

Matrix

NO.60

2022年1 - 3月

AI挺进生命科学领域，
分子动力学模拟加速新冠
病毒致病机理研究进程

ACM/IEEE Fellow 邱铨力博士
担任微软亚洲研究院（上海）负责人

Swin Transformer 迎来30亿参数的
v2.0，我们应该拥抱视觉大模型吗？

01 焦点

- ACM/IEEE Fellow 邱锺力博士担任微软亚洲研究院(上海)负责人 2
- 微软亚洲研究院副院长刘铁岩博士获选 2021 ACM Fellow 3
- 微软亚洲研究院常务副院长张冬梅获 2022 年度 MSR Foundational Contribution Award 3

02 前沿求索

- AI 挺进生命科学领域, 分子动力学模拟加速新冠病毒致病机理研究进程 4
- 微软亚洲研究院多模态模型 NÜWA: 以自然语言创造视觉内容 7
- Swin Transformer 迎来 30 亿参数的 v2.0, 我们应该拥抱视觉大模型吗? 10
- 语音识别的快速纠错模型 FastCorrect 系列来了! 14

科研第一线

- 你真的了解计算生物学和 AI for Science 吗? 16
- 如何亿点点降低语音识别跨领域、跨语种迁移难度? 16

03 文化故事

- 科学匠人 | 白静: 拥抱变化, 不断发现计算机科学中的新天地 17
- 她们非同凡响的“另一面” 20
- 星跃重洋 | 刘国栋: 非典型理工男在微软亚洲研究院的科研“旅”记 23

04 媒体报道

- 周礼栋: 成熟的 AI 不再只是谈“颠覆” 26
- 微软云强劲增长的背后, 是全新的人工智能黑科技 28
- 中国科学家利用计算生物学开展新冠病毒系列研究揭示病毒侵染机理 31

ACM/IEEE Fellow 邱锺力博士担任微软亚洲研究院（上海）负责人

2022年1月17日，微软亚洲研究院宣布，无线及移动网络领域的国际顶级专家邱锺力博士正式加入微软亚洲研究院，担任副院长一职，主要负责微软亚洲研究院（上海）的研究工作，以及与产学研各界的合作。加入微软亚洲研究院之前，邱锺力博士在美国得克萨斯大学奥斯汀分校担任计算机系教授，并担任国际计算机学会无线及移动系统专委（ACM SIGMOBILE）的主席。

对于邱锺力博士的加入，微软亚洲研究院院长周礼栋博士表示，“邱锺力博士拥有世界级的卓越技术实力和杰出的团队领导能力。她将为微软亚洲研究院带来新的国际化视野和与众不同的领导活力，并且与我们优秀的同事们一道，在研究院多元和包容的环境中打造令人兴奋的新想法、新合作和新技术。我们非常期待邱博士能够带领微软亚洲研究院（上海）塑造致力于长期成功的文化，设立驱动突破性创新和影响力的研究方向，培养优秀人才，确保他们在追求理想的过程中持续成长。”

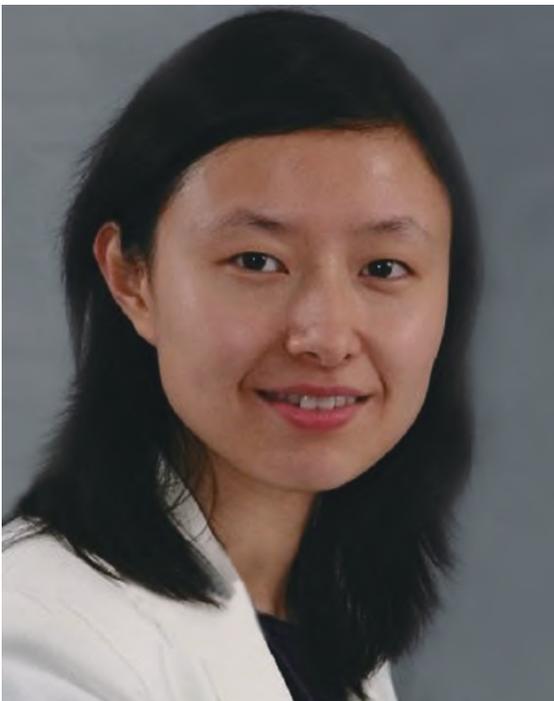
邱锺力博士生长于上海，在美国康奈尔大学先后获得计算机硕士及博士学位。2001-2004年，邱博士曾任微软雷德蒙研究院

系统和网络组研究员。2005年她加入美国得克萨斯大学奥斯汀分校任计算机系助理教授，之后晋升为终身正教授、博士生导师。邱锺力博士是全球为数不多同时拥有国际计算机学会院士（ACM Fellow）和国际电子电气工程师学会院士（IEEE Fellow）称号的华人。

邱锺力博士表示，“微软亚洲研究院人才济济，通过广泛的合作发表了众多卓越的科研成果，并为技术落地提供了广阔的平台，持续赋能全球产品和整个社会。我很荣幸、也很高兴加入微软亚洲研究院。在AI+5G这个振奋人心的时代，我期待与优秀的科研人员一道共同推动技术创新的边界，充分发挥本地优势，促进更多跨领域、跨行业的深度合作、融合发展。”

微软亚洲研究院（上海）自2018年宣布落户上海市徐汇区以来，持续吸引全球顶尖的计算机基础研究和应用创新人才的加盟。此次，邱锺力博士的加入将加速微软亚洲研究院（上海）的发展，为进一步吸引全球顶尖人才起到带头作用，同时也为科技创新与转化、产学研合作、人才培养做出重要贡献。

邱锺力博士简介



邱锺力博士，现任微软亚洲研究院副院长，主要负责微软亚洲研究院（上海）的研究工作，以及与产学研各界的合作。

邱锺力博士是无线及移动网络领域的国际顶级专家，曾在2001-2004年任微软雷德蒙研究院系统和网络组研究员。2005年，她加入美国得克萨斯大学奥斯汀分校（UT Austin）任计算机系助理教授，之后因其在互联网和无线网络领域的卓越成就被晋升为终身正教授、博士生导师。

邱锺力博士是国际电子电气工程师学会院士（IEEE Fellow），国际计算机学会院士（ACM Fellow）。同时，她还担任国际计算机学会无线及移动系统专委（ACM SIGMOBILE）的主席。邱锺力博士曾获得 ACM 杰出科学家（ACM Distinguished Scientist）、美国国家科学基金会杰出青年学者奖（NSF CAREER Award）等多项荣誉。

邱锺力博士生长于上海，在美国康奈尔大学先后获得计算机硕士及博士学位。



微软亚洲研究院副院长刘铁岩博士 获选 2021 ACM Fellow

2022年1月，国际计算机学会（Association for Computing Machinery，简称 ACM）公布了2021年度国际计算机学会院士（ACM Fellow）名单，以表彰来自世界各地高校、企业和研究机构的研究人员，在算法、计算机科学教育、密码学、数据安全和隐私、医疗信息学以及移动和网络系统等领域的贡献。微软亚洲研究院副院长刘铁岩博士因其在机器学习算法及应用方面所做出的卓越贡献获选 2021 ACM Fellow。

扫描二维码了解更多信息



微软亚洲研究院常务副院长张冬梅 获 2022 年度 MSR Foundational Contribution Award

微软亚洲研究院常务副院长张冬梅与北京大学教授谢涛，于2022年4月共同获得了由软件工程领域数据科学和 AI 方面的重量级学术会议 Mining Software Repositories (MSR) 2022 所颁发的“奠基性贡献奖”（Foundational Contribution Award）。两位科学家非常有远见地携手开创了软件分析学（Software Analytics）这一新领域，并在之后的十余年中与团队不断耕耘，持续推动该领域在研究和实践上的发展，在学术界和产业界都产生了深远影响。

扫描二维码查看张冬梅博士个人页面



AI 挺进生命科学领域

分子动力学模拟加速新冠病毒致病机理研究进程

我们身体的三分之一是蘑菇？因为人类与真菌共享三分之一的 DNA。我们一直在与异类共生？因为人体内有一半外来细胞。

这些看似不相关的现象其实都有着深层次的联系，随着科学家们的研究探索，生命神奇的本质正在逐渐被揭开。而近年来大数据、AI 等技术的发展和应用，更是为生命科学研究开启了新范式。利用新技术，科学家们可以模拟瞬间变化的生命现象、发现生命机理的规律、降低研究成本、获得更好的研究结果。近日，微软亚洲研究院就与清华大学合作，利用分子动力学模拟技术，取得了新冠病毒机理研究的重要成果。

不同领域的科学家协同合作的秘籍是什么？如何在 AI for Science 的趋势中拔得头筹？让我们从微软亚洲研究院与清华大学的合作分享中一探究竟吧。

新冠疫情自爆发以来，已造成全球范围内近 2.8 亿人感染，540 多万人死亡，给全球的经济和社会生活带来了巨大的损失和伤害，且至今仍未有缓和的迹象。相比之下，2003 年的 SARS 疫情持续一年多，累计报告病例 8000 多例，死亡 900 多人；2012 年的中东呼吸综合征则主要在中东地区流行。同样是冠状病毒所引起的传染病，为什么新冠病毒有如此高的传染性？它又是如何感染人体的？

面对这场病毒遭遇战，全球的科学家们迅速行动对新冠病毒展开研究，同时也推动了人工智能等新技术与生命科学之间的加速融合。近两年来，微软亚洲研究院的研究员们也一直在思考，如何利用自身在人工智能、深度学习等计算机领域的优势，与生物学、病毒学专家深度合作，结合生命科学的专业知识，为缓解新冠疫情贡献自己的力量。就在不久前，微软亚洲研究院与清华大学生命科学学院以及传染病研究中心合作，在新冠病毒的跨领域、跨学科研究中取得了两项重要成果，为厘清新冠病毒机理提供了新的方向。

新冠病毒致病机理研究两开花

计算生物学潜力凸显

研究发现，COVID-19 新型冠状病毒是由 SARS-CoV-2 病毒所引起的。和其他冠状病毒一样，它的表面由刺突糖蛋白结构组成，也就是 S (Spike) 蛋白。若病毒想要进入人体细胞，S 蛋白就需要与人体细胞的受体结合。S 蛋白的构型很像英文字母“Y”，竖着的 S2 区域起支持作用，向上伸出的两个枝杈，一个是 RBD，另一个是 NTD。科学家们已经认识到直接造成感染的是 RBD 区域，而且它的状态是站立 (up) 还是躺平 (down) 会直接影响受体结合，只有站立时 RBD 才能进行受体结合，从而感染人体。

基于这些背景知识，微软亚洲研究院的研究员们产生了一连串的疑问：RBD 的功能已经清楚了，那 NTD 在感染过程中扮演着怎样的角色？在病毒侵染的过程中 NTD 对 RBD 的状态变化是

否有协同作用？如果找到了 RBD 站立与躺平的规律，是不是就有可能抑制病毒的入侵？因此，研究员们希望利用计算生物学，特别是分子动力学模拟技术对 NTD 展开深入研究。当他们把这一想法与清华大学生命科学学院龚海鹏教授讨论后，双方立即开启了合作研究。



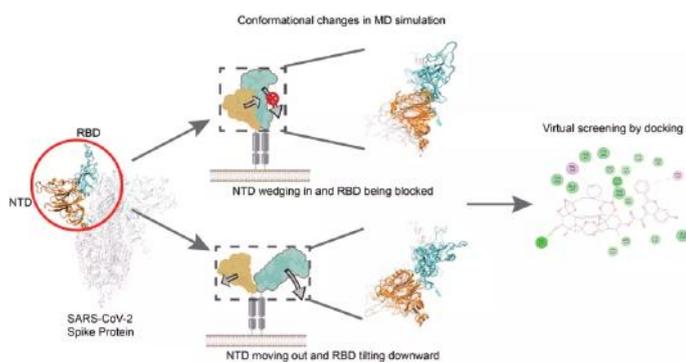
龚海鹏教授在微软亚洲研究院做研究分享

经过分析，研究员们发现，以往很多研究只对 RBD 或 NTD 的一小部分进行了模拟，只见树木，不见森林，无法还原其在整个 S 蛋白上的变化情况，模拟精度也有所欠缺。虽然只是对一个蛋白质进行模拟，但其中包含了百万级的原子数，计算量和复杂度可想而知。对此，微软亚洲研究院的研究员们采用了增强采样、加速算法等手段，基于强大的计算平台，建立了大体系、全原子的分子动力学模拟模型，实现了长时间的计算。

1、大体系、全原子是指构建拥有百万级原子的完整 S 蛋白，而不是只对 10 万个或 1 万个点进行抽象模拟，从而提升模拟精度。

2、长时间指研究员们通过数十亿步的计算，每步代表 1 飞秒 (1 秒的一千万亿分之一)，模拟运行了 20 微秒。不能小看这个数字，20 微妙相当于 $2 * 10^{11}$ 步，在分子动力学模拟中这属于相当长的时间，以此可以更真实地模拟 NTD 和 RBD 之间的相对运动。

最终，微软亚洲研究院首次提出了 NTD 在病毒侵染过程中发挥调控作用的“楔形”模型，相关成果于 2021 年 10 月在著名期刊《Advanced Theory and Simulations》上作为封面文章发表。“其实 RBD 是倾向于躺平的，这和人一样，躺着肯定更舒服，但当 RBD 想躺下的时候，NTD 会像楔子一样堵住 RBD 下方的空隙，从而使其维持站立的状态，感染人体。”微软亚洲研究院主管研究员王童形象地解释了他们从模拟中取得的发现。



NTD 在 SARS-CoV-2 的 S 蛋白构象变化中发挥调控功能示意图

利用这种“楔形”模型，研究员们进一步再对中草药数据库 TCMS 中的中药化合物进行虚拟筛选，检测到了 8 种中药中的 18 种化合物与 NTD 作用的该位点具有很强的结合能力，从而为新冠病毒药物研发提供了一定的参考价值。

像这样利用计算机模拟的方式去做生物学实验，甚至去预测和推论，被称为“干实验”。但生物学研究还是不能离开“湿实验”，也就是基于分子、细胞、生理等层面的生物实验。在开展 NTD 探索性研究的同时，王童了解到清华大学王新泉教授和张林琦教授的团队正在合作开展新冠病毒致病机理的研究。于是三方一拍即合，通过清华两位老师团队的结构生物学和免疫学实验发现，与其他冠状病毒相比，新冠病毒 S 蛋白 372 号位点的突变使得 370 号位点缺失了糖基化。这一变化促使 RBD 更多处于站立状态，增强了病毒的感染性。而微软亚洲研究院利用分子动力学模拟等计算手段进一步分析了 S 蛋白 370 位点糖基化对 S 蛋白构象变化和病毒感染能力的影响。最终，通过干湿结合的手段验证了结论的正确性，相关论文也已被生物学领域的顶级期刊《Cell Research》接收。

对于这项三方合作开展的前瞻性科研工作，张林琦教授表示，“我们在众多信息和生命活性的相互作用中找到了一个极其重要的点。它是在大量数据分析、实验验证以及预测的基础之上得到的结果。通过与微软亚洲研究院合作，我们看到将计算机科学与生命科学系统对接，可以加速找到生命现象的关键环节，解决一些生命科学的问题，并进一步了解生命科学本身，从而对研发新药物来阻断或者促进某些生命现象起到了标杆性的作用。”

AI 为生命科学研究 开辟新方向、开拓新产业

正如张林琦教授所言，AI、大数据等创新手段与生命科学的深度融合正在为生命科学研究开辟新的方向，甚至改变生命科学研究范式。生命科学研究发展至今，经历了不同的阶段，从 20 世纪前的描述观察，到 20 世纪的实验分析，在科学家们的努力下，生命的密码正在逐渐被破解。但这些传统生物学研究方式依赖于不断地试错和积累，不仅耗资巨大，周期往往也很长。同时，基因组学等底层数据采集技术的发展以及药物试验中持续产生的数据等等，也让生物数据呈现爆发式增长。虽然这为个性化的靶向药物研发、精准医疗提供了可能，但海量数据也注定了单靠人力完成数据的整理、分析和挖掘已是不可能完成的任务。



张林琦教授（左），刘铁岩博士（右）

如今，随着算力的提升、机器学习等模型的精进，大数据使得计算生物学的研究条件越来越完善，在基础科学研究中扮演着越来越重要的角色。对于 AI 与生命科学的结合，龚海鹏教授说道，“我们能不能从湿实验得到的数据中发现规律？人的逻辑思维可以有一个大致的判断，但还不够细致，AI 在这方面就能体现出它的优势。”对此张林琦教授也表示认同，他认为生命科学不能只靠感觉，而是要朝定量化和精准化的方向发展，“湿实验看到的结果往往是静态的，但所有的生命过程都是动态的，分子结构变化更是瞬时反应，在自然条件下一闪而过，人的肉眼连看到的机会都没有。在模拟分子动态变化以及定量评判方面，一些新的算法和技术能发挥非常大的作用。”张林琦教授说。

跨领域、交叉学科协同合作 打破次元壁的秘籍

尽管计算机科学与生命科学的跨界合作大有可为，但协作过程还需要更多的磨合。两个领域的科学家所面对的是两类不同的知识结构、语言体系，如何打破行业壁垒、共建合作生态是关键。微软亚洲研究院与清华大学通过上述两项合作研究，为跨学科交叉实践积累了一定的经验。

那么不同背景的科学家协同合作的秘籍是什么？首先，明己之长，知己所短，优势互补。张林琦教授长期专注于艾滋病等人类重大病毒性传染病的致病机理、抗病毒药物、抗体和疫苗的研究；王新泉教授的主要研究方向是结构生物学；龚海鹏教授则致力于把分子动力学模拟等新方法用于分析生物大分子的大尺度构象变化。他们及团队在各自的领域都有着深厚的积淀和世界级的影响力。这些专家对生命科学专业、前沿的洞察为算法提供了实现基础，可以帮助算法专家理解数据背后的科学意义。而微软本身是以计算机技术为核心能力的平台公司，在人工智能、云计算等领域能为其他学科提供强有力且最先进的计算机科学加持。



王新泉教授在微软亚洲研究院做研究分享

“微软亚洲研究院在生物学、材料科学、物理和化学方面并不是专家，所以我们需要与真正的领域专家共同努力、密切合作。在这个过程中，双方会互相影响，相互改变。AI 科学家可以提供基于数据的端到端解决问题的思路，提供比传统科学计算更加高效的解决方案；自然科学领域的学者则可以提供独到的领域知识，让这些计算能力以一种符合科学规律的方式用到刀刃上。”微软亚洲研究院副院长刘铁岩表示。

其次，跨领域合作需要提出最具有前瞻性、挑战性的科学问题。只有前沿课题才能发挥双方实力，激励科研人员克服困难，合理调配资源。刘铁岩表示，“虽然人们认为 AI 能够在任何领域发挥作用，但如何找到关键的科学问题才是关键所在，这需要领域专家与 AI 专家坐下来细致地讨论，不断淬炼出真正重要的问题。”合作伊始，微软亚洲研究院的研究员们与清华大学的师生团队也遇到了预期不匹配、沟通鸿沟等问题。通过随后定期的会议与学术讨论，双方逐渐明确了彼此的优势所在，找到了“最难啃的骨头”。当实验结果出现差异时，大家会从不同角度共同分析问题产生的原因，不断磨合，增强了彼此的信任。

最后，是要有耐心与恒心。生命科学研究是一个漫长而枯燥的过程，很多基础研究短时间内都无法带来直接的收益。对此龚海鹏教授认为“做科研需要踏踏实实。解决生物学的实际问题，要以推动科学发展为目标，而不是以发论文为目标。微软亚洲研究

院在提供强大的计算资源、AI 算法的同时，在合作研究中也极具耐心，这是跨领域合作的基础。”

在双方的合作中，大家也加深了对彼此所在行业和机构的理解。在合作之前清华大学的老师们还有些疑虑，“在我们眼中，企业的研究部门更多的是以短期业绩为导向的。但合作之后我们发现微软亚洲研究院是一个真正的学术机构，尤其是‘顶天立地’的价值取向和学术定位与清华大学的理念非常吻合，也只有这样才能开展更具学术性的研究合作。”王新泉教授说。

无论是用深度学习优化大气污染排放量、把 Graphormer 用于催化剂设计、神经网络用于新物理发现，还是近期 AI 领域顶会 NeurIPS 上火热的科学相关主题演讲，都昭示着 AI for Science 已经成为一种趋势。计算机科学、人工智能与生命科学、生物医药、量子科学、天文学等一系列基础科学研究交织碰撞，将为科学发展注入新的强劲动力。而在这一浪潮中，微软亚洲研究院也将继续与科学界合作，取得更加亮眼的成绩。

相关阅读

[扫描二维码查看文章](#)

NTD 的深度研究，为厘清新冠病毒机理提供新方向！



科学匠人 | 用 AI 打开生物学研究的一扇窗



公开催化剂挑战赛冠军模型，通用 AI 分子模拟库 Graphormer 开源！



微软亚洲研究院多模态模型 NÜWA：以自然语言创造视觉内容

此前我们曾提出了一个问题：从文字脚本生成创意视频一共分几步？微软亚洲研究院的开放领域视频生成预训练模型给出了答案：只需一步。现在，我们追问：除了文字生成视频之外，还有哪些途径可以生成视频？我们能否使用自然语言对视觉内容进行编辑？微软亚洲研究院最新推出的多模态模型 NÜWA，不仅让视觉内容创造多了一条路，甚至还让 Windows 经典桌面有了更多的打开方式。

人类对于信息的感知有五种途径，包括视觉、听觉、嗅觉、触觉和味觉，其中视觉是接受信息的最主要渠道，也是创造力的源泉。在推动人工智能发展的道路上，计算机视觉已经成为一个重要的研究领域，尤其是近几年视觉创作类应用的频繁涌现，让创作变得越来越便捷，越来越多的用户可以用这些工具制作和分享身边的美好生活。与此同时，视觉类应用的广泛使用也促进了计算机视觉领域的研究。

然而，尽管这些工具功能强大，但仍有不足之处：其一，它们需要创作者手动收集和处理视觉素材，导致现有的大规模视觉数据中所包含的视觉知识无法自动地有效利用。其二，这些工具往往是通过图形界面与创作者交互，并非自然语言指令，因此对于一些用户来说具有一定的技术门槛，他们需要拥有丰富的使用经验。在微软亚洲研究院看来，下一代可视化内容创建工具应该能够利用大数据、AI 模型帮助用户更便捷地进行内容创作，并使用自然语言作为更加友好的交互界面。

在这样的理念下，微软亚洲研究院在视频生成预训练模型的基础上进行再创新，开发了多模态的 NÜWA (Neural visUal World creAtion) 模型。通过自然语言指令，NÜWA 可以实现文本、图像、视频之间的生成、转换和编辑，帮助视觉内容创作者降低技术门槛，提高创造力。同时，开发者也可以利用 NÜWA 构建基于 AI 的视觉内容创造平台。

凭“一己之力”完成多项视觉内容创造任务

NÜWA 模型提出了一种全新的 3D 编码器 - 解码器框架。编码器可以支持包括文本、图像、视频，或者草图等多种不同的输入条件，甚至是部分图片或部分视频，让模型补全后续的视频帧；解码器则将这些输入条件转换为离散的视觉标记，根据训练数据输出图像、视频内容。

在预训练阶段，研究员使用了自回归模型作为预训练任务来训练 NÜWA，其中 VQ-GAN 编码器将图像和视频转换为相应的视觉标记，作为预训练数据的一部分。在推理阶段，VQ-GAN 解码器会基于预测的离散视觉标记重建图像或视频。

NÜWA 还引入了三维稀疏注意力 (3D Nearby Attention, 3DNA) 机制来应对 3D 数据的特性，可同时支持编码器和解码器的稀疏关注。也就是说，在生成特定图像的一部分或者一个视频帧时，NÜWA 不仅会看到已经生成的历史信息，还会关注与其条件所对应位置的信息，比如，在由视频草图生成视频的过程中，生成第二帧时，模型就会考虑第二帧草图对应的位置是什么，然后按照草图的变化生成满足草图变化的视频，这就是编码器和解码器的同时稀疏。而此前的工作通常只是一维或二维的稀疏关注，而且只在编码器稀疏，或只在解码器稀疏。通过使用 3DNA 机制，NÜWA 的计算复杂度得到了简化，提升了计算效率。

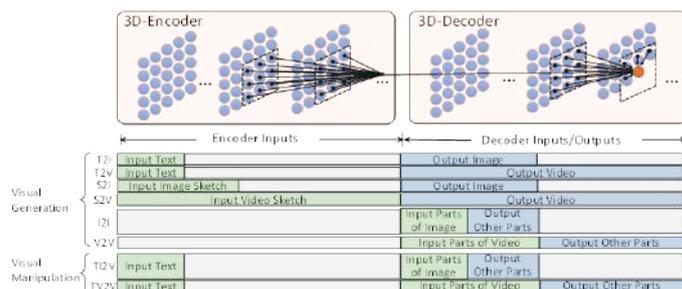


图 1: NÜWA 基于 3D 编码 - 解码架构

为了支持文本、图片、视频这些多模态任务的创建，跨越不同领域数据的鸿沟，研究员采用了逐步训练的方式，在预训练中使用不同类型的训练数据。首先训练文本 - 图片任务和图片 - 视频任务，待任务稳定后，再加入文本 - 视频的数据进行联合训练，而且研究员们还使用了视频完成任务，根据给定的部分视频作为输入生成后续视频，使得 NÜWA 拥有强大的零样本视觉内容生成与编辑能力，实现图像、视频内容的增、删、改操作，甚至可以对视频的未来帧进行可控调整。

微软亚洲研究院高级研究员段楠表示，“NÜWA 是第一个多模态预训练模型。我们希望 NÜWA 可以实现真实世界的视频生成，但在训练过程中模型会产生大量的‘中间变量’，消耗巨大的显存、计算等资源。因此，NÜWA 团队与系统组的同事们联手协作，为 NÜWA 在系统架构上设置了多种并行机制，如张量并行、管道并行和数据并行，使得我们的跨模态训练成为可能。”

NÜWA 覆盖了 11 个数据集和 11 种评估指标。在文本到图像生成的弗雷切特起始距离 (Frechet Inception Distance, FID) 指标上, NÜWA 的表现超过了 DALL-E 和 CogView, 在视频生成的 FVD 指标上超越了 CCVS, 均取得了当前 SOTA 结果。其中, 测试结果如下:

Table 1. Qualitative comparison with the state-of-the-art models for Text-to-Image (T2I) task on the MSCOCO (256×256) dataset.

| Model | FID-0↓ | FID-1 | FID-2 | FID-4 | FID-8 | IS↑ | CLIPSIM↑ |
|--------------|------------|-------------|-------------|-------------|-------------|-------------|---------------|
| AttnGAN [47] | 35.2 | 44.0 | 72.0 | 108.0 | 100.0 | 23.3 | 0.2772 |
| DM-GAN [52] | 26.0 | 39.0 | 73.0 | 119.0 | 112.3 | 32.2 | 0.2838 |
| DF-GAN [36] | 26.0 | 33.8 | 55.9 | 91.0 | 97.0 | 18.7 | 0.2928 |
| DALL-E [33] | 27.5 | 28.0 | 45.5 | 83.5 | 85.0 | 17.9 | - |
| CogView [9] | 27.1 | 19.4 | 13.9 | 19.4 | 23.6 | 18.2 | 0.3325 |
| XMC-GAN [50] | 9.3 | - | - | - | - | 30.5 | - |
| NÜWA | 12.9 | 13.8 | 15.7 | 19.3 | 24 | 27.2 | 0.3429 |

表 1: 文本到图像任务测试结果

NÜWA-LIP: 让视觉编辑更精细

NÜWA 模型已基本包含了视觉创作的核心流程, 可在一定程度上辅助创作者提升效率, 但在实际创作中, 创作者还有很多多样且高质量的需求。为此, 微软亚洲研究院的研究员们在 NÜWA 的基础之上更新迭代, 提出了 NÜWA-LIP 模型, 并且在视觉领域的典型任务——缺陷图像修复中取得了新突破。

尽管此前也有方法完成了类似的图像修复, 但是模型的创作却比较随意, 不能完全符合创作者的意愿, 而 NÜWA-LIP 几乎可以按照给定的自然语言指令修复、补全成人们肉眼可接受的图像。下面, 让我们直观感受一下 NÜWA-LIP 神奇的图像修复效果。

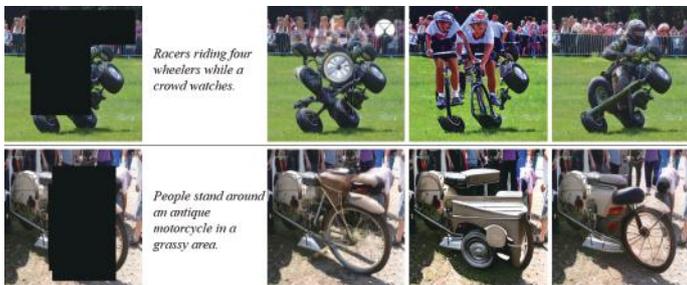


图 2: 在图像编辑任务上, NÜWA-LIP 展现出优秀的性能

图 2 给出了两个例子。第一个例子是希望模型可以按照 “Racers riding four wheelers while a crowd watches” (一群人在看摩托车手骑四轮车) 来补全黑色区域。已有工作 GLIDE 虽然可以补全, 但是可以看到边界处有明显的白线, 并且补全的区域比较模糊。NÜWA 模型使用自回归的方式从左到右依次扫描生成,

边界处相比于 GLIDE 更加自然。但是由于在补全黑色区域时看不到右侧的车轮, 因此标准的 NÜWA 模型存在补全边界衔接不对的问题。NÜWA-LIP 修复了 NÜWA 这一不足, 它会提前预览整个图像, 并创新地使用无损编码技术, 然后再自回归地生成, 因此可以做到黑色区域边界处衔接自然, 并且补全区域也很清晰。

在 FID 指标测试中, 通过将修复图与原始图对比, NÜWA-LIP 在自然语言指示图像修复的任务上取得了最好分数。(注: FID 分数越低表示修复图像的质量越高。)

| MODEL | MASKCOCO | MASKFLICKR | MASKVG |
|---------------------|-------------|-------------|------------|
| GLIDE | 13.5 | 51.9 | 9.0 |
| NÜWA | 21.4 | 59.5 | 18.5 |
| NÜWA-P | 20.6 | 54.2 | 17.7 |
| NÜWA-LIP | 12.0 | 42.5 | 8.5 |
| NÜWA-LIP (FINETUNE) | 10.5 | - | - |

表 2: NÜWA-LIP 在图像编辑任务中的 FID 指标达到 10.5

NÜWA-Infinity: 让视觉创作趋于 “无限流”

除了图像修复之外, 微软亚洲研究院在高分辨率、大图像的横向延展方面也进行了持续研究, 提出了 NÜWA Infinity 模型。顾名思义, NÜWA Infinity 可以根据给定的图像生成无限连续的高清 “大片”。“一开始 NÜWA 能够生成、编辑的图像和视频的分辨率相对较低, 一般是 256×256 分辨率的小图。我们希望通过模型可以生成更高清的大图, 形成更大的视觉冲击, 满足不同创作者的实际需求。简单来说, NÜWA Infinity 会根据图像的不同层次内容扫描每一帧窗口, 不断渲染形成高像素、连续的大图, “微软亚洲研究院研究员吴晨飞介绍说。

段补充说, “表面看 NÜWA Infinity 解决了之前 NÜWA 生成图片不高清, 以及视频帧数有限的问题。但其实 NÜWA Infinity 从底层形成了一套生成机制, 不仅可以对图片进行延展式的生成, 也可以应用于视频预测创作, 而这也是我们接下来要攻克的研究课题。”

自此, NÜWA-LIP 让机器接受语言指令自动修图成为了可能, 而 NÜWA-Infinity 则使得图像生成质量向高清、无限的真实世界迈进了一大步。按照这样的迭代创新步伐, 未来创作者拥有一套趋于 “无限流” 的视觉创作辅助工具指日可待。

NÜWA 多模态模型连锁反应： 或将带来更多“杀手级”应用

未来，随着人工智能技术的发展，增强现实、虚拟现实等沉浸式的人机交互界面将会得到更广泛的应用，数字世界和物理世界的结合也将越来越紧密。而不同类型的多模态内容则是拉近虚

拟空间与现实世界的强力胶，因此，虚拟内容的创建、编辑和交互将至关重要。

NÜWA 提供的视觉内容生成和编辑技术，为这些应用提供了无限的想象空间。当多模态技术成为未来人工智能应用发展的方向时，多模态模型将会为学习、广告、新闻、会议、娱乐、社交网络、数字人、脑机交互等领域带来更多的下一代“杀手级”应用。

NÜWA 在八大任务中的效果



1、文本到图像任务

例如，给定文本“A wooden house sitting in a field”(一个小木屋坐落在田野间)。NÜWA 创作了4种不同拍摄角度的小木屋，这些小木屋不仅朝向风格多样、而且真实性很好。



2、草图到图像任务

例如，给定一张公共汽车的草图（第一行第一列），NÜWA 创作了3种满足草图形状和位置的图像，包括窗户的反光也清晰可见。



3、图像补全任务

例如第1行，输入上方的塔尖（50%的原图），NÜWA 可以补全出塔下方的样子，圆柱甚至屋顶。对于第2行，当仅仅给5%的图像区域时，NÜWA 依然可以做到图像补全。



4、图像编辑

例如第1幅图，给定待编辑的图像、需要编辑的图像区域（红框）以及图像上方的文本“Beach and sky”（海滩和天空），第2幅图则给出了编辑后的结果。



5、图像到视频任务

NÜWA 不仅可以依据常见的文本“Play golf on grass”（在草地上玩高尔夫球）来生成视频，而且可以生成现实中不可能的视频，例如“Play golf on the swimming pool”（在泳池玩高尔夫球）。



6、视频草图到视频

输入视频草图，NÜWA 可以生成帧和帧连续的视频。



7、视频预测

输入静止图像，NÜWA 可以输出将其“动”起来的视频。



8、视频编辑

输入编辑文本、视频，NÜWA 可以输出编辑之后的视频。例如，原视频潜水员在水平游动，经过第二幅图“The diver is swimming to the surface”（潜水员在向水面游去）的控制，生成的视频潜水员在向上游。

扫描二维码
查看原文视频



Swin Transformer 迎来 30 亿参数的 v2.0 我们应该拥抱视觉大模型吗？

2021 年，获得 ICCV 最佳论文奖的 Swin Transformer，通过在广泛的视觉问题上证明 Transformer 架构的有效性，加速了计算机视觉领域基本模型架构的变革。2021 年末，微软亚洲研究院的研究员们又进一步提出了 Swin Transformer v2.0 版本，新版本训练了迄今为止最大的稠密视觉模型，并在多个主流视觉任务上大大刷新了记录，相关论文也已被 CVPR 2022 接收。研究员们希望借助 Swin Transformer v2.0 展现视觉大模型的“强悍”能力，呼吁整个领域加大对视觉大模型的投入，并为之提供相应的训练“配方”，从而为视觉领域的科研人员做进一步探索提供便利。那么，Swin Transformer v2.0 有哪些不同？今天就让我们来一探究竟吧！

人脑是大模型的一个典范。人的大脑拥有着千亿量级的神经元数目，和百万亿量级的连接数（参数）。而这样超大规模的模型为人脑成为目前唯一的通用智能“机器”提供了坚实的基础。在大容量下，人脑不仅在通常的智能任务中表现卓越，还具备极强的零样本和少样本迁移能力，从而可以快速适应新的环境和技能。

最近几年，自然语言处理（NLP）领域令人难以置信的成功就主要得益于对模型容量的大幅度扩展。短短几年时间，其模型容量扩大了几千倍，从 3.4 亿参数的 BERT 模型，进化到了拥有 5300 亿参数的 Megatron-Turing 模型，这些大型语言模型在语言理解和语言生成任务上都取得了长足的进步。同时，语言大模型还被证明具有极强的小样本甚至零样本学习能力。

与人脑和 NLP 的模型相比，计算机视觉领域的模型规模仍相对较小。视觉 Transformer 的出现为视觉模型的扩大提供了重要的基础，此前最大的稠密视觉模型是 18 亿参数的 ViT-G 模型和 24 亿参数的 CoAtNet 模型，它们都曾在 ImageNet-1K 图像分类任务上刷新了新的记录。但在更广泛的视觉任务中，大模型的效果仍然未知。

因此，探索如何进一步扩大视觉模型的规模，以及如何将其应用在更广泛的视觉任务上，是探索视觉大模型的重要问题。基于此目的，微软亚洲研究院的研究员们在 Swin Transformer 的基础上设计了 Swin Transformer v2.0，它具有 30 亿参数，是迄今为止最大的稠密视觉模型，可以有效地迁移到需要更高分辨率图像的各种视觉任务中。通过扩展模型容量和分辨率，Swin Transformer v2.0 已在四个具有代表性的基准上刷新了纪录，证明了视觉大模型在广泛视觉任务中的优势。

Swin Transformer 打破 视觉研究由 CNN “统治” 的局面

Swin Transformer 是一个通用的视觉 Transformer 骨干网络，它在物体检测和语义分割任务中大幅刷新了此前的纪录，并被广泛应用于众多视觉任务中，如图像生成、视频动作识别、视觉自

监督学习、图像复原、医疗图像分割等。Swin Transformer 打破了计算机视觉领域被 CNN（卷积神经网络）长期“统治”的局面，加速了计算机视觉领域基本模型架构的变革，这一工作也因此获得了 2021 年 ICCV 最佳论文奖——马尔奖。

Swin Transformer 的核心思想在于将具有很强建模能力的 Transformer 结构与重要的视觉信号先验结合起来。这些先验包括层次性、局部性以及平移不变性等等。Swin Transformer 的一个重要设计是 shifted windows（移位的不重叠窗口），它可以大幅降低计算复杂度，让计算复杂度随着输入图像的大小呈线性增长；同时不同于传统的滑动窗，不重叠窗口的设计对硬件实现更加友好，从而具有更快的实际运行速度。



图 1: Sliding windows (滑动窗口) vs. Shifted windows (移位窗口)

事实上，Swin Transformer 不是一个一蹴而就的工作，而是研究团队四年多在相关方向上不断坚持的结晶。“人脑成功的一个关键密码就是其拥有大量的新皮质，新皮质中的神经结构是统一和通用的，这就使得人类不用通过生物进化就可以实现和适应各种新的智能或者环境。在这方面，我们一直看好 Transformer 或者其中的注意力模块，三年前我们首次尝试将 Transformer 应用于视觉骨干网络的设计，并提出了局部关系网络 LR-Net，但当时的实用性还不足。Swin Transformer 通过引入移位窗口，终于达成了实用的视觉 Transformer 骨干网络，”微软亚洲研究院视觉计算组高级研究员胡瀚说。

Swin Transformer 的目标是希望证明视觉 Transformer 能在广泛的视觉问题中超越此前占据主导地位的 CNN。如今该目标已达成，那么下一步做什么？胡瀚认为，“过去几年 NLP 领域最重要的发现之一就是扩大模型容量可以持续帮助各种 NLP 任务，并且模型越大，零样本和少样本学习的能力越强。所以我们希望探索计算机视觉中能否拥有同样的性质。”于是，Swin Transformer v2.0 诞生了。

在探索过程中，研究员们发现如下三个问题对于视觉大模型格外重要：

1. 如何解决大模型训练稳定性的问题
2. 如何将大模型用于拥有更高分辨率的下游视觉任务的问题
3. 如何减少大模型对标注数据的要求

针对上述三个问题，Swin Transformer v2.0 给出了自己的回答。下面就让我们来详细了解一下。

30 亿参数的 Swin Transformer v2.0 稳定性与准确性双提升

在进一步扩大模型容量的过程中，微软亚洲研究院的研究员们发现训练过程存在严重的不稳定性问题。如图 2 所示，随着原始 Swin Transformer 模型从小变大，网络深层的激活值会急剧增加，拥有 2 亿参数的 Swin-L 模型，其幅值最高和最低层之间的差异可以达到 10^4 。当进一步将模型容量扩大到 6.58 亿参数，它会在训练过程中崩溃。

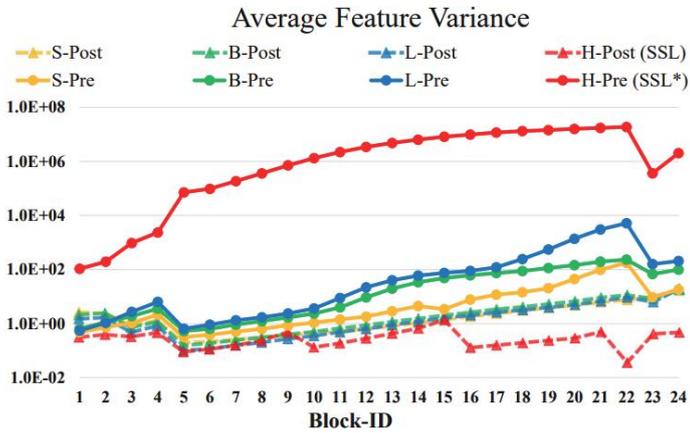


图 2: 各种大小模型各层的激活值。H 大小的模型是在自监督学习阶段训练的，而其他规模的模型则是由分类任务训练的。
* 表示使用了模型崩溃前第 40 轮的模型权重。

仔细观察原始 Swin Transformer 的架构，研究员们发现这是由于残差分支的输出直接加回主分支而导致的。原始的 Swin Transformer (以及绝大多数视觉 Transformer) 在每个残差分支

的开始都使用了预归一化 (Pre-normalization)，它可以归一化输入的幅值，但对输出没有限制。在预归一化下，每个残差分支的输出激活值会直接合并回主分支，并被逐层累加，因而主分支的幅值会随着深度的增加而越来越大。这种不同层的幅值差异很大程度上导致了训练的不稳定性。

为了缓解这一问题，研究员们提出了一种新的归一化方式，称为残差后归一化 (Residual-post-normalization)。如图 3 所示，该方法将归一化层从每个残差分支的开始移到末尾，这样每个残差分支的输出在合并回主分支之前都会被归一化，当层数加深时，主分支的幅度将不会被累加。实验发现，这种新的归一化方式使得网络各层的激活值变得更加温和。

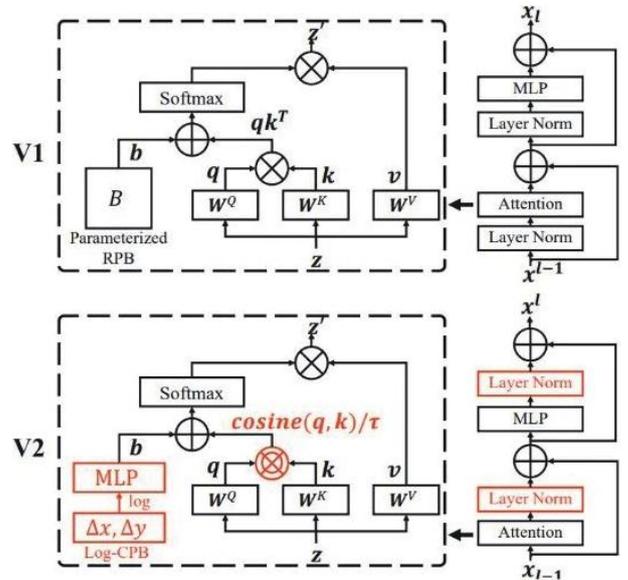


图 3: Swin Transformer v1.0 vs. v2.0

除此之外，研究员们还发现，随着模型变大，在原始的自注意力计算中，某些层的注意力权重往往会被几个特定的点所支配，特别是在使用后注意力的情况下。为了缓解这一问题，研究员们还提出了缩放的余弦注意力机制 (Scaled cosine attention)，它可以取代之前的点乘注意力机制。在缩放的余弦注意力机制中，自注意力的计算与输入的幅值无关，从而可以产生更温和的注意力权重。

实验证明，以上两种技术不仅使大模型的训练过程更加稳定，而且还提高了准确性。

从低分辨率向高分辨率迁移 看 Swin Transformer v2.0 如何克服不良反应

视觉大模型的另一难题在于许多下游视觉任务需要高分辨率的输入图像或注意力窗口。由于预训练往往在低分辨率下进行，所以在高分辨率的下游任务上进行微调的窗口大小会产生显著变化。目前常见的做法是对位置偏置 (position bias) 进行双立方插值，这是一种随意的简单处理方式，其效果并不是最佳的。

为了解决这一问题，研究员们提出了对数空间的连续位置偏置 (Log-spaced continuous position bias, Log-spaced CPB)。通过对对数空间的位置坐标应用一个小的元网络，Log-spaced CPB 可以产生任意坐标范围的位置偏置。由于元网络可以接受任意坐标，因此通过共享其权重，一个预训练好的模型可以在不同的窗口大小之间自由迁移。另一方面，通过将坐标转化到对数空间，在不同的窗口分辨率之间迁移所需的外推率要比使用原始线性空间坐标的外推率小得多，如图 4 所示。

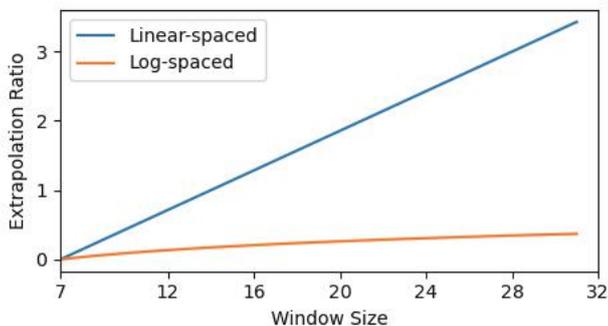


图 4: Log-spaced CPB vs. Linear spaced CPB

借助 Log-spaced CPB, Swin Transformer v2.0 实现了模型在不同分辨率之间的平滑迁移。当把预训练分辨率从 224 像素缩小到 192 像素时，其结果也不会受到影响，并且计算速度还提升了 50%，显著降低了训练类似体量的模型的成本。

模型容量和分辨率的扩大也导致了现有视觉模型的 GPU 显存消耗过高。为了解决显存问题，研究员们结合了几个重要的技术，包括零冗余优化器 (zero-redundancy optimizer)、后向重计算 (activation check-pointing) 以及新提出的顺序自我注意计算力机制 (sequential self-attention computation)。有了这些技术，大模型和大分辨率下的 GPU 显存消耗明显减少，而其对训练速度的影响却很小。

自监督学习 SimMIM

解决视觉大模型的数据饥饿问题

训练越大的模型往往需要越多的数据，而相比 NLP，计算机视觉领域缺乏蕴含人类监督信息的数据来支撑大模型的训练。这就要求视觉领域在训练大模型时，要减少对标注数据的依赖，需

要在更少量数据的情况下探索大模型。对此，研究员们通过引入自监督学习的掩码模型 SimMIM 来缓解这一问题。如图 5 所示，SimMIM 通过掩码图像建模 (masked image modeling) 来学习更好的图像表征。它采用随机掩码策略，用适度大的掩码块对输入图像做掩码；同时，通过直接回归来预测原始像素的 RGB 值；由于该模型的预测头很轻，所以只需要一层线性层即可。

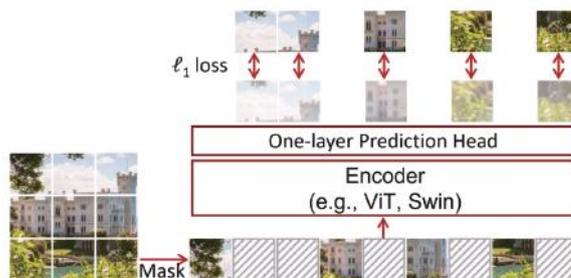


图 5: SimMIM 自监督学习方法示意图

SimMIM 非常简单且高效，借助 SimMIM, Swin Transformer v2.0 降低了对标注数据的需求，最终只用了 7000 万张带有噪声标签的图像就训练了 30 亿参数的模型。

有图有真相: v2.0 性能“强悍” 在四大基准上均创新纪录

通过扩展模型容量和分辨率，Swin Transformer v2.0 在四个具有代表性的基准上均刷新了纪录，证明了视觉大模型在广泛视觉任务中的优势：在 ImageNet-V2 图像分类任务上 top-1 准确率为 84.0%；在 COCO 物体检测任务上为 63.1/54.4 box/mask mAP；在 ADE20K 语义分割上为 59.9 mIoU；在 Kinetics-400 视频动作分类的 top-1 准确率为 86.8%。

| Method | param | pre-train images | pre-train length (#im) | pre-train im size | pre-train time | fine-tune im size | ImageNet-1K-V1 top-1 acc | ImageNet-1K-V2 top-1 acc |
|----------------|--------|------------------|------------------------|-------------------|--------------------|-------------------|--------------------------|--------------------------|
| SwinV1-B | 88M | IN-22K-14M | 1.3B | 224 ² | <30 ¹ | 384 ² | 86.4 | 76.58 |
| SwinV1-L | 197M | IN-22K-14M | 1.3B | 224 ² | <10 ¹ | 384 ² | 87.3 | 77.46 |
| ViT-G [65] | 1.8B | JFT-3B | 164B | 224 ² | >30k | 518 ² | 90.45 | 83.33 |
| V-MoE [44] | 14.7B* | JFT-3B | - | 224 ² | 16.8k | 518 ² | 90.35 | - |
| CoAtNet-7 [11] | 2.44B | JFT-3B | - | 224 ² | 20.1k | 512 ² | 90.88 | - |
| SwinV2-B | 88M | IN-22K-14M | 1.3B | 192 ² | <30 ¹ | 384 ² | 87.1 | 78.08 |
| SwinV2-L | 197M | IN-22K-14M | 1.3B | 192 ² | <20 ¹ | 384 ² | 87.7 | 78.31 |
| SwinV2-G | 3.0B | IN-22K-ext-70M | 3.5B | 192 ² | <0.5k ¹ | 640 ² | 90.17 | 84.00 |

表 1: ImageNet 图像分类的性能

| Method | train | | test | | mini-val (AP) | | test-dev (AP) | |
|------------------|-----------|-----------|------|------|---------------|------|---------------|------|
| | I(W) size | I(W) size | box | mask | box | mask | box | mask |
| CopyPaste [17] | 1280(-) | 1280(-) | 57.0 | 48.9 | 57.3 | 49.1 | - | - |
| SwinV1-L [35] | 800(7) | ms(7) | 58.0 | 50.4 | 58.7 | 51.1 | - | - |
| YOLOR [53] | 1280(-) | 1280(-) | - | - | 57.3 | - | - | - |
| CBNet [32] | 1400(7) | ms(7) | 59.6 | 51.8 | 60.1 | 52.3 | - | - |
| DyHead [10] | 1200(-) | ms(-) | 60.3 | - | 60.6 | - | - | - |
| SoftTeacher [60] | 1280(12) | ms(12) | 60.7 | 52.5 | 61.3 | 53.0 | - | - |
| SwinV2-L (HTC++) | 1536(32) | 1100(32) | 58.8 | 51.1 | - | - | - | - |
| | | ms (48) | 58.9 | 51.2 | - | - | 60.8 | 52.7 |
| SwinV2-G (HTC++) | 1536(32) | 1100(32) | 61.7 | 53.3 | - | - | - | - |
| | | ms (48) | 61.9 | 53.4 | 62.5 | 53.7 | 63.1 | 54.4 |

表 2: COCO 物体检测的性能

| Method | train I(W) size | test I(W) size | mIoU |
|--------------------|-----------------|----------------|--------------|
| SwinV1-L [35] | 640(7) | 640(7) | 53.5* |
| Focal-L [61] | 640(40) | 640(40) | 55.4* |
| CSwin-L [14] | 640(40) | 640(40) | 55.7* |
| MaskFormer [8] | 640(7) | 640(7) | 55.6* |
| FaPN [22] | 640(7) | 640(7) | 56.7* |
| BEiT [3] | 640(40) | 640(40) | 58.4* |
| SwinV2-L (UperNet) | 640(40) | 640(40) | 55.9* |
| SwinV2-G (UperNet) | 640(40) | 640(40) | 59.1 |
| | | 896 (56) | 59.3 |
| | | 896 (56) | 59.9* |

表 3: ADE20K 语义分割的性能

| Method | train I(W) size | test I(W) size | views | top-1 |
|-------------------|-------------------------|-------------------------|-------|-------------|
| ViViT [1] | -(-) | -(-) | 4×3 | 84.8 |
| SwinV1-L [36] | 480×480×16 (12×12×8) | 480×480×16 (12×12×8) | 10×5 | 84.9 |
| TokenLearner [45] | 256×256×64 (8×8×64) | 256×256×64 (8×8×64) | 4×3 | 85.4 |
| Video-SwinV2-G | 320×320×8 (20×20×8) | 320×320×8 (20×20×8) | 1×1 | 83.2 |
| | | 384×384×8 (24×24×8) | 1×1 | 83.4 |
| | | 384×384×8 (24×24×8) | 4×5 | 86.8 |

表 4: Kinetics-400 视频动作分类的性能

视觉大模型发展的三道坎： 数据匮乏、学习方法和通用性

Swin Transformer v2.0 在众多视觉任务中的“强悍”性能证明了视觉大模型的潜力，同时，为视觉大模型提供了一个可行的训练“配方”。微软亚洲研究院的研究员们期望 Swin Transformer v2.0 可以为视觉领域的科研人员们做进一步前沿探索提供便利，并借此激励整个领域加大对视觉大模型的投入。

“我还是比较看好视觉大模型的”，胡瀚表示，“大模型已经证明通过增加模型容量可以持续提升性能的能力，其小样本甚至零样本能力也十分惊艳，而小样本能力对于实现通用智能非常关键”。

当然，胡瀚也认为视觉大模型仍旧面临一些挑战，“一是数据问题，视觉领域可用于训练的有效数据相比 NLP 领域还是有不小的差距。自监督是一个潜在的解决方法，但目前的自监督方法包括 SimMIM/BEiT/MAE/PeCo 等等都还不能利用好更大的数据，也就是证明更多的图像数据能帮助训练更好的模型。”

“二是，学习方法上还需要突破。现在在训练大模型方面，学界和业界对于自监督、图像分类和多模态方法均有所尝试，也取得了一些效果，但这些离我们真正解决问题还差的很远”。

“三是，如何建立视觉通用模型的问题还未知。如今的视觉大模型大多还是依赖于预训练和微调的方式，不同的视觉应用仍

需要依赖于不同的模型，如何能真正建立视觉的通用模型，用一个模型解决大部分问题，这还需要科研人员做大量的探索。”

视觉大模型的未来很光明，但也充满了挑战。微软亚洲研究院的研究员们期待更多同仁一起努力，推进视觉大模型的更多进展。

相关论文：

Swin Transformer V2: Scaling Up Capacity and Resolution
<https://arxiv.org/abs/2111.09883>

SimMIM: A Simple Framework for Masked Image Modeling
<https://arxiv.org/abs/2111.09886>

Swin Transformer: Hierarchical Vision Transformer using Shifted Windows
<https://arxiv.org/abs/2103.14030>

Local Relation Networks for Image Recognition
<https://arxiv.org/abs/1904.11491>

相关 GitHub 链接：

<https://github.com/microsoft/Swin-Transformer>

<https://github.com/microsoft/SimMIM>

相关阅读

为何 Transformer 在计算机视觉中如此受欢迎？

近年来，Transformer 在计算机视觉领域所带来的革命性提升，引起了学术界的广泛关注，有越来越多的研究人员投入其中。本篇文章介绍了 Swin Transformer 的优越性，并讲述了过去几年间学术界不断挖掘出的 Transformer 建模的各种优点，包括通用的建模能力、和卷积形成互补、更强的建模能力、对大模型和大数据的可扩展性以及更好地连接视觉与语言任务。



扫描二维码查看文章

语音识别的快速纠错模型 FastCorrect 系列来了!

语音识别支持着许多生活中的常见服务，而语音识别模型往往并不完美，需要纠错模型来纠正语音识别中的错误。目前，大部分纠错模型采用了基于注意力机制的自回归模型结构，虽然能够提升语音识别的准确率，但是延迟较高。一个直接的做法是简单利用当前的非自回归模型来提升速度，但无法同时降低错误率。为此，微软亚洲研究院机器学习组与微软 Azure 语音团队合作推出了 FastCorrect 系列工作，提出了低延迟的纠错模型，相关研究论文已被 *NeurIPS 2021* 和 *EMNLP 2021* 收录。

纠错是语音识别中的一个重要后处理方法。许多纠错模型采用的是延迟较高的自回归解码模型，但是语音识别服务对模型的延迟有着严格的要求，在一些实时语音识别场景中，纠错模型无法上线应用。为了加速语音识别中的纠错模型，微软亚洲研究院的研究员们提出了一种基于编辑对齐 (Edit Alignment) 的非自回归纠错模型——FastCorrect，FastCorrect 在几乎不损失纠错能力的情况下，将自回归模型加速了 6-9 倍。考虑到语音识别模型往往可以给出多个备选识别结果，研究员们还进一步提出了 FastCorrect 2 来利用这些识别结果相互印证，从而得到了更好的性能。FastCorrect 1 和 2 的相关研究论文已被 *NeurIPS 2021* 和 *EMNLP 2021* 收录。当前，研究员们还在研发 FastCorrect 3，在保证低延迟的情况下，进一步降低语音识别的错误率。

FastCorrect：快速纠错模型

语音识别的纠错实际上是一个文本到文本的任务，模型训练的输入为语音识别结果文本，输出为真实文本。在自然语言处理领域（如机器翻译和文本编辑），已经有一些非自回归的快速模型被提出。但初步实验结果（如图 1）显示，简单地将这些模型应用到语音识别的纠错任务中，并不能取得令人满意的结果。经过对语音识别的分析，研究员们发现语音识别中的错误比较稀疏，通常错误的单词数不到总单词数的 10%，而模型必须精准地找到并修改这些错误，同时还要避免修改正确的单词，这是语音识别的纠错任务中最大的挑战。而机器翻译中非自回归模型的主要问题是修改了太多原本是正确的单词，模型修改了原有错误的同时又引入了较多新的错误，因此无法提升语音识别的精度。

| Model | AISHELL-1 | | | |
|-------------|------------|------------|-------------|-------|
| | P_{edit} | R_{edit} | P_{right} | WERR |
| LevT | 71.7 | 14.4 | 38.4 | 2.07 |
| FELIX | 76.7 | 18.3 | 40.6 | 4.14 |
| FastCorrect | 83.7 | 34.5 | 50.1 | 13.87 |

图 1: FastCorrect 和基线方法的错误检测率和错误改正率对比

考虑到语音识别的纠错输入输出的对应关系是单调的，所以如果可以在词的级别对齐输入和输出，就可以得到细粒度的错误信息：哪些词是错误的，这些错误的单词应该怎样修改。

基于两个文本序列的编辑距离，研究员们设计了编辑对齐 (Edit Alignment) 算法（如图 2）。

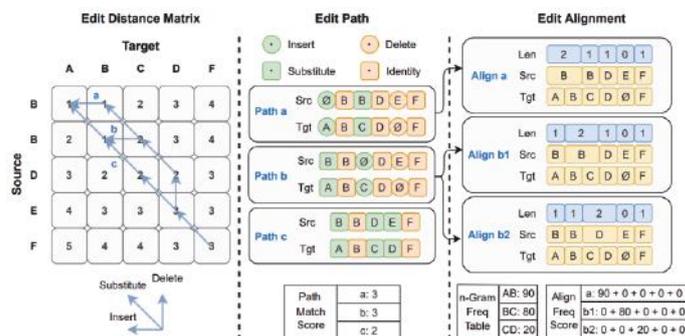


图 2: 编辑对齐 (Edit Alignment) 算法

给定输入（语音识别结果）和输出（真实文本），第一步是计算两个文本的编辑距离，然后可以得到数条编辑路径 (Edit Path)，路径中的元素为增加 / 删除 / 替换 / 不变四种操作之一。为了避免修改正确单词，包含“不变”操作最多的编辑路径会被选择。最终，基于编辑路径可以得到：对于每个输入的单字，哪些输出的单字与之对应。如果对应的输入输出单字不同，那么就表明输入单字是错误单字。

基于细粒度的输入输出对应关系，研究员们针对性地设计了快速纠错模型 FastCorrect。模型包含了三个主要部分——编码器、长度预测器和解码器（如图 3）：

1. 编码器可以学习输入文本的特征，这些特征会被长度预测器和解码器利用。
2. 长度预测器基于编码器的输出，预测每个输入单字有多少个输出单字与之对应，即 Duration。如果输入单字 Duration 为 0，那么没有输出单字与之对应，它也将被删除，如果 Duration 大于 1，那么有多个输出单字与之对应，意味着解码器需要插入数个单字。
3. 解码器除了利用编码器的输出之外，还会基于长度预测器的结果，调整输入文本的长度，使之和输出文本长度一致。将长度调整一致后，解码器可以并行地同时解码出所有单字。

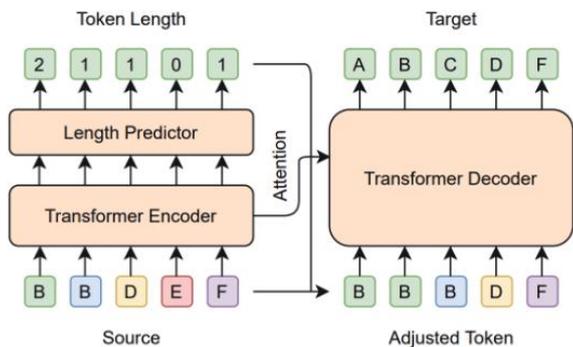


图3: 非自回归语音纠错模型结构

在开源学术数据集 AISHELL-1 和微软内部的产品数据集上, FastCorrect 都取得了跟自回归模型几乎相同的纠错能力。模型的衡量准则为纠错后的词错误率 (Word Error Rate, WER) 和相对错误率下降 (WER Reduction)。相比于自回归的模型, FastCorrect 的解码速度可以提升 7-8 倍。相比于其它非自回归的基线模型, FastCorrect 取得了更好的纠错精度。从实验结果可以看出, FastCorrect 很好地解决了语音识别模型中纠错模型的高延迟问题。

FastCorrect 2 : 基于多输入的快速纠错模型

FastCorrect 已经取得了不错的性能和较大的加速比, 那么还有没有进一步的提升空间呢? 由于语音识别模型往往会在语音模糊不清或者存在同(近)音字的情况下出错, 所以如果纠错模型能够得到这方面的信息, 那么就可以在很大程度上帮助纠错模型进行错误检测。

通过分析语音识别模型的输出, 研究员们发现语音识别模型在进行 BeamSearch 解码的时候, 会得到多个相似的结果, 这些结果的大部分文本相同, 只在有限的几个位置不同, 而这些位置正是语音识别模型不确定的位置, 也往往是模型比较容易出错的位置。通过多个识别结果的相互印证不仅能够得到语音识别错误位置的信息, 还能得到正确单词的读音信息 (例如, 从不同的词 cat、hat、mat 可以推断出纠错后的词大概率会以 /æt/ 这个音结尾)。为了能够让纠错模型更好地利用多输入 (多个语音识别结果) 信息, 研究员们对 FastCorrect 的对齐算法和模型结构进行了一系列调整。

首先是对齐算法。考虑到语音模型的多个输出可能长短不一, 如果简单地通过补零来让这些输出拥有同样的长度, 那么会导致每一个对齐的位置, 单词不相同 (如图 4), 读音也不相同。这样的话, 模型将很难从“错位”的输入中得到错误位置和正确单词读音的信息。为了解决上述问题, FastCorrect 2 在计算编辑对齐的时候考虑了输入输出的读音相似度, 在选择“不变”操作最多的编辑路径的基础上, 选择读音相似度最高的编辑路径, 并根据读音相似度来插入“空”单词。这样得到的对齐结果在每个位置上的单词相似度和读音相似度都很高, 可以很大程度上帮助模型进行

错误检测 (如果一个位置上包含不同的单词) 和估计正确单词的读音 (如果一个位置上的不同单词读音相同)。

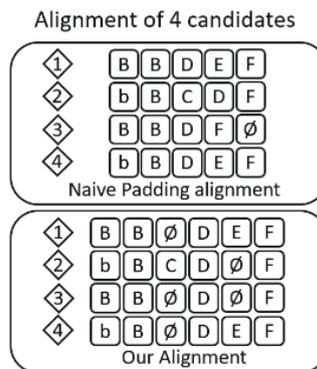


图4: 简单补零和读音相似度对齐的对比, 可以看出 FastCorrect 2 在每个位置上的单词一致性更高。

对于模型结构, 研究员们也进行了改动, 并引入了一个新模块: 选择器 (如图 5)。

1. 在编码器之前, 新引入了一个 PreNet, 用来融合每个位置上不同单词的信息。
2. 长度预测器需要对每个位置上的每一个单词, 预测 Duration。
3. 选择器则用来选择一个输入, 这个输入会被调整 (基于 Duration) 并被送进解码器, 选择器的训练目标是预测解码器的损失, 因此选择器选择的是解码器损失最小的, 也就是最容易被解码器修改的输入。

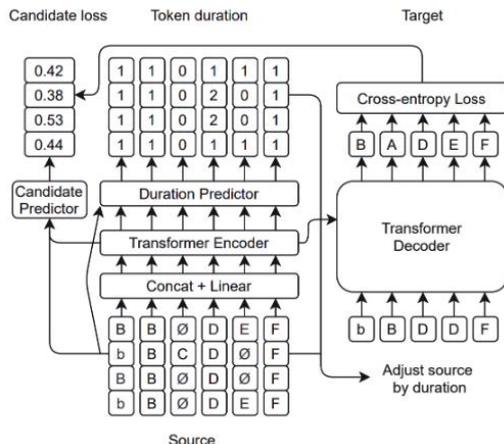


图5: FastCorrect 2 模型结构

实验表明, 在开源学术数据集 AISHELL-1 和微软内部的产品数据集上, FastCorrect 2 相对于 FastCorrect 都取得了较大的提升, 而且模型的提升会随着输入的增多而不断增大。

扫描二维码
文末获取相关论文链接



科研第一线



你真的了解计算生物学和 AI for Science 吗？

近年来，计算生物学无疑是人工智能领域的一大热门话题。但，计算生物学究竟是什么？目前进展如何？未来又蕴藏了怎样的机遇？

在量子位对撞派推出的“计算生物学”专题直播中，微软亚洲研究院副院长刘铁岩、首席研究员邵斌和主管研究员王童介绍了微软亚洲研究院计算生物学领域的最新研究，并对未来 AI for Science 的发展和融合进行了分享。

扫描二维码查看回放视频
以及精华版文字内容！



如何亿点点降低语音识别跨领域、跨语种迁移难度？

随着深度学习的不断发展，语音识别技术得到了极大的提升，同时为人们的日常生活提供了许多便利。然而，一个语音模型的训练并非易事，因为语音数据天然存在着获取难、数据标注耗时昂贵的问题，而且还会面临模型漂移、标注数据不足等难题。因此，迁移学习技术对于语音数据非常重要。为了解决语音识别的跨领域和跨语言问题，微软亚洲研究院机器学习组和微软（亚洲）互联网工程院提出了跨领域和跨语言语音识别的 CMatch 和 Adapter 方法。这两项技术是如何提升模型迁移学习性能的？他们又利用了哪些创新技术？

扫描二维码查看文章



科学匠人 | 白静：拥抱变化，不断发现计算机科学中的新天地

在计算机领域，研究与产品的关系往往十分微妙。一方面，二者相辅相成、互相推动；另一方面，它们追求的目标又不尽相同——产品需要精确的 KPI 及短期落地目标，研究则更注重宏观的、长期的影响力。今天科学匠人的主人公，微软亚洲研究院首席研究员白静博士在产品与研究领域都深耕多年，她希望自己的研究能成为这两个领域之间融会贯通的纽带。在她看来，让研究成果与产品产生共振，进而推动彼此进化，是研究的价值所在。

从蒙特利尔大学博士期间从事信息检索（IR）和自然语言处理（NLP）的研究，到加入微软硅谷研发中心推动多个重要产品落地，再到微软亚洲研究院带领系统和算法等新领域的研究，微软亚洲研究院首席研究员白静的职业经历看似是“研究 - 产品 - 研究”的轮回，但实际上却是她在计算机科学领域的不断“螺旋上升”。每一次转型，她都自己的工作惠及更多产品和用户——无论是将创新研究转化为用户最满意的产品，还是以产品需求为驱动力投身于研究，都是如此。



微软亚洲研究院首席研究员白静

从全局和长远视角做研究 创造全新机会不断激发产品潜力

2019 年，白静从微软 Azure AI 团队加入了微软亚洲研究院。尽管此前她在 Azure AI 和微软必应（Bing）团队潜心研究机器学习和自然语言处理，但这次她希望挑战业界还未深入涉足并且能在微软产品中获得充分应用的领域——图深度学习（Graph Learning）。对于一直喜欢挑战未知的白静来说，这是一个足够“新”，且成果值得期待的领域。

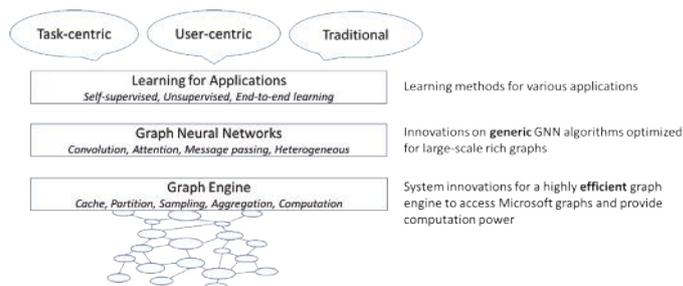
“当时深度学习技术已经很流行了，但大多被应用于自然语言处理、计算机视觉等领域，在图学习领域还没有被广泛使用。微软的许多产品都与图相关，该如何将机器学习的热门技术应用到图领域，进而提高微软产品的性能和效率？”这是白静初入微软亚

洲研究院时就在思考的问题。

在白静看来，图（Graph）作为一种通用数据结构，可以清晰地表现出多个元素之间的有机关联。在微软的很多产品中，图有着丰富的应用场景，例如 Office 的企业图谱中，用户与繁多的会议、文档、邮件之间的关联；领英（LinkedIn）的社交网络中，复杂的社交关系和给求职者的职位推荐；Ads 广告业务中，广告主与受众需求、点击预测、关键词之间的关系，以及大规模知识图谱等，这些隐藏在众多产品中的有机关联，在广义上都是图结构。

2019 年底，在微软亚洲研究院院长周礼栋的协助下，白静团队主导并在微软公司内部成立了图神经网络工作组，为公司提供了一个从研究到产品的全局视野平台，以及和图学习相关的工具、算法，便于研究、产品、工程团队的跨部门协作、沟通和知识共享，从而提升图学习效率，促进系统和算法的研究创新。这一工作组的构建结束了各部门图深度学习研究和应用各自为战的局面，将“百花齐放”的算法统一在一个高效的平台上，不仅有利于激荡新思路，找到正确的研究方向，推动大规模的图学习应用和技术进步，还能将图学习算法上的创新成果快速迭代应用在相关的产品中。

两年来，白静与公司多个部门合作创新图深度学习算法，探索适合的应用场景。她带领团队与微软广告、Office、Azure 及 LinkedIn 等多个团队展开积极的合作，并将图学习算法应用到不同的产品中，提高了多个产品的运营效率，给公司业务带来了直接的商业效益。她说，“我们希望通过这些合作来带动核心研究，进一步提升平台和算法的性能和效率，继而推动全公司更大规模的产品和应用，而不只是某个单一产品。”



图深度学习及其应用的全局展示

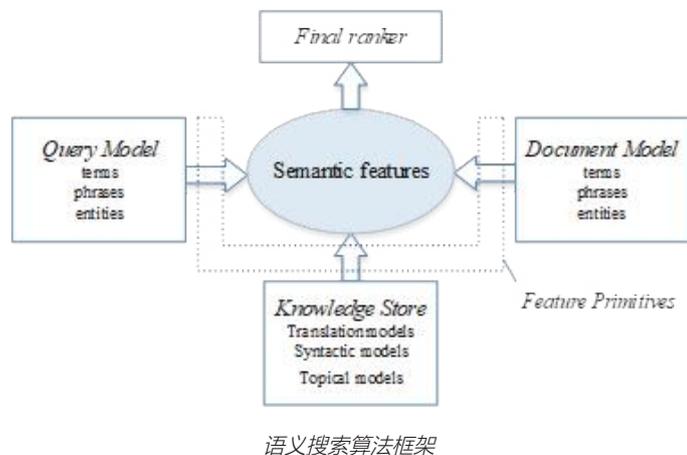
虽然研究创新的最终目标之一是服务于产品，但是在“研究 - 产品 - 研究”的迭代中，白静认为，相比于产品快速落地的诉求，研究需要有长远和全局视角。她说，“每个产品的需求不尽相同，而我们做研究是希望尽可能把全公司的产品都推动起来，建立一个系统性的长期规划，这正是做研究和做产品的区别。”而这也是白静选择加入微软亚洲研究院的主要原因，“我希望可以从更广阔和长远的视角来思考研究问题，从个性化的产品需求中抽象出共性问题，从更底层赋能众多产品的研发和最终成果。”

深入一线的科研人员 如何理解研究与产品之间的联系

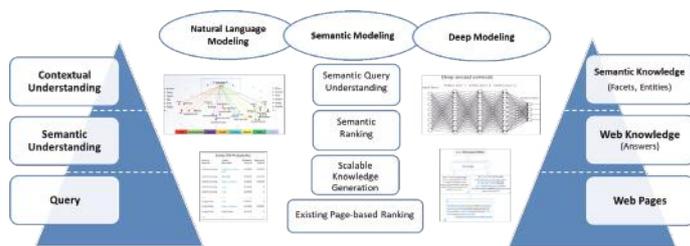
白静对于研究与产品间相辅相成关系的深刻认识，来自于她多年在这两方角色转换的切身体会和经验积累。如何跳脱固有的思维框架，站在更加全局的角度思考问题并非易事。

在蒙特利尔大学获得计算机科学博士学位后，白静被硅谷多元开放的文化和创新热情所吸引，同时她也希望可以将自己的研究成果转化到产品中，服务千万用户。怀揣这样的理想，白静在2010年加入了微软硅谷研发中心，任职高级研究科学家。彼时微软正布局拓展搜索引擎业务，而白静的研究方向正好是信息检索，这让她的研究有了用武之地。

也是在这一时期，白静与微软中国团队“结缘”。2012年她受邀参与了新一代搜索引擎系统的研发工作，新系统采用了全新的网页索引结构和系统设计。作为微软必应搜索相关性的主要贡献者之一，白静敏锐地发现学术界前沿的诸多语义检索方法可以赋能新系统，从而给搜索相关性带来突破性创新，而这正是传统搜索引擎所不具备的。由此，她为微软必应开创了全新的语法语义搜索算法框架（Semantic Ranking Framework），并研发出了基于PDI（Per Document Index）正排前瞻索引的一系列大规模语义模型，实现了全文语义检索，进而显著提高了必应搜索结果的相关性。此后多年，这项成果一直都是提升搜索相关性的最有效技术之一，许多相关模型应运而生。该项目也受到当时多位微软公司高管的高度评价，并通过微软的产品服务于上亿用户，同时还获得了多项国际专利。



在实现了全文语义检索的基础上，白静进一步思考能否通过更精确的用户意图分析，让搜索引擎直接给用户想要的答案，而不仅是列出相关网页链接。为了实现这一目标，她提出了深度搜索的构想，利用互联网动态信息和人工智能算法直接生成用户满意的答案。这个想法得到了当时微软全球执行副总裁陆奇的支持，一个新项目由此诞生。由于原算法的搜索结果包含的网页数量巨大，再去检索网页中相关度更高的信息，搜索空间会呈指数级增长。为了解决如何从海量的网页中匹配出用户需求这一问题，白静和团队首创了基于全网的大规模段落检索系统（Web-scale Passage Retrieval System），大胆地引入和改进了学术界前沿的研究成果，实现了从网络内容中直接生成搜索答案的目标，并推动了该系统在微软必应中的成功运用。该系统的上线开创了微软搜索引擎智能化的新方向，这项成果也成为了必应 AI 计划的一个重要里程碑。



基于全网的大规模段落检索系统

随着计算机技术的不断发展，云计算和 AI 时代的到来吸引白静加入了微软 Azure AI 团队，迎接一个新的挑战——如何利用 AI 技术自动设计出更高效的深度学习模型。她带领团队与微软雷德蒙研究院合作，研发了 Azure 第一个适用于机器学习模型的自动化超参调优产品“HyperDrive”。该系统可以将用户从手工调参的繁琐工作中解放出来。正式上线仅半年，“HyperDrive”就成为了当时 Azure ML 中流量最大的产品，服务于第一方和第三方用户，在智能性和通用性上都处于业界领先地位。

在这几个重要产品的研发过程中，白静的工作都聚焦在产品与研究的有机结合，这让她更加深刻地理解到研究与产品之间的关联和差异。她认为产品和研究各有乐趣，“好比盖房子，产品开发可以聚焦在房子的具体建造，不同的产品部门分工合作，各自将某个房间或局部打造到极致，确保其准确和稳定，却往往没有很多机会跳出来思考；而研究工作则可以从全局视角思考整个房子的设计和构建，实现不同布局之间的融会贯通，甚至可以开创超越传统的设计，这样才能提供超出用户想象的最佳体验。”

“其实，从研究到产品不只是简单的产品落地，更多的时候需要从产品的角度重新审视研究成果，看它们能否给产品带来预期的效果。这需要我们结合研究和产品开发的双重洞察力，从多个不同视角和维度去思考问题，并不断积极主动地寻求解决问题的方案，”白静说。正是这种研究和产品有机结合所产生的独特价值激发了白静热情，使她在工作中保持着极大的好奇心。她相信研究和产品应该相互推动彼此进化，只有将它们很好地融合在一起才能使效益最大化。

计算机科学的常态是变化 这是它最具吸引力的所在

如今，在微软亚洲研究院包容、开放、多元的研究氛围中，白静有了更大的施展空间，虽然工作重点从产品开发变成了科研创新，但对如何让研究产生更大价值的追求始终如一。她让自己变成研究与产品之间的纽带，既可以沉浸于抽象、共性课题的研究，也可以与不同产品组合作进行成果转化。

过往项目的成功经验让白静深切体会到了跨团队、跨领域合作的优势，她清楚地看到，“团队合作的本质就是优势互补，创造双赢局面，从而达到事半功倍，1+1>2 的效果。只有集思广益，突破传统的思维和认识，才能不断地实现跨界创新。”现在以研究员的视角与产品组合作，白静会更多地鼓励团队做核心技术的研究，用开创性思维拓展技术的边界，Be bold and be creative（大胆创新），从不同的角度提供新鲜的灵感，给产品带来本质的提升，从而更好地激发产品组合作的积极性。



从搜索引擎到 AI，再到图深度学习，每一次的岗位变换都是在寻求新的挑战并不断超越自己，但白静并没有感到走出“舒适圈”的迷茫与困惑，反而更多地增加了探索新领域的新鲜感与好奇心。对此她表示，“不要抵触对新领域的陌生感，因为过往沉淀的方法与经验，可以让你快速适应变化。就像计算机知识会过时，但其核心和本质却万变不离其宗，比如过去我们研究了多年的 NLP，尽管现在有了 BERT、GPT 等新技术，但是我们之前积累起来的研究方法依然有效。”

在白静看来，计算机科学的常态就是变化，而这也是它最具吸引力的地方。就像业余时间，她喜欢通过旅行见识不同的文化和风景，新的变化带给白静的是新的景观和由此而激发出来的新的热情和动力。“在计算机科学领域，你永远不用担心一直重复做同样的事情。因为这里永远不乏新鲜感，而且能深切体会到你是在真的在改变人们的生活，”白静说。

相关阅读

扫描二维码查看文章

科学匠人 | 秦涛：以独立、深度的视角看世界，做有意义、创新的研究

自加入微软亚洲研究院以来，秦涛已在众多等业界顶级期刊、会议上发表了 100 余篇论文，他所带领的团队也在 2018 年将中英机器翻译提升至人类专业水平，在 2019 年国际机器翻译大赛上获得 8 项冠军，并且还开发了迄今为止最强大的麻将 AI 系统 Suphx，所研发的语音合成技术被广泛应用于微软 Azure 认知服务、Windows 等多个微软核心产品和服务中。如何持续不断地产出高质量的科研创新成果？或许你可以从秦涛的科研经历中获得一份“锦囊”。



科学匠人 | 申博野：以数据为“武”铸机器之智

2012 年，申博野开启了在微软亚洲研究院的科研生涯。他认为自己是一个天生适合做系统研究的人，乐于研究如何让系统、机器变得更加智能化，让技术更好地服务于人类。在申博野眼中，微软亚洲研究院是一个拥有众多有趣灵魂，坚持开放、包容与协作精神的多元化研究机构。申博野认为，多元与包容是一个科研机构不可或缺的重要因素，因为只有包容才能多元，只有多元的人和想法才能让研究为更广泛多样的大众所服务，解决真正的实际问题。



科学匠人 | 高睿：行我所爱，做追逐“月亮”的幸运儿

在微软已有 17 年开发经验的高睿最感兴趣的工作就是“摆弄代码”。他所在的创新工程组是一个介于研究组与产品部门之间的“独特”存在。“由于专业背景的差异，研究员与工程师之间难免会有意见分歧，但大家都秉承包容合作的态度，平等沟通，以最终达成当下的最优方案。”高睿表示，自己喜欢微软的原因之一，是工程师文化在这里已经融入到了“DNA”中。微软亚洲研究院开放、包容的环境则是他能够安心、专心地研究技术的重要原因。



她们非同凡响的“另一面”

“人类经常少年老成、青年迷惑、中年喜欢将他人的成绩与自己相比较，因此觉得受挫，好不容易活到老年还是一个没有成长的笨孩子。我们一直粗糙地活着，而人的一生，便也这样过往了，”：女作家三毛曾这样说过。

这些描述是否与你的人生轨迹有些许重合？我们都不约而同地走上了同样的路，忘记了最初的梦想，失去了曾经的兴趣，与那个不一样的自己渐行渐远。不过，在微软亚洲研究院有这样一群女性，她们在看似重复的科学研究和日常琐事中不断探索内心，始终保持自我，绽放出不一样的烟火。

享受专注，保持对价值的审视

姓名：吴文珊
软龄：2.5年
组别：创新工程组



A面：
连接科研与产品的后端工程开发者，缩短技术与大众的距离是工作的价值所在。



B面：
摔出来的滑雪爱好者，从双板到单板，只选择最合适自己的。

Q：为什么会喜欢滑雪这项运动？

A：爱上滑雪源于一次摔倒，当时我仰躺于滑雪场，望着天、看着雪，那一刻的我很自由、平静，就在那一瞬间我爱上了滑雪。我非常喜欢专注做某件事的感觉，而滑雪恰恰也是需要集中注意力的一项运动。

现在我滑单板，因为它可发挥性更强、更自由，更符合我的个性。我很享受滑雪时沉浸其中的感觉，甚至暑期也会在室内滑雪场练习。我的下一个目标是学会单板立刃，同时未来也计划尝试冲浪。

Q：目前在微软亚洲研究院的主要工作是什么？

A：我所在的创新工程组主要负责连接科学研究与产品落地。大多数时候模型或是先进的技术能力，都离普通人较远。缩短这个距离，让技术服务更多人，是一个很有价值的过程。

目前我在做语言学习应用相关的工作，主要以微软小英为载体。看到微软小英能够真正帮助人们提高英语口语和写作水平，让我成就感爆棚。微软亚洲研究院的氛围让我们可以专注做正确的事，有更笃定的方向，而不是试错后半途而废。这种工作模式很适合我。

Q：如何兼顾高效工作与兴趣爱好共同发展？

A：我会努力将工作与兴趣爱好隔离开来。工作之外是很宝贵的个人时间，我会全身心投入到兴趣中。

Q：工作中是否遇到过一些印象深刻的挑战或难题，如何克服？

A：“这件事情真的有价值吗？”我经常自问。做一件事时，我会有三问：What、Why、How，通过这样的思考方式去做事情，你就会感觉自己一直在成长。而成长的过程就是不断发现问题、定义问题、拆解问题的过程，我会按照优先级逐一攻破。同时，我也会定期复盘工作、整理思路，在推演过程中也会有新的视角和收获并且经常与同事沟通探讨，可以从多角度思考问题，这样看似纷繁复杂的挑战也能迎刃而解。

Q：如何理解“我是我自己”？

A：我们的每一个想法、每一个思路都构成了自己的个性。在科研领域，并不存在男女能力之别，我们各有优势，找到自己的不可替代性，深入理解所做之事、清楚的规划，优势就能得以发挥。

追求完美，热诚、热爱

姓名：朱波
软龄：16年
组别：行政部门



A 面：
严谨细致的执行助理，即使多项任务并行、与多人同时沟通，也能游刃有余。



B 面：
热爱音乐、舞蹈，因一段口哨的即兴表演，受邀加入微软乐队。

Q: 为什么选择口琴这种乐器以及喜欢吹口哨?

A: 我从小是被放养在大自然里的“野孩子”，当需要表达愉悦之情时就吹几声口哨，吹着吹着就成了调。学习口琴则是受我父亲的影响，他喜欢拉手风琴和吹口琴，家里总是萦绕着音乐之声。直到现在，尤其是在感到疲劳和低落时，我都会通过音乐来直叙胸臆，音乐能帮我迅速调整到最佳状态，让我豁然开朗，也能给我的生活和工作带来灵感。

Q: 非技术背景的职能部门在研究院扮演着怎样的角色?

A: 如果将微软亚洲研究院的管理层比作大脑，那么各研究部门就是核心的内脏，职能团队是四肢，而我所在的行政部门则更像遍布全身的经络，它看不见，摸不着，但却是调节整体机能不可或缺的一部分。我们的职责就是以各种方式打造适宜的环境和平台，让同事们高效、愉悦地沟通和工作，创造更大的价值。

Q: 在研究院这么多年，你一直保持热情和新鲜感的秘诀是什么?

A: 加入微软亚洲研究院之前，我就知道这里拥有一群顶级的科学家，所以一直心怀敬仰。2006 年我加入时，研究院被麻省理工学院《技术评论》誉为“全球最火的计算机实验室”。16 年间，我与研究院一起成长，它让我遇见了最好的自己。怀揣着对微软亚洲研究院的满腔热爱，助力有趣又满怀激情的天才同事们做改变世界的事业，就是我每天活力满满的动力来源。

Q: 作为执行助理，每天都会面对繁杂多样的任务，你认为职场人需要具备哪些能力?

A: 与时俱进、快速学习，做事要极尽全力追求完美、做到极致。每一次会议、每一场活动，除了有详密的计划，还要不断创新，给参与者提供最佳的体验。例如，受疫情影响，我们的很多全球

性年度会议都转为线上 + 线下的形式，为了呈现出更好的会议效果，我们要充分掌握和利用 Teams、Forms、SharePoint 等诸多微软产品以及 Azure 上的功能，搭建完美的远程和现场会议平台，沉浸式的体验让远程参会者如同亲临现场；同时，每次重要会议的视频、评论和资料也会被及时整理上传到公司内网，方便大家会后查看，这两点都是与疫情前有所不同的新变化。其实，任何逆境的发生都有它的原因和规律，当你以一种乐观、积极的态度面对时，挑战通常都会转化为机遇。

Q: 最近有在关注什么领域或者话题吗?

A: 非常荣幸，微软是可持续发展的先行者，加之近年来地球上发生的各种各样的灾难也让我越来越多地关注环保问题。除了公司层面具有全球影响力的承诺和举措，我也一直坚持从自身点滴做起，物尽其用，杜绝浪费，保护环境。我希望通过自己的行动，可以尽可能多地影响周围的人。我们每个人的每个小小的善举，集合在一起，长此以往，都会产生意想不到的效果，反之亦然。

Q: 如何理解“我是我自己”?

A: 我仍然在寻找那个理想中的自己，为此我每天坚持以自己的信仰引导、反思、校准自己，以期与我心目中的肖像一步步靠近。我很庆幸自己有几个能够互相促进、彼此激励的良师益友，为此我非常感恩。其实，每个人都有自己的小宇宙，但不是每个人都能找到那把启动它的秘密钥匙，我希望能像我心中的榜样们一样具有足够的能量与智慧，用光和爱温暖照亮身边人，这是我的长远目标。

始于好奇，忠于逻辑和自治

姓名：孟琪
软龄：3.5 年
组别：机器学习组



A 面：
数学专业出身的硬核跨界研究院，不断发现数学之美。



B 面:

AI+ 物理, 致力于解决更多实际问题。

Q: 作为一名数学专业的研究人员, 为什么选择加入微软亚洲研究院?

A: 读博期间, 我曾在微软亚洲研究院实习, 那时我就深切地感受到了这里自由、开放的研究氛围, 所以毕业后这里就成了我的第一选择。在研究院, 你的视野会更广阔、研究课题可以更多元, 研究思路不会受限, 不论遇到什么类型的问题, 你都能找到相应的研究学者共同探讨。

数学往往给人的感觉是非常抽象的, 仿佛飘在天边, 但在研究院, 我们所做的研究离实际问题和产业应用更近, 可以直接感受到研究的价值, 看到研究对业界所产生的颠覆性影响。

Q: 你当初为什么选择数学专业?

A: 起初选择数学专业是因为数学是很多学科的基础, 并且实用性强, 无论是在物理、工程或社会科学中, 它都可以作为一个通用工具, 我们可以利用数学解决许多问题。而且, 数学的逻辑性和自洽性也是吸引我的地方。在学习的过程中, 我时常能感受到人类在不断突破智慧的极限。我非常欣赏、崇拜前人所构建的理论, 也希望自己能够掌握数学的奥秘。

Q: 数学在你目前的研究领域中发挥着怎样的作用?

A: 我现在的研究方向是机器学习理论, 我们希望通过数学建模, 来了解深度学习背后的机理, 例如, 我们用动力系统的工具来分析深度学习中的优化器, 刻画随机噪声或二阶近似算法对模型性能的影响; 我们也曾利用群论的知识研究神经网络的群对称性, 提出在路径空间描述神经网络的新算法。也有一些数学工具是我在工作中新学习的, 对我来说, “活学活用, 即用现学”也是研究的乐趣之一吧。

Q: 从数学到计算机再到物理学, 跨界研究是一种怎样的体验?

A: 在微软亚洲研究院, 我们所进行的 AI for Physics 的研究是希望通过 AI 技术助力基础科学领域的发展, 例如发现新的物理定律等。但这其中最大的挑战是找到物理学中需要被解决的那些关键问题。跨领域研究通常都存在知识壁垒。身为计算机领域的科研人员, 我们并不了解物理研究的前沿方向, 所以必须与众多领域专家深入交流, 先理解跨界领域的知识, 再来评判利用 AI 技术解决未知问题的可能性。

Q: 对希望从事计算机跨界工作的学生或研究员们有什么建议?

A: 跨界研究是非常有挑战的工作, 要摆正跨界研究的心态, 怀有敬畏之心。要对跨界的问题感兴趣, 以兴趣为驱动, 与领域专家深入交流, 才不至于浅尝辄止。

Q: 如何理解“我是我自己”?

A: “我是我自己”大概意味着学会独立思考, 在工作、生活、人际交往中找到属于自己的平衡点。在成长的过程中, 当我的内心所想与周围环境有冲突时, 我会尝试通过不断学习和积极沟通来找到平衡。工作后, 微软亚洲研究院多元、包容的研究文化也给了我们自由选择的空间, 所以在研究院坚持自我并不是一件难事。

相关阅读

[扫描二维码查看文章](#)

女性 × 科技: 撕掉标签, 乘风破浪

2021 年国际妇女节时, 我们和三位在微软分别工作了 10 年、15 年和 20 年的女性聊了聊。她们在微软亚洲研究院从事着不同岗位的工作, 一位是“学霸”研究员, 一位是对外合作“大使”, 还有一位超级“大管家”。从她们丰富的经历中, 我们看到了科技行业女性的独特魅力和无限可能。



微软乐队: 我的 BGM 我做主

微软亚洲研究院二十周年庆典前夕, 来自不同部门、不同研究领域的几个人聚在一起, 组建了一个集东西方乐器于一体的乐队。此后, 在微软内部的各种活动中, 总能看到他们活跃的身影。正如微软推崇的多元包容文化一样, 微软乐队涵盖了十几种乐器, 乐队成员也身处不同的岗位。本篇文章带我们了解四位成员与微软乐队的故事。



星跃重洋 | 刘国栋：非典型理工男在微软亚洲研究院的科研“旅”记



结束一天的科研工作，中国科学院计算技术研究所博士生刘国栋走出微软大厦，抬头是繁星点点，耳机里装的是苏打绿乐队温暖的旋律。

通过“星跃计划”在微软亚洲研究院实习一年，刘国栋在微软亚洲研究院主管研究员苗又山、微软雷德蒙研究院高级研发工程师 Saeed Maleki 两位 mentor 的指导下，围绕加速深度学习模型的训练进行着科研探索。

恰如设立初衷，“星跃计划”在优秀人才与微软全球两大研究院的研究团队之间架起桥梁，为他们创造了一起聚焦真实前沿问题的机会。对刘国栋而言，这场“跨越重洋”的科研之旅不仅让他实现了自己的科研设想，也让他在思考方式和科研品味上有了新的顿悟。

一名科研小生，如何在科研道路上披荆斩棘？

一位非典型理工男，如何汲取科学与艺术的魅力？

在微软亚洲研究院，刘国栋找到了属于自己的诸多答案。

用简单漂亮的思想，解决复杂的问题

“想通过简单但漂亮的思想，解决复杂的问题。”这是刘国栋加入微软亚洲研究院“星跃计划”的机缘与初衷。

在一次国际会议上，他读到一篇令他印象深刻的文章：微软亚洲研究院和北京大学、上海科技大学合作提出了一种可以提升深度学习计算速度的编译框架 RAMMER。“研究员们通过一些很漂亮的系统层面的思想，解决了一个很有意义的问题。”刘国栋非

常期待自己也能做出这样的工作。

他与微软亚洲研究院的故事便由此拉开序幕。去年一月，当看到“星跃计划”招募文章中的 High-performance Distributed Deep Learning 项目时，刘国栋感到眼前一亮，“我之前就一直关注这个方向的研究，相关课题对于硬件资源的要求比较高，我在学校做有点困难，而微软恰好拥有非常丰富的软硬件资源。没有过多犹豫，我直接申请了这个项目。”刘国栋的导师、中国科学院计算技术研究所副所长包云岗教授也很支持刘国栋更多地进行有价值的科研探索。去年三月，刘国栋正式入选“星跃计划”。



一年的“星跃”之旅已颇有成果。由于深度学习模型的训练往往非常耗时，很多研究者都在尝试用各种方法加速训练的过程，以节省时间成本和经济成本并节约能源，帮助其他研究者更快地实现他们的模型。刘国栋及其所在研究组提出的优化方法，已经在诸如 T5 这样百亿参数量的模型上，实现 1.5 倍的加速。

“这个数字听起来比较小，但实际上一个模型的训练不仅需要很多硬件资源，也常常耗时数十天。所以哪怕能加速 10%，带来的经济收益和节约的能源也是非常可观的。”刘国栋介绍道，接下来他们会将这些思想和方法形成论文，并期待与其他课题组合作，将实验性的想法应用到一些大规模模型的部署中，如 GPT、T5、Swin Transformer 等。

在 mentor 苗又山看来，刘国栋的科研潜质不仅在于扎实的专业知识，更重要的是勤于思考、专注钻研的特质与探索科学问题的热情，因此能以较快的成长速度在后期承担重要的实验工作。另一位 mentor Saeed Maleki 也表示，刘国栋加入“星跃计划”以来做出的成果给他留下了深刻的印象：“国栋刚开始对我们正在研究的课题没有太多经验，但就在一年的时间里，他优化的技术在 Transformer 模型上实现了比 Megatron 更好的并行效果，而 Megatron 是目前分布式机器学习公开领域最先进的系统。”

“1.5 倍”加速背后是现实对理想的“超车”

创新性和飞跃式的工作从不是一蹴而就的。刘国栋在提出分布式训练的优化方法之前,进行了大量基础繁琐的工作:修改参数、收集数据、图表分析……正如学校导师包云岗教授对他的教导,“系统性的研究,需要全力投入、亲自动手——先把自己的手弄脏’(get your hands dirty)。”在微软亚洲研究院,他也践行着这样的原则。刘国栋坦言,这个过程非常耗时且单调,但基础工作必不可少,然后才能找到研究的突破口。

在研究初期,为了提高模型总体的训练效率、进行分布式的训练加速,刘国栋提出了一种更加灵活的模型分布式部署方案。“这个点子的内核思想是很简单的,但我期待它可以漂亮地解决问题,”刘国栋说。然而工作推进一段时间后,刘国栋发现:大部分模型都比较规整,模型里的神经网络层都非常相似,这就使得灵活的分布式部署的优化空间非常有限。

“当时比较沮丧,觉得想法很美好,但现实并不那么好。”正在刘国栋想要放弃的时候,mentor 苗又山认为他的点子很有意思,并引导他换一种思路:这个方法能不能用到别的模型上?用于别的场景会不会有更高的收益?于是,刘国栋开始尝试其他模型以及硬件异构的场景,并使之实现了更高的收益,让这个点子发挥了更大的价值。

一波未平一波又起,接踵而至的是第二个困难。当刘国栋试图用一种更加灵活的方式来分布式部署模型时,内存的开销也随之变大,内存开销变大限制了性能的提升。“又是理想跟现实不一样的状况,”这一次刘国栋主动和 mentor 沟通,苗又山提醒他:问题变复杂并不是一个坏事。

确如苗又山所说,刘国栋意识到,自己发现了一个前人没重点关注的现象——对模型的灵活部署导致内存开销增加的问题,或许是他未来研究的新空间。“如果这个问题只需想一层就能解决,那这个工作本身的意义就没那么大了。”刘国栋认为把沮丧转化成机会,就能经常在科研之中感受到一些比较有趣的部分。

如“过关斩将”一般,经过这两次挫折与突破,刘国栋的心态愈渐稳定:做科研常常面临理想与现实的差异,但挫折反而会让他感到兴奋。再面对意料之外的问题时,刘国栋会选择去追问问题背后的原因,“一旦追问为什么,可能会发现其实它不是你实现上有错误,而是你在某个地方就没有想清楚。”刘国栋认为,“回溯”能帮助自己更好地理解问题,同时也可能会给研究带来新的可能性;更不要害怕理想与现实的差距,因为出现差距很可能是一个新的机会。

星跃重洋,“1+1>2”的 mentor 指导

对于刘国栋来说,无论是科研工作上的进展,还是科研心态上的成长,他都在微软亚洲研究院收获颇丰。“星跃计划”采取双导师制,项目中的实习生都会在微软亚洲研究院、微软总部两方

导师的指导下进行科研工作,并有机会与不同研究背景的科研人员深度交流。在双导师的带领下,实习生们将聚焦工业界的真实前沿问题,致力于做出对学术及产业界有影响力的成果。

刘国栋的两位 mentor 拥有不同研究背景,他们从不同视角为刘国栋带来了“1+1>2”的科研助推。作为在系统方向深耕的研究员,苗又山常在思路层面对刘国栋进行引导。刘国栋说,“曾经我特别容易陷入问题的细枝末节里,但现在我意识到:如果这个问题本身没那么重要,就应该赶紧跳出来往前走,并学会转换视角去思考问题。”苗又山的指导让他形成了主线意识和发散性思维。



刘国栋(左二)与 mentor 苗又山(右二)及组内同事合影

Saeed Maleki 则在分布式机器学习领域有着丰富的工程经验,常常会从另一个视角来看问题。“他会很敏锐地捕捉到通信效率提升的可能性,并提示我从通信的角度看待我所做的优化。”刘国栋说。

在刘国栋看来,两位 mentor 都是温和的人,但对研究又高度严谨。“苗老师做研究的要求高,所以常常会很直接地对我的工作提出疑问。他会问我,你提的这一点它背后的原因是什么?这么做到底有没有意义?”刘国栋回忆道,自己长期以来处在较为安全的环境中,不会面对太多的质疑和批评;在微软亚洲研究院面对 mentor 不断提出的疑问,让他在思考问题时更加深入和清晰,也能更加开放、从容地面对质疑了。

对于刘国栋的成长,mentor 苗又山并不意外。在他看来,刘国栋拥有很强的内驱力,能够主动地探寻新的思路,并且通过不断地反思和调整提升自己的认知。这样的思维方式使刘国栋的思考逐步深入,从而揭示更本质的问题,并引导他有效地制定进一步探索的计划。与此同时,在一次次跨洋组会中,刘国栋的英文演讲能力也获得了明显提升,无论是宏观思路的流畅性还是细节的清晰性。

与研究员、实习生们的交流、碰撞让刘国栋受益颇多。他所在课题组每周都有两次常规组会来更新工作进展,所有实习生进行工作汇报并开展广泛的讨论。对于具体的问题,刘国栋时常会和两位 mentor 发起临时讨论,解决急迫的问题或进行头脑风暴。除此之外,每周还会进行一场大组会,除了刘国栋的两位

mentor, 微软亚洲研究院和微软雷德蒙研究院相关方向的许多资深研究员都会参与。

“在大组会上, 我有机会听到更多实习生和研究员做的工作, 这常常对我很有启发。其他研究员也会对我的课题提供建议和想法, 这能有效弥补我的缺漏。”刘国栋表示, 在线下与线上的交流合作中, 他也愈渐感受到微软两大研究院国际化、开放的科研氛围, 以及多元与包容的文化。



刘国栋参加微软亚洲研究院活动

非典型理工男, 科学与艺术的碰撞

这样一个热衷于与科研“较劲”的理工男, 却常常被朋友们评价为: “文艺青年”。在很多个天朗气清的午后, 刘国栋会选择抱着相机, 走在北京的大街小巷, 看晨曦暮霭, 看浮云悠游; 或是穿梭于各个艺术展览, 在每一件欣赏的展品前流连驻足。

这样一位“非典型理工男”为什么会选择计算机专业? “因为我感觉计算机领域蕴含了许多有趣的思想, 这和艺术一样都很吸引人。”刘国栋认为科学思想也可以非常迷人, 并且也是具有艺术性的——科学与艺术的碰撞, 促使他更愿意去做科研工作。

对于理性与感性的平衡, 刘国栋也曾在微软亚洲研究院受到启发。在微软亚洲研究院院长周礼栋博士的“咖啡时间”分享会中, 周礼栋院长“要去找你的内在动力”的观点令刘国栋印象深刻——只有找到自己的内在动力, 才可能真正对自己做研究的状况和生活状态感到满意。

周礼栋院长分享的很多观点与思考, 直到现在仍影响着刘国栋, “后来我也一直在寻找自己的内在动力, 我会一直把这件事情作为生活中的一个小计划和任务。”刘国栋认为, 科学中的艺术性算得上是他的动力之一, 但他仍将继续寻找和追求其他的源动力。

关于未来, 刘国栋计划继续完善在微软亚洲研究院的研究, 并希望有机会与微软亚洲研究院的研究员持续合作, 实现更多科研新想法。“在学界与业界之间, 微软亚洲研究院提供了一个完美兼顾两者的新选择——在工业界做研究。”刘国栋说。“在企业里面做研究, 是一件挺有意思的事情, 我体会到了一种做科研的新模式。”刘国栋说, 在微软亚洲研究院他不只是接收到更多前沿资讯和技术, 更重要的是观念层面的变化: 避免增量式而无创新的



刘国栋(后排左二)参加微软亚洲研究院院长周礼栋博士(前排右三)“咖啡时间”分享会

工作、提升自己的研究品味、做能影响和帮助更多人的科研……

“做真正有意义的工作, 并将之完成得很漂亮, 内心会感到踏实、满足。”刘国栋表示, 从这里再度出发, 他希望自己今后能做出让更多人受益的工作。

Mentor 寄语



苗又山
微软亚洲研究院
主管研究员

刘国栋同学是一名非常优秀的实习生。他专业知识扎实, 且思维敏捷、勤于思考、认真专注。他入职近一年以来, 我们的合作非常愉快。实习期间, 他不断学习提高自我的同时, 也在我们 Redmond、Asia 两个研究院合作的科研项目中发挥了重要的作用。我希望刘国栋同学能够保持对于科研的好奇, 不断进步、再接再厉, 在研究的道路上获得更大的成功。



Saeed Maleki
微软雷德蒙研究院
高级研发工程师

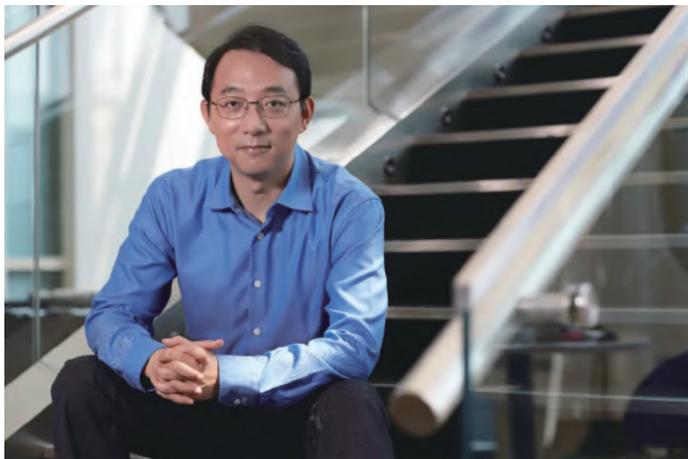
Guodong's greatest ability is to catch up to the state-of-the-art for a research problem. We have been working with Guodong and been impressed with his accomplishments given that he didn't have much experience with the topic we are working on. In the span of one year, he has developed techniques that beat Megatron parallelism for transformer models which is the state-of-the-art for distributed machine learning. Currently, he is expanding his work to utilize AI-based tuning techniques to further optimize the machine learning workloads. I look forward to continuing working with him and seeing him thrive.

周礼栋：成熟的 AI 不再只是谈“颠覆”

来源：文汇报

身为世界级计算机专家，IEEE（电气与电子工程师学会）与 ACM（国际计算机学会）双料院士，新任微软亚洲研究院（MSRA）院长周礼栋博士与上海颇具渊源。他曾上海生活、学习多年，后来虽在海外长期留学工作，但是直到现在，周礼栋仍对上海充满感情。目前，作为顶级研究机构的微软亚洲研究院已在上海搭建团队，希望能在这片人工智能（AI）的沃土上做出世界级成果，并建立更强的产学研合作体系。

人工智能未来会向何处发展，上海该怎样利用自身优势打造人工智能高地，广大学子该如何提升自己以适应 AI 时代。围绕这些话题，周礼栋近日接受了《文汇报》记者的专访。



微软亚洲研究院院长周礼栋

应该特别关注“跨界”

“我在上海待了很多年，又在复旦读书，所以回到上海就有一种特别的亲切感。而且直到现在，用上海话交流依然是‘零障碍’。”采访一开始，周礼栋就谈起了自己的“上海情结”。他回忆说，上世纪 90 年代，他在复旦大学计算机系读本科时，虽然国内的计算机教学、科研以及设备与世界顶尖水平有较大差距，但教授们兢兢业业、呕心沥血的培养，让周礼栋他们在理论方面完全不逊色于世界顶尖学校。当他留美进入康奈尔大学后，第一个学期，就在素有“魔鬼课程”之称的“计算机算法设计与复杂性分析”和“计算机语言理论”两门课上拿到双 A。作为该届唯一的双 A 博士生，周礼栋对复旦老师的培育感恩至今。

自从 3 年前首届世界人工智能大会在上海举行，参与其中的周礼栋就见证了上海各级政府对于发展 AI 的决心。他认为，上海打造 AI 高地的条件得天独厚：一是人才资源充分，二是产业基础深厚，三是应用前景广阔。目前，微软亚洲研究院已在上海与本

地的金融、制造、物流等行业合作，一起探讨人工智能解决方案；同时，上海的国际化特色也帮助他们吸引和培养更多尖端人才。

在周礼栋看来，现阶段发展 AI 最重要的就是要“跨界创新”。不仅要关注 AI 企业，更要让 AI 与千行百业结合，打破技术界限，深入应用场景，从而真正发挥潜力。

过去两年，微软亚洲研究院在上海积极推动产学研合作，与上海仪电、华东师大、复旦大学等建立合作，将自身研究能力与特定行业对接。周礼栋说，跨界合作重要，但并不容易，因此必须坚持。只要坚持，终究会有收获。通过与行业合作，微软亚洲研究院希望从具体案例中抽象出新 AI 逻辑，从而打造出通用的 AI 平台，并通过开源方式让更多企业受益。在周礼栋看来，这种合作机会在国内普遍存在。

足够尊重才能“融入”

谈到人工智能，周礼栋认为，这个行业越来越成熟，一个重要标志就是 AI 从业者不再像前两年那样，动辄就想“颠覆”。大家意识到，AI 并非无所不能，传统行业自有其积累，AI 应该放下身段，给予传统行业足够的尊重。只有真正融入，才能发现痛点和问题，然后想办法利用技术加以改进。

周礼栋说，如今 AI 的主攻点，应该是为各个行业赋能。比如可以通过强化学习，提高原有系统的效率；或是通过智能预测，给出更合理的规划。无论如何，AI 与传统行业的结合，离不开充分的了解、磨合，只有拿出足够的耐心与尊重，才能让 AI 落地。

事实上，AI 行业从“颠覆”到“融入”，这种轨迹与周礼栋职业生涯似乎有一定的契合度。20 年前，他毕业后来到微软的硅谷研究院。回忆起博士阶段，身边很多人似乎都有些“急功近利”，希望尽快成为本领域的“大拿”。但在硅谷研究院，周礼栋结识了许多一流研究员，包括两位图灵奖得主，他发现，这些人做研究，不是为了发论文或是拿奖，对他们来说，研究是一种对真善美的长期追求，这个过程让他们非常享受。

硅谷的经历给周礼栋很多感触，让他在日后不断提醒自己：不要为了追逐短期目标而去牺牲研究的本来意义，同时，在心态上要从竞争转到合作。此后，当他回国进入微软亚洲研究院，工作重心再次发生变化，从之前近似“独立研究员”转而成为团队负责人。虽然这一度让他经历了痛苦的转型，需要努力学会如何从挫折中挣脱、在不同环境下快速适应，但最终，他带出世界级的研究团队，体会到了比个人成功更强的成就感。

而今，执掌微软亚洲研究院的周礼栋更多考虑的是社会责任。他希望微软亚洲研究院能为中国乃至全球社会以及学术界作出自己的贡献；同时，能在面对疫情、气候、环境等全世界共同的挑战时，施展自己的抱负。他说，AI 研究要“立地”、也要“顶天”。AI 要解决根本性问题，体现出革命性价值，说到底需要通过物理、化学、生物等基础学科相融合。比如，预测蛋白质结构的人工智能 AlphaFold 已对本行业形成了颠覆性影响；将来，在基因组学、新材料、气候变化、环境保护等方面，AI 都有很大的施展空间。



文科生也能做一流 AI 研究

正因为 AI 越来越强的“跨界”与“赋能”属性，使得该领域对人才的需求也相应地出现了改变。周礼栋说，如今，狭义的 AI 人才——那些擅长调参数、做模型的人才，实际上并不见得有多短缺；我们缺的是广义的 AI 人才，这些人不仅要懂算法、懂编译器，还要具备系统知识，了解计算机结构，甚至懂硬件。

周礼栋最看好的 AI 人才，可能是那些非计算机专业背景的学生，也就是具备各种专业背景，同时了解人工智能的跨领域人才。他再次以预测蛋白质结构的 AlphaFold 为例解释说，这个研究团队内，有许多生命科学背景的人，他们了解该领域有哪些问题值得挖掘，在哪个点上最有可能突破。如果离开这些人，可能连选题都很难确定。

微软亚洲研究院近年来也顺应跨界趋势，吸收了多位非计算机专业人才。除了理工科背景的之外，他们甚至还招募过文科生，



周礼栋说，他并不建议文科或者理工科学生将来都转到计算机专业；但是他希望大学生接受一定的计算机专业教育，从而具备计算思维，进而可以从本专业出发，更好地应用 AI，推动跨领域的合作。他认为，上海拥有许多一流的大学，面向下一代 AI 人才的培育，各高校都可以尝试从“跨界”出发，对现有的人才培养体系进行创新，比如把 AI 及计算机课程变成一种通识教育，让计算机学院的教授能帮到各个院系。

此外，同样需要接受计算思维训练的还包括企业管理人才。周礼栋说，决策层特别需要建立人工智能意识，推动数字化转型，“那些有远见的企业已开始朝着这个方向大踏步前进。在这方面起步越早的公司，将来就会越有优势。”

相关阅读

[扫描二维码查看文章](#)

周礼栋：用长期主义，回归科学研究的本质

在升任微软亚洲研究院院长后，人才培养及与之相适配的文化建设，成为周礼栋眼下最重要的“长期目标”，他认为，微软亚洲研究院的理念一向以人才培养为核心，而不是将短期目标或某个项目的成功作为优先选项。在人才培养的方式上，周礼栋也同样秉持着“长期主义”。在他看来，“真正的创新不能靠施加外部压力，而更多源于内在动力。”面向未来，他希望能与微软亚洲研究院的同事以及各界合作伙伴，共同回到研究的核心和本质，持续引领科技创新，为整个社会的发展做出贡献。



打造一流创新环境：协作、开放、可持续

无所不在的协作成就了我们所处的世界与文明。创新需要的是深层次的、可持续的协作，让分工、协调、合作链环上的每个人都可以成为新的发现与贡献的核心主体。在微软亚洲研究院院长周礼栋看来，研发机构打造多元与包容的独特文化至少需要依循以下三个原则：每个人都有发言权，管理者应当倾听和重视所有不同的声音，让思想在辩论和建设性批评中演进，让最好的想法获胜。



微软云强劲增长的背后，是全新的人工智能黑科技

来源：云科技时代

众所周知，近年来微软 Azure 云计算平台一直保持着强劲的增长势头。在上一个财季的财报中，微软 CEO Satya Nadella 表示，数字技术是通胀经济中的去通胀之力，不论是大企业还是小企业，都能通过构建自己的技术强度而提高生产力，以及让自己的产品和服务被更广泛的采用。端到端交付数字平台和工具的微软云，正帮助企业和组织在当前的转型和巨变中稳步前进。

在上一个财季，微软云贡献了 207 亿美元的营收，年增长 36%。而就在新一季财报即将发布的近期，微软又宣布斥资 687 亿美元收购动视暴雪，收购后的微软将成为世界第三大游戏公司，数以亿计的用户将在微软云上获得更加沉浸的游戏和元宇宙体验，而微软也将更加丰富的内容，通过微软云带给所有人——这无疑将进一步加大微软云的运维难度。

如今，微软云有上千万台物理服务器，部署在世界五大洲上千个数据中心，运行着上百万客户的应用和服务，其中包括 95% 以上的财富 500 强企业的服务，每个月都有新的硬件上线使用，几乎每天甚至每分钟都有新的软件更新部署到云上，微软云的复杂程度已经远远超出人类的理解范围。那么在微软云的背后，又有怎样的黑科技呢？

近日，微软亚洲研究院常务副院长、微软杰出首席科学家张冬梅和微软亚洲研究院首席研究员林庆维，以及微软云计算与人工智能事业部首席数据科学家党映农，联袂向外界披露了微软云背后的黑科技——人工智能运维 AIOps。

软件工业的数字化转型

近 15 年来的云计算迅猛发展，为人工智能运维 AIOps 提供了广阔的发展空间。所谓人工智能运维 AIOps 就是以大数据和人工智能的方式，替代原先人工对数据中心的运维方式。在公有云数据中心出现之前，都是企业自建数据中心，相比来说都是中小型数据中心而且位置分散，出现的故障数量不高、故障类型也难以全面，因此很难形成用于人工智能分析的大数据。而公有云的出现，改变了一切。

在数字经济大发展的前提下，公有云也获得了大发展，以微软云为代表的公有云，出现了全球化的超级云，例如微软云就有上千万台服务器，因此在微软云一个云平台上就有可能获得全局、全量、全面的数据中心运维大数据。其次，由于公有云成为了全社会的基础设施，全社会对于公有云运营商的要求也水涨船高，这就倒逼公有云运营商必须提高数据中心运维的自动化和智能化

水平，从人工方式切换到更加稳定、高效而又安全的人工智能运维 AIOps。

张冬梅介绍，微软亚洲研究院在人工智能运维 AIOps 研究方向的起步非常早。可以追溯到 2008 年到 2009 年（2008 年微软云 Azure 上线），那个时候微软亚洲研究院成立了 Software Analytics Group（软件分析组），希望从数据驱动的角度对软件领域进行研究，在软件场景下最重要的问题包括运行系统问题、用户体验问题、开发效率问题等，这与今天的 AIOps 非常类似。



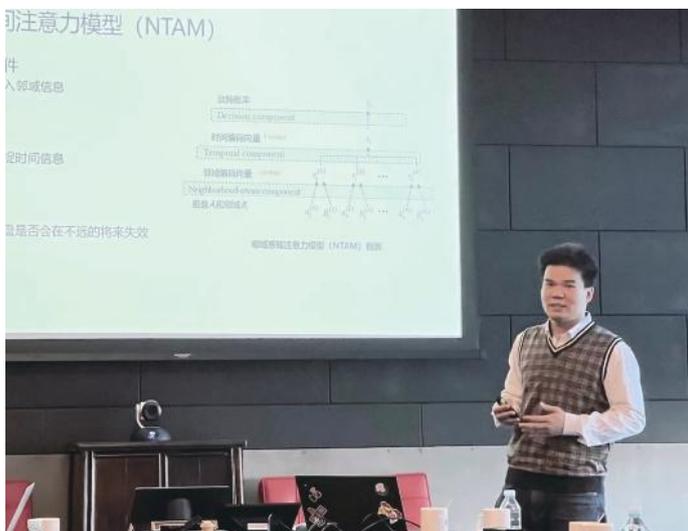
微软亚洲研究院常务副院长、微软杰出首席科学家张冬梅

云计算成为了过去 10 到 15 年内软件和软件工业的范式转变动因和平台。当云计算系统变成了软件运行的主要形态时，Software Analytics 自然地就会把研究焦点集中到云计算系统。随着软件从以前面向单机和服务器到面向超大规模云数据中心的延展，软件工程面临着深刻变革，从之前主要关注程序员，向关注系统、用户和开发者的扩展——这是因为在公有云中的软件，最终将运行起来变成系统，从而对外提供云服务，同时，作为系统还要关心用户体验和开发部署效率，这是三位一体、不可或缺的模式。

总结下来，云计算不仅是企业数据中心的转型，更推动了软件和软件工业的数字化转型——软件的设计、开发、运维等由人工方式转向数据和人工智能驱动，由此带来了人工智能运维 AIOps 的可行性和可能性。而微软亚洲研究院自十多年前开始的 Software Analytics 研究，保障了今天微软云的稳定、可靠、安全等特性，进而帮助奠定了微软云全球第二大超级云的地位，以及业绩不断上涨的强势局面。

AIOps 的三大方向

目前，业界对于 AIOps 的研究以及实践，仍处于初级阶段。对比自动驾驶的五个等级，AIOps 可以说是数据中心的“自动驾驶”，也可以分为 L0-L5 等级，其中 L5 对应于完全“无人驾驶”的最高级，可以说是 AIOps 的终极目标。当前，市场有不少 AIOps 的供应商，已经可以部分地实现从 L0-L4，但在追求 L5 方面，仍是一个十分前沿的领域。



微软亚洲研究院首席研究员林庆维

对于微软云这样一个超级庞大而复杂的分布式系统来说，L5 是终极追求目标，而且 L5 这样的终极目标也只有微软云这样的全球化超级公有云才有可能实现。张冬梅介绍，AIOps 通过创新的人工智能和机器学习技术，有效且高效地设计、构建和运营大规模复杂的云服务。从微软亚洲研究院的角度，AIOps 主要有三大研究方向：系统、用户与开发者。

首先是系统服务，即 AI for System。云服务由部署了软件的数据中心系统为客户提供服务，软件必须要在云数据中心上运行起来，与云数据中心一起组成可运行的系统，才能对外提供云服务。而从可运行的系统角度来看，系统的性能、稳定性、安全性等各方面问题，都是需要研究的对象。其次是客户，即 AI for Customer。与互联网服务不同的是，云服务不仅服务于个人，更服务于企业，既然是为用户服务，一定要注重用户体验。即使系统再稳定，但用户体验不好的话，大家也不会选择云服务。第三是开发者，即 AI for DevOps，主要是面向开发人员和运维人员，用智能技术提高他们的生产率，让日常工作更顺畅。

在系统服务方面，经常遇到的场景有两个：异常行为检测，也就是当系统出现问题了，必须要能够检测出来；预警，即在问题出现之前预测可能要出现的问题。对于这两个场景，主要基于数据和机器学习的方法，再结合专业领域知识，就能够进行比较好地判断和预判。

在用户方面，一个场景就是加强与用户的互动，从而让用户获得很好的体验。例如，当用户遇到问题的时候，经常会联系云服务的客服人员，如果事先能够为用户提供一些工具，让用户自动获得云上服务的构建情况以及到底出现了什么问题，当与客服人员沟通的时候，客服人员也能第一时间掌握具体情况，双方在同等知识层面上交流，这样才能更好地帮助用户。

在开发和运维方面，主要是帮助开发或运维人员更好完成诸如 CI/CD 持续集成和持续部署等工作。而当出现了问题，一方面要快速找到解决办法，让云服务尽快地回到正常状态，但系统恢复正常并不等于根本性解决问题。很多时候，因为系统非常复杂，要花很多时间去研究、发现和调试，才能找到根因。由于日志量非常庞大，就需要通过智能的方法，帮助开发人员尽快高效、准确地完成诊断。

除了面向系统、用户和开发者外，AIOps 需要解决的问题类型方面，又可以分为检测、诊断、预测与优化等四大方面。在每一个问题方面，都有很多挑战以及多个解决问题和挑战的细分研究领域。归根结底，云平台非常复杂，不仅规模大而且为分布式架构，因此 AIOps 是一个需要长期投入的研究领域。

微软云背后的黑科技



微软云计算与人工智能事业部首席数据科学家党映农

对于类似微软云这样的大规模高度复杂并承载大量客户应用的云计算系统，很难运用传统的非智能的软件开发和运维技术进行高效地开发、部署、运营和管理。微软云计算与人工智能事业部首席数据科学家党映农介绍，早在五六年前，微软云计算部门的工程团队就深切意识到了实施智能运维的巨大必要性和迫切性，开始建立专门的数据科学家团队，并且与微软研究院深度合作，进行智能运维方面的研发以及部署工作。

通过过去几年和微软研究院的深度合作，以及微软云工程师们的不懈努力，今天微软云已经在智能运维方面积累了不少重要的技术创新，包括云服务系统的智能化和管理自动化、云开发和

部署的智能化，以及智能化客户响应等等。具体来讲，人工智能和机器学习技术已经深度集成到微软云基础设施的管理软件中，包括智能监控、智能预测、智能修复等等。

党映农强调，自动化和智能化一起推进，一方面使得云服务的可用性、可靠性以及效率得到提升，另一方面使得云服务运行的自主性得到提高，需要进行人工维护的场景不断减少，维护代价不断降低。而机器学习技术也极大改善并增强了微软云的开发和运行维护解决方案的功能，比如像智能测试、智能诊断、智能部署等等，大幅提高了开发和运营工程师的效率。

以常见的硬盘故障为例。林庆维介绍，硬件问题是虚拟机宕机的原因之一，而硬盘故障是造成硬件问题的主要原因之一。为此，工程师们希望在出现故障之前，能更早地预测出故障，然后采取措施将用户的虚拟机迁移到其它机器上，或是通过软启动的方式等解决问题，从而让用户不受影响。然而，在硬盘故障预测中，大规模复杂的云计算平台上故障磁盘与健康磁盘的比例可能有 3: 10,000，这样极度不平衡的正负样本为传统的机器学习预测带来了极大的挑战。此外，在硬盘彻底无法使用之前，其上层应用已经受到影响，因此单靠硬盘自身数据是无法做到即时预测的。

解决硬盘故障，就遇到了 AIOps 研究中的一大难题，这就是小数据样本。那么，微软亚洲研究院的研究员又是怎么解决问题的呢？首先，不能光看硬件本身的数据，而要将硬件相关的上下游数据都连在一起看问题，这样就极大扩展了可用数据。例如，远在硬盘出现故障前，可能硬件之上运行的虚拟机性能就已经受到了影响，这时就可以通过监测虚拟机性能来提前判断是否硬盘将要出现故障；另外，硬盘基本上都放置在同一个磁盘阵列中间，如果主板电压不太稳定，当一个硬盘坏掉之后，可能同时影响到其它硬盘，或者工作负载也会受到影响，因此把邻近的硬盘当作一个整体，也可以很好的进行预测。

基于上述思路，微软亚洲研究院提出了邻域 - 时间注意力模型 (NTAM)，包括邻域感知组件、时间组件、决策组件等，在时间和空间上都能够捕捉更多信息，让模型预测能力更强。通过大量的数据实验，与过去 10 年到 20 年最前沿期刊或者会议上的论文方法进行比较，微软亚洲研究院提出的模型预测能力从精确度和召回率方面都有更佳效果。

在大规模服务故障预测方面，为了最大程度避免大规模服务中断 (outage)，减少服务停机时间，确保云服务的高可用性，微软亚洲研究院开发了一种智能的大规模中断预警机制 AirAlert，可以在云服务大规模中断发生前预测中断的发生。AirAlert 收集整个云系统中的所有系统监控信号，检测监控信号之间的依赖性，并动态预测整个云系统中任何地方发生的大规模中断，然后使用一种叫做鲁棒梯度提升树 (robust gradient boosting tree) 的技术，预测潜在的大规模中断。研究团队在微软云系统收集了超过 1 年的服务中断数据，并在数据集上验证了该方法的有效性。

在系统实际运行中时而会发生某些系统故障，导致系统服务质量下降甚至服务中断，通常称为服务事故 (Service Incident)。云服务事故往往会带来巨大的经济损失，并且严重损害公司的商业形象。过去几年，微软亚洲研究院采用软件解析的方法来解决在线系统中的事故管理问题，开发了一个系统云服务分析工作室 (SAS) 来帮助软件维护人员和开发人员迅速处理、分析海量的系统监控数据，提高事故管理的效率和响应速度。SAS 在 2011 年 6 月被微软某在线产品部门所采用，并安装在全球的数据中心，用于大规模在线服务产品的事故管理。通过分析半年的 SAS 使用记录发现，工程师在处理大约 86% 的服务事故中使用了 SAS，并且 SAS 能够为其中约 76% 的服务事故处理提供帮助。

微软亚洲研究院长期深耕数据智能领域，利用大规模数据挖掘、机器学习和人工智能技术对纷繁复杂的运维大数据进行实时分析，为系统维护提供有效的决策方案。如今，微软亚洲研究院的研究成果已经应用到了微软 Skype、OneDrive、Office 365、Azure 等诸多在线服务中，成为保障微软云和在线服务高质量运行背后的黑科技。未来，微软亚洲研究院也希望打造更加通用化的 AIOps 技术，帮助更多的用户以及业界提升云服务的整体运维水平，夯实云计算作为新型社会基础设施的“地基”。

相信 AI 的力量「AI 中国」机器之心 2021 年度评选结果公布

自 2017 年至今，机器之心「AI 中国」年度评选已连续举办 5 届，每一届评选都得到了人工智能领域内公司、机构和专业从业者的积极响应和大力支持。

本次评选中，微软亚洲研究院成功获选“最强技术实力公司 Top 10”。该榜单主要关注企业的技术实力、对新一代人工智能技术的研发布局与技术储备。入选企业具有强大的技术实力，重视对新一代人工智能技术的研发布局，在研发人员、论文发布、专利储备上拥有强大的储备，并在相应人工智能技术细分领域中处于顶尖水平。

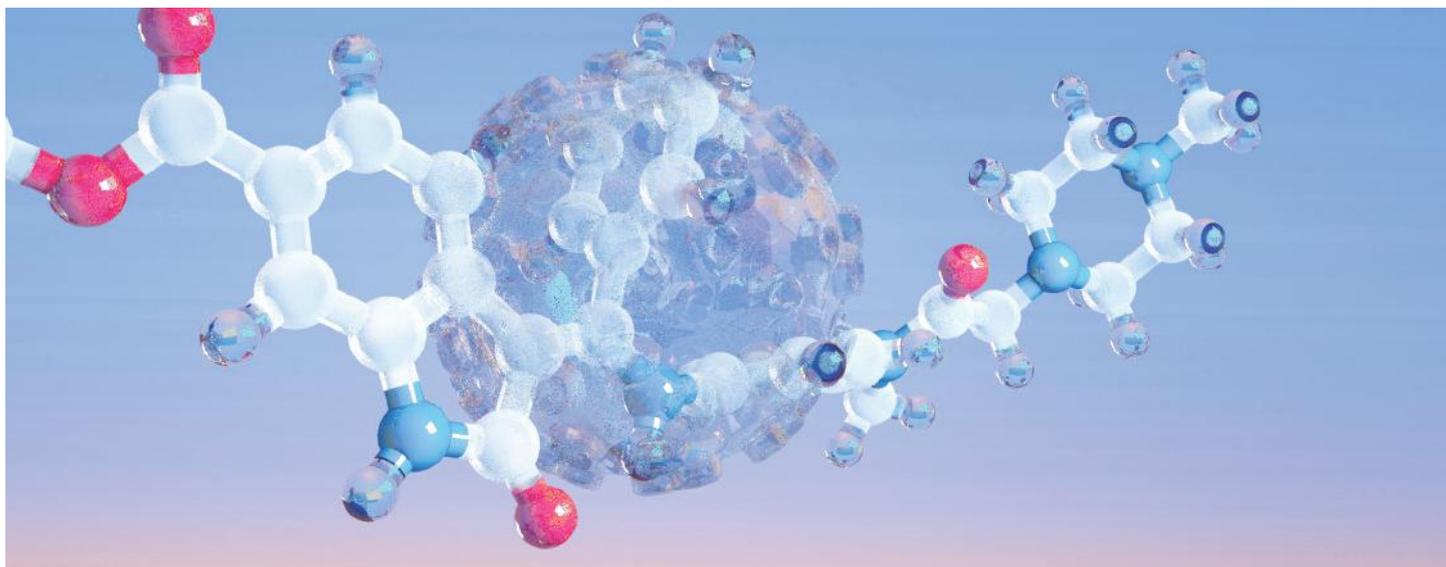
最强技术实力公司 TOP 10
相信AI的力量「AI 中国」机器之心 2021 年度评选

本榜单主要关注企业的技术实力、对新一代人工智能技术的研发布局与技术储备。入选企业具有强大的技术实力，重视对新一代人工智能技术的研发布局，在研发人员、论文发布、专利储备上拥有强大的储备，并在相应人工智能技术细分领域中处于顶尖水平。
(按拼音首字母顺序排列)

微软亚洲研究院

中国科学家利用计算生物学开展新冠病毒系列研究 揭示病毒侵染机理

来源: DeepTech 深科技



近年来, 微软亚洲研究院在新冠病毒方面进行了一系列研究。

2021 年, 微软亚洲研究院与清华大学在 Advanced Theory and Simulations 上共同发表了封面文章《通过分子动力学模拟探索 SARS-CoV-2 刺突蛋白 NTD 的调控功能》(Exploring the Regulatory Function of the N-terminal Domain of SARS-CoV-2 Spike Protein Through Molecular Dynamics Simulation)。

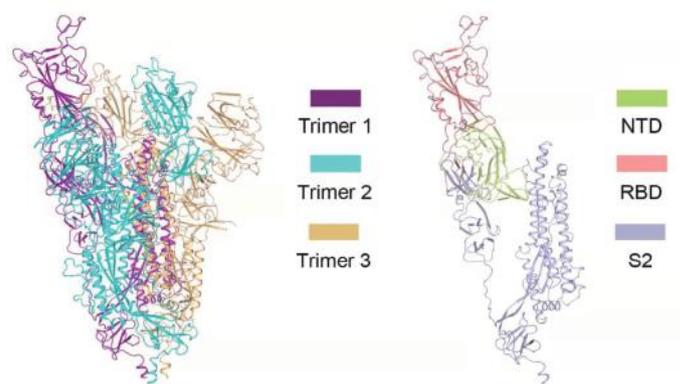
这是微软亚洲研究院和清华大学生命科学学院龚海鹏教授团队共同开展的一项计算生物学研究, 主要通过分子动力学模拟、研究新冠病毒侵染人类机理中 N 端结构域 NTD (N-Terminal Domain) 的作用。

起感染调控作用的 NTD

据了解, 在新冠病毒表面有一层 S 蛋白, 只有在被人体的受体蛋白 ACE 2 识别以后, 病毒才能进入人体。两者结合能力越强, 就越容易造成感染。S 蛋白的形态有点像字母“Y”, 向下的一竖即 S2 区域嵌入在病毒表面, 起到固定的作用。伸出来的两个“枝丫”分别叫作 NTD 和 RBD (受体结合域, Receptor Binding Domain)。

在该研究之前, 人们已经通过实验知道 RBD 可以直接起到一个识别人体受体蛋白的作用, 而 NTD 的具体作用还不十分清晰。

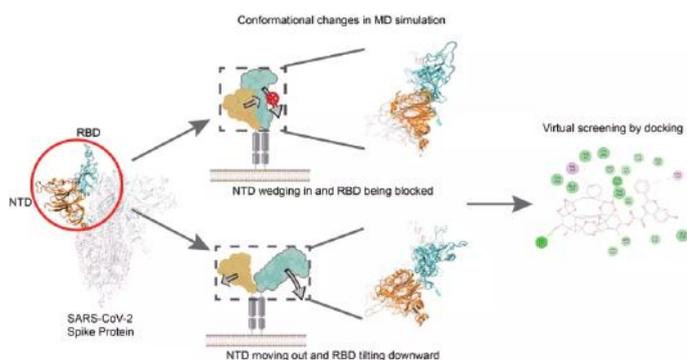
微软亚洲研究院和龚海鹏教授实验室合作搭建了一个有百万个原子的稳定模拟体系, 在进行了数十亿步 (1 步是 1 飞秒, 即一千万亿分之一秒) 的分子动力学模拟之后, 提出了一个“楔形”模型假说, 简单来说, 可以把 NTD 看作是安插在了 RBD 和“Y”字底部的一个“楔子”。



新冠病毒 S 蛋白结构 (左图为 S 蛋白三聚体结构, 由三条链组成; 右图为 S 蛋白单体结构, 由 NTD、RBD 以及 S2 组成)

研究发现, RBD 有 down 和 up 两种状态, 只有 up 状态才有可能与人的受体蛋白做识别。该项研究发现, RBD 倾向于从 up 的状态转变为 down 的状态, 而当 NTD 像楔子一样塞入 RBD 下

端时，阻止了 RBD 的这一倾向，即 NTD 使得 RBD 保持 up 状态，进一步让其拥有了感染的能力。可以说，NTD 在病毒的侵染过程中起到了一个调控作用，调节 RBD 的形态，从而改变了病毒感染人体的倾向。



NTD 在 SARS-CoV-2 的 S 蛋白构象变化中发挥调控功能示意图

由此可以预见，NTD 调控 RBD 的界面可能成为潜在的药物靶点。若能设计一个药物分子可以把 NTD 向外拉出，也就是把这个“楔子”拔出来，那 RBD 就倾向于 down 的状态，变成没有感染能力的构象，病毒感染人的能力就会减小很多。

“我们这项工作在世界范围内首次提出了 NTD 在病毒侵染过程中的调控模型。此外，我们根据这个潜在的药物靶点，设计了药物的虚拟筛选算法，并对新冠病毒的一些潜在性药物做了一定筛选，为新冠病毒药物研发提供了一定的参考价值。”微软亚洲研究院主管研究员王童说道。

新模型揭秘奥密克戎强感染性的原因

近期，微软亚洲研究院和清华大学生命科学学院王新泉教授以及该校医学院张林琦教授课题组开展合作，在奥密克戎变异株强传染性的机理解释方面有了新的突破，其论文已被生物学领域的顶级期刊《Cell Research》接收。

据了解，在此次合作中，王新泉教授课题组率先解析了奥密克戎高分辨率的晶体结构，为奥密克戎感染机理研究奠定了基础。相比于静态的晶体结构，王童团队利用分子动力学模拟从“动态视角”模拟和分析奥密克戎结构变异及其侵染机理。

研究中，微软亚洲研究院的研究员们先构建了两个模拟体系，分别以原始的新冠病毒结构和奥密克戎的结构作为起始结构，再对两个结构平行地进行长时间、数亿步的分子动力学模拟，模拟出的是原子级别的运动过程，从而观察病毒在人体中的真实变化。

“对分子运动进行模拟的方法可以分成两种，经典的分子动力学模拟和基于第一性的量子模拟，”王童说，“对于研究新冠病毒中蛋白大分子、大尺度的构象变化运动，经典动力学模拟方法是一种更适合的手段。”

据王童介绍，经典模拟适用于大体系，像蛋白质成千上万的原子，比如此次模拟体系就包含上百万的原子。从时间上来说，经典模拟做的都是数亿到数十亿步的模拟。相比之下，量子模拟相对准确，但计算过程的耗时使其只能适用于非常小的体系，比如对只有十余个原子的体系计算性质或进行短暂的模拟。

构建了模拟体系之后，微软亚洲研究院利用自研算法，分析了原始新冠病毒和奥密克戎感染人体的能力，以及它们结构上非常细微的差异。

值得注意的是，研究过程中产生的数据量极其庞大，整个模拟下来，至少有数亿个结构，产生了数亿到数十亿帧的结构变化。

为此，研究人员创造性地提出一种全新的马尔科夫模型算法对数据进行分析。其对分子动力学模拟的数据分析有着更强的表征能力，可以更加真实地模拟出蛋白的动态变化过程。

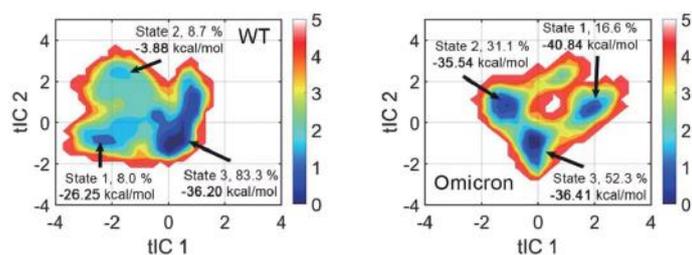
在对海量的数据进行筛选和分析后，研究人员将数十亿的结构聚合成几类代表性的结构，再去分析它们之间的区别和如何相互转换，以及与人受体蛋白的结合能力，从而更好地理解奥密克戎感染的分子机理。

这些聚合出来的代表性构象被称为“亚状态”，研究人员对其做了定性的结合自由能计算，以此在一定程度上反映病毒的感染能力。具体来说，在研究中，聚类以后的奥密克戎和原始病毒都有三个亚状态。由于奥密克戎本身只是原始病毒的变体，所以它们各自占比最高的一个亚态，无论是从结构上，还是和人的受体蛋白结合的能力上来说都是非常类似的。

由此可见，奥密克戎结构的关键不同之处在其他两个亚状态上。相比原始病毒的两个亚状态，奥密克戎的两个亚状态和人的受体蛋白 ACE2 有更多的相互作用，结合能力更强（结合自由能更低），因此可能导致了更强的感染能力。

同时研究还发现，奥密克戎的三个亚状态之间相互转换非常快，很容易从类似于原始病毒的主亚状态，转换成结合能力更强的另外两个亚状态。这也从动态视角解释了奥密克戎感染性这么强的一个原因。

在最新有关奥密克戎的研究工作中，王童表示，“这项研究在世界上首次使用‘干湿结合’的方式和动态视角，从结构生物学和计算生物学两个角度出发，提出了奥密克戎感染性强的分子机理。”



分子动力学模拟揭示奥密克戎的状态变化及感染机理（左图为新冠病毒 S 蛋白，右图为奥密克戎变异株的 S 蛋白，其中两个体系的 State3 基本一致且为主要构象，奥密克戎的另外两个亚态结合自由能明显高于新冠病毒的两个亚态）



微软亚洲研究院主管研究员王童

“干湿结合”促进科学研究

上面提到的“干湿结合”，也就是“干实验”（计算实验）和“湿实验”（传统科学实验）的结合。

据了解，“干实验”可以对“湿实验”现有的一些结论做验证和补充，还能做一些“湿实验”做不了的或者是需要花很多时间、人力去做的研究。

比如，在奥密克戎的研究中，若在“湿实验”实验室里一个个去分析突变位点结构的变化，可能需要数月时间，在疫情如此严峻的情况下，显然不利于疫情防控工作。而使用计算的手段来模拟这种突变造成的影响，可能只需要一周到两周的时间。

另外，“湿实验”在前期很多时候都是探索性质的，甚至在研究方向上是非常迷茫的。这时“干实验”可以在一定程度上引导“湿实验”的实验设计，帮助其找到一个更容易成功的方法。

值得注意的是，两者并不是互相取代，而是同等重要、相互指导、相辅相成的。“我们在和研究伙伴合作的过程中就有这样的感觉，我们认为干湿结合会对整个科学研究起到一个螺旋式上升、波浪式前进的推动作用。”王童说道。

他还补充说，在新冠病毒的研究方面，与清华大学多位老师的合作是一个非常好的起点。微软亚洲研究院非常希望能和海内外高校、科研院所的专家学者们有更多的深入合作和跨领域交流。

据了解，微软亚洲研究院后续还会就新冠病毒为何在整个冠状病毒进化过程中感染性如此强的原因方面开展相关研究。

跨学科发展对人才提出了更高要求

最后，关于计算生物学等跨学科领域对人才的要求和发展趋势，王童也发表了自己的看法。

他认为跨学科发展主要是围绕一个关键词，就是“交叉”。交叉不是简单的“加法”，而是“乘法”。两个学科之间要发生一定的化学反应和深度的融合。

拿计算生物学来说，其本质上就是计算机科学与生物学的交叉，需要结合两个领域的知识，并真正消化理解，才能在计算生物学领域产生真正的学术影响力，解决真正的科学问题。

而这也对人才的培养或者说对人才的需求，提出了一个更高的要求。如果研究人员只掌握某一方面的知识，对另一学科的认识不够深入，只停留在简单的工具使用层面，那么研究工作就只能浮于表面，也很容易造成计算参数设置有误等问题，导致最终计算结果与实验结果完全不一样的情况。

未来，各学科之间的交叉融合一定会越来越多，在具体的科研工作中，不同背景的人才在合作交流过程中，需要更多倾听对方领域的知识和诉求，只有建立在互信、开放心态的基础上，才能推动跨学科间的研究持续向前发展。



周礼栋

微软亚洲研究院院长

“二十多年来，微软亚洲研究院始终秉承开放、积极的心态，致力于打造自由、平等、可持续的科研协作环境，让分工、协调、合作链环上的每个人都成为新的发现与贡献的核心主体，为各种创造性想法的星星之火提供形成燎原之势的催化剂。

一个创新型组织的成长是不断拓展视野并承担更大社会责任的过程。微软亚洲研究院从创立伊始就持续与国内外计算机科研机构展开深度合作，携手进步，共同发展。在面对当下可持续发展、碳中和、医疗健康等人类社会亟待解决的关键问题时，微软亚洲研究院将守正创新，践行所有有利于激发创新力的原则，大胆接受和改造各种新的范式，与各界伙伴共同推动计算技术的跨界融合发展。”

关于微软亚洲研究院

微软亚洲研究院是微软公司在亚太地区设立的研究机构，也是微软在美国本土以外规模最大的一个。从1998年建院至今，通过从世界各地吸纳而来的专家学者们的鼎力合作，微软亚洲研究院已经发展成为世界一流的计算机基础及应用研究机构，致力于推动整个计算机科学领域的前沿技术发展，并将最新研究成果快速转化到微软全球及中国本地的关键产品中，帮助消费者改善计算体验。同时，微软亚洲研究院着眼于下一代革命性技术的研究，助力公司实现长远发展战略和对未来计算的美好构想。

微软亚洲研究院一直努力营造奋发、进取的科研环境，倡导对技术进步怀有远大抱负，推崇大胆创新以及富于冒险的极客创新精神。研究院鼓励研究人员树立长远眼光，加强与外界的交流，在第一时间接触世界领先的研究成果，对行业发展、技术趋势、及消费者需求高度敏感、迅速反应。微软亚洲研究院强调成员之间的相互信赖、相互尊重与开放合作，并承诺与高校和科研机构开展持久而有效的合作，促进协作、激发创新、推进教育，变梦想为现实！

微软研究院全球布局





微信



知乎



电话：86-10-59178888

网址：<http://www.msra.cn/>

微博：<http://t.sina.com.cn/msra>