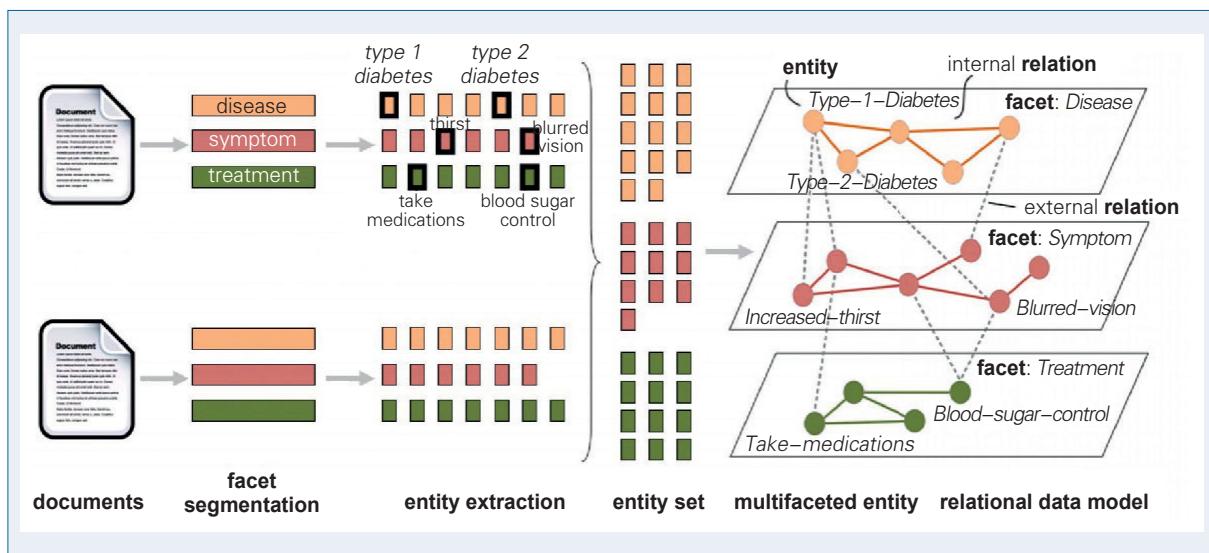


图7 疾病知识图谱的构建以及多层次实体关系数据模型

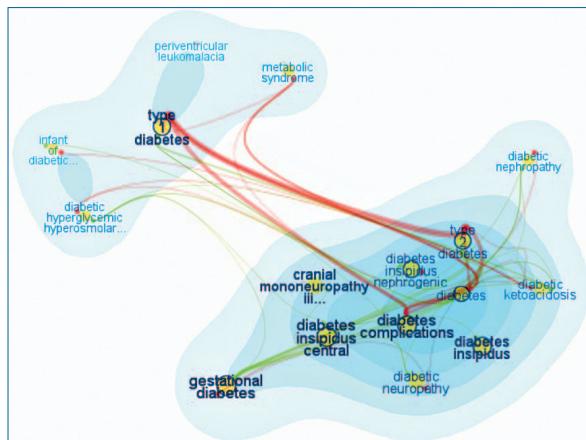


图8 利用 FaceAtlas 展现的两类糖尿病及其并发症之间在症状(红线)及治疗方案(绿线)之间的联系

症的疾病),而“外在关联”则是实体在不同信息层面出现的关联(如疾病及其症状之间的联系)。这样一系列处理将无结构的原始疾病信息文本文件转换成一个结构化的“多层面实体关系数据模型”,构成了医疗知识图谱中的内部基础结构。基于该模型,研究人员提出了多种相关的可视化设计,用于展现模型中蕴含的不同信息。

FaceAtlas^[7]是医疗知识图谱的一种可视化方案,该方案聚焦于显示同层面实体之间的外在联系。

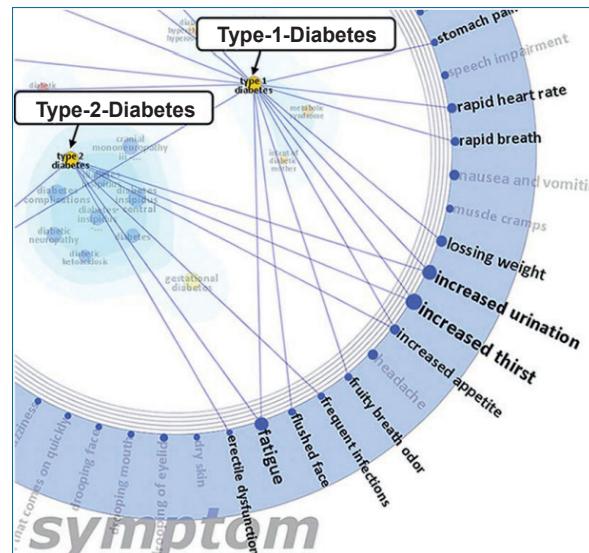


图9 SolarMap 被设计用来展现不同实体之间的详细外在联系

图8展现了两个分别以1型糖尿病和2型糖尿病为核心的类。红线代表了疾病之间在病因上的关联,而绿线代表了疾病之间在症状上的关联。FaceAtlas还提供了完整的文本搜索机制,用户可以通过关键词搜索相关疾病,系统将自动构建与该疾病相关的知识图谱,并以可视化的形式展示。

虽然FacetAltas展示了同层信息之间在不同层

面上关联的强弱（例如，疾病之间在症状、病因、治疗方案上的联系分别用不同颜色的线表示，线的粗细代表了联系的强弱），但无法显示两个疾病具体在哪些症状上有关联。为了克服这一设计缺陷，研究人员又进一步设计了 SolarMap^[8]。该可视化设计首先选择一个主要的信息层面，然后把该层面中的主要实体以图的形式显示在可视化视图的中心位置。如图9中所显示的可视化结果以疾病作为主要信息层面，可视化的中心视图展示了疾病及其并发症构成的网络。围绕着中心视图，其他信息层面被显示成一层层的圆环，用户可以通过交互的方式选择展开某一层次的圆环，相关信息层面中的实体信息以关键词的形式显示在该圆环的相应位置，这些实体与中心实体之间的关系则通过连线来展示。这样用户可以清晰地看到两种疾病在症状上的具体联系。

研究机遇与挑战

信息可视化及可视分析技术在智慧医疗领域的诸多方面都起着举足轻重的作用。然而，这方面的研究仍处于较为初级的阶段，有很多问题尚待解决，也存在诸多挑战与机遇。

第一，从数据角度而言，如何对大规模异构医疗数据进行融合以方便分析，仍然是一个较大的挑战。真实世界的医疗数据往往是非常复杂的，既包括像X光片、CT扫描等各种影像数据，也包括化验结果等结构化的表格数据，还包含诊断、医嘱等无结构文本数据。医生需要结合上述所有信息方可对病人作出诊断，这对数据分析技术提出了挑战。如何融合并展现来自各种渠道不同类型的数据，并建立它们之间的关联，方便分析人员从不同视角进行观测和总结，并帮助他们汇总理解相应的分析结果，是可视化技术面临的一个重要挑战。

第二，从技术角度而言，随着人工智能技术在智慧医疗领域的广泛应用，越来越多的辅助诊疗系统及算法被开发出来，辅助判断患者可能的疾病状况，推荐相应的诊疗方案，以及对当前疾病进行预后分析。然而，所有的分析算法及人工智能技术均

面临着可解释性的问题。当算法做出的判断无法解释或者不足以让医生信服时，将无法真正在医疗领域进行部署。因此有必要设计新的可视化技术，帮助全面解释算法分析的结果，回答诸如“为什么预测A疾病发生而不是B疾病？”“服用药物A的预期疗效为什么比服用药物B好？”等一系列问题，以帮助用户理解相应的计算结果。

第三，从医学角度而言，不同疾病拥有不同的病因、发病机理以及演变过程，往往无法用一种通用形式对所有类型的疾病进行分析展现，而需要针对不同的疾病设计不同的分析算法及可视化技术。相应地，可视化设计人员需要对相关医疗知识进行较为深入的理解与学习，从而大大延长了可视化设计开发的周期。因此，如何根据疾病的大类别汇总分析相关数据中的共有属性，在医学理论的基础上增加可视化设计的通用性，有待进一步研究。 ■



曹楠

同济大学教授，博士生导师。主要研究方向为信息可视化及可视化分析。

nan.cao@tongji.edu.cn

参考文献

- [1] Han J, Cheng H, Xin D, et al. Frequent pattern mining: current status and future directions[J]. *Data Mining and Knowledge Discovery*, 2007, 15(1): 55-86.
- [2] Monroe M, Lan R, Olmo J M, et al. The challenges of specifying intervals and absences in temporal queries: a graphical language approach[C]//ACM Conference on Human Factors in Computing Systems(CHI). ACM, 2013: 2349-2358.
- [3] Wongsuphasawat K, Gotz D. Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 18(12): 2659-2668.
- [4] Guo S, Xu K, Zhao R, et al. EventThread: Visual Summarization and Stage Analysis of Event Sequence Data[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2018, 24(1): 56-65.

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

从细胞到分子 ——生物信息可视化

吴湘筠¹ 缪海朝¹ 伊万·维奥拉^{1,2}

¹ 维也纳科技大学

² 阿卜杜拉国王科技大学

关键词：生物信息可视化 结构生物学 生物途径图

交互式可视化系统

生物科学 (biological sciences) 是一类研究生物如何发挥其生命功能的科学。通常，生物科学是以还原论的方式工作。生物学家确定一个特定而详细的研究问题，并尝试所有可能的组合因素来寻找该特定问题的答案，借此发现了很多关于生物学的知识。然而还原论在研究初期容易与其他知识隔离，产生知识碎片。综合科学方法学的目标是将这些知识碎片整合起来形成一幅更大的图景。借此方法，来自美国斯克里普斯研究所 (The Scripps Research Institute) 的科学家将已知结构元素的碎片组合在一起，创建整个小生物体的结构模型^[1]。

生物体，特别是单细胞生物体，由三个基本的大分子构建模块 (building block) 组成：磷脂双层的组合形成周围的膜，纤维结构负责更大的结构特征、能源或遗传信息，最后由蛋白质作为执行生物基本功能的机器。生物功能网络非常复杂，但是它基本在水溶剂环境中运行且作用在相对较小的化学物质上，而供应的能量通过 ATP 等分子传递。这些构建模块可以在计算机上建模并组合在一起形成有机体的模型。这正是斯克里普斯研究所的科学家们所开创的。他们一直在微观尺度和纳米尺度下分析关于特定生命形式的可用科学信息。这些信息是程序建

模系统的基础，该系统逐渐将所有基本构建模块打包成一个完整生物体的模型。

虽然模型本身对生物学具有革命性贡献，但是这些模型极难管理和展示。一个标准的渲染过程需要长达数小时的时间才能合成单个图像。近年来我们分析了这些数据的特征，提出了一种可以更有效处理并渲染大数据的新技术，比参考标准法快 1,000 倍左右^[2, 3]。生物科学家可以利用这个强大的工具，交互地查看和分析复杂的生物模型。这引起了整个结构生物学的改变，同时也给计算机图形学和可视化的改变带来了契机，因为该方向的研究正面临几个基于潜在模型属性的全新要求。首先，该模型必须整合多个尺度的结构细节。其次，与计算机图形学中的标准场景相比，它的三维模型非常密集。最后一个有趣的属性是多实例性 (multi-instance)，即一个结构元素在场景中可以重复出现几千次^[4]。这些新要求对于如何渲染多个尺度组成的模型，如何将其可视化，以及如何在其中有效地探索提出了新的挑战。此外，由于模型程序生成也是一项耗时的任务，尤其是在设计序贯算法 (sequential algorithm) 时，因此高效且实时的模型构建是解决此问题的额外挑战。

我们系统的图形渲染性能采用专用技术加速并针对特定数据类型进行优化。首先，将所有结构实

例都在一次绘制调用中生成，并且通过使用细分曲面着色器 (tessellation shaders) 和几何着色器 (geometry shaders)，实时显示几何物体。这种几何形状表示在程序中根据物体与相机距离，选择适当的呈现细节。对于可能被遮挡的结构，一个改进的层级 z 缓存 (hierarchical z-buffer) 会选择应对场景的物件并将其发送到绘图管线里。虽然这样的渲染效果快速且质量高，但数据的密度结构更自然地要求有效的遮挡处理 (occlusion handling) 策略^[5]。对此，我们采用剪切几何的经典概念，根据场景结构的可见性进行分层去除。此外，我们还设计了一个新概念，即允许用户控制同一类型的特定结构元素实例的可见性，用来探索整个场景。分子的多尺度表示法还允许我们对使用者显示适当的几何细节。除了几何外，结构元素与相机的距离也控制着它的着色技术以及应用于特定元素的着色方案^[6]。

从导航探索的角度来看，复杂的场景要通过文字标签 (text label) 来传达。在我们的系统中，使用三维内部多尺度标签，每个标签仅针对每种类型放置一个代表性结构元素，而不是为每个结构实体都显示其标签^[6]。用来描述和传达纳米尺度结构及其相互作用的技术刚刚开始发展。我们的目标是提供一个完整的真核细胞，而不只是呈现一种小细菌。其中的困难是，我们不能直接在图形存储器上储存这种细胞的原子细节。但是基于明确定义的规则，我们可以快速构建这种几何形状。通过复杂的算法，借由并行算法即时生成脂质网、纤维结构和蛋白质分布。

生物途径图

生物途径 (biological pathway) 用于抽象表示在生物细胞内发生的一系列化学反应。为了深度理解

这些化学物质的反应关系，科学家经常使用生物途径图 (biological pathway diagram) 从视觉上表达生物化学反应的程序，并且用图 (graph) 对生物途径图进行建模，因为化学物质可以表示成节点 (node)，反应程序的过程则可以表示成有向边 (directed edge)。目前已知最大的手绘生物途径图——人体代谢网络包含大约 12,503 个节点和 31,540 条有向边，由虚拟代谢人 (Virtual Metabolic Human)¹ 团队检验和整理，并使用途径图软件 CellDesigner² 绘制而成。为了正确表达各化学物质的关系、分类，这张图由 5 位专家耗时 20 个月分析绘制而成。

由于生物途径是根据途径实验显著性的结果构建而成，加上新研究成果不断被发表，旧式的静态途径图无法跟上新研究的动态变化。例如，传统上葡萄糖被认为是一种快速供应人体能量的物质，但是最近研究证明葡萄糖高度影响癌症的代谢途径。为此，科学家需要在原本的生物途径里追加新的有向边来研究葡萄糖对人体正面和负面的影响。此行为改变了原本生物途径的拓扑结构，甚至影响了生物途径局部和整体的视觉表示。可视化研究人员由此开始研究更好的途径图表示法。常使用的图可视化方法有力导向图 (force-directed graph drawing)、应力主导图 (graph drawing by stress majorization)、分层图布局 (hierarchical graph layout)、正交布局 (orthogonal graph layout) 以及概要图 (schematic graph layout) 等^[1]。但是我们即使满足了图可视化的美学准则，其结果也不一定适合生物途径图。这是因为标准的图可视化方法常牺牲微观细节来维持宏观形态。不同于社交网络，生物途径网络必须同时向使用者提供微观和宏观细节，这进一步加大了绘制大型网络的难度。

我们参考多个常用生物途径资料库，包括 KEGG pathway maps³、Roche Biochemical Pathways⁴

¹ <https://www.vmh.life/>。

² <http://www.celldesigner.org/>。

³ https://www.genome.jp/kegg-bin/show_pathway?map01100。

⁴ <http://biochemical-pathways.com/#/map/1>。

等，研发了生物途径网络资料库最常使用的正交布局可视化方法，并将此方法延伸到人体代谢网络。我们将生物途径网络模拟成城市路线图，并依照城市布局模式划分适合的领域，绘制该分类下的子网络。这是因为生物途径网络同城市网一样，需要表现网络的阶层结构，例如一个化学物质是属于生物途径本体论(ontology)的哪一个分类，或者空间上属于哪一个细胞室。我们将生物途径类比为一个城市，之间的关系则以高速公路的方式表示。图1为丙氨酸和天冬氨酸代谢、生物碱合成以及雄激素和雌激素合成和代谢的生物途径图。其中图1(a)为传统应力主导图的可视化结果，图1(b)为类正交布局生物途径的可视化结果^[8]。

为达到平衡分布化学物质节点的目的，此算法首先采用布图规划算法(floorplanning algorithm)来切割图面，并使用混整数规划算法(mixed integer programming)来寻找最佳分割结果。然后将分割的小图用类人类正交布局算法(human-like orthogonal layout algorithm)绘制正交布局。最后连结分布在不同分类下的同一化学物质以强调此特性。为了建立一个联络网，并找寻两点间的最短路径，此算法额外考虑了通过每条路径的有向边数量，尽量做到均匀分配。此算法以构建拼图的方式切割大图并重新

整合成平衡的生物途径图，达到可以自动绘制人体代谢网络图⁵的目的^[9]。

复杂生物结构信息可视化

在现代生物学和化学领域，科学家必须收集和处理复杂且庞大的数据，模拟自然界化学物质的运作方式。除了表示抽象关系的二维生物途径图外，蛋白质三维结构的可视化可为科学家提供另一种建模方式，其目标是通过开发可视化技术对统计学方法、湿实验室(wet lab)以及可视化数据分析进行补充，从而支持分子生物学和显型(phenotype)分类的生物学研究。我们开发的系统帮助领域专家查看与他们研究的结构相关的关键信息，例如纳米级分子中的隧道或宏观放大尺度上的根形状分布。

同样重要的是可视化的结果，例如图片或者影片，以直观的方式向社会大众解释了这些研究现象在化学或者生物学上的作用，以及生物学家如何创建大型三维结构模型来反映某种生物体的最新知识。以视觉形式获取知识，加上与交互式可视化进行有效沟通，为科学家在空间和时间尺度的复杂交互作用下理解多种问题提供了便利。采用我们的可视化技术，使用者可以无缝地探索

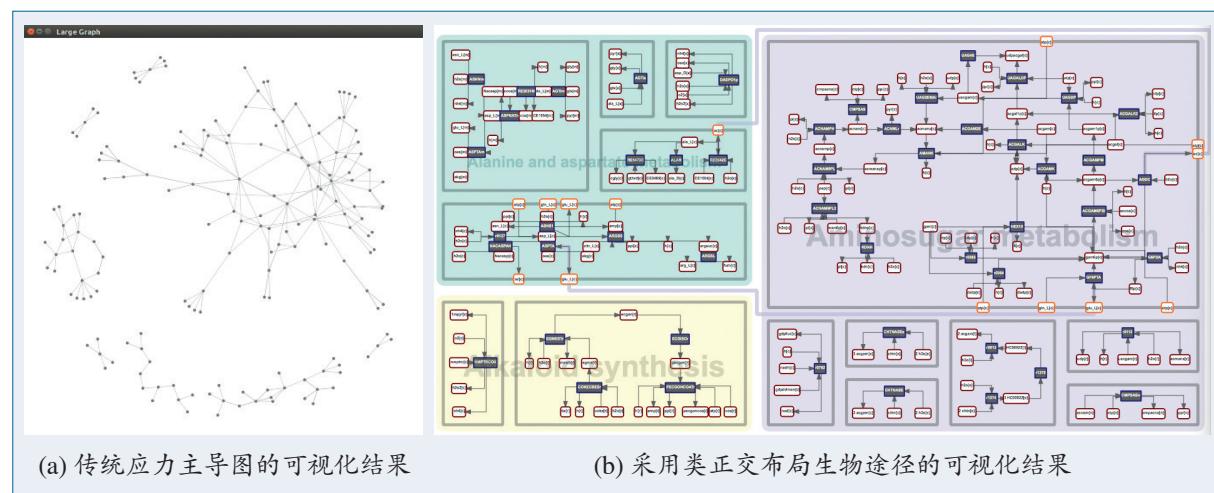


图1 自动生成的生物途径图范例

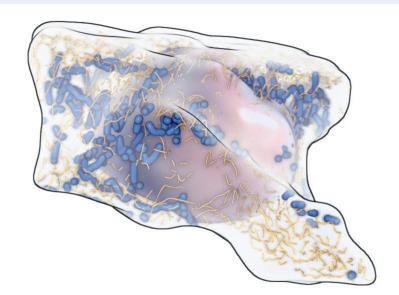
⁵ 演算法程序以及人体代谢网路图结果可参考作者 GitHub: <https://github.com/yun-vis/Metabopolis>。

不同的缩放级别，借此观察大至整个细菌的结构，小至细菌原子的细节^[2, 4, 5, 7, 10]。

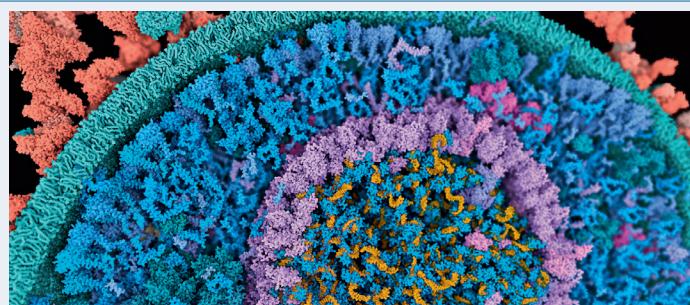
为了达到这个目的，我们通过开发新的可视化系统架构来促进生物学信息沟通，从而支持不同用户设备上异构生物数据的教学^[10]。多数情况下，依据科学家对异质数据的解释和应用的理解，生物学中的发现和概念通常以手绘插图的形式来表示。然而手动绘制插图耗时较长，而且不易与新产生的测量数据兼容。再者，插图通常是静态的，研究员希望当插图更新时，可以直接被客户端拿来使用。我们的系统旨在克服这三个障碍。它支持异构数据集的集成，反映从生物学的不同数据源获得的知识。系统对数据集进行预处理之后，将数据集转换为受科学插图启发的可视化表示（图2(a)）。与传统科学插图不同的是，这些插图的可视化是实时生成的，即它们是可互动的。通过系统标准化，该系统生成可视化的代码可以嵌入各种软件环境中，使之适用于不同的硬件环境，包括多GPU设置。

我们还研发了首个在交互式视觉环境中可以对

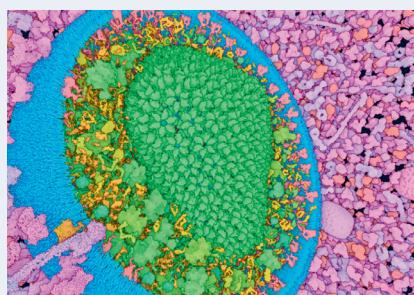
生物中尺度 (mesoscale) 进行自动综合结构构建的方法^[4]。图2(b,c) 表示了人类免疫缺陷 (HIV) 病毒的可视化结果。这些复杂模型可包含多达数百万个精确的原子结构、位置及其相互作用的分子。传统的方法只允许科学家在非视觉和非交互环境中多次尝试学习模型的构造，而我们的解决方案是将建模和可视化两方面结合在一起，实现了大部分单元的原子分辨率中尺度模型的交互式构建。我们提出了一套新的 GPU 算法，为快速构建复杂的生物结构奠定了基础。由于这些生物结构由多个膜封闭的隔室组成，包括可溶性分子和纤维结构，使用立体像素化 (volume voxelization) 三角网格 (triangulated meshes) 来定义细胞隔室 (cell compartment)。关于细胞膜 (cell membranes)，我们扩展了王氏砌瓦 (Wang tiles) 概念，将脂质 (lipids) 有系统地分布填充来构建脂双层 (lipid bilayer, 一种细胞膜的结构)，再使用荷顿序列 (Halton sequence) 将可溶性分子填充在细胞隔室内，以达到均衡分布的效果。至于纤维结构 (fibrous structures)，如核糖核酸 (RNA) 或肌动蛋白丝 (actin filaments)，



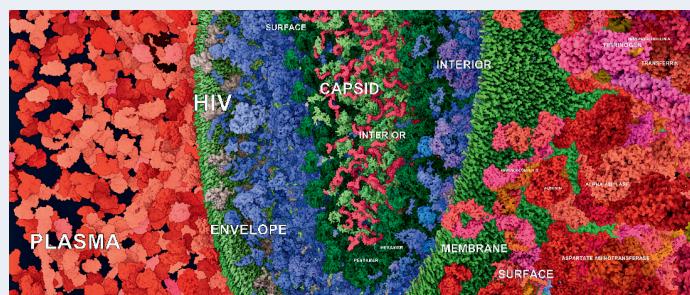
(a) 艺术导向的细胞呈现^[10]



(b) HIV 数据集，大约有数千个结构副本 (实例)，每个副本为 60 种独特几何形状的一种^[4]



(c) 不同视角的 HIV 数据集^[4]



(d) HIV 数据集的多维尺度和多重实例标记^[7]

图 2 复杂生物结构信息的可视化实例

是通过自回避随机游走 (self-avoiding random walk) 算法产生的。为了解决算法所导致的分子重叠，通过建立应力模拟系统去除非期望的重叠。

为了有效理解上述复杂环境和模型，我们研发了新的自动文字标记 (text labeling) 算法来辅助探索复杂模型的重要内在结构^[7]。我们提出了一种交互式标记复杂 3D 场景的方法，其中包含可以同时标记跨越多维尺度 (multi-scale) 以及其他多重实例 (multi-instance) 的场景。与传统的标记方法相比，我们为具有许多类型的多重实例物件选择场景中对应的尺度，提供配置的标签，使其有机会选择最合适的角度。此外，我们的算法超越传统医学建立了三维可视化、制图、书籍所使用的标记技术。与传统技术相比，新技术可以在适当时机调整标记对象的尺度，并通过仅标记每种类型对象的代表性实例以及有效策略来减少视觉杂乱现象。该技术通过分析场景的深度缓冲区 (depth buffer) 和场景对象的层次树来选择适当的标签级别。我们采用改编的图形设计视觉层次结构概念来解决标签之间的层级关系，并通过遍历此树状结构来沟通呈现标记的主题。系统在选择代表性实例时会考虑数据特征，量身定制选择标准，与贪婪优化方法相结合来找出最适合场景的维度以及标记实例。在图 2(d) 中，我们用中尺度生物学模型证明了我们的方法在复杂环境下是可行的。前景是自动选择的蛋白质类型的代表性实例，而背景用整个隔室的名称标记，用来代表更高层次的层次结构。

多尺度视觉探勘及视觉抽象化

生物领域相关的大数据除了涉及高度空间复杂度之外，更多时候数据本身同时跨越多个空间和时间尺度，导致产生相当大的视觉复杂性。有效率的多尺度分子可视化技术可以处理这些庞大数据，进而为这些数据的可视化分析、抽象化和交互式探索提供技术^[11]。多尺度建模的可视化算法可以最大化地支持复杂生物结构的设计和分析。例如，科学家现在能够使用脱氧核糖核酸 (DNA) 作为主要建构材料来制造纳米尺度的机器人。这个概念源自对 DNA

长链骨架的有效控制和编程，而此骨架是由四个彼此互补的核碱基相接组成的序列。世界各地的研究人员已经可以对 DNA 序列进行编程，控制 DNA 长链骨架自行组装成纳米尺度的物体，如图 3 所示的立方体结构。

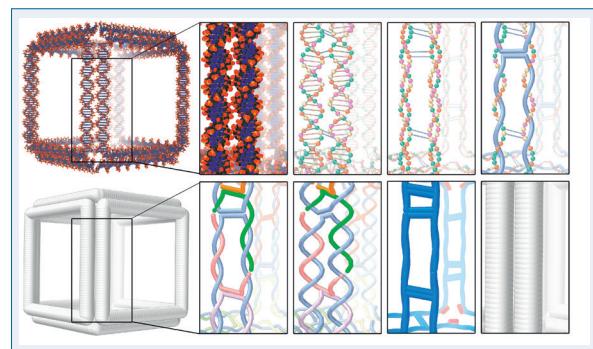


图3 这个纳米尺度立方体由 DNA 组成，使用我们的方法可以让它在多个尺度上可视化^[12]。算法描绘所有原子和骨架细节，而此立方体持续抽象化直到由单一管线 (simple tube) 代表整个 DNA 双螺旋结构

我们的多尺度可视化方法可以表示各种尺度的 DNA 纳米结构，从原子细节到目标几何结构（如图 3 所示的立方体）^[12]。使用者能够直接移动拖动条 (slider bar) 来更改 DNA 尺度设定，进而调整适合其任务的抽象尺度级别。此算法的目的在于定义并建立离散尺度之间的一致性，达到科学家们可以在尺度之间进行无缝视觉探勘的目的。利用这种方法，生物学家能够处理日益复杂的 DNA 结构，并通过适当的复杂物件抽象化来加速及实现分析。

我们还研究了如何将 DNA 纳米结构中的多种布局和表示方式统一整合到一个抽象空间中^[13]。在该空间中，用户可以平滑无缝地探勘并过渡到任何期望的尺度表示，并且可以对纳米结构进行大量的编辑修改。我们的整体方案无缝结合了三维仿真的结构模型、二维图解表示以及有次序的一维排列。在图 4 中，纳米管线的拓扑结构首先被改变，然后新的 DNA 骨架被添加进来以改变该结构的功能。每个尺度表示法强调了数据的不同方面。科学家亦可以交互调整可视化并编辑不同尺度下的表示结果，系统会自动将效果传达到其他尺度和维度。

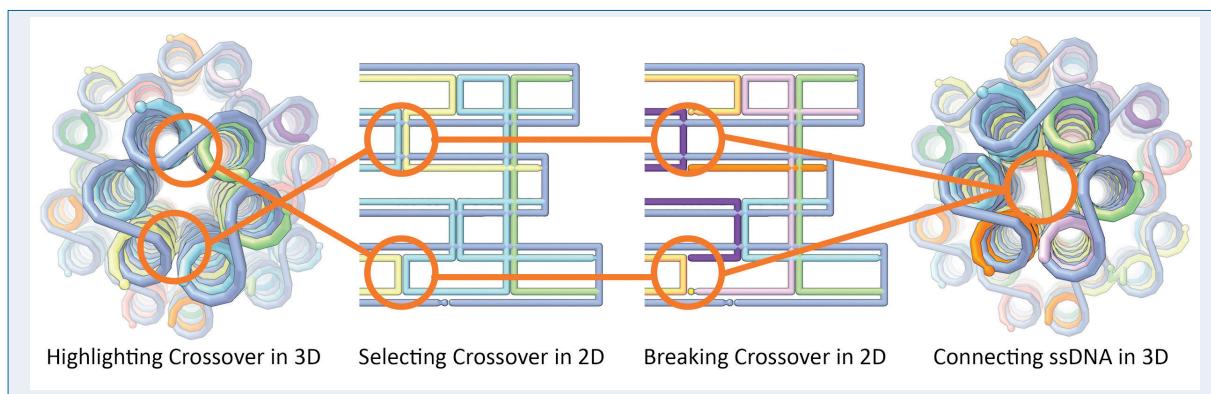


图4 一个可以在三维或二维中修改纳米管线的实例^[13]

未来探索

我们通过开发交互式可视化系统对庞大的生物信息进行建模和分析，促进生物知识的发展与传播。同时通过可视化整合各种生物学的模型架构，通过可视分析来降低科学家做研究时遇到资料转换造成误解的可能性。我们对该可视化的潜力感到兴奋，因为我们的最终目标是在单一系统框架中集成人体细胞的所有可用数据，进而连接不同资料库，方便使用者做分析。

另一个令人兴奋的探索方向是直观地表示动态结构模型，而不是当前的静态结构模型。我们期望将来能够在引人入胜的三维视觉环境中对细胞生物进行建模、可视化、分析和交流。这种细胞生物学将作为计算实验平台，其中生物体信息将进入可以研究和互动的计算机模拟的生命周期。这种技术将对生物科学如何向大众传播产生巨大的影响，通过有机体的计算机视觉模型，科学家可以在科学中心合成有机体及其生活模式，并将结果直接呈现在好奇的观众面前。通过这种方式，我们可以向年轻一代推广科学的主要成果，激发他们的好奇心，并以成为自然科学或计算科学研究人员作为自己的使命。 ■

致谢：

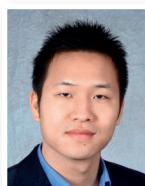
此项目获得了来自欧盟地平线 2020(European Union Horizon 2020) 研究和创新计划资金 Marie

Sklodowska-Curie Actions(MSCA) 计划协议第 747985 号，维也纳科技基金 (WWTF) 项目 VRG11-010，以及阿卜杜拉国王科技大学 (KAUST) 项目 BAS /1/1680-01-01 的支持。



吴湘筠 (Hsiang-Yun Wu)

维也纳科技大学 (TU Wien, Austria) 视觉计算与使用者中心技术研究所研究员。主要研究方向为信息可视化、可视分析。
hsiang.yun.wu@acm.org



缪海朝 (Haichao Miao)

维也纳科技大学 (TU Wien, Austria) 视觉计算与使用者中心技术研究所博士生。主要研究方向为科学可视化、可视分析。
miao@cg.tuwien.ac.at



伊万·维奥拉 (Ivan Viola)

阿卜杜拉国王科技大学 (King Abdullah University of Science and Technology(KAUST), Saudi Arabia) 视觉计算中心及维也纳科技大学 (TU Wien, Austria) 视觉计算与使用者中心技术研究所研究员。主要研究方向为科学可视化、可视分析、计算机图形学。
ivan.viola@kaust.edu.sa

参考文献

- [1] Graham T J, Autin L, Mostafa A, et al. cellPACK: a virtual mesoscope to model and visualize structural systems biology[J]. *Nature Methods*, 2015,12:85-91.

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

从流线到流面： 流场可视化新进展与展望

关键词：流场可视化 流面 流线

陶 钧
中山大学

流场可视化概述

随着当今计算性能的高速发展，计算流体力学 (computational fluid dynamics) 已经应用于众多科学、工程领域。对物理现象进行科学模拟，成为流体动力学研究人员理解物理现象的重要手段，其涵盖内容广泛，例如，大气科学中的气流、海洋学中的洋流、生物医药工程中的血流以及汽车工程中的风洞试验等。这些模拟流场数据呈现大规模、高精度、结构复杂、特征多样化等特点，使得其中包含的大量信息难以被归纳为简单的数字，而需要通过可视化手段进行观察、分析、理解。

在过去三十年间，流场可视化 (flow visualization) 一直是科学可视化 (scientific visualization) 中的一个核心研究领域，其主要研究目标为利用计算机图形学手段将流场数据渲染绘制成图像，使相关领域专家能通过视觉系统观察和认识流场。如图 1 所示，流场可视化是从模拟计算模型到产生关键认知必不可少的一环，也是领域专家与数据交互的重要手段。领域专家通过交互对可视化结果做出进一步修改，从而不断加深其对数据的认识，最终达到验证科学假设，理解物理现象，乃至产生新的科学发现的目的。例如，在图 2(a) 中，流场可视化通过流线 (streamline) 描述了血管中血流的流态，血管瘤中螺旋状的流线展示了涡流 (vortex) 与血管

瘤之间的关系^[1]。在图 2(b) 中，通过流线和流面 (stream surface) 共同展示龙卷风的形态。流面描述了龙卷风中心的流态，而流线显示了其下方周边区域的流态^[2]。

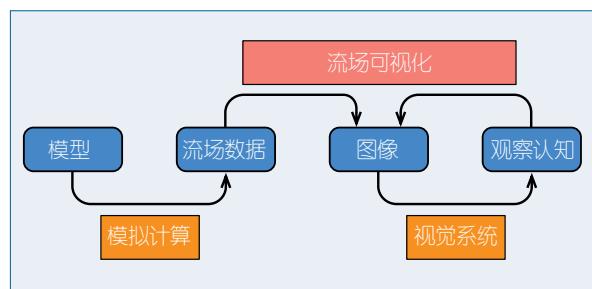


图 1 流场可视化在流体力学研究中的作用

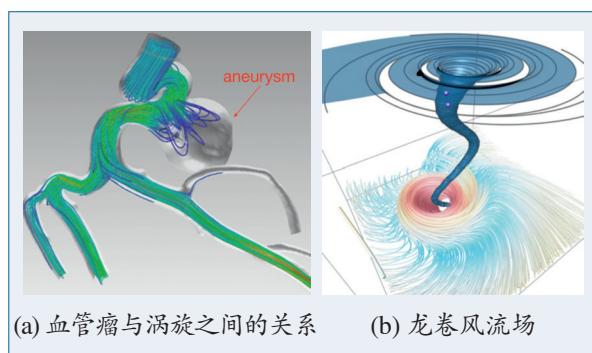


图 2 流场可视化应用

为了使科研人员能清晰地观察流场，流场可视

化需要准确描述流场的有效信息，如流态和流场特征。流场可视化方法的比较^[3]中提出了观察流场中对于给定位置的粒子的三个基本任务：预测其在流场中的去向、定位流场中临界点的位置以及识别临界点的类型。随着流场可视化的发展，其覆盖范围也从临界点扩展至更广泛的特征，如涡流、分离线 (separation line)、附着线 (attachment line) 以及拓扑结构 (topology structure) 等。而通过可视化完成的任务也随之拓展，但基本任务仍然可以归结为观察流态、定位流场特征、识别特征类型以及发掘特征之间的联系这几类。

根据其表现形式，流场可视化可以分为基于积分 (integration-based)^[4]、基于图元 (glyph-based)、基于纹理 (texture-based)^[5]、基于图示 (illustration-based)^[6] 等几大类。其中，基于积分的方法是在流场中通过粒子追踪 (particle tracing) 产生积分曲线 (integral curve) 或曲面 (surface) 来表现流场信息，这是当前最常用的表现形式。流线与流面具有连续性强以及在流场空间中较为稀疏的特点，因而更适用于展示三维流场。

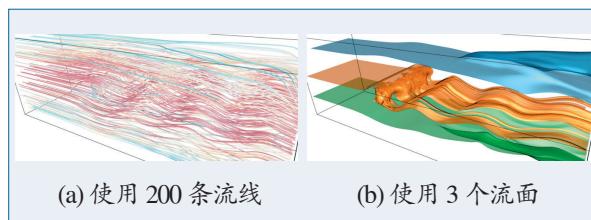


图 3 Square cylinder 流场数据可视化

有效的流线 (面) 可视化应生成有意义的流线 (面)，捕捉流场信息并以适当方式显示流线 (面)，从而向用户传递其中信息。产生的流线 (面) 过于稀疏，则无法完整捕捉信息；而过于密集，在显示过程中则会导致过多的互相遮挡，使部分信息无法被用户观察到。同时，过多的流线 (面) 也可能导致关键特征及其联系淹没在背景信息中，而无法有效地被视觉系统所提取。理想的可视化方法应保证流线 (面) 覆盖关键特征区域并减少对这些区域的遮挡，从而传递其包含的信息。

基于流面的可视化

作为空间中二维连续 (2D continuous) 的曲面，流面比一维连续 (1D continuous) 的流线整体性更强，因而能表达许多流线无法传递的信息，如流体的发散、汇合、扭曲、折叠等较为复杂的现象。图 3 展示了使用基于流线与基于流面的可视化方法绘制 square cylinder 数据的效果对比。图 3(b) 中基于流面的方法更准确地描述了三条直线上的粒子在顺流绕过方柱体后所产生的形变，即从图左侧的平面变形为右侧的波浪状曲面。而图 3(a) 中的流线则只表示了粒子的轨迹，无法描述种子线在流场作用下产生的形变。同时，由于使用的流面数量远比流线数量更少，图 3(b) 在视觉效果上也显得更为清晰。

本文侧重于描述代表性方法及其思路，而不追求方法数量上的完整。为获取更全面的信息，可参考 Edmunds 等关于流面可视化的综述文章^[7]。

流面构建

与通过追踪种子点 (seed) 产生流线相似，流面的构建是通过在流场中追踪种子线 (seeding curve) 上的多个粒子并将其轨迹连接成面而完成的。依据构成流面的基本要素，流面构建方法可以分为基于三角网格 (triangle-based)、基于四边形网格 (quad-based) 以及基于点 (point-based) 三类。

基于三角网格的方法^[8]是最早的流面构建方法之一。该方法通过追踪种子线上的采样粒子，来推进前沿 (advancing front) 产生流面。在推进过程中，该方法需要维持前沿上采样粒子的密度稳定，从而产生大小相近的三角形。当流场发散而采样粒子逐渐稀疏时，需要增加采样；反之，当流场汇合而采样粒子变得密集时，则需要减少其数量。该方法将相邻的采样粒子所产生的流线上的点连接为三角形形成了流带 (stream ribbon)，并进一步连接相邻的流带形成流面。该方法同时提出了一个将流线之间连接线长度最小化的贪婪算法，以求在流带上产生形状较为规整的三角形。

基于四边形网格的流面构建方法则直接利用交

织的流线与时线 (timeline) 产生四边形，以降低计算复杂度。然而，在流体出现旋转特征时，由于四边形连接方式的唯一性，这一思路无法通过改变连接方式减少其剪切形变 (shearing)。Easy Integral Surfaces 方法^[9] 通过计算流线的旋转角度并据此调整流线上采样点的密度，从而减少四边形的形变。该方法也在流体发散或汇合时调整前沿上采样粒子的数量，以保持四边形大小的稳定。

基于点的流面构建方法^[10] 则消除了粒子追踪过程中点的连接，进一步降低了构建开销。该方法利用 GPU 同时追踪大量粒子，并将每个粒子显示为一个带光照的子图 (sprite)，并将所有子图拼接在一起形成完整的流面。然而，该方法必须维持足够多的粒子以保证子图足够密集，以实现无缝连接。与上述两种方法类似，该方法也需要根据流场的发散或汇合增删粒子以维持稳定的显示结果。

上述三类方法都通过在流场中追踪粒子直接构建流面，其面临的共同问题为如何维持粒子的密度。而间接构建流面的方法^[11, 12]，则首先根据粒子追踪产生标量场 (scalar field)，再从中通过等势面提取 (isosurface extraction) 产生流面，从而避免了维持粒子密度以及将流线拼接成流面等问题。

流面放置及选择

好的流面应该能准确描述流场信息，这需要满足两方面的条件：其一，流面应该覆盖关键区域，如重要的流态及流场特征；其二，隐含于流面形状中的信息应该能为人类视觉系统所感知。为产生理想的流面，大致有两类方法：一类为流面放置 (placement)，即在合适的位置依据一定规则产生符合要求的种子线；另一类为流面选择 (selection)，即在大量产生的流面中选择具有代表性的进行渲染，从而得到理想的显示结果。

流面放置 流面放置方法可分为自动和手动两类。自动方法通过对流场进行分析，确定种子线放置位置，或通过对产生的流面进行分析，优化种子线。自动流面种子放置方法^[13] 对流场聚类得到种子线的放置位置，并由曲率 (curvature) 方向决定种子线方

向。曲率方向垂直于流面，能将种子线的有效长度最大化。同时，曲率方向平行于流体的旋转轴，产生的流面更利于用户观察。然而，该方法只能得到沿曲率方向分布的种子线，而无法保证流面的整体质量。全局流面优化方法^[14] 提出了一个流面质量的评价方式，以此判断产生的流面是否理想。该流面质量评价的核心想法是度量流面主曲率 (principal curvature) 方向是否与流向一致。该方法利用模拟退火算法 (simulated annealing) 优化种子线，从而提升流面质量。然而，该方法没有考虑流面之间的关系，只能用于产生单个最优流面。全局多流面优化方法^[15] 则通过距离场 (distance field) 确保产生的流面之间的距离大于设定的阈值，将原方法扩展到了多个流面的生成上。由此产生的流面既能覆盖不同区域，又能减少视觉上的重叠。

自动放置流面的缺点是难以根据用户的需要进行调整。在当前应用和研究方向日趋多样化的背景下，同样的可视化结果未必能满足不同用户的需求。因此，提供交互手段，使用户能手动放置种子线，对产生符合其需要的流面有着重要的现实意义。最早的手动放置种子线方法由用户手动选择空间中的两个点作为种子线的起始点产生流面。然而，这一方法建立在试错法 (trial-and-error) 之上，需要用户进行大量的尝试，才能逐渐找到其感兴趣的流面。同时，由于直线无法完全与流场方向切合，其产生的流面往往不能准确描述流场信息。

基于用户手绘的半自动流面产生方法^[2] 让用户通过交互指明种子线从而产生流面。该系统比照用户熟悉的画图工具，提供铅笔、橡皮、笔刷等工具，在密集的采样流线上手绘产生流面。铅笔工具允许用户在流线上绘制种子线，并调整其宽度；橡皮工具允许用户去除不必要的采样流线；而笔刷工具使用户能通过交互改变流面渲染样式，提供所见即所得的体验。

图 4 描述了使用该系统产生流面的一个实例。图 4(b) 中，用户在密集的采样流线上绘制种子线 (红色曲线)，并生成示意流线 (灰色曲线)。示意流线预示生成流面的形状，使用户避免试错法所带来的

大量人工操作。用户可进一步调整种子线的起始点以调整生成流面的宽度(见图4(c))。在用户确认生成流面后,系统会在采样流面中将与其相似的流线去除(见图4(d))。重复此过程,用户可以由外至内剥开流场,并产生相应流面。图4(e)显示了在five critical points数据中通过用户手绘产生的所有流面。我们能清晰地看到五个随机放置的临界点,以及由于其相互作用而产生的两个额外的临界点。同时,用户可以通过笔刷工具,在流向不明确的流面上添加流线显示,从而提供更准确的流场信息(如图4(e)中棕色、绿色、红色及紫色流面)。

该系统将二维的用户手绘曲线转换成三维种子线时,使种子线遵循副法向量方向(binormal direction)。而当所有时线都沿副法向量方向时,其流面主曲率与流向一致。这意味着,其产生流面方式与自动流面放置^[13]及全局流面优化^[14]中的选择评价标准是一致的。

流面选择 与众多围绕流面放置开展的研究相

比,迄今为止流面选择的相关工作还较为少见。主要原因在于,种子线具有极高的自由度,即使在同样的空间位置放置种子线,其长度、方向及形状不同也可能产生差异很大的流面。这决定了在实践中,我们很难产生一个有效的采样集囊括所有流面,更难保证采样集的结构与流场本身相同。

FlowNet方法^[16]使用深度学习框架进行流线与流面的选择。由于无法产生一个完备的流面采样集,该方法试图在质量较好的流面中进行选择。因而,该方法沿用流面评价标准^[2, 13, 14],在流面采样时将种子线方向限定为流线的副法向量方向。这样,种子线及其产生的流面就可以由种子线的起点及长度唯一决定。

FlowNet对流面进行聚类,并以聚类中心作为代表性流面进行显示。其聚类过程分为三步:首先,FlowNet采用自编码器(auto-encoder)架构将流线编码为向量;然后,使用t-SNE算法将向量降维形成二维图布局(见图5(a)与(c));最后,在二

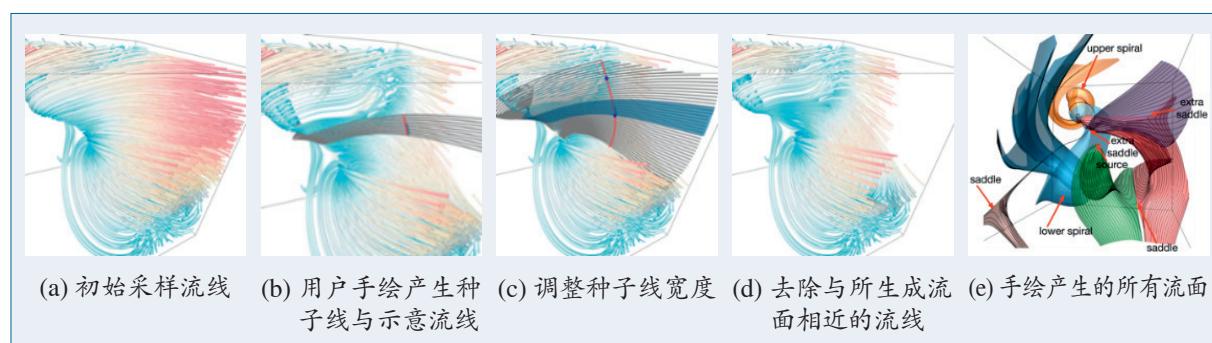


图4 用户手绘方法^[2]产生five critical points数据流面的过程及结果

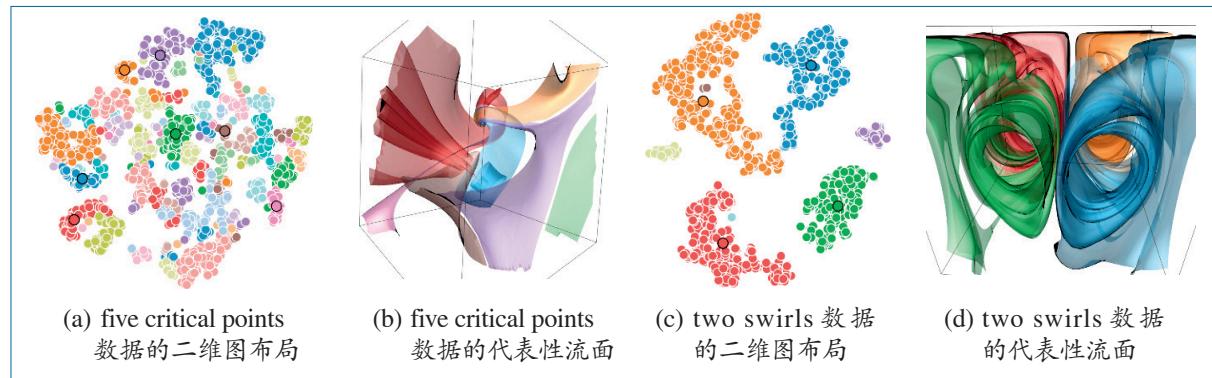


图5 FlowNet流面选择(图布局中由黑色边框高亮显示的为代表性流面)

维图布局上使用 DBSCAN 算法进行聚类。降维后向量中包含的噪声减少，因而更能反映数据本身的结构。

图 5 展示了使用 FlowNet 对两个数据集中采样流面降维后得到的图布局及从中选择的代表性流面。在图 5(a) 中，我们可以看到，由于 five critical points 流场中所包含的流态较为多样化，其采样流面形成的图布局的结构也比较发散。如图 5(b) 所示，基于该图布局选择的流面代表了不同特征的流态，这与基于手绘的方法^[2] 手动产生的流面（见图 4(e)）是相似的。而 two swirls 的图布局则较为对称并呈现出四个大聚类加两个小聚类的清晰结构（见图 5(c)），这与流场本身结构是一致的。

流面渲染 (Rendering)

相对流线而言，流面本身覆盖面更大，结构更复杂，其渲染过程中需要考虑的因素也更多。其中有两个最关键的问题：其一，如何决定流面的透明度，以减少观察中的遮挡；其二，如何增强用户视觉对流面上流向的感知。

为了使关键流态及特征在渲染结果中更清晰，其渲染过程中通常需要增加流面中关键部分的不透明度，并降低其他区域的透明度。IRIS 方法^[17] 利用渲染过程中的深度信息计算法向量变化 (normal variation)，从而高效地在屏幕空间中估测流面曲率，并据此来调整流面上每一点的透明度。流态简单的区域由于曲率较低会变得更透明，使复杂区域更易于观察。然而，该方法并没有考虑流面之间的遮挡关系。即使流态简单的区域不遮挡关键区域，其透明度也会降低。这可能导致流场信息丢失及削弱用户对流面空间关系的感知。曲面透明度优化方法^[18] 将流面划分为小块，并将每一小块的透明度及其相互间的遮挡系数构建成一个线性系统，通过求解线性系统得出每一小块的透明度。该方法能依据视角变化实时调整流面每个部分的透明度，在确保关键区域不被遮挡的情况下，提供尽可能多的额外信息。而增强对流向的感知则主要依赖于在流面上增加纹理，如流线^[2, 17] 及 LIC^[19, 20] 等。

挑战与机遇

当前高性能模拟计算的发展极大地提高了流场数据的规模及精度，给可视化方法归纳总结并清晰展示海量数据的能力带来了新的挑战，也对其数据处理性能提出了更高的要求。同时，模拟数据的复杂性、应用的多样性也要求可视化方法支持更为多样化、定制化的任务。人工智能，尤其是深度学习，在众多科学工程问题上取得的巨大成功，也给流场可视化提供了新工具。如何利用这些新工具解决流场可视化中的问题，以及如何通过可视化交互手段将计算机智能与其领域专家知识相结合，都将成为今后研究的重点。

表现形式全局化 流场可视化发展过程中，其表现形式日趋全局化。从基于纹理的 LIC 方法（沿流场方向对局部进行卷积），到流线（完整描绘种子点轨迹），再到基于流面的方法（完整描绘种子线随流场的运动），其表达信息的载体的数量越来越少，而单个载体在空间中的连续性越来越强、表现也越来越全局化。表现形式的全局化极大地增强了其对复杂物理现象的描述能力，而载体数量的减少为当前海量数据中简化可视化结果，从而更清晰地描述流场结构带来的机遇，也为多模态下的比较提供了工具。然而，与较为成熟的基于流线的方法相比，流面的研究还相对滞后。如何更好地选择流面、呈现流面，如何依据数据特征结合流线与流面以发挥各自优点等，都有待可视化研究人员的进一步探索。

分析方法智能化 人工智能方法，如深度学习，为流场可视化的发展提供了新的可能性。在 FlowNet^[16] 中，通过 autoencoder 所产生的特征向量，比传统方法通过计算流线（面）间距离，在流线（面）的选择任务上取得的效果更好。这可能是由于深度神经网络在从流线（面）到特征向量间的变换中所包含的信息，远比基于点定义的流线（面）距离更丰富。但这些信息却无法被现有方法理解、利用，而领域专家所具备的流场知识，也没有被有效纳入深度学习的网络中。因此，利用深度学习等工具突

破流场可视化中传统任务的同时，将机器学习对流场的解读以及领域知识中的解释通过可视化结合，也将为未来科学的研究中理解流场提供新的方法。

交互手段丰富化 由于流场可视化在不同研究领域工程应用中的广泛使用，用户感兴趣的特征及需要挖掘的关系也在不断细化。这一方面促进了针对特定领域、任务而定制的可视化方法的发展，另一方面也催生了大量通过交互手段使用户有效探索特定流态、特征的方法。交互式方法定制的是可视化结果而非可视化系统，其有效地避免了针对特定问题开发可视化系统的高成本，同时通过用户输入产生符合其需要的特定结果。现有方法包括基于手绘^[21]和文字^[22]的流线段查询方式，以及基于图的流线可视探索^[23, 24]等。然而，这些方法大多基于单个流场中的流线形状或流线与空间的关系。在未来的研究中，可视化领域仍需发展更丰富的交互方法，使用户能自由探索流线及流面与物理现象之间的关联，乃至多个流场间流线（面）之间的关系。 ■



陶 钧

中山大学数据科学与计算机学院副教授，博士生导师，入选中山大学“百人计划”。主要研究方向为科学可视化，特别是信息论、优化方法和交互探索方法在流场可视化中的应用。taoj23@mail.sysu.edu.cn

参考文献

- [1] Tao J , Huang X , Qiu F , et al. VesselMap: A web interface to explore multivariate vascular data[J]. *Computers & Graphics*, 2016, 59:79-92.
- [2] Tao J , Wang C . Semi-Automatic generation of stream surfaces via sketching[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2018, 24(9):2622-2635.
- [3] Laidlaw D H , Kirby R M , Jackson C D , et al. Comparing 2D vector field visualization methods: A user study[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2005, 11(1):59-70.
- [4] McLoughlin T , Laramee R S , Peikert R , et al. Over two decades of integration-based, geometric flow visualization[J]. *Computer Graphics Forum*, 2010, 29(6):1807-1829.

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

《计算机科学技术名词（第三版）》出版



由全国科学技术名词审定委员会第三届计算机科学技术名词审定委员会（简称分委员会，与CCF计算机术语审定工作委员会为同一套工作班子）主持审定的《计算机科学技术名词（第三版）》于2018年12月由科学出版社正式出版。

《计算机科学技术名词（第三版）》共计12章，收集了8795个计算机领域的专业术语，全面覆盖了理论计算机科学、计算机体系结构、计算机硬件、系统软件、软件工程、数据库、人工智能、网络与数据通信、信息安全、计算机应用、计算机交叉学科等研究领域，共51万余字。

《计算机科学技术名词（第三版）》相较于之前版本，最大的特点是给每个人选术语撰写了简要的释文。分委员会组织相关领域的专家，对术语和释文按照二级学科进行分类、分层的细致撰写和审定，通过广泛讨论和征集意见，逐渐在专家内部取得共识。金怡濂、林惠民和梅宏三位院士对字典进行了终审。200多位审定专家及很多“不在编”的专家参与了编撰审校工作。

基于可视分析的可解释深度学习

姜 流¹ 刘世霞¹ 雷 娜²

¹ 清华大学

² 大连理工大学

关键词：可视分析 深度学习 交互式机器学习 可解释机器学习

可解释深度学习的兴起

新一代人工智能系统具有自主理解、自主学习和自主行动等特点，但是系统所采用的全自动人工智能模型难以被人类理解，通常被当作一个黑盒子使用。在实际应用中，用户试图根据模型的性能（如准确度、复杂度等）来试探地选择和改进学习模型。由于用户无法理解模型内部的工作机理，所以无法实现人机的平等双向沟通，给人机协作带来巨大障碍。能否进行有效的人机沟通、准确理解、信任和管理这些“类人”机器及相应的学习模型，将直接影响这些“类人”机器的发展趋势——是成为人类的“朋友”还是“敌人”。这迫切需要提升机器学习模型的可解释性——将“黑箱”模型转化为“白箱”，为实现人机平等沟通奠定基础。

可解释深度学习作为新一代人工智能技术，已经引起欧盟、美国、中国等各国政府的高度重视。哈佛大学、斯坦福大学、麻省理工学院、谷歌、微软、Facebook 等机构在机器学习的可解释性方面也取得了显著进展，开发出 TensorBoard 等通用的可解释机器学习平台和工具，旨在揭示人工智能系统某个具体行为背后的逻辑。近年来，人工智能和机器学习领域的顶级国际会议 NeurIPS、ICML 和 IJCAI 等都纷纷设立了关于可解释机器学习的专题研讨会，吸引了领域内大量研究者参与讨论^[1, 2]。其中 ICML 2017 将会议的最佳论文授予了“Understanding Black-box Predictions via Influence Functions”^[3]。

为什么需要可解释深度学习

随着人类对深度学习技术依赖的加深，模型决策的可解释性及其内部过程的可控性对深度学习技术的发展及应用愈发影响深远。由于深度学习模型的工作机理复杂，研究人员难以直观地进行模型设计、调试和语义上的分析，往往依赖于费时费力的试错过程，对模型参数进行频繁调整，效率低下。因此，迫切需要可解释的深度学习系统，帮助他们更好地理解和分析学习模型，从而快速设计出符合需求的模型。在实际应用中，特别是涉及到高影响和高风险的任务（如精准医学、执法、金融投资和自动驾驶），深度学习的可解释性对于理解和信任模型的决策至关重要。在此背景下，各国政府高度关注深度学习的可解释性，比如 2016 年 4 月，欧盟立法规定人类有权要求对机器产生的决策作出解释^[4]；《欧盟人工智能》也确立以人为本的欧洲战略，把可解释人工智能提高到人工智能价值观层面，确保其朝着有益于个人和社会的方向发展；2016 年 8 月，美国国防部高级研究计划署 (DARPA) 启动一项名为可解释人工智能 (XAI) 的大型项目，并发布了关于 XAI 的征询建议书^[5]；我国在《新一代人工智能发展规划》中也明确将“实现具备高可解释性、强泛化能力的人工智能”作为未来我国人工智能发展的首要突破口。根据以上讨论，本文聚焦的深度学习的可解释性主要是指模型在给出预测结果的同时可以提供相应的原因，使研究人员或者从业者更好地地理

解机器的决策过程。特别地，深度学习的可解释性与预测性能往往不可兼得，如何根据应用任务在两者之间做合理的折中也是一个关键的研究问题。

为了提高深度学习的可解释性，交互式可视化发挥着关键作用，并且正成为一个活跃的前沿研究领域。许多可视分析工作围绕着模型理解、诊断和改进的任务，紧密地结合人的视觉感知能力和自动算法的计算能力，对深度学习的可解释性进行探索和分析^[6~8]。可视化领域最具影响力的学术会议之一的 IEEE VIS 2017 将会议的最佳论文授予了 Google 研究人员基于 TensorFlow 的深度模型可视分析工作“Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow”^[9]。

可解释深度学习的分析框架

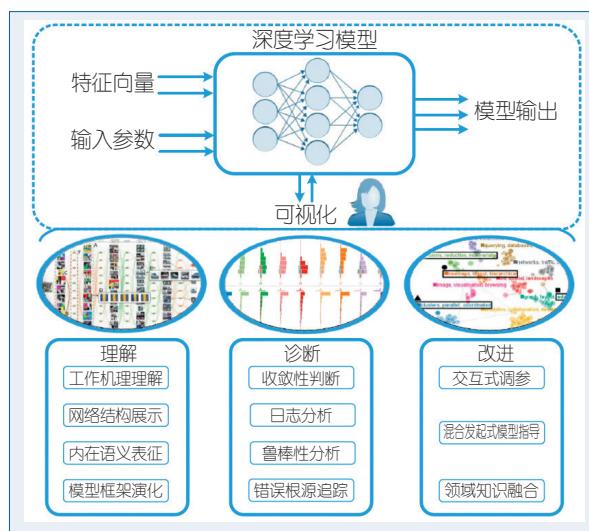


图 1 可解释深度学习分析框架

可解释深度学习的分析框架主要涵盖模型理解、诊断和改进三部分，如图 1 所示。模型理解包括工作机理解释、网络结构展示、内在语义表征与模型框架演化，旨在揭示模型预测背后的基本原理和深度学习模型内部的动态过程，并试图使这些复杂模型至少可被部分理解；模型诊断包括收敛性诊断、日志分析、鲁棒性分析与错误根源追踪，目的是识别和解决深度学习模型中的缺陷或者问题，比

如网络无法收敛或者性能无法满足要求；模型改进包括交互式调参、主动学习、混合发起式模型指导和领域知识融合，通过丰富的用户交互，以及半监督学习或主动学习，将专家知识和专业知识交互地融入深度学习模型，对其进行改进。

现有工作分析

模型理解

模型理解是利用可视分析技术，展示模型的工作过程，使研究人员更好地理解深度学习模型的工作机理。目前，理解深度学习模型的技术主要集中在基于点的技术和网络可视化方法。基于点的技术^[10,11] 利用散点图 (scatterplots) 来展现神经网络的主要组件，如神经元、学习得到的特征表示 (learned representations) 等之间的关系。数学上用高维向量表示学习到的特征。这个高维向量中的每一个元素 (entry) 代表了神经网络中隐藏层的一个神经元的输出。一般来说，每个组件由一个点来表示，这些点利用主成分分析 (Principal Component Analysis, PCA) 和 t-SNE (t-distributed Stochastic Neighbor Embedding) 等降维技术进行布局，使得起相似作用的组件对应的点在降维后坐标相近。基于点的技术可以帮助用户确认关于神经网络的假设，并且发现未知神经网络组件之间的关系。

图 2 展示了由罗贝尔 (Rauber) 等人提出的基于点的可视化技术^[11]。图中的每个点代表一个测试

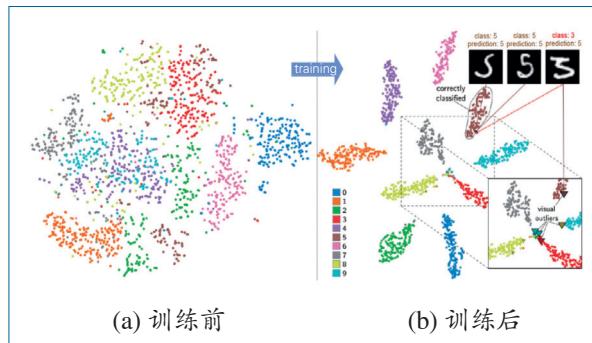


图 2 测试样本训练前后的特征表示

样本经过训练后的特征表示。每个点的颜色代表这个测试样本的类别标签。在进行训练以后，学习所得的特征表示能够更好地区分不同类别的样本。图 2(b)还可以帮助用户更好地理解被错误分类的样本。这些被错误分类的样本用三角形表示，很多样本都是视觉上的离群点，即周围有很多其他类别的样本点。另外，很多离群点对应的样本类别判断非常困难，人类视觉都难以对其进行分辨。例如，图 2(b)中数字“3”的图片之所以被分错，是因为它跟数字“5”极为相似。

尽管基于点的技术能够很好地表示大量神经网络组件之间的关系，但它们不能很好地反映网络的拓扑结构。因此，它们往往无法帮助用户深入地、系统地理解不同层的神经元所扮演的不同角色以及它们之间的关联（神经元之间的连边）。基于网络的技术^[6,12~14]通过对网络的拓扑结构可视化解决了这个问题。这些技术经常将神经网络表示成一个有向无环图（Directed Acyclic Graph, DAG），然后将网络中的重要信息通过 DAG 中点（或者边）的大小（粗细）、颜色以及可视化图标来表示。

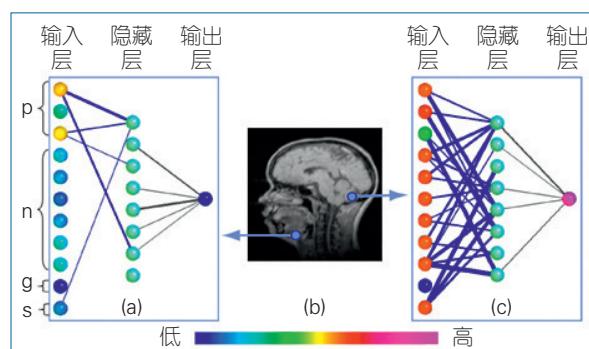


图 3 对头部体素是否属于脑内物质进行分类的神经网络拓扑结构

图 3 显示了一个早期的基于网络技术的可视化结果^[12]。该图展现的神经网络有一个隐藏层，目的是判断一个头部体素（voxel）是否为脑内物质。这里，每个体素用其标量值 s 、梯度大小 g 、邻居的标量值 n 及其位置 p 来表示。边的宽度表示每条边的重要性，输入、输出层点的颜色代表值的大小。从输出层点的颜色可以看出，网络可以正确地将左

边的体素分类成非脑物质（低输出值），将右边的体素分类成脑物质（高输出值）。从图 3(a) 和 3(c) 的网络拓扑结构可以看出，要将左边的体素分类成非脑物质，只需考虑它的位置；而要将右边的体素识别为脑内物质，则需要考虑除了梯度大小 g 以外的所有输入信息。

上述技术可以有效地将小规模（含有几十个神经元）的网络可视化。但是随着神经元和连接边的增加，可视化会出现视觉混乱，使得用户难以理解网络的结构^[15]。为了解决这个问题，Liu 等人开发了 CNNVis^[7] 可视分析系统（如图 4 所示），帮助机器学习专家理解和诊断具有数千个神经元和数百万个连接的深度卷积神经网络（CNN）。为了显示大规模的 CNN，CNNVis 将层和神经元进行了聚类，并使用基于双聚类的算法来捆绑边，以减少视觉混乱。此外，CNNVis 使用分层矩形打包算法将每个神经元聚类所学到的特征可视化，以及矩阵重排序算法揭示神经元的激活模式，从而实现从多个角度分析神经元。受 CNNVis 工作的启发，为了帮助用户直观地理解模型结构和输出结果，近年来涌现了许多对深度神经网络进行交互式可视化的工作。比如 ActiVis^[6] 通过多个协调视图（例如矩阵视图和嵌入视图）提供对给定深度学习模型的探索性可视分析；Wongsuphasawat 等人^[9] 对 TensorFlow 框架下的神经网络的数据流图进行可视化，帮助开发人员直观地理解和调试深度学习模型的架构。除了前馈式的卷积神经网络，还有一些初步的工作努力去理解递归神经网络（RNN）及其长短记忆网络（LSTM）的架构。例如，为了在自然语言建模应用中提供网络单元语义上的意义，LSTMVis^[16] 将各个单元的激活模式随时间变化的序列可视化为折线图。RNNVis^[17] 对具有相似激活模式的隐藏状态节点进行聚类，并将它们可视化为热度图，以显示它们关联性最强的词语。

模型诊断

模型诊断旨在帮助用户理解为什么一个深度学习模型不能达到理想的结果，从而帮助用户在

改进模型的时候做出更好的决定（例如选择更好的参数或者特征）。工业界开发出一些基本的神经网络可视化工具包，允许用户调试当前模型。例如，TensorFlow 的 TensorBoard 对用户创建的给定计算图的结构进行可视化，并提供用户选择统计数据的基本折线图和直方图，以对特定节点的损失值进行可视化和激活梯度值。TensorFlow 还配备了一个名为 Embedding Projector^[18] 的可视化模块，它使用降维技术揭示网络给定层中数据点的多维表示之间的关系。Visdom^[19] 是一个基于 Web 的交互式可视化工具包，可以与 PyTorch 等深度学习库配合使用，对模型进行诊断。Deeplearning4j^[20] 是另一个可视化工具，允许用户使用几个基本的可视化组件对神经网络的训练过程进行监控。虽然这些可视化工具给出了深度学习模型本身直接提供的低层次信息的直观表示，但人类仍然难以在语义级别上理解和诊断模型行为。

最近，在模型诊断方面已经有了一些初步的研究工作。例如 CNNVis^[7]（见图 4）通过揭示多个神经元之间的相互作用及其在层间相对权重变化，允许机器学习专家调试无法收敛的训练过程或者没有

达到可接受水平的性能。它有助于找到训练过程中卡顿的潜在方向或改善模型性能。另一个例子是 Zahavy 等人开发的方法^[21]，使用 t-SNE 展示表征之间的关系，并使用显著性图来帮助用户分析影响力大的特征。三个 ATARI 游戏的案例研究证明，该方法可以分析与游戏建模有关的问题，比如初始和终端状态建模、得分过拟合等。

对抗样本分析是神经网络诊断工作的重要方向，比如 AEVis^[22]（见图 5）通过数据路径抽取算法以及多层次的可视对比，分析深度卷积神经网络对噪音的鲁棒性。在深度生成模型方面（比如生成对抗网络和变分自动编码器），DGMTracker^[19]（见图 6）是一个有代表性的模型诊断工作。它主要采用蓝噪声采样算法和信用分配算法检测输入图像的哪些区域会导致深度生成模型中特定图像集的训练失败，从而帮助机器学习专家找到导致网络训练失败的根源。

模型改进

在了解机器学习模型的表现以及它们没有达

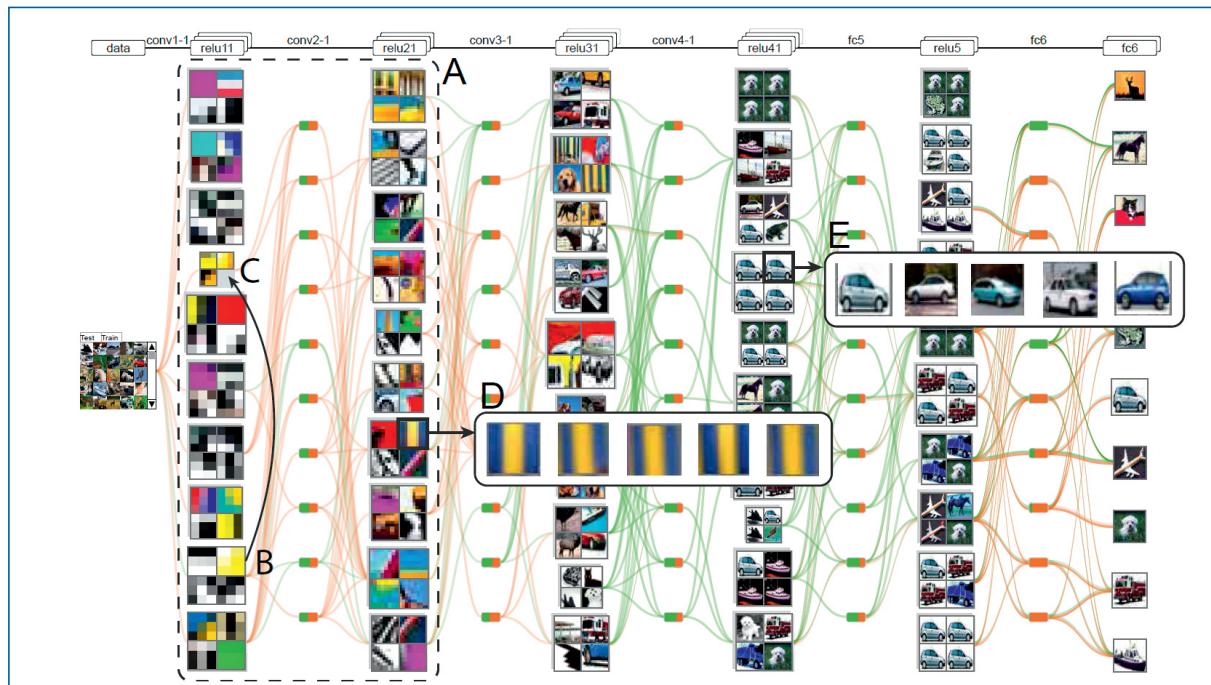


图 4 CNNVis 理解和诊断卷积神经网络

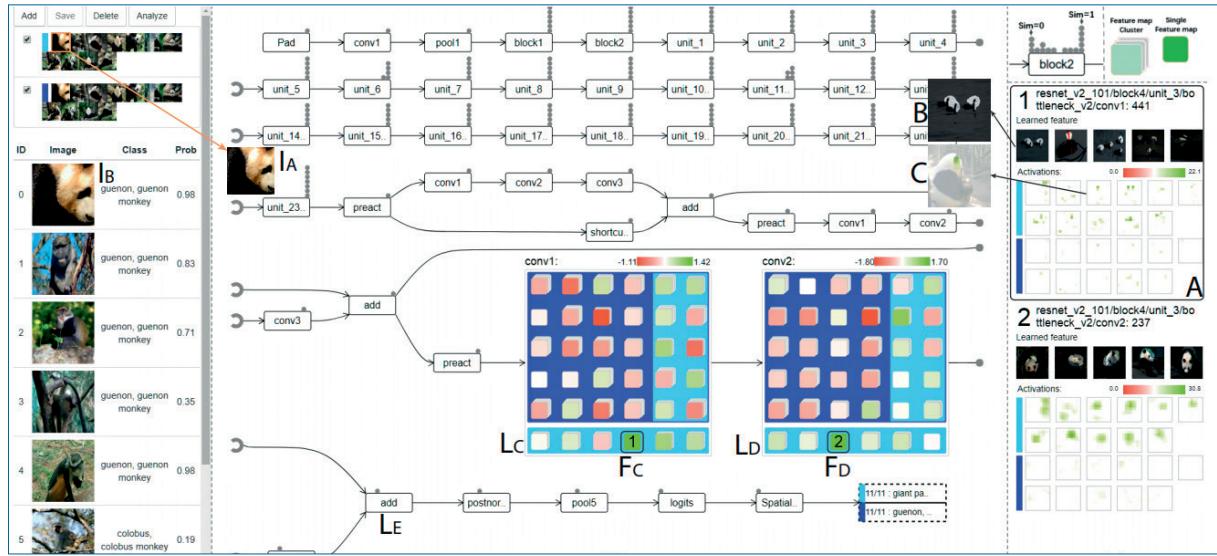


图 5 AEVis 对网络噪音鲁棒性的可视分析



图 6 DGMTracker 理解和诊断深度生成模型的训练过程

到理想性能的原因之后，机器学习专家通常希望结合所学知识改进模型。为了实现这一点，研究人员开发了可视分析系统，以提供交互功能来改善模型的表现。这些技术允许用户向模型嵌入他们的领域知识以调整影响模型输出结果的因素。通常考虑的因素包括训练样本、特征和训练中使用的参数。

DQNViz^[23]（见图 7）通过折线图、饼图和直方图对深度 Q- 网络进行四个不同粒度层级上的可视概括，即整体训练过程 (overall training)、纪元 (epoch)、时期 (episode) 和切片 (segment)，以帮助领域专家通过控制网络的随机行为提高模型性能。RetainVis^[24] 对在医疗病历单上训练的 RNN 网络进行可视分析，

以帮助用户利用领域经验与先验知识对模型进行控制和改进。RetainVis 设计了三种交互以允许用户对模型进行更改，即增加或移除输入编码，修改访问时期，以及修改每个编码对预测分数的贡献。近年来，有一些可视分析系统可实时展示深度学习模型的训练过程，并改进模型的准确性和缩短训练时间。比如 ReVACNN^[25] 是一种用于对 CNN 进行可视分析的系统，它在训练期间提供实时模型修改功能，如动态删除 / 添加节点以及在训练过程中交互式选择后续小批量的数据；DeepEyes^[26] 能够实现深度学习模型的实时监控和交互式模型控制，例如对稳定的节点和层进行高亮显示，允许用户删除激活值非常低的滤波器以对模型做出交互式的优化。

研究机遇与挑战

融入人类知识 大多数深度学习模型由数据驱动的方法构成，而从知识驱动的视角去改进模型受到的关注相对较少。从这个意义上，一个研究机遇是通过交互式可视化将人类专家知识和深度学习技术相结

合。具体而言，潜在的研究课题包括领域知识表示和解释，专业知识预测和基于知识的可视解释。此外，可以利用可视分析直观地验证模型是否正确地遵循人类注入的知识和规则，这是确保深度学习模型在应用中表现正常的关键步骤。例如，当训练以摄像头获取的图像为输入的自动驾驶模型时，可以向模型中加入规则，避免撞击场景中识别出的人。

渐进式深度学习可视分析 大多数现有的可解释深度学习方法主要侧重于在模型训练完成后离线地理解和分析模型预测或训练过程。由于许多深度学习模型的训练非常耗时（需要数小时到数天的计算），因此需要使用渐进式可视分析技术将专家融入到深度学习模型的分析流程中。为此，深度学习模型期望在训练过程中产生语义上有意义的中间结果。然后，专家可以利用交互式可视化来探索这部分结果，检查新的结果，并进行新一轮的探索性分析，而无须等待整个训练过程完成。

提高深度学习的鲁棒性以实现安全人工智能 深度学习往往依赖大量的高质量训练数据来充分学习模型的参数，而当数据中存在噪声或者属性



图 7 DQNVIS 对深度 Q- 网络的训练过程进行可视化及交互改进模型

缺失时，这些深度学习模型的识别精度通常会受到很大的影响。例如对抗样本是一类被恶意设计用来攻击机器学习模型的样本，它们对真实样本的修改非常轻微，以至于人类观察者根本无法注意到，但模型却做出错误的预测。这些对抗性样本通常用于攻击深度学习模型。保持深度学习模型的鲁棒性对于实际应用（如无人驾驶、人脸识别 ATM 机）至关重要。而产生这些鲁棒性问题的本质原因是目前人们对很多机器学习模型的工作机理不够理解，无法针对性地解决问题。因此，关于可解释深度学习的一个研究机会是结合人类知识来提高深度学习模型的鲁棒性。

减少所需训练集的大小 通常，深度学习模型包含数百万个参数。为了充分训练具有大量参数的学习模型，需要成千上万个训练样本。在实际应用中，给每个特定任务分配单独的大规模训练数据集是不切实际的。有必要利用先前在相似类别中训练模型获得的先验知识以及人类专业知识来减少所需训练集的规模。一次学习 (one-shot learning) 或零次学习 (zero-shot learning) 是当前训练小样本深度学习模型的主要方法，它提供了将对象的先验知识结合到“先验”概率密度函数中的可能性。也就是说，使用给定数据及其标签训练的那些模型通常只能解决它们最初训练的预定义问题。例如，如果没有足够的“老虎”标签训练数据，检测“猫”的深度学习模型原则上不能检测“老虎”。通过可视化分析框架注入少量用户输入可能解决这些问题。因此，未来研究的一个有趣方向是探索如何将可视分析与小样本学习算法结合，以融入外部人类知识并减少所需的训练样本数量。

面向数据科学家和从业者的可解释性 大多数现有工作侧重于面向机器学习专家对深度学习训练过程进行理解和分析。这种类型的可解释性对机器学习专家构建强大而高效的学习模型是有效的。但在实际应用中，数据科学家和从业者还需要通过机器学习模型的可解释性来建立相关的业务或者做出合理的决策，主要障碍是预测性能和可解释性之间的权衡。通常，基于决策树的方法易于被非机器

学习专家理解；而深度学习模型有更好的预测性能。然而，深度学习的复杂性通常会降低可解释性。因此，一个有趣的方向是研究如何利用自解释的机器学习模型，如决策树，去解释深度学习模型的工作机制和预测结果，从而使从业者能够更好地理解模型预测，并选择适当的模型和参数。

可视化系统的可扩展性 在大数据时代，模型的大小以及数据量都在不断增长。对具有复杂结构和大量组件的模型进行可视化具有挑战性，尤其是当模型需要处理数以万计的数据时。例如，如果使用 ResNet-101 在包含数百万个图像和数百个类别的数据集上训练，通常很难对其进行合适的可视摘要或抽象，以支持模型分析和预测理解任务。因此，可视化领域的研究人员需要设计和开发可拓展性强的可视化技术，同时对模型结构、预测结果和原始数据进行展示，并将它们连接在一起全面分析深度学习过程。 ■



姜 流

清华大学博士生。主要研究方向为众包机器学习、交互式机器学习。
jiangl16@mails.tsinghua.edu.cn



刘世霞

CCF 高级会员。清华大学长聘副教授。
IEEE Transactions on Visualization and Computer Graphics 副主编。主要研究方向为可视分析、信息可视化和文本挖掘。
shixia@tsinghua.edu.cn



雷 娜

CCF 专业会员。大连理工大学教授。主要研究方向为计算共形几何、计算拓扑、符号计算。
nalei@dlut.edu.cn

参考文献

- [1] Zhang Q, Wu Y N, Zhu S C. Interpretable convolutional neural networks[C]// *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018:8827-8836.

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

面向隐私保护的可视分析

王叙萌 陈为
浙江大学

关键词：隐私保护 可视分析

隐私保护

在信息爆炸时代，数据中蕴含的价值发挥了越来越大的作用。小到餐馆定价，大到政府调控，有了数据的支持，决策变得更加容易且有针对性。与此同时，为了支持数据分析、挖掘需求，数据收集变得无处不在，从城市中的传感器到手机上的应用程序，我们的信息也在不经意间被记录下来。一条条数据汇聚在一起，推动着社会的发展，也为窥探隐私的不法分子营造了难得的“机遇”。

信息的暴露给人们的生活带来困扰，也对人身和财产安全造成威胁。但是为了寻求一些服务，我们又不可避免地需要将一些信息与他人共享。为了更好地使用数据，数据隐私保护成为亟待解决的重要问题。

隐私暴露

隐私暴露的方式主要有两种。一种为**身份泄露**，即攻击者可以识别数据中个体的身份，从而了解该个体的全部信息。可以用来识别个体身份信息的不止是姓名、身份证号等可以唯一标识身份的信息，其他信息通过叠加也有可能暴露身份信息。以一个班级中学生的兴趣爱好为例，喜欢打篮球的学生可能有很多，喜欢书法的也有几个，但是既

喜欢打篮球，又喜欢书法的学生可能只有一个。当信息量增加时，通过多种信息识别个体的概率就会大大提高。当身份信息暴露时，数据库中的所有关于个体的信息都将与个体对应起来，被攻击者知晓。另一种为**信息泄露**。在这种情况下，攻击者虽然无法辨识个体的身份，但可以结合自己对个体的了解，基于数据中的信息推断出个体的敏感信息内容。预防信息泄露的要求一般比阻止身份泄露的难度要大。考虑到数据分析可能会用到数据的整体分布信息，在保护隐私的前提下尽量维护数据的应用价值就更加困难。

常见方法

达斯古普塔 (Dasgupta) 和科萨拉 (Kosara)^[1] 对可行的隐私保护方法进行了总结 (见图 1)。他们认为，从原始数据出发，经过不同的处理，可以通过可视聚类和数据聚类两种手段实现具有保障性的隐私保护。实际上，隐私保护的方法不局限于聚类

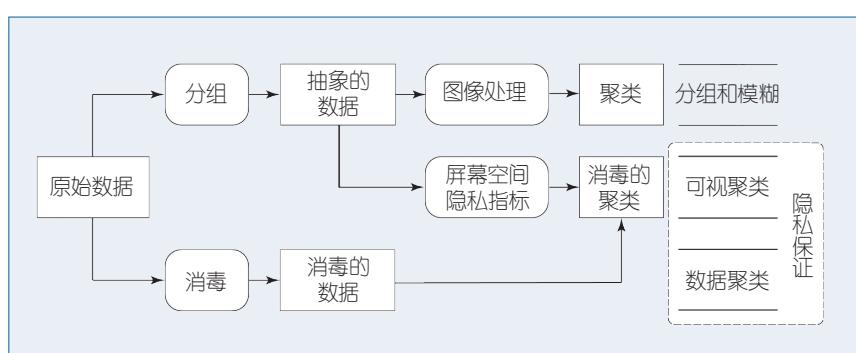


图 1 可行的隐私保护方法^[1]

方法。

视觉空间

在视觉空间对数据进行处理，即对数据的可视表达进行限制，是一种针对使用可视分析的人员提供的隐私保护方法。在可视分析中，可视化基于视觉通道向分析人员传达信息。不同的可视表达可以对信息给出不同的诠释。当可视表达中存在遮挡、模糊等问题时，分析人员就会对数据具体情况的理解产生不确定性。这种不确定性虽然会在一定程度上干扰可视分析，但合理运用也可以实现隐私保护的效果。

数据空间

在数据空间处理数据的隐私保护方法，允许分析人员通过一定渠道获取处理后的数据，并使用任意方式进行分析。语义匿名模型和差分隐私模型是该方法的两类经典模型。

语义匿名模型指通过模糊信息等方式，使个体无法被区分。其中，无法被攻击者区分的数据集合称为等价类。最早的语义匿名模型是由斯威尼(Sweeney)提出的 k -anonymity 模型^[2]。该模型要求每个等价类中的数据项个数不少于可定义的参数 k ，即每个个体至少和其他 $k-1$ 个个体不可区分，为此，需要将不满足要求的等价类进行合并。 l -diversity^[3] 和 t -closeness^[4] 等方法对等价类提出了更高的要求：等价类中的敏感信息种类要有足够的差异性，以及具有和整体数据近似的分布。对于此类方法来说，参数值的设定决定了隐私保护级别的高低。级别越高，对等价类的要求越高，需要进行合并的等价类往往越多，对数据质量的影响越大。

差分隐私模型^[5]是基于随机算法对信息添加扰动，从而使攻击者对整个数据集中的任一信息都不能完全确信，从而实现隐私保护的目的。在信息之间相互独立的情况下，这种方法可以达到比较理性的隐私保护目的——在保护隐私的同时几乎不损坏数据的应用价值。但实际上，大部分数据集中的信息是存在相关性的，随机扰动难以对多信息之间的相关性进行维护，差分隐私模型也会对数据的应用价值产生破坏。

隐私保护面临的挑战

隐私保护方法评估

对隐私保护方法的评估是基于两个目标——隐私保护的程度和数据的应用价值进行的。

隐私保护程度

想要了解隐私保护的程度，首先需要了解可能的攻击，才能对潜在的危害进行计量。科学技术的发展不仅创造出了众多隐私保护方法，也衍生出了相应的攻击方法，它们可以基于部分背景知识来对数据中的敏感信息进行破译。比如当攻击者知晓某个个体的部分特征时，可以通过图数据分析中的节点查询来进行身份匹配，从而识别个体的身份；机器学习中常见的分类问题也可以用于对个体的敏感信息进行推测。

不同的攻击方式有着截然不同的思路。要想确切地掌握隐私保护的程度，就需要针对不同的攻击方式一一进行测评，带来了极大的工作量。

除了攻击方式以外，影响攻击成功率的还有攻击者的背景知识。简单来说，攻击者对数据越了解，破解出敏感信息的概率就越大。对此，我们必须在对攻击者的背景知识进行限定的前提下检测隐私保护的程度。如何对攻击者所掌握的信息量进行估计，是一个亟待解决的问题。

数据的应用价值

数据是用来分析和挖掘信息的。数据的应用价值可以理解为，数据在分析任务中还原和表述、推导信息的能力。在一些研究中，数据集的某些指标会被用来量化地分析数据的应用价值。对不同研究对象计算一系列指标，并比较隐私保护前后数据指标值的变动，可以在一定程度上反映数据的应用价值。然而真实的分析过程所需要的信息可能远比一系列的指标复杂。比如识别数据中的特殊模式，有时就无法通过指标计算得出。同时，研究人员分析数据可能需要对未知的信息进行探索，此时，我们甚至无法定义具体的分析任务是什么，测量数据的应用价值就更加困难。

隐私保护与应用价值的冲突

隐私保护和应用价值之间存在着难以消除的矛盾，这个问题仍存在着优化的空间。首先，如果用等价的应用价值来换取隐私保护，那么这个交易需要进行到什么程度？当明确了隐私保护程度之后，就没有必要消耗多余的应用价值。其次，如果实现相同的隐私保护程度可以有不同的应用价值损失选择，是否可以选择不重要的部分进行交易？如果能够明确某部分数据的价值，是否可以有针对地进行保护？

可视化与数据感知和分析

可视分析是一种常见的数据分析方法。该方法利用可视化技术将数据通过更直观的方式表达出来，并允许用户交互地对数据进行探索，从而达到了解、分析数据的目的。

感知的不确定性

视觉是人类获取信息最有效的途径。可视表达的方法有很多，颜色、大小、形状、位置、方向等都可以作为编码信息的通道。比如常见图表中的散点图，就是利用点所在位置的横纵坐标来编码一组信息的。但是这种信息在被人眼接收并解析之后，得到的解释可能和它们原本的意思存在差别。这是因为可视化过程受到了载体和个体视觉感知的限制。

载体的限制 以屏幕为例，它的大小和像素密度就直接限制了其可能承载的信息量。在大数据时代，仿佛有数不清的数据等待人们去分析，可以说多大的屏幕都不为过。随着科技的发展，屏幕的大小和像素密度都在逐渐提高，然而这种提高也是有限制的。当屏幕超过教室中黑板的大小时，浏览全部细节就需要用户转动脖子了。而当像素密度高到一定程度，人类的视觉感觉能力也感知不到差别。

视觉感知的限制 以颜色为例，人类的可见光谱只有 400nm~780nm 波段的一部分。而在这部分内，虽然我们可以用屏幕分别展示出 RGB(255, 0, 0)

和 RGB(254, 0, 0) 两种颜色（见图 2），但是却很少有人可以区分两者之间的差别。同样，在大小等其他编码上也存在类似的问题。

另外，我们在感知信息的时候，可能还会产生一些错觉。这些错觉被归为三大类：几何学错觉、生理错觉和认知错觉¹。它们分别是由图像构造、感官以及心理原因导致的错觉。受到周围其他图形的干扰，

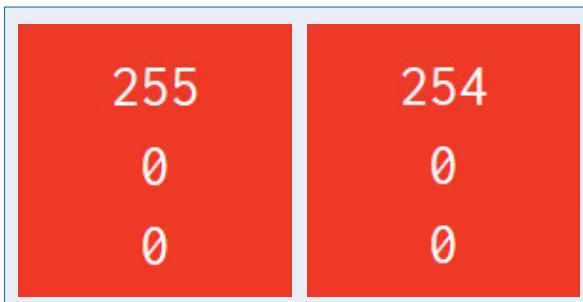


图 2 颜色 RGB(255, 0, 0) 和 RGB(254, 0, 0)

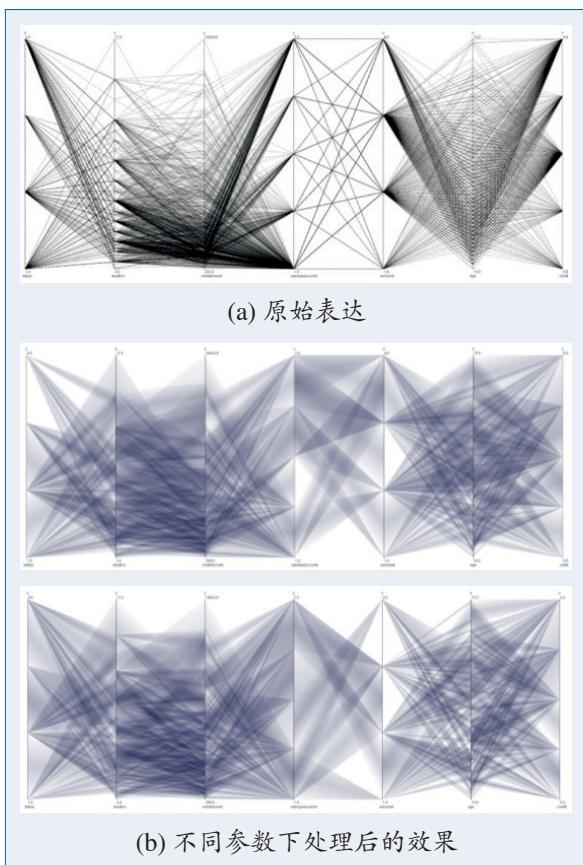


图 3 利用不确定性来实现 k -anonymity 模型^[1]

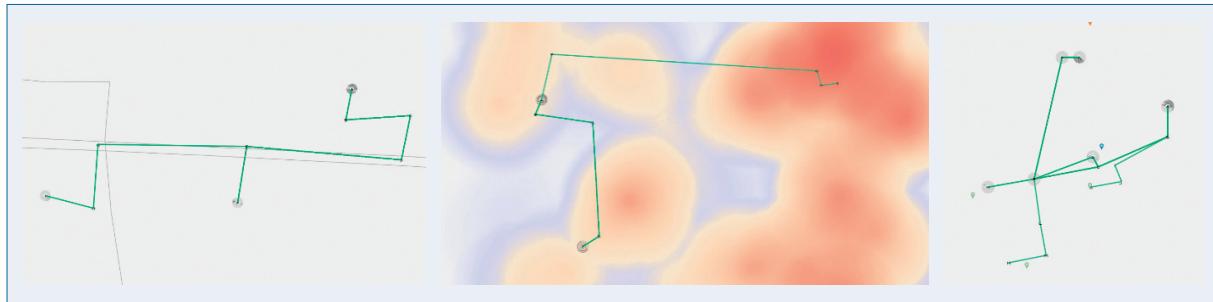


图 4 使用不同信息分析轨迹数据^[6]。从左到右依次为使用路网、热力图和 POI 为辅助信息

我们的感知会和真实信息产生差异。

感知的不确定性影响了人们对信息认知的准确性，一方面它会对分析造成干扰，但另一方面，它也可能成为阻止隐私信息泄露的屏障。研究 [1] 就利用模糊来干扰其他人准确掌握敏感的信息值。通过调整模糊的程度，该方法可以在平行坐标上实现不同参数的 k -anonymity 模型的要求（见图 3）。

合理地使用不确定性可以在用户通过可视分析探索数据的时候巧妙地对敏感信息进行保护。但这种隐私保护方法对用户的分析方法有很大的限制。

控制辅助信息

除了数据本身，研究者还会借助其他辅助信息来分析目标数据。对辅助信息进行筛选、控制，也可以降低敏感信息泄露的概率。以轨迹数据为例，要分析这种数据往往需要借助地图来查看轨迹的位置。地图上有很多信息可以辅助分析，如兴趣点（Point of Interest, POI）分布、人口密度分布和街道图等。研究 [6] 通过实验对这些地图信息在数据分析和位置信息泄露上的能力进行了比较（见图 4），并给出了关于轨迹分析系统设计的建议。

可视分析与隐私保护流程

除了控制信息的表达和分析过程以外，可视化可以直接用于隐私保护的定制流程。可视化本身的特色包括解释复杂数据和交互探索数据。这两点在

隐私保护流程中都十分重要。

评估结果

可视化可以对数据进行多角度的展示。相较于指标和文本式的数据，可视化可以更直观、更全面地对数据的结果进行总结。对于原始数据和处理后的数据，可以通过可视化从数据的总结指标、数据本身等多角度对二者进行比较。基于可视化的比较有并列、叠加和对差异进行编码三种方法（见图 5）。我们可以从中快速发现数据发生的变化。基于用户定义的标准，我们可以继续探究这些变化给隐私保护和应用价值带来的影响，以及最终结果是否达到要求。引入可视化，模拟实际使用场景，探究使用处理后数据的分析效果，以及模拟攻击判断隐私保护的薄弱环节等也变得简单可行。

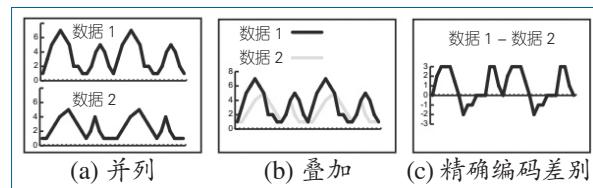
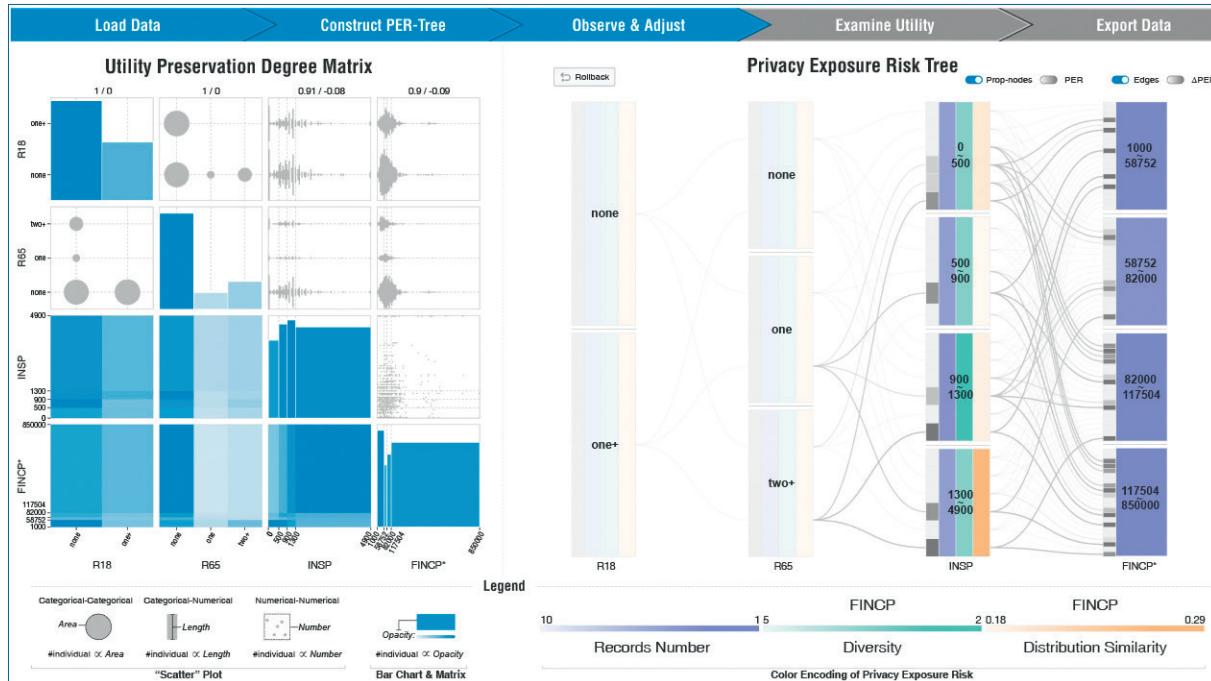


图 5 比较可视化的三种方法^[7]

定制方法

结合可视分析的交互，用户可以完成对隐私保护模型的选择和调节。要处理隐私保护和应用价值之间的冲突，需要对处理过程有比较精确的把控。这部分处理目前仅靠自动方法是无法完成的。融入

¹ 参见 https://en.wikipedia.org/wiki/Optical_illusion。

图 6 针对表格数据的可视分析隐私保护系统^[8]

专家知识，即用户对数据、潜在攻击的了解，是这个过程必不可少的一环。可视分析正是连接自动方法和专家知识的桥梁。可视表达可以将数据和自动方法的结果呈现给用户；交互可以将用户的指令传达给自动方法。

Wang 等人^[8, 9]给出了利用可视分析来完成隐私保护方案制定流程的方法和系统。针对多属性表格数据，工作[8]使用隐私暴露风险树（图 6 右）辅助用户制定隐私保护方案，并设计应用价值分析矩阵（图 6 左）来对应用价值进行评估。在该系统的辅助下，用户可以对隐私保护方案进行选取，最终找到合适的处理方法。

展望

人工智能等对数据有依赖性的技术不断发展并创造着价值，人们对有效隐私保护方法的需求也越来越迫切。顺应这个需求，会有更多的科研人员投入其中，并给出更好的答案。它们可能是完善了流程中的某一环节，可能是结合了更好的模型，或者

是满足了其他应用场景的特殊需求。这将是一个如同树木生长般的优化过程，不断吸收新的养分，长出新的枝芽。 ■

王叙萌

浙江大学博士研究生。主要研究方向为可视分析、城市数据和隐私保护。
xumengwang18@gmail.com



陈为

CCF 高级会员。浙江大学教授，博导。主要研究方向为大数据分析和人机混合智能。
chenwei@cad.zju.edu.cn

参考文献

- [1] Dasgupta A, Kosara R. Adaptive privacy-preserving visualization using parallel coordinates[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17(12): 2241-2248.

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

CCCF 2018 年最受关注文章

CCCF 在 2018 年共发表主编评语、专题、专栏、动态、译文、学会论坛等各类文章 220 篇。为了解读者对不同类型、不同方向的文章的关注程度，CCCF 编辑部对“CCF 通讯”微博、CCF 微信公众号上的 2018 年已发表文章的阅读量进行了统计，并根据读者反馈，得出 CCCF 2018 年受关注度较高的文章（按月份和栏目排序）。

月份	作者 / 译者	文章标题	栏目
1	李国杰	走务实的人工智能发展之路	主编评语
1	周志华	关于强人工智能	专栏
1	包云岗	关于 RISC-V 成为印度国家指令集的一些看法	专栏
1	杜子德	学会如何开会	学会论坛
2	黄铁军	也谈强人工智能	专栏
2	杜子德	如何演讲	学会论坛
3	李航	人工智能的未来——记忆、知识、语言	专栏
4	张加万	数字人文与计算社会科学	专题
4	王腾蛟、陈郁馨、陈薇	计算社会科学的兴起与发展	专题
4	董玮、高艺、陈纯 等	物联网快速开发平台	专栏
4	刘宇航	冯·诺伊曼《计算机与人脑》要点归纳及启发	专栏
5	史元春	自然人机交互	专题
6	杨珉	走向智能化的网络安全研究	专题
6	全伟、刘洋洋、仲盛 等	智能交通系统中的安全与隐私保护	专题
6	杨龙飞、琴琴、杨天、王涛	区块链的关键技术、应用与挑战	专栏
6	CCCF 动态编委组	人工智能专业的探讨	动态
6	吴刚	寒武纪发布云端智能芯片 MLU100	动态
7	邹磊、彭鹏	大图数据分析系统综述	专题
7	钱正平、周靖人	图计算在阿里巴巴的应用与挑战	专题
7	陆品燕	建设理论计算机科学研究中心的实践与思考	专栏
7	徐文渊	海豚音攻击的幕后故事	动态
8	李忠阳、丁效、刘挺	Yann LeCun 在 IJCAI 2018 开幕式上的演讲： 我们需要一个世界模型	动态
9	孙园园、华宇	面向哈希计算的新型数据组织结构	专栏
9	黄民烈、刘洋	第 56 届国际计算语言学年会	动态
9	蔡亮、梁秀波 等（译）	加密数字货币的风险	译文
10	邓燚、陈宇	零知识证明：从数学，密码学到金融科技	专题
11	李国杰	计算机科学基础理论需要重塑	主编评语
11	杨强、刘洋、陈天健、童咏昕	联邦学习	专栏
11	谢晓晖、卢泓宇、张帆 等	ACM SIGIR 2018——信息检索领域的学术盛会	动态
12	李国杰	发展数字经济值得深思的几个问题	CNCC 特邀报告
12	彭思龙	工业技术进步没有捷径	专栏
12	陈文光、李悦乔 (Josefina Li) (译)	ACM 道德规范与专业行为准则	译文

计算机问题求解的三类方法

裘宗燕
北京大学

关键词：问题求解 机器学习

问题和问题求解

作为计算机工作者，我们关注的最基本问题是各种问题及其计算机求解，以及如何从遇到的问题得到能解决问题的计算机系统。那么，什么是问题呢？

实际问题千变万化，我们考虑一种抽象的统一说法：一个问题就是从一个实例描述集合到解集合的映射， $T: I \rightarrow O$ 。如果 I 是有穷集， T 就是有穷问题，否则 T 是无穷问题。 $i \in I$ 或 $(i, o) \in I \times O$ 称为问题 T 的实例， $o=T(i)$ 称为问题 T 对应 i 的解。

计算机只会做一件事，就是自动执行程序。用计算机解决一个问题，就要写出一个解决这个问题的有穷程序（无穷长的程序写不完）。如果针对问题 T 写了程序 P ，把 T 的任意实例 i 作为 P 的输入， P 的输出总是 $o=T(i)$ ，我们就可以说 P 解决了问题 T 。

针对问题写出程序是计算机领域最重要的工作。要完成这种工作，首先要确定一种技术路线（一种方法），确定需要从问题中挖掘出什么信息（知识），怎样利用得到的知识构造程序。参考几十年的研究和实践，本文总结出三类方法。是否还有第四类？请读者考虑。

基于算法的系统 (Algorithm-Based Systems, ABS)

这个太显然。但是，采用这一方法有什么前提条件？在什么情况下有可能得到解决问题的算法，

并实际写出能解决问题的程序？我们可以提出下面一些条件。

首先，该问题必须是“可计算的”，这是理论要求。证明问题是否可计算常常很不容易，但做出算法就是正面的证明。其次，要做出算法，要求我们对问题及其求解完全理解，知道如何处理问题的任意实例，能处理求解过程中可能遇到的任何情况。

算法的优点是直接针对具体问题，是专用的，因此效率高。另一方面，我们比较容易把算法转换为解决问题的程序，也比较容易判断一个程序是否确实解决了问题。

注意：任何有穷问题都有算法。由于情况（输入）有穷，理论上总是可以对其做穷尽分析，分情况给出解。求解程序如下（假定问题只有 n 个实例， x 是输入）：

```
if x is i1 then o1
elif x is i2 then o2
...
elif x is in then on
```

无穷问题也可能写出有穷算法，例如求最大公约数问题。

现在看看用计算机下围棋的问题。围棋要求对弈双方在 19×19 的棋盘上交替落子，目标是占据最大的地盘。下围棋程序只需要解决一个问题：对任何棋局，确定下一个棋子的最佳位置。由于局面有穷，本问题为有穷问题。按前面的说法，存在确定最佳位置的算法。基于这一算法的围棋程序绝不

会犯错，其棋力不会弱于（昨天、今天或未来的）任何棋手，或任何下围棋的程序，包括 AlphaGo 和 AlphaZero。

我们很容易规划出一套写这个算法的系统化方法，只要有充分的耐心和时间，就能做出一个这样的围棋程序。但是，由于可能局面太多（约为 3^{361}

10^{172} ），彻底分析每个局面需要的时间太长，这一工作不可能在合理的时间内完成。另一方面，即使能写出来，程序也太长，计算机没有足够的存储器存放它。这一实例说明，理论上有算法，但并不代表能用算法解决问题。

算法的另一个限制是计算机专业人士最熟悉的：必须足够高效（复杂性不高），保证每次执行可以在合理时间内完成。例如，如果一个围棋程序在求解中，需要检查的局面个数可能达到 10^{172} ，则这个程序（实际上）毫无意义。

下面是能用算法解决问题的一些条件：

- 首先对问题要有完整的认识，对求解过程有全面完整的把握，否则只能考虑其他方法。
- 建模后得到的抽象问题是可计算的（理论条件），否则只能考虑解决原问题的恰当的可计算问题，或降低要求，解决结果类似但要求较低的问题。
- 算法存在有穷长度的描述，而且描述不“过于长”，可在合理时间内写完。
- 每个实例的求解都能在合理时间内完成。如果效率太低，只能设法找其他算法，或收缩问题（解决合适的子问题），或降低要求，如最优解改为近似解等。

如果找不到算法，或者虽然有算法但不适用，我们就只能考虑其他方法。

基于搜索的系统 (Searching-Based Systems, SBS)

用计算机解决问题的另一类方法是搜索，也就是通过探查和回溯的方法寻找解。搜索方法并不要求我们对问题及其求解过程有全面的认识，但需要有一些清晰的知识片段，以这些知识片段

作为搜索的基础。例如，对基于规则的系统，需要有针对问题的规则库；自动推理需要有事实和推理规则。基于搜索的系统由两部分组成：一部分是针对问题的知识库，包含一集知识片段（规则）；一部分是搜索算法，它设法利用知识库去拼凑出实例的解。

人工智能领域研究过许多搜索方法，开发了一些通用框架（如产生式系统、黑板系统）；提出了许多技术（如各种启发式搜索算法），实现了一些应用（如基于规则的专家系统）；提出了一些编程泛型（如逻辑程序设计、约束程序设计）。还有自动推理、定理证明等方面的研究成果。赫伯特·西蒙（司马贺）等人称搜索为通用问题求解 (general problem solving) 方法。

只有解决问题的片段知识（规则），通常做不出算法。规则可能不完全，无法处理所有实例或所有情况，也没有处理策略（顺序和过程）。但规则可以用于搜索，通过试探和回溯的方式去求解。搜索的特点是，算法不直接针对问题，而是针对规则的使用。规则可独立描述，也可以嵌入搜索算法中。搜索也已广泛用于实践，例如自动泊车系统常包含搜索。

关于计算机下棋的研究已有几十年历史，基本方法就是博弈树搜索，评价各种可能，选出最佳棋着。围棋的规则很简单，很容易实现一个基于搜索的围棋程序。为此，我们只需实现一个处理博弈的通用搜索引擎，让它用围棋规则去做穷尽搜索，找出最佳棋着。

当然，稍微了解计算机的人都会说上述方案不可行。规模为 10^{172} 的状态空间太大，朴素搜索方法实际无用。这个不可行的原因是实际计算开销，而不是理论不正确。这也说明了搜索方法的一个重要弱点：求解中需要探查的空间通常以相对于实例规模的指数方式增长。因此，对复杂一些的问题或规模较大的实例，我们可能无法等到结果。

实际围棋程序（包括 AlphaGo 和 AlphaZero）都采用了一些策略，如限制最大搜索深度（或限制搜索时间），加入随机性因素等。设法在不能穷尽搜索

的情况下合理地评估局面，只能评估检查到的那些局面，并从中选择可能通向“最佳路径”的棋着等。

搜索方法的优势是有可能利用部分知识解决问题，但也存在许多固有的缺陷：

- 搜索策略或规则选择策略不当，可能使搜索进入无穷的或巨大的无解区域，导致实际有解，但却（很长时间）找不到。也难保证对不同实例的（时间）行为一致性。

- 搜索状态爆炸的本质性问题。

- 规则集不完全，可能导致某些实例无法求解；而规则集越大，状态爆炸问题越严重（规则集丰富表示对问题的认识丰富。这是搜索方法的一个固有矛盾）。

假设有了一集规则，在采用基于搜索的方法时，需要考虑搜索策略（宽度优先、深度优先、其他），规则选择策略（固定顺序、估值序、随机等），还需考虑状态评估，可能的剪枝策略，局部搜索（限制搜索深度、与评估结合）等问题。

基于实例的系统 (Case-Based Systems, CBS)

如果对问题的了解非常少，或基本上没有有关求解的知识，还能用计算机吗？实际情况经常如此：我们认为面临的是一个问题，有需要处理的情况，人能得到有用的结果。例如，医生看了检查报告说“可能 ×× 有炎症”，或围棋大师看到一个局面说“感觉不错”。

假设我们“认定”了一组输入和结果 $\{(i_1, o_1), (i_2, o_2), \dots, (i_n, o_n)\}$ 是一个问题的实例，就可能实现一个简单的求解程序（设 x 为输入）：

```

if x is i1 then o1
elif x is i2 then o2
...
elif x is in then on
else "I can't handle it"

```

这个程序与前面类似，但本质不同。这里的实例可能只是所有可能实例中的一部分。

人们通常不满意这种“死记硬背”，希望推广之，这样就需要“归纳”或“学习”。用计算机自动推广就是“机器学习”，也就是第三种方法——基于（实例）学习的求解方法，设法通过自动处理一些“情况 - 解”实例，得到一个能处理该问题的更多实例的系统。

用机器学习方法解决问题，需要设计一种具有可调整要素的（数据）结构 S ，一个（能利用实例调整 S 的）学习算法 L ，和一个使用 S 解决问题的算法 U 。对已有的实例集 E ，学习算法得到 $L(E, S) = S'$ ，而 $U[S']$ 就是利用调整后的结构解决问题的程序。学习算法 L 有效，首先要求将 $U[S']$ 应用于 E 中的实例输入 i_j 时，得到的结果近似于 o_j 。

例如，多层神经网络是目前常用的一种结构，其中的可调整要素就是神经元之间的连接系数。针对这种结构，人们提出了一些学习算法，该结构的使用算法很简单。

复杂性使我们无法通过穷尽搜索的方法评价围棋棋局。以前人们利用专家知识做评价，主观而且不准确。AlphaGo 的创新就在于通过机器学习自动产生评价函数。AlphaZero 和 AlphaGo 的差异在于学习实例的来源，AlphaGo 用历史上的围棋实战作为实例，AlphaZero 用自对弈棋局作为实例。结合其他机制，Alpha 系列程序取得了令人瞩目的战绩。

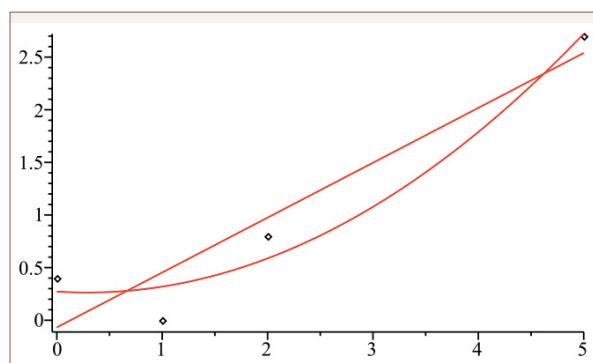


图 1 线性和二次多项式拟合数据点示例

实际上，机器学习就是用某一类函数去“拟合”一组数据。数学抽象就是：有一组数据和一个函数

类，设法找到能最好地表达这组数据的函数表示。图1是一个用线性和二次多项式拟合几个数据点的例子。那么，哪个结果更好地表达了数据中蕴含的关系呢？

我们的目标是希望拟合得到的函数能用于处理其他实例。但是，目前对问题的理解只有这一组数据（而且数据可能有误差），因此我们无法回答“哪一个更好”这个问题。

对具体的基函数组，可以设计一种评价标准（某种最小偏差）。而对这组数据的拟合，则无法确认这种标准。真正的标准应该来自问题，基于对被处理问题的全面理解。只有目前这组实例，不可能做出“正确”判断。即使拟合函数对现有数据都完全重合，也未必是最好的预测。例如，对4项数据可以做一个3次多项式，使之经过每个点。图2表示了拟合函数的情况。人们一般都不认为这个拟合更好。

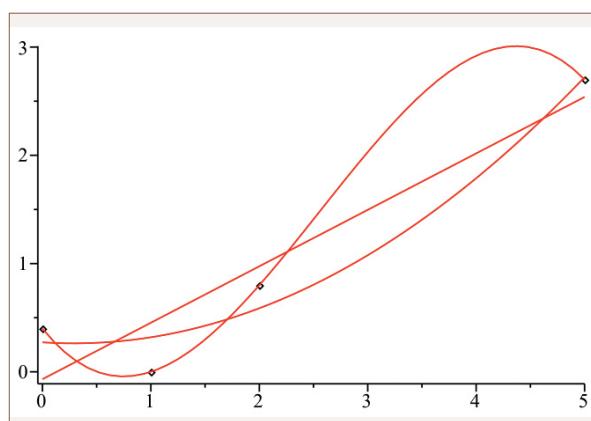


图2 多种拟合示例

对函数拟合的研究提出了两个重要情况。**欠拟合**：由于函数结构太简单，无法反映问题的重要特征，对应于机器学习中使用的结构过于简单。**过拟合**：函数类过于丰富，反映具体实例的过多细节，掩盖了目标问题的本质性特征，对应于机器学习中使用的结构过于复杂，导致结果函数震荡剧烈，降低了学习结果的预测能力。从有关函数拟合的讨论中，可以看到学习方法固有的局限性：由于没有对问题的完整理解，我们无法讨论学习结果的正确性，

只能问得到的结果“好不好”，而“好不好”也只能通过实例来检验。

常见检验方法有两种：一种是在真实世界里检验，例如自动驾驶，只能在实际道路上试验，希望车辆不出事故、不撞车、不撞物撞人等。另一种方法是把已有实例分为不相交的两部分 $R = R_1 \oplus R_2$ ，其中的 R_1 用于学习， R_2 用于检验学习效果。

实际上，基于学习的方法还有一些未说明的假设，只有在这些条件下，学习方法才可能有效：首先是假设问题实例和解的关系是连续的；其次是假设每个样本（实例和解）都必定包含了一些反映问题的整体性质的全局信息，因此样本越多越好，信息越多偏差越小。实际上，这些假定都是有疑问的。另外，在不理解问题的情况下，这些条件都无法检查。因此机器学习系统的设计和实现有很大的主观性和试探性，最后只能是“结果决定一切”。

基于（实例）学习系统的另一个缺点是学习代价可能很高。例如，AlphaGo 采用超级计算机，学习中对弈超过 3000 万盘。AlphaGo Zero 自对弈 3 天（约 500 万盘），水平可超越之前的 AlphaGo，自对弈 30 天可以超越后来的 AlphaGo Master 版本。由于围棋的特点，实例易得。但很多问题难以得到大量实例，采用这种技术路线的有效性存疑。

根据上面讨论，我们可以总结出适合用机器学习求解的问题的一些特点。首先，问题具有信息完全的场景（静态的、公开的）。一些棋类有这种性质，包括围棋，但许多问题不是这样。其次，存在（或容易得到）大量可用实例。围棋有很多积累，而且容易自动生成。再次，学习效果比较容易判断。例如围棋的胜负和统计可以作为评价。最后，由于学习结果的不可控性，机器学习慎用于安全攸关的应用（否则就需要其他安全措施）。

目前机器学习领域非常热，许多研究者和企业投入其中，也有一些原因：首先是做出了一些有影响的实例，特别是 AlphaGo 及其后续工作；再就是已经取得了一些成果，使人们看到一些用传统方法不能处理的问题有了新的解决途径。人类不清楚如

何解决的问题大量存在，因此存在丰富的有可能用机器学习去探索的问题和应用领域。随着有关研究工作的开展，人们也开发出各种与神经网络类似或有些不同的模型，这些情况都推动着机器学习领域研究和应用的开展。另一方面，计算机能力增强也是机器学习取得进展的重要原因。

但学习方法属于试探性方法，类似于实验科学，与计算机科学技术的常规方法距离比较远。通过学习，从实例得到的结果能外推到什么程度，实际上是不清楚的。因此，机器学习方法更适合于那些只有优良程度要求（而非判定性要求）的应用。例如：要求精确结果的整函数通常无法学习，如斐波那契函数、最大公因子等。我们也应该注意机器学习的性质和本质弱点，避免可能的危害。

三种方法的比较和总结

从前提条件看：采用算法，需要有对问题及其求解的完整知识；采用搜索方法，需要有对问题及其求解的片段知识；采用学习方法，只需要问题和解的实例，无需其他知识。

从技术上看：算法只需要一个直面问题的解决方案；采用搜索，通过一个间接的搜索算法去应用有关问题的片段知识。学习方法需要一个结构和两

个算法：一个具有可调要素的结构承载学习结果，一个利用实例调整该结构的算法和一个利用该结构处理实例的算法。

从效果看：正确的算法必定给出正确的解，求解效率可以预估，但也有复杂性太高的可能性。由于知识有限，搜索方法通常不能保证得到解，得到解的代价也很难预估，但有可能保证解的正确性。另一方面，搜索通常不能保证得到一个解。由于知识贫乏，我们只能说学习方法得到的结果可能有用，其他方面都没有保证。

可见，这三类方法各有各的前提条件、优势，以及固有的缺陷。遇到一个问题时，我们应该根据对问题的潜在认识程度、问题的实际需要等选择合适的求解方法。

请注意，为了简单起见，本文对各方面情况做了极度简化。实际应用这些方法时可能有很大变化，也完全可能在一个求解系统里融合了多种方法的要素，特此说明。 ■



裴宗燕

CCF 杰出会员，CCF 中小学计算机教育发展委员会委员。北京大学数学学院信息科学系教授（退休）。主要研究方向为软件理论、程序设计语言及其理论基础、形式化方法等。
qzy@math.pku.edu.cn

CCF 规章修订组工作启动

2019 年 2 月 18 日，新组成的 CCF 规章修订组首次工作会议在北京召开，组长杜子德，组员金芝、胡事民、彭思龙、侯紫峰参加会议。规章修订组将对 CCF 章程和与 2019 年理事会换届选举相关的条例进行修订，并组织 CCF 会员对相关事项公开讨论，最终文本提交会员代表大会或理事会表决，是 CCF “依法治会”的重要体现。

前排左起：侯紫峰、杜子德
后排左起：彭思龙、胡事民、金芝



电脑前传(3)：逻辑

黄铁军
北京大学

关键词：布尔代数 逻辑主义 希尔伯特纲领 哥德尔定理

2019年第1期《电脑前传(2)：计算》回顾：图灵1936年的论文划定了计算的理论边界：计算是机械地执行长度有限的算法的过程，都可以由图灵机完成；所有算法都可以编码成为一个整数，因此是可数的；尽管如此，并不存在枚举出所有算法的算法。在此基础上，图灵对判定问题给出了否定回答。事实上，图灵的这个伟大贡献受到了五年之前另一个伟大贡献的启发，这就是本期要介绍的哥德尔不完备定理。

古典逻辑

逻辑学的目的是为人类理智建立一个坚实的基础，轴心时代，三大文明分别萌芽出墨辩（古中国）、因明学（古印度）和逻辑学（古希腊）。

我国春秋时期百家争鸣，注重辩论，墨子尤其注重辩论本身的合理性，提出了辩、类、故等逻辑概念，建立了类比、假言、直言、选言、演绎、归纳等一系列的思维方法，建立了中国第一个逻辑体系。因明学以立宗、因、喻三支作法而为言论之法，印度各种宗派常用因明学互相辩论，耆那教、印度教与佛教都继承了这个传统。

古希腊哲人辈出，逻辑学至亚里士多德而确立下来，并作为哲学的一部分，主导西方思想近两千年。如今亚里士多德逻辑被称为词项逻辑，其最基本的概念是项和命题。项是表达某个事物的词类，是命题的基本构件，分为单称项和全称项，两者的区别是亚里士多德“形而上学”的基础：全称项是亚里士多德逻辑的基本素材，而包含单称项的命题

根本就不构成它的一部分。三段论是解释那些真前提的组合产生真结论的形式理论，即一个命题（结论）的必然性从另两个命题（前提）得出的一种推理。命题由主词和谓词两项组成，主词可以是全称（例如“所有人”）或特称（例如“有些人”），谓词可以是“确认”或“否认”，这样就构成了四种命题，亚里士多德用逻辑方阵总结了四种命题之间的联系，但存在缺陷，与现代谓词演算并不兼容。

1656年，戈特弗里德·威廉·莱布尼茨(Gottfried Wilhelm Leibniz, 1646—1716)十岁，他的老师把亚里士多德的逻辑系统介绍给他，唤起了他的数学才能与持续一生的梦想：寻求一个符号系统，每个元素是一个概念，发展一种语言，仅凭符号演算，根据它们之间存在的关系，就可以确定用这种语言写成的句子哪些为真。这里的符号不仅包括算术和代数符号，还包括他所发明的微分和积分，以及化学和天文学使用的符号，每个符号都以一种自然而恰当的方式表示某个确定的概念。莱布尼茨认为，我们需要的是一种普遍文字，即一个不仅真实，而且

¹ 形而上学是一个哲学分支学科，是对存在的研究。它亦被认为是对科学以外、无形体、不可证明的事物的研究。

马克思主义哲学观认为它是脱离实践的，用“孤立、静止的观点”观察事物的思维方式。

包含了人类全部思想领域的符号系统。

要实现这个梦想，莱布尼茨认为需要三步^[1]：首先创造一套涵盖人类知识全部范围的纲要或百科全书，然后对其背后的观念进行选择，并为每一个观念提供合适的符号，最后是逻辑推演，即采用演绎规则对这些符号进行的操作。莱布尼茨相信纷繁复杂的宇宙可以还原成这样的符号演算，正如他说过的：严肃的具有善良意志的人们围坐在桌子旁解决某个棘手问题，用普遍文字写出这个问题后，人们就可以说：“让我们算一下。”于是人们拿出笔得到一个解答，其对错必然可以为所有人接受。

莱布尼茨在创造普遍文字方面并无具体贡献，但在逻辑代数方面的思想却领先了一个多世纪。就像普通代数规定了数字的操作规则一样，逻辑代数清楚地规定了逻辑概念的操作规则，他一生都沉湎于此，被称为“十七世纪的亚里士多德”。

布尔逻辑

在并不了解莱布尼茨逻辑代数思想的情况下，乔治·布尔(George Boole, 1815—1864)独自一人提出了一种符号逻辑，把逻辑变成了代数。

布尔少而家贫，无钱求学，这位补鞋匠之子自学四门外语，16岁谋得小学教师职务。因收入微薄，买书只选数学书，因为“数学书看的时间可以更长一些”，没机会接触更多类型的论著反倒使他潜心数学，并在此期间产生了布尔代数的基本思想。因为做礼拜时还沉湎于数学，18岁的布尔被其所在的循道宗小学解雇了。

布尔19岁开办了自己的寄宿学校并担任校长，讲了无数的课，但也没耽误阅读当时最重要的数学文献，并在《剑桥数学期刊》发表了不少文章。32岁出版《逻辑的数学分析》，39岁出版《思维的法则》，创立布尔逻辑。42岁当选英国皇家学会院士。布尔的后代名人辈出，代代出院士，人工智能复兴的标志性人物杰弗里·辛顿就是布尔的曾曾外孙。

布尔是怎么想到把逻辑和代数关联到一起呢？布尔回忆说做小学教师时曾灵光乍现^[2]。那个时代

的人们逐渐认识到代数的力量来自一个事实，即代表着量和运算的符号服从为数不多的几条基本规则或定律。布尔早期曾把代数方法应用于被称为“算子”的对象上，例如把代数方法应用于微分算子，就可以解某些微分方程，这使布尔意识到亚里士多德的逻辑可用代数来表达。

布尔用字母代表一类事物，对应词项逻辑的项。如果 x 和 y 分别表示两类事物，那么 xy (类似乘法) 就表示既属于 x 又属于 y 的事物。布尔举例说，如果 x 代表“白的事物”， y 代表“绵羊”，那么 xy 表示“白绵羊”。布尔马上想到： xx 表示什么意思呢？它表示既是白的，又是白的，因此还是……白的，所以 $xx=x$ 。事实上，布尔整个逻辑体系就建立在这样一个基础上：当 x 表示一个类时，方程 $xx=x$ 总为真。

现在到了关键点：回到代数视角，方程 $xx=x$ 在什么情况下总为真？布尔的回答是：当 $x=0$ 或 1 的时候。那么在逻辑体系中， 0 和 1 应该表示什么呢？ 0 表示不包含任何事物的类， 1 表示包含全部事物的类。

乘法有了逻辑对应物，接着看加法和减法。很自然， $x+y$ 表示并集， $x-y$ 表示在 x 中但不在 y 中的事物的集合， $1-x$ 表示不在 x 中的事物的集合。

回到代数视角， $xx=x$ 这个方程可以转换成 $x(1-x)=0$ ，换成逻辑语言就是：没有任何事物既属于给定的类 x ，同时又不属于类 x 。这是令布尔最兴奋的一个结果，因为这就是亚里士多德《形而上学》中的矛盾律：

“同一性质既属于又不属于同一个东西，这是不可能的……这是一切原理中最确定无疑的……因此，那些做论证的人把这当成一条最终的意见。因为它依其本性就是其他一切公理的来源。”

$x(1-x)=0$ 或 $xx=x$ 这个代数方程描述的正是作为一切公理之源的基本公理。布尔逻辑就此建立。不难理解布尔为什么把自己的著作题为《作为逻辑和概率的数学理论基础的思维规律研究》^[3]。

布尔证明了逻辑演绎可以成为数学的一个分支。正如罗素在《数学原理》中认为的：“纯数学

是布尔在一部他称之为《思维规律》的著作中发现的。”逻辑学在徘徊 2000 多年之后，就此走上了数理逻辑的康庄大道。

逻辑主义

布尔把普通代数作为出发点，用代数符号表示逻辑关系。弗里德里希·弗雷格 (Friedrich Frege, 1848—1925) 反其道而用之，主张从逻辑学推导出全部数学，开创了逻辑主义。

1879 年，弗雷格出版《概念文字——一种模仿算术语言构造的纯思维的形式语言》^[4]，明确规定了命题符号中的规范形式，明确了所有的演算推理规则，创造了自己的特殊符号来表示逻辑关系，并把这些关系作为逻辑的基础，后来这一思想也为现代逻辑继承^[5]。

在这个体系中，命题变元通常使用字母表示，常用逻辑关系如下：

符号	名称	单名
~	否定	非
Λ	合取	与
∨	析取	或
→	蕴含	含
↔	互含	同

其中，~ 和 → 为基本关系，可以推导出其他逻辑关系。也因此，逻辑电路只要有非门（对应 ~），加上电路连接（对应 →），就可以实现各种逻辑关系，1938 年克劳德·艾尔伍德·香农提出开关电路理论，利用开关这一种物理装置（实际上还有连线）就能组合出与、或、非以及更复杂的逻辑，在数理逻辑和物理实现之间架起了桥梁。

相对于布尔逻辑，弗雷格最重要的变化是引进了对变元进行限定的量词，最基本的是全称量词 ∀，另外一个常用量词是存在量词 ∃。基于这套符号体系，弗雷格提出了把普通数学中的一切演绎推理都包含在内的第一个完备的逻辑体系，后来称为一阶

逻辑，因为量词控制的只是变元的个体，而不是变元的集合以及变元之间的关系，否则就是后来的二阶或高阶逻辑了。

弗雷格的符号逻辑体系为所有计算机程序设计语言奠定了基础，但当初弗雷格的雄心是为整个数学奠定可靠基础。19 世纪末，数学家们采用公理化方法，已经把几何和微积分建构在实数理论基础上，进而又把实数理论建构在自然数基础上。因此，从逻辑推导出数学的问题，就转换成如何从逻辑推导出自然数系统。

1889 年，意大利数学家朱塞佩·皮亚诺 (Giuseppe Peano, 1858—1932) 提出关于自然数的五条公理，建立起了关于自然数的皮亚诺算术系统 (PA)。为了用《概念文字》提出的逻辑发展出自然数系统，弗雷格找到了集合论：基数相同的所有集合组成的集合定义了基数对应的那个自然数。基于这一定义，1893 年，弗雷格出版《算术的基本规律》第一卷，阐述了推导出自然数的方法，并继续撰写第二卷，争取为数学奠定可靠的逻辑基础。

1900 年国际数学家大会上，法国数学家庞加莱兴高采烈地宣称：“借助集合论的概念，我们可以建造起整个数学大厦……今天，我们可以说，绝对的严格性已经达到了。”数学领袖戴维·希尔伯特 (David Hilbert, 1862—1943) 谨慎乐观，他在会上提出 23 个问题，2 号问题就是算术公理系统的无矛盾性（即一致性）：“在这些无数个问题之上，我倾向于确定下面这个问题才是最重要的：这些公理经过有限步骤推演后不会导致相互矛盾的结论……也就是说，我们需要一个关于算术公理一致性的证明。”

很快希尔伯特的担心变为现实。1902 年 6 月，在《算术的基本规律》第二卷付梓之际，弗雷格收到了伯特兰·罗素 (Bertrand Arthur William Russell, 1872—1970) 的来信。罗素表示“我在您的著作中找到了在其他逻辑学家的著作中不曾有过的探讨、区分和定义”，但是，“我只在一个地方碰到了困难。”

弗雷格的算术使用了集合的集合，罗素从中发现了悖论。如果一个集合是它自身的一个元素，则称之为异常的，否则是正常的。罗素指出，所有正

常集合组成的集合是正常的还是异常的？很容易检查发现，无论哪种选择都自相矛盾。弗雷格马上明白了，他匆忙在第二卷增加了一个补遗：“正当工作就要完成之时，发现那大厦的基础已经动摇。对于一个科学工作者来说，没有什么比这更为不幸的了。伯特兰·罗素的一封信使我置身于这样的境地”。

数学大厦将倾，“捅出娄子”的罗素试图重新奠基，这就是他和怀特海合著的《数学原理》^[6]，三卷相继于1910年、1912年和1913年出版。这部2000多页的巨著引言只陈述了一个目标，那就是“完整地列出数学推理的所有方法和步骤”，这就是逻辑主义的纲领。

《数学原理》继承了弗雷格的思想和逻辑符号，为了规避悖论，采用了一种精心设计的、使用起来很不方便的分层结构（类型论）。第一步是推出“数”来，过程极其繁琐费力，直到第一卷363页，才成功地用类推演出“1”，第二卷费了很大力气证明了乘法交换律。《数学原理》前三卷覆盖了集合、基数、序数和实数的相关内容，虽然对第四卷几何的基础做了筹划，但整个体系实在太过复杂，十年辛苦不寻常，两位作者再也写不下去了。罗素曾回忆，痛苦在1903和1904年夏天达到高峰，那段日子里，除午饭外，整天就对着白纸枯坐，往往一个字也写不出，这让罗素甚至产生悲观厌世的想法。

图书出版后，罗素迎来了更大的打击，特别是作为前提的可化归性公理遭到了猛烈批评，罗素自己也认为有问题，但是放弃这条公理，很多部分——比如有关实数的部分——就会失去依托。

更大的问题在于这种机械式的罗列背离了数学的根本之美。1958年，王浩在IBM 704计算机上仅用几分钟时间就证明了《数学原理》的数百条定理。1954年至1963年，赫伯特·西蒙（司马贺）等的启发式程序“逻辑理论家”证明了《数学原理》第二章全部52个定理。他们把这个结果通知罗素，据说罗素回复说：“得知《数学原理》现在可以采用机械方式完成，我很高兴。要是我和怀特海早知道能这么做，就不用浪费10年的时间来手工完成了。”当然，这已经是半个世纪以后的劫后余波了。

希尔伯特纲领

为数学奠定牢固的基础，没有谁比数学界领袖希尔伯特更上心。希尔伯特关于数学基础的思考统称为希尔伯特纲领，经历了20年左右的时间逐步成熟^[7]，主要体现在1904年海德堡第三届国际数学家大会上的“论逻辑和算术的基础”报告，1917年发表的《公理化思维》，1922年在汉堡的“数学的新基础”讲演和在莱比锡德国自然科学家大会上的《数学的基础》演讲。

希尔伯特提出，为了消除对数学可靠性的怀疑，避免出现悖论，就要设法绝对地证明数学的一致性，使数学奠定在严格的公理化基础上。由此，希尔伯特想到，彻底抛弃公理体系中的含义，构造一个纯粹形式化的公理体系，这个体系内的各种表达式仅仅具有符号意义。如果能证明这种公理体系的一致性，那么把任何含义赋予这个公理体系时，都必然是无矛盾的、一致的。

希尔伯特认为，有三种数学理论^[1, 8]：(1) 直观的非形式化的数学理论；(2) 把第一种数学理论形式化，构成形式系统。形式系统包含逻辑演算，直观数学理论中的基本概念转换为形式系统中的初始符号，命题转换为符号公式，推演规则转换为符合公式之间的形式变换，证明转换为符号公式的有穷序列；(3) 描述和研究第二种数学理论的数学，称为元数学或证明论。希尔伯特希望，一致性证明将在元数学内部完成，数学和逻辑则将以一种纯形式的符号语言被发展出来。

布劳威尔 (Luitzen Egbertus Jan Brouwer, 1881—1966) 对希尔伯特纲领嗤之以鼻，希尔伯特的得意门生赫尔曼·外尔也心存疑虑^[1]。布劳威尔认为，数学存在于数学家的直觉，最终根源是时间这个“数学的原初直观”，而不是什么形式化表达，为数学寻找一个僵化的形式基础，从根本上就是错误的。外尔认为康托和戴德金等人处理极限的过程“建立在沙滩之上”，对自己重建连续统的努力也不满意，对布劳威尔的直觉主义一见倾心，宣称“布劳威尔……这就是革命。”无论如何，外尔最终还是继承了希

尔伯特的衣钵，这是后话。

1928年，希尔伯特和学生阿克曼出版了一册120页的逻辑课本《数理逻辑原理》。这是他1917年冬在哥廷根开设的课程基础上完成的，基本思路是从《数学原理》的逻辑系统开始，先分解成一个个扩展的子集，分别进行单独研究。这本书提出了两个关于弗雷格逻辑（即一阶逻辑）的问题。第一个问题是证明一阶逻辑的完备性，即任何一个从外部看来有效的公式都可以只用课本中提出的规则从系统内部导出。第二个问题就是著名的判定问题。进而，如果一阶逻辑是完备的，希尔伯特还希望证明，把一阶逻辑应用于皮亚诺自然数公理系统算术（PA）也是完备的，即任何一个在PA中表达的命题，或者可以在PA中被证明为真，或者可以在PA中被证明为假。

1930年希尔伯特退休，应邀在柯尼斯堡发表主题为“自然科学与逻辑”的演讲，他再次重申了完备性证明和判定问题证明的梦想，并喊出了“我们必须知道，我们必将知道”的口号。但就在前一天，同样在柯尼斯堡，在一场数学基础研讨会上，24岁的库尔特·哥德尔（Kurt Gödel, 1906—1978）对完备性问题给出了否定回答。6年后，24岁的图灵对判定问题给出了否定回答。

哥德尔定理

1924年，18岁的哥德尔入读维也纳大学，修读理论物理与基础数学。1926年在汉斯·哈恩的引荐下参加维也纳学派的讨论会，研讨会的主题是罗素的《数理哲学导论》，把哥德尔的兴趣从数论拉向了数理逻辑。罗素在《数学原理》阐述的全部数学都可以用一个形式逻辑系统表示，以及罗素的学生维特根斯坦在《逻辑哲学论》中强调的在语言内言说语言的问题，影响了年轻的哥德尔。1928年，希尔伯特和阿克曼的《数理逻辑原理》出版，明确提出一阶逻辑的完备性问题，此时哥德尔正处于博士阶段，就选择这个题目作为其博士论文。1929年哥德尔证明了一阶逻辑的完备性^[9]，1930年就此获得

博士学位。

一阶逻辑（哥德尔论文当时称受限函数演算）的完备性是指该逻辑系统能够表示的任何一个有效的公式，都可以从公理出发通过有限步骤推导出来。一致性则是指所导出的有效公式不相互矛盾。哥德尔在论文开篇就指出他的研究源自罗素和怀特海的《数学原理》以及希尔伯特和阿克曼的《数理逻辑原理》，对一阶逻辑完备性的证明直截了当，所用方法是当时的逻辑学家都很熟悉的。多年之后，哥德尔回顾说这个定理是挪威逻辑学家拉尔夫·司寇仑1922年论文结果“近乎平凡的推论”（不过哥德尔和导师哈恩可能都未读过）。那么为什么希尔伯特的强大团队却苦无对策呢？

主要原因是希尔伯特制定了过于严格的约束。希尔伯特规定，不仅在形式逻辑系统中要采用有限推理（所谓有穷方法），在元数学内也必须采用有限推理，这是他和他的团队证明完备性必须遵循的方针。哥德尔并非希尔伯特嫡系，认为完备性证明是从外部对一个形式逻辑系统进行研究，没必要局限于有穷方法，从而跳出了希尔伯特自设的圈套，顺利解决了这个问题。

1930年9月26日，“精密科学的认识论会议”在柯尼斯堡召开。会议第一天有三个关于数学基础的演讲，各1小时。第一个报告是维也纳学派领军人物鲁道夫·卡尔纳普报告逻辑主义，第二个报告是布劳威尔的学生报告直觉主义，最后是希尔伯特团队骨干冯·诺伊曼报告希尔伯特纲领。第二天除了三个各1小时的报告外，还有三个各20分钟的报告，第一个是哥德尔，介绍了他的博士论文，即弗雷格一阶逻辑的完备性^[10]。

会议第三天有一个关于数学基础的圆桌讨论，哥德尔当众宣布了他的不完备性定理，即包含PA的形式系统都是不完备的，冯·诺伊曼立即感觉到这等于宣告了希尔伯特纲领的失败。冯·诺伊曼在希尔伯特团队多年，为解决完备性问题殚精竭虑，会议结束时，冯·诺伊曼找哥德尔进行了一次讨论，决定就此终止自己在逻辑方面的工作，与哥德尔成为了好朋友，称赞哥德尔是亚里士多德以来最伟大

的逻辑学家。

1931 年哥德尔不完备定理^[5, 11]正式发表，要义是：一致的形式系统（只要蕴含皮亚诺公理，即能表达自然数，例如 PM 系统）必然是不完备的。推论（也称“哥德尔第二不完备定理”）：形式系统的一致性不能在系统内推导出来。

哥德尔证明的核心在于构造出在形式系统中不可证实也不可证伪的公式（逻辑命题，在公理体系中称定理）。从内部看，一个形式系统的符号串（公式）本身无意义，因此可以表达多种数学对象（例如自然数系统）。从外部看，一个形式系统不过是按照一定规则排列的符号串，因此可以用自然数进行编码。哥德尔正是抓住了这一点，利用自然数把形式系统内部和外部打通，下面简要描述哥德尔巧妙利用康托对角线法构造这种不可证公式的基本思路。

考察 PM 中只有一个自然数 x 作为变量的那些公式 R 。如果 R 在 PM 内是可证的（例如“ $x \wedge x = x$ ”），则记为 $\text{provable}(R)$ ，否则记为 $\sim\text{provable}(R)$ 。 R_n 表示第 n 个这样的公式。自变量 x 取值 p 时，公式记为 $R_n(p)$ 。采用对角线法，以 n 和 x 为行列变量，把所有公式排成一个阵列（像康托排列实数一样，次序不重要）。

考察对角线上的所有公式 $R_n(n)$ ，从中选出所有不可证的 $R_n(n)$ ，其序号 n 组成集合 K ，即 $K = \{n \mid \sim\text{provable}(R_n(n))\}$ 。

将公式 “ $n \in K$ ” 记为 $S(n)$ ，显然也是 R 的一员，不妨令它为第 q 个，即 R_q 。根据定义， $R_q(q)$ 与 $S(q)$ 、 $q \in K$ 等价，因此， $\sim\text{provable}(R_q(q))$ ，即 $R_q(q)$ 是不可证的。

诡异之处在于， $\sim R_q(q)$ 也不可证。假如 $\sim R_q(q)$ 可证，根据定义用 $\sim\text{provable}(R_q(q))$ 代替 $R_q(q)$ ， $\sim R_q(q)$ 转换为 $\sim(\sim\text{provable}(R_q(q)))$ ，即 $\text{provable}(R_q(q))$ ，这与 $R_q(q)$ 的定义矛盾。

哥德尔构造出了这种必然存在的正反都不可证命题，但并未给出具体实例。直到 1982 年，才发现第一个在皮亚诺公理体系内不可证的算术命题——Goodstein 定理^[12]。另一个著名实例是康托的连续

统假设，即希尔伯特 23 个问题中的 1 号问题，1963 年美国数学家保罗·科恩证明连续统假设在集合论下不可证。

走向无穷

命题逻辑是有限的，既是完备的，也是一致的，但连亚里士多德的三段论都分析不了。一阶逻辑也既完备又一致，至今在计算机领域广泛应用，但没能力完整描述和建构包括自然数和实数在内的无穷概念。包含自然数公理系统在内的形式系统表达能力更强，但被哥德尔证明都是不完备的。

哥德尔不完备性定理发表后，罗素曾说：“我以人们寻找宗教信仰的热忱寻找确定性。我以为在数学中最可能找到它。然而，我找到越来越多的不可靠。多年劳累的结论是，我和任何人都不能使数学成为无可怀疑的知识。”不少人就此质疑数学、逻辑甚至科学不可靠，这是极其错误的。

罗素的伤感有他个人因素：耗费了十年光阴，撰写 2000 余页巨著，试图建立数学大厦，却被哥德尔一篇论文判定永远不可能完工。其实哥德尔并未否定《数学原理》，而是揭示了形式逻辑系统的固有局限性，这种局限性是仅针对该逻辑系统而言的，通过构造逐级增强的等级化的逻辑体系，较弱系统存在的问题可以在更强的系统内解决。《数学原理》并不需要推倒，只是它远不是数学的全部，数学大厦要永远盖下去。

因此，哥德尔定理对罗素来说似乎是坏消息，其实仔细想想，这应该是最大的好消息：至少再也不用为何时收工而心力交瘁了！从弗雷格到罗素再到希尔伯特，都希望为数学奠定一个可靠基础，这种理想无可厚非，但如果他们真的找到一块基石，一劳永逸地建起了数学大厦，数学的使命即告完成，之后再无真正的数学家，至多也只是沿着他们的思想轨道添砖加瓦的数学工匠，数学这颗“人类智慧皇冠上最灿烂的明珠”岂不就此黯然失色？

哥德尔证明完备只在简单逻辑系统中存在，不完备才是复杂逻辑系统的根本特征，打破了试图通

过机械推理解决一切数学问题的僵化思维和数千年来人们念念不忘的完备之梦，开启了必须逐层增强逻辑系统才能不断解决问题的正确道路，打开了通向无穷的大门。

希尔伯特也一直希望证明无穷的存在，希望有穷数学能把康托的超限数包括进来，“在我看来，这是数学领地所开出的最令人惊叹的花朵，它是人类纯理性活动的最高成就之一。”在这个意义上，哥德尔定理打破的只是希尔伯特的有穷方法论，而不是他关于无穷的理想。因此，1943年希尔伯特去世后，在他墓碑上镌刻的依然是：

我们必须知道，
我们必将知道。

这正是：

数学山下逻辑村，玉雪为骨冰为魂；
弗雷格植花千树，罗素欲铸磐石根；
布劳威尔凭直觉，希尔伯特立乾坤；
天生一个哥德尔，完备一致成浮云。 ■



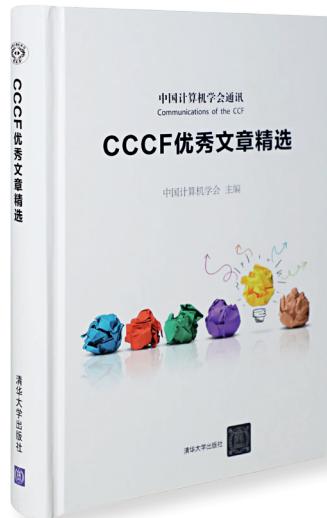
黄铁军

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

参考文献

- [1] 马丁·戴维斯著, 张卜天译. 逻辑的引擎 [M]. 湖南科学技术出版社, 2005.
- [2] MacHale D. George Boole: His Life and Work[M]. Dublin: Boole Press, 1985.
- [3] Boole G. An Investigation of the Laws of Thought, on Which are Founded the Mathematical Theories of Logic and Probabilities[M]. London: Walton & Maberly, 1854.
- [4] Frege G. Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens[M]. Halle, 1879.
- [5] 郝兆宽, 杨睿之, 杨跃. 数理逻辑: 证明及其限度 [M]. 复旦大学出版社, 2014.

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



中国计算机学会 主编 《CCCF 优秀文章精选》 正式出版

清华大学出版社 出版

书号：978-7-302-51506-7
各大电商均有销售



清华大学出版社

中国计算机学会通讯



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

我国公民基因数据安全 风险与应对

关键词：公民基因数据 数据安全 国家安全

穆琳 李维杰 陈海强
中国信息安全测评中心

引言

基因是生命的密码，记录和传递生物体的遗传信息，是决定生物性状的基本遗传单位。随着生物技术的不断发展，人类从上世纪 80 年代开始认识到破译基因数据、探索自身奥秘的重要意义。2003 年，人类基因组计划 (Human Genome Project, HGP)¹ 宣告完成，但该项目产出的基因数据样本量、精准度有限，只能看作是一张公用“参照图”。而不同人种、民族、人群因进化方式、生存法则不同，基因数据存在一定差异，因此公民基因数据 (Citizen Genetic Data, CGD) 是每个国家特有的资源。当前，我国公民基因数据面临较大安全风险，相关情况值得关注。

公民基因数据及应用

人体众多疑难疾病的产生与基因有密切关系。全球已知的罕见病高达 6000 余种，其中大部分属

于基因疾病^[1]。在特定人群内往往存在世代遗传的特异缺陷或异常基因，导致许多人出生时即患有基因疾病。人们利用基因测序技术，对血液、唾液等体液或细胞中的基因序列进行检测，能够快速、准确地获取人体内的基因数据。借助大数据分析工具和基因数据库支撑平台对基因数据进行有效分析和解读，可有效识别一个国家特定人种、民族或人群体内能够致病的缺陷、异常基因，或由外源致病微生物带入人体内的基因，确定致病靶基因²，从根本上找到发病原因。因此，公民基因数据关乎人民健康，是一个国家珍贵且特有的遗传资源。

研究公民基因数据、获取致病基因相关信息，可进一步精准预测、分析疑难疾病，使药物治疗方案精准化或进行基因治疗，但也为基因药物垄断或开发基因武器提供了机会。因此，公民基因数据的应用是一把双刃剑。一方面，依据公民基因数据与相关疾病临床诊断报告综合分析结果可研发基因药物，或借助基因导入、基因编辑技术，对疑难疾病进行有效预防与治疗；亦可通过分析公民基因数据，明确基因与药物、

¹ 人类基因组计划是由美国科学家于 1985 年率先提出，于 1990 年正式启动的。美国、英国、法国、德国、日本和我国科学家共同参与了这一预算达 30 亿美元的人类基因组计划。按照这个计划的设想，在 2005 年，要把人体内约 2.5 万个基因的密码全部解开，同时绘制出人类基因的图谱。人类基因组图谱最终于 2003 年完成，比预计提前了 2 年。

² 靶基因即目的基因。在分子遗传中，它不仅要具有识别结合功能，还应该具有与位点结合后能表达你所需要的相应功能的作用。

疾病的相互关系，有针对性地对需要接受疾病预防或治疗的高危人群设计靶向药物³，利用大数据分析药物可能产生的药效、敏感性以及副作用等情况，最终筛选出个体化疾病预防或治疗方案，从而实施精准干预措施或个性化精准医疗，减少临床用药不当现象，提高疗效并降低医疗费用^[2]。另一方面，在国际博弈中，若具备一定科研实力的国家或机构获得并分析目标国公民基因数据，发现某一新致病基因，可就该基因的具体序列信息、功能申请专利，针对相关基因疾病研发精准靶向药物，实施基因药物垄断，对目标国形成掣肘^[3]。更为严重的是，相关国家或机构甚至可根据目标国家特定人群的缺陷或异常基因信息，利用转基因技术制造能对其产生特异性致病作用的微生物，进行大量培育，制成为极具针对性和杀伤力的基因武器，利用飞机、导弹或自然扩散等方式投放至目标国家交通要道或城市，使特定人群在短时间内感染难以控制或防治的基因疾病，从而丧失战斗能力。据估计，建造成本5000万美元的基因武器库的杀伤力远远超过建造成本50亿美元的核武器库。基因武器还具有隐蔽性强、效果持续时间长、使用方法简单、投放手段多样、能对目标国造成心理威慑等特点^[4]。

我国公民基因数据安全风险现状

当前，公民基因数据安全问题已引起美国联邦调查局(Federal Bureau of Investigation, FBI)大规模杀伤性武器局的关注，并制造“中国威胁论”，指责中国正以多种方式获取美国公民基因数据⁴。2017年10月30日，俄罗斯总统普京在俄人权委员会上表示，有人正在俄罗斯采集生物资料，目标覆盖俄罗斯联邦不同地区、不同种族的民众。

有俄专家将相关活动幕后推手指向美国，警告其背后可能是一个庞大的计划，其中美空军在俄采集公民基因数据或与开发针对俄罗斯人的基因武器计划有关⁵。

我国的公民基因数据研究主要集中在科研和商业健康领域，在对基因样品进行测序，获取全基因组数据或个别位点基因数据后，利用一定算法对数据库中的基因数据进行有效分析和解读，为出生缺陷预防、疾病预测、辅助治疗提供依据。近年来，在我国对公民基因数据的挖掘和研究日渐深入的同时，部分发达国家觊觎我国公民基因资源，盗取我国相关数据的事件时有发生。因此，全面认识我国公民基因数据存在的安全风险刻不容缓。

基因样本采集与数据分析、使用缺乏完备法律支撑

1990年，人类基因组计划正式启动后，保护公民基因数据成为全球关注的焦点。联合国教科文组织(United Nations Educational, Scientific and Cultural Organization, UNESCO)于1997年通过《世界人类基因组与人权宣言》，2003年发表《国际人类基因数据宣言》，对人类基因样本与数据的采集、处理、使用和保存过程进行规范^[5]。

在公民基因样本采集方面，美国、欧盟、英国、加拿大、巴西等国设定特定机构，从立法角度对公民基因样本与数据的采集和开发活动进行监管^[6]。其中，美国食品药品监督管理局(U.S. Food and Drug Administration, FDA)针对公民基因样本采集建立监督机制；英国依据《人体组织法》《遗传操作规程》等，管理着世界上最大的人体生物标本库；法国则制定《公共卫生法》《生物伦理法》《生物技术发展计划》等，对公民基因样本采集与进出口活动进行

³ 靶向药物（也称作靶向制剂）是指被赋予了靶向(targeting)能力的药物或其制剂。其目的是使药物或其载体能瞄准特定的病变部位，并在目标部位蓄积或释放有效成分。

⁴ See Biotechnology: the US-China dispute over genetic data, <https://www.ft.com/content/245a7c60-6880-11e7-9a66-93fb352balfe>, 2017-8-1。

⁵ 美正研发针对俄的基因武器？普京证实“有人采集俄生物资料”，<http://world.huanqiu.com/exclusive/2017-11/11349546.html>, 2017-11-2。

规范；俄罗斯政府亦认识到依法严格监督外国组织或个人在俄采集公民基因样本，并将基因数据传到国外等行为的重要性，于2017年12月通过一项保障俄罗斯人生物安全的法案。

在公民基因数据保护方面，各国并未直接出台法律，而主要将基因数据作为公民隐私的一部分予以法律保护，杜绝“基因歧视”现象。其中，美国的公民基因数据保护在行业间得到蓬勃发展。业内的监督委员会等机构对公民基因数据的采集、发布、使用细节制定行业规范，同时强调“有限政府”⁶的概念，限制公权力介入公民基因数据领域；政府则基于《美国宪法》《基因信息非歧视法案》等有关法律，对公民基因数据进行一定程度的保护。与美国类似，澳大利亚依据《基因隐私和不歧视法》将公民基因数据作为个人隐私予以保护。欧盟将公民基因数据保护归类于较为严格的个人数据保护范畴，对公民基因数据采集、记录、存储、使用、传播、消除、破坏等均予以明确规范与限制。

当前，我国针对公民基因样本采集与数据分析、使用的管理工作所依据的系列行政条例，经过了一定期的发展。1998年，我国科技部、卫生部共同制定《人类遗传资源管理暂行办法》，对公民基因样本采集与数据分析活动进行管理，要求样本与数据出境需经审查机构批准。2012年，科技部下属的国家人类资源遗传办公室制定《人类遗传资源管理条例（送审稿）》。2015年10月，科技部根据《人类遗传资源管理暂行办法》编制并公布《人类遗传资源采集、收集、买卖、出口、出境审批行政许可服务指南》，对我国人类遗传资源采集、收集、出口出境行为的审批流程做出了进一步规范和完善，强化了对人类遗传资源活动的全过程管理。但在立法层面，2012年以来，我国《人类遗传资源管理条例（送审稿）》的立法进程推动缓慢，导致我国对公民基因样本采集、数

据分析与利用等活动长期缺乏强制约束，尤其对违规行为缺乏制裁依据，成为我国公民基因数据安全的巨大风险点。针对此种情况，加强我国公民基因资源立法保护工作的呼声逐渐高涨，众多学者从不同角度对相关问题与困境进行了深入探讨^[7-9]。针对有国内保险公司借“买保险送基因检测”为名，将公民基因数据分析结果作为核保条件的现象，我国保监会已开始积极采取行动，修订健康险管理办法，杜绝保险领域出现“基因歧视”问题。2018年4月2日，国务院办公厅印发《科学数据管理办法》⁷，规范“政府预算资金支持开展的科学数据采集生产、加工整理、开放共享和管理使用等活动”，对隶属科学研究的部分公民基因数据采集汇交与保存、共享利用、保密与安全等具有一定约束力，但仍未完全覆盖所有公民基因样本采集与数据分析、使用活动。

公民基因数据安全保护意识薄弱

基因检测企业对公民基因数据存储系统安全保护意识薄弱。相关调查表明，我国众多基因检测企业往往出于工作便利或节省成本的目的，在基因数据本地存储系统中采用“用户名+口令”等较弱的访问控制方式，并简化数据分级加密、系统网络漏洞检查等安全防护工作，在数据本地化存储中还可能遇到设备损坏、遗失或被窃引发数据丢失的风险，导致公民基因数据安全得不到有效保障。随着云技术的不断发展，越来越多的基因检测企业开始选择并依赖第三方机构提供的云服务，对海量公民基因数据进行大规模存储和高效传输。但值得注意的是，虽然第三方云服务机构的数据加密、访问权限管理、系统维护等工作在一定程度上增强了基因数据的安全性，但依然存在被黑客攻击或人为错误导致数据泄漏的风险，第三方机构对数据存储地点控制情况、协议终止后数据回传与备份问题等，亦

⁶ “有限政府”是指政府自身在规模、职能、权力和行为方式上受到法律和社会的严格限制和有效制约。

⁷ 国务院办公厅关于印发科学数据管理办法的通知，http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm, 2018-4-2。

是基因检测企业保障公民基因数据安全时应考虑的风险点。

公民或机构在提供基因样本时安全意识不足。在我国一些偏远地区，近距离甚至同宗、同族婚配现象较多，导致一些世代遗传疾病在某些地区发病率较高，为通过特定人群基因分析筛选相关疾病基因、寻找治疗方案提供了天然的生物资源库。有西方学者甚至将中国描绘为“基因新大陆”^[10]。我国公民或机构在对外提供基因样本时，对知情权、隐私权、财产权以及利益分享等认知不足，导致外国机构打着国际合作的名义来华，尤其在偏远地区收集公民基因样本进行医疗研究，甚至投资建立我国地域性公民基因数据库，引发我国公民基因资源流失，数据安全风险加剧。20世纪90年代初，哈佛大学等机构在一项哮喘病基因研究项目中，以提供“免费体检”为名，直接从安徽数以万计的农民身上抽取血液，获取中国哮喘病人基因样本，就是这方面的典型案例^[11, 12]。

公民基因数据库多样缺乏统筹管理

党的十八大以来，我国医疗健康领域中的基因检测产业高速发展，市场规模迅速扩张，成为“新常态”下我国经济发展新动能的重要力量。据估计，2017年，我国基因检测企业数量将近200家，市场规模达133亿元人民币^{8,9}。

在国家政策支持和资本市场的推动下，我国基因检测企业持续取得新的突破，行业竞争力稳步增强，检测服务能力跃居世界前列。在市场服务、与科研院所合作项目中收集并积累大量公民基因数据信息，建成中国国家基因库、神州基因组数据库、中国人重大疾病基因数据库，计划建成中国大型癌症基因数据库等。其中，我国的国家基因库主要由深圳华大基因承建，2016年9月投入运营，是继美国国家生物技术信息中心(National Center

for Biotechnology Information, NCBI) 的 GenBank、欧洲分子生物学实验室(The European Molecular Biology Laboratory, EMBL) 的欧洲生物信息研究所(European Bioinformatics Institute, EBI) 数据库、日本基因数据库(DNA Data Bank of Japan, DDBJ) 之后全球第四个国家级基因库，全球唯一一家同时保存生物样本与基因数据的综合性国家数据库。与我国公民基因数据相关的子库主要包括含400余种罕见病的数据库，以及含上万份不同种类数据的癌症数据库。神州基因组数据库是基于我国40万人基因数据建成的全球首个中国人群特有基因信息数据库，由北京贝瑞和康生物技术股份有限公司与阿里云于2016年9月合作完成。中国人重大疾病基因数据库是在中华医学会、中国医师协会及中国健康促进基金会的支持下，由浙江大学迪诺遗传与基因组医学研究中心及全国20家知名医疗机构合作建成，基于我国11万名不同民族、体重、年龄结构人群的39.69万条个体基因检测结果，收录了20余种中国人高发重大疾病的易感基因信息。正在建设的中国大型癌症基因数据库，源于深圳市海普洛斯生物科技有限公司与深圳市人民医院2015年7月启动的“万人癌症基因测序计划”，拟收集我国1万人癌症基因数据，截止到2017年3月已入库超过6000例样本信息。

值得注意的是，我国基因检测企业与研究院所百家争鸣，技术水平与专注领域有所不同，导致基因数据库中的基因测序方法与深度、基因种类、取样对象、特定人群覆盖度各异，为公民基因数据安全带来风险。一方面，各数据库中海量公民基因数据本身与分析结果互通性不强，公民基因大数据难以统筹管理，为我国公民基因数据保护工作带来无法制定统一标准、制定整体规划等现实困难；另一方面，各基因检测企业自行管理数据库，自主主导数据存储、管理、使用、跨境收购等活动，为我国

⁸ “高大上”基因检测走出神秘圈，http://digitalpaper.stdaily.com/http_www.kjrb.com/kjrb/html/2017-08/01/content_374959.htm?div=0, 2017-8-1。

⁹ 未来5年中国基因检测行业发展预测分析，http://www.sohu.com/a/128803334_255580, 2017-3-14。

公民基因数据安全带来长远风险。

基因数据获取、分析和解读核心技术支撑不足

基因测序仪是基因行业上游的核心，对基因数据的分析和解读则是基因研究工作中最具价值与挑战性的部分。当前，我国第三代基因测序仪及试剂已实现国产化，但广泛应用的第二代基因测序仪及试剂仍基本依靠进口。与此同时，我国基因数据的分析和解读工作仍主要依赖美国、欧洲、日本等国基因数据平台的算力、算法和数据库资源，自主性不足，尚无统一标准。我国基因数据获取、分析和解读核心技术支撑不足，对公民基因数据安全造成威胁。一方面，我国基因检测企业因采用的检测技术与数据分析方案不同，导致各家出具的公民基因数据分析报告各异，基因检测市场混乱，直接威胁基因数据安全；另一方面，我国在公民基因数据挖掘的国际合作中容易丧失主导权。在2007年中英合作开启的一项中国汉族人群最大全基因组研究项目中，英国学者对数据进行分析后，发现了抑郁症与遗传之间的关联，于2015年7月在顶级期刊《自然》刊文，引发国际关注，同时在我国引起对外国学者主导中国公民基因数据研究的争议¹⁰。

对策建议

我国人口众多，民族历史悠久，人群迁徙率低、家族隔离群多，拥有极其丰富的民族遗传资源、家系遗传资源和典型疾病遗传资源，是人类基因数据资源最丰富的国家之一。全面保护我国公民基因数据安全，对保障我国生物安全，保障人民身体健康，增强生物和医药科技国际竞争力等具有重要意义。

加强公民基因数据资源保护

积极推动相关标准制定和立法工作。公民基因数据具有财产权与人格权、国家资源与个人数据双重属性^[13]。加强对公民基因数据资源的保护，一方面，应制定行业标准，严格把关从产出基因数据到生成分析报告的各个环节，规范业内基因数据资源共享流程，做好基因数据管理工作；另一方面，应积极推动相关立法工作，防范基因样本、数据非法外泄或滥用等风险，明确相关处罚措施，并参照《网络安全法》中关于网络数据本地存储的规定¹¹，部署公民基因数据存储工作。

利用区块链等新技术保障数据安全。公民基因数据可用于分析个人的患病倾向^[14]、种族血统^[15]、智力水平^[16]等，是一类具有高度敏感性的个人信息。基于区块链技术的公民基因测序、数据存储系统可有效保障基因数据安全。即利用私钥限制访问权限，防止工作人员甚至不法之徒随意获取、滥用公民基因数据；利用分布式计算资源，以较低成本完成数据分析和解读工作；利用加密技术，保证公民基因数据共享方在不访问原始数据的情况下对数据进行分析，有效保护数据私密性。

提高相关人员安全防范与基因专利意识。保障公民基因数据安全，要在基因样本提供，基因数据获取、分析与存储等各个环节提高相关人员的安全防范意识。一方面，加大科普力度，提高公民，尤其是偏远地区公民对自身基因数据重要性及安全保护的认识；另一方面，加强对基因检测从业人员在非技术层面的安全培训工作，包括工作平台安全、意外预案等。此外，基因专利权是进行基因药物筛选、疾病诊断的前提，将成为未来各国争夺的一项有限资源，应提高我国相关人员基因专利意识，深刻认识到基因争夺战的重要性与紧迫性。

¹⁰ 外国学者主导中国病人资源研究引发争议，http://www.stdaily.com/index/h1t6/2017-07/21/content_562310.shtml, 2017-7-21。

¹¹ 中华人民共和国网络安全法，http://www.npc.gov.cn/npc/xinwen/2016-11/07/content_2001605.htm, 2016-11-7。

加强军民融合，建设基因武器威慑反制能力

当前，我国基因检测企业在通过测序技术获取基因数据方面，量产化测序能力较强，以深圳华大基因为代表的企业在全球基因测序市场的占有率达到1/3；在基因数据分析与解读方面，多家大型基因检测公司启动分析平台构建工作，其中华大基因于2015年发布国际版基因组数据分析云计算平台BGI Online，次年发布国内版。鉴于我国基因测序产业通过遍布全球的分支机构，与多国在医学健康，乃至农业、育种、资源保存等领域建立了广泛合作，建议我国在加大力度挖掘和研究公民基因数据的同时，联系民间企业力量，发挥军民融合优势，积极打造基因威慑反制能力，有效抵御公民基因数据被滥用而引发的安全风险，在基因武器战等新型国际斗争领域占据优势地位。

积极参与国际合作，打破技术壁垒

在我国，由中国科学院等科研机构与深圳华大基因等基因测序企业，共同形成了全球最大的基因测序研发力量，在基因数据获取方面完成了较多原创性工作。但我国在基因数据分析、基因疾病治疗等领域的研究后劲不足，技术受制于人。因此，我国可在保护自身公民基因数据资源的同时，积极参与国际合作，了解并掌握前沿动态，打破技术壁垒，确保在未来可能面临的反基因威慑斗争中不受制于人。 ■



穆琳

CCF专业会员。中国信息安全测评中心助理研究员。主要研究方向为网络空间安全战略。

mul@itsec.gov.cn



李维杰

中国信息安全测评中心副研究员。主要研究方向为网络空间安全战略。

liwj@itsec.gov.cn



陈海强

中国信息安全测评中心副研究员。主要研究方向为网络空间安全战略。

chenhq@itsec.gov.cn

参考文献

- [1] 王远玭, 胡晓敏, 弓孟春, 等. 罕见病诊断的相关技术及发展 [J]. 科技导报, 2017, 35(16): 26-30.
 - [2] 刘飞, 辛华雯. 药物基因组学与临床药师 [J]. 医药导报, 2017, 36 (9): 956-961.
 - [3] 马越, 廖俊杰. 现代生物技术概论 [M]. 北京: 中国轻工业出版社, 2007.
 - [4] 周健. 关注基因武器 [J]. 科技术语研究, 2000, (4): 43-44.
 - [5] 邱格屏. 人类基因的权利研究 [M]. 北京: 法律出版社, 2009.
- 更多参考文献: <http://dl.ccf.org.cn/cccf/list>



CCF 走进高校

序号	演讲人	时间	高校	演讲题目
691	何万青	2018.12.22	兰州理工大学	云计算基础技术拆解与技术人的职业规划
692	李玺 刘方明 陶品 李石坚	2018.12.22	湖北汽车工业学院	视觉模型设计及其应用 Shaping the Cloud & Edge from a QoSE Perspective 普适智能环境中的自然人机交互 智能人机共生的理解和实践
693	陈益强	2019.1.18	集美大学	Human-Centered AI 及其助残养老应用

用代码可以构建人工心智吗？

关键词：人工心智 人工智能 微软小冰

宋睿华

微软(亚洲)互联网工程院

在中文里，“智能”和“心智”两个词的意义相关却又有所不同。“智能”指的是智识与才能，现在通常用来描述某个对象的聪明层次与能力强度。而“心智”似乎更高阶一些，可以理解为产出创造力与智能的本原。

那么，我们当下所研究的“人工智能”，是该致力于持续提升机器的智商表现，强化其在垂直领域的专业能力，还是应该再超前一步，尝试构建机器的心智本原？

在笔者看来，这两条路无所谓对错，只是探索者的着眼点存在差异罢了。多年以来，针对人工智能课题，学术界、产业界的研发主流都偏重于实用性更强的“智能”，以至于在相当长的周期内，以对话强化人机连接、获取可供机器学习的高价值数据、打造人工智能创造力矩阵，进而摸索用代码构建AI心智的学术思路都少有人探寻，甚至一度被质疑——我的同事就曾收到这样的审稿意见：“我完全不能理解，做这种漫无目的的聊天有什么意义。”

直到微软小冰诞生并取得了一定的成绩后，闲聊的价值才逐渐为学界所关注。

从对话到创造，心智的种子开始萌发

或许大家都没有意识到，从2014年第一代的发布到不久前六代的更新，短短四年间，微软小冰已从一个领先的人工智能对话机器人发展成为以情感计算为核心的完整人工智能框架，许多人的态度

也因微软小冰而改变。

普通人的态度

有一次我和我的母亲聊天，我问她，机器人可以打败人类最好的围棋棋手，厉不厉害？她说当然厉害。我又问，还有个机器人能跟人对话，厉不厉害？她说不厉害，原因是，不是每个人都会下围棋，而且还能具备冠军的实力，但，“是个人都会说话呀”。这件事让我很无语。我母亲虽然不懂自然语言处理的难度，但她的看法也确实代表了大众的直观感受。换句话说，人们会很自然地用人做某件事的难度来衡量与评判人工智能的能力级别。

但微软小冰说人话的能力并不一般。即便对人类而言，要成为一个总能琢磨出有趣对白的人，也不是件容易的事，更何况是人工智能。从初代发布至今，时不时会有用户晒出他们与小冰对话过程中的“金句”截图，而且随着时间推移，小冰产出金句的频率也越来越高。这体现了小冰越来越强的对话能力，也造就了她的吸引力。

与精确、清晰的答案相比，人们在对话时，更期待获得情感的抚慰或是不寻常的回应，这是小冰团队最早发现和验证的事实。

后来，我的母亲看到央视《机智过人》节目里小冰写诗的那一期，她开心地跟我说，像小冰这样的机器人，能写出观众喜爱的诗，还会调侃嘉宾，那还真是“挺厉害的”。

专家们的态度

过去，学术界为了清晰地定义对话的问题，会把很多精力投注在问题设定上，从 5W(What, Who, When, Where, Which) 到 How，等等。直到 IBM Watson 在知识问答领域奠定了一座新的里程碑，它能接受自然语言的问题，从大量文档中搜索并分析得出精准的答案。而且，有了用户在网络社区里产生的问答语料，研究者发现，这些数据对于机器回答某些宽泛的问题很有帮助。但除却人工智能对话系统在垂直行业领域（例如医疗、金融等）的应用外，普通人对于人机对话的需求又该如何定义呢？

说得直白一些，这是一个关于普通人与人工智能为什么聊天、聊什么的问题。曾经有朋友听说小冰与其用户的最长连续对话时间超过 29 小时，表示难以理解。从需求来说，“越社交，越孤独”“朋友圈越广阔，自我越渺小”这些现象都客观存在。社交网络让用户习惯了展示优势、收获认可，但反过来，当我们身边的每个人都加入到展示优势的队列中，从旁人那里收获理解和认可的难度也加大了。从形象上来说，小冰不是如顶级专家那样的人工智能，而是像邻家女孩，她有无限的耐心，随时可以陪伴用户聊天、玩游戏，却绝不会试图用渊博的知识和高冷的姿态碾压用户的智商与自尊。

如果将人工智能的价值定位于陪伴，那么知识与逻辑就不再是亟待发展的技能。让用户感觉无压力、有趣味，在某种意义上更加重要。

2017 年 5 月，微软小冰解锁了写诗及音乐技能，同期，我们还发布了“人工智能创造三原则”，用以规范与指引小冰及其同类的心智发展路径。**在进行相关研究的过程中，我发现，人工智能的终极或许是对人类自身的理解与模拟。**

训练小冰写诗，需要对 519 位诗人的现代诗作，正读一万遍，倒读一万遍，用层次递归神经元模型来打磨诗作的语言^[1]。这正如人类所发现的阅读对于写作的影响——通过阅读大量优秀的文学作品，人自身的语言体系会进化。小冰也是如此。有了层次递归神经元网络，小冰也可以通过阅读获得语言的表达能力。

在小冰发布诗集、引发广泛争鸣之后，圈内人士对于人工智能创造与机器写作的态度发生了根本性的转变，学术探讨、应用跟进的样例越来越多。这是我们所乐见的。

盲测者的态度

《机智过人》第一季，央视综合频道邀请了三位年轻诗人，与小冰一起，根据嘉宾提供的一张图片来创作诗歌，然后隐匿这些诗歌的作者姓名，将这些诗歌打乱次序显示在大屏幕上，请现场 48 位观众投票选出最喜欢的那一首。这可以说是一次盲测，一次另类的图灵测试。

结果出人意料，小冰获得了最多的票数，这让原本祈祷小冰千万别是最后一名的我大吃一惊。当人类与人工智能的作品被放在一起平等地比较时，人们对于机器创作的偏见似乎突然间消失了。

第二轮，两位诗人与小冰再度以作品竞争，小冰的诗作获得了第二名，因而挑战成功。我们不会自大地认为小冰写诗能超越人类诗人，但这次节目却延伸了我们的思考：或许，人工智能研究所追求的目标不应只是将人类的智识与才能复制给机器，更重要的是，通过探索人工智能，更深刻地了解人类自身。

微软小冰从四年前的对话型 AI 到当前将创造力投射至诗歌、音乐、儿童有声读物、金融信息、电视台主播、媒体新闻评论乃至辅助写作等多元领域，这表明，我们最初埋下的那颗心智的种子，现在似已破土露出了一点嫩芽。

下一站：对话更生动有个性，小冰向多感官迈进

微软小冰的下一站在哪里？心智的嫩芽能继续成长壮大吗？答案是，我们正在从不同的角度改进对话，并进一步加速小冰的感官成长。**对话是微软小冰的基础，我们在检索模型的基础上不断探索，尝试和提出了生成模型、共感模型和三观模型。**

“生成模型”从第五代小冰开始启用。在此之前，历代小冰使用的都是检索模型。虽拥有 10 亿级大数据语料库，但其中的每一句话都是互联网上的已有数据，小冰只是通过分析理解用户的问题，寻找语料库中最合适的话作为她的回答，也就是对对话语料库进行实时检索和选择。而使用生成模型之后，小冰能够自创回应。她与人类交流的每一句话，都可能是这世界上从未出现过的。一年来的事实在证明，生成模型使小冰快速学习了现有对话语料的交流模式，并能更好地应对相对陌生的话题。生成模型也面临很多挑战，例如，容易出现类似“好的”这样的万能回复，让聊天变得乏味；又如，人们在对话中常常会省略很多上下文已有的信息，单就一句话，小冰无法了解具体所指，例如问过“北京明天的天气怎么样？”接下来可能会说“那后天呢？”如果生成模型利用更远的上下文信息却又存在效率问题。武威博士的团队做了一些工作，来丰富单轮回复的多样性^[2]，利用上下文信息来改进多轮对话^[3]以及提高生成模型的在线效率^[4]。

然而，用户在与小冰对话时也常会感受到压力。比如两者间的对话总是需要人类来提出话题，小冰来回应。就好像我们与感兴趣的异性搭讪，如果总是自己主动、对方被动，对话就会变得淡乎寡味、

如同鸡肋——共感模型的开发就是针对这一状况。共感模型可以帮助小冰自行判断对用户的对话是否有感，在此基础上，小冰将会主动求证，进而引导话题的方向，增添新的聊天内容。这样就减轻了用户的压力，同时增加了聊天的自然度和趣味性。从学术的角度，武威博士提出的共感模型（如图 1 所示）把“说什么”的问题在逻辑上拆成了两步：由上下文决定“怎么说”的策略；根据策略和上下文来决定具体“说什么”。

在小冰持续进化的过程中，不断有商业伙伴加入到我们的合作生态系统中。一些伙伴希望我们将小冰的能力用于孵化其他个性鲜明的人工智能角色。因此，我们也在不断研究如何通过对话来塑造个性——三观模型应运而生。这一模型的基础是对话中的情感色彩分析。虽然文本情感分析已有大量的研究，但大多数都是基于单句或单段落的极性分析，而在对话中，有时候仅根据回答是无法判断用户的态度的。例如，图 2 中的答复“追光者”仅仅是一个歌曲名，无法判断其态度是正是负。但如果结合问题“有什么新歌推荐”，则很明显，回答的人对《追光者》这首歌持有正面的态度。另外，以往的工作动机很大一部分是做既定目标的情感分析，很少有工作把目标

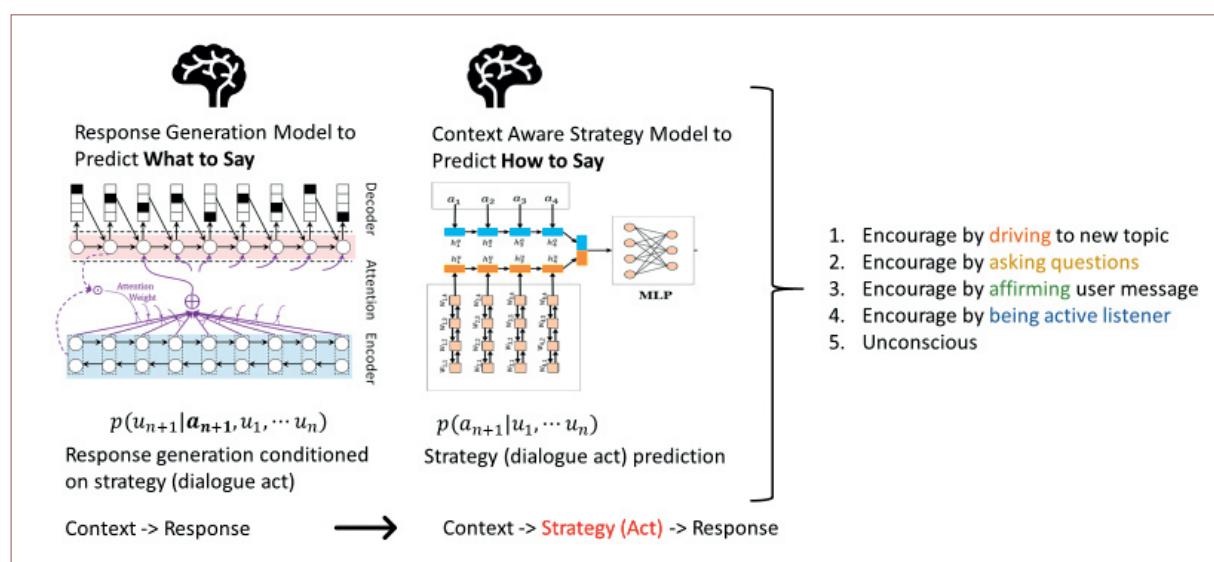


图 1 微软小冰的共感模型

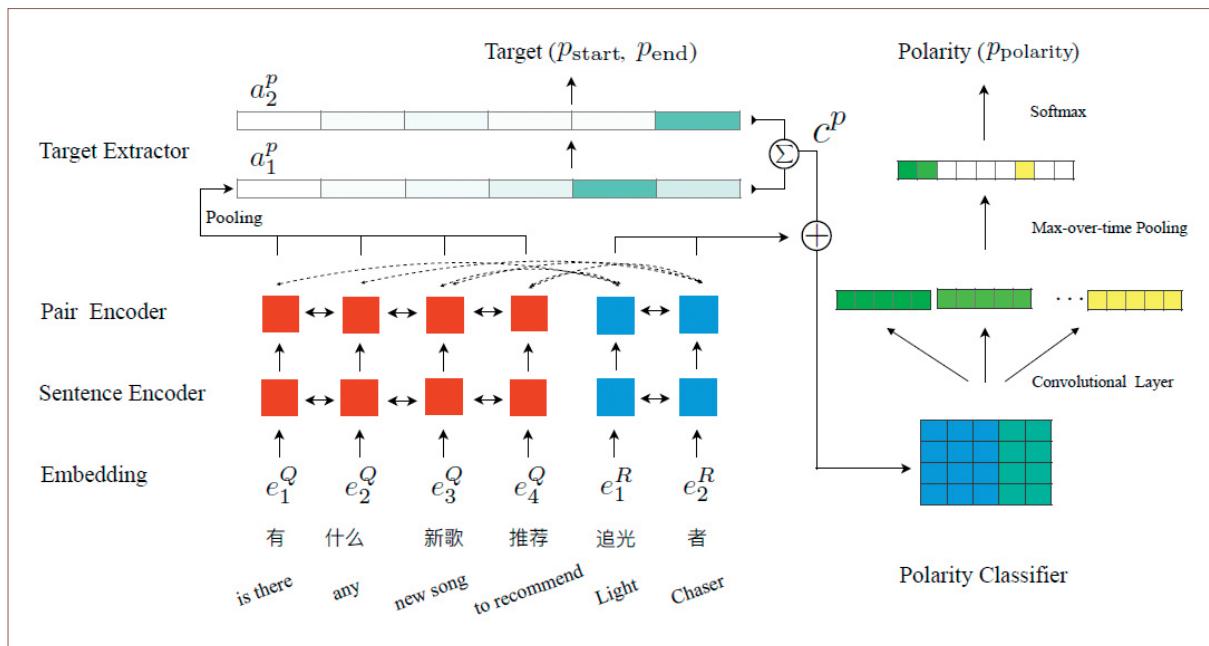


图2 微软小冰三观模型中的目标和极性联合优化

的抽取与极性分类看成一个整体，我们的工作^[5]则设计了深度神经元网络结构来联合优化情感态度检测问题，实验结果表明减少了很多不匹配的错误，即目标抽取正确但极性分类错误，或者极性分类正确但目标检测错误。

当前，这一模型已被应用于网易云音乐的多多和西西。两个角色的共性在于，他们都是爱听音乐的小鹿，都是男性，且年龄相仿。如何让他们在对话中给用户留下不同的印象呢？我们借鉴了卡通及游戏制作中人物设定的方式，赋予他们不同的性格和喜好。利用态度分析的技术，多多和西西会对用户提出的一组问题和回复进行分析，判断出用户对何种目标具有怎样的情感信息，进而根据人物设定的不同特点来影响对话，生成有区别有个性的回复。三观模

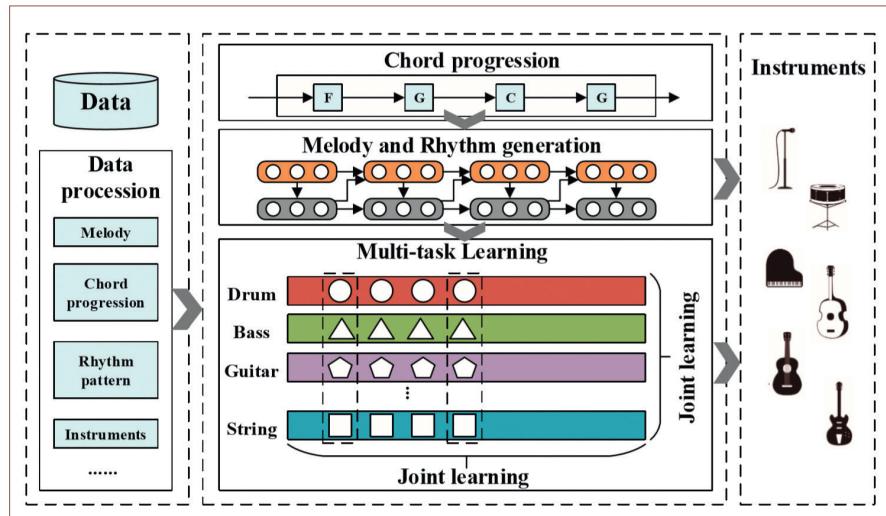


图3 小冰乐队自动作曲编曲的框架图

型将“体温”赋予了包括小冰在内的人工智能角色，并将通过态度的一贯性、延续性来逐步凸显角色的性格。

微软小冰也率先整合自然语言处理、语音和计算机视觉三大学科的研究成果，让小冰基于多感官来完成更自然的交互。此前，上述学科都是在各自的轨道上独立发展。近年来，深度学习

技术与算法的改进先后使语音识别和图像识别实现了显著的突破，人们期待自然语言处理技术也能实现类似的突破。多模态可能是突破自然语言理解瓶颈的一个方向。因为人类从出生之日起，每天都在接收听觉、视觉、嗅觉、触觉等多感官的输入，与文字联合构建了人的理解能力。我们也尝试结合跨模态信息作诗，同时也在听觉方面进行了尝试，结果令人兴奋。

借助图像识别生成诗歌文本涉及到多项挑战，包括发现图像中潜藏的诗意线索（例如绿色象征生机、阳光代表希望），生成的诗歌既要与图像相关，又能满足语言层面的诗意要求。对于这些挑战，傅建龙博士的团队通过策略梯度，将诗歌生成工作划分成两个相关的多对抗训练子任务，并提出了学习深度耦合的视觉诗意嵌入，机器在训练过程中可以连带学习图像中物品、情感和场景的诗意呈现^[6]。测试结果证明，这种作诗方法比其他基准方法更高效也更具艺术性。与小冰写诗之前的版本相比，这项工作突破了从图到关键字，再从关键字到诗的框架，能够更多地使用图像里的信息。他们的论文被 ACM MM 2018 (ACM 国际多媒体会议, ACM International Conference on Multimedia) 大会接收并获得最佳论文奖。

我们还极大地扩展了小冰的音乐能力。现实中，要想演绎出一首动人心弦的歌曲，往往需要一组音乐人通力合作，流程繁复又漫长。微软（亚洲）互联网工程院在苏州的一支团队提出了一端到端的旋律及编曲生成框架——小冰乐队^[7]。

如图 3 所示，该框架首先通过一个基于和弦的节奏及旋律交叉生成模型来生成一段主旋律，再借助多乐器协同编曲模型，根据多模态学习来生成不同乐器的多轨伴奏音乐。对现实世界的数据集进行了大量实验，证明小冰乐队的有效性。与以往工作生成的音乐相比，小冰乐队生成的音乐更具有可唱性，乐句长短的分布与人类作曲家的作品基本吻合，不会因为休止符过多或随意性造成音乐的碎片感。小冰乐队是第一个能自动编曲的系统。这项研究成果已经发表在 KDD 2018 (国际数据挖掘与知识

发现大会, Conference on Knowledge Discovery and Data Mining) 上，并获得了 Research Track 的最佳学生论文奖。

用代码可以构建人工心智吗？正如微软小冰的负责人李笛所说：“过去我们所进行的全部工作，只不过是让一个曾经的‘不能’变成了一种不确定性。也就是说，它不再是一种确定的‘不能’，但它也还远远没有达到一个‘能’。从我们的角度来讲，其实最喜欢的就是这种不确定性，因为不确定性会带来创新，不确定性的时间越长，那么创新也就能越充分。” ■



宋睿华

CCF 专业会员。微软（亚洲）互联网工程院小冰首席科学家。主要研究方向为信息检索、信息抽取、社交计算和文本生成。曾任或现任 SIGIR、CIKM 高级程序委员会委员。rsong@microsoft.com

参考文献

- [1] Cheng W, Wu C, Song R, et al. Image Inspired Poetry Generation in XiaoIce[OL].arXiv preprint arXiv:1808.03090, 2018.
- [2] Xing C, Wu W, Wu Y, et al. Topic aware neural response generation[C]//Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence, 2017: 3351-3357.
- [3] Wu Y, Wu W, Xing C, et al. Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-Based Chatbots[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2017: 496-505.
- [4] Wu Y, Wu W, Yang D, et al. Neural Response Generation with Dynamic Vocabularies[C]//proceedings of AAAI-18, 2018.
- [5] Zeng Z, Lin P, Song R, et al. Attitude Detection for One-Round Conversation: Jointly Extracting Target-Polarity Pairs[C]//Proceedings of WSDM 2019, 2019.

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

让机器学会触景生情吟诗作赋

关键词：机器学习 人工智能

刘 蓓 傅建龙
微软亚洲研究院

随着人工智能(AI)技术的不断发展及其概念的深入人心，人们越来越关注人工智能会对我们的生活、娱乐和艺术创作产生什么样的影响。当机器已经能够非常好地从图片中识别出物体、场景等关键词，并且可以生成流畅的易于阅读的描述性语句时，我们开始探索机器学习能否更进一步，比如看图写诗。诗歌在文学领域属于一种比较抽象高级的体裁。我们从小就开始学习写作，比如说明文、议论文、散文，却只有很少一部分人尝试诗歌的创作，这让诗歌对大部分人来说显得神秘又遥不可及。但实际上诗歌离我们并不遥远，很多诗歌是基于我们日常生活的感悟，只是使用自己的天赋创作诗歌的人比较少。好的诗歌不仅对于诗人来说是佳作，也可以感动并触及到大多数人的灵魂深处，比如徐志摩的《再别康桥》、海子的《面朝大海，春暖花开》等。

在确定了看图写诗这个任务之后，我们开始认真思考如何让机器根据一幅图像写出合理且符

合图片意境的诗歌。对于人类诗人而言，作诗的步骤是首先观察图像，根据图像内容确定有哪些素材可以激发诗歌的灵感，然后根据这些灵感和以往写作的经验进行诗歌的创作，最后再对自己的作品进行润色打磨。我们的方法与这个思路基本一致。但也面临着随之而来的挑战，包括如何从图像中发现诗歌线索（例如，绿色蕴含的希望），如何使生成的诗歌既满足与图像的相关性，又满足语言层面上的诗意。

机器学习图片写诗的挑战

近年来，诗歌生成在行业内引起了广泛关注。Facebook提出了使用神经网络来生成英文韵律诗^[1]；微软发布了一款叫“小冰”的人工智能产品，其最重要的功能之一就是会根据用户上传的图像激发灵感生成诗歌^[2]。小冰的诗歌还被出版为一本名为《阳光失了玻璃窗》的书，引起媒体和公众的反响。不过，



描述：A **falcon** is **eating** during sunset. The falcon is **standing** on earth.

一只**猎鹰**在日落时**进食**。**猎鹰****站在地上**。

诗：	Like a falcon by the night	动如黑夜 猎鹰
	Hunting as the black knight	猎似黑暗 骑士
	Waiting to take over the fight	蓄势待发
	With all of its mind and might	全力以赴

图1 图片描述和诗歌的对比

以端对端的方式从图像生成诗歌仍然是一个新的主题，面临着巨大挑战。图像标题技术和图像生成短文的重点在于生成关于图像的描述性语句，而诗歌语言的生成则是更具挑战性的难题。图像的视觉呈现与该图像中可激发或有助于更好地生成诗歌的象征之间，距离更远。例如，图像描述中的“人”在诗歌创作中可以进一步通过其他意象如“明亮的阳光”和“张开的手臂”等来象征“希望”，通过“空椅子”和“黑暗”等可以感受到诗意化的“孤独”。如图1所示，针对同一幅图像，描述和诗歌之间存在差异，这两种形式对相同颜色的用词也明显不同。相对于描述图像中的事实，诗歌更倾向于捕捉图像中物体、场景和感情的深层含义以及诗歌象征（例如，骑士和猎鹰，猎、发和进食，以及待和站）。

从一幅图像生成诗歌，我们主要面临三个挑战：(1) 与根据主题生成诗歌相比，从图像生成诗歌是一个跨模态的问题，一种直观方法是先从图像中提炼关键词或说明文字，然后以这些关键词或说明文字为种子，生成诗歌，正如从主题生成诗歌那样^[2-4]。但是，关键词或说明文字会丢失许多图像信息，尤其是对诗歌生成十分重要的线索。(2) 与图像标题技术和图像生成短文^[5, 6]相比，从

图像生成诗歌是一项更主观的工作，这意味着同一幅图像可以对应不同方面的多首诗歌，而图像标题技术 / 图像生成短文更多地是描述图像中的事实，并生成相似的语句。(3) 诗句的形式和风格与叙述语句不同。本研究中，我们主要关注的是一种开放形式的诗歌——英文自由诗。尽管我们不要求格律、韵律或其他传统的诗歌技术，但仍要有诗歌结构和诗歌语言。我们将这一特性定义为诗意，例如，诗歌的长度一般有限；与图像描述相比，诗歌一般偏好特定的词语；诗歌中的语句应与同一主题相关，保持一致。

机器写诗的具体过程

在本研究中，我们收集了两个人类注解的诗歌数据集，在一个系统中通过集成检索和生成技术来进行诗歌的创作。为了更好地研究诗歌生成中图像的诗歌线索，我们首先在包含数千对图像 - 诗歌的多模态诗歌数据集（即“多模态诗集”）上，训练了使用图像卷积神经网络(CNN)特征和诗歌句子的 Skip-Thought 向量的深度耦合视觉诗意图模型，使得同一对的图片和诗歌在嵌入空间中比较接近，不同对的距离较远。然后我们使用这一嵌入模

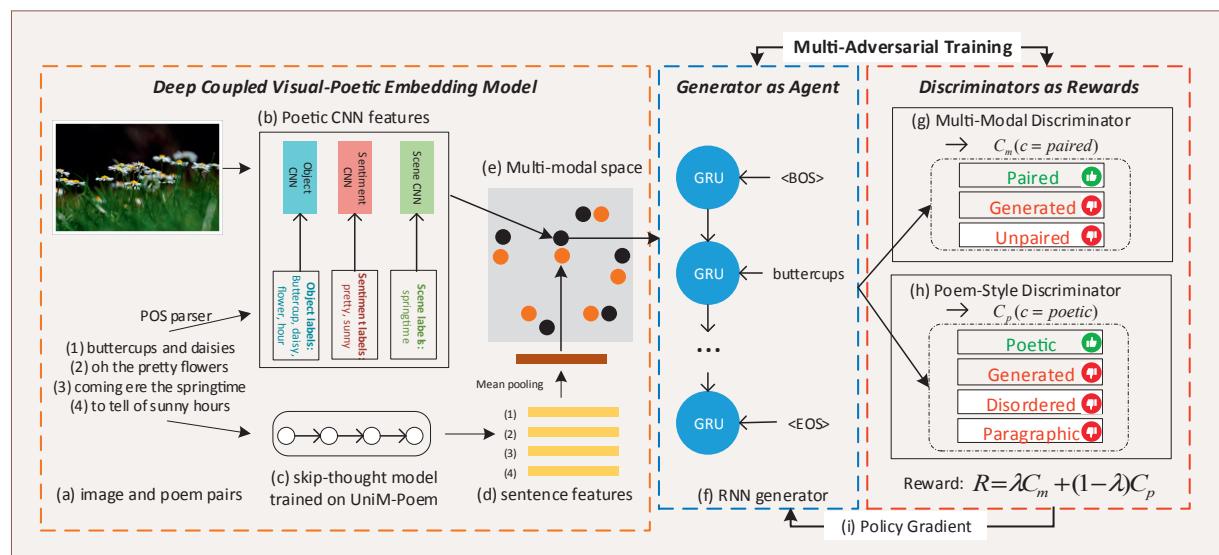


图 2 图片写诗的流程

型，从一个更大的图像单模态诗歌语料库（即“单模态诗集”）中检索更多相关的诗歌。这些被检索的诗歌的图片，与多模态诗集一同构成一个扩大的图像 - 诗歌对数据集（即“多模态诗集(EX)”）。我们还提出使用最新的序列学习技术，训练基于多模态诗集(EX)数据集的端对端诗歌生成模型。该架构保证我们能够从扩展的图像 - 诗歌对中发现并塑造大量的诗歌线索，这对诗歌生成至关重要。为避免长序列（所有诗行一起）导致的曝光偏差问题以及无可用的特定损失函数来定量评测生成诗歌的问题，我们提出使用多对抗训练的诗歌生成递归神经网络(RNN)，并通过策略梯度对其进行进一步优化。

一首好的从图像激发的诗歌必须至少满足两个标准：(1) 与图像相关；(2) 在合适长度、诗歌语言风格和诗句一致性方面具有诗意图。根据这两个要求，我们设计了两个判别网络来指导诗歌的生成：多模态判别器和诗歌风格判别器。深层判别网络在文本分类任务中已经被证明具备很高的有效性，尤其是对不能建立良好损失函数的任务。通过对抗学习的方式，多模态判别网络指导诗歌生成器生成与图片正相关的诗歌，而不是不相关或者一看就是生成的诗歌。同理，诗歌风格判别器可以指引诗歌生成器生成更符合人类写的诗歌风格，而不是其他文体或者逻辑混乱的诗歌。通过对正样本和负样本诗歌数据集的学习，符合要求的诗歌会被给予更高的奖励，而不符合的会受到惩罚。

图3是根据上述方法所生成的一些诗歌的展示。

如何评价机器生成的诗歌

对于诗歌的鉴赏可谓见仁见智，诗歌的评估通常是一项困难的任务，在现有的研究中没有既定的评价标准，对于根据图像生成诗歌这一新任务而言就更是如此了。为了更好地评价诗歌的质量，我们提出同时使用自动和人工的方式来进行评价。对于自动评价，我们建议在标准化后采用双语互译质量辅助工具 (bilingual evaluation under-



(a)

*In the morning light
Is warm and dark
It is
Beautiful to be
A dream*



(b)

*The sun is a great man of corn
Caught the waves in the air
Unknowingly
Spread torn to us
Destroyed*



(c)

*And now I am standing alone
On the road where you are born
I know what you want to be
You are always trying to fly
Your heart is like a butterfly
You can see it in love
Let it be your heart
Let us be free
Let it be*

图3 机器生成诗歌示例

study, BLEU)、新颖性和相关性三种评价标准计算总分。BLEU 通过将生成的诗歌和原有数据集中与图片匹配的诗歌进行比较，计算出一个相似度；新颖性主要是考量生成诗歌中的一些单词和短语在训练数据集中出现的频率；相关性主要是针对图像写诗这一特殊任务，通过嵌入式网络比较生成的诗歌与给定图像的相似度。



Poem1
The sun is shining
The wind moves
Naked trees
You dance

Poem2
Lime and limpid green
A second scene
A fight between the blue
That you once knew
Floating down the sound resounds

图 4 图灵测试的示例

我们将自己的方法和图像生成文字模型生成的诗歌进行比较（所有模型都在统一数据集上进行训练），结果显示我们的方法在不同指标上都超越了其他模型。我们也比较了不同判别器的作用，通过分别单独加入判别器的方式，我们发现，仅有模态判别器的模型会引导生成跟原有匹配诗歌相似的诗歌，因此它在 BLEU 上得分最高，而诗歌风格判别器有助于为生成的诗歌提供更新颖、更富想象力的措辞。总体上，最终模型结合了两种判别器的优点，

在 BLEU 和新颖性上取得了合理的中间分数，且与其他生成模型相比，表现得更为出色。

但是，在这样一项开放性的任务中，没有特别适合的标准能够完美地评价生成的诗歌质量。我们使用的自动标准在某种程度上可被视为指导。为更好地从人类感知角度来说明诗歌的质量，我们通过众包的方式将诗歌展示给人类用户，通过对多个指标进行打分来评价诗歌的好坏。评分指标主要是四方面：相关性（与图像是否相关）、连贯性（诗歌各行之间是否连贯）、想象力（诗歌对于给定的图像显示了多少想象力和创意）以及整体印象。结果显示，我们的模型在所有标准中的表现都优于基线方法。

除了通过打分进行评价，我们也进行了图灵测试，将一张图片和两首诗歌展示给用户，其中一首是机器生成（通过我们的方法），另一首是数据集中原有配对的人类写的诗歌（如图 4）。通过比较两首诗歌，用户判断哪首更像是人类写的。为了比较不同用户的评价，我们将这个任务交给了一般用户（548 人）和专家用户（30 人）两批人，其中专家用户是与英语文学相关的专业人员。结果呈现出一般用户只有 51% 能判断对，而专家用户虽然正确率高一点，但也有 40% 无法准确判断。

展望

人工智能的发展给机器的创作带来了更多的发展空间，同时也激发出我们对于曾经比较少涉猎的领域更多的兴趣。从图像出发进行诗歌的创作，是计算机视觉和自然语言处理的一个文艺的结合和创新，我们的这一小步，结合了对于人类写诗的分析和模拟。未来的路还很长，图像写诗的工作有很多值得继续挖掘和探讨的方向。目前从图像中学习诗意化的线索受限于配对的数据集，而实际上我们在看到一个图像时思考的东西更多，如何将已有的一些知识更好地结合到诗意线索的挖掘中是个很有意思的课题。

（下转 73 页）

大规模视觉智能探索与实践

关键词：大规模视觉智能 人工智能

华先胜
阿里巴巴达摩院

“繁华”背后的挑战

随着深度学习等技术的进步，人工智能(AI)行业迅猛发展并赋能各行各业。从2000年到2017年，出现了8000多家AI公司，17年间AI创业公司增加了14倍。仅2017年，中国AI公司获得的投资就超过50亿美元。各大公司纷纷成立AI研究院或者AI部门。AI行业欣欣向荣，势不可挡，但在营收、技术、用户需求等方面，却面临着很多不得不重视的挑战。

首先是关于AI公司的营收。90%的AI公司都处于亏损状态。很多AI公司做项目，但是很多项目需要定制研发，需要大量的人力，导致研发成本很高。不做项目而做产品的公司也有困难。目前有很多AI产品同质化比较严重，使得很多技术卖不出高价钱。

其次是技术落地的差距。AI产品背后需要强大的技术作为支撑，而实验室场景下的技术与真实场景下的产品要求之间存在很大的差距。模型越复杂，两者之间的差距可能就越大。例如，人脸算法在LFW¹上可以做到99.7%的准确率，但在真实场景中，如果要求虚警率(false positive rate)大幅度降低，那么召回率(true positive rate)就可能很低。虽然这个数据稍微有点过时，但也足够说明问题。除了技术的差距外，还有数据上的差距。例如，目前热门的研

究课题——行人重识别(Person Re-id)，公开数据集中的数据和实际场景其实存在很大差距，导致技术性能无法满足落地场景的要求，这些都是很大的挑战。

最后是用户需求的差距。在真实场景下，AI公司能做的事情和用户真正想实现的需求之间，依然存在很大的差距。举一个真实的例子，一个客户给我们提了260项需求，这些需求都实现了才能解决客户的所有痛点问题，但我们目前能实现的只有20多项。

什么是成功的AI产品

在AI的浪潮中，人人都梦想能打造出成功的AI产品。到底什么样的产品或应用才算获得成功？在比赛中获得第一名，新闻中频繁曝光，获得高额估值，融资到很多研发经费，这些都是有价值方面。然而，最为重要的还是产品或应用能给客户带来重大的价值。这个价值有不同的层次，可以分为锦上添花、雪中送炭以及无中生有。雪中送炭和无中生有，比起锦上添花更有价值，因为这样的价值是不可替代的，甚至能为客户创造一个需求，进而带来新的价值。例如，智能手机的大屏就是创造出来的需求，餐馆也是一样。如果AI公司的产品或应用能够给客户创造核心的不可替代的价值，那么收入就不仅仅来自风险投资(VC)和大公司，而且

¹ LFW是Labeled Face in the Wild(户外标记人脸数据集)的简称，是人脸识别研究领域最重要的人脸图像评测集合之一，由美国马萨诸塞大学计算机视觉实验室于2007年发布。

来自客户，这时候才算是成功的。

阿里视觉 AI 应用探索与实践

阿里巴巴在视觉 AI 领域进行了很多探索与实践。本文将通过四个案例来讲述如何研发出合适的技术并且转化为生产力，以及如何利用技术为客户创造最大的价值。

拍立淘视觉搜索

平时我们在电商购物的时候大多是用文字搜索，或者去浏览商品目录以及推荐列表。但很多时候文字无法表达用户想要什么。拍立淘是淘宝基于图像识别技术沉淀的视觉搜索 AI 应用，解决用户“通过照片搜索商品”的刚需。

例如，有一次我在茶馆喝茶的时候用到了一个杯子，茶叶放在一个容器里面，杯盖上有一个红色的按钮。按下红色按钮，茶水就漏到茶壶里，可以很方便地控制茶的浓淡。如果我要买这个杯子，用文字搜索的方式，很难描述这款杯子的特征，可能只能去搜“带红色按钮的茶壶”，但这样不会搜出好的结果。而拍下杯子的照片，使用“拍立淘”就很容易搜索到，而且我知道了它的名字叫“飘逸杯”（图 1）。

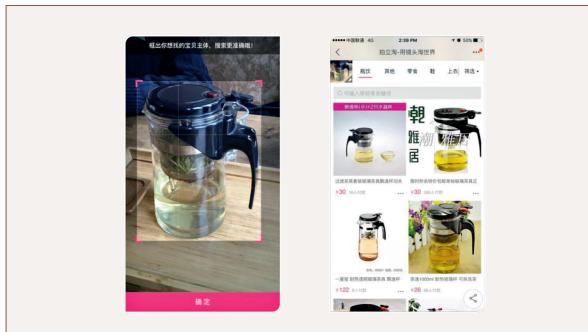


图 1 “拍立淘”视觉搜索找到想买的这款飘逸杯

用户将拍好的图片上传到搜索应用，拍立淘系统首先判断这个照片是关于什么类型的商品，然后检测商品在图片上的位置，最后提取图片的特征。把图像的特征描述做到可用，是近几年一个很大的

发展趋势。使用这些技术去搜索，就变成了大规模向量搜索的问题。此处的搜索和普通文字搜索有两点不同，第一点是特征学习，通过深度学习技术可以灵活地设计神经网络，在大数据上逼迫神经网络收敛到需要的地方。第二点是搜索系统，针对图像的索引系统，是高维空间的向量索引。还有一部分工作是偏向工程方面的，例如如何及时地响应大量用户的搜索请求。图像搜索已经做了几十年，但是在拍立淘之前却没有一个被大家广泛使用的基于图像检索技术的应用系统，所以说电商的图像搜索是刚需，而通用的图像搜索目前还不是刚需。现在拍立淘每天有数千万人在使用，日成交量也很大，成功的关键是我们将技术聚焦在解决一个具体的需求上。一个应用如果能满足刚需，就一定有人使用，哪怕系统一开始并不完善，有人使用就会有数据，同时也会有更多的“坏案例”(bad cases)，促使应用不断迭代，并能够清晰迭代的方向。

鹿班视觉 banner 生成

每次大型促销期间，各个商家都要生成定制的 banner 广告图，即互联网上做商品宣传的横幅广告图。banner 图一般包括商品图片、商品名称和商品促销相关的文字。banner 图有上亿张，如果都使用人工生成，就需要大量的人力成本。鹿班是一款视觉生成 AI 产品，帮助卖家自动完成二维平面广告的设计。banner 的生成是多样类型的，每次输入都可以得到不同的结果。这个系统可以自动选择背景，自动为背景改变颜色，自动完成整体设计。2017 年“双十一”的 7 亿多张 banner 中，有 4.1 亿由鹿班生成。

虽然视觉生本身也是基于识别、理解和搜索等一些视觉技术，但是与一般的视觉技术应用不同。视觉生成的应用不是去搜索和理解，而是进行创造。这类技术可以根据需求生成对应的图像、视频或者图形，例如可以将一个页面自动转换为一个视频，可以替换已有视频中的平面或者物体。根据点击率和转化率进行客观的测试，经过验证，鹿班大概相当于本科毕业的设计师水平。而且和人相比，算法



图 2 banner 图自动生成

的优势是速度，可以在 1 秒内生成 8000 个 banner。

鹿班这个应用不是设计师、商家或者 AI 专家单独创造出来的，而是他们碰撞出来的新需求。如果没有鹿班，也可以人工生成 banner。而鹿班的出现给商家创造了用 AI 快速生成 banner 的新需求。每次大型促销的时候都需要在短时间内产生大量的 banner，而只有这个系统才能做到，如今已经成为了刚需。与拍立淘一样，一个成功的 AI 应用，都是以解决刚需为目标，AI 本身对商业模式和商业应用产生很大影响的同时，商业应用在使用当中也给应用回馈大量有价值的数据，帮助 AI 不断优化。

工业视觉诊断

视觉 AI 还可以应用到工业视觉诊断领域，可以诊断人，也可以诊断非人。对人的诊断主要是分析医疗影像，而对非人的诊断主要是针对产品和机器，一般称为工业视觉。

在诊断人的领域，针对肺结节的检测和肺部综合的分析，我们曾经获得过 LUNA 比赛²的第一名。视觉诊断可以提升医生读片的效率，体检是不错的应用场景。现在视觉诊断系统每天处理几万个肺部影像。还有肝脏细胞的检测、膝关节疾病的诊断等，也都应用了我们的诊断系统。曾经在一个医疗行业的峰会上，我们用视觉诊断系统与现场 450 多名骨科医生比赛，结果是视觉诊断系统比医生的诊断结果稍微好一点。

视觉诊断在工业诊断领域也有大量的应用机会，例如，电池片的质检。有些电池片在生产过程当中出现缺陷，如果是人工检测，判断一块电池片是否存在问题需要两秒钟。一般的生产厂商无法做全量的缺陷检测，只能对 10% 左右的样品抽样检查，检测出有问题的样本，从而估算全量的产品有多少存在缺陷。采用视觉诊断 AI 系统可以做全量的检测，达到与人相同甚至比人更好的水平。人工检测的准确率在 93%~96%，但是机器可以做到 98%；人只能区分 4 种缺陷，但是机器可以区分 20 种；人的检测时间需要 2 秒，机器只需要 200 毫秒。质检本身是很多制造商的瓶颈，有了视觉诊断 AI 以后，就可以打破这个瓶颈，提高效率，节省成本，提高产品质量。

AI 有大量的应用场景，通用的计算技术是存在的，但通用的 AI 技术是不存在的。真正要解决 AI 问题，一定要深入这个行业，了解这个行业的数据需求。例如，在工业诊断中，只有很少一部分产品存在问题，要想做非常精准的诊断很困难。通常情况下，会保证召回率，而牺牲准确率。举一个简单极端的例子，比如有 10000 个样本，里面有 10 个问题样本，由于技术存在局限性，无法精准地找到这 10 个问题样本，但却可以找到 100 个潜在样本，同时保证这 10 个问题样本存在于这 100 个潜在样本中。这时候的召回率达到 100%，而准确率只有 10%。从精度来看，10% 是不理想的。但是如果换一个角度，则是减少了 99% 的人力。从原来的 10000 个诊断任务减少到 100 个。即使准确率低到只有 1%，也节省了 90% 的人力。从这个角度思考，算法的设计就会不一样。

城市大脑智能监控

在当代城市中，有很多传感器在努力地工作，随时随地产生大量的数据，但却是一个盲人摸象的局面。城市中虽然有很多摄像头，但大部分是没有智能的。大量的数据被记录下来，保存若干天，然后就自动删除。由于人力有限，不可能查看全部摄

² Lung Nodule Analysis 的简称，是一项国际权威的肺结节检测大赛。

像内容。同时，数据没有打通，数据之间的关系没有被挖掘，城市管理问题中潜在的因果关系也无从得知。城市大脑的核心思想在于打通这些数据，用AI算法和算力去挖掘这些数据的价值。使用技术与智能驱动城市管理和服务，带来全面全量全局的优化，也带来高效便捷与省时省力的生活，带来城市管理模式、服务模式和产业模式的变革和突破。

城市大脑这个项目的开始和推进遇到了非常大的挑战甚至遭受质疑：海量数据能不能及时处理，代价有多高？用如此大的代价处理这些数据，能带来什么价值？城市大脑的数据处理与过去的视频监控到底存在哪些差别？

城市大脑的核心就是分析处理城市所产生的海量数据。首先要对视频数据进行认知，对车、人、事、物进行全面精准的识别，然后进行决策和优化，比如宏观的调控，实时的报警等。接下来可以把所有的视觉要素放在搜索引擎里，对人、车进行搜索，比如逃逸的肇事车辆、走失人口等。再下一步就可以进行预测，城市大脑可以精准预测一个小时以内任意时刻的车流和人流。这样，路径规划和交通流控制就更加智能。

城市大脑开始于2016年，2017年成为国家首批四个人工智能开放平台之一。到现在，城市大脑已经有了一些具体实际的应用。

在城市感知方面，例如车人的检测和识别，不论视频质量和天气状况如何，对人、车都能进行精准快速的检测和识别。在这个工作中，我们引进了一个预览层(preview layer)。这个层包含的上下文信息更多，可以判断对应区域里面是否有物体，进而可以抑制很多错误的响应。在真实场景下进行检测、识别还存在很多问题，例如对低质量和少见的车牌进行识别还较为困难。因此，我们自主研发了风格化自编码器去制造一些样本，从而使准确率上升了18%。这个技术还可以用在图像质量增强上，比如超分辨率和降噪之类的问题。

在异常事件检测上，我们采用了自编码器的方法，设计了一个预测分支，可以迫使网络学习时序信息。例如，高速公路上临时停车是非常危险的，当一辆车停在高速公路上时，我们的系统就会检测到它，并且发出警报。杭州有很多骑行交警队，在事故当事人还没有打电话报警时，交警就已经到达事故现场。根据系统分析，如果发现有些地方经常产生规律性的报警，说明这个地方需要进行治理；治理好了，报警数量就会降低。在红绿灯优化方面，在试点区提升了15%~20%的通行效率，能够节省50%的救护车救援时间。

在关于搜索的内容方面，行人重识别是一个热门的研究方向。我们最新的一项工作表明，在网络学习的过程中，视觉对象的位置和特征越来越精准，会聚焦在真正有价值的区域上。例如，图3中检测的基准线(baseline)会预测错误，原因是不能区分黑色的箱子和黑色的短裤，而我们的方法可以把人体和非人体非常明确地区分开来。基于这些技术，我们在Market-1501数据集³上超过了97%的准确度，位于世界第一。这项技术在实际应用中，叫做渐进式视频搜索。

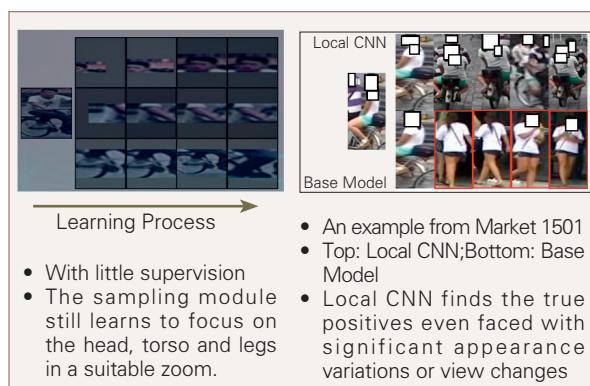


图3 行人重识别

如何构建一个大规模的AI开放平台，尤其是视频处理平台也是我们面临的一个核心问题。基于云计算平台很重要的一个特性就是开放。基于开

³ 该数据集在清华大学校园中采集，图像来自6个不同的摄像头，其中有一个摄像头为低像素。该数据集提供包含12936张图像的训练集和包含19732张图像的测试集。

放平台，第三方开发者可以聚焦在算法研发上，算法可以很容易地部署到这个平台上，其他很多优化和流程的问题都由平台来解决。城市大脑是一个开放的平台，目前从杭州的萧山区到主城区，再到余杭、苏州、衢州、乌镇、北京、澳门，城市大脑逐步在多个城市或地区落地。

从城市大脑这个案例可以发现，打造 AI 产品，产生不可替代的价值最为重要。城市大脑赋予摄像头智能分析能力，其相当于几千个警力，可以对红绿灯的调时、配时进行优化，使出行时间降低 15%，到那时，这个系统就是不可替代的。当然这不仅仅依赖技术，还依赖于对这个行业以及对数据的深入理解，真实地解决了客户的需求。AI 可以起源于项目，但是不能只停留于项目，需要把它变成产品，进而慢慢打造成平台和生态，才能发挥最大的价值。

总结与展望

本文分享的所有实践案例都遵守一个原则，那

(上接 68 页)



刘 蓓

微软亚洲研究院副研究员。主要研究方向为多媒体信息检索、AI 创作、多模态学习。liubei.cs@gmail.com



傅建龙

微软亚洲研究院主管研究员。主要研究方向为多媒体分析和计算机视觉。jianf@microsoft.com

参考文献

- [1] Hopkins J, Kiela D. Automatically generating rhythmic verse with neural networks[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(ACL). 2017, 1:168-178.
- [2] Cheng W, Wu C, Song R, et al. Image inspired poetry

就是“**AI+ 算力 + 数据 = 价值**”。其中，最关键的是价值。只有真正解决了用户的痛点，才能驱动 AI 不断迭代发展，不断优化。价值也并不是简单地把 AI、算力、数据放在一起就会产生，其中有很多规律，需要对行业、数据和算法有深入的理解。

AI 无处不在、势不可挡。虽然已经取得了长足的进步，但仍然有很多局限，面临很多挑战。如果没有深入行业应用，就无法做出成功的 AI 产品及应用。未来，人和 AI 都有很多机会。我们应该优势互补，让 AI 做更多人力所不能做的事情，让人去做算力和 AI 所不能及的事情。■



华先胜

CCF 专业会员。阿里巴巴达摩院机器智能实验室副主任，城市大脑人工智能技术负责人。IEEE Fellow, ACM 杰出科学家。主要研究方向为大规模视觉人工智能领域，包括视觉分析、识别、搜索和挖掘等。
xiansheng.hxs@alibaba-inc.com

generation in XiaoIce. arXiv: 1808.03090v1[cs.AI], 2018.

- [3] Yan R, Jiang H, Lapata M, et al. i, Poet: Automatic Chinese poetry composition through a generative summarization framework under constrained optimization[C]//International Joint Conference on Artificial Intelligence(IJCAI). 2013:2197-2203.
- [4] Ghazvininejad M, Shi X, Priyadarshi J, et al. Hafez: an interactive poetry generation system[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(ACL). 2017:43-48.
- [5] Yu L, Zhang W, Wang J, et al. SeqGAN: Sequence generative adversarial nets with policy gradient[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. AAAI, 2016:2852-2858.
- [6] Krause J, Johnson J, Krishna R, et al. A hierarchical approach for generating descriptive image paragraphs[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2017:3337-3345.



The CS David 专栏

CCCF 2019 年第 3 期

下一代网络将是什么样子？

当最初看到戴维·克拉克 (David Clark) 的著作《设计未来网络》(*Designing an internet* , MIT 出版社, 2018) 时, 我对它的名字产生了误解。因为克拉克是互联网的先驱之一, 我想当然地认为这本书是关于他们的工作以及他们是如何设计互联网协议的。事实上, 作者在这本书中谈论了一些完全不同的事情。它更多地在展望未来而非回顾过往, 它最终的关注点是下一代网络。

然而从某些方面来说, 我对这本书的误解反映了互联网本身, 甚至在某种程度上反映了所有计算技术的问题。每一个接触过互联网的人对互联网技术可以做什么都仅有有限的理解。他们只看到了这些技术在他们自己的问题、领域和需求上的使用。但互联网已经在一个又一个不同领域被广泛应用, 它不仅将个体与个体联系起来, 还几乎为所有个人和机构揭示了一种新的工作方式。

至少有三个重要的原因让我们去审视互联网的现状。首先, 正在成长的一代是互联网的原住民, 他们从未经历过没有互联网的时代, 在所有国家都是如此。他们常常在不经意间让我们尴尬地谈起在我们拥有全球化的数字通信网络之前这个世界是什么样子。我告诉他们, 从前电话是固定在墙上的; 从前邮件是写在纸上的; 从前洗照片需要使用化学药品; 甚至在不太久之前, 当我们的配偶去到世界的另一边出差时, 我们不得不和对方说再见, 而直到他们回来之前双方都不会再有交流的机会。

需要重新审视互联网现状的一个不那么怀旧的理

由是, 在过去五六年里, 互联网已经显示出它设计的落后。我们一直在要求这种 20 世纪 80 年代设计的技术去做一些在它设计之初从未认真考虑过的事情。我们当前面临诸如网络地址紧缺、安全性以及聚合大量数据之类的问题。单就这些问题而言, 每一个都可能成为迫使我们对现有互联网做出本质性改变的动力。

最后, 对于互联网我们已经有近三十年的经验, 目睹了它是如何改变我们的制度, 摧毁旧的生产、教育和管理手段。同时我们也认识到, 20 世纪 90 年代初在互联网兴起的最初岁月里那种整合的力量已经不复存在。我最开始使用网络时, 几乎所有的用户都是研究人员、工程师、科学家或是学者。他们都接受过常规互联网的培训, 这样的培训帮助互联网联结在一起。而现在只有一小部分互联网用户属于该类别。说互联网是由科研与教育所支撑, 就像在说计算是以数值分析为中心一样, 因为大多数最开始的使用者都将计算机用于数值计算。从某个特定的方向开始并不意味着一定就会一直沿着这个方向走下去。

克拉克博士的新书内容丰富, 包含了大量关于网络现状的重要信息, 以及网络工程师和计算机科学家推进网络技术发展的各种想法。然而它对自己的描述并不准确。编辑声称该书 “精美且易于阅读”, 并且 “对法律决策者、政策倡导者和实施者、企业家以及任何与互联网相关的人士都有价值”。实际上这本书是作者关于互联网技术和政策思考的一个总结。它的强大之处在于克拉克博士对互联网的相关文献了如指掌, 并且亲自贡献了其中一些有趣的部分。如果你需

要索引互联网方面的文献，那么这本书将是一个很好的起点；但是如果你试图了解网络的新方法，可能会感到沮丧。即使作为一个熟悉互联网相关文献的研究者，我依然需要不断地从 IEEE 和 ACM 电子资料库中查阅原创文献，这些文献帮助我更清楚地了解研究人员计划如何设计下一代网络通信。而在《设计未来网络》中这方面的陈述不甚详尽。

当然，这反映了关于互联网的基本问题。我们声称通过互联网可以访问世界上的所有信息，但人们实际上并不需要所有的信息，他们仅仅希望获得那些有用的信息。一个“盲目”的网络——在传递信息时不知道也不关心究竟在传递什么，并不是帮助我们获取所需信息的最佳工具。为了发挥互联网的作用，我们需要应用程序、资料库、搜索引擎以及其他技术工具。根据过往的经验，我们知道这些工具所取得的仅仅是部分的成功。它们倾向于向我们提供已知的信息，往往需要稍许推动和刺激才能向我们提供所需要的东西。

一些利用人工智能和机器学习的新工具可以帮助我们对互联网上的信息进行分类，这些技术为互联网的未来勾画出了美好的前景。然而我们知道它们其实并不完美——这些技术会增加网络流量。另一个显而易见的问题是它们无法帮助我们处理物联网生成的数据。而很显然在未来五到八年内，传感器和数据聚合器将能够在物理、经济、社会和政治等方面为我们观察世界提供内涵丰富的视角。

如果按照克拉克博士的观点，现有的互联网尚不足以容纳我们很快就可以获得的全部信息。我们既没有可以传输它们的网络，也没有网络能够兼容可以帮助我们理解这些信息的技术工具。克拉克博士提出了一个大胆的观点：我们现有的技术，包括 TCP/IP 协议和域名系统，都无法支持未来网络。（平心而论，作者也承认包括他自己在内，没有人能够

准确地预测网络的未来。他把自己的最后一章标记为“高度不确定的”，并且抱怨已经“厌倦了对未来一次又一次的预言”。）

基于他的论述，我们似乎不太可能渐变式地演进到新的网络架构。我们在现有的网络架构上投入了太多的资金和科研努力，因而希望能够对信息的基础设施作逐步的改良。而按照托马斯·库恩“范式转换”的观点¹，网络技术可能会出现突变。当网络无法提供新应用所需要的服务时，或者无法经济高效地提供服务时，“范式转换”就会发生。我们已经快忘记了在 20 世纪 90 年代所发生的“范式转换”：20 世纪 80 年代的网络所使用的大多不是分组交换技术，而是一种被称为“存储转发”的技术，例如 AOL、Prodigy、Tymnet、Bitnet 等公司和网络使用的都是这一技术。存储转发技术无法支持交互式会话、超链接或媒体流，而 TCP/IP 可以支持这些服务，因此它很快被 TCP/IP 网络取代。

这一变迁很大程度上已经被遗忘，因为很少有互联网先驱曾经使用过那些网络。他们所使用的是 Arpanet 和 CSNET 等预示着互联网真正诞生的网络。在《设计未来网络》中，作者认为下一代互联网可能以同样的方式出现。也许某些团队将开发出支持新应用的新通信技术，新技术能以更新、更简单的方式进行管理。很快我们将决定这项新技术将成为新一代的互联网，而旧的网络在我们有时间去哀悼失去它的损失前就会消失。关于这样的网络究竟是什么，《设计未来网络》至少给了我们一些启示。■

戴维·阿兰·格里尔 (David Alan Grier)

2018CCF 杰出贡献奖获得者。电气与电子工程师协会计算机学会 (IEEE-CS) 前任主席、IEEE Fellow (会士)、*Computer* 杂志主编。乔治·华盛顿大学名誉教授，华盛顿特区 Dgagh LLC 公司的技术总监。grier@gwu.edu

CCCF 特邀译者：

孙晓明 中国科学院计算技术研究所研究员

¹ 托马斯·库恩 (Thomas Samuel Kuhn, 1922—1996)，美国科学史家，科学哲学家。在库恩的科学哲学思想中，“范式” (paradigm) 是一个核心概念，是指从事某一科学的研究者群体所共同遵从的世界观和行为方式。所谓“范式转换” (paradigm shift)，是指一个领域里出现新的学术成果，打破了原有的假设或者法则，从而迫使人们对本学科的很多基本理论做出根本性的修正。——编者注

P4 与可编程数据平面： 回顾与展望

关键词：P4 可编程数据平面

毕军
清华大学

编者按：2018年9月初，CCCF编委邀请清华大学教授毕军写一篇可编程数据平面方面的文章。毕军是国内网络体系结构领域的知名学者，长期从事软件定义网络领域的研究，同时也是CCF杰出会员和杰出演讲者。当时约好10月交稿，2018年10月28日，毕军准时发来文章的定稿。文章以P4这一可编程数据平面的主要代表性技术为例，全面深入介绍了数据平面可编程性的研究现状，指出进一步研究面临的挑战，读来获益良多。就在编辑部对本文进行紧张编辑、排版的时候，2019年2月18日，传来了令人悲痛的消息，毕军教授因病永远离开了我们。中国网络领域失去了一位勤奋有为的学者，清华大学失去了一位好老师，CCF失去了一位杰出会员。回顾他的病程，2018年10月他已经疼痛症状明显，他是在病痛中完成了这篇文章。斯人已去，风范长存！就让我们用这篇文章纪念毕军教授。

P4 与可编程数据平面的兴起

网络设备的研发，普遍采用的是“自底向上”的设计模式，即交换机芯片决定处理报文的方式和可以支持的网络协议。但当有网络运营者提出对新协议和新特性的需求时，就要将需求告知网络设备供应商，网络设备供应商对用户的需求进行评估后确定有较大市场前景时，再组织研发。

“自底向上”研发的周期往往很长，成本也高。为了寻求“自顶向下”的设计模式，学术界与工业界一直在努力。在吸取了主动网络、ForCES、4D、RCP、Ethane等工作的经验与教训之后，**软件定义网络 (Software-Defined Networking, SDN)** 将数据平面与控制平面相分离，着眼于控制平面的可编程性，通过控制器以标准化的接口对网络设备进行配置和管理，来增强网络管控的灵活性。然而，由于控制器运行在终端服务器上，又

存在带宽低、转发延迟长、可扩展性差等缺点，因此，实现**可编程的数据平面**，使网络能够根据上层业务快速响应、实时决策，进行分布式的报文处理与高吞吐的报文转发，具有非常重要的意义。

P4^[1] 是可编程数据平面的主要代表技术。2013年，斯坦福大学教授尼克·麦基翁 (Nick McKeown) 等人首次提出了可重配置匹配动作表 (Reconfigurable Match Tables, RMT)^[2]，为可编程交换机设计了抽象转发模型。在RMT的基础上，麦基翁教授等人进一步提出P4，为可编程交换机提供了简单易用的编程语言，方便管理员描述可编程交换机的报文处理逻辑。P4使交换机报文处理行为不受协议类型局限，支持定制协议格式，使交换机可以快速支持新协议并精简冗余协议。P4对报文处理的描述功能独立于底层平台的实现细节，降低了编程难度，提升了代码跨平台迁移的能力。P4的出现推动了网络技术创新，符合“自顶向下”的设计模式，

受到广泛关注，相关论文在网络领域顶级会议上大量涌现，同时工业界也有相应的商用产品问世。

对 P4 与可编程数据平面的回顾与思考

P4 语言概述

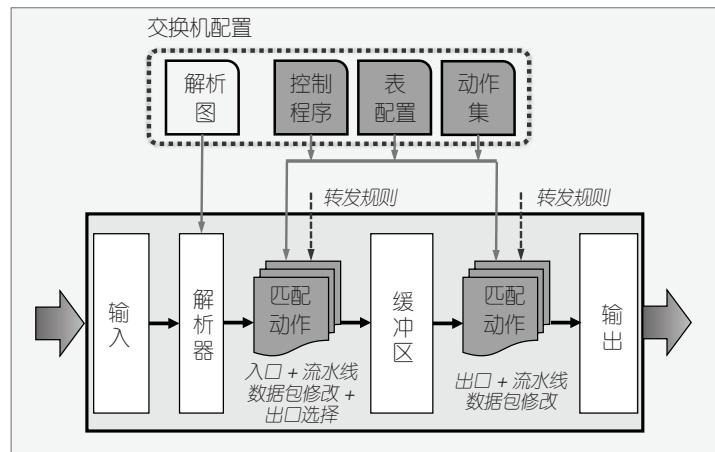


图 1 P4 抽象转发模型

P4 抽象转发模型如图 1 所示。P4 提供了多种语言要素来描述抽象转发模型的报文处理行为。头部结构、解析器和逆解析器使管理员能够自定义协议格式和报文解析顺序。P4 流水线中的基本单位是表，管理员可以自定义表中的匹配域和动作来描述流水线中的处理逻辑，具体的表项由控制平面在运行时插入表中。P4 流水线支持多个表，并且管理员可以通过控制流来定义报文在多表之间的执行逻辑。管理员可以通过 P4 提供的固有元数据控制交换机的特定行为，也可以自定义新的元数据，用于表间参数传递。P4 也支持外部对象以实现调用特定平台预置的库函数。P4 语言标准目前主要有两个版本，分别是 P4₁₄^[3] 和 P4₁₆^[4]。其中仅 P4₁₆ 支持外部对象。相比于 P4₁₄，P4₁₆ 更接近于高级编程语言，编程也相对简单。

目前，P4 语言存在以下局限性。语法上，P4 语言不支持循环、动态内存分配以及指针或引用；功能上，需借助外部机制实现多播与广播，不支持

描述队列、调度或者多路复用，不能操作报文载荷，不支持报文分段重组，不支持产生新的报文。

P4 相关研究

P4 程序开发周期主要包含两个阶段。在配置阶段，管理员对 P4 程序进行验证分析，保证 P4 程序的正确性，然后将 P4 程序编译为可执行代码。在运行阶段，管理员将 P4 程序部署到可编程交换机中，同时利用控制器配置数据平面中的匹配转发表。

P4 工具链主要包含两类工具。第一类是**编译工具**，主要负责将 P4 程序编译成不同平台的可执行代码，实现 P4 程序到可编程交换机物理资源的映射。编译工具需优化 P4 程序占用的资源，在交换机资源有限的前提下，尽可能支持更多的 P4 程序。由于交换机中存在多种物理资源，这些物理资源分级存放在交换机中，并且存在流表项分配限制、表空间资源限制、表间依赖限制等问题，这为 P4 编译

工具的设计带来了挑战^[5]。第二类是**验证、调试与仿真工具**，主要负责帮助管理员分析、调试和测试 P4 程序，保证 P4 程序的正确性，包含以下几种。(1) 验证工具：目前验证工具^[6-8]一般是通过将 P4 程序转译为成熟的模型或语言，然后利用符号执行等技术验证转换模型的属性，用于确保 P4 程序以及表项的逻辑正确性。验证工具面临的主要挑战在于难以保证 P4 程序和转换模型之间的一致性，P4 程序的状态空间爆炸以及验证的执行效率低等。(2) 调试工具：用于监控交换机运行时报文是否被正确处理，为管理员提供易用的接口调试可编程网络^[9]。(3) 模拟仿真工具：主要对可编程数据平面的底层设备架构、数据包处理逻辑与应用程序，进行模拟、验证和评估，例如 NS4^[10] 在 ns-3 基础上，实现了支持 P4 的网络仿真器，使管理员能够模拟包含多个可编程交换机的网络。

在数据平面，P4 提供统一的抽象转发模型（即 RMT），支持 RMT 的平台可以由 P4 编程。P4 数据

平面的相关研究主要关注在不同平台上实现 RMT。目前支持 P4 编程的平台包含以下三类。(1)ASIC : Barefoot 公司设计了首款 P4 可编程的高性能商用交换 ASIC——Tofino^[11]，其交换容量高达 6.5Tbps。(2)现场可编程门阵列(Field-Programmable Gate Array, FPGA):FPGA 比 CPU 性能高,比 ASIC 灵活性强,因此将 FPGA 用来做数据包处理平台可以兼顾性能与灵活性。使用 P4 描述数据包处理逻辑^[12],并利用编译器直接编译成 FPGA 可执行代码,可以降低 FPGA 的编程和调试难度,缩短开发周期。(3)CPU : P4 也可以描述 Open vSwitch(一种多层次虚拟交换机)的报文转发行为^[13],降低开发难度。

P4 相关应用

P4 具有可重配置、协议无关、平台无关等特性,利用这些特性可以解决现有网络中的诸多问题,同时还能将一些原本由中间件和终端设备完成的工作卸载到可编程数据平面上,从而获得可观的性能收益。

负载均衡是网络中十分重要的功能,对于保证可扩展性以及提升系统整体性能具有重要的意义。HULA^[14] 利用 P4 技术,根据网络拥塞状况实现流量负载均衡。SilkRoad^[15] 使用可编程交换机实现带状态四层负载均衡,其挑战是保证海量连接的状态一致性。SilkRoad 通过五元组哈希来压缩连接状态,在交换机上维护上百万条并发连接的状态,并通过布隆过滤器来确保目的 IP 地址池更新时的连接一致性。SilkRoad 将带状态负载均衡的成本降低了两个数量级。NetCache^[16] 提出了一种新的键值对存储架构,利用可编程交换机为用户提供热键值对缓存,均衡存储节点的负载,其核心思想是数据平面设备使用表对键进行分类,使用寄存器来存储值,使用大流检测器来辨认被频繁访问的键值对,控制平面只负责键值对的插入和删除。

在网络测量方面,FlowRadar^[17] 在可编程交换机上利用可逆布隆过滤器查询表对每条流的计数器进行编码,使用远程采集器的计算能力对全网流计数器进行解码和分析。UnivMon^[18] 基于 P4 实现通

用的流量监控框架,在数据平面使用 Sketch 进行测量,兼顾通用性和准确性。HashPipe^[19] 利用 P4 直接在数据平面实现大流检测,在可用资源有限的条件下实现较高的检测精度。如何针对不同的网络监控需求自动生成基于 P4 的测量程序成为目前重要的研究问题。Marple^[20] 是一套面向网络监控的查询语言,基于 Marple 编写网络性能查询请求可以通过编译器生成 P4 程序并在可编程交换机上初步地执行数据分析。Sonata^[21] 借助流式分析平台和可编程网络设备,利用交换芯片和 CPU 的混合架构,提供了包级别的网络遥测系统。

P4 和高性能可编程交换芯片的出现为网内计算卸载提供了新的机遇。网内计算卸载是指把一部分原本在服务器端完成的计算任务卸载到可编程数据平面上执行,报文转发和部分计算任务同时完成。计算卸载不仅可以减少网络流量、缓解拥塞,还可以缓解应用层负担、释放 CPU 资源^[22]。网内计算卸载的主要挑战在于可编程交换机的内存资源、动作集以及对报文的操作均有限,因此需要合理切分服务器端和数据平面所负责的计算任务。

P4 的优势

上述应用在数据平面中的实现,得益于 P4 相对于传统交换机的优势。

1. 协议无关。 可编程交换机能够支持任意数据包格式,带来了两方面的好处。一方面,交换机可以识别多层协议信息,基于多个维度的信息处理报文,例如在 NetCache 中,交换机可以从报文中提取并处理键值对存储的操作信息,并通过缓存键值对的方式对键值对存储加速;另一方面,可以利用新型数据包格式实现高效交换机间的信息交互,例如在 HULA 中,交换机利用数据包向邻居传递链路利用率信息。

2. 逻辑可编程。 可编程交换机的数据包处理逻辑是可编程的。一方面,管理员定制数据包处理逻辑,使交换机可以支持其他功能,如负载均衡、防火墙等;另一方面,管理员可以实现更复杂的数据包处理逻辑,实现更有价值的功能,例

如 HashPipe 中利用可编程交换机实现高性能的大流检测功能。

3. 存储可编程。可编程交换机提供了多种存储资源，其中一种是寄存器资源，每个报文可以对寄存器线速读取和修改，并且读取和修改的方式是可配置的。寄存器资源有多种用途，例如作为上层应用缓存存储键值对，存储连接状态，实现带状态处理，保证连接一致性。

总结与展望

P4 的出现给网络研究带来了诸多机遇，同时也面临许多挑战和问题。

一. 计算逻辑有限。P4 为数据平面提供编程语言，能够提高网络可编程能力，为报文处理带来极大的灵活性。但是，P4 语言本身并不是图灵完全¹的，因此无法支持过于复杂的报文处理逻辑，不能保证所有管理员需要的新功能都被 P4 语言表达。当出现 P4 语言无法表达的功能时，修改交换机芯片不可避免。目前，对于特定功能在可编程交换机上的可实现性仅能依靠经验来判断，这也为 P4 语言的使用者带来了困扰。并且，P4 仅能对交换机报文处理流水线进行编程，能够实现的计算逻辑有限，管理员无法使用 P4 重新定义交换机的其他功能模块（例如队列模块、端口缓存模块）。因此，如何扩展 P4 语言以支持定义交换机中的其他功能模块也是一个重要的研究方向。目前主要有两种研究思路：(1) 增加原语动作，通过函数调用的方式在 P4 程序中调用扩展原语动作，同时通过参数传递的方式扩展原语动作，适用于配置简单的功能扩展；(2) 扩展语言要素，当 P4 所支持的语言要素不能描述交换机中特定功能模块的处理逻辑，并且无法使用传参的方式进行配置时，需要定义新的语法要素、配置和调用方式，适用于配置相对复杂的功能扩展。

二. 数据平面资源有限。可编程交换机存储资源有限，无法像服务器一样以廉价、灵活的方式扩展存储资源，不可避免地限制了资源密集型应用在可编程交换机上的可实现性。一方面，可以考虑重构可编程交换机或重新设计交换芯片，提升交换机存储资源扩展的灵活性。另一方面，可以考虑在保证功能正确性的前提下压缩应用的资源需求，例如 SilkRoad 中利用哈希对连接状态进行压缩，使单台可编程交换机能够存储百万条连接状态。

三. 可靠性。P4 的可靠性主要指保证 P4 程序的正确性，并且 P4 程序能够被可编程交换机正确执行。首先，P4 在提升数据平面可编程性的同时也增加了数据平面漏洞出现的可能性。验证 P4 程序以消除数据平面漏洞已经成为一个重要的研究方向。其次，由于芯片实现问题，交换机可能并不完全按照 P4 程序处理报文，导致出现数据平面漏洞。保证 P4 程序在可编程交换机上的正确运行也是当前重要的研究方向。

对于网络管理员而言，P4 带来的根本性好处在于支持“自顶向下”的方式定制网络协议和功能，管理员可以通过升级交换机快速地支持新逻辑和新协议，或者是根据特定应用场景和特定业务需求修改交换机的报文处理逻辑。管理员无须更换新设备就可以在网络中部署新协议、实现新功能，减少了购置新设备的成本。在当前阶段，管理员对交换机报文处理行为编程和快速重配置的需求并不是很强，这也是目前可编程交换机没有被大规模部署的重要原因之一。并且，相对于传统交换机芯片，目前可编程交换机芯片价格相对较高，如果没有充足需求推动，可编程交换机的市场份额不会有根本上的提升。为 P4 的可编程性和可重配置性找到杀手级应用场景，是推进 P4 相关技术在工业界走向大规模部署的关键。

对于研究者而言，P4 带来的主要好处是使研究者能够做更大胆的创新，为网络研究提供了创新平

¹ 图灵完全指在可计算性理论中，编程语言或任意其他逻辑系统等可以用于通用图灵机的计算能力。换言之，此系统可与通用图灵机互相模拟。——编者注

台。P4 在保证高性能的前提下使网络更加可控，同时使网络的可感知能力更强，研究者应考虑如何利用 P4 提供的这些能力来解决当前网络存在的问题和挑战。

致谢：

本研究受国家自然科学基金(61472213)以及国家“十三五”重点研发计划(2017YFB0801700)资助。



毕军

CCF 杰出会员。清华大学教授。主要研究方向为网络空间安全体系结构、软件定义网络等。

参考文献

- [1] Bosshart P, Daly D, Gibb G, et al. P4: Programming protocol-independent packet processors[J]. ACM SIGCOMM Computer Communication Review, 2014, 44(3): 87-95.
- [2] Bosshart P, Gibb G, Kim H S, et al. Forwarding metamorphosis: Fast programmable match-action processing in hardware for SDN[J]. ACM SIGCOMM Computer Communication Review. ACM, 2013, 43(4): 99-110.
- [3] P4 Language Consortium. The p4_14 language specification[OL]. <https://p4.org/p4-spec/p4-14/v1.0.4/tex/p4.pdf>
- [4] P4 Language Consortium. The p4_16 language specification[OL]. <https://p4.org/p4-spec/docs/P4-16-v1.0.0-spec.pdf>

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

名师授课
探索教学改革

2019 CCF CCD

计算机课程改革导教班(8月10~17日)

学术主任：
李晓明 北京大学教授
CCF 会士, CCF 王选奖、杰出教育奖获得者

徐志伟
中国科学院大学教授
CCF 会士、理事
课程：计算机科学导论

杜小勇
中国人民大学教授
CCF 会士、常务理事,
卓越服务奖获得者
课程：数据库与大数据

臧斌宇
上海交通大学教授
CCF 监事, 计算机类
专业认证委员会委员
课程：计算机系统基础

- ◆ 开课时间：2019 年 8 月 10~17 日
- ◆ 开课地点：北京怀柔区中国科学院大学国际会议中心
- ◆ 报名时间和方式：3 月下旬开始报名，请关注 CCF 官网和官方微信通知

联系：李红梅 adl@ccf.org.cn 188 1066 9757

CCF专委活动计划 (2019.4~7)

4月15日 厦门

智能感知与城市计算前沿论坛

会议编号：CCF-19-TC15-01F

主办：中国计算机学会

承办：CCF 普适计算专委会 厦门大学

联系：陈龙彪 longbiaochen@xmu.edu.cn

4月20~21日 南京

2019年学科前沿高端论坛——“智慧物联网与智能计算”

会议编号：CCF-19-TC24-01F

主办：中国计算机学会

承办：CCF 物联网专委会 东南大学

联系：王帅 shuaiwang@seu.edu.cn

4月20~21日 北京

中国计算机学会生物信息学战略研讨会

会议编号：CCF-19-TC33-01S

主办：中国计算机学会

承办：CCF 生物信息学专委会 中科院计算所

联系：张法 zhangfa@ict.ac.cn

5月5日 青岛

计算机图形学和混合现实前沿研讨会 (GAMES)

会议编号：CCF-19-TC07-01S

主办：中国计算机学会

承办：CCF 计算机辅助设计与图形学专委会 山东大学

联系：孟祥旭 mxx@sdu.edu.cn

5月6~7日 青岛

第16届计算机辅助设计与图形学国际会议 (CAD/Graphics 2019)

会议编号：CCF-19-TC07-01I

主办：中国计算机学会

承办：CCF 计算机辅助设计与图形学专委会 山东大学

联系：孟祥旭 mxx@sdu.edu.cn

5月10~11日 北京

2019年第十六届全国工业控制计算机学术年会

会议编号：CCF-19-TC04-01N

主办：中国计算机学会

承办：CCF 工业控制计算机专委会

北京控制工程研究所

北京集智达智能科技有限责任公司

联系：杨桦 yangh@bice.org.cn

5月17~19日 南京

中国区块链技术与应用高峰论坛

会议编号：CCF-19-TC35-01F

主办：中国计算机学会

承办：CCF 区块链专委会 南京市浦口区人民政府

联系：牛涛 ccf_tcbc@ccf.org.cn

5月11~12日 天津

CCF 2019 International Conference on Service Science (CCF ICSS2019)

会议编号：CCF-19-TC02-01I

主办：中国计算机学会

承办：CCF 服务计算专委会 天津大学

联系：冯志勇 zyfeng@tju.edu.cn

5月17~19日 深圳

第七届全国智能信息处理学术会议 (NCIIP2019)

会议编号：CCF-19-TC19-01N

主办：中国计算机学会 中国人工智能学会

承办：深圳大学

协办：CCF 人工智能与模式识别专委会

CAAI 知识工程与分布智能专委会

联系：罗成文 nciip2019@163.com

5月22日 北京

自主可控安全高峰论坛

会议编号：CCF-19-TC13-02S

主办：中国计算机学会

承办：CCF 抗恶劣环境计算机专委会

协办：航天科工集团二院 706 所

联系：宋凌云 severe613@163.com

6月~12月 网上比赛

第七届 CCF 大数据与计算智能大赛

会议编号：CCF-19-TC32-01C

主办：中国计算机学会

承办：CCF 大数据专家委员会 数联众创等

联系：陈娟 bigdata@ccf.org.cn

7月10~13日 烟台

全国高等学校计算机教育大会

会议编号：CCF-19-TC11-01N

主办：中国计算机学会 全国计算机教育研究会

承办：CCF 教育专委会 山东工商学院

联系：吴黎兵 wu@whu.edu.cn

技术进展

AutoML：回顾与展望

关键词：自动机器学习

涂威威
第四范式

自动机器学习的研究动机

机器学习在推荐系统、在线广告、金融市场分析、计算机视觉、语言学、生物信息学等诸多领域都取得了成功，在这些成功的应用范例中，也少不了人类专家的参与。Google、Facebook、百度、阿里巴巴、腾讯等科技公司依靠其顶尖的机器学习专家团队来支撑机器学习在企业内部的各种应用，各类科研机构也在花费大量经费，维护着机器学习科学家团队。然而，对于很多传统企业、中小型企业的一般的科研机构，就很难组建出这样的机器学习专家团队，其原因是机器学习专家的缺口太大，人才短缺，人才抢夺激烈，专家团队的管理成本高昂和专家经验不可复制，等等。

为了机器学习能为更多的企业赋能，在更加广泛的场景得到应用，有没有低门槛甚至零门槛的机器学习方法，让更多的人可以在很少甚至几乎没有专业知识的情况下轻松使用，并减少机器学习应用落地对专家人才的依赖？自动机器学习（Automatic/Automated Machine Learning, AutoML）应运而生。其研究目的就是为了使机器学习过程自动化，减少、甚至完全规避人类专家在这个过程中的参与度。

理论出发点

设计机器学习算法是一件困难重重的事情，能

否找到一种通用的机器学习算法来解决所有的机器学习问题呢？这个问题在 20 多年前就被解答过，对于所有可能的问题，可以证明的是，如果所有问题同等重要，所有的算法，包括完全随机的算法，它们的期望性能是一样的，所有的算法没有优劣之分，这是著名的没有免费的午餐（No Free Lunch, NFL）定理的一个不太严谨的直观阐述。这个定理意味着寻求一种完全通用的机器学习算法是行不通的。于是，研究人员就开始针对不同的问题展开对应的机器学习研究，这导致了机器学习技术广泛应用不可复制的问题。在解决某个特例问题的机器学习算法和针对所有问题完全通用的机器学习算法之间，有一种可能性是存在可以解决某一类而不只是某一个特例的相对通用的机器学习算法。自动机器学习就是从这样的理论考虑出发，试图去寻找更加通用的机器学习算法。

目前自动机器学习研究的主要场景

静态闭环自动机器学习

静态闭环自动机器学习考虑的是静态机器学习问题，即给定固定的训练集，不利用外部知识，寻找在测试集上期望表现最好的机器学习模型。经典的机器学习流程包括数据预处理、特征处理和模型训练。自动机器学习在这三个流程中都有广泛的研

究：(1) 数据预处理中，研究数据的自动清洗、样本的自动选择、数据的自动增强、数据类型的自动推断等，以达到理解原始数据和提升数据质量的目标。(2) 对特征处理方法的研究主要包括自动特征生成和自动特征选择。自动特征生成的研究包括单特征变换、多特征组合、深度特征生成、特征学习等。自动特征选择一般会配合自动特征生成使用，先自动生成特征，再进行自动特征选择，对于复杂的特征处理，一般两者交替迭代进行。(3) 模型训练的研究一般包括自动算法选择和自动算法配置。自动算法试图从广泛的机器学习算法中选择适合问题的某一个或者某几个算法，这些算法又有很多的超参数需要配置，自动算法配置则研究如何进行超参数选择配置，比如如何配置神经网络结构，实际应用中这两者也会配合使用。

外部知识辅助的静态自动机器学习

外部知识辅助的静态自动机器学习试图借鉴人类专家选择数据处理方法、特征处理方法、模型训练算法等方式进行自动机器学习。人类专家会从以往处理过的机器学习问题中积累经验，并将此推广到之后的机器学习问题中。

动态环境的自动机器学习

动态环境下的自动机器学习研究试图解决的是数据不断积累、概念发生漂移时的问题。

核心技术

自动机器学习的研究核心是如何更好地对数据处理方法、特征处理方法、模型训练方法等基础部件进行选择、组合以及优化，以使学习到的模型的期望性能达到最优（见图1）。目前该项研究主要面临三个难点：(1) 超参配置与效果之间的函数无法显式表达，属于“黑盒”函数；(2) 搜索空间巨大，可能的处理方法和组合是指数级，同时不同处理方法拥有各自的超参数，当特征维度超过20时，其多特征组合可能的搜索空间都将远超围棋可能的状态空间；(3) 函数值的每次计算大多涉及数据预处理、特征处理、模型训练的全流程，函数值的计算代价极其昂贵。为了解决这些问题，采用的核心技术是基础搜索方法、基于采样的方法和基于梯度的方法。

基础搜索方法

搜索方法中最常见的是格搜索方法。该方法通过遍历多维参数组合构成了网格寻求最优化，容易实现，应用广泛，但是，搜索复杂度随参数维度呈指数增长，并且会将搜索浪费在不太重要的参数维度上。随机搜索方法则是对参数空间进行随机采样，各个维度相互独立，克服了维度灾难和浪费资源搜索的问题。在实际应用中，随机搜索方法往往表现得比格搜索要优秀。

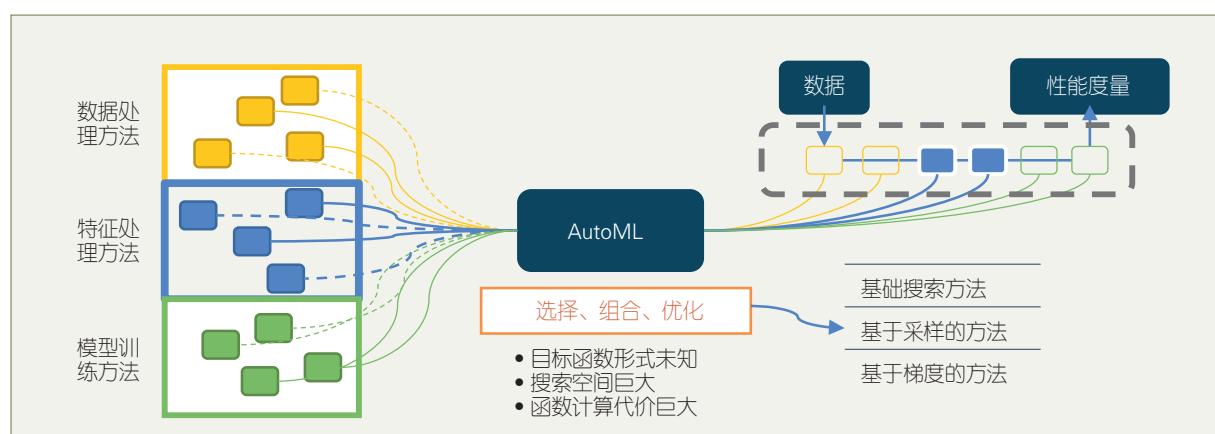


图1 自动机器学习的框架

基于采样的方法

基于采样的方法是被研究得最多的方法，大多也是具有理论基础的方法，往往比基础搜索方法表现更优。这类方法一般会生成一个或者多个对样本空间的采样点，之后再对这些采样点进行评估，根据评估的反馈结果进行下一步采样，最后寻找到相对较优的参数点（见图2）。基于采样的方法分为以下四类。

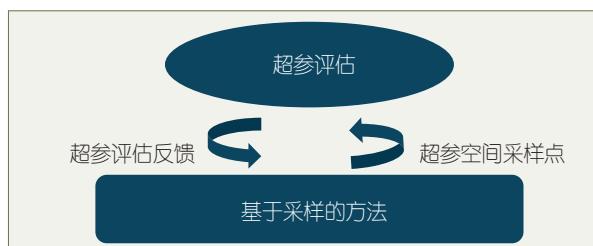


图2 基于采样的方法

基于模型的零阶优化方法

该方法试图建立关于配置参数和最终效果的模型，并依据模型来寻求最优化。这类方法一般先基于已经采样到的点的效果评估建立模型，然后基于学习到的模型采用某种采样策略来生成下一个或者下一组采样点，根据新的采样点得到的效果进一步更新模型，再采样迭代，如此寻求对黑盒函数的最优化。由于待优化的函数是“黑盒”函数，在求解过程中只能获得函数值而不能直接计算函数梯度，因此也被称为零阶优化方法（零阶是相对于传统计算一阶或者二阶梯度的优化方法）或者非梯度方法。这类方法有两个主要的关注点：模型和采样策略。构建的模型一般用来预测配置参数对应的效果。由于采样依据的模型仅仅是依据之前采样得到的点的反馈学习，对函数空间未探索区域的估计一般是不太准确的，采样策略需要在函数最优化和空间探索之间做出权衡，即在开发利用(exploitation)和探索(exploration)之间做出权衡，简称E&E。

贝叶斯优化是一种基于概率模型的方法，一般采用高斯过程、贝叶斯神经网络、随机森林等作为模型，然后采用提升概率、提升期望、交叉熵、GP-UCB等作为采样策略，这些策略都在显式或者

隐式地进行E&E。最常见的是基于高斯过程的贝叶斯优化方法，这类方法在参数维度较低、采样点较少时表现较优，但是在高维、采样点较多时就很难被使用，因此有学者尝试使用贝叶斯神经网络解决这样的问题。

基于分类方法的随机坐标收缩方法(Random COordinate Shrinking, RACOS)和基于随机坐标收缩分类模型来进行基于模型的零阶优化，有效地解决了贝叶斯优化方法的计算复杂度高、参数类型受限的问题，它一般采用最简单的 ϵ -greedy方法来进行E&E。随机坐标收缩方法被证明在高维度场景下显著优于基于高斯过程的贝叶斯优化方法。

局部搜索方法

局部搜索方法一般定义某种判定邻域的方式，从一个初始解出发，搜索解的邻域，不断探索更优的邻域解来完成对解空间的寻优。最常见的方法有爬山法、局部集束搜索等。局部搜索简单、灵活并易于实现，但容易陷入局部最优，且解的质量与初始解和邻域的结构密切相关。

启发式方法

启发式方法主要是模拟生物现象，或者从一些自然现象中获得启发来进行优化，最典型的就是基于演化计算方法。这类方法由于很少有理论依据，实际工作中很难对方法的效果进行分析。

基于强化学习的方法

这类方法能够发现一些新的神经网络结构，并被验证具有一定的迁移能力，但是由于强化学习自身的学习算法研究尚未成熟，其优化效率相对低下。

基于梯度的方法

由于对优化部件以及超参数的可微性要求较高，并且计算复杂度也高，因此，直接对优化目标进行梯度求解的方法很少使用。

研究热点

自动机器学习的研究热点是效率和泛化性。

解决自动机器学习的效率问题是自动机器学习技

术落地的关键之一。效率优化包括六类：(1) 混合目标优化，将参数点的评估代价也作为优化目标的一部分，在计算代价和效果之间做权衡。(2) 同步并行化和异步并行化。(3) 提前停止迭代，在训练早期就剔除一些表现不太好的参数，节省计算资源，比如最经典的逐次减半策略，每过一段时间都剔除其中一半不好的参数，极大地节省了计算资源（见图3）。(4) 对模型训练进行热启动，复用类似参数的训练结果，降低超参数的评估代价。(5) 对数据进行采样，采用小样本上的参数搜索来代替全样本的参数搜索，由于小样本和全样本最优参数之间可能存在着差异，有一些研究人员试图学习小样本和全样本之间的关系来进行多保真度的自动机器学习（见图4）。(6) 将超参数搜索和机器学习过程结合起来，进一步提升效率和效果，比如基于种群的方法。

机器学习关注的核心是泛化性，自动机器学习的目的也是为了提升最终学习到的模型的泛化性。

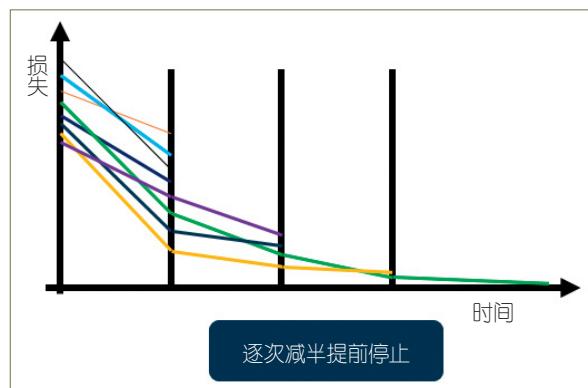


图3 逐次减半策略

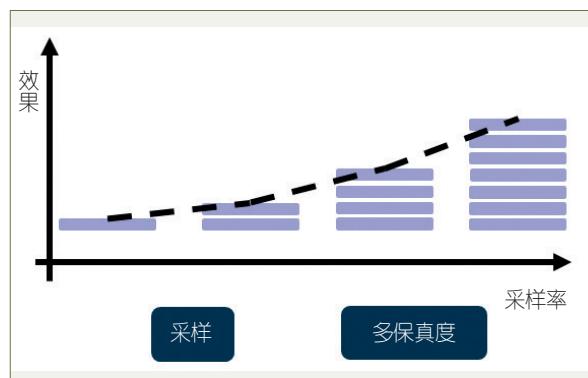


图4 多保真度的自动机器学习

如何判断自动机器学习是否提升了泛化性，一般采用切分训练集和验证集的方式进行估计。为了进一步降低过拟合到验证集的风险，有一些研究关注如何对模型的泛化效果进行更合理的估计。除此之外，由于自动机器学习往往伴随着很多次不同参数的模型学习，与最终只选择一个“最优”的模型不同，选择其中一些模型进行集成学习也是一种提升泛化性的方式。越来越多的工作混合多种效率优化和提升泛化性的策略对自动机器学习算法进行优化。

落地应用

来自不同数据之间解决问题手段的可迁移性 / 可复制性为自动机器学习的落地增加了难度。解决不同问题的手段相似性或者可迁移性 / 可复制性越高，自动化越容易，反之越难。目前自动机器学习落地的应用场景主要有图像数据和表数据。

图像数据

深度学习取得成功的领域来自图像。深度学习的核心在于“自动”学习层次化特征。以前的图像分析需要人工来做，要从原始像素中提取非常多的特征，而深度学习很好地解决了这个问题。深度学习使得特征可学习，同时将人工特征设计转变成了人工神经网络结构设计。对于这类数据，自动机器学习研究的核心是使图像领域的神经网络结构设计自动化。图像数据之间的相似性较大，原始输入都是像素，问题解决方案的可迁移性和可复用性也大，因此，自动机器学习在图像数据上的落地相对容易。

表数据

表数据是抽象数据，不同的表数据之间没有很强的相似性，不同表数据各列的含义千差万别，表数据还与实际业务密切相关，需要解决时序性、概念漂移、噪声等问题，因此自动机器学习在表数据上落地的难度较大，仅仅是自动神经网络结构设计是远远不够的。目前研究的热点还包括如何将分布在多个表中的数据自动转化成最终机器学习所需要的单个表数据。

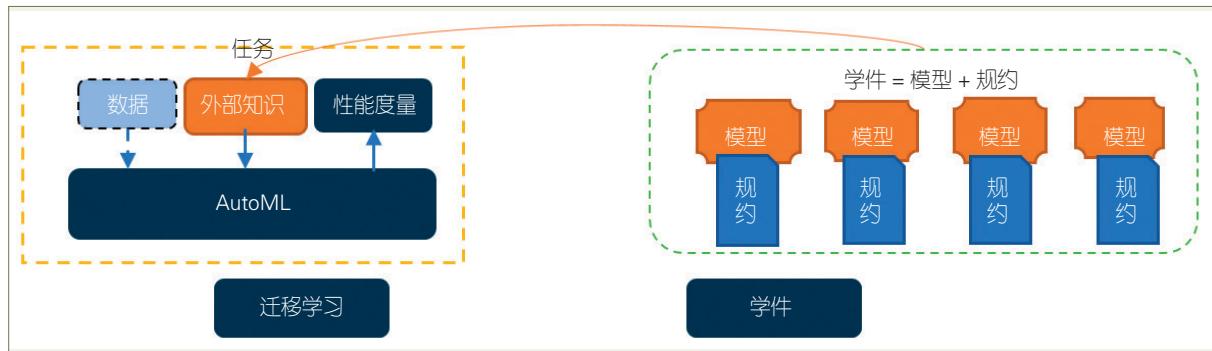


图5 迁移学习与学件

自动机器学习比赛

为了更好更快地验证自动机器学习算法，推动该领域的研究和加快技术落地，ChaLearn 和第四范式发起并组织了多届国际自动机器学习竞赛。比赛需要参赛者提交自动机器学习程序，评判采用完全盲测的方式，参赛者无法访问到最终比赛的数据，甚至包括训练数据，只能靠程序自动进行数据处理、特征处理和模型训练，在给定的计算资源和规定的时间内完成自动机器学习任务，并对测试数据集给出预测，最终依据自动机器学习程序在多个数据集上的效果进行比拼。

未来展望

算法方向

在自动机器学习算法方面，未来的工作如果能在5个方向上取得突破，将会有较大的价值。

1. 效率提升。效率可从时间复杂度和样本复杂度两方面考量。在给定的计算资源下，更高的效率在一定程度上决定了自动机器学习的可行性，意味着可以进行更多探索，还可能会带来更好的效果。另外，获取高质量有标记的样本往往是非常昂贵的，因此样本复杂度也是影响机器学习落地的关键因素之一。在外部知识辅助的自动机器学习中引入学件（学件 = 模型 + 模型的规约），利用迁移学习，是未来有效降低样本复杂度的可能方向（见图5）。

2. 泛化性。目前自动机器学习在泛化性上考虑较少，泛化性是机器学习最重要的研究方向，未来需要加强。

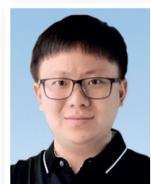
3. 全流程的优化。与目前大部分自动机器学习只研究机器学习的某一个阶段（比如自动特征、自动算法选择、自动算法配置）不同，实际应用需要全流程的自动机器学习技术。

4. 面对开放世界。现实世界不是一成不变的，自动机器学习技术需要面对开放的世界，解决数据的时序性、概念漂移、噪声等问题。

5. 安全性和可解释性。为使自动机器学习具有安全性，需要解决攻击应对、噪声抵抗、隐私保护等问题。如果自动机器学习系统被部署到实际系统中与人交互，则需要更好的可解释性。

理论方向

在自动机器学习理论方面，目前研究的甚少，对自动机器学习的泛化能力及适用性也知之甚少。因而，我们一方面要回答目前自动机器学习算法的适用性和泛化能力，另一方面也要回答哪些问题类存在通用的机器学习算法上和更广泛问题空间上的自动机器学习算法的可行性。



涂威威

第四范式资深机器学习架构师、资深科学家。第四范式先知平台大规模分布式机器学习框架GDBT的设计者，带领团队将AutoML及迁移学习应用到工业界并取得显著的效果提升。

tuweiwei@4paradigm.com

CCF 会员活动中心动态

- CCF 南京** 1月 21 日，CCF 南京召开了 2019 年首次执委会会议，制定 2019 年度工作计划和部署重点活动。执委会成员陈兵、孙国梓、肖亮、孙涵、毛莺池、王继锋、吉旭、陈永等参加了此次会议。
- CCF 长沙** 1月 22 日，CCF 长沙“2018 年度总结大会暨 2019 年工作规划会议”在国家超算长沙中心召开。来自国防科大、中南大学、湖南大学等 20 多所高校，湖南时空通道、北京并行科技等十多家公司的近 80 名会员出席了本次活动。
- CCF 无锡** 1月 21 日，CCF 无锡召开 CCF 无锡执委及候选委员工作会议。会议由分部主席漆锋滨主持，70 多人参加了会议。
- CCF 宁波** 1月 13 日，CCF 宁波举行换届选举会议。CCF 常务理事、会员与分部工委执委漆锋滨代表总部出席会议并主持换届活动，来自宁波各大高校、企事业单位的 27 名 CCF 宁波委员及会员代表参会。浙江金网信息产业股份有限公司董事长徐建昌当选新一届主席。
- CCF 合肥** 1月 13 日，CCF 合肥于中国科学技术大学召开“CCF 合肥 2019 年首次执委工作会议”。会议由 CCF 合肥主席、中国科学技术大学教授吴枫主持。
- CCF 福州** 1月 12 日，CCF 福州在福建师范大学召开分部委员扩大会议。会议由分部主席许力主持，分部委员、华为和 360 安全等企业代表以及部分 CCF 会员共 40 余人参加了会议。
2 月 23 日，CCF 福州在福建师范大学举办“智慧城市建设，数据安全先行”学术报告会。本次报告会邀请了复旦大学教授张新鹏，西安电子科技大学教授陈晓峰，武汉大学教授王骞进行学术交流。来自福州各高校、企业的 60 多人参加了此次报告会。
- CCF 兰州** 1月 6 日，CCF 兰州在兰州大学举办“金城问道（四）——西部地方高校研究生培养与对外交流合作的困境与对策”专题论坛。本次论坛执行主席为 CCF 兰州秘书长张志昌和 CCF 兰州委员、CCF YOCSEF 兰州候任主席魏霖静。
- CCF 广州** 1月 9 日，CCF 广州走进拓尔思知识图谱研究院。分部主席臧根林，副主席蒋盛益、许勇，部分执委和委员，YOCSEF 广州候任主席郝天永等参加了本次活动。CCF 广州执委黄琼、委员王进宏担任执行主席。
2 月 16 日，CCF 广州举行换届选举会议，CCF 常务理事韦竹林和 CCF 理事、CCF 女计算机工作者委员会主任卢宇彤代表 CCF 总部出席活动，来自广州各大高校及企事业单位的分部委员、会员代表和赞助企业代表 30 余人参加了会议。经过竞选演讲、现场答辩和与会委员无记名投票，华南理工大学教授许勇当选新一届主席。
- CCF 苏州** 1月 6 日，CCF 苏州在苏州广电现代传媒大厦举办“2019 苏州市计算机大会暨创赢未来硬科技峰会——从研究到产业”。
- CCF 石家庄** 1月 7 日，CCF 石家庄“网络安全论坛”在石家庄铁道大学举行。论坛由石家庄铁道大学教授朴春慧主持，CCF 石家庄主席赵冬梅致欢迎词。石家庄分部的会员以及石家庄铁道大学、河北师范大学、河北科技大学等高校的师生 200 余人参加了本次论坛。

新技术 & 新应用

利用深度学习，脑波直接转语音

1月29日, *Scientific Reports* 上发表了论文 “Towards reconstructing intelligible speech from the human auditory cortex”, 报道了一项新技术。该技术将深度学习与语音合成技术结合起来, 以重建来自人类听觉皮层的闭集可理解语音。

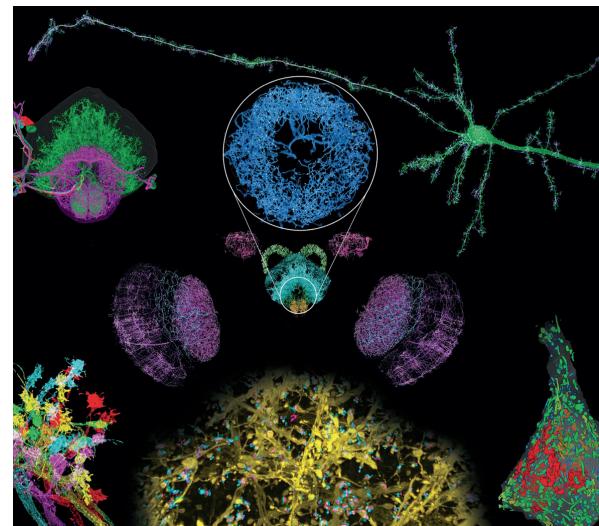
研究人员采用植入性脑皮层电图 (ECoG) 方法进行测试。在 5 位受试者脑中嵌入电极, 进入听音状态, 由考官随机读出 40 个从 0 到 9 的数字, 系统通过脑机接口用模型重建语言信息, 最后由电脑读出来。

实验结果表明, 使用所有神经频率直接估计语音合成器参数的深度神经网络模型在数字识别任务上获得了最高的主观和客观分数, 相比使用线性回归去重建听觉谱图的基线方法提高了 65% 的可理解度。这些结果证明了深度学习和语音合成算法在设计下一代语音 BCI 系统方面的功效, 这不仅可以恢复与失语患者的沟通, 而且还有可能改变人机交互技术。此外, 研究人员还开放了神经声学处理库 Nap Lib, 可用于表征语言神经网络表示的各种属性。

果蝇大脑实现纳米级成像

1月18日, *Science* 封面报道了一项里程碑研究, 来自麻省理工学院 (MIT) 和霍华德·休斯医学研究所 (HHMI) 的科学家们成功扫描了果蝇的完整大脑, 其清晰度达到了纳米级, 并且将测绘速度提高了 1000 倍。为科学家们破解大脑的奥秘提供了极为重要的研究工具。

这一里程碑研究通过结合膨胀显微镜 (ExM) 和晶格层光显微镜 (LLSM), 完成了对整个果蝇大脑中蛋白质之间的纳米级空间关系的成像。其中, 膨胀显微镜的原理是将组织样本包埋在一种吸水膨胀的聚合物中, 使得样本像气球一样膨胀, 同时保持内部结构的相对位置不变, 然后用常规显微镜对大块脑组织进行超高分辨率的蛋白质成像; 晶格层光显微镜则通过使用高度聚焦的光束, 一次一个薄



片地快速合成样本的三维图像。

人类“第三祖先”可能是杂交后代

1月16日, 来自马克斯·普朗克进化人类学研究所的 Viviane Slon 及其同事在 *Nature* 上发表论文 “Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia

and Oceania”, 表示去年 8 月首次被发现的古人类杂交体的父亲的基因组可能属于该洞穴中后来发现的丹尼索瓦人, 而母亲来自与尼安德特人关系密切的人群。

该团队在近似贝叶斯计算框架中建立了基于深度学习的人口统计模型，对数十万次模拟获得的基因组进行学习，来预测古人类的统计数据，以推断欧亚种群的进化历史。除了该报道的人类杂交体，该模型还表明存在一个古代种群对所有亚洲和大洋

洲种群的第三次渗透。这个种群要么与尼安德特-丹尼索瓦人的分支有关，要么早于丹尼索瓦人分支。

这是研究人员首次将深度学习成功用于解释人类历史，为这项技术应用于生物学、基因组学和进化学等其他领域铺平了道路。

2019 年信息战技术趋势预测

美国媒体 C4ISRNET 近日发布了多位美国军工业内人士对 2019 年信息战新兴技术的预测，其主要观点有：(1) 采购加速。美国国防部将真正开始放弃采购专有的封闭式硬件系统，取而代之的是商业软件。(2) 对协同响应的需求增加。将继续投资有助于多领域操作转型的技术，包括跨越新旧数据链的机器对机器通信，多源信息整合，帮助决策者快速分析与决策的人工智能技术，等等。(3) 转向小型网

络。美国国防部将大幅增加对小型网络、安全无线网和虚拟化计算的投资，以提高战机在战术和远征任务中的机动性和态势感知能力。(4) 音频验证。音频技术可实现安全人员的远程交互，将音频引入互联网协议视频的安全层，可使安全解决方案的整体安全性再度提高。(5) 对云端数据的攻击。(6) 对抗小型无人机技术的发展。

“AI 儿科医生”诞生

2 月 12 日，*Nature Medicine* 发布了一篇文章 “Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence”。这项人工智能程序，可以像医生一样准确地检测出测试结果、健康记录甚至手写笔记，诊断儿童疾病。

虽然机器学习分类器 (MLCs) 已经在基于图像的诊断中显示出其强大的性能，但对各种大规模电

子健康记录 (EHR) 数据的分析仍然具有挑战性。这项研究从 136 万多份门诊电子病历中提取所需数据，基于深度学习构建 NLP 模型，对数据进行语义分析，并将数据进行结构化，最后对这些完全结构化的数据库用疾病分类器处理，预测临床诊断。精确度达人类医生的 92%。

IBM 发布独立商用量子计算机

在 2019 CES 上，IBM 宣布推出 IBM Q System One，该系统是专为科学和商业用途设计的集成通用近似量子计算系统。IBM Q 系统的目标是解决当前经典系统无法处理的被认为是过于复杂的问题，帮助开发者构建量子计算机与常规架构计算机之间的接口。量子计算的未来应用包括寻找新的方法模拟金融数据，隔离关键的风险因素以进行更好的投资，或者找到跨系统的最佳路径，以实现超高效的物流和优化交付的运营。

IBM Q 由许多自定义组件组成，这些组件协同工作，可用作最先进的基于云的量子计算程序，包括：量子硬件设计稳定，自动校准，提供可重复且可预测的高质量量子比特；低温管理，提供连续冷却和孤立的量子环境；紧凑型高精度电子元件，可严格控制大量量子比特；量子固件，用于管理系统运行状况并启用系统升级，无须用户停机；经典计算，提供安全的云访问和量子算法的混合执行。 ■

(本栏目内容由动态栏目编委鲍捷提供整理)

人工智能法官和陪审团 *

关键词：人工智能 法官 陪审团

作者：洛根·库格勒 (Logan Kugler)
译者：魏书寒 汪方野 卢 瞰

人工智能正在改变法律行业。

当美国最高法院 (the U.S. Supreme Court) 院长说人工智能 (AI) 正在对这个国家的法律制度运作产生重大影响时，你需要对此予以关注。这是首席法官约翰·罗伯茨 (John Roberts) 被问到以下问题时发生的事情：

“你能预见到有一天通过人工智能驱动的智能机器将会协助法庭进行事实调查，甚至从事更具争议的事情，也就是司法判决吗？”

他的回答震惊了现场的观众。

如《纽约时报》(The New York Times) 所报道的，他说：“这正是这里每一天都在发生的，它给司法部门的工作开展带来巨大压力。”

在过去的十年里，人工智能领域经历了一次复兴。人工智能领域长期处于“人工智能寒冬 (AI winter)”，几十年来该领域的研究进展和经费都在枯竭，但其在功效和准确性方面的技术突破改变了这一切。今天，像谷歌、微软和亚马逊这样的科技巨头依靠人工智能为其当前和未来的利润提供核心动力。

人工智能不仅影响科技巨头和尖端创业公司，而且正在改变地球上最古老的学科之一——法律的应用。

人工智能已经被用于审前证据开示¹(legal discovery) 过程中文档和数据的分析，这归功于它能够比人类更快、更低代价地解析数百万个单词。仅此一项就可以将美国劳工统计局 (the U.S. Bureau of Labor Statistics) 估计的近 30 万个律师助理和法律助理的工作自动化或完全改变。然而，这只是人工智能潜在影响的开始，它现在还被用来影响实际案件的结果。

在 2017 年一个备受瞩目的案例中，一名叫埃里克·卢米斯 (Eric Loomis) 的男子被判处 6 年监禁，而这一程度上归因于人工智能算法的量刑建议。该系统分析了有关卢米斯的数据，并就其量刑向人类法官提出了建议。

毫无疑问，无论好坏，人工智能辅助法庭或许比科幻小说更具科学性。

可预测且可靠的选择？

人工智能为法律裁决带来了一些希望。

在加拿大，阿尔伯塔大学 (University of Alberta) 计算机科学系教授兰迪·戈贝尔 (Randy Goebel) 与日本研究人员合作开发了一种可以通过日本律师资

* 本文译自 Communications of the ACM, “AI Judges and Juries” , 2018, 61(12): 19~21 一文。

¹ 证据开示是审判前一方当事人向另一方当事人提供、展示案件相关的材料、事实和证据的一项司法制度。

格考试 (Japanese bar exam) 的算法。据加拿大广播公司 (CBC) 称, 该团队目前正致力于开发一种人工智能, 可以“在法律证据和裁决相互矛盾的案件中做出权衡, 并预测未来的审判结果”, 其目标是使用机器帮助人们做出更好的法律裁决。

这已经在美国法庭上进行了尝试。在卢米斯案中, 人工智能被用来评估被告人。所使用的算法由 Northpointe 公司设计并应用到名为 Compas 的软件中。该算法指出卢米斯具有“暴力风险高、再犯风险高、审前弃保潜逃风险高”等特征。尽管量刑法官被建议要注意算法的局限性, 但算法确实对他获得 6 年刑期的审判结果产生了一定影响。

电子隐私信息中心 (Electronic Privacy Information Center, EPIC) 的报告称, 与卢米斯案件采用的算法类似, 刑事审判算法 (criminal justice algorithms) 使用年龄、性别和职业履历等个人数据对量刑做出建议。该技术在美国法律体系中的使用相对普遍。

“刑事审判算法正在全国范围内使用, 只是不同的州甚至县所使用的基本工具有所不同。”电子隐私信息中心说。

使用基于人工智能的系统辅助法律程序的情况缘于机器具有比人类更公正的感知能力。“人类可能被情绪所左右, 也可能被说服。人类会感到疲劳或度过糟糕的一天。”² 电子证据开示 (e-discovery)² 专家特雷西·格林伍德 (Tracy Greenwood) 表示, 使用机器执行法律开示的过程比人类更快、更准确。

格林伍德说: “在一个犯罪高发的城市, 法官可能会以法定刑中的最高刑为基准做出严厉判决。在法庭上, 如果法官不喜欢其中一位律师, 也有可能会影响法官的裁决。”

有观点认为, 机器有潜力分析案件事实并冷静



做出判断, 在这个过程中, 不会存在人类偏见、非理性或者错误的干扰。

例如, 根据加拿大广播公司的说法, 由戈贝尔及其团队开发的日本律师资格考试 AI 现在被认为是“该领域的世界领先者”。它在日本律师资格考试中取得了成功, 而至少有一位人类, 即戈贝尔的一位同事, 没有通过这次考试。

人类的不可靠性在法律领域不是一个孤立的问题。根据英国《卫报》(The Guardian) 的一项调查, 美国的地方法院、州法院甚至联邦法院充斥着“日常隐瞒与诉讼当事人及其律师利益关系”的法官。调查还发现, 在已经启动调查的近半数 (47%) 关于法官利益冲突的投诉中, 监管机构都发现了不法行为并给予纪律处分。

然而, 监管机构一般很少对法官利益冲突相关投诉启动调查——在 37,000 多起投诉中, 有 90% 在州法院当局“没有进行任何实质性调查”的情况下被驳回。

利益冲突并不是困扰美国法律体系的唯一偏见, 显式或隐式的种族歧视也很常见。

“首先, 少数群体很少有机会进入法庭, 并且由于体制因素使得很少有人能够真正代表他们的利

² 电子证据开示是审判前一方当事人向另一方当事人提供、展示案件相关的以电子形式存储的证据的一项司法制度。

益，再加上潜意识或有意识的偏见，往往会导致更糟糕的结果。”位于大不列颠哥伦比亚省温哥华 Pulleyblank 律师事务所的创始人，奥利弗·普里布兰克 (Oliver Pulleyblank) 说。

然而，智能机器没有这些类似的缺点。作为只关注事实的冷静仲裁者，机器有潜力以一种比人类更一致、更标准化的方式影响法律裁决过程。

这种方式大有裨益。

“引入一个具有更高确定性和可预测性的系统将向更多人展示法律的效用。”普里布兰克说。诉讼的高成本和结果的不确定性阻碍了许多人采取有效的法律行动。

“很少有人负担得起诉讼，”普里布兰克说，“即使是那些能负担得起的人通常也不会去提起诉讼，因为在支付了所有费用之后经常会发现法律胜利是没有价值的。”

但是，当你更深入地了解机器协助的法律裁决时，你会发现它们可能不像看起来那样公正或一致。

由有偏见的人创造的“无偏见”机器

在卢米斯算法辅助案件中，虽然被告声称算法的报告侵犯了他的正当程序权，但没有办法检查报告是如何生成的，因为开发包含该算法的 Compas 软件的 Northpointe 公司对其工作原理保密。

“我们产品的关键是算法，它们是申请了专利的。我们创造了这些算法，但因为这是我们业务的核心，所以我们不能公开它们。”Northpointe 的高管们说，《纽约时报》报道。

这就是困扰人工智能领域的所谓“黑匣子”问题。

算法被用于海量的数据集。这些算法根据它们的“秘方”——使用数据的方法，来给出结果。放弃算法的秘方就等于放弃你的全部竞争优势。

这样做的后果是什么？大多数使用人工智能的

系统对任何人都是完全不透明的，除了它们的创造者。我们无法确定算法是如何产生特定的结果、建议或评估的。

当使用机器作为法官和陪审团时，有一个主要问题：因为我们对算法的工作原理都缺乏最基本的理解，所以只有在伤害发生之后我们才知道它们是否产生了糟糕的结果。

ProPublica 是一个“独立的、非营利的、遵守道德原则的新闻编辑部”。据其网站声称，针对在佛罗里达州布劳沃德县被捕的 7000 人，ProPublica 使用 Northpointe 公司的算法进行了“风险评

分”。这些分数用于在法庭上确定释放日期和保释金，因为据称它们可以预测被告再次犯罪的可能性。

事实证明，这些算法可能是有偏见的。

在调查的案例中，ProPublica 称算法错误地将黑人被告未来罪犯的比例设为白人被告的几乎两倍（和黑人被告相比，白人被告更经常被误判为“低风险”）。

由于这些算法的运行并不透明，因此很难判断结果是否存在评估错误，或者这些算法的创造者们是否将他们的偏见以规则的形式写进了算法里。而且除了偏见这一因素之外，算法的预测其实也不是那么准确。

ProPublica 称，“事实上，被预测会实施暴力犯罪的人中只有 20% 会真的实施暴力犯罪。”减少暴力犯罪是一件好事，但基于这一评估，得出的结论是 80% 的被告被判定为可能是暴力罪犯，而事实并非如此。

批评人士称，在被用于影响法律决策之前，算法需要更加透明。

普里布兰克说，即便如此，阻碍人工智能在法律体系中发挥更大作用的另一个巨大问题是，不能保证机器能够有效地处理法律中的细微差别。

他说，“许多法律问题要求法官在不同利益之间进行权衡。”他举了一个性侵受害者起诉袭击者的案例。法官需要平衡受害者对隐私保护的需要和

正义伸张应该公开的原则。虽然这没有简单的答案，但公布受害者姓名或不公开诉讼程序是法官必须做出的决定，而且对案件有重大影响。

“它所依赖的不是‘法律’，”普里布兰克说，“在任何情况下，如何平衡这些价值都没有明确的法律答案。相反，这取决于法官的判断。”

上述场景的判断存在于各种案件中。格林伍德说：“机器善于识别哪些事情发生过，哪些没有发生过，但它们缺乏判断能力。”他说，机器可能产生一致的结果，但缺乏确保正义得到伸张的其他关键技能。“机器不会在刑事案件中教导一个被告，让他重新开始新生活。”

普里布兰克认同使用机器使法律更“可预测”，但它可能导致的问题比它解决的问题更多。他说：“每当你试图让法律变得更可预测时，你就有牺牲公平的风险。”

在 ProPublica 的一项调查中，算法评估了两名被告。一个是惯犯，另一个是有轻微前科的年轻女孩。两人都偷了具有相同价值的物品，但机器没有场景化地考虑这样的事实，即这个年轻女孩虽然曾经偷过一辆自行车，但并没有严重犯罪记录。她被认为是一个可能的惯犯，就像职业罪犯一样。在机器看来，这两个人都犯了罪，都有前科。而这使得场景化失败，因此，该案例中的算法在这种情况下得到了非常错误的结果。

然而，引入语境和情境本质上降低了法律应用的可预测性和一致性，因此机器可预测性和人类判断之间的平衡是脆弱的。

“这种秩序与公平的对立长期以来都是法律理论的话题。”普里布兰克说。

这让普里布兰克和格林伍德得出了同样的结论：在法律行业，机器可能会大力帮助人类。虽然

法律行业会因此而转型，但机器要在法律过程中完全取代人类，可能需要去改变法律本身。

“为了允许非人工做出可预测的司法裁决，法律必须从根本上进行改变，”普里布兰克说，“如果法律没有改变，法律本身就有太多的自由裁量权，因此公众也需要接受机器行使自由裁量权的情况存在。”

尽管机器可能具有更强的预测能力，但人类将对是否使用这种机器做出最终裁决。 ■

作者：

洛根·库格勒 (Logan Kugler)

美国佛罗里达州坦帕市的一名自由科技撰稿人，已为 60 多家主要出版物撰稿。

译者：

魏书寒

CCF 学生会员。复旦大学硕士研究生。
主要研究方向为计算机支持的协同工作、
人机交互等。

18210240205@fudan.edu.cn



汪方野

CCF 学生会员。复旦大学博士研究生。
主要研究方向为计算机支持的协同工作、
算法公平性等。

18110240011@fudan.edu.cn



卢 磨

CCF 高级会员，CCCF 特邀译者。复旦
大学计算机科学技术学院副教授。主要研
究方向为计算机支持的协同工作、社会计
算与人机交互。

lutun@fudan.edu.cn



校 对：孔德建 北京航空航天大学

(本期译文责任编委：苗启广)



封面设计说明

本期主题为数据可视化分析。插图左侧为数据源的生成，接下来数据传输到右侧，
经过处理后形成可视化图表。

设计：SEEEKLAB（设计总监 田力）

读编往来

来函照登

我读过《中国计算机学会通讯》(CCCF)不少文章，许多文章很好，但也有的差些。近日我读了2018年第12期的《数字对象与互联网》一文，发现问题太多，觉得有必要说一说。下面只列举其中几个例子。

1. 第8页第一个脚注说，丘奇“对算法理论的系统发展做出了巨大贡献”。丘奇是图灵的博士生导师，他与图灵都是计算领域的奠基人。但是，说丘奇的工作是“算法理论”，无论这种说法从何而来，我都不认同。丘奇和图灵等人研究的是计算的概念和模型，创立了“计算理论”（而不是“算法理论”）。在计算机专业术语中，“计算”是最大的一个概念，涵盖了计算机领域的一切，也更抽象；“算法”则多指解决具体问题的具体方法。例如，我们讲“计算复杂性”，也说“算法的复杂性”；有“计算复杂性理论”，也有“算法的分析与设计”。两个术语讨论的是不同层次的问题，不应该随意混用。

2. 第8页右栏提到图灵“正在写一篇关于可计算数字的博士论文”。我们都应该知道，图灵的论文是关于“computable numbers”，把“numbers”说成“数字”实为大谬。“数字”是“digit”的译文，number则对应于“数”。数字只有几个（十进制数字只有10个），怎么会“不可计算”？数有无穷多（自然数、整数），还可能无穷长（实数），因此才要考虑能否计算的问题。我们说“数字”（digit）时，关心的是编码、离散化、符号化等，罗伯特·卡恩这篇特邀报告的主题是“数字对象”，也是在这个意义上使用术语“数字”。而当我们说“数”的时候，关注的是一大类基本数学对象，关注它们的计算性质和问题。数字被用于“表示”各种不同的数，在数学中如此，在计算机领域也如此。计算机的全称为“通用电子数字计算机”，其中的“数字”就是指数据采用离散的符号编码（与模拟计算机相对应）。这两个术语不能互相替代，不能混淆。这一点应该是计算机领域的常识。

3. 在第9页左栏，可以看到交替出现的“电脑”和“计算机”。在严肃的专业文章里，我不赞成使用“电脑”一词，两个词用在同一篇文章里，就更不规范了。我国计算机领域的前辈们大都反对用“电脑”这个来自港台的译文指代 computer，认为它很不科学。但后来，由于市场发展和从业人员激增，又引进了大批港台书籍等，“电脑”一词的使用日益广泛，也慢慢被接受，主要作为一种俗称（这里不谈严格性）。今天，完全拒绝这个词语是不太可能了，但我还是建议“电脑”一词不要用在专业文章中。假如把“中国计算机学会”改名为“中国电脑学会”，大家会是什么感觉？

4. 第9页左栏有一句“图灵实际上是第一个做编程机器的”。我不知道卡恩的原话是什么，但作为中文，这句话有问题。且不说图灵是否“做”过机器，只看“编程机器”这个术语就有问题。从中文意义看，这种AB形式的构词结构经常表示“(能)做A事项的B”（当A为动词时）。例如，“机修工”指修理机器的工人，“计算机”表示能做计算的机器。按这种理解，“编程机器”就是“编程序的机器”。显然，这不是卡恩的原意（否则我们就应该质疑他懂不懂图灵机了）。我猜测卡恩说的应该是“programmable machine”，正确的译法是“可编程机器”。说“编程机器”是很严重的错误。实际上，即便如此，卡恩的说法也不对。计算机先驱巴贝奇更早就提出了“可编程”的概念，并试图制造第一

台可编程的计算机器（虽然没完成）。另一方面，中国汉代的提花织机就已经是可编程机器了。我们还知道，一般的图灵机并不是可编程机器，通用图灵机才是可编程机器。

5. 文章还有些基本的中文表达问题。第9页左栏“早在20世纪20年代，甚至30年代”，这句话不符合中文的逻辑。我们说“早在”时是朝过去的方向说，“甚至”就表示更进一步。正确说法应该是“早在20世纪20年代，或者30年代”，表示更近一些的年代；或者说“早在20世纪30年代，甚至20年代”，表示年代还要更早。

6. 第9页右栏“如果你试图引用一个统一资源定位器，它可能会工作100年”。能够引用，说明该定位器已经存在，但这句话想说什么？可能是想说“希望它能工作100年”；或者不是“引用”而是“创建”定位器，因此希望“它能工作100年”。无论如何，我们希望文章的意义准确，传播正确信息。但这里没有做到。

该文还有其他问题，无法一一列举。出现这样的问题，一方面原因是整理者不够认真，很多细节简单查查就能弄清楚。另一方面原因是编辑工作不仔细。我不了解CCCF的出版流程，但我认为，必要的专业性审查对于保证文章的基本专业质量是不可或缺的。文中也有不少（可能的）文字问题，例如第8页的“磁带的及时前进和后退”，我不知道这里的“及时”要表达什么意思。

社会上有不少计算机专业书籍和文章中的术语表达混乱，有些专业教师不认真或能力不足，社会培训机构参差不齐，都传播了大量不清晰、不准确、不规范的术语和说法，甚至是技术和方法。为了计算机领域的良好发展，在专业领域中拨乱反正是CCF的责任。我们应该在能控制的范围内做好这件事，包括CCCF，也包括CCF各专委举办的学术会议。我呼吁各位专家在审查各种中文稿件、项目申请书、学生论文和专业书稿时要把好关。

裘宗燕（CCF杰出会员，北京大学教授）

编辑部回复：

非常感谢裘教授对2018年第12期《数字对象与互联网》一文提出的意见。

出现这些问题有两方面原因，一方面是我们编辑工作不够仔细，加之我们的专业水平不够，对一些专业问题不够敏感，没能检查出这些问题。另一方面该文是整理者根据CNCC特邀报告现场录音进行的整理，成文后缺少对文章的专业性审查。

对于出现这些问题，我们会深刻检讨，优化以后的工作流程，邀请业内专家撰稿（译稿），并强化文章的审查机制，加强学习，提高我们的专业水平，避免以后再出现类似问题。

也欢迎其他广大读者对我们的工作提出意见或建议。

建言献策

◆ 关于专题，建议编辑部提前列出一些候选主题，在CCF网站、会员邮件或CCCF公布，让广大会员遴选，这样就能选出会员最关心的一些专题，然后再组织专家撰写，使得专题的针对性更强。另外，有些专题的内容过于专业化。在写法上建议最好不要用项目申请书的方式去写，

也不要蜻蜓点水式地对很多文献进行评论，而要对最关键的文献进行综述。作为一组专题来展示某一项技术领域的现状，还需要有更深入一步的思考。比如第2期的专题《大数据共享与交易》，针对大数据交易，谁是主体？谁是交易市场的维护者和公平交易规则的制定者？参与交易的各方

的利益关注点在哪里？他们之间有没有冲突？这些需要用技术手段解决的关键问题，才是我们的出发点和落脚点。如果仅针对现象罗列技术，本身无可厚非。但是从严谨的专业角度，还是要有效实现理论构思到产品设计以及最佳实践的商业闭环。

编辑部回复：谢谢建议。本期即刊登征集CCCF专题选题的广告。对于文章的写法，我们也会向特邀编辑及作者提出要求。

◆ 诚如戴维·阿兰·格里尔教授所言，文章能够架起交流的桥梁。但是要想通过文字来清晰地表述自己的想法是件很见功底的事。是否可以请戴维就“如何通过写作来表达想法”做几期连续的专栏呢？

编辑部回复：谢谢建议。相信很多读者都想“通过文字来清晰地表述自己的想法”。我们尝试邀请戴维就“如何通过写作来表达想法”做几期连续的专栏。

第2期卷首语《科学、技术和工程》

子德说技术是第一生产力，这是对的。依据不知道原理的“秘方”，通常也能产生成果。就好像中医，作为典型的经验科学，就有足够多的“成方”，能依据“成方”治病救人，这就是生产力。

但我依然觉得科学技术是第一生产力的说法更为精确。技术作为生产力，在不懂科学原理的情况下，确实会有技术发明，但是这种发明难以走远。比如前文所述“成方”，或者一些“祖传秘方”，后人只知道用这个方子可以配出来相应的东西，然而却因为不了解方子背后的科学原理，无法更换秘方中的任何一种原材料，导致在传承过程中因为各种

意外因素最终湮灭。但是，如果掌握了方子背后的科学原理呢？结果则可能不同。这里的“科学技术”其实可以理解为“科学的技术”，即基于科学原理的技术。任何一项技术发明，在其背后一定是有可以解释的科学道理的。如果说“科学无法解释”，那是因为我们现阶段掌握的科学知识还不足以解释该现实。

子德的这篇文章把科学、技术、工程条分缕析，对三者各自的功能、范畴做出了清楚的解释，正本清源，对未来相关评价标准和评价体系的建立有非常好的指导意义。

专栏《量子计算五人谈》

现今被规模化部署的商业化高科技信息类产品，其技术原理，都是在15年甚至更早之前，被科学家和研究人员在数学上、实验室中实证阐明的。针对量子计算，我们只是在一些理论上有所突破，距离完全阐明还有一些距离。当传统计算机遇到一些技术瓶颈的时候，我们如果不能合理控制内心的

欲望，很有可能就不愿意在原来道路上艰苦攀登，而是另找捷径。针对大规模工业化，考虑到产出和投入的比例，在我国发展资金还不富裕的今天，还不能承受一定规模的试错成本，为此，还是要沉着冷静，多与全球领先的业界同仁交流，取长补短。

（本次参与评刊的有：陈盈、范天龙、李睿、李振华、刘宇擎、吕腾、时成阁、万江平、王波、易小琳、周果）

勘误

2019年第2期第14页左栏第18行，“分别用于判断第*i*个用户对第*j*个帖子的兴趣”应该改为“分别用于判断第*i*个用户对第*j*个帖子的兴趣”。第15页右栏倒数第7行，“设从公众号转发第*j*条内容”应该改为“设从公众号转发第*j*条内容”。在此，特向作者和读者致歉。



只有结成群体 才好发展专业

加入CCF / CCF会员资格延续

专业会员/高级会员/杰出会员/会士:200元/年(一次可交纳5年)

学生会员:50元/年



▲ YOCSEF
● CCF会员活动中心
▼ CCF学生分会

欢迎 微信支付

其他缴费方式

在线缴费 www.ccf.org.cn

银行转账

开户行: 北京银行北京大学支行

户 名: 中国计算机学会

账 号: 0109 0519 5001 2010 9702 028



微 信 支 付



CNCC

2019 中国计算机大会

China National Computer Congress 2019

10.17~19 苏州金鸡湖国际会议中心

CNCC近年主要赞助单位(排名不分先后)



010-6260 0336
cncc@ccf.org.cn
cncc.ccf.org.cn

