微软亚洲研究院院刊

Matrix

NO.63
2022年 10 - 12 月

人工智能开启甲骨文整理研究新范式

科学匠人 | 对话邱锂力：一起探索未知的科技之美

2022“微软学者”奖学金获奖名单公布

01 焦点

02 前沿求索

科研第一线

03 文化故事

04 观点

05 媒体报道

2022 微软研究峰会在线举行

扫描二维码查看回放视频

当前，我们正在经历一波又一波计算机技术的突破，这些突破几乎改变了我们生活的方方面面。人工智能让我们的开发和创造方式产生了变革，人类语言技术彻底改变了医疗专业人员的工作流程，深度学习加速了我们理解和预测从原子到星系规模的自然现象的能力。云计算的基础也在历经着一场彻头彻尾的重塑。

要让这些新的技术突破对社会发展有所裨益，就需要全球科研界以一种全新的方式连接到一起。从高度理论到即时可用，发明和创新的活力越来越多地体现在传统研究学科之间的交叉点上。要确保科技的持续发展可以让每一个人受益，需要创造新技术和使用新技术来改善生活的群体充分沟通、合作，并且共同创新。

2022 微软研究峰会于 10 月 18 日至 20 日在线举行，来自全球的科研人员汇聚于此，共同探索新兴研究将如何更好地应对社会挑战，并在未来对我们的生活产生重大影响。在为期三天的微软研究峰会中，每天都以一个主题演讲开启并展开深入讨论，包括探讨深度学习对科学发现的潜在影响；如何利用技术使医疗更精准、更普惠；基础技术的发展如何使未来的云计算成为可能；从更高效、适应性更强的人工智能，到赋能人类创造力和助力可持续社会发展的技术。

2022 “微软学者”奖学金获奖名单公布

扫描二维码查看文章

10 月 6 日，2022 年“微软学者”奖学金获得者名单正式出炉！经过激烈的角逐，来自亚太地区的 12 名优秀博士生最终被授予 2022 年“微软学者”称号，另有 21 名博士生获得提名奖。

2022 年 “微软学者” 奖学金共吸引了来自亚太地区顶尖研究型大学及机构的近两百名优秀博士生申请，申请者的研究领域广泛分布于计算科学、硬件与软件系统、人类与机器智能，及感知、识别与交互等领域。获奖者们具有广阔的视野，致力于做出有价值的科研工作，并且期待以科研带来切实的社会价值，造福世界。

“微软学者” 奖学金是微软亚洲研究院 1999 年启动的一项面向亚太地区计算机科学以及相关专业和交叉学科的优秀博士生的项目。该奖学金项目旨在发掘、支持和鼓励优秀的、有潜力的低年级博士生更好地开展研究工作。截至 2022 年，已有数百名优秀博士生获得 “微软学者” 称号。其中多位 “微软学者” 已成为学术界中流砥柱或耀眼新星，也有多位 “微软学者” 成为工业界翘楚。

人工智能开启甲骨文整理研究新范式

在甲骨学研究中，甲骨“校重”整理是一项费事费力但又极其重要的基础性研究工作。微软亚洲研究院与首都师范大学甲骨文研究中心莫伯峰教授团队合作开发的甲骨文校重助手 Diviner，第一次将自监督 AI 模型引入到甲骨文“校重”工作中，并取得数百项新成果，为甲骨文整理领域开创了人工智能与人类专家协作（AI+HI）的全新研究范式。

入选《世界记忆名录》的甲骨文，是迄今为止中国发现的年代最早的成熟文字系统，对中国历史乃至世界文化的发展研究具有非凡意义。有人曾说“东周之前无信史”，因为《春秋》一书记录了 2000 多年前的东周历史，而之前的商文明曾被认为是传说，直至甲骨文被发现，才有力地证明了殷商王朝的存在，把中国信史向上推进了约 1000 年。

从甲骨文首次被发现至今，出土的甲骨实物约有十五万片。因为收藏、流转的缘故，大部分的甲骨都留下了多张拓本图像，被称为“重片”。甲骨重片数量繁多，效果互有参差，对其整理成为了一项重要的基础性研究工作，称作“校重”。然而，人工校重只能一一对照，费时费力，是甲骨文研究的一大痛点。正如《甲骨文合集补编》“前言”中所述：“这种对重、选片的工作，其烦琐、费工是局外人难以想象的。”

近期，微软亚洲研究院主管研究员武智融与首都师范大学甲骨文研究中心莫伯峰教授团队合作，提出了基于自监督学习的甲骨文校重助手 Diviner，大幅提升了甲骨文校重工作的效率。系统穷尽比对了 18 万幅拓本，辅助甲骨学家在上百个甲骨文数据库中发现了大量甲骨重片，不仅复现了专家过去所发现的数万组重片，而且经过初步整理，已发现了三百多组未被前人发现的校重新成果。这项研究为甲骨文整理领域开创了人工智能与人类专家协作（AI+HI）的全新研究范式。本项目全面成果的甲骨学解读已发布于中国社科院先秦史研究室网站 https://www.xianqin.org/blog/archives/17264.html。

小知识：为什么同一片甲骨的不同拓本有时会差异巨大?

两个原因导致。一是早期制作拓本时，只拓下了有字的部分，而后来制作拓本，则将所有部分全部拓下，这就会导致早期拓本显得更小。二是早期甲骨比较完整，但随着时间推移甲骨出现了破损，后来所做的拓本就不再完整。也正是因为这些原因，使得甲骨校重工作变得愈发困难。

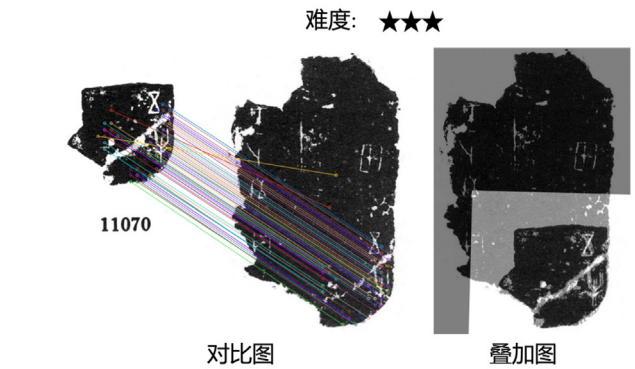
小知识：只有不同时期的拓本间会出现重片现象吗?

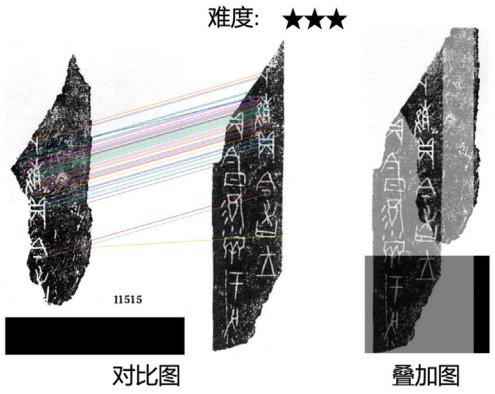
并非如此。在早期甲骨著录书中就已出现，同一片甲骨的拓本被同一本书反复收录的情况。所以校重工作并不都是在不同著录书之间进行，在同一本书内部也是需要的。

AI 在甲骨文“校重”中令人惊喜的新发现

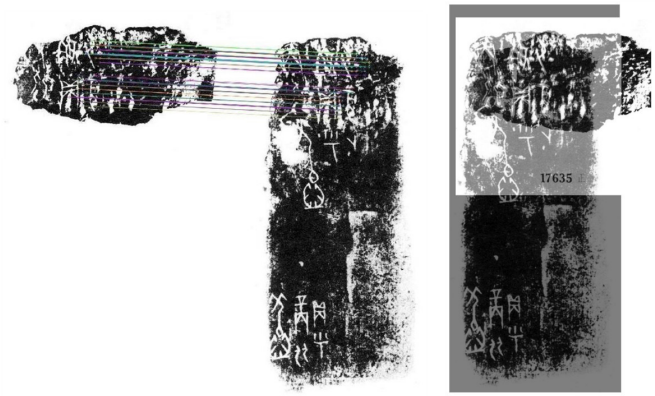
“校重”是甲骨学领域的一个老题目，此前已有很多甲骨学家为这项工作倾注了大量心血。比如甲骨文领域最重要的两部著录书《合集》、《合补》，在编著过程中花费大量功夫的工作就是校重。理论上来说，完成一张甲骨拓本的校重工作，应该将它与其余所有甲骨拓本逐一比照，才能确保没有遗漏。尽管可以利用文字信息和分类方法缩小对比范围，但对于甲骨学家而言，这仍然是一项十分艰巨的工作，且难以保证全面性和准确性。

对比和处理海量数据，并从中挖掘有用信息正是 AI 的专长。大规模的校重，穷尽性的比对，都难不倒校重助手 Diviner。接下来就让我们一起看看 Diviner 的效果。你也可以测试一下自己能否发现其中的异同?



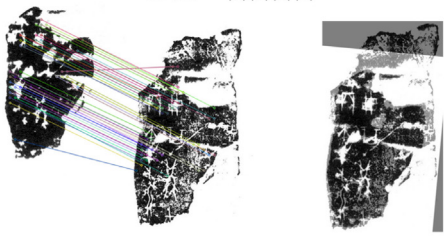


难度: ★★★★



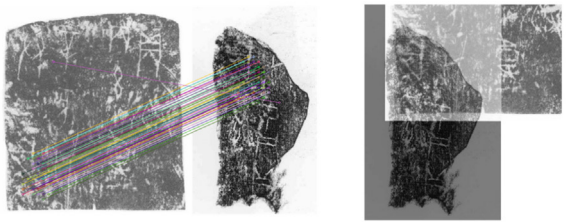
对比图 叠加图

难度: ★★★★



对比图 叠加图

难度: ★★★★★



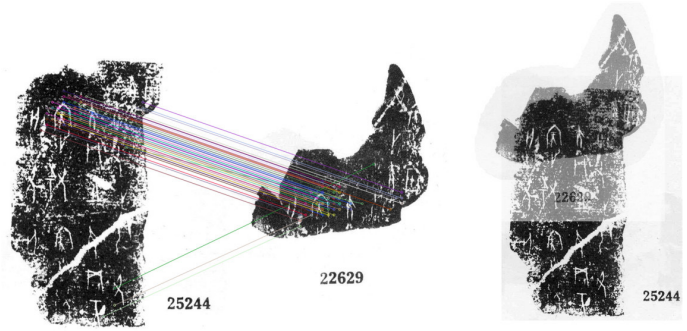
对比图 叠加图

这些校重结果对甲骨文研究有什么作用?

作为三千年前古人留下的一份礼物，每一片甲骨都弥足珍贵。但目前甲骨的研究主要依靠拓本图像，而非甲骨实物，所以拓本就是甲骨研究的根本出发点。很多时候一个字形、一条卜辞的清楚认知，就来自更全、更清的拓本材料。所以甲骨学家从不放过哪怕只有一个字的拓本。

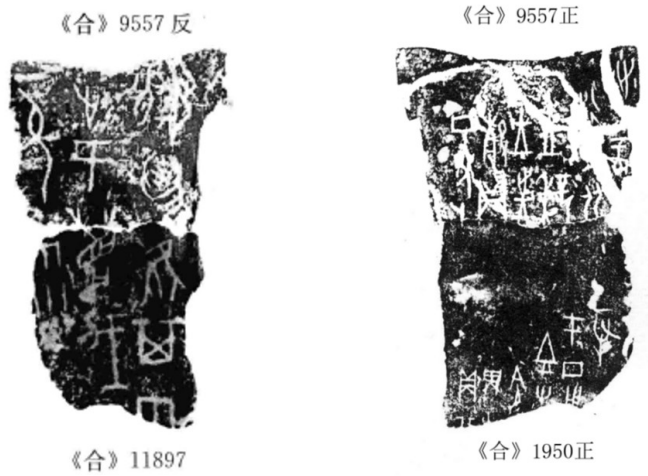
本次校重助手 Diviner 就提供了一批更全、更清晰的甲骨图像，更有不少堪称惊喜的新发现：

"重而不同" 的新图像。左侧是时间较早、未拓全的甲骨拓本。右侧是时间较晚的拓本，甲骨残破只余下一部分，尽管拓全了但很不完整。通过将两个拓本重叠，获得了一张最完整的甲骨图像，特别是右上部分的一段甲骨文字的完整展现，为甲骨文研究直接提供了一条新材料。



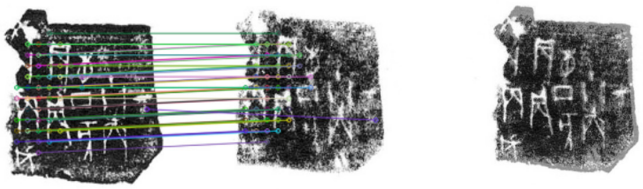
对比图 叠加图

"有里有面" 的新图像。有些甲骨正反两面皆有文字，但有时只有一面留下了拓本。比如马保春先生曾发现这两版反面拓本可以缀合在一起。但其中一片的正面图像一直没有找到。Diviner 校重发现了下面一片甲骨的正反完整拓本，正面缀合复原的图像也终于得以呈现。

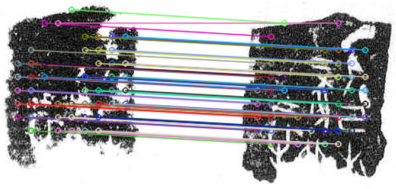


原反面缀合 正面新缀合

从模糊到清晰。由于有些甲骨拓本质量不高，上面文字难以辨认，给甲骨学家带来很多困扰。比如这几组重片，模糊拓本上的文字让人难以辨认，直到这次发现了清晰的重片，才把过去的很多疑惑解决了。



对比图 叠加图



对比图

叠加图

从“重片”到“缀合”。左侧图像是过去由张宇卫先生缀合在一起的两片甲骨。通过 Diviner 发现，下部拓本还有一片更完整的重片。如此，两片甲骨的缀合就扩展成了三片甲骨的缀合。

原缀合　　新缀合

Diviner 在甲骨校重工作中的出色表现和展现出的巨大潜力得到了很多甲骨学家的认可。复旦大学出土文献与古文字研究中心研究员蒋玉斌认为：“甲骨校重与指出互见，是甲骨学重要的基础性工作。同一甲骨片，可能经过多次著录，各版本有早有晚，清晰度、完整度存在差别，需要加以关联、比对、研判。过去，这种工作完全靠学者凭经验、记忆零星举列，虽颇有得，但总体上耗时费力，也仍有大量未能指出的重出、互见现象。莫伯峰教授团队与微软亚洲研究院合作开发的人工智能甲骨文校重助手 Diviner，实现了大范围的校重，效率高，成果多，令人振奋。我坚信，在甲骨校重与指出互见方面，校重助手 Diviner 已经远胜人力，今后此项工作的大规模开展，或将完全由校重助手 Diviner 这样的工具取代。近年，有多支学术团队致力于甲骨文等古文字研究与人工智能的融合创新，先进的技术手段将为古老文字的研究插上腾飞的翅膀。但人工智能助力古文字研究的着力点在哪里，是首先要解决的问题。校重助手 Diviner 很好地契合了甲骨文研究的需要与人工智能的专长，功效显著，成果突出，我认为是人工智能辅助甲骨文研究的成功典范。”

小知识：什么是甲骨缀合?

甲骨缀合是甲骨文研究中另一项重要的整理研究工作。由于材质坚硬容易破碎，原本完整的甲骨很多都碎裂为多个碎片，只有将它们恢复原样才具有更大的研究价值，这种复原工作就是甲骨缀合。

小知识：甲骨文考释有多难?

甲骨文中已知不重复的单字数量约为四千五百个，在过去的 120 多年中，甲骨学家前赴后继也只破译了一千个左右，大部分甲骨文字仍待破译。中国文字博物馆曾在 2017 年推出了一项甲骨文考释竞赛，单字破解悬赏 10 万元。竞赛推出以来，只有来自复旦大学出土文献与古文字研究中心研究员蒋玉斌和清华大学出土文献研究与保护中心教授王子杨成功拿到过这笔奖金。与西方表音体系的古文字相比，甲骨文字的破解难度无疑要大得多。

自监督学习首次在甲骨文中应用，AI 模型泛化性显著

校重助手 Diviner 能有如此出众的效果，技术上是如何实现的？有哪些创新之处？

近两年，不依赖人工标注数据的自监督学习是 AI 研究的热门方向，但很多前沿技术仍停留在研究阶段。Diviner 不仅第一次将自监督 AI 模型引入到甲骨文“校重”工作，也是自监督 AI 模型在真实场景中的一次成功应用。

“尽管自监督研究热度很高，但是很多问题最终还是要通过人工数据标注来解决。我们一直希望使用完全无标注的数据进行自监督学习，甚至是人工根本上无法标注的数据。”微软亚洲研究院主管研究员武智融说，“甲骨校重需要两两比对十八万张数据库中的所有拓片，这为基于完全无标注数据的自监督学习模型应用提供了一个绝佳的落地场景。”

计算两张拓片的视觉相似度，通常的方法会从全局特征出发。然而，在甲骨文研究中，即使是重片，外观上也可能有很大差异，这是由于拓印范围、拓印方式、磨损等多方面原因造成的。考虑

到一块完整的甲骨可能会碎裂成多片，校重时经常需要从大骨片中找出小骨片。因此，基于全局外观表示的传统方法并不能很好地发挥作用。面对这一挑战，研究员想到了甲骨拓片的特性，因其是从同一块甲骨而来，重片之间存在着精确的点与点的对应关系。基于这一特性，校重助手 Diviner 从局部寻找匹配关系，再拓展到全局。

局部匹配。Diviner 使用的局部描述符（local descriptor）是经过自监督训练的深度神经网络。模型应用了对比学习的自监督技术，使用图像增强，让特征在训练时不受甲骨拓片上清晰度、对比度、噪音、旋转等因素的影响。在甲骨图像上训练的局部描述符能够检测和匹配局部块之间的关键点，并进行点对点匹配。

全局优化。基于密集的点与点的匹配结果，通过使用鲁棒的优化算法 RANSAC 估计全局的几何仿射变换。仿射变换允许模型在内容重复的情况下拼合或拼接已有图像。这种局部到全局的方法对检测大量的甲骨碎片至关重要。

Diviner 模型一个特点是具有强大的泛化能力，这归功于其自监督学习的匹配算法。模型通过图像增强技术模拟同一块甲骨在不同时期制作成拓片或者因年深日久造成的图像变化，例如磨损、模糊等。在大规模无标注数据上获取的密集的自我监督，远远比稀疏的基于整体的人工监督更有效。

Diviner 模型另一个特点在于能精确的预测出重片之间点对点的对应关系，并将重片拼合或拼接在一起。这种可以被专家快速解读的结果大大方便了人类与人工智能的协同合作。对于甲骨文这种冷门绝学，人机合作尤为重要。在校重结果中，专家可以看到局部匹配细节和重叠图，极大地帮助并加速了他们验证的过程。

“过去的甲骨校重工作中，对拓面差异较大的不同拓本之间的认同存在现实困难。甲骨文校重助手 Diviner，既不受文字信息的限制，也不受图像数量的限制，直接运用图像比对就可以完成精准的图像校重，并取得了显著的成果。可以预期，随着 Diviner 模型功能的不断完善，甲骨学界一定会取得更大、更多的科研成果。”清华大学出土文献研究与保护中心教授王子杨如此评价 Diviner。

小知识：甲骨文图像的著录方式有哪些?

甲骨文有三种主要的著录方式。一是拓本，这是甲骨最主要的著录方式，应用了中国传统的墨拓技术；二是照片，早期利用照片进行甲骨著录的情况较少，近年来已经成为甲骨著录的主要方式；三是摹本，采用目视手绘的方式临摹甲骨文字，主要是著录那些没有条件做拓本，也没有条件拍照的甲骨。此外，甲骨 3D 成像技术近年来也开始进行实验。Diviner 模型现在主要针对拓本进行校重，今后将尝试扩展到更广阔的范围。

“AI+HI”为古文化研究打开新大门

“甲骨学是一个系统性的科学，一方面它是一种语言文字研究资料，另一方面它是一种历史研究资料，其研究涉及方方面面，研究者需要了解文字在古代的形、音、义等等，因此我们解决一个问题也要从不同维度探讨。此次与微软亚洲研究院的合作只是甲骨文和人工智能交叉研究的一个小序幕，推开了甲骨学研究的一扇新大门，为后续的研究起到了示范作用。未来，人工智能与古文字研究的结合将具有更广阔的前景。”莫伯峰教授表示。

首都师范大学甲骨文研究中心莫伯峰教授（左）与微软亚洲研究院主管研究员武智融（右）

“我们很高兴看到人工智能模型 Diviner 能够为甲骨学专家节省用于甲骨文数据整理的时间，让他们更专注于其他方面的研究。甲骨文是兼具象形图像属性和文字属性的神秘语言，多模态的人工智能在甲骨文研究上有着广阔天地。未来，我们希望能够与甲骨文专家一起探索更多有趣的课题。”武智融表示。

计算机图形图像领域知名学者、微软亚洲研究院常务副院长郭百宁表示，“甲骨文作为世界文化的瑰宝，其研究已经发展成为国际性的学术课题。多年来，微软亚洲研究院一直致力于将最前沿的计算机技术应用于文化遗产保护与传承等具有社会意义的研究中，并取得了诸多成果。我们希望可以与更多研究机构、研究学者共同合作，为推进世界文化、历史的保护和传承贡献一份力量。”

预见 AIOps：让云计算更自主、更前瞻、更全面，也更易于管理

“建立一个可供数百万人每天使用，但只需一名兼职人员管理和维护的系统。”这是吉姆·格雷（Jim Gray）在 1999 年获得图灵奖时对无故障服务器系统的畅想。他设想了一个自管理的“空中服务器”，可以存储大量数据，并按需刷新或下载数据。如今，随着人工智能（AI）、机器学习（ML）、云计算的出现和快速发展，以及微软对云智能 /AIOps（智能运维）的开发，我们比以往任何时候都更接近这一愿景，并有望超越这一愿景。

在过去 15 年间，软件行业最重要的范式转变是向云计算的迁移，这一转变给企业、社会和人们的生活都创造了前所未有的数字化转型机遇，并带来了巨大的利益。如今，云计算已经成为全球基础设施的一部分，因此，云计算的服务质量，包括可用性、可靠性、性能、效率、安全性和可持续性等也变得愈发重要。但云计算平台的分布式特性以及超大规模和高度复杂性的特点，贯穿于存储、网络、计算和其他各个方面，给系统的设计、构建和运维带来了巨大挑战。

什么是云智能 /AIOps？

云智能 /AIOps（简称“AIOps”）旨在通过创新的人工智能（AI）或机器学习（ML）技术，帮助人们有效且高效地设计、构建和运营大规模的复杂云平台及服务。AIOps 有三大支柱，每个支柱都有其各自的目标：

AI for System：让智能内置于云系统，实现更少的人工干预，并同时保障系统的高可靠、高效率、自控制和自适应。

AI for Customer：利用 AI/ML 创造无与伦比的服务体验，实现非凡的用户满意度。

AI for DevOps：将 AI/ML 注入软件开发的全生命周期，以提高开发的质量和效率。

AIOps 研究始于软件分析

全球咨询分析机构 Gartner 在 2017 年首次创造了 AIOps（人工智能 IT 运维）一词。Gartner 称，AIOps 是将机器学习和数据科学应用于 IT 运维问题上。Gartner 的 AIOps 概念仅聚焦开发运维（DevOps），而微软的云智能 /AIOps 的研究范围更为宽泛，还包含 AI for System 和 AI for Customer 两个方面。

微软在云智能 /AIOps 方面的研究可追溯到 2009 年微软亚洲研究院提出的“软件分析（Software Analytics）”研究，这项研究旨在让软件从业者能够通过探索和分析软件相关数据，获取深入且具有指导性的洞见，并应用于与软件和服务相关的数据驱动任务。2014 年，微软研究员们开始将软件分析的研究重点放在云计算上，并将新的研究主题命名为云智能（Cloud Intelligence）。回看过去，软件分析主要是关于软件行业自身数字化转型研究，例如让从业者利用数据驱动的方法和技术来开发软件、运维软件系统以及提升用户体验。而当下在进行云智能研究时，也同样以数字化转型的观念来看待云计算平台的发展，就是将最先进的 AI 技术内嵌于云计算平台，以服务于云系统、开发运维人员、以及云客户，从而更好地推动云计算平台的智能化发展。

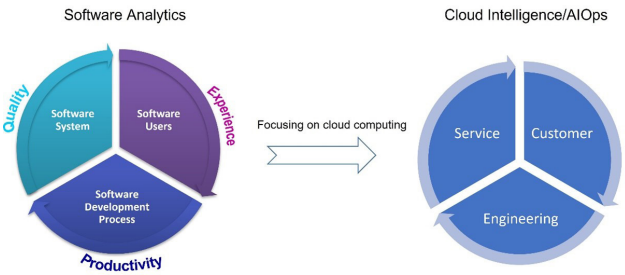


图 1：从软件分析到云智能 /AIOps

AIOps 中的主要研究领域：检测、诊断、预测和优化

在 AIOps 的三大支柱中，每一个都有许多场景。例如，在 AI for System 方面，包括为高效和可持续服务进行前瞻性预测、监测服务的健康状况和及时发现系统健康问题；在 AI for DevOps 方面，确保代码质量并防止有缺陷的软件被部署到线上；在 AI for Customer 方面，如何有效提升客户体验等。

在所有这些场景中，有四个主要的问题类别，它们一起构成了 AIOps 的主要研究领域：检测、诊断、预测和优化。具体而言，“检测”的目的是及时地识别异常的系统行为或状态。“诊断”的目标是找出服务系统问题的原因并定位其根源所在。“预测”则旨在对系统行为、用户工作负载或 DevOps 活动等进行预判。最后，“优化”是找出最佳策略或决策，以全面优化与系统质量、客户体验和 DevOps 生产力相关的特定性能。

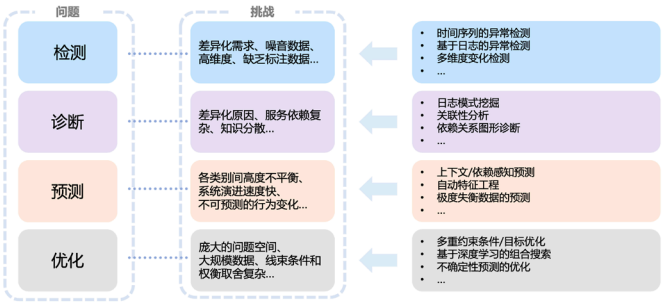


图 2：AIOps 的研究领域及面临的挑战

从图 2 中可以看到，每类问题都面临着各自不同的挑战。以“检测”为例：为了确保服务运行时状况良好，工程师必须不断监控各种指标并及时检测异常情况；在开发过程中，为了确保持续集成 / 持续发布（CI/CD）等业务实践的质量，工程师需要创建一些机制识别有缺陷的软件版本，并防止它们被部署到生产环境中。这两种情况都需要及时检测，在应用 AI/ML 方法时也存在着共同的挑战。例如，时间序列数据和日志数据是最常见的数据输入形式，然而它们通常是多维度的，数据中可能存在噪声，这都会对可靠的检测构成重大挑战。

微软研究院 AIOps 的愿景：更自主、更前瞻、更全面、更易管理

微软正在针对 AIOps 的每一类问题进行持续研究，目标是让整个云系统的每一层都变得更自主、更前瞻，更全面、更易于管理。

“让云系统更自主”：AIOps 致力于让云系统变得更加自动并更有自主性，最大限度地减少人工操作，降低维护成本，结合全局信息做出更合理的决策，从而避免系统问题对用户的影响。为此，要尽可能地实现 DevOps 自动化和自主化，包括开发、部署、监控和问题诊断。例如，安全部署的目的是及早发现有缺陷的软件版本，防止其部署到线上对用户造成重大影响。对于工程师而言，人工识别有缺陷的软件版本是非常耗时耗力的，因为异常行为有多种模式，随着时间的推移这些模式还会发生变化，而且并不是所有的异常行为都是由新版本引起的，这就可能会导致误报。

微软研究院的研究员们借助迁移学习和主动学习技术，开发了一套安全部署解决方案来克服上述挑战。这套解决方案已运行在微软 Azure 云计算平台中，在帮助识别缺陷版本方面非常有效，在 18 个月内，实现了 90% 以上的准确率和几乎 100% 的召回率。

相关论文：
An Intelligent, End-To-End Analytics Service for Safe Deployment in Large-Scale Cloud Infrastructure
https://www.microsoft.com/en-us/research/publication/an-intelligent-end-to-end-analytics-service-for-safe-deployment-in-large-scale-cloud-infrastructure/

AIOps 自动自主化的另一种方法是自动根因分析。为了缩短处理时间，工程师必须快速确定云系统故障发生的根源所在。然而，由于云系统结构的复杂性，故障告警通常只包含部分信息，并且一个故障可能同时触发多个服务和组件，因此工程师在采取有效措施之前，不得不花费大量的时间来诊断问题的根本原因。通过先进的对比挖掘算法（contrast-mining algorithms），微软研发了包括多层次故障定位（Hierarchy-aware Fault Localization）和故障影响范围评估（Outage-impact Scope）在内的故障自主诊断系统，在缩短响应时间的同时，提升了故障诊断任务的准确性。这些系统现已集成至微软 Azure 云计算平台和 Microsoft 365（M365）中，提高了工程师在云系统中快速准确处理故障的能力。

相关论文：
Fast Outage Analysis of Large-scale Production Clouds with Service Correlation Mining
https://www.microsoft.com/en-us/research/publication/fast-outage-analysis-of-large-scale-production-clouds-with-service-correlation-mining/

HALO : Hierarchy-aware Fault Localization for Cloud Systems
https://www.microsoft.com/en-us/research/publication/halo-hierarchy-aware-fault-localization-for-cloud-systems/

Onion : Identifying Incident-indicating Logs for Cloud Systems
https://www.microsoft.com/en-us/research/publication/onion-identifying-incident-indicating-logs-for-cloud-systems/

“让云系统更具前瞻性”：AIOps 通过引入“前瞻性设计”的概念，让云系统变得更具前瞻性。前瞻性系统设计是在传统系统中添加了基于机器学习的预测组件。预测系统通过对大量历史数据的学习获得预测模型，并结合当前的系统状态，以预测系统的未来状态。例如，预测某个服务器集群下一周的资源容量状态，磁盘是否会在未来几天内出现故障，或者未来一小时内需要创建的特定类型虚拟机的数量。

了解了未来的状态，就能够主动避免对用户的负面影响。例如，工程师可以把未来将会出故障的计算节点上的服务实时迁移到健康的计算节点上，以减少虚拟机的停机时间，或者在未来一小时预配置特定类型和数量的虚拟机，以减少配置虚拟机所需的等待时间。此外，AI/ML 技术还可以使系统随着时间的推移不断学习来调整以达到适应当前系统的最优决策。

作为前瞻性设计的范例之一，微软研究院的研究员们构建了一个名为 Narya 的系统，它能够主动处理潜在的硬件故障，以减少服务中断并把对用户的影响降至最低。运行在微软 Azure 云计算平台中的 Narya，可以对硬件故障进行预测，并使用增强学习算法来决定采取何种最优化的处理措施。

相关论文：
Correlation-Aware Heuristic Search for Intelligent Virtual Machine Provisioning in Cloud Systems
https://www.microsoft.com/en-us/research/uploads/prod/2020/12/AAAI21_Provisioning.pdf

Intelligent Virtual Machine Provisioning in Cloud Computing
https://www.microsoft.com/en-us/research/uploads/prod/2020/04/UAHS_IJCAI_2020_updated.pdf

Predictive and Adaptive Failure Mitigation to Avert Production Cloud VM Interruptions
https://www.microsoft.com/en-us/research/publication/predictive-and-adaptive-failure-mitigation-to-avert-production-cloud-vm-interruptions-2/

“让 AIOps 在云堆栈中的应用更全面”： AIOps 还能扩展到整个云堆栈，从底层的基础设施层（如网络和存储）到服务层（如调度器和数据库）再到应用层，从而让 AIOps 变得更全面。广泛应用 AIOps 的好处在于可显著提高整体诊断、优化和管理能力。

构建在 Azure 之上的微软服务被称为第一方（1P）服务。微软 1P 服务的范例包括 Office 365 等大规模的成熟服务、Teams 等相对较新但规模较大的服务，以及 Windows 365 Cloud PC 等即将推出的服务。1P 服务中的单个实体可以看到并控制云堆栈的各个层，通常会占用大量的资源，比如广域网（WAN）流量和计算资源等，是应用更全面的 AIOps 方法的重要场景。

作为将更加全面的 AIOps 方法应用于 1P 设置的示例，由 Azure、M365 和微软研究院联合开发的 OneCOGS 项目考虑了三类广泛的优化机会：

1. 使用跨层信号对用户及其工作负载进行建模，例如使用用户的消息活动与固定工作时间的对比来预测 Cloud PC 用户何时将处于活动状态，从而提高准确性，实现系统资源的合理分配。

2. 通过应用程序和基础架构联合优化，实现成本节约等好处。

3. 控制数据和配置的复杂性，实现 AIOps 的通用化。

用于云计算平台和 1P 服务的 AIOps 方法、技术及实践，同样适用于云堆栈上的第三方（3P）服务。若要实现这一目标，就需要进一步的研究和开发，让 AIOps 方法和技术变得更通用和更易适配。例如，在运行云服务时，检测多维数据中的异常以及随后的故障定位，就是常见的监控和诊断问题的方法。

基于 Azure 云和 M365 的实际需求，微软的研究员们提出了 AiDice 技术和 HALO 技术，前者可以自动检测多维时间序列中的异常，后者是一种层次感知方法，使用从云系统中收集的运行数据对故障组合进行自动定位。除了在 Azure 和 M365 中部署 AiDice 和 HALO 之外，研究员们还与产品团队合作，正在开发可供第三方服务使用的 AiDice 与 HALO AIOps 服务。

相关论文：
Efficient incident identification from multi-dimensional issue reports via meta-heuristic search
https://www.microsoft.com/en-us/research/publication/efficient-incident-identification-from-multi-dimensional-issue-reports-via-meta-heuristic-search/

“让云系统更易管理”： 通过引入分层自治的概念，AIOps 让云系统变得更易于管理。层级包括了从自动日常操作的顶层，到需要深厚的专业知识来应对罕见和复杂问题的底层。由于系统的复杂性，AI 驱动的自动系统管理通常无法处理所有不同类型的问题，因此，研究员们通过构建针对每一层的 AIOps 解决方案，让云平台更简便地管理复杂系统中不可避免的长尾罕见问题。此外，分层设计也确保了自主系统从开发阶段起，就能够评估确定性和风险，并在自动化故障或平台面临前所未有的情况时具有安全回滚的保障，例如 2020 年由新冠疫情导致的意外的需求增加。

分层自治的范例之一是微软的研究员们构建的安全的节点在线学习（Safe On-Node Learning，SOL），这是一个在顶层服务器节点上进行安全学习和驱动的框架。另一个范例是，研究员们正在探索如何预测运维人员在处理故障时应执行的命令，同时考虑在顶层自动化无法防止故障发生时，衡量与这些命令相关的确定性和风险。

“AIOps 已在 Azure 和 M365 中大显身手”： AIOps 是一种迅速兴起的技术趋势，也是结合系统、软件工程和 AI/ML 等领域的跨学科研究方向。通过多年的云智能研究，微软研究院在检测、诊断、预测和优化等方面积累了丰富的经验和成果。通过与 Azure 和 M365 团队的密切合作，相关研究成果也已经在云系统中进行部署，在改善 Azure 和 M365 的可靠性、性能和资源效率的同时，也提高了产品开发人员的生产力。此外，微软研究院正在与学术界和工业界的同行展开合作，推进 AIOps 的研究和应用实践。例如，在各方的共同努力下，以 AAAI 2020、ICSE 2021 和 MLSys 2022 等顶尖学术会议为依托，微软研究院的研究员们已经组织了三期 AIOps 研讨会。

展望未来，微软相信云智能 /AIOps 作为一个全新的创新维度，将发挥越来越重要的作用，使整个云系统变得更加自主、前瞻、全面和易于管理。云智能 /AIOps 终将助力实现人们对于云计算的未来愿景。

扫描二维码查看微软亚洲研究院常务副院长张冬梅介绍云智能 /AIOps 研究的主题演讲视频

微软 T-ULRv6：引领基础模型向多语言“大一统”迈进

近日，微软通用语言表示模型再创新佳绩。最新的 T-ULRv6 在谷歌 XTREME 和 GLUE 排行榜上摘得双榜冠军，证明单个多语言模型可以同时在英语和多语言理解任务上达到 SOTA 性能。这也是多语言理解模型首次在两个排行榜上同时夺魁，力压专用于英语或专用于多语言任务的模型，从而有助于消除“多语言诅咒”。

微软亚洲研究院自然语言计算组首席研究员韦福如表示，“T-ULRv6 是我们推进大规模预训练语言模型以及 AI 模型‘大一统（The Big Convergence）’研究的重要里程碑。我们第一次发现通过规模化预训练语言模型，可以让多语言基础模型在高资源（rich-resource）语言（例如英文）上，取得与专门为这些语言设计和训练的单语言预训练模型在对应语言的下游任务上一样好的效果。之前的研究曾表明多语言预训练模型在低资源（low-resource）语言的下游任务上有很大的性能提升并具有支持跨语言迁移的能力。这也说明未来可以专注于规模化多语言基础模型，并结合我们所推进的多模态基础模型大一统方面的研究（如 BEiT-3），为接下来推进多语言、多模态模型的统一提供经验与参考。”

基于“XY-LENT”的 T-ULRv6 XXL 模型是微软图灵团队和微软亚洲研究院通力合作的成果，其平均分比 XTREME 排行榜目前位居第二的模型高出 0.5 分，在 GLUE 排行榜上也占据首位。与此同时，微软必应搜索引擎庞大的索引库对于检索效率有非常高的要求。微软亚洲研究院提供的快速预训练蒸馏方法有效地将预训练大模型的索引能力迁移到轻量化模型中，在识别准确率上将现有模型提升了 14%，同时极大地优化了模型的计算效率，实现了百亿图片的快速推理。

Rank	Model	Participant	Affiliation	Attempt Date	Avg	Sentence-pair Classification	Structured Prediction	Question Answering	Sentence Retrieval
0		Human	-	-	93.3	95.1	97.0	87.8	-
1	Turing ULR v6	Alexander v-team	Microsoft	Sep 6, 2022	85.5	91.0	83.8	77.1	94.4
2	MShenNonG+TDT	Cloud Xiaowei AI	Tencent	May 29, 2022	85.0	90.4	83.1	76.3	94.4
3	Turing ULR v5	Alexander v-team	Microsoft	Nov 24, 2021	84.5	90.3	81.7	76.3	93.7
4	CoFe	HFL	iFLYTEK	Oct 26, 2021	84.1	90.1	81.4	75.0	94.2
5	InfoXLM-XFT	Noah's Ark Lab	Huawei	Oct 5, 2021	82.2	89.3	75.5	75.2	92.4
6	VECO + HICTL	AliceMind + MT	Alibaba	Sep 21, 2021	82.0	89.0	76.7	73.4	93.3
7	Ensemble-Distil-XFT (ED-XFT)	Huawei	Huawei Ireland Research Center	May 5, 2022	82.0	89.2	74.6	75.2	92.4
8	Polyglot	MLNLC	ByteDance	Apr 29, 2021	81.7	88.3	80.6	71.9	90.8
9	Unicoder + ZCode	MSRA + Cognition	Microsoft	Apr 26, 2021	81.6	88.4	76.2	72.5	93.7

图 1：T-ULRv6 XXL 位居 XTREME 排行榜首位

Rank	Name	Model	URL	Score
1	Microsoft Alexander v-team	Turing ULR v6		91.3
2	JDExplore d-team	Vega		91.3
3	Microsoft Alexander v-team	Turing NLR v5		91.2
4	DIRL Team	DeBERTa + CLEVER		91.1
5	ERNIE Team - Baidu	ERNIE		91.1
6	AliceMind & DIRL	StructBERT + CLEVER		91.0
7	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.8
8	HFL iFLYTEK	MacALBERT + DKM		90.7
9	PING-AN Omni-Sinitic	ALBERT + DAAF + NAS		90.6

图 2：T-ULRv6 XXL 位居 GLUE 排行榜首位

T-ULRv6 能够取得如此优异的成绩，是因为它在 XY-LENT 研究的基础之上，利用了不同语言之间的多向 (X-Y) 平行文本对 (bitexts)，并整合了 T-ULRv5 的关键创新，其中包括 XLM-E 架构、MRTD 和 TRTD 的新型预训练任务、改进的训练数据和词汇，以及高级微调技术 xTune。此外，为了能够扩展到 XXL 大小的模型，微软还借助了 ZeRO 的内存优化优势。

超越以英语为中心的平行文本对范式，更好地学习多语言表达

T-ULRv6 的关键改进在于摒弃了以英语为中心的 (EN-X) 平行文本对，直接利用不同语言之间的多向 (X-Y) 平行文本对（如法语 - 德语、印地语 - 乌尔都语，或斯瓦希里语 - 阿拉伯语）。尽管在多语言机器翻译中利用这种平行文本对数据属于常规操作，但这是由问题的性质所决定的，研究员们的此次尝试表明，利用平行文本对数据进行多语言编码器训练会带来意想不到的性能提升。虽然 EN-X 平行文本对有助于学习跨语言对齐和共享表示，然而这种方式在语言和领域的覆盖范围及多样性上会受到制约。另一方面，X-Y 平行文本对可以为学习多语言表示提供更丰富、更均衡的信息，从而可以更好地推广到更广泛的语言和任务中。

为了有效地利用 X-Y 平行文本对，研究员们采用了一种新颖的采样策略，以确保数据在多语言之间有效分布，同时保持语言边际分布一致。反言之，这也确保模型仍能维持强大的英语性能。

在编码器中有一个值得注意的特性，就是参数效率。XY-LENT XXL 明显优于 XLM-R XXL 和 mT5 XXL，同时规模较后两者分别缩小了约 2 倍和 3 倍。即使在 Base、Large 和 XL 三个类别中，与同类的其他模型相比，XY-LENT 也是最先进的，并且展现出了跨类别的竞争优势。强大的性能和较少的参数，在产品开发场景中非常实用。

模型	参数量	XNLI 性能
XLM-R XXL	10.7B	83.1
mT5 XXL	13B	85.0
T-ULRv6-XXL	**4.6B**	**86.1**

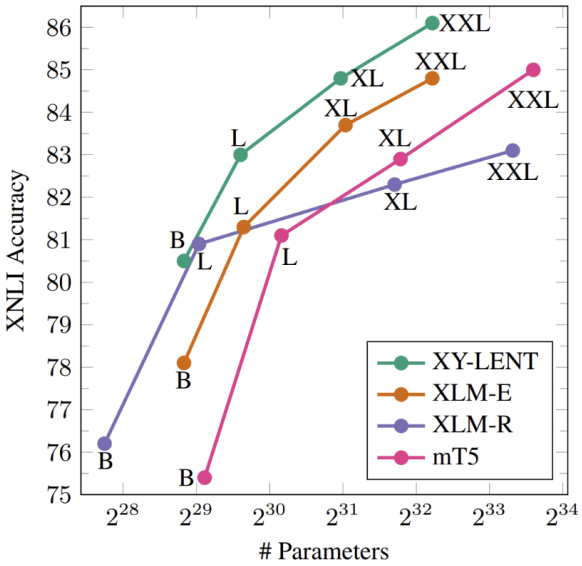


图 3：T-ULRv6 (XY-LENT) 在模型规模范围内具有 SOTA 水平，同时具有参数效率

在 T-ULRv6 中，微软亚洲研究院自然语言计算组的研究员们与微软图灵团队紧密合作，为预训练模型的研究和开发以及下游任务的微调算法，提供了关键技术。基于 XLM-E 工作中提出的多语言预训练方法，研究员们成功实现了 130 倍的收敛提速，为 T-ULRv6 提供了方法框架。此外，针对多语言预训练特有的语种竞争问题，研究员们还提出了 VoCap 准则，以此动态决定多语言词表的分配额度，从而更好地对多语言输入进行表征。基于多语言的一致性准则，微软亚洲研究院的研究员们提出的多语言微调框架 xTune，也更好地实现了跨语言迁移性能。

只需一个模型就能应对英语和多语言任务

T-ULRv6 XXL 的另一个显著优势，是它在不牺牲质量或效率的前提下，凭借单一模型即可在英语和多语言任务上同时实现 SOTA 性能。这意味着用户不用再根据自然语言处理任务来选择使用哪个预训练模型，因为 T-ULRv6 XXL 可以很好地处理这两种情况。这就简化了模型选择和部署的过程，也降低了维护多个模型所需的计算和存储成本。

为了实现这一点，T-ULRv6 利用其扩展能力和非英语平行文本对 (non-English bitexts) 优势消除了"多语言诅咒"，即在权衡英语和多语言性能时，常常给多语言模型造成困扰。T-ULRv6 不仅在涵盖一系列英语自然语言理解任务的 GLUE 基准测试中优于专门的英语模型，在覆盖 40 种不同类型语言和 9 种跨语言任务的 XTREME 基准测试中也优于专门的多语言模型。此外，T-ULRv6 模型规模也要小得多，这保证了其参数效率和可扩展性。

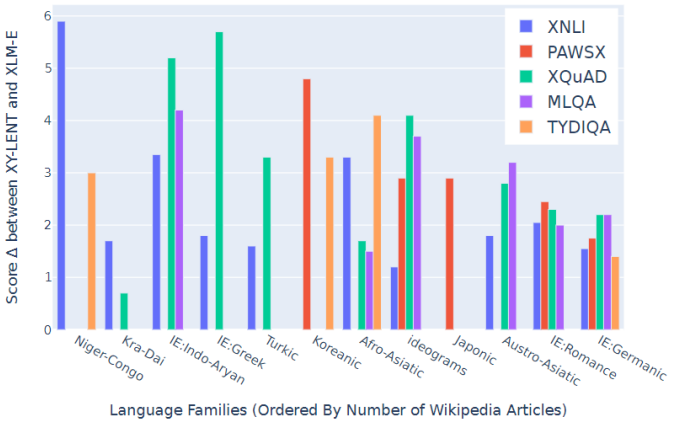


图 4：T-ULRv6 (XY-LENT) 在多语言任务中展现出了强大的性能

开放共享，共同推动领域发展

目前，T-ULRv6 已应用于微软必应 (Bing) 中，为必应的国际化提供支持，使用户能够使用不同语言在不同地区搜索信息。T-ULRv6 还将会把最先进的多语言功能赋能微软其他产品，通过其跨国别和跨语言的能力，助力微软践行 "予力全球每一人、每一组织，成就不凡" 的使命，为更多用户提供帮助。

微软一直认为 AI 技术要在学术界开放共享，进而促进合作与创新。因此，微软启动了"微软图灵学术计划" (MS-TAP，Microsoft Turing Academic Program)，允许科研人员提交研究方案，从而获得 T-ULRv6 和其他图灵模型的详细资料。微软邀请所有人共同探索多语言理解和生成的潜力，一起应对挑战，同时也欢迎大家提供宝贵的反馈和见解。未来，微软还将开源 Base 和 Large 模型，进一步推动该领域的研究工作。

以多语言技术为锚点，让 AI 更具包容性

多语言技术不仅是一个技术挑战，更是一项社会责任。微软一直致力于通过消除限制 AI 易用性和包容性的障碍，例如缺乏训练数据、语言建模成本过高以及多语言系统过于复杂等问题，实现 AI 的普及化。T-ULRv6 让 AI 向着这一目标迈出了重要一步，它为跨语言系统开发提供了一个更为高效和可扩展的框架，仅使用一个模型就能同时处理英语和多语言任务。微软很高兴有机会进一步提高技术水平，开发新的多语言能力，让世界各地的更多人和组织从中受益。希望这些工作能够推动社会进步，让 AI 更具包容性，并惠及所有人。

相关链接：

XY-LENT 论文链接：
Beyond English-Centric Bitexts for Better Multilingual Language Representation Learning
https://arxiv.org/pdf/2210.14867.pdf

XLM-E 论文链接：
XLM-E: Cross-lingual Language Model Pre-training via ELECTRA
https://arxiv.org/abs/2106.16138

xTune 论文链接：
Consistency Regularization for Cross-Lingual Fine-Tuning
https://arxiv.org/pdf/2106.08226.pdf

ZeRO 论文链接：
ZeRO: Memory Optimizations Toward Training Trillion Parameter Models
https://arxiv.org/pdf/1910.02054.pdf

VoCap 论文链接：
Allocating Large Vocabulary Capacity for Cross-lingual Language Model Pre-training
https://arxiv.org/pdf/2109.07306.pdf

微软图灵学术计划网页：
https://www.microsoft.com/en-us/research/collaboration/microsoft-turing-academic-program/

相关阅读

扫描二维码查看文章

文档基础模型引领文档智能走向多模态大一统

自 2019 年以来，微软亚洲研究院在文档智能领域进行了诸多探索，开发出一系列多模态任务的文档基础模型 (Document Foundation Model)，包括 LayoutLM (v1、v2、v3) 、LayoutXLM、MarkupLM 等。这些模型在诸如表单、收据、发票、报告等视觉富文本文档数据集上都取得了优异的表现，获得了学术界和产业界的广泛认可，并已应用在包括 Azure Form Recognizer、AI Builder、Microsoft Syntex 等在内的微软产品中，赋能企业和机构的数字化转型。

通用多模态基础模型 BEiT-3：引领文本、图像、多模态预训练迈向“大一统”

近年来，基础模型（foundation models，也被称为预训练模型）的研究从技术层面逐渐趋向于大一统（the big convergence），不同人工智能领域（例如自然语言处理、计算机视觉、语音处理、多模态等）的基础模型从技术上都依赖三个方面：一是 Transformers 成为不同领域和问题的通用神经网络架构和建模方式，二是生成式预训练（generative pre-training）成为最重要的自监督学习方法和训练目标，三是数据和模型参数的规模化（scaling up）进一步释放基础模型的潜力。

技术和模型的统一将会使得 AI 模型逐步标准化、规模化，从而为大范围产业化提供基础和可能。通过云部署和云端协作，AI 将有可能真正成为像水和电一样的"新基建"赋能各行各业，并进一步催生颠覆性的应用场景和商业模式。

微软亚洲研究院持续迭代 BEiT，为通用基础模型的大一统发展奠定基础

近期，微软亚洲研究院联合微软图灵团队推出了 BEiT-3 预训练模型，并在广泛的视觉及视觉 - 语言任务上，实现了 SOTA 的迁移性能。BEiT-3 创新的设计和出色的表现为多模态研究开创了新的范式，更预示着人工智能大一统渐露曙光。BEiT-3 的构建思路是什么？大规模预训练又将通向怎样的未来？在深科技近日的采访中，微软亚洲研究院首席研究员韦福如详细介绍了生成式自监督视觉预训练模型 BEiT 和通用多模态基础模型 BEiT-3 背后的技术，并探讨了大模型开发与训练中需要探讨和深思的问题，以及该领域的未来发展方向。

SPINE：高拓展性、用户友好的自动化日志解析新神器

在计算机系统与软件的实践和研究中，可靠性是至关重要并且经久不衰的课题。如何自动化地分析日志所记录的系统状态并让数据“说话”，受到了广泛研究。日志解析是自动化日志分析中的关键起步。如何将日志解析应用于大规模复杂的云环境往往面临诸多现实挑战，如数据不均衡，数据漂移等。

为了解决这些挑战，微软亚洲研究院的研究员们提出了支持用户反馈且具有高可扩展性的日志解析方法 SPINE。该方法被软件工程领域顶级会议 ESEC/FSE 2022 接收，并荣获“杰出论文奖”(ACM SIGSOFT Distinguished Paper Award)。SPINE 是如何提升日志解析效果和性能的呢？让我们从文章中获得答案吧！

在云计算时代，软件系统的可靠性至关重要，一点小问题就可能引发蝴蝶效应，影响百万用户。为了了解并保障软件系统的稳定，日志被广泛用于观测并忠实记录系统的内部状态，是分析与解决系统故障的基础。然而，使用人工分析体量巨大的日志并不现实，因此自动化日志分析日渐兴起，而日志解析是关键且基础的步骤。在实践中，日志数据往往存在着数据量巨大、极度不均衡、数据漂移且没有标注等问题。为了解决这些问题，并将日志解析真正落实到复杂的云环境中，微软亚洲研究院的研究员们和微软 Azure 的工程师们提出了支持用户反馈的大数据场景下的日志解析方法 SPINE，并将其落地到了产品线中。

近日，SPINE 被软件工程领域的全球顶级会议 ESEC/FSE 2022 接收，并荣获“杰出论文奖”(ACM SIGSOFT Distinguished Paper Award)。

ESEC/FSE 大会全称为 ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)，与 ICSE、ASE 并列为软件工程领域三大顶级会议，在学术界和工业界都具有极大的影响力。今年的 ESEC/FSE 大会有效投稿量为 449，最终接收 99 篇，接收率约为 22%，会议于 2022 年 11 月 14 日至 18 日在新加坡举办。

日志解析：智能日志分析的关键核心

自动化日志分析在近年来逐渐成为研究热点，例如基于日志的异常检测、故障诊断、故障预测等。几乎所有的自动化日志分析技术，都依赖于日志解析这一关键的前置步骤。经过日志解析，将半结构化文本形式的原始日志转换为结构化的日志数据之后，下游的各类日志分析任务才能自动化地执行。

日志解析可从形式上被定义为：从原始日志信息中提取日志模板和日志参数的任务。日志信息的主体通常由两部分构成：(1) 模板：描述系统事件的静态的关键字，通常为一段自然语言，这些关键字被显式地写在日志语句的代码中。(2) 参数：也称为动态变量，是在程序运行期间的某个变量的值。

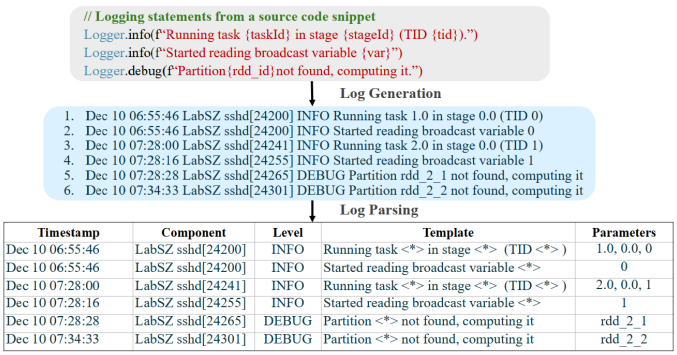


图 1：日志解析示例

现阶段，大量的自动化日志解析工作致力于准确高效地分离日志中的模板和参数部分。尽管这些日志解析器在公开的基准日志数据集上取得了良好成效，但它们在实际应用中仍然面临诸多挑战。微软的研究员和工程师们通过在实际工业环境中进行的大量例证研究，揭示出其中的两个核心挑战。

大规模、不平衡的日志数据

首先，大多数现有的日志解析器只能在单线程模式下运行。然而，现实世界的日志数据量极为庞大。例如，在例证研究中，仅微软某个内部服务，平均每天就会产出约 50 亿条日志，合每小时约 2 亿条。如此规模的数据量超出了任何单一计算核或节点的处理能力，尤其难以满足实时日志分析的需要。

表面上，日志解析似乎是一项很容易并行化的任务。然而，工业实践中日志数据的内在不平衡性将大大降低并行化的效率。

这促使研究员们设计一种能够在多个计算单元上进行更有效的横向扩展的日志解析器。

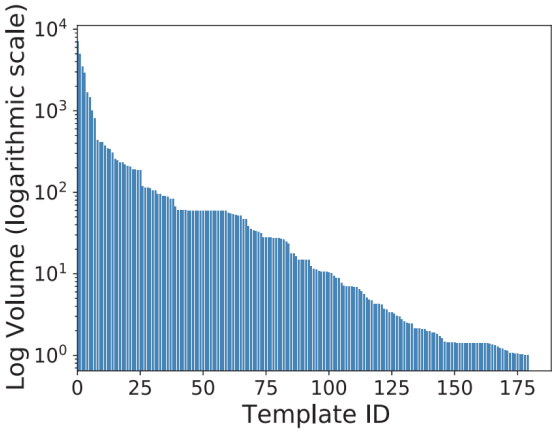


图 2：不同日志模板下日志数量分布：X 轴表示模板 ID，Y 轴表示对应于该模板的日志数量（Y 轴为对数标度）。

日志漂移与解析器的快速适应

另一个挑战来自于日志伴随着软件系统的迭代而不断发生变化。研究员们在微软某内部服务中收集了 8 周的日志，并计算随着时间推移而新出现的日志模板的数量，结果如图 3 所示。由于持续集成 / 交付（CI/CD）的开发范式，日志模板的数量会随时间增加，日志解析器也应不断地更新，以适应数据的漂移，否则解析的准确度会随时间流逝而逐渐下降。

遗憾的是，因为缺少足够的有标签数据，现有的日志解析器大多采用无监督的方法，例如聚类、频繁模式挖掘、最长共同子序列提取等来识别日志的公共部分作为模板。这需要大量的人工标注来进行繁琐的模型超参数调整，并且要求用户对日志解析方法的内部原理极为熟悉。因此，研究员们认为日志解析应当降本增效，尽可能地降低用户反馈机制的成本，提高用户体验，以达到快速调整日志解析器参数的效果。

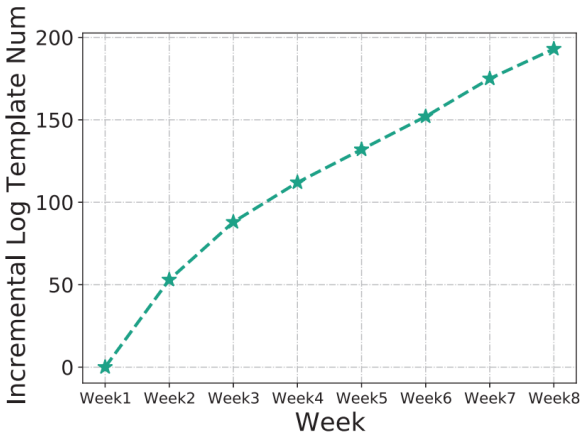


图 3：新日志模板数量增加曲线

反馈支持的高扩展日志解析器 SPINE

针对上述问题，微软亚洲研究院的研究员们设计了 SPINE。SPINE 具体分为两个阶段：离线训练阶段（红色箭头）和在线解析阶段（绿色箭头）。在离线训练阶段，SPINE 会基于收集的日志数据训练一个初始模型。随后，在在线解析阶段，应用训练得到的日志解析模型，处理不断更新的在线日志数据。

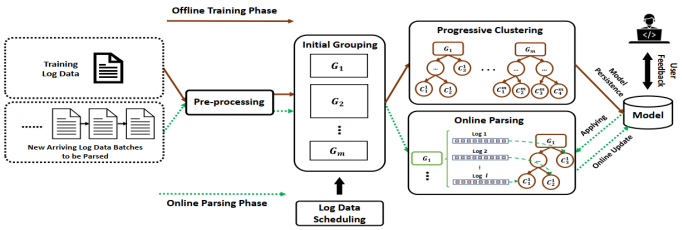


图 4：SPINE 模型总体架构图

SPINE 包含四个核心组件：日志数据预处理（Pre-processing）、初始分组（Initial grouping）、渐进式聚类（Progressive clustering）和在线解析（Online parsing）。首先，对原始的日志分词，并进行必要的日志清理。在此之后，初始分组模块会将日志快速分割成粗粒度的、互不重叠的多个日志组（log group）。再将渐进式聚类算法应用于每个日志组，把相似的日志进一步划分为细粒度的日志簇（log cluster）。一个日志簇中的日志，可以认为诞生于同一个日志打印语句。因此，可以提取其共同的 token 作为模板，将其余部分视为参数。在线解析阶段，SPINE 会将学习到的模型应用于新到来的日志数据。基于这些日志和模型中已有的日志模版之间的相似度，将其归属为最相似的日志簇中，并解析出其模板和参数。

SPINE 可以灵活地扩展到多个并行计算单元，以应对极大规模的工业日志数据。为了应对工业日志数据的极端不平衡性，研究员们设计了一种特殊的日志数据调度算法来平衡不同计算单元上的工作负载，以节约总体运行时间。此外，SPINE 还设计了专门的用户反馈机制来维持在漂移日志数据下的解析精度。

并行化日志数据调度

在前置步骤中，日志被划分成不同的日志组。然而，工业日志数据的不平衡性会导致日志的解析时间往往受制于最大的日志组。这一挑战促使了新调度算法设计的诞生，将日志解析任务更均匀地分配给多个计算单元，以达到最佳性能。假设 m 个日志组 g_i ∈ G 被分配到 n 个计算单元 e_i ∈ E。理想情况下，每个计算单元将处理相等数量的 avg = ∑_(i=1)^m|g_i |/n 的日志消息，其中 |gi| 是日志组 g_i 的大小。为了实现这一目标，研究员们将数量高于平均水平的日志组分割成更小的子组，使其日志数量接近平均值，同时合并数量少于平均水平的日志组，使其所产生的超集的规模也接近于平均水平。算法细节可参考图 5 中的伪代码。

Algorithm 1: Log Data Scheduling for Parallelization

Require: Set of Executors, $\mathcal{E}$
Set of Log Groups to be Allocated, $\mathcal{G}$

Sort $\mathcal{G}$ according to their sizes in descending order;
$\mathcal{E}_{left} = \mathcal{E}$; $\mathcal{G}_{left} = \mathcal{G}$;
$l_{left} = l_{total} = \sum_{g_i \in \mathcal{G}} |g_i|$; $avg = \frac{l_{left}}{|\mathcal{E}_{left}|}$
for g_i *in* $\mathcal{G}$ **do**
  if $|g_i| > avg$ **then**
    $h = floor(|g_i|/avg)$
    Split g_i into subsets $g'_1, g'_2, ..., g'_h$ uniformly and schedule them to
    $\mathcal{E}_{used} = \{e_j, e_{j+1}, ..., e_{j+h-1} | \forall e_j \in \mathcal{E}_{left}\}$;
    $\mathcal{E}_{left} = \mathcal{E}_{left} \setminus \mathcal{E}_{used}$; $\mathcal{G}_{left} = \mathcal{G}_{left} \setminus g_i$; $l_{left} = l_{left} - |g_i|$
    $avg = \frac{l_{left}}{|\mathcal{E}_{left}|}$
  else
    break;
if $|\mathcal{G}_{left}| > 0$ **then**
  $BestFit(\mathcal{E}_{left}, \mathcal{G}_{left})$
return

图 5：日志数据并行化调度算法

用户反馈融合

SPINE 模型架构中使用了渐进聚类算法进行细粒度的日志聚类。在每个日志组中，日志信息首先被转换为 one-hot 向量，再利用常见聚类方法（例如 K-Means 或高斯混合聚类）对日志向量进行聚类，并将其分为两个子组。同样的过程在每个子组上迭代进行，直到满足停止条件。结果可得一棵二叉聚类树，树上的叶子结点即为日志簇，如图 6 所示。此算法的优点在于：（1）其算法本身具有极高的聚类效率。（2）该算法非常容易控制聚类的粒度。其核心在于是否需要继续沿着聚类树进行二叉分裂。这为用户决定何时停止聚类提供了可能。

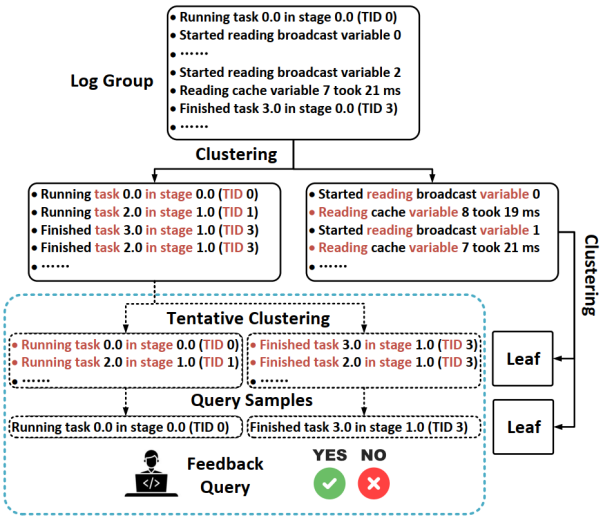


图 6: 渐进聚类与用户反馈查询示意图

用户的反馈过程涉及多轮查询，在每一轮中，SPINE 在一定条件下会向用户推荐来自同一日志簇的一对日志，用户对于这对日志是否共享相同的模板给出反馈。用户的反馈有助于 SPINE 决定是否将相应的日志簇分成两个子簇。随后，SPINE 选择下一个需要反馈指导的日志簇决定是否要拆分。图 6 显示了一个反馈查询的例子。

如何尽可能少地标注日志数量，并同时最大限度地提高模型的准确性，是设计反馈机制的核心问题。SPINE 的反馈查询的选择基于饱和度增益（saturation gain）进行的。我们约定，若一个 token 在某一 log cluster 的每条日志消息中都出现过，则视为一个 saturated token。一般来说，平均 token saturation 越高，意味着这组日志共享了更多的单词，所以它们越有可能是来自同一个日志生成语句。饱和度增益是指，如果我们把一个叶子节点分成两个子叶子节点，饱和度的增量。我们计算一个叶子节点 ln 的饱和度增益 G_ln，如公式 1 所示。该公式反映出继续向下分裂聚类树是否能所带来的更多的 saturated token，从而有较高的提高解析精度的可能性。

$$G_{ln} = \left(\frac{|cln_1| \cdot \bar{S}_{cln_1} + |cln_2| \cdot \bar{S}_{cln_2}}{|ln| \cdot \bar{S}_{ln}} - 1 \right) \cdot \frac{|ln|}{|l_{total}|}$$

公式 1: 饱和度增益计算公式

实验结果

研究员们在 16 个公开的日志数据集上对 SPINE 进行了大量实验，以回答以下研究问题。

RQ1: 在没有反馈指导和并行化加速的情况下，基础版本的 SPINE 效果和效率如何?

RQ2: 有反馈指导的 SPINE 效果如何?

RQ3: 有并行化加速的 SPINE 效率如何?

基础实验

从对比 SPINE-base 和五个 SOTA 日志解析器的解析精度实验（表 1），以及解析运行时间的对比结果（图 7）中可见，SPINE 的效果优于或不逊于 SOTA 方法。另外，在没有并行加速优化的情况下，SPINE 的解析效率也可以达到领先水平。

Dataset	AEL	LenMa	Spell	IPLoM	Drain	SPINE-base	Best
HDFS	**0.998**	**0.998**	**1***	**1***	**0.998**	**0.998**	**1**
Spark	**0.905**	0.884	**0.905**	**0.920**	**0.920**	**0.925***	**0.925**
BGL	0.758	0.690	0.787	**0.939**	**0.963***	**0.948**	**0.963**
Windows	0.690	0.566	**0.989**	0.567	**0.997***	**0.990**	**0.997**
Linux	0.673	0.701*	0.605	0.672	0.690	0.676	0.701
Android	0.682	0.880	**0.919**	0.712	**0.911**	**0.932***	**0.932**
Mac	0.764	0.698	0.757	0.673	0.787	0.789*	0.789
Hadoop	0.538	0.885	0.778	**0.954***	**0.948**	**0.946**	**0.954**
HealthApp	0.568	0.174	0.639	0.822	0.780	**0.988***	**0.988**
OpenSSH	0.538	**0.925***	0.554	0.802	0.788	0.681	**0.925**
Thunderb.	**0.941**	**0.943**	0.844	0.663	**0.955**	**0.964***	**0.964**
Proxifier	0.518	0.508	0.527	0.515	0.527	**0.967***	**0.967**
Apache	**1***	**1***	**1***	**1***	**1***	**1***	**1**
HPC	**0.903***	0.830	0.654	0.824	0.887	0.871	**0.903**
Zookeeper	**0.921**	0.841	**0.964**	**0.962**	**0.967**	**0.989***	**0.989**
OpenStack	0.758	0.743	0.764	0.871*	0.733	0.757	0.871
Average	0.754	0.721	0.751	0.777	0.865	**0.901***	N.A.

表 1：解析精度对比结果

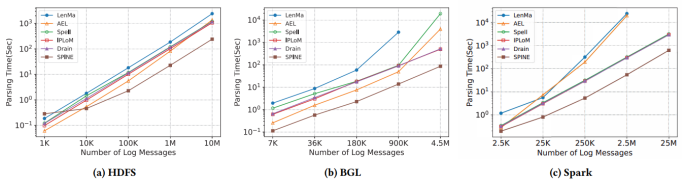


图 7：解析运行时间对比结果

用户反馈实验

研究员们将 SPINE-feedback 版本应用于三个日志数据集，即 Linux、Mac 和 OpenStack 后，如表 1 中所示。注意，在这三个数据集上，原有的日志解析器的解析精度都很不理想。为此，研究员们首先为每个数据集训练了一个初始解析模型，随后查询了具有最高饱和度增益的日志簇的用户反馈。

为了显示 SPINE 反馈机制的有效性，实验中研究员们使用了两个基线推荐策略。第一个基线是一个随机策略（表示为 random），即随机地从任意一个日志簇中抽出一对日志推荐给用户标注。第二条基线表示为 LNS，即仅仅考虑叶子节点的大小进行推荐。实验结果如图 8 所示。随着反馈次数的增加，以上策略都可以在 SPINE 模型架构上提高解析的准确性。与其它两个基线相比，SPINE-feedback 取得了最高的准确度提升。例如，Linux 和 OpenStack 日志数据集上的解析准确率可以在不到 10 次的用户反馈后提高到 0. 90 以上。即使是在非常复杂的 Mac 日志数据集上，解析准确率也可以从 0.78 提高到 0.85 以上，而这仅仅需要 20 次用户反馈即可。

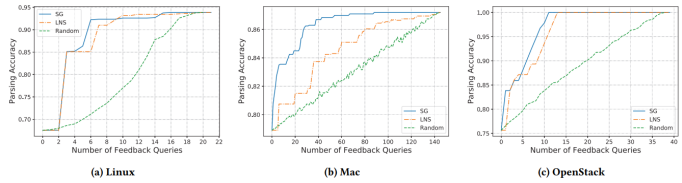


图 8：用户反馈实验结果

并行化实验

在三个最大的数据集——HDFS、BGL 和 Spark 上，研究员们评估了日志数据调度方法的有效性。研究员们在不同数量的计算单元下运行 SPINE-parallel，并记录了其吞吐量（每秒处理的日志数量）变化，其结果如图 9 所示。

SPINE 日志调度方法可以显著提高解析的效率。例如，SPINE-parallel 在 BGL 数据集上可以实现 5 倍左右的吞吐量提升，大约每秒处理 225,000 条日志。对比简单的 BestFit 方法，SPINE 的日志数据调度算法可以更充分地利用更多的计算单元，以到达更高的效率。

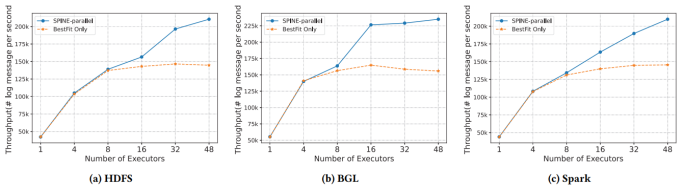


图 9：并行加速实验结果

相关链接：

论文链接：
https://www.microsoft.com/en-us/research/publication/spine-a-scalable-log-parser-with-feedback-guidance-2/

参考文献

[1] Towards automated log parsing for large-scale log data analysis.

[2] Onion: identifying incident-indicating logs for cloud systems.

[3] Public datasets for log parsing. https://github.com/logpai/logparser

[4] Tools and benchmarks for automated log parsing

微软亚洲研究院深入探索图深度学习领域两大挑战，以图深度学习赋能知识计算

在 NeurIPS 2022 联合 Open Graph Benchmark 举办的大规模图学习竞赛（Open Graph Benchmark Large-Scale Challenge，OGB-LSC）上，微软亚洲研究院数据、知识与智能（DKI）组的研究员们聚焦知识图谱的链接预测任务，通过更好的知识图谱补全方案，实现了知识图谱更高的 "时效性"、"准确性" 和 "完备性"。而在今年 2 月 WSDM 2022 联合亚马逊举办的动态异质图上的链接预测竞赛中，DKI 组的研究员们也取得了优异的成绩，其研究成果强调了对异质图信息和时序信息的建模。

在图深度学习领域的持续深耕，让微软亚洲研究院 DKI 组提出了一系列新方法和新思路，为多项研究成果的突破奠定了基础。那么对于图深度学习技术在知识计算领域的应用，微软亚洲研究院的研究员们有哪些独到的理解？又预见了哪些前沿的研究方向？

图（Graph），作为一种通用的数据组织方式，被广泛应用于建模实体间的联系，例如知识图谱、社交网络、交通路网、引文网络、互联网以及云服务依赖关系网络等。随着深度学习技术的快速发展，由深度学习与图数据处理相结合，催生出了图深度学习这一热门的研究方向，并以图嵌入、图神经网络等技术为代表对图数据进行学习和分析。从数据的角度来看，图深度学习如今已成为图数据分析背后的重要技术。

微软亚洲研究院数据、知识与智能（Data, Knowledge and Intelligence，DKI）组一直致力于发掘数据的价值，因此，DKI 组的研究员们希望从数据分析和知识提取中获取洞见，以更有效的图深度学习技术来赋能企业级的数据分析和知识计算。所谓知识计算就是指利用计算机程序来处理人类知识的过程，而且在这个过程中要将人类的知识转化为计算机可以理解的形式，并用这些信息解决复杂的问题。

知识计算领域的数据对象往往很有特点，其中的典型数据包括知识图谱和根据领域特点自定义的异构网络，这类图中的节点和边有更明确的语义，而且往往有确定的实体名和关系类型名，还常具有详细的文本描述。图的结构和语义信息都是对分析结果有明显影响的要素，基于这两种信息融合的知识表示也非常具有挑战性，所以图学习模型的设计也要更有针对性。

目前对知识的建模手段主要有两类：一类是通过大规模语言模型隐式建模知识，但这类模型的可控度和可解释性较低，有些回答真假难辨，比如 ChatGPT；另一类是通过显式的知识建模，利用结构化的知识表达，将其存储于知识库中，可以显式进行问答、推理等任务，然而如何更好地利用知识库中的知识却是个难题。

微软亚洲研究院 DKI 组的研究员们认为可以利用图深度学习，从以下几个方面增强显式知识建模的能力：

（1）增强知识表示能力。通过相应技术学习得到知识的向量表示，让现有的智能模型可以更好地利用知识库中的知识。

（2）提升知识挖掘能力。图深度学习技术可以用来挖掘知识图谱结构中的隐藏关系，以更好地理解知识中的含义和关联性。

（3）扩展知识应用范围。图深度学习技术能够应用于多种领域，如自然语言处理、推荐系统、知识图谱构建等，为知识计算的应用提供了更多的可能性。

知识图谱是最为常用的显式建模知识的方式，它是一种用节点表示实体，用连边表示关系的图结构组织方式。针对知识图谱的图深度学习技术是知识计算中很重要的一环。目前，知识图谱上的图深度学习方法以嵌入技术为主，该类技术将实体和关系映射到低维向量空间，用来表示知识图谱中实体和关系之间的相似度，从而进行知识图谱的推理、推荐和分类等任务。在应用外部知识解决各类智能任务的过程中，图深度学习也发挥着重要作用。

"我们希望利用图深度学习来增强显式建模知识的能力，并结合知识图谱和图深度学习进行更多探索。针对知识图谱，我们通过图深度学习来挖掘更多潜在的隐藏关系，力争得到更全面、完善的知识表达，这也是我们在 NeurIPS 2022 大规模图学习竞赛 OGB-LSC 上的课题，比赛结果表明我们的研究已经取得了阶段性成果。" 微软亚洲研究院 DKI 组主管研究员杜仑表示。

系列研究让图深度学习模型更通用、更稳定

图深度学习领域的研究内容非常广泛，微软亚洲研究院 DKI 组将系列研究聚焦在了图深度学习需要持续攻克的几个课题上：设计更通用、更具泛化性的图深度学习模型和更稳定有效的模型训练策略，以及探索更广泛的图模型应用场景。

从模型设计的角度，目前很多模型都擅长处理具有同配属性

的数据。同配属性是指图上节点具有相邻相似性，这种性质在传统的图研究对象中存在较多，例如社交网络、交通路网等等，然而图数据的覆盖面非常广，例如企业中团队协作的关系网络就有更明显的优势互补倾向，或者推荐系统中用户对于内容不喜欢的反馈网络显然不具备同配关系。那么如何建模更广泛类型的图，并挖掘更多图中的有效信号，是目前模型设计上的一个挑战。

从模型训练的角度来看，由于图数据中节点和节点的连边导致训练过程中无法简单地流式遍历数据，需要配合图采样等技术才能进行有效的训练，因此如何在保证高效训练的同时又尽可能减少信息损失，是真实大规模图数据场景中的重要问题。除了图特有的问题外，图深度学习模型的训练也会遇到其他深度学习模型所面临的类似的问题，比如如何保证训练的稳定性、效率和最终模型的泛化表现等。

除此之外，图模型的过压缩（oversquashing）、过平滑（oversmoothing），以及一般深度学习的模型初始化、过拟合等也都是需要逐一解决的问题。

经过近几年的持续研究，DKI 组的研究员们在适用范围更广、可解释性更强的图模型设计，以及一些通用的提高模型训练稳定性和泛化性的设计等方面都取得不少突破性成果。

在更具泛化性的模型结构设计方面，研究员们提出了针对图同配性和异配性同时建模的双核图网络模型，和针对邻域特征分布建模的混合矩图网络模型：

针对图同配性和异配性同时建模的双核图网络模型：研究员们发现无法建模异配关系的部分原因是，对同一阶邻居的向量表征使用了相同的核做变换所致，即使使用类似于图注意力网络（GAT）的注意力机制，但由于注意力计算的权重总是一个正值，所以一个核无法同时对节点表征之间的相似性和相异性（如正负相关性）进行建模。针对这个问题，研究员们分析发现，无论是在同配图还是异配图的数据集上，都存在着相当数量的异配子图，且子图的异配度参差不齐，而传统模型如 GCN（图卷积神经网络）在同配子图上往往表现优异，但在异配子图上发挥较差，这充分说明了同时建模同配和异配性模型的必要性。因此，研究员们提出了一种基于双核特征转换和门（gate）机制的新型 GNN（图形神经网络）模型——GBK-GNN。通过具有不同同质异质特性的七个真实数据集的广泛实验表明，与其他 SOTA 方法相比，GBK-GNN 有稳定且显著的提升。

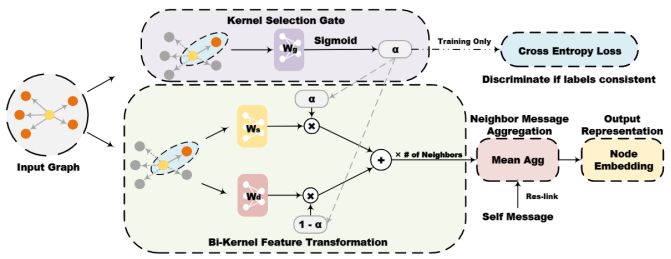


图 1：GBK-GNN 模型架构图

针对邻域特征分布建模的混合矩图网络模型：GNN 是一类通过聚合邻居信息来对图上的节点、边或者子图进行表示的机器学习模型。然而，大多数现有的 GNN 都使用单一的统计量，如平均数、最大值和求和，来聚合邻居的特征，丢失了与邻居特征分布相关的信息，降低了模型的性能。为了解决这个问题，研究员们借鉴统计学理论的矩方法，提出了新的 GNN 模型——混合矩图神经网络 MM-GNN。在 15 个真实世界图数据集（包括社交网络、引文网络和网页网络等）上进行的广泛实验表明，MM-GNN 优于现有的最先进的模型。

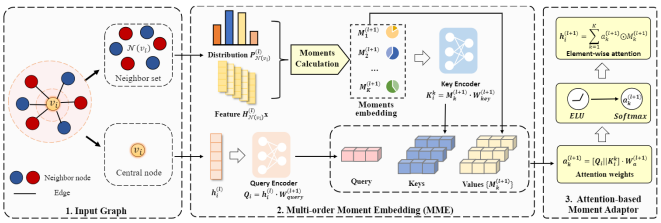


图 2：MM-GNN 模型架构图

在探索稳定的图深度学习模型的过程中，微软亚洲研究院 DKI 组还发现了稳定神经元的响应对模型泛化能力提升的帮助，提出了基于信息瓶颈理论的神经元竞争初始化策略：

稳定神经元响应以提升模型泛化性能：研究员们从神经元级别的细粒度出发，分析了单个神经元在神经网络训练和测试中的响应特性，发现提升神经元对同类输入样本响应的稳定性能够有效地提高神经网络的泛化性能。据此，研究员们提出了一种通用的正则项，用于控制神经元在激活状态下响应的类内方差。该正则项简单高效，不仅显著提高了图学习领域的图神经网络的泛化能力，还在计算机视觉领域中为卷积神经网络和多层感知机模型带来了显著提升。

基于信息瓶颈理论的神经元竞争初始化策略：在深度神经网络的复杂系统中，稳定的训练过程往往依赖于有效的初始化机制。现有的初始化机制研究工作主要关注于如何更好地缓解训练过程中所出现的梯度消失或爆炸问题，但缺乏对提升模型最终泛化效果的关注。受信息瓶颈理论（information bottleneck theory）的启发，研究员们定义了两个初始化目标，保证初始模型具有一定分类效果的同时能尽可能多地保留两种模型输入的信息量。此外，通过一种新颖且高效的神经元竞争算法，模型的初始化在上述两个目标之外还能保证初始化参数的多样性。该方法的新颖性和有效性得到了 CIKM 委员会的青睐，并获得了最佳短文奖。

DKI 组还利用图建模方法赋能了更多领域，提出了基于图模型增强的表格理解深度网络。表格数据结构的自动化理解是对文档表格和网页表格进行数据分析的重要步骤。然而，表格数据类型多样，包括便于存储的数据库表格、为了利于展示的电子表格以及结构更为灵活的问卷式表格，这大大增加了表格理解的难度。对此，研究员们利用图结构灵活、泛用性强的特点，引入了图建模的思路，兼顾了建模表格结构以及表格中文

本的语义信息，设计了一个面向表格的通用深度网络，可以有效地理解表格结构。此外，网络中还引入了行粒度和列粒度上的双向循环神经网络模块，以更好地理解表格不同区域间的边界关系。在两种不同数据粒度的真实表格理解任务中，该方法都取得了最优表现。

加强合作，推动图深度学习赋能更多场景

微软亚洲研究院 DKI 组在图深度学习研究中所取得的阶段性技术突破，现已开始应用在众多业务场景中。例如，在 Excel 中，通过图建模的方法引入 WordNet 作为建模表格语义信息时的外部知识，对表格结构识别任务有明显提升。而在领英（LinkedIn）的工作推荐功能中，一个很重要的问题是如何把合适的工作推荐给合适的人。领英与 DKI 组合作通过异构图建模包括行业信息、教育背景、技能等在内的领域知识，并结合异构图 GNN 模型同时建模领域知识与用户行为等信息，当前已在线下实验中取得了明显的推荐准确率提升。

除此之外，微软亚洲研究院 DKI 组还与学术界的高校和科研机构合作，一道推进图深度学习领域的进步与应用。通过微软亚洲研究院铸星计划，DKI 组的研究员与中科院计算所的学者共同探索了结合图模型的交通轨迹数据的表示学习，借由层级图模型建模数据点的物理距离，有效提升了轨迹表示学习模型的效果。在与上交所的研究合作中，研究员们对大规模图处理进行了研究，提出了新的图模型加速推断方法，使推理过程更高效。

对于图深度学习未来的研究规划，微软亚洲研究院首席研究员韩石表示，“下一步，微软亚洲研究院 DKI 组将持续推进企业级知识计算领域与相关基础研究的探索，包括文档智能、显式知识表示和大规模语言模型的结合、以及图深度学习模型等。同时，我们也希望可以与更多学术机构和专家学者合作，共同探索图深度学习的前沿发展方向。”

感谢微软亚洲研究院 DKI 组图深度学习研究团队（成员包括：杜仑、陈旭、马晓君、付强、韩石）对本文的贡献。

相关链接：

论文链接：

1. Solution for NeurIPS 2022 OGB-LSC
https://ogb.stanford.edu/paper/neurips2022/wikikg90mv2_DNAKG.pdf

2. HTGN-BTW: Heterogeneous Temporal Graph Network with Bi-Time-Window Training Strategy for Temporal Link Prediction
https://www.wsdm-conference.org/2022/wp-content/uploads/2022/02/Task2_nothinghere_2nd.pdf

3. Neuron with Steady Response Leads to Better Generalization，NeurIPS'22
https://openreview.net/forum?id=9YQPaqVZKP

4. MM-GNN: Mix-Moment Graph Neural Network towards Modeling Neighborhood Feature Distribution, WSDM'23
https://arxiv.org/abs/2208.07012

5. Neuron with Steady Response Leads to Better Generalization，NeurIPS'22
https://openreview.net/forum?id=9YQPaqVZKP

6. Neuron Campaign for Initialization Guided by Information Bottleneck Theory，Best Short Paper at CIKM'21
https://dl.acm.org/doi/abs/10.1145/3459637.3482153

7. TabularNet: A Neural Network Architecture for Understanding Semantic Structures of Tabular Data, KDD'21
https://dl.acm.org/doi/abs/10.1145/3447548.3467228

TSRFormer：复杂场景的表格结构识别新利器

近年来，各大企业和组织机构都在经历数字化转型。将文档转换成计算机所能识别的样态，是数字化转型的关键步骤，如何识别出图片中表格具体的结构与内容，并直接提取其中的数据和信息是学术界和工业界共同瞩目的焦点。然而，目前的表格识别算法多用于识别横平竖直的表格，对于全无边界和实线的表格、行列之间存在大片空白区域的表格等日常生活中常见的表格还没有较好的解决方案，对于拍摄角度倾斜而表格边框弯曲等情况更是束手无策。今天我们将为大家介绍微软亚洲研究院在表格结构识别方向的最新进展，研究员们提出了一种新的表格结构识别算法 TSRFormer，能够较好地识别复杂场景中不同类型的表格。

如今，各行各业正在向数字化转型，海量的文档型数据也源源不断地生成。用人工处理这些蕴含着丰富信息的文档，存在如耗时长、成本高、易出错等缺陷，在实际应用中难以高效执行。因此，社会对于自动化文档处理技术的需求日益增加，智能文档处理（IDP）成为了近几年的热点。与此同时，市场上也涌现出了许多相关产品，例如微软就提供了全方位的 IDP 服务及解决方案（https://adoption.microsoft.com/intelligent-document-processing/）。如图 1，智能文档处理通过光学字符识别（OCR）、文档图像分析、计算机视觉、自然语言处理等技术，将复杂的非结构化文档数据转变为能被计算机直接理解和使用的结构化数据，从而帮助企业或个人更加高效地获取文档中的有用信息。

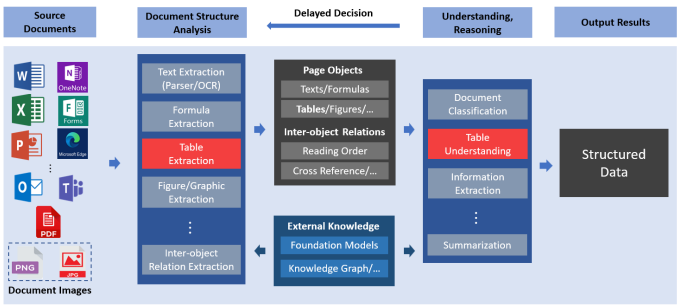


图 1：智能文档处理（IDP）的流程示意图

在各类文档中，表格作为一种高效的信息表达形式，通常被人们用来呈现结构化的数据，例如公司财报、发票、银行流水、实验数据、医院检验报告等等。如何抽取及理解表格的技术一直都是 IDP 中的重要组成部分。

表格抽取技术解决的主要问题是如何自动地将图像中的表格数字化，其包含两个子任务：表格检测和表格结构识别。其中，表格结构识别旨在从表格的图像中还原表格的结构信息，包括每个单元格的坐标位置以及每个单元格所属的行列信息。如图 2 所示，在实际场景中，表格结构识别是一个极具挑战性的问题。其挑战的难度主要在于表格的结构与内容的复杂多样性，例如存在完全无边界和实线的表格、包含许多空白单元格或者跨行跨列单元格的表格、行列之间存在大片空白区域的表格、嵌套的表格、密集的大表格、单元格包含多行文字内容的表格等等。不仅如此，在相机拍摄的场景中，有些表格的边框可能因拍摄角度而倾斜或弯曲，这都大大增加了表格结构识别的难度。

图 2：表格图像的多样性与复杂性

近年来，表格结构识别领域受到了学术界与工业界的广泛关注，其中涌现出了大量研究成果。但这些研究成果的视角大多仅限于简单的应用场景，例如 PDF 或扫描文档中横平竖直的表格或分割线均为实线的表格，而对于图 2 中这些在实际场景中经常出现的情况，尤其是倾斜、弯曲且没有实线的表格关注度较低。因此，现有的算法距离完全解决实际场景中的表格识别问题还

存在很大差距。为了让表格识别技术适用于更广泛的应用场景，微软亚洲研究院的研究员们提出了一种新的表格结构识别算法 TSRFormer[1]，该算法能够较好地识别复杂场景中不同类型的表格。

TSRFormer: 提供表格结构识别新思路

现有的表格结构识别算法大致分为三种范式：编码 - 解码范式、自底向上范式和拆分 - 合并范式。编码 - 解码范式下的模型在输入表格图像后可以直接预测表示表格结构的编码序列（如 HTML、LaTeX 等）。该范式即使在识别较为容易的横平竖直表格的任务中，仍然需要远超于其他范式的训练数据才能产出较好的效果。若要进一步支持倾斜或弯曲的表格，则还需额外收集大量的数据，因此研发成本较高。此外，目前基于该范式的方法在处理单元格较为密集的大表格时，精度相对较低。

自底向上范式一般需要依赖额外的模块预先检测文本或单元格作为基础单元，再预测这些基础单元是否属于同一行、列或单元格从而定位表格结构。所以该范式难以处理包含大量空白单元格或空行空列的表格。

不同于以上两种范式，微软亚洲研究院的研究员们发现基于拆分 - 合并范式的方法具有更强的可扩展性，在复杂场景中只需要较少的训练数据就能达到很高的精度，而且可以鲁棒地处理包含空白单元格以及空行空列的表格。因此，基于该范式研究员们提出了 TSRFormer。如图 3 所示，对于输入的表格图像，TSRFormer 先由拆分模块预测出所有行、列的表格分割线，求交点后，生成 N x M 个单元格，再由合并模块预测相邻单元格是否需要合并从而恢复出跨多行、多列的单元格。

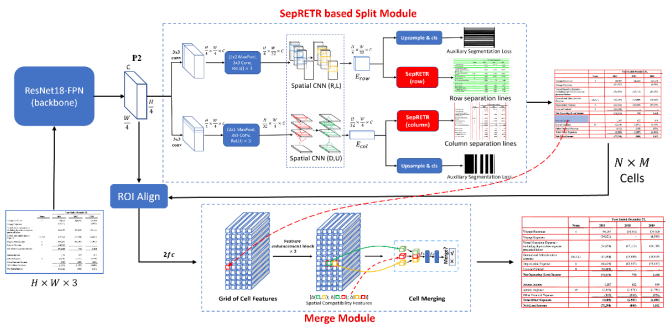


图 3：TSRFormer 的整体结构图

在以往基于拆分 - 合并范式的方法中，预测拆分模块的表格线一般通过图像分割模型结合从分割图中提取表格分割线的后处理模块完成（如 [2][3] 等），而基于规则设计的后处理模块难以处理低质量的分割图，这严重降低了模型针对诸如倾斜、弯曲的表格识别的精度以及泛化能力。不同于既有设计，TSRFormer 提出了一种不需要后处理模块的全新思路：通过直接回归的方式来预测分割线。具体来说，该方法采用每条分割线上的若干采样点来表示该分割线，并让模型直接回归每条分割线上采样点的坐标，从而得到分割线的位置信息。

为了让 TSRFormer 能够精确且高效地预测表格分割线，研究员们还提出了一套新的基于两阶段 DETR [4] 的分割线回归算法：SepRETR。如图 4 所示，在第一阶段中，SepRETR 先用参考点预测模块，为每一条表格分割线预测出一个参考点（reference point）；在第二阶段，由这些参考点的视觉以及空间信息组成的特征向量集合作为查询特征（query）输入进一个解码器（Transformer decoder）来回归对应的完整分割线。

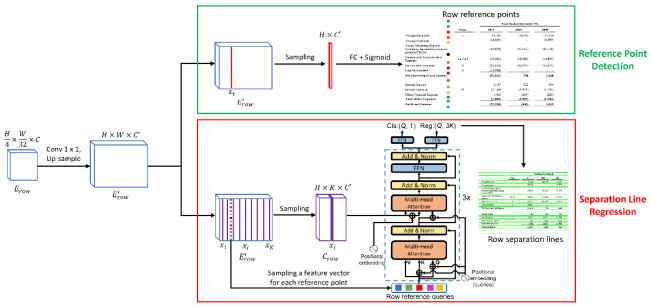


图 4：基于 SepRETR 的表格分割线预测模型（此处以行分割线为例）

在此基础上，研究员们进一步提出了两个改进算法来提升模型性能：（1）提出了基于先验增强的匹配策略来解决原始 DETR [5] 训练收敛慢的问题；（2）仅采样少量像素的特征作为解码器交叉注意力（cross attention）模块的输入，该方案可以使模型事半功倍，利用较少的计算量即可达到高定位精度。

实验结果及可视化效果

目前，学术界的绝大部分公开数据集都只包含 PDF 或者扫描文档图像中完全横平竖直的表格（如 SciTSR [6]、PubTabNet [7] 等）。与实际应用场景相比，这类数据集较为简单，不能涵盖日常生活中的所有表格类型。近一年，复杂场景中的表格结构识别问题逐渐受到关注，例如去年新发布的 WTW 数据集 [8] 就开始考虑实际自然场景中的表格。在该数据集中，由于相机拍摄引起的干扰，一些表格会出现倾斜或弯曲，这大大增加了表格结构识别问题的难度。但 WTW 数据集只考虑了分割线均为实线的表格，而没有包含无实线的表格。为了能够更全面地测试模型在各类场景下的性能，研究员们收集了一个更加复杂的数据集，该数据集包含了各式各样复杂场景的样本，例如结构复杂、包含大量空单元格或长跨行跨列单元格的无实线表格，以及倾斜甚至弯曲的表格等等。

研究员们首先在三个较大规模的公开数据集 SciTSR、PubTabNet 以及 WTW 上验证了 TSRFormer 的性能。从表 1、表 2 以及表 3 的结果可以看出，无论是在横平竖直的简单场景（SciTSR、PubTabNet）还是在分割线均为实线的自然场景（WTW）表格识别任务上，TSRFormer 均比现有的方法表现得更加优秀。

Methods	SciTSR (%)			SciTSR-COMP (%)		
	Prec.	Rec.	F1	Prec.	Rec.	F1
TabStruct-Net [9]	92.7	91.3	92.0	90.9	88.2	89.5
GraphTSR [6]	95.9	94.8	95.3	96.4	94.5	95.5
LGPMA [10]	98.2	99.3	98.8	97.3	98.7	98.0
FLAG-Net [11]	**99.7**	99.3	99.5	98.4	98.6	98.5
TSRFormer	99.5	99.4	99.4	99.1	98.7	98.9
TSRFormer*	**99.7**	**99.6**	**99.6**	**99.4**	**99.1**	**99.2**

表 1：TSRFormer 与现有方法在 SciTSR 上的性能对比

Methods	Training Dataset	TEDS (%)	TEDS-Struct (%)
EDD [7]	PubTabNet	88.3	-
TableStruct-Net [9]	SciTSR	-	90.1
GTE [12]	PubTabNet	-	93.0
LGPMA [10]	PubTabNet	94.6	96.7
FLAG-Net [11]	SciTSR	95.1	-
TSRFormer	PubTabNet	-	**97.5**

表 2：TSRFormer 与现有方法在 PubTabNet 上的性能对比（其中 TEDS[7] 指标同时考虑表格结构识别和表格内容 OCR 识别的精度，而 TEDS-Struct [10] 仅评测表格结构识别，因此后者更适用于公平比较表格结构识别模型的精度）

Methods	Prec. (%)	Rec. (%)	F1-score (%)
Cycle-CenterNet [8]	93.3	91.5	92.4
TSRFormer	**93.7**	**93.2**	**93.4**

表 3：TSRFormer 与现有方法在 WTW 上的性能对比

为进一步验证 TSRFormer 的有效性，研究员们在更具挑战性的内部数据集上开展了实验，并将 TSRFormer 与另外两个基于拆分 - 合并范式的代表算法——SPLERGE [2] 和 RobusTabNet [3]，进行了对比。为了使对比更加公平，在实现这三个方法的时候仅有表格分割线预测的部分不同，其余部分模型结构均保持一致。从表 4 可以看出，由于 SPLERGE 假设表格是横平竖直的，其在同样是横平竖直场景的数据集 SciTSR 和 PubTabNet 上都能取得接近 SOTA 的结果，但在包含倾斜甚至弯曲的内部数据集上则大幅度落后于 TSRFormer，F1-score 相差了 11.4%。图 5 的可视化效果展示了 SPLERGE 与 TSRFormer 在复杂场景中的明显差距。一个非常平等的学术交流环境，在这里我有机会见到很多非常优秀的研究学者，他们也都非常愿意和同学们交流学术经验。"

Methods	Dataset	Prec. (%)	Rec. (%)	F1. (%)	TEDS-Struct (%)
SPLERGE	SciTSR	99.3	98.9	99.1	-
TSRFormer	SciTSR	**99.5**	**99.4**	**99.4**	-
SPLERGE	SciTSR-COMP	98.8	98.0	98.4	-
TSRFormer	SciTSR-COMP	**99.1**	**98.7**	**98.9**	-
SPLERGE	PubTabNet	-	-	-	97.1
TSRFormer	PubTabNet	-	-	-	**97.5**
SPLERGE	In-house	85.4	82.3	83.8	-
TSRFormer	In-house	**95.1**	**95.3**	**95.2**	-

表 4：TSRFormer 与 SPLERGE 在多个数据集上的性能对比

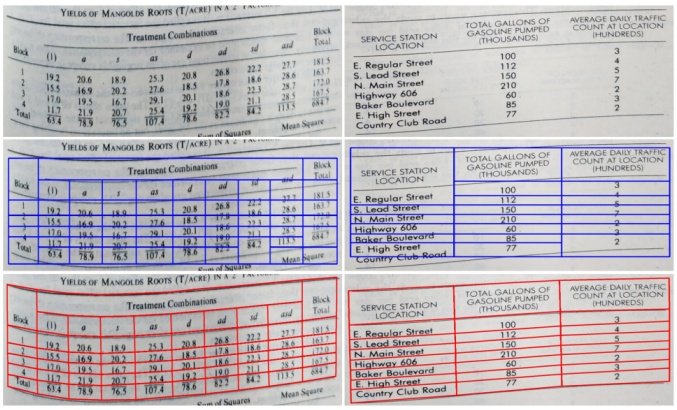


图 5：TSRFormer（红）与 SPLERGE（蓝）的可视化效果对比

在表 5 的消融实验中，研究员们将基于直接回归的 TSRFormer 与目前基于图像分割的最优方案 RobusTabNet 进行了对比。TSRFormer 与 RobusTabNet 均能处理倾斜或弯曲的表格。根据表 5 的实验结果，在更具挑战性的内部数据集中，相比 RobusTabNet，TSRFormer 的 F1-score 高出 2.9%。关于消融实验的其他细节，可见论文 [1]。

	SCNN	Aux-seg.	SepRETR	Cell Merging	F1. (%)	
Segmentation based		✓			83.5	
	✓	✓			90.0	
	✓	✓		✓	92.3	RobusTabNet
Regression based			✓		88.6	
	✓		✓		91.0	
	✓	✓	✓		**92.6**	
	✓	✓	✓	✓	**95.2**	TSRFormer

表 5：TSRFormer 与 RobusTabNet 在内部数据集上的对比，以及各模块的消融实验

图 6 中的可视化结果展示了基于直接回归方法的优势。对于图 6 这种单元格密集、弯曲且含有大面积空白区域的困难样本，基于图像分割的结果并不鲁棒，这使得后续的后处理模块难以提取出正确的分割线。而与之相反，基于直接回归思想的 TSRFormer 并不需要任何后处理模块，对表格中的数据和内容识别得更为精确。

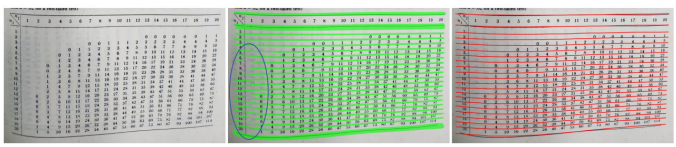


图 6：TSRFormer 与 RobusTabNet 的可视化结果对比

图 7 展示了 TSRFormer 在多个场景表格图像上的可视化结果，看到该方法对于大部分复杂场景表格的识别呈现高鲁棒性。

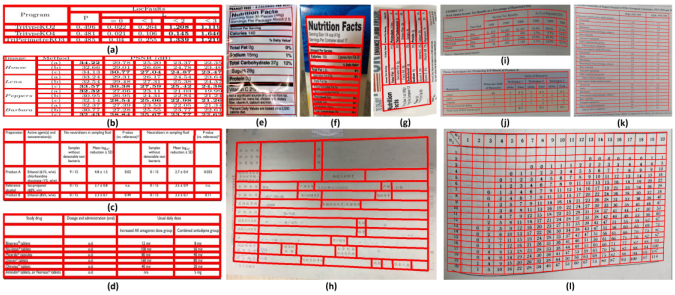


图7：TSRFormer 在各个数据集上的可视化结。(a-b)来自 SciTSR, (c-d) 来自 PubTabNet，(e-h) 来自 WTW，以及 (i-l) 来自内部数据集

未来的挑战

虽然 TSRFormer 在识别大部分场景的表格图像中取得了可喜成果，但要完全解决所有场景的表格结构识别问题道阻且长。主要问题在于，目前的算法只考虑了视觉图像单一模态的信息，而对于内容极为复杂的表格，例如单元格包含多行文字内容或存在极长且无实线的跨行跨列单元格，不仅需要利用图像信息，还需要充分理解图中文字的语义后，才能正确地识别表格结构。 此外，现有的方法仍然无法解析多层级的嵌套表格。微软亚洲研究院的研究员们将不断推进表格结构识别的性能，也欢迎同行共同交流、探索该领域更好的技术！

参考文献：

[1] Weihong Lin, Zheng Sun, Chixiang Ma, Mingze Li, Jiawei Wang, Lei Sun, Qiang Huo. TSRFormer: Table structure recognition with Transformers. In ACM Multimedia, 2022.

[2] Chris Tensmeyer, Vlad I. Morariu, Brian Price, Scott Cohen, Tony Martinez. Deep splitting and merging for table structure decomposition. In ICDAR, 2019.

[3] Chixiang Ma, Weihong Lin, Lei Sun, Qiang Huo. Robust table detection and structure recognition from heterogeneous document images. Pattern Recognition, 2023.

[4] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, Jifeng Dai. Deformable DETR: Deformable Transformers for end-to-end object detection. In ICLR, 2021.

[5] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, Sergey Zagoruyko. End-to-end object detection with Transformers. In ECCV, 2020.

[6] Zewen Chi, Heyan Huang, Heng-Da Xu, Houjin Yu, Wanxuan Yin, XianLing Mao. Complicated table structure recognition. arXiv:1908.04729, 2019.

[7] Xu Zhong, Elaheh ShafieiBavani, Antonio Jimeno Yepes. Image-based table recognition: Data, model, and evaluation. In ECCV, 2020.

[8] Rujiao Long, Wen Wang, Nan Xue, Feiyu Gao, Zhibo Yang, Yongpan Wang, Gui-Song Xia. Parsing table structures in the wild. In ICCV, 2021.

[9] Sachin Raja, Ajoy Mondal, CV Jawahar. Table structure recognition using top-down and bottom-up cues. In ECCV, 2020.

[10] Liang Qiao, Zaisheng Li, Zhanzhan Cheng, Peng Zhang, Shiliang Pu, Yi Niu, Wenqi Ren, Wenming Tan, Fei Wu. LGPMA: Complicated table structure recognition with local and global pyramid mask alignment. In ICDAR, 2021.

[11] Hao Liu, Xin Li, Bing Liu, Deqiang Jiang, Yinsong Liu, Bo Ren, Rongrong Ji. Show, read and reason: Table structure recognition with flexible context aggregator. In ACM Multimedia, 2021.

[12] Xinyi Zheng, Douglas Burdick, Lucian Popa, Xu Zhong, Nancy Xin Ru Wang. Global table extractor (gte): A framework for joint table identification and cell structure recognition using visual context. In WACV, 2021.

科研第一线

微软亚洲研究院研究员研究工作获“ACM SenSys Test of Time Award”（时间检验奖）

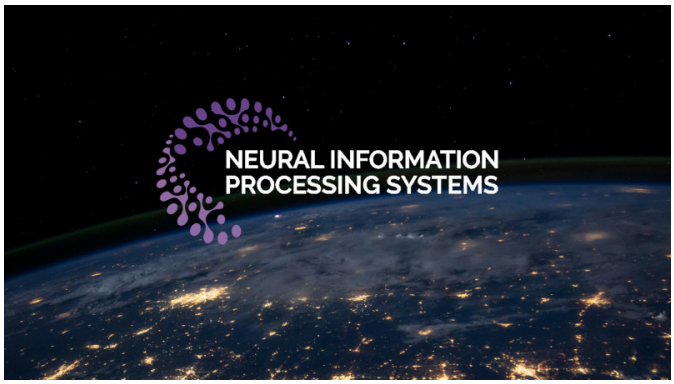
微软亚洲研究院高级研究员梁傑然此前的研究工作“Design and Evaluation of a Versatile and Efficient Receiver-Initiated Link Layer for Low-Power Wireless”荣获了国际移动计算和感知领域顶级会议 ACM SenSys 2022 时间检验奖（Test of Time Award），得到了研究界的肯定。正如 ACM SenSys 大会对这项工作所做的评价：“2010 年，该研究工作率先实现了在低功率无线通讯中利用同步传输在 MAC 层的优势，来突破低功率无线电的极限。在过去 12 年的时间里，这项成果为许多物联网和嵌入式系统奠定了无线通讯协议的基础。”

扫描二维码了解更多信息

文档基础模型引领文档智能走向多模态大一统

自 2019 年以来，微软亚洲研究院在文档智能领域进行了诸多探索，开发出一系列多模态任务的文档基础模型 (Document Foundation Model)，包括 LayoutLM (v1、v2、v3) 、LayoutXLM、MarkupLM 等。这些模型在诸如表单、收据、发票、报告等视觉富文本文档数据集上都取得了优异的表现，获得了学术界和产业界的广泛认可，并已应用在包括 Azure Form Recognizer、AI Builder、Microsoft Syntex 等在内的微软产品中，赋能企业和机构的数字化转型。

扫描二维码了解更多信息



NeurIPS 2022 | 微软亚洲研究院精选论文

作为目前全球最负盛名的人工智能盛会之一，NeurIPS (Conference on Neural Information Processing Systems) 在每年年末都是计算机科学领域瞩目的焦点。被 NeurIPS 接收的论文，代表着当今神经科学和人工智能研究的最高水平。2022 年 NeurIPS 大会于 11 月 28 日至 12 月 9 日举行，大会共收到 10411 篇有效投稿，其中 2672 篇获接收，最终接收率为 25.6%。相比 2021 年，投稿数量继续增加。在本届大会中，微软亚洲研究院也有诸多论文入选，内容主要涵盖人工智能五大热点话题：人工智能走向大一统、计算机理论、赋能产业界的人工智能、负责任的人工智能、人工智能赋能内容与设计生成。

扫描二维码了解更多信息

微软亚洲研究院理论中心前沿系列讲座回顾

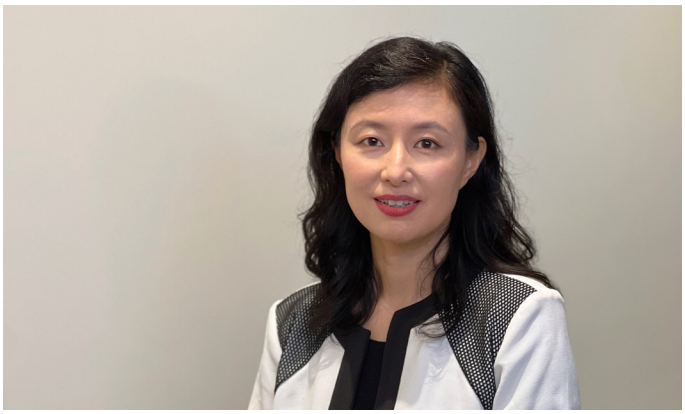
微软亚洲研究院理论中心前沿系列讲座，作为微软亚洲研究院的常设系列直播讲座，将邀请全球站在理论研究前沿的研究者介绍他们的研究发现，主题涵盖大数据、人工智能以及其他相关领域的理论进展。通过这一系列讲座，我们期待与大家一起探索当前理论研究的前沿发现，并建立一个活跃的理论研究社区。欢迎对理论研究感兴趣的老师和同学们参与讲座并加入社区，共同推动理论研究的进步，加强跨学科研究的合作，助力打破 AI 发展瓶颈，实现计算机技术实质性发展！

扫描二维码了解更多信息

科学匠人 | 对话邱锂力：一起探索未知的科技之美

2018 年世界人工智能大会上，微软宣布成立微软亚洲研究院（上海）。成立至今，微软亚洲研究院（上海）都做了哪些研究，取得了怎样的进展？未来会重点投入哪些研究方向？有哪些人才引进的新计划？让我们对话微软亚洲研究院（上海）负责人邱锂力博士，一起了解微软亚洲研究院（上海）的成长步伐和未来规划。

2022 年 1 月，邱锂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，邱锂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。她也是全球为数不多同时拥有国际计算机学会会士 (ACM Fellow) 和电气电子工程师学会会士 (IEEE Fellow) 称号的华人学者。



微软亚洲研究院副院长邱锂力

Q：您为何选择回国并加入微软亚洲研究院?

邱锂力：一直以来我对微软研究院都有很深厚的感情。读博期间，我曾在微软雷德蒙研究院实习，这让我有机会与之前拜读过的很多优秀论文的作者近距离接触、交流。我发现他们不仅治学严谨而且非常平易近人，从他们身上我学到了很多，之后我一直激励自己要成为他们那样的人。实习期间我参与的项目在领域内的顶会上获得了不错的成绩，让我倍受鼓舞。作为实习生我还有幸受邀到比尔 • 盖茨家做客，盖茨给了我们很大的鼓舞，我也因此更加了解到微软研究院是一家真正致力于推动科学研究进步的机构。

2001 年 1 月毕业后我就加入了微软雷德蒙研究院，担任系统和网络组研究员。这里有世界顶级的科学家团队，研究员们还可以自由选择研究方向，这样的学术环境给年经人创造了极大的提升空间。虽然我后来去了得克萨斯大学奥斯汀分校，但与微软在全球各地的研究院仍保持着紧密的合作。所以，当 2021 年有机会加入微软亚洲研究院时，我欣然答应，并希望可以在国内做一些具有全球影响力的、有意义的工作。

Q：请介绍一下微软亚洲研究院（上海）成立至今的发展情况。

邱锂力：自 2019 年 5 月微软亚洲研究院（上海）正式揭牌成立至今的三年多时间里，我们陆续吸引了系统、网络、通信、感知和游戏等多个领域的人才。现在的微软亚洲研究院（上海）有 20 余名全职员工，以及来自国内外高校的 30 多位实习生。整个团队朝气蓬勃、充满活力，也有很多新想法，在与他们的交流中，我感受到了他们对科学研究的热爱，以及用科技改变人类社会的激情。

微软亚洲研究院（上海）完全延续了微软亚洲研究院包容、多元、自由、开放的文化，让我仿佛回到了学生时代。在这里我们可以自由地定义科研议题，也会定期和不定期地组织不同的学术讨论，分享前沿技术，探索创新突破点。工作之外，小伙伴们也是多才多艺，口琴、电子琴、吉他、架子鼓都玩得转，研究院采购了一批乐器，来支持团队的业余兴趣爱好，有音乐爱好的同事们还组成了微软乐队上海分队。尽管我加入微软亚洲研究院仅几个月的时间，但我已经充分感受到了这个大家庭的温馨。

Q：作为负责人，您对微软亚洲研究院（上海）的研究领域有哪些规划?

邱锂力：微软亚洲研究院（上海）的三大主攻研究方向分别是：AI、系统和网络。

AI 研究主要关注 AI for Health（智慧医疗）和机器学习算法。目前，我们在 AI for Health 上已取得一些不错的成绩，发表了不少顶会论文，与一些医院的合作也有了令人欣喜的成果。在此基础上，我们将进一步推进与各大医疗机构的交流与合作。尤其是，上海有众多一流的医院和医学院，给我们提供了充分的合作环境，有利于将 AI 应用到医疗健康领域。我们也将更深入地开发先进的机器学习算法，以支持 AI for Health 和其他更广泛的应用场景。

系统领域主要着眼于 System for AI 和 AI for System。AI 落地到真实生活中，除了好的算法和模型外，还需要强大的系统支持。在实际场景中，人们对系统的实时性、高效性、可靠性、绿色节能等有了更高、更新的要求。我们也发表了许多有影响力的论文，并且已经把这些技术转化到了微软的产品中，不断推动 AI 在系统中的应用。

网络方面我们将集中精力于 5G、6G 及无线感知的研究。我们计划探索毫米波、太赫兹，低轨道卫星通信，以及利用无线信号感知运动、温度、湿度以及人体健康状况的特征。

除此之外，我们还将加强与内外部的合作，不仅与微软内部各个研究组和产品团队进行合作，也会组织学术界的科研交流以及学术访问。同时，在与产业界的合作方面，我们会在已有的与医疗、金融、物流航运企业机构合作的基础上，发掘更多合作机会。

邱锂力（第二排右四）及上海团队与康奈尔大学教授、图灵奖得主 John Hopcroft（第二排左四）进行交流

Q：请介绍一下您所在的无线和系统领域的研究现状、前沿方向以及面临的挑战。

邱锂力：无线技术是一个非常基础且重要的研究领域，当前我们的智能设备离不开随时随地的网络连接，未来的无人驾驶则需要更强大的 5G、6G 网络的支撑。现阶段对无线技术的研究还有许多问题需要解决，我举几个例子：

第一，当前 5G 网络建设才起步，没有实施经验可循，部署过程复杂，经常出现难以理解的问题。而且每一个基站都有大量参数、多种配置，会出现各种不可预知的问题。怎样将 AI 算法与领域知识结合来解决问题，是当前的一个研究重点。

第二，网络每十年有一次更新换代，因此现在也需要着手 6G 网络的研究。我们认为太赫兹是实现 6G 的一种方法。但是太赫兹在频率上升时覆盖范围就会缩小，业界一直在探索新技术来保证太赫兹在较大范围内的稳定覆盖。同时 6G 与机器学习的结合会更紧密，需要探索如何用 AI 更好地支持 6G 网络。

第三，就是给 5G、6G 网络寻找合适的应用场景。在落地应用时，如何克服环境的影响，怎样更好地支持边缘计算、云游戏等应用场景，都需要深入研究。这些研究对实验设备和硬件条件都有较高的要求。幸运的是，在微软亚洲研究院的支持下，我们正在陆续引进顶尖的科研仪器，全力支持无线和感知领域的研究。

Q：目前，微软亚洲研究院（上海）有哪些与 AI、无线以及感知等技术应用场景相关的研究项目？

邱锂力：我们一直在进行跨领域交叉融合的研究。其中一项研究是与上海的一家医院合作，对唇腭裂病人的发音问题进行研究。针对这个问题，当前的方法需要专业医生诊断后再做手术治疗，但现在这方面的医学专家相对较少，所以我们希望借助机器学习算法可以达到自动诊断的效果，让更多人获益。

我们还与一家医院合作，利用 EEG 信号来诊断婴儿是否患有癫痫病。成年人身上的 EEG 信号通常有 20 多个频道，对癫痫的检测相对容易，但是婴儿只有 4 个频道，且波动较大，检测比较困难。我们的研究才刚开始，数据量也极其有限，并且数据质量较低，所以我们不仅要开发新的癫痫检测算法，还要用算法对数据进行加工处理。

以上提到的这些研究都还在初始的阶段，还需要更详细的规划和扎实的研究。我们诚挚邀请更多志同道合的伙伴加入微软亚洲研究院（上海），和我们一起努力，用科技改变生活！

感谢大家对微软亚洲研究院的支持，我们会以此为动力，脚踏实地做更多更具影响力的科研工作。

Q：对于想要加入微软亚洲研究院（上海），从事计算机和网络科研的各路英才，您有怎样的期待？

邱锂力：无论是 AI、系统还是网络，现阶段所有领域都还需要更深入的探索，尤其是无线感知、6G 的研究才刚刚起步，因此我们需要更多的人才与我们共同研发新技术。

我们希望加入的伙伴拥有过硬的计算机、编程等专业实力，同时还要有很强的学习能力，可以不断吸收新技术和跨领域知识，因为计算机科学的发展日新月异。此外，除了硬实力，想要做科研还需要有毅力和耐心，可以在漫长而枯燥的研究过程中持续抱有好奇心，勇于直面如家常便饭一般的科研失败。只有对科学充满好奇，并希望用科技推动社会发展，才会真正享受科研的过程。

借此机会，我诚挚邀请各位有志于计算机科研的伙伴加入微软亚洲研究院（上海），尤其是系统和网络领域的人才，虽然这些领域没有 AI 研究火热，但却是现代社会必不可少的技术基础设施，同时由于领域内参与人数相对较少，所以也更容易脱颖而出。

科学匠人 | 杨玉庆：AI 系统研究需要硬件和软件的“双向奔赴”

从微电子、集成电路到系统架构、软件设计，再到 AI 模型、算法研究，从复旦大学到微软（亚洲）互联网工程院，再到微软亚洲研究院（上海），现任高级研发经理的杨玉庆如何转变不同角色？有着软硬件跨领域研究背景的他，对 AI 研究有什么不一样的理解？作为微软亚洲研究院（上海）最早一批研究员，又对上海研究院有怎样的感受和期望？让我们一起走近杨玉庆的“立体”研究世界。

从本科到博士，杨玉庆在复旦大学学的是微电子专业，职业生涯的起点做的是和数字电路、SoC（系统级芯片）架构设计相关的工作，可谓是一名实打实的硬件工程师。然而，如今的他却带领微软亚洲研究院的团队在人工智能和分布式系统领域取得了多项重要的研究成果，他和团队参与的 OpenPAI、NNI、nn-Meter、SparTA 等软件工程项目均获得了学术界和工业界的肯定。

微软亚洲研究院（上海）高级研发经理杨玉庆

软硬件协同，“全栈式”视角看待系统与 AI 研究

谈及最初的专业选择，杨玉庆认为自己更多的是“顺势而为”。当时正是通信技术从 2G 向 3G，再到 4G 的高速发展阶段，市场对集成电路有了更多的需求，相应的研究从通信制式、算法到电路的实现和通信标准等也呈现井喷态势。半导体行业蕴藏的新机会为个人的发展提供了施展拳脚的空间，也让杨玉庆看到了更多可能。因此，他持续钻研无线通信系统的集成电路加速和数字信号处理方面的研究，并在之后几年的工作中深耕半导体行业。

随着研究的深入，杨玉庆对数字电路产生了新的认知。他认识到，无论是数字电路还是架构设计，本质上都在寻求一种权衡，这就需要超越电路本身，将视角从底层的硬件向上层的系统扩展，归根究底在于用户需求不只局限在电路层面，也就是所谓的“工夫在诗外”。

顺着这一思路，杨玉庆的职业规划开始发生了变化。2017 年，杨玉庆加入了微软（亚洲）互联网工程院，工作方向从底层系统的架构设计逐渐走向系统软件设计，并且开始站在开发者和用户的视角看问题。“系统的优化不能只停留在硬件或软件本身，而是要从更广阔的视野去理解。从底层你会更清楚地理解模型会以怎样的形式被计算硬件执行，从而做出有针对性的优化。反之，从上层你会看到系统优化中最关键的部分，包括它的发展趋势。微软给我提供了这样的机会，让我可以从‘全栈式’的视角来看待这些研究工作。”杨玉庆说。

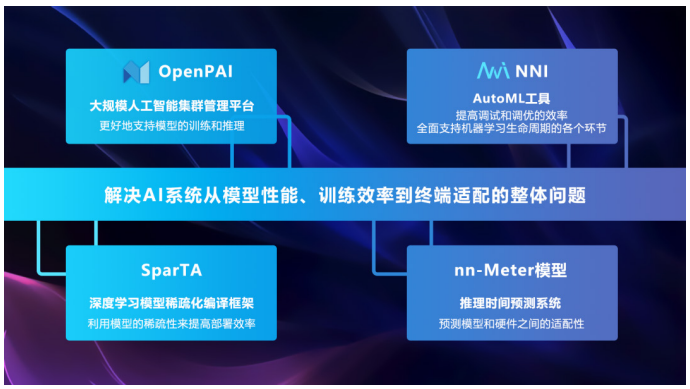
汇溪成流的 AI 基础设施研究，全过程提高 AI 开发效率

2019 年，微软亚洲研究院（上海）正式成立，杨玉庆成为了第一批员工。“一方面，我可以更好地兼顾家庭；另一方面，研究院希望能够与更广泛的行业伙伴合作，推动 AI 技术的真正落地，这一点对我有极大的吸引力。”杨玉庆说，“作为上海研究院的第一批员工，我们还有点像一个‘创业团队’，大家从头开始规划做一件事情非常有趣。”

在微软亚洲研究院（上海），杨玉庆的研究工作聚焦于如何在深度学习模型越来越大时提升计算效率，让模型能够更好地在各种终端落地，得到更广泛的应用。广受业界好评的 AutoML 工具 NNI、大规模人工智能集群管理平台 OpenPAI、推理时间预测系统 nn-Meter 模型，以及深度学习模型稀疏化编译框架 SparTA 等都在沿着这条研究主线向前推进。

杨玉庆表示，在设计系统来支持深度学习或 AI 模型时，它们不是由独立的点构成的，解决的也不是一个个独立问题，而是需要从整体的、系统化的角度进行研究。例如，一个模型从训练到部署，再到最终用户的多次使用，其中多层级的调用都会影响模型的效率、性能，而微软亚洲研究院的这四个项目就是先通过每个点的研究逐个解决不同环节出现的问题，再由点连成线，形

成面，解决 AI 系统从模型性能、训练效率到终端适配的整体问题。



首先，在 AI 模型训练时，基本上都是基于集群训练，在这个过程中要充分计算 GPU 或硬件集群如何更好地支持模型的训练和推理。OpenPAI 平台的发布就是为了解决这个问题，它支持多种深度学习、机器学习以及大数据任务，可提供大规模 GPU 集群调度、集群监控、任务监控、分布式存储等功能，且用户界面友好，易于操作。

其次，AI 模型的整个生命周期中充满了大量的迭代任务，开发人员需要不断调优，包括架构调优和参数调优，这并不是一次性的任务，而是需要大量的调试才能让模型实现更好的性能。AutoML 工具 NNI 可提高调试和调优的效率，对机器学习生命周期的各个环节全面支持，包括特征工程、神经网络架构搜索（NAS）、超参调优和模型压缩在内的步骤，都可以使用 AutoML 算法完成。同时，NNI 2.0 还加入了对 “探索性训练” 框架 Retiarii、基于掩码的模型压缩加速工具的支持，提供了利用 Python 发起实验（预览功能）与多种算力混合训练的能力，并简化了自定义算法的安装方法。NNI 是目前 GitHub 上最热门的 AutoML 开源项目之一。

第三，模型基本确定以后，还需要适配到不同的终端设备上，这就会引出部署效率的问题。因为在大模型训练时，研究人员更关注的是模型的性能，单个模型往往具有高达万亿级别的参数，但端侧的内存、算力和功耗的限制对深度学习模型的大小和推理延迟提出了更高的要求，这就需要对模型 “瘦身”，在性能和模型大小上做出平衡。 为了解决这一问题，除了 NNI 的大模型压缩功能外，微软亚洲研究院的研究员和工程师们又研发了 SparTA，从而利用模型的稀疏性来提高部署效率。

最后，在端侧部署 AI 模型，还要考虑硬件的多样性。针对不同的设备，每个人对模型都有不同的理解，比如有人认为 CNN 更好，有人喜欢 Transformer 架构，那么在硬件适配时就会有不同的取舍。相应地，硬件也会针对模型特点进行优化，致使同一个模型在 A 设备上效率很高，在 B 设备上则会很低。这就需要以自动化的方式解决硬件多样化问题，否则适配时间和成本可能超乎想象。这其中关键的一点就是理解模型在硬件上的表现，再进行相应的优化。推理时间预测系统 nn-Meter 模型能预测深度学习模型在不同边缘设备上的推理延迟，也就是对模型和硬件之间的适配性进行预测。

“AI 模型的训练、部署到应用需全栈、全生命周期、多层次、系统化的思考，并不是一两个项目就能解决的问题。” 杨玉庆说，“在微软亚洲研究院的这几年让我体会最深的是，之前我们基于个人的认知和好奇心，从一个点上出发进行的研究和尝试，最终就像不同的支流一样，慢慢汇合成了一条让系统可以更好支持 AI 模型的江河，为研究者和开发者们提高生产效率。”

与此同时，杨玉庆认为每个项目的成功，都离不开团队成员的齐心协力，其中最重要的一点就是所有人都有一致的愿景。在项目开始时，大家清晰沟通、明确分工，只有确保每个人都能 “believe in”（相信与信任），才能保障后续工作的顺利进行。

以数据为脉络，赋能百行千业

经过三年多的建设，微软亚洲研究院（上海）的研究领域已稳定成型，杨玉庆所在的系统组主要为计算密集型任务开发高效技术，包括大规模 AI 模型、实时视频流处理，主要实现模型、编译器和系统平台级工具的整体优化。此外，杨玉庆和团队成员也关注视频流传输的跨级优化，包括实时视频会议、云游戏等，以及医疗行业的多模态数据处理，如心电图、脑电图等非结构化数据的检索和挖掘工作。

现阶段，微软亚洲研究院（上海）的所有业务基本以数据为核心脉络，从数据获取，到借助算法对数据进行处理和挖掘，再到系统层面的数据存储、搜索、查询，围绕着更普适的、多模态的数据形成完整的闭环工作链。“大数据时代，任何数据以及数据与数据之间的关联都是有价值的。但在实际应用中真正能够被检索、被人们所处理的数据只是冰山一角，大量非结构化、稀疏的数据好似沉在水下的冰山，并没有被真正挖掘出来。如果能利用算法、系统针对特定的场景，比如在医疗健康领域去构建医生所认可的、医学意义上的数据关联和理解，那么将能释放出医生和医疗研究者的更多潜能。这样的方式也能推广到更多的行业中，对整个社会都将有巨大的意义。“ 杨玉庆说。

随着近年来交叉学科研究课题的逐渐增多，杨玉庆认为，AI 需要跳出传统的计算机科学领域，与更广泛的行业研究建立关联，从纯粹的 AI 研究，到把 AI 变成一种能力和工具与其他学科和领域共同推进研究边界，从定义问题开始就像双螺旋一样，缠绕攀升，进而赋能百行千业。“这种赋能是合作式赋能（collaborative empowerment），与单方面提供服务或工具的销售式赋能有很大的区别。” 为此杨玉庆也欢迎各类学科背景、志同道合的伙伴加入微软亚洲研究院（上海）这个年轻、充满朝气与活力的团队，一起用科学研究改变世界，让世界更美好！

实习派 | 姜雪：我适合做科研吗？我在微软亚洲研究院找到了答案

“我适合做科研吗？” 带着这个问题，在中国传媒大学通信与信息系统专业读研一的姜雪来到微软亚洲研究院寻找答案。当时的她还没有丰富的科研经验，对深度学习的了解也有限。

在研究院的“沉浸式”科研环境中，姜雪在多媒体计算组从头探索“基于深度学习的音频编码”这一新方向。一年半的实习成果丰硕，她在包括 ICASSP、INTERSPEECH 等在内的语音研究领域顶级会议以第一作者的身份发表了多篇学术论文。

2022 年 9 月，姜雪选择继续攻读博士学位，这段实习经历带领她找到了开头问题的答案。“不管是 mentor 们还是小伙伴们，我都在他们身上看到了自己想成为的样子。” 姜雪说。

从科研小白到顶会文章作者，背后是姜雪工作日里风雨无阻的早出晚归，也是她与 mentor 亲密无间的合作。自我定义为 “非天赋型选手” 的她证实了 “一直全身心投入一件事，总会有收获”。

生活中的姜雪

在微软亚洲研究院从头探索新方向

姜雪与微软亚洲研究院的缘分始于她在中国传媒大学的导师推荐。中国传媒大学张远教授认为，微软亚洲研究院的科研氛围非常好，学生能够在此接触前沿技术、接受专业指导，因此她很建议学生来到研究院实习。此前，姜雪的同门师姐郑澄瑜也曾在研究院实习且收获颇丰。在导师的全力支持下，姜雪开始了在微软亚洲研究院的科研之旅。

初入研究院，姜雪用“一张白纸”来形容自己——对深度学习的了解有限，也没有完整经历过全流程的科研工作。2021 年恰逢基于深度学习的音频编码蓬勃发展，微软亚洲研究院也开始在该领域进行持续探索，多媒体计算组的很多项目都围绕实时通讯场景下传输音频中出现的各种问题展开。

遇上全新的方向，姜雪开始从头探索。在微软亚洲研究院高级研究员彭秀莲和研究员薛华颖的指导下，姜雪聚焦音频通讯场景下的编解码问题，研究在同等码率下，如何使解码出来的音频质量更好。当时，这个方向在领域内尚属蓝海。

创新总是与挑战并存。在没有源码、参考文献也很少的情况下，不管是写代码还是搭框架，一切都需要从零开始。姜雪和研究院的新方向一起成长，一步步解锁科学研究新地图。目前，在多媒体计算组的努力下，音频在 3kbps 下已达到 near-transparent 的听觉质量，也最先在实时通讯的场景下实现了 1kbps 还能达到远高于 Lyra 3kbps 的听觉质量。

姜雪（右四）和小伙伴们

“姜雪是我们这个方向探索的中坚力量，和我们一起完成了许多重要工作。” 彭秀莲这样评价姜雪的贡献。姜雪参与了抗丢包恢复算法、回声检测算法等多个 Teams 语音通信项目的研究，后期还设计实现了基于 TFNet 框架的新一代实时语音编码器，性能远超越传统语音编码器，并以此展开了超低码率及 scalable 语音编码器研究。在与 Teams 团队合作优化的过程中，姜雪也做了很多模型训练的工作。与美国 Teams 产品团队的协作也让姜雪收获颇丰，产品团队立足产品工程的视角，为科研提供了新的思路。“真

正感受到了产品与科研的相互促进。" 姜雪说。

实习的一年半转瞬即逝，姜雪硕果累累，在包括 ICASSP、INTERSPEECH 等在内的语音研究领域顶级会议以第一作者的身份发表了多篇学术论文。姜雪坦言，自己最初并未想好是否读博，这段实习经历直接影响了她的决定。"我适合做科研吗？" 回想起一年半之前的问题，姜雪找到了答案。

亦师亦友，绽放女性科研光芒

从学术小白到独当一面，mentor 们在姜雪的成长中扮演着重要角色。实习期间，姜雪的两位 mentor 彭秀莲和薛华颖扮演着亦师亦友的角色。从学术指导到情绪辅导，她们关心实习生方方面面的成长，也作为榜样引领着姜雪的科研之路。而这个全女性的组合，也绽放着她们独特的科研光芒。

彭秀莲关注视频编码领域，在编码方向上非常有经验。在姜雪看来，秀莲是一个细致的 "实干派"，会基于自己的经验提出方向和目标，不断提出一些可以尝试的地方。从初期调研，到写代码完成实验，再到分析实验结果，每一次进展间都穿插着密集的讨论，二人共享着探索成功后的兴奋感与成就感。

由于秀莲此前也未曾涉猎音频编码领域，二人在新方向的探索初期难免遇到一些问题。秀莲会根据自己的经验，耐心帮助姜雪分析当前问题和解决思路，姜雪提起一次令她印象深刻的瓶颈期，尽管她参考了过往相关论文的方法，解码出来的音频质量却仍不理想。姜雪疑惑：明明感觉所有的地方都很合理，为什么结果就是不对呢？是解码端不够强大，还是量化做得不合理？在秀莲的建议下，她重新阅读了所有相关论文，仔细对比了不同论文输入音频数据的处理细节，最后将问题定位到音频数据的分帧步长。做出调整后，问题迎刃而解。"她真的很牛"，姜雪反复赞叹。

至今，姜雪仍对遇到困难时与秀莲一起头脑风暴的场景记忆深刻。秀莲站在白板前梳理难点的身影让姜雪觉得"做科研的女生真的很酷，我也想成为这样的人"。

姜雪与 mentor 彭秀莲（右）、薛华颖（左）合影

另一位 mentor 薛华颖则是姜雪心中"最有能量的姐姐"，从如何使用编码工具到如何做好学术分享，薛华颖亲手带她推开学术的大门。在共同制定研究方案、分析问题、验证结果的过程中，研究思路在思维的碰撞中迸发。

作为女性研究员，华颖对情绪的感知也十分细腻。科研进展缓慢时，姜雪初期容易陷入低落情绪中。华颖敏锐地察觉到姜雪的焦虑，主动开导姜雪静下心来分析问题。两人时常进行朋友般的聊天。华颖将自己定位为忠实的聆听者，"我会先以同理心开导她，表示感同身受，其次是鼓励以平常心对待，化焦虑为动力，最后是跟她一起直面瓶颈。我们会跳出来去看一些别人的工作，试图跳出当前的思维困境。"

除了科研能力的提升，姜雪的另一大收获便是培养了稳健的技术心态。"做科研就是这样，不可能一直很顺，翻过一个坎之后一定会是一个瓶颈期。只需要关注自己问题的本质，别的不用考虑太多。"

姜雪将学术路上遇到的导师视为自己的榜样。在学校，导师张远是姜雪的科研启蒙。她关注行业前沿，并鼓励自己的学生多探索，对图像、编码等每一个领域都有所了解，打好基础再选择自己感兴趣的领域。在秀莲和华颖的影响下，姜雪也越来越坚定对自己科研想法的信心，持续探索、耐心调整。导师们对科研的极致追求也激励着姜雪更加努力，未来，她希望自己也能成为 "科研品味很好、能主动有一些新发现" 的女性研究员。

一直全身心投入一件事，总会有所收获

在科研方面，姜雪从不认为自己是"天赋型选手"，但她相信天道酬勤。无论是本科四年保持综测第一，还是一年半发表多篇论文，她将这些都归为"习惯性努力"带来的收获。

姜雪就读于距离微软亚洲研究院 30 公里远的中国传媒大学。从研一下学期开始，为了避免八通线脚不沾地的早高峰，她会在大多数工作日的七点多钟出现在地铁站，在地铁上听听歌放空大脑。在研究院的一天里，她或是在开会，或是在做实验。晚饭后回到学校，她还会前往实验室，总结自己一天的收获。姜雪认为自己最大的优势就是一旦专注某件事，就会全身心投入，秀莲和华颖也不约而同地用刻苦勤奋、坚韧不拔来形容她。

科研占据了姜雪生活的绝大部分时间，但她并不因此而疲惫。姜雪喜欢简单的生活，也越来越习惯在研究院形成的科研节奏。读博期间，她还打算保持这样不松不紧的节奏，"否则再想捡起来就比较困难了"，姜雪说。在闲暇时间里，姜雪还会和学校实验室以及研究院的伙伴们一起约饭、逛街、运动。未来她也计划在继续全身心投入科研的同时多多参加活动，丰富自己的读博生活。

回顾自己在微软亚洲研究院的时光，姜雪充满感激。她也希

望未来能延续和研究院的缘分，继续音频 / 语音 +AI 领域的研究，带着在此培养的科研能力和技术心态继续自己的研究旅程。

生活中的姜雪

Mentor 寄语

彭秀莲
微软亚洲研究院
高级研究员

在这一年多的时间里，姜雪在 MSR Asia Media Computing 组基于深度学习的音频编码方向做出了非常优秀的工作。她以第一作者的身份发表了多篇文章，将高质量的实时语音编码推到了很低的码率，并不断地尝试更多的突破。姜雪身上所体现出来的扎实的态度，勇于探索和永不放弃的精神给我留下了很深的印象，希望未来她继续努力不断探索，在音频 / 语音 +AI 领域做出更大的成就，成为一个杰出的女性研究员！

薛华颖
微软亚洲研究院
研究员

在一年半时间里，姜雪在 MC 组做出了一系列优秀的研究工作，并在包括 ICASSP、INTERSPEECH 等在内的语音研究领域顶级会议以第一作者的身份发表多篇学术论文。她参与了组里多个 Teams 语音通信项目， 包括抗丢包恢复算法、回声检测算法的研究。后期作为主力成员， 主导了深度学习的语音编码器算法研究。

她设计实现了基于 TFNet 框架的新一代实时语音编码器，性能远超越传统语音编码器，并以此展开了超低码率及 scalable 语音编码器研究。她对科研的热情以及无畏科研困难的勤奋刻苦的态度令我印象深刻，祝愿姜雪在未来的研究道路中更加自信地前行，做出更有影响力的工作，成为该领域杰出的女性研究员！

相关阅读

扫描二维码查看文章

星跃重洋 | 刘国栋：非典型理工男的科研“旅”记

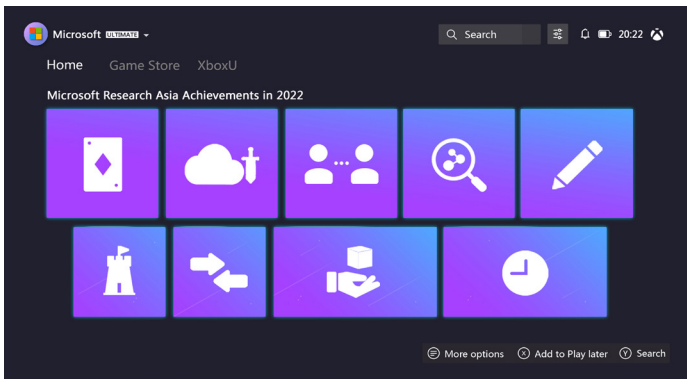
通过 “星跃计划” 在微软亚洲研究院实习一年，中国科学院计算技术研究所博士生刘国栋在微软亚洲研究院主管研究员苗又山、微软雷德蒙研究院高级研发工程师 Saeed Maleki 两位 mentor 的指导下，围绕加速深度学习模型的训练进行着科研探索。

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。

微软亚洲研究院 2022 年度成就图鉴

亲爱的微软亚洲研究院服务器的玩家们：

恭喜你，通关 2022 赛季中的所有挑战！本赛季，你在微软亚洲研究院服务器人工智能赛道取得的成就奖励现已达成 1000/1000 * 的完美结局！ 2023 赛季版本即将更新，在此，我们希望你带上 2022 年的经验与进展，点亮跨越计算机领域内外的更多版图，开启全新的冒险！



扫描二维码查看文章

对话微软 CTO Kevin Scott：人工智能的未来之路

如今，从为软件开发人员生成代码到为图形设计师绘制草图，由大型语言模型驱动的人工智能系统正在改变人们的工作和创作方式。微软执行副总裁兼首席技术官 Kevin Scott 认为，未来，无论是帮助应对气候变化及儿童教育等全球挑战，还是彻底变革医疗健康、法律、材料科学甚至科幻小说等领域，这些人工智能系统的复杂度和规模都将继续增长。

近期，Kevin Scott 就人工智能对知识工作者的影响以及人工智能下一步发展等话题分享了他的看法，核心观点包括：人工智能大模型和生成式人工智能的发展将继续提高人们的生产力、创造力和满意度；人工智能将助力实现科学突破，并帮助世界解决一些重大挑战；随着人工智能模型的平台化，以及微软继续以负责任的方式为客户推动人工智能的进步，云、基础设施投资和以极其负责任的方式发展人工智能变得至关重要。

下面就让我们一起来看一看 Kevin Scott 对人工智能未来之路的展望吧！

Q：在你看来，今年人工智能领域最重要的进步有哪些？

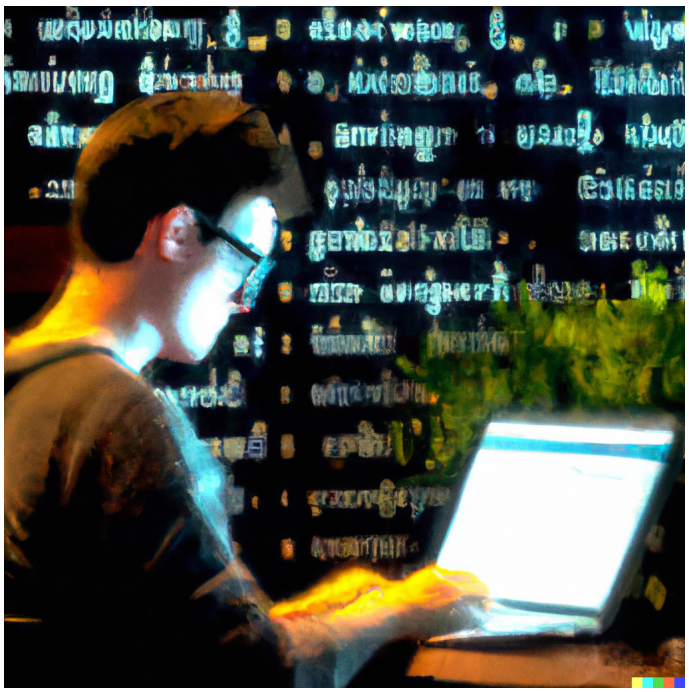
Kevin Scott：当我们进入 2022 年时，我想人工智能领域的几乎每个人都期待在接下来的 12 个月左右的时间里能发生引人瞩目的大事。现在，这一年马上就要结束了，即使当初的期望颇高，回顾我们在人工智能领域各个方面取得的创新规模依旧令人兴奋。科研人员和其他同行们为推进前沿技术所取得的成果，仅在几年前都几乎是无法想象的。而几乎所有这些都是人工智能大模型飞速发展的结果。

今年有三项成果让我印象最为深刻。首先是 GitHub Copilot 的发布，这是一个基于大型语言模型的系统，它能将自然语言提示词转化为代码，给开发人员的工作效率带来了非常积极的影响。未来的发展将在很大程度上取决于我们编写软件的能力，因此 GitHub Copilot 史无前例地让更广泛的人群可以拥有编码技能，这一点非常了不起。

第二个是 DALL · E 2 等生成式图像模型广受欢迎且变得更易使用。素描、绘画以及掌握所有的平面设计、插图和艺术工具都需要相当高超的技能。像 DALL · E 2 这样的人工智能系统尽管不能把普通人变成专业的艺术家，但它给了很多人视觉表达的能力，一种他们从未想过自己会拥有的全新超能力。

我们还看到，人工智能模型变得越来越强大，并为其所要解决的问题带来了更多实质性的收益。纵观今年整个科技行业，我认为蛋白质折叠方面的研究非常出色，包括微软与华盛顿大学蛋白质设计研究所大卫·贝克实验室利用 RoseTTAFold 所做的项目，以及利用一系列先进的人工智能技术帮助其开展变革性的工作。任何能对科学和医学有增效作用的事情对世界都是有益的，因为我们面临的最大、最棘手的问题就在这些领域中。

2022 年是一个令人印象深刻的科技大年。我认为明年会更好。



Q：你认为接下来几年，人工智能技术在哪些方面会产生最大的影响？

Kevin Scott：我可以很有把握地说，2023 年将会是人工智能领域有史以来最激动人心的一年。之前我也曾真心实意地相信 2022 年是有史以来最令人激动的一年。创新的步伐一直在快速向前。

前面我已经谈到了 GitHub Copilot，但这也只是人工智能大模型潜在能力的冰山一角，如果把同样的理念外推到各种不同的

场景中，那么我们就可以帮助到编程以外的其他脑力劳动。整个知识经济将会见证人工智能如何帮助人们解决工作中的重复性问题，并让工作更愉快、更有成就感。 这将适用于几乎所有的工作，比如设计新分子来创造药物、根据 3D 模型设计制造 “配方”，或者是单纯的写作和编辑。

例如，我一直在使用一个我基于 GPT-3 为自己构建的实验系统，来帮助我完成一件从十几岁起就想做的事情——写一本科幻小说。我的笔记本上写满了我为设想中的书编写的概要，描述了书中的大致内容和这些故事将发生在什么样的宇宙中。如果我用传统方式写书，一天能写 2000 个字，我就会觉得自己很不错了。但有了这个工具，我就可以打破僵局，我经常可以一天写出 6000 个字，这对我来说已经很多了。与之前相比，这是一个更加充满活力的过程。

这就是 “一切皆有副驾驶（copilot for everything）”的梦想——当你做任何类型的认知工作时，都会有一个 “副驾驶” 坐在你身边，它不仅可以帮助你完成更多的工作，而且还能以新颖有趣的方式增强你的创造力。

Q：这种生产力的提高显然也会提升满意度。为什么这些工具能给工作带来更多乐趣?

Kevin Scott：我们所有人都需要使用工具来完成工作。其中一些人非常乐意获得、掌握这些工具，并且弄清如何以超级有效的方式用它们来做事。在很多情况下，人们已经有了全新、有趣且从根本上比以前更有效的工具。我们做过一项研究，发现使用无代码或低代码工具对用户的工作满意度、总体工作量和员工士气产生了 80% 以上的积极影响，特别是对那些处于相对早期阶段的工具，这是一个巨大的好处。

对于一些员工来说，这实际上是在强化工作的核心流程，它会让你加速。就像穿着一双更好的跑鞋去跑步或参加马拉松。我们发现这正是开发人员使用 Copilot 时的体验，据他们反馈，Copilot 可以帮助他们保持心流状态，并且在面对曾经看起来枯燥重复的任务时依然头脑清醒。当人工智能工具可以帮助人们消除工作中的苦差事，也就是那些重复的或令人讨厌的或妨碍他们做真正喜欢的事情的工作，毫无疑问这会提高满意度。

就我个人而言，这些工具让我可以比以前更长久地处于心流状态。创意流的天敌是分心和思维停滞。例如，当我不太清楚该如何解决下一个问题，或者下一问题是 “我得去查一下这个问题，我不得不从正在做的工作中切换出来，去解决一个从属性问题。” 这些工具越来越多地为我解决了这些从属性问题，我则可以一直保持在心流状态中。

Q：除了 GitHub Copilot 和 DALL • E 2 之外，人工智能还以其他方式出现在了微软的产品和服务中。那么下一代人工智能如何改进 Teams 和 Word 等现有产品?

Kevin Scott：这是一个人工智能不为人知的故事。迄今为止，人工智能带来的大部分益处都分散在 1000 种不同的地方，你甚至可能都没有意识到你获得的产品体验中有多少来自机器学习系统。例如，在 Teams 视频通话功能的系统中，所有这些参数都是通过机器学习算法学习的；音频系统有抖动缓冲器使沟通顺畅；屏幕上显示的模糊的背景效果也是机器学习算法在起作用。有十几个机器学习系统协同工作，才让我们的交流体验变得更加愉快。而整个微软公司的产品和服务都是如此。

我们已经将机器学习的应用从几个地方扩展到遍布不同产品的上千个场景，从 Outlook 电子邮件客户端的运作、Word 中的文本预测、必应（Bing）搜索的体验，到用户在 Xbox Cloud Gaming 和 LinkedIn 中看到的信息流是什么样的，无处不在的人工智能正在让这些产品变得更好。

过去两年发生了很大变化的一件事是，曾经我们需要为所有产品针对每项任务专门定制一个模型，现在一个模型可以用在很多地方，因为它们拥有了很强的泛化性。能够投资于这些随着规模扩展而变得更强大的模型，然后让所有构建在模型之上的东西同步受益于你所做的改进，这是十分了不起的。

Q：微软通过 AI4Science 和 AI for Good 等倡议持续推进人工智能的研发。人工智能领域最让你兴奋的是什么?

Kevin Scott：我们的社会现在面临的最具挑战性的问题都在科学领域。如何治疗那些难以治愈的复杂疾病？如何为下一场大流行病做好准备？如何为逐渐老龄化的人口提供负担得起的高质量医疗？如何帮助更多的孩子接受他们未来所需的技能教育？如何通过开发技术来抵消碳排放产生的一些负面影响？我们正在探索如何将人工智能中一些令人兴奋的发展成果用于解决这些问题。

这些基础科学应用中的模型具有与大型语言模型相同的规模扩展特性。当你建立一个模型，让它进入某种自监督模式，它就可以从模拟中学习，或者通过自身观察特定领域的能力来学习，然后得到的模型可以让你显著改变所应用领域的表现，无论你是在做计算流体力学模拟，还是药物设计的分子动力学研究。

这其中蕴含着巨大的机遇。这意味着我们能够找到更好的药物，意味着也许我们可以找到解决碳排放问题的新催化剂，意味着全面加快科学家和其他有着远大想法的人们努力解决全社会最严峻挑战的速度。

Q：计算技术和硬件的突破如何促进人工智能的进步?

Kevin Scott：我们在人工智能领域看到的几乎所有最新进展背后的根源，是我们验证了模型规模的重要性。事实证明，基于更多数据和更多计算能力训练出来的模型具有更丰富和更通用的能力。如果想继续推动这一进步——需要明确的是，目前我们还没有看

到扩大规模所带来的好处的边界，我们要做的是尽可能优化和扩展计算能力。

两年前，微软推出了第一台 Azure AI 超级计算机，在今年的 Build 开发者大会上，我曾说我们现在拥有多个超级计算系统，而且我们非常确信这些系统是当今全球规模最大、功能最强的 AI 超级计算机。我们和 OpenAI 使用这些基础设施来训练我们几乎所有最先进的大模型，其中包括微软的图灵（Turing）、Z-code 和 Florence 模型，以及 OpenAI 的 GPT、DALL · E 和 Codex 模型。最近，我们还宣布与英伟达（NVIDIA）的合作，打造一台结合了 Azure 基础设施和英伟达 GPU 的超级计算机。

这其中的一些进展就是通过使用越来越大的 GPU 集群实现大规模强力计算而取得的。然而，更大的突破或许在于软件层，它优化了模型和数据在这些巨型系统中的分布方式，既可以训练模型，又可以让这些模型为客户提供服务。如果我们希望将这些大模型作为人们可以用来创作的平台，那么它们就不能只被世界上极少数拥有足够资源来建造巨型超级计算机的科技公司所使用。

因此，微软对一些软件进行了大量投入，例如用 DeepSpeed 来提高训练效率，用 ONNX Runtime 来加速推理。这些软件针对成本和延迟进行了优化，能帮助我们让这些人工智能大模型更容易为人们所用，也更有价值。 我为我们的这些技术团队感到自豪，因为微软在这一领域确实处于行业领先地位，而且我们对所有这些成果都进行了开源，以便其他人也能够不断提升。

Q：与这些进步相伴的是人们对“人工智能将影响就业”的担忧。你如何看待人工智能和就业的问题?

Kevin Scott：我们生活在一个异常复杂和宏观经济历史性变革的时代，展望未来 5 到 10 年，我们需要全新的生产力形式，让所有人都能够继续享受进步。我们希望将这些人工智能工具打造成平台，人们可以使用这些平台来构建业务和解决问题。我们相信，这些平台可以让更多的人使用人工智能。而有了这些平台，我们就能够解决更多的问题，就会有背景更加多元的人们参与到技术的创造中来。

此前，人们需要大量的专业知识才能开始人工智能的实例化。但现在你可以调用微软 Azure 认知服务和微软 Azure OpenAI 服务，并在这些服务的基础上构建复杂的产品，而你不必是 AI 方面的专家，也不需要从零开始训练自己的大模型。

随着所有这些巨型人工智能系统的不断增长和演进，我想我们可以预期，这些进步将从根本上改变工作的性质，每个领域被影响的程度会有所不同，在某些情况下甚至还会创造出大量以前没有的新工作岗位。回顾过去可以看到，历史上重大的技术范式转变都伴随着相同的情况：电话、汽车、互联网。我认为就像这些例子一样，我们需要用新的方式思考工作和技能，并专注于确保我们有足够的人才且接受过培训，能够承担起真正关键的工作。

Q：与人工智能技术相关的另一个担忧是技术被滥用和误用的可能性。微软正在采取哪些具体措施来确保其人工智能工具和服务是以负责任的方式开发和使用的?

Kevin Scott：我们一直非常严肃地对待这个问题。微软的人工智能系统必须通过“负责任的人工智能（responsible AI）”流程，而且我们还在继续改进这个流程。我们会与一个由多学科专家组成的团队一起仔细审查正在进行的工作，确保我们充分了解可能发生的所有潜在危害，并尽可能降低它们的负面影响。例如，改进用于训练模型的数据集、部署限制有害内容生成的过滤器、集成拦截敏感主题查询的技术帮助防止不良行为者的滥用，或者应用可以返回更有用和更多样化响应和结果的技术。我们为人工智能系统制定的计划还包括在发布后尽快发现并减轻任何我们没有预料到的危害。

另一个非常重要的保护措施是有意识的迭代部署。我们所做的大部分工作都是针对具有广泛能力的模型。我们将这些模型托管在我们的云中，并通过 API 或我们的产品提供给公众。任何开发者都可以访问 API，但他们必须遵守服务条款才能使用，如果他们违反了服务条款，那么他们的访问权限将被取消。对于其他产品，我们可能会先向一些有明确用例设想的客户提供有限的预览版。与这些早期客户的合作将帮助我们确保负责任的人工智能保障措施能够在实践中发挥作用，以便我们能够在更大的范围内推广应用。我们坚信安全和责任是重要的，希望我们能为整个行业提供一些激励。为此，微软将通过我们的“负责任的人工智能标准与原则（Responsible AI Standard and Principles）”，与业界共享在开发某些解决方案时应用的全部资源和专业知识。

对话 | AI+ 病毒学研究：跨界合作就像无影灯，减少跨领域认知阴影

在地球上，任何一种生物都摆脱不了病毒的纠缠。经过多年的研究，人类是否认清了病毒的本质？病毒究竟从何而来，又将去往何处？人类能否真正消灭病毒？它与人类只是敌对关系吗？ AI+ 病毒学的跨界研究能否追本溯源？

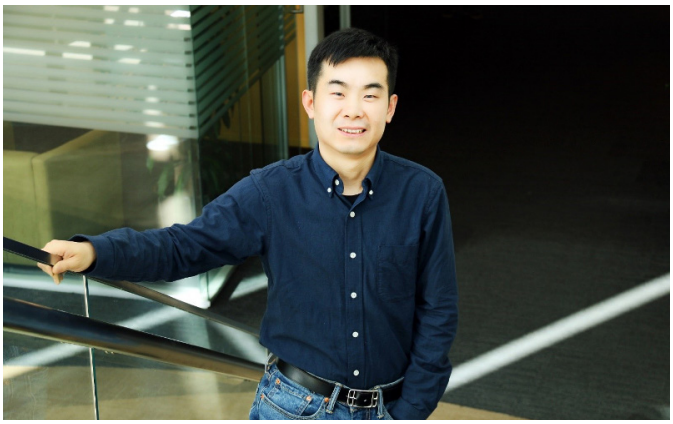
来自西湖大学生命科学学院的助理教授魏磊博士与微软研究院科学智能中心主管研究员邓攀进行了一场精彩的跨界对话，就上述问题进行了分享，并共同探讨了病毒学的核心前沿，以及大数据、AI 等计算机技术对病毒学领域和抗病毒研究可能产生的革命性影响。

扫描二维码了解更多信息

如何走好在学术界的发展之路、选择科研方向并做出有价值的研究？

如何走好在学术界的发展之路？如何选择科研方向并做出有价值的研究？如何更好地在研究社群中发挥价值？在微软亚洲研究院院友会“思享云客厅”系列活动之“我的学术之路”中，三位在海内外学术界深耕的微软亚洲研究院院友：加州大学伯克利分校教授马毅博士、中国人民大学高瓴人工智能学院执行院长文继荣博士、清华大学电子工程系主任汪玉博士，在微软亚洲研究院学术合作经理王婧雯主持下分享了他们在学术之路上的故事与经验。

扫描二维码了解更多信息



深科技｜MSRA 持续迭代 AI 大模型 BEiT，为通用多模态基础模型开创全新方向

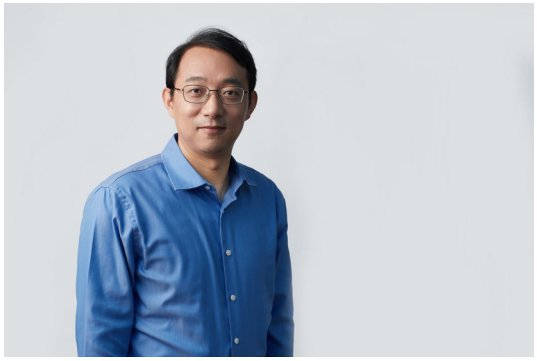
基于 2021 年推出的视觉预训练模型 BEiT，2022 年微软亚洲研究院进一步丰富了自监督学习的语义信息，发布了 BEiT-2，并随后将其升级为 BEiT-3，为多模态研究打开了新思路，也预示着 AI 大一统渐露曙光。微软亚洲研究院首席研究员韦福如认为："只有模型标准化，才可能实现规模化，进而为大范围产业化提供基础和可能。'大一统'中很重要的一点是，技术会变得越来越通用，只有通用才有可能更接近本质，也更利于不同领域的深度合作和相互促进。"

扫描二维码了解更多信息

环球网｜BEiT3 &"小花狮"入选 2022 环球趋势案例

为了总结 2022 年中国在各领域、各行业的取得的发展成果和未来趋势，积极引导经济和社会发展和谐共进，环球网开展了 2022 环球趋势案例征集展示活动。微软亚洲研究院联合微软图灵团队推出的 BEiT-3 预训练模型，以及华东师范大学和微软亚洲研究院联合推出的作文智能辅导系统教育产品"小花狮"成功入选科技创新的典型案例。

扫描二维码了解更多信息



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

微软亚洲研究院成立于 1998 年，在北京和上海拥有 300 多位科学家和工程师，是微软公司在亚太地区设立的、美国本土以外最大的研究机构。通过来自世界各地不同学科和背景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，将最新研究成果快速转化到微软的关键产品中，并且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。

作为微软研究院全球体系的一员，微软亚洲研究院拥有广阔的国际视野，同时扎根中国，辐射亚洲，通过融合东西方创新文化的精髓，以高度的社会责任感，持续开展有影响力、有温度、面向未来的基础科学研究和技术创新。微软亚洲研究院始终秉持相互信赖、相互尊重以及开放合作的理念，承诺与高校和科研机构开展持久而有效的合作，激发创新潜力、推进行业发展。

微软亚洲研究院倡导对技术进步怀有远大抱负，推崇富于冒险的极客创新精神，鼓励研究人员拓展研究的深度与广度，跨越计算机领域的界限，把视野拓展到解决具有广泛社会意义的问题上：提高人类的知识水平，推动基础研究的发展；增强人类的创造力和成就；培育有韧性、可持续的社会；支持健康的全球社会；确保技术值得信赖，让每个人都可以受益。

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