



01 全体大会

科技部副部长李萌在 2020 北京智源大会开幕式上的致辞

2020 年 6 月 21 日，由北京智源人工智能研究院主办的 2020 北京智源大会在线上成功开幕。在开幕式上，科技部副部长李萌在线出席并致辞。下面是致辞全文。

尊敬的殷勇副市长、各位专家，女士们、先生们、朋友们，大家上午好！

去年，我参与了 2019 北京智源大会，对大会提出的打造“内行的 AI 盛会”印象深刻，今年智源大会围绕“人工智能的下一个十年”开展广泛研讨，将会为全球人工智能的发展走向提供更多前瞻思考。首先，我谨代表科技部，对 2020 北京智源大会的召开表示热烈的祝贺！

人工智能是新一轮科技革命和产业变革的战略性技术，正在对科技进步、经济发展和社会治理等方面产生重大而深远的影响。中国高度重视人工智能的创新发展，2017 年 7 月，国务院印发《新一代人工智能发展规划》，明确提出中国新一代人工智能发展的战略目标：到 2030 年，人工智能理论、技术与应用总体达到世界领先水平，成为世界主要人工智能创新中心。

近年来，科技部加快推进规划实施，围绕人工智能重大科学问题研究、关键技术研发和产业化应用，进行了一系列重点部署。一是加快实施科技创新 2030—“新一代人工智能”重大项目，围绕**大数据智能、跨媒体智能、群体智能、混合增强智能、自主智能系统**等五大方向持续攻关，加快突破人工智能基础理论和关键核心技术瓶颈。二是推动国家新一代人工智能创新发展试验区建设，已支持北京、上海、深圳、天津、济南、浙江德清县等建设 11 个创新发展试验区，到 2023 年，将布局建设 20 个左右。三是建设国家新一代人工智能开放创新平台，依托百度、京东、小米、旷视等领军企业建设自动驾驶、智能供应链、智能家居、图像感知等 15 个开放创新平台，推动基础软硬件和技术的开放共享，服务人工智能创新创业。四是高度重视人工智能治理问题，成立国家新一代人工智能治理专业委员会，发布《新一代人工智能治理原则》，提出和谐友好、公平公正、包容共享、尊重隐私等八条原则，并探索建立人工治理国际交流平台，推动全球发展负责任的人工智能。五是推动人工智能与实体经济深度融合。培育发展人工智能新兴产业，促进传统产业智能化升级改造，推动产业链、价值链的高端化。特别是在应对此次全球新冠疫情中，推动人工智能技术在病区管理、流行性病学调查、药物研发、辅助医疗诊断、社区防控、疫情监控、风险分析预警等方面的广泛应用，为全球疫情防控工作积累了丰富的经验与成功案例，可为世界各国利用人工智能等技术建立应对疫情机制提供借鉴和参考。

北京是首个国家新一代人工智能创新发展试验区。在试验区建设过程中，希望北京市依托自身科技资源和产业发展优势，持续加大投入、深化改革、完善机制、强化保障，突出把握好“三个引领”：

一是原创引领，希望北京市依托智源研究院等新型研发机构，建立完善适应于人工智能基础前沿研究的科研制度，持续加大人工智能顶尖科学家培养，在人工智能原始创新方面取得更多突破性成果，引领全球人工智能创新方向，形成具有全球影响力的人工智能“北京学派”。同时要积极推进人工智能“新基建”发展，强化人工智能创新发展的基础支撑条件。

二是改革引领，希望北京市持续围绕数据开放与保护、应用场景开放、成果转化、知识产权保护等方面开展政

策先行先试。深入推进公共数据有序开放，加快智能医疗、智能交通、智能政务、科技冬奥等优势领域应用场景建设，推动领先人工智能技术示范应用，并加快制订适应人工智能发展的政策框架和法规标准体系，为人工智能科学研究、技术开发、产品创新、产业发展和社会应用营造良好环境。

三是开放引领，希望北京市以更加包容、开放的心态融入全球人工智能创新生态，推动智源研究院等研究机构积极参与人工智能领域国际重大科研计划，加强与国际顶尖人工智能研究机构的国际合作，推动建设链接全球人工智能创新网络的重要节点。同时要积极参与国际人工智能治理，推动《新一代人工智能治理原则》和《人工智能北京共识》落地，为未来打造“负责任的、有益的”人工智能，为开放合作和包容共享的智能时代贡献更多的“北京智慧”。

为进一步推动北京构建开放协同的人工智能生态系统，科技部正联合相关部委共同编制《关于支持北京加快建设具有全球影响力的人工智能创新中心行动计划》，从人才培养、基础前沿研究、创新服务平台建设等方面强化支持，助力北京打造世界一流的人工智能发展高地。

最后，希望智源研究院加快建设，在人工智能基础理论与共性技术上产出具有重要国际影响的原创成果，早日建设成为汇聚全球顶尖科学家研究力量的重大战略平台。

预祝大会圆满召开！谢谢大家！

北京市副市长殷勇在 2020 北京智源大会开幕式上的致辞

2020 年 6 月 21 日，由北京智源人工智能研究院主办的 2020 北京智源大会在线上成功开幕。在开幕式上，北京市副市长殷勇在线出席并致辞。下面是致辞全文。

尊敬的各位嘉宾，各位新老朋友，大家好！

在全球抗击新冠肺炎疫情的特殊时刻，我们举行第二届北京智源大会，同世界人工智能领域的顶尖专家学者通过视频相聚，共同探讨未来十年人工智能的发展走向，意义十分重大。首先，我谨代表北京市人民政府，向出席大会的专家学者和业界精英，表示诚挚的欢迎。

当前，新冠肺炎疫情仍在全球蔓延，给世界各国经济社会发展带来严重冲击，夺走了数十万人的宝贵生命。面对疫情，我们最需要做的，就是坚持人民至上、生命至上，全面加强国际合作，特别是大力推动科技领域的国际交流合作，发挥科技防疫的支撑作用，凝聚全球战“疫”的强大合力。

人工智能作为引领新一轮科技革命和产业变革的战略性技术，具有溢出带动性很强的“头雁”效应，正在对经济发展、社会进步、国际政治经济格局等产生重大而深远的影响。在这次疫情防控中，人工智能的优势得以凸显，在中国和北京产生了很多典型应用场景。比如，我们利用人工智能，在高密度人流中快速、准确地识别出体温异常者；我们利用 AI 技术开展远程问诊，有效降低了医护人员近距离接触感染的风险；我们利用人工智能影像辅助诊断技术，大幅提高诊断效率，缓解了医护人员紧张问题；我们通过人工智能有效减少病毒检测中的重复性工作，提高了检测效率；等等。

同时我们也认识到，当前全球人工智能发展整体上仍处于起步阶段，许多基础性的科技难题尚未突破，在行业应用、生态构建等方面仍面临一系列难题与挑战。我们认为，克服这些难题与挑战，需要统筹各个国家的创新资源，加强知识与信息共享、学术与市场互动，推动多元文化碰撞，促进创新要素互补与整合，从而降低创新成本，提高创新效率，加速科学发现。而在所有这些工作中，起决定性作用的，还是人的交流与合作。我们应该本着对人类负责、对科学负责的精神，采取更加务实的举措，为各类人才开展国际交流合作创造条件、提供支撑。

北京作为中国的政治中心、文化中心、国际交往中心和科技创新中心，是中国人工智能资源最丰富、创新最活跃、人才密度最高的城市。北京拥有人工智能相关人才近 4 万人，占全国总量的近 60%；北京的人工智能相关企业超过 1 千家，占全国近 30%，旷视、寒武纪、地平线、云知声等 8 家公司入选 CB Insights 2020 年全球 AI 独角兽公司榜单，第四范式、创新奇智、禾多科技 3 家公司入选 CB Insights 2020 年全球 AI 初创公司 100 强榜单；同时，北京还拥有中国首个新一代人工智能创新发展试验区，布局建设了 7 个国家新一代人工智能开放创新平台。

近年来，我们明确将人工智能作为北京科技创新和产业发展的重点领域之一，采取一系列举措，推动人工智能加速发展。人才方面，我们持续加大投入和支持力度，制定出台一系列针对性人才政策，不断拓宽人才宽度和厚度，努力为各类人才在京发展提供最好的条件。科研方面，我们聚焦原始创新和核心技术，建立自由探索与

目标导向相结合的科研体制，组建北京智源人工智能研究院等新型研发机构，支持科学家勇闯人工智能科技前沿“无人区”，开展原创性强、非共识的探索性研究，挑战最基础的问题和最关键的难题。资金方面，我们设立了总额为 300 亿元、放大后规模可达 1500 亿元的科创母基金，用于长期稳定支持包括人工智能在内的各类前沿科技的研发。应用场景方面，我们积极推动人工智能在医疗、交通、教育、城市治理、政务服务等领域广泛应用，并且将全市的重大项目、重大工程向人工智能技术应用开放，努力构建智慧城市。下一步，我们还将进一步加大投入力度、工作力度和创新力度，最大限度消除制约人工智能发展的各类障碍，构建更加包容、开放的人工智能创新生态，为全世界人工智能人才在北京开展学术交流、科学研究、创新创业营造更加优良的环境。

各位专家学者、行业精英，得力于你们的卓越工作和杰出贡献，人工智能技术在过去十年间取得重大突破，人工智能产业取得爆发式发展。面向下一个十年，如何实现从专用人工智能向通用人工智能的跨越式发展，如何加速人工智能的商业化应用，如何有效应对人工智能发展涉及的法律、伦理问题，如何让人工智能在促进全球社会、经济和环境可持续发展方面发挥更加积极的作用，都有待于大家去共同探索。

我们举办北京智源大会，就是希望为国内外人工智能领域的优秀人才搭建一个开展前沿学术研讨的交流平台，进行灵感的碰撞、思想的交锋。希望大家借助这个平台，深入研讨未来全球人工智能的发展趋势，积极开展科学研究、技术研发、人才培养等方面的交流合作，同时也欢迎各国人工智能人才来北京创新创业，共同为全球人工智能的赋能共生、繁荣发展贡献力量！

最后，预祝大会取得圆满成功！谢谢大家！

智源研究院院长黄铁军：智源进展报告

各位专家、朋友们好！

我是北京智源人工智能研究院的院长黄铁军，下面我将代表智源研究院简要汇报我们在过去一年中的主要进展和未来一段时间的主要发展计划。

首先，让我们简单回顾一下智源研究院的发展愿景，它分为四个层次：

1. 聚焦原始创新和核心技术，建立自由探索与目标导向相结合的科研体制。
2. 支持科学家勇闯人工智能科技前沿“无人区”，挑战最基础的问题和最关键的难题，推动人工智能理论、方法、工具、系统和应用取得变革性、颠覆性突破。
3. 营造全球最佳的学术和技术创新生态，推动北京成为全球人工智能学术思想、基础理论、顶尖人才、企业创新和政策的源头，率先成为国际领先的人工智能创新中心。
4. 推动人工智能产业发展和深度应用，改变人类社会生活，促进人类、环境和智能的可持续发展。

下面，我从五个方面报告智源研究院近一年来的主要进展。

一、智源学者

首先是“智源学者”计划，这是2019年4月开始执行的，目的是遴选最优秀的科学家，挑战最基础的问题和最关键的难题。在过去的一年里，我们先后设立了人工智能的数理基础、机器学习、自然语言处理、智能信息检索与挖掘、智能体系架构与芯片五大方向，遴选并支持了以“智源研究员”和“智源科学家”为主的学者69人。2019年底我们还遴选了以“自由探索”为使命的16位青年科学家，总数达到了85人，他们分别来自于北京大学、清华大学、中国科学院等高校院所和在京优秀企业。

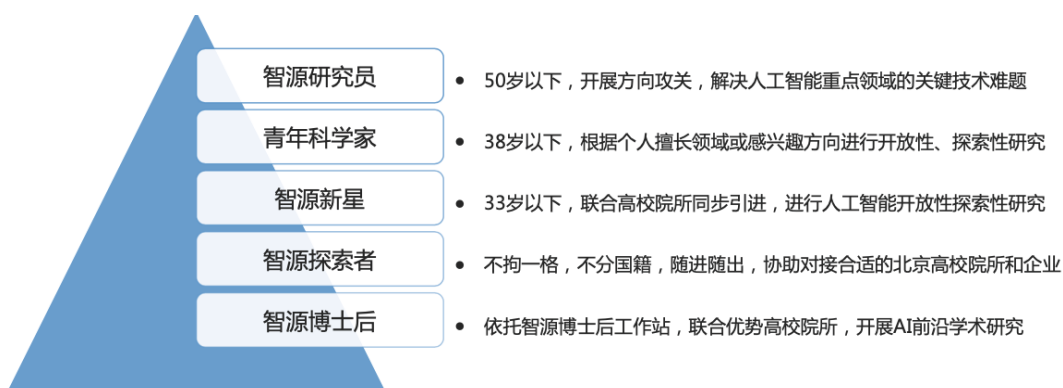


图1：以一流学者为核心，实施“智源学者计划”。智源学者计划扩充为五个层次，培育和支持人工智能优秀科技人才，建设世界人工智能人才高地，人数再增加100名，其中智源研究员和青年科学家20+、智源新星20+、探索者30+、博士后30+。

今年我们将智源学者计划扩充为五个层次，继续再增加支持 100 人，除了去年已经执行的“智源研究员”和“青年科学家”保持不变之外，新增了“智源新星”，这个层次的人才要求 33 岁以下，由智源研究院和在京的高校、院所同步引进，进行人工智能的开放性、探索性的研究。我们还设立了“智源探索者”这样一个新层次，它面向所有有意到北京发展、但尚未确定目标单位的人工智能优秀人才，不拘一格。智源探索者不仅可以在智源进行潜心研究，而且智源也会协助他们对接合适的北京高校院所和企业。在博士后层次，智源今年正式设立了博士后工作站，联合优势高校院所，招收优秀的博士后开展前沿研究。

方向 任务	数理基础	机器学习	自然语言处理	信息检索挖掘	芯片系统	认知神经基础	决策智能	机器感知	开放探索
智源社区	●	●	●	●	●	●	●	●	●
学术突破	●●	●●	●●	●	●	●●	●●	●	○
成果转化	○	●	●	●●	●●	○	●	●●	○
智能医疗	○	●●	●●	●	●	●●	○	●	○
智能交通	●	●	○	●	●●	○	●	●●	○
智能政务	○	●●	●●	●●	●	○	●●	●	○
智能机器	●●	●	○	○	●●	●●	●●	●	○
认知图谱	○	●	●●	●●	●	●●	●●	○	○
伦理与可持续发展	●	●	●	●	●	●	●	●	●

图 2：智源研究院的重大方向布局

图 2 中大致描绘了智源在重大研究方向的布局，以及这些方向与我们重要任务、目标的关系。纵向就是我们的重大方向，比如刚才说的已经成立的五个重大方向，以及今年已经在酝酿、争取尽早公布的三个新方向：认知神经基础、决策智能和机器感知。此外，我们还专门为非共识的、完全开放的探索研究设置一个“开放探索方向”。所有这些研究方向和我们现在做的各方面任务有密切的关系，比如所有这些方向都可以参与“智源社区”的建设，所有这些方向都会追求“学术突破”。当然各个方向关注的重心可能会各有侧重，图中大的实心圈就表示这样的关系，比如数理基础的重心是学术突破，芯片系统和信息检索挖掘的重心在成果转化等。应用方面，我们目的面向三大应用场景：智能医疗、智能交通和智能政务，它们也在各大方向中各有侧重。我们还在规划面向未来的一些重大研究任务，比如如何制造能力更强的智能机器，如何设计更接近人类认知的认知图谱，以及从去年开始的人工智能伦理和科学应用发展方面的一些工作等。所有这些方向都会参与这些任务的实施。

二、智源社区

智源社区是互联网上的智源研究院，经过一年多的建设努力，今天，我们正式发布智源社区，目的是构建高度合作的人工智能社区，营造活跃的学术和技术创新生态，通过进一步加强社区成员的协同效应，培养培养出新一代问题的发现者和解决者。



图 3：智源社区——构建人工智能学者社交网络，营造学术和技术创新生态

正在观看智源大会的数万名朋友，你们现在已经体验了智源社区的部分功能——活动系统和线上论坛，大家可以输入网址 baai.org，或扫描屏幕上的二维码，注册成为第一批社区成员，与本次大会报告嘉宾线上交流，更多功能将在智源大会结束后对外开放，包括以展示研究成果，探讨新发现，参与智源大会、学术报告会、主题沙龙、论文分享等活动；社区还将帮助学者、学生、开发者、创业者找到学术交流与技术合作的伙伴，社区将紧密联系 3000 名以上的人工智能领域顶尖学者，辐射 10 万以上人工智能科研和技术人员。

symbol	meaning	LaTeX	simplified
d	input dimension	d	
d_o	output dimension	$d_{\text{rm o}}$	
m_l	the number of l th layer neuron, $m_0 = d, m_L = d_o$	m_1	
$\mathbf{W}^{[l]}$	the l th layer weight	$\mathbf{W}^{[1]}$	$\mathbf{W}^{[1]}$
$\mathbf{b}^{[l]}$	the l th layer bias term	$\mathbf{b}^{[1]}$	$\mathbf{b}^{[1]}$
\circ	entry-wise operation	\circ	
$\sigma: \mathbb{R} \rightarrow \mathbb{R}^+$	activation function	σ	
θ	$= (\mathbf{W}^{[0]}, \dots, \mathbf{W}^{[L-1]}, \mathbf{b}^{[0]}, \dots, \mathbf{b}^{[L-1]})$, parameters	$\mathbf{\theta}$	$\mathbf{\theta}$
$f_{\theta}^{[0]}(\mathbf{x})$	$= \mathbf{x}$		
$f_{\theta}^{[l]}(\mathbf{x})$	$= \sigma \circ (\mathbf{W}^{[l-1]} f_{\theta}^{[l-1]}(\mathbf{x}) + \mathbf{b}^{[l-1]})$, l -th layer output		
$f_{\theta}(\mathbf{x})$	$= f_{\theta}^{[L]}(\mathbf{x}) = \mathbf{W}^{[L-1]} f_{\theta}^{[L-1]}(\mathbf{x}) + \mathbf{b}^{[L-1]}$, L -layer NN		

图 4：世界首个“机器学习通用数学符号集”，部分样例。

作为智源社区的公共服务之一，今天我们将发布由多所高校和多位领域学者紧密协作的成果——世界首个“机器学习通用数学符号集”。目前，机器学习是一个热门的研究方向，但研究学者们往往来自不同领域，而论文中所用记号差别很大，便造成论文阅读困难。“通用机器学习符号集”针对常用且易混淆的符号，给出一套标准化建议并提供 LaTeX 包，辅助论文写作过程中的常用符号选择，以解决符号混乱导致的交流障碍问题。

本次会议，智源研究院将与清华大学人工智能研究院、创新工场共同发布《人工智能下一个十年》报告。报告

通过科技情报大数据平台 AMiner 对人工智能领域的学术论文进行数据挖掘，并结合行业投融资等数据，回顾展望 AI 技术、人才、产业、政策、治理等方面的发展情况。报告可以通过注册成为“智源社区”成员获得。以下是一些主要研究发现，更详细的分析内容大家可以阅读报告来了解。

首先是科研方面，报告对比了中美的 AI 科研实力。借助 AMiner 系统，基于 2009–2020 年 AI 领域 45 个顶级期刊和会议所收录的论文进行数据挖掘和分析，美国发表的论文数和学者数都是我国的 2–3 倍，所以我们应该更加努力。在报告分析的 20 个人工智能领域里面，人机交互、知识工程、机器人、计算机图形等领域，中国与美国的差距尤其大。

	论文量（篇）	学者量（人）	人均产出（篇/人）
美国	69764	49116	1.4
中国	25418	17368	1.5

图 5：2009–2020 年间 AI 领域，中、美两国论文量和学者量对比。中国学者量和论文量都约是美国学者量的三分之一，但人均论文产出量略高。

中国国内我们也对城市分布做了统计，北京拥有国内 AI 高校和企业的数量资源优势，在 AI 领域的论文量和人才量最多。相对而言，深圳和香港 AI 人才在发表论文方面更努力，其人均论文产出量更多。

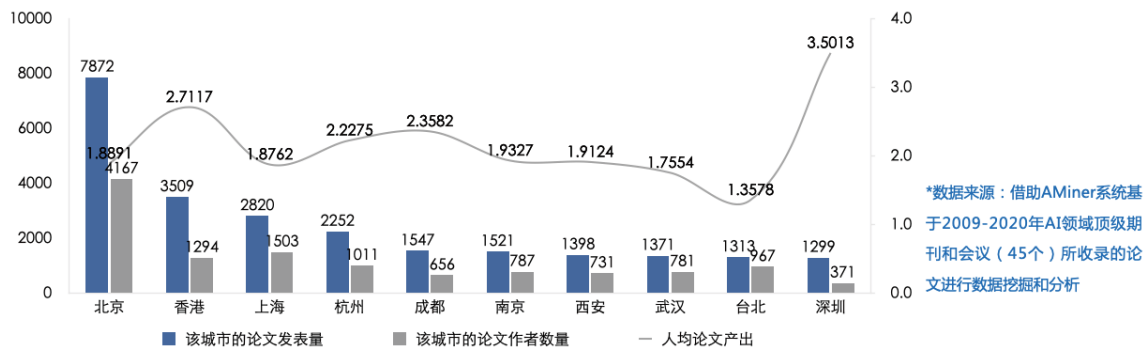


图 6：人工智能领域论文发表量 TOP10 国内城市。

AI 投融资方面，大家可能已经注意到了，中国 AI 创业热潮自 2013 年起兴起，每年新成立的 AI 创业公司数量增速明显，直到 2015 年爆发达到近 10 年的顶峰，此后两年 AI 创业虽有下滑趋势但仍保持较高水平，直到 2018、2019 年呈现断崖式下跌。目前中国 AI 企业创业方向集中在机器学习、计算机视觉、自然语言处理等技术层，以及金融、教育、交通、工业等行业应用层。

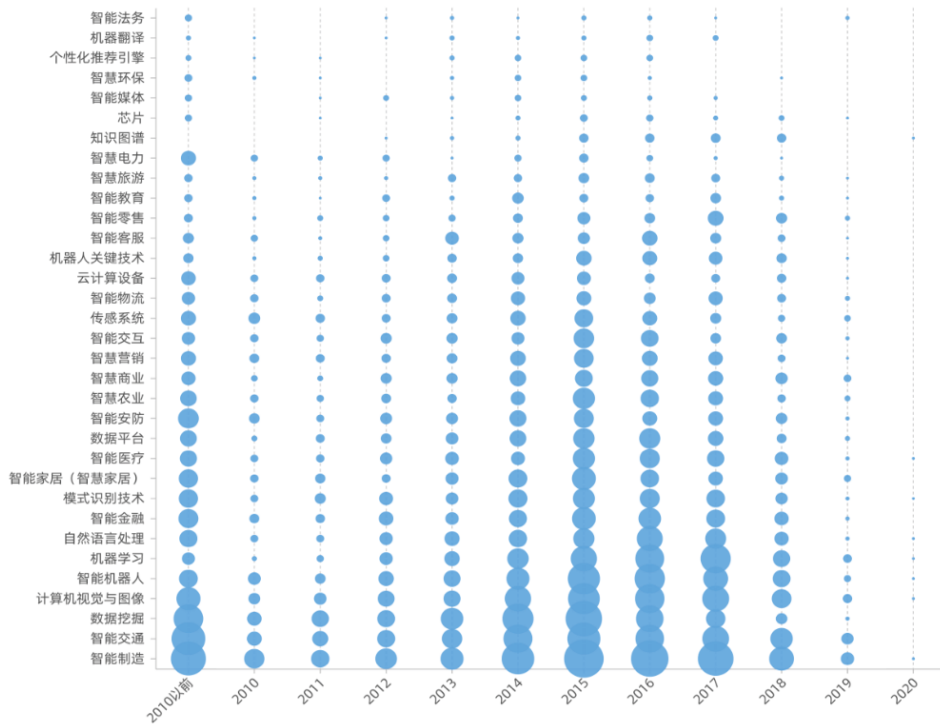


图 7：中国 AI 创业企业数量在各领域的分布情况。数据来源：IT 桔子、北京智源人工智能研究院

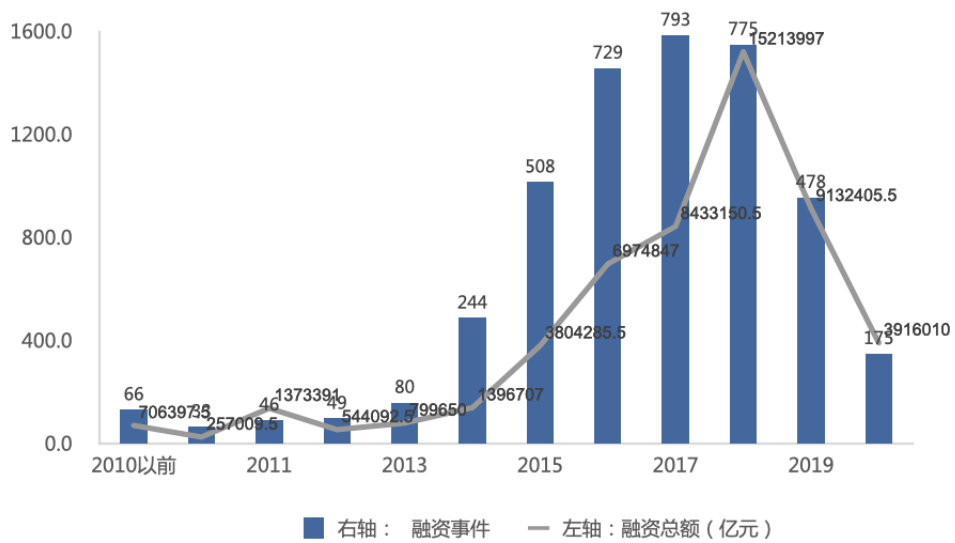


图 8：中国 AI 领域历年投融资事件数量及融资总额。数据来源：IT 桔子、北京智源人工智能研究院

报告也对人工智能技术的未来发展进行了分析，我们认为，机器学习与深度学习在人工智能领域应用广泛并取得显著成果，但受限于底层算法，使得人工智能技术发展达到瓶颈，未来需要与数学、脑科学等结合实现底层理论的突破。同时，要实现具有能理解、会思考、有认知、会决策的强人工智能，离不开数据支撑和知识驱动，将知识图谱、认知推理、逻辑表达结合的认知图谱也将是未来的一个重要研究方向。

三、数据中心和智能医疗

数据中心和智能医疗方面，智源研究院成立了数据开放研究中心，围绕“开放数据”、“知识图谱”、“算法平台”三大主题，建设数据与知识开放平台。在“开放数据”方面，我们推出行业基准系列数据集和智源大赛系列数据集，旨在推动人工智能领域的数据与知识共享。今年，自疫情初始，我们发布了最全的“新冠肺炎开放数据源”，从疫情、科研、知识、媒体和政策这五个方面收集了来自世界各地各种类型的相关开放数据，已吸引来自全球各地近四万人次访问和下载。除此之外，我们在疫情应对方面持续加力，通过构建一个大规模、结构化的中英文双语新冠知识图谱，用知识支撑AI，提供更多智能服务与应用。在“算法平台”方面，通用智能平台CogDL的构建为研究人员提供了基于图的模型开发与测试工具，在健康医疗方面给予安全可靠、高效便捷的智能预测服务。接下来，我们将继续联合旷世发布全球最大的物体检测数据集，和京东联合发布多模态人机对话数据集，和智源学者朱军团联合发布AI安全数据集，并举办相应的竞赛，推进人工智能行业数据共享。



图 9：智源数据开放研究中心

在“数据研究”方面，针对智能医疗领域面临的数据孤岛和数据安全隐私的挑战，我们研发了智能医疗共享计算平台。在“数据共享”和“数据竞赛”方面，我们推出行业基准系列数据集和智源大赛系列数据集，旨在推动人工智能领域的数据与知识共享。为了推进新冠肺炎药物研发，我们发布了智源抗疫 - 药物研发小分子性质预测赛。



图 10：智能医疗开放平台

刚才讲到，新冠疫情的抗疫是过去半年左右的时间里智源研究院的一个重要任务，我们新冠疫情数据计算平台上线了“知识疫图 – 全球新冠疫情智能驾驶舱”，基于知识驱动、全球疫情统计数据 and 预测模型，对世界各地的疫情发展及风险状况进行量化评估和预测，提供复工、复产各方面的辅助决策支持，包括地区疫情风险评估、政府政策推荐、个人生理和心理健康自评评估等。此外，我们智源联合清华、协和、首都医科大学共同发布的乳腺癌智能计算与预测平台 SonoBreastX，可以自动预测病变肿瘤良 / 恶性及其分子亚型，辅助医生进行乳腺癌诊断。为解决智能医疗领域的医疗数据共享问题，智源研发了联邦学习多方安全计算平台，接下来会基于乳腺癌数据集进行智能医疗领域联邦实验，突破医疗数据共享关键技术的研发。

四、创新中心

创新中心是智源为推动人工智能产业发展和深度应用设立的成果转化创新载体，也是支撑人工智能产业发展的重要一环。创新中心通过开放智源的生态资源，支持关键核心技术攻关，推动 AI 原始重大创新和关键技术落地和深度应用。



图 11：智源创新中心——人工智能研究成果转化、创新载体

具体来讲，分为三个层面：

1. 围绕智源重大学术方向，支持智源学者或高校院所的 AI 科学家，基于自身的原始重大创新成果进行技术开发创新，实现成果落地；
2. 支持已初步实现落地成果的重大技术或行业共性关键技术进行工程化验证和应用探索，实现跨越式发展；
3. 支持新兴行业发展过程中对人工智能关键技术需求开发，加速人工智能面向实际应用的广泛渗透和应用。

创新中心将把自己的创新生态资源开放出来支持大家，提供包括研究者、开发者以及创业者共同合作的良好土壤，共同探索人工智能领域前沿技术的落地，共同推动中国人工智能产业的发展。欢迎大家加入创新中心。

五、伦理与可持续发展

最后一个方面是伦理和可持续发展。去年，智源牵头发布了的《人工智能北京共识》，得到国内外广泛关注。我们也严格按照《北京共识》进行实践。下面，我讲两个例子。

第一个典型的例子，是保护隐私的传染链密切接触者精准追踪系统。智源研究院团队从3月份开始就着手研发常态化防控技术，研发一种保护隐私的传染链密切接触者精准追踪系统，称为智源蓝保，Blue Bubble。能够精准追踪感染人员相对位置，快速排查与感染者有近距离接触风险的人员，大幅降低隔离人员比例，降低经济损失。

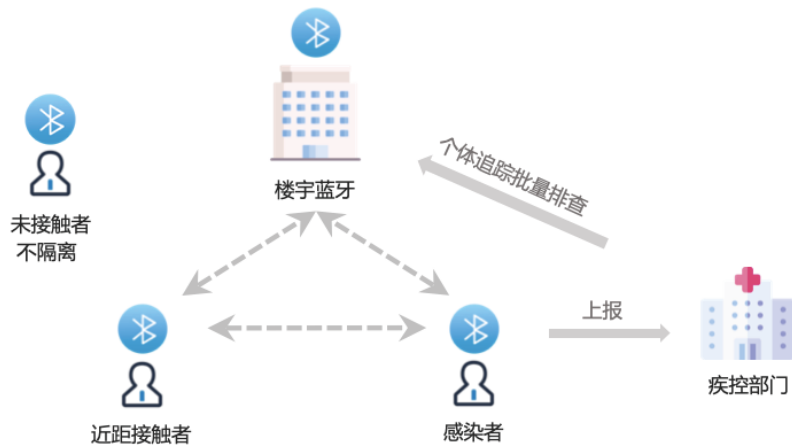


图 12：智源蓝保技术框架——蓝牙 ID 碰撞 + 近距离判断 + 本地边缘存储

使用到的技术，就是手机、手环上随手可得的蓝牙设备，通过设备 ID 之间的相互碰撞和记录近距离接触。当感染者被确诊，他手机上记录的行动轨迹由疾控部门公开，其他人和安装量蓝牙设备的场所通过比对发现密切接触者，做到精准追踪。未有近距离时空交叉者，不必隔离，不用现在这样发现一名感染者，整个街道或整个社区全部隔离。

需要特别说明的是：除了蓝牙 ID 匿名化处理之外，我们记录的蓝牙两两发现数据，自始至终没有离开过该场所，或者存在用户手机上，全程不上网，不上传商业数据库，所以不会数据外泄。

总结起来就四句话：场所数据不离场，个人隐私不离机，患者轨迹供比对，能自证者不隔离！最近人民日报报道了这项进展。

我们的智源蓝保有四大优势：

- GPS、手机网络定位，粒度太粗，无法精确到场所内部、大楼内部，而我们设计的基于蓝牙发现的技术可以精确到一两米。
- 前面提到，我们的数据全程不离场、不离机、不上网，就和大楼内的视频监控数据一样在本地保护起来。
- 支持常态化防控，比如大楼内人员聚集热度图；或者确诊病例出现时，可以恢复他在楼内的位置轨迹，这一过程还是本地化操作，匿名化。
- 我们已经在量子芯座实现示范楼，充分利用了大楼内的已有无线网络和边缘存储，1000 多平米一层楼成本不到 1000 元，两三个小时就可以完成部署安装调试。

智源蓝保系统，未来有望和视频监控一样，成为智慧楼宇的标配，比视频监控成本低很多，而且不涉及人脸图像等隐私信息。

联合国可持续发展目标 (Sustainable Development Goals, SDG) 由联合国 2015 正式通过, 旨在从 2015 年到 2030 年间以综合方式彻底解决社会、经济和环境三个维度的发展问题, 转向可持续发展道路。

人工智能是能够改变社会的颠覆性技术, 也是能够推动社会进步的使能技术。应当通过新一代人工智能的发展促进全球可持续发展目标的实现, 这既是写入国家《新一代人工智能发展规划》、《国家新一代人工智能治理原则》的目标, 也是全球人工智能产、学、研界的共同目标。人工智能的科研机构、教育机构、企业有责任通过人工智能推动社会、经济、环境的可持续发展。



图 13: 建立面向可持续发展的人工智能智库, 发布面向可持续发展的人工智能公益研究计划

为此, 北京智源人工智能研究院拟建立面向可持续发展的人工智能智库, 发布面向可持续发展的人工智能公益研究计划。

我们将联合全球科研机构相关学者, 并邀请“对采用人工智能实现可持续发展有共同愿景”的国内、国际企业作为伙伴单位, 成立面向可持续发展的人工智能公益研究计划。智源研究院将作为管理机构执行该公益研究计划的资助, 组织全球相关专家成立科学委员会, 与参与共建的企业共同拟定年度资助方向并评审、并向全球发布相关课题。组建面向可持续发展的人工智能智库平台, 负责计划和研究成果的全球发布和推广。智源研究院作为组织管理机构, 从中不收取任何费用。

研究项目将向全球开放申请, 研究成果在公益研究计划官网向全球公开, 支持在遵循开源许可协议条件下无偿使用。目前接受企业申请, 作为战略合作伙伴和共同发起单位资助相关研究计划, 旷视、滴滴、小米三家企业已经表示准备支持, 希望更多企业能够共同支持这项计划, 为推动人工智能技术对全球的可持续发展做出更多的贡献。

最后再次感谢大家长期以来对北京智源人工智能研究院的支持, 让我们共同努力, 建设人工智能创新的活跃生态, 共同推动人工智能技术的发展以及人类社会的进步。

卡内基·梅隆大学教授 Ruslan Salakhutdinov: 将领域知识融入深度学习模型

整理：智源社区 钱小鹅

在第二届智源大会“全体会议”中，Ruslan Salakhutdinov 教授分享了主题为《Incorporating Domain Knowledge into Deep Learning Models》(将领域知识融入深度学习模型)的报告。

Ruslan Salakhutdinov，是卡内基梅隆大学教授，加拿大统计机器学习研究主席，微软学者和斯隆奖获得者，主要研究领域为深度学习概率图模型和大规模优化，师从 Geoffrey Hinton，是剪枝、深度编码等著名学习方法的提出者。他于 2009 年获得多伦多大学的博士学位，随后又在麻省理工大学开展了为期两年的博士后研究。2016 年，入职卡内基梅隆大学，同年获得英伟达人工智能先驱奖，并加入苹果公司担任 AI 研究总监。目前 Ruslan 已累计发表学术论文 300 多篇，是深度学习领域年轻一代领军科学家，主要的研究方向为：语音识别、计算机视觉、推荐系统、语言理解、药物研发及生物图像分析。

Ruslan 认为，虽然我们用 AI 算法可以解决很多问题，但还是面临很多局限性和挑战，主要分为这四大类：

1. 自然语言理解与推理
2. 人工智能：深层强化学习和控制
3. 将领域知识融入到深度学习模型中
4. 多模态、半监督学习、自监督学习

本次讲座中，Ruslan 主要为我们深入分享了他的团队在上述第二、第三项挑战领域中，所进行的探索和最新成果。

一、将领域知识融入到深度学习模型

首先，Ruslan 为我们剖析了一个非常有意思的例子，如下图 1，如果我们试图回答如下这个问题：据我们所知，哪些冠状病毒可以感染人类呢？答案有三类：MERS-CoV、SARS-CoV 以及 Covid-19。想要回答这个问题，我们可以从多个渠道亦或者我们自身现有的知识库中挖掘到答案，但我们不难发现的是，这是一个关于生物防疫或病毒方向的专业问题，无论我们调用上述所说的哪种渠道回答这个问题，我们最终所调用的必然是与该问题相同或相关的领域知识，也就是 Ruslan 本次讲座中的关键词——Domain Knowledge。



图 1：哪些冠状病毒可以感染人类？

领域知识在回答复杂问题是非常重要的，尤其是我们现在的工业场景中所需要的交叉知识更为丰富，例如自动驾驶，在现行的自动驾驶算法方案中，依赖的领域知识是非常丰富的：我们需要从摄像机中采集图像，利用计算机视觉的一些算法来处理图像；同样，我们也需要从其他传感器，例如激光雷达、GPS、IMU 等中收集数据，来构造图像的辅助算法。显然，构造一套完整的自动驾驶算法需要多个领域的知识支撑，同样回答上述图 1 中提出的问题，也需要多个领域的知识支撑，如下图 2。因此，如何收集相关的数据、如何找到数据间的逻辑关系，如何将这种推理关系抽象成模型，甚至将其抽象成标签依赖性非常弱的弱监督学习模型，一直是该领域的关键挑战。

Key Challenges

- Heterogeneous Data
- Reasoning
- Weak Supervision

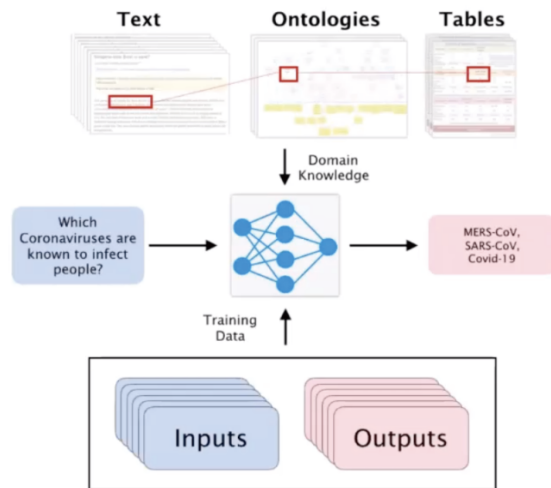


图 2：领域知识在深度学习推理中应用时的关键挑战

解决上文中提到的“关键挑战”有很多不同的方法，Ruslan 为我们分享了其中一个研究方向——使用知识库和文本结合的方法建立开放域问答系统 (如图 3)。首先，我们先为大家介绍两个比较核心的关键词：开放域以及知识库。

- **开放域**。开放领域问答系统的概念虽然提出的时间并不长，但已经形成发展出了一些比较成熟的系统。它起源于 1993 年麻省理工学院人工智能实验室开发的 START 系统。所谓开放域问答系统，即采用基于知识库和指定的检索方式，对用户提出的问题进行自动的答案搜索的系统。如果用户提出的问题在系统的知识库中，则直接返回正确答案。反之，如果用户提出了超出系统知识库的问题，则系统将会分析问题的关键词，并通过搜索引擎搜索关键词得到相关信息，对信息进行后处理后得到较为准确的答案。
- **知识库 (Knowledge Bases, KBS)**。所谓知识库，其实就是将互联网上的信息经过专家人工提取和构造，以三元组的形式存储下来 (Subject, Relationship, Object)，是一种非常结构化的信息，比较知名的有 FreeBase。

在人工智能兴起之前，传统的 QA 都是通过知识库进行检索或者信息抽取等方式进行建模的。有了知识库，那么给定一个问题我们自然可以进行检索记忆并作出回答。该问题的挑战在于，由于问题往往是以自然语言的方式提出，而知识库是高度结构化的组织，所以如何进行检索或者信息抽取其实是一个很困难的工作，同时，面对 KBS 这么庞大的知识库应该如何解决内存容量问题也同样是值得不断优化方向。

Knowledge Base as a Knowledge Source

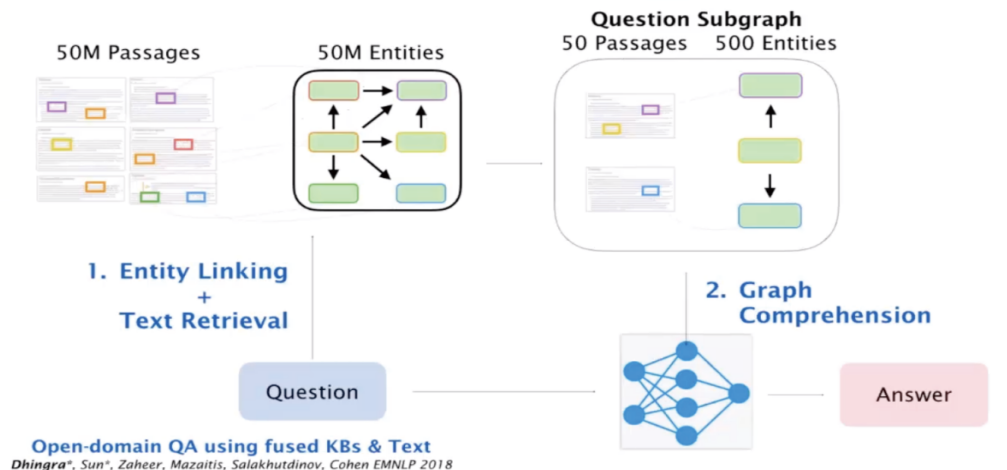


图 3：使用知识库和文本结合的方法建立开放域问答系统

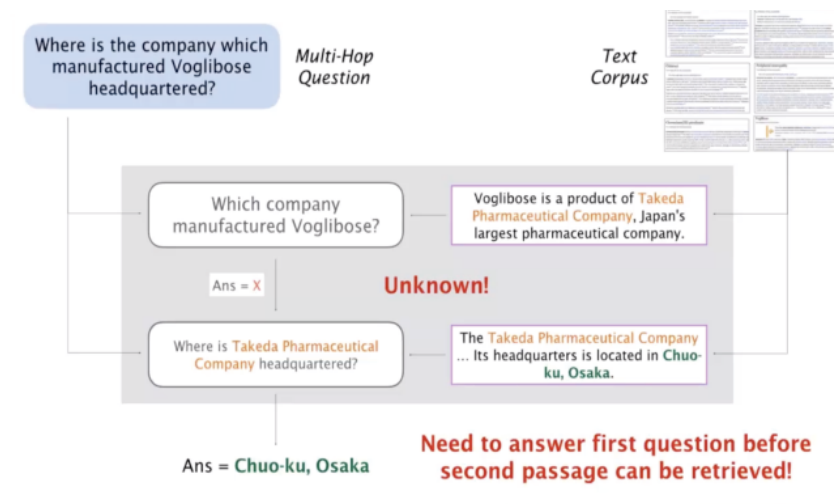
自人工智能兴起后，在知识库中快速搜索答案成为了可能。通常，整个过程可以描述为（如图 3）：首先，我们在知识库中确定具体的关键词；其次，将这些词进行文本挖掘；最后，结合文本信息建立知识图谱。如果将神经网络引入知识图谱模型，其本质即是我们通常所说的图神经网络。

使用图神经网络进行答案挖掘（换句话说即为，整合文本信息获取知识图谱信息）是一件非常有趣的事情，学界的研究也非常火热。随着研究的深入，大家把方向转向了更为复杂的问题——Multi-Hop Question。Multi-Hop Question，多跳问题，顾名思义也就是回答问题需要多步推理过程。与 Single Face 相比，多跳问题属于复杂问题的一类子场景。为了方便大家理解，我们不妨先为大家分享一个实例。如下图 4 所示，如果我们试图让回答系统回答一个较为复杂的问题：**生产 Voglibose 的公司总部设在哪里？那么**，其实我们可以把这个问题分为两个问题回答：

1. 哪家公司生产 Voglibose ？
2. 该公司的总部设在哪里？

整合大量的文本信息后，我们了解到，Voglibose 这种药物的生产公司为武田制药公司，该公司是日本最大的制药公司，该公司的总部设在日本大阪。

用这种方式，我们通过连续的回答两个简单问题，最终回答了一个复杂问题。当然，多跳问题也能分解为三个或更多的简单问题。Ruslan 随即提到，对于该问题能否做到端到端、更有效、组合更具灵活性？



Relational Following

Differentiable Reasoning over a Virtual KB
 Dhingra, Zaheer, Balachandran, Neubig, Salakhutdinov, Cohen
 ICLR 2020 (oral)

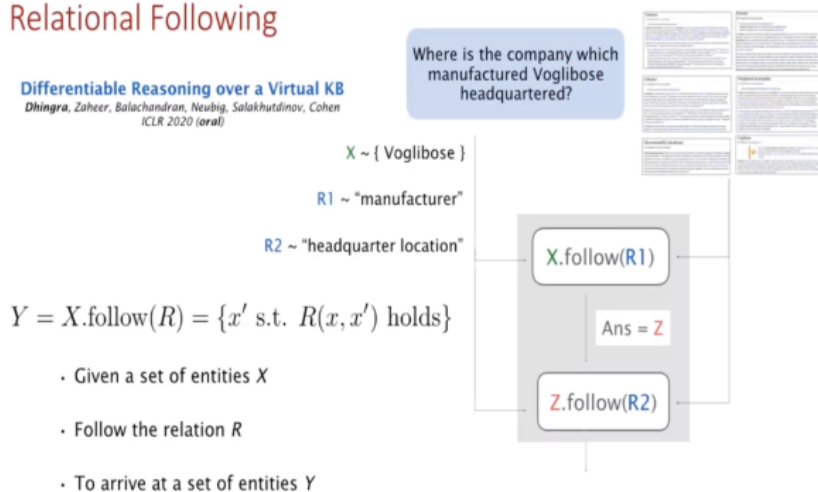


图 4: 多跳问题实例分析图

Ruslan 及其团队成员在该领域已经做出了一些成果，他介绍到，整个多跳问题可以通过图神经网络解决，该神经网络大体可以分为四个模块：

- **Graph Construction**

该模块的核心目标建立图，将与问题及答案相关的数据存储为图类的数据结构，主要包括：1) 根据关键词检索相关段落，寻找到与关键词匹配度最高的 N 个段落；2) 寻找能够链接到其他相关短多的 facts and entities; 3) 为上一步中相关的段落和实体添加边。

- **Context Encoding**

该模块的核心目标是为了把所有的搜索到的段落或实体拼接成一个整体，我们称之为 Context，接着我们将原始的问题 Question 一同拼接起来，进行编码，将编码后的结果输入到神经网络中。通常，我们使用 Bag-Of-Symbol 的方法对段落或实体进行编码，将知识库存储成一个 N 维的向量，并使用 Bag-Of-Ngrams 的方法对原始的问题进行编码，存储为 N_v 维向量。

• Graph Reasoning

该模块的核心目标是利用图神经网络，进行图中信息的传播，训练各实体类型节点中连接边的权重，同时学习图神经网络中各门的注意力和表达能力，用于做答案区域抽取。

• Multi-task Prediction

该模块的核心目标是图推理之后，更新过的节点表示被用于不同的子任务，由于答案可能不在实体节点中，因此我们需要构造一个有效的损失函数，对整体编码的内容进行回归，从而进行多任务的预测。

在整个的流程中，我们注意到，以实体 (Entity) 作为节点，关系 (Relationship) 作为边，记忆作为矩阵进行结构化存储，使得问题更具有计算性。我们可以根据问题找到相关记忆，并对其进行加权求和，同时，我们考虑记忆对问题的稀疏性，使得在计算的过程中可以直接使用大量稀疏矩阵的算法模块进行存储和计算，有效的加快了推理速度，减少了内存的损耗。同时在端到端记忆传播的过程中，Ruslan 引入得分机制将实体与问题的相关性分数与实体共同存储在数据结构中，最终得到答案。Ruslan 谈到，他对这个项目是非常感兴趣的，因为我们通常在开放式回答系统背后嵌入的知识库是基于维基百科的，或基于某个知识库的侧面。但 Ruslan 所建立的知识库是非常与众不同的。由于记忆矩阵的嵌入，以及这些特定的数据结构和网络架构，使得多个问题的“软连接”（问题之间的逻辑关系）更容易得到，同时鲁棒性也更强，基于此，Ruslan 可以建立一个完全由他个人“控制”的知识库，这个数据库并不来源于维基百科，也不来源于单纯的一些知识片段，这个知识库可以非常大，内容涉及的更加广阔。最后，Ruslan 为我们展示了新算法模型的精度及速度，如下图 5。

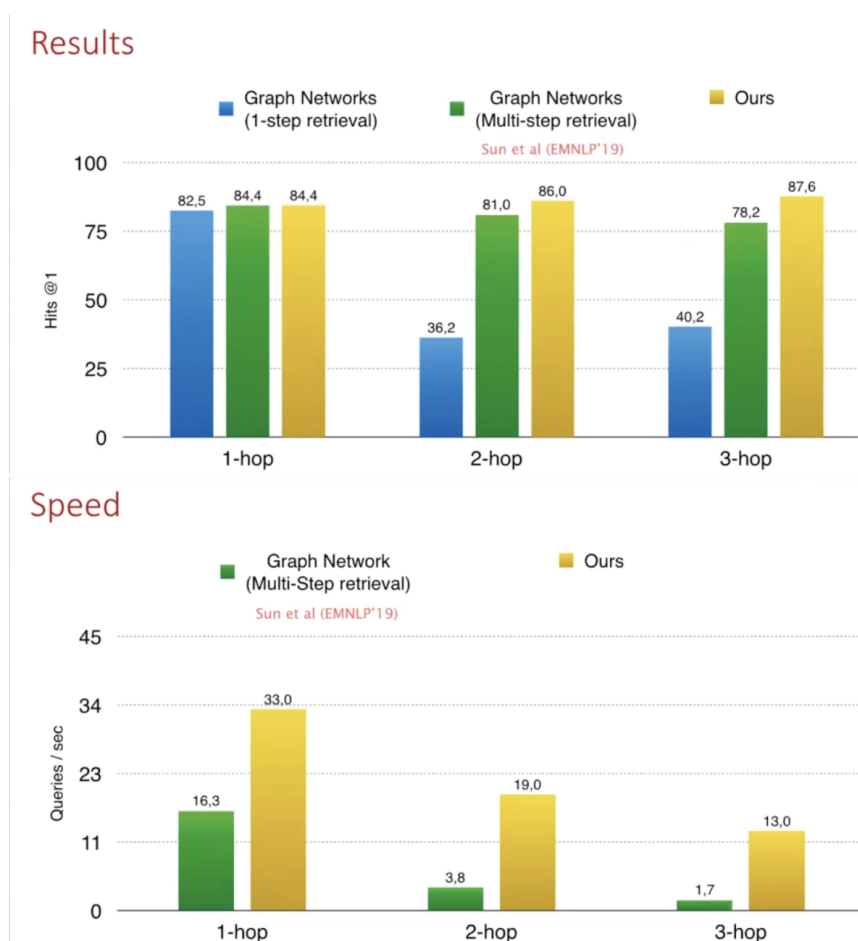


图 5: Ruslan 与合作者的最新研究成果（精度 / 速度）对比图

二、深层强化学习与控制

讲座的第二部分，Ruslan 为我们分享了四大挑战（请参阅文章开头介绍）中他对第二个挑战——深层强化学习与控制的思考和研究。他指出，我们心中所构想的所谓的人工智能，其更偏向于增强学习，也就是机器真正的学习能力或所谓的“智慧”。Ruslan 首先为我们列举了一个非常有产业价值的实例——机器人导航。如下图 6 中，如果我们讨论机器人导航的真正学习能力，我们期望机器人会自动的构建地图，在构建的地图中寻找人类需要的目标，并自动进行路径规划获取（或抵达）目标所在地。

Navigation

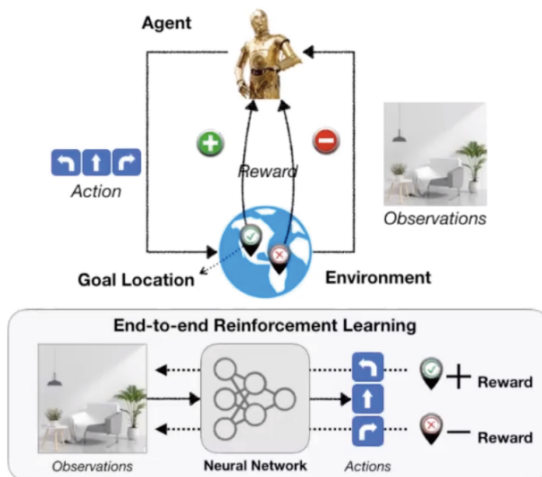


图 6：导航系统中的增强学习

Ruslan 指出，导航是构建智能代理的关键任务。导航任务可以用多种形式表示，例如，点目标任务，即到导航到特定的坐标；语义导航，即查找特定场景或对象的路径等。但不管任务是什么，在未知环境中进行导航的核心问题是探索，即如何有效地访问尽可能多的环境。这有助于最大限度地扩大覆盖范围，以便在未知环境中找到目标，或者在有限的时间预算内高效地预映射环境。在 2019 年，学术界的一些研究人员已经开始探索使用端到端的强化学习方法来解决这个问题。研究人员认为，使用强化学习方法有诸多好处：1) 对于模型的输入而言，经典模型的输入大部分为传感器观察到的实际几何体，而强化学习模型可以直接从图像中推断几何体，使得输入模式的选择更加灵活；2) 强化学习可以有效地利用现实世界的结构规律；3) 较经典的机器学习算法，强化学习可以提高对显示状态估计的精度，因此更具有鲁棒性。然而，相对于从前的经典模型，强化学习模型也具有不容忽视的缺陷：算法算力大，所需的存储空间大。

Exploration

- How to efficiently explore an unseen environment?
- Limitations of end-to-end reinforcement learning:
 - Learning about mapping, pose-estimation and path-planning is expensive
 - Sample inefficiency
 - Poor generalization
- Our solution:
 - Incorporating the strengths of learning
 - Modular and hierarchical system



图 7：导航算法的挑战及解决方法

鉴于此，Ruslan 提出了一种改进的新算法——Active Neural SLAM (ANS)。这项工作提出了一个模块化和层次化的方法来学习三维环境中的策略，称为“主动神经系统”。Ruslan 的方法利用了经典方法和基于学习的方法的优势，通过使用带学习 SLAM 模块的分析路径规划器，以及全局和本地策略。强化学习的使用使得输入模式更具有灵活性（在 SLAM 模块中），在全局的导航决策中利用了物理世界的结构规律，在局部的导航决策中并增加了对状态估计的鲁棒性。该算法在每个模块中使用强化学习方式，并保留了它的优点，同时，在训练过程中进行分层分解和模块训练，如图 8 为该算法的整体框架图：神经 SLAM 模块根据传入的 RGB 观测值和传感器读数预测地图和智能体姿态估计。

全局导航策略使用此映射和姿势输出长期目标，并使用分析路径规划器将其转换为短期目标。局部导航策略被用于训练短期目标，使得我们避开在训练过程中高样本的复杂度。该算法模型可以容易的转移到 PointGoal 任务中，并在 CVPR 2019 Habitation PointGoal 导航挑战赛中获得非常好的成绩。

Active Neural SLAM: Overview

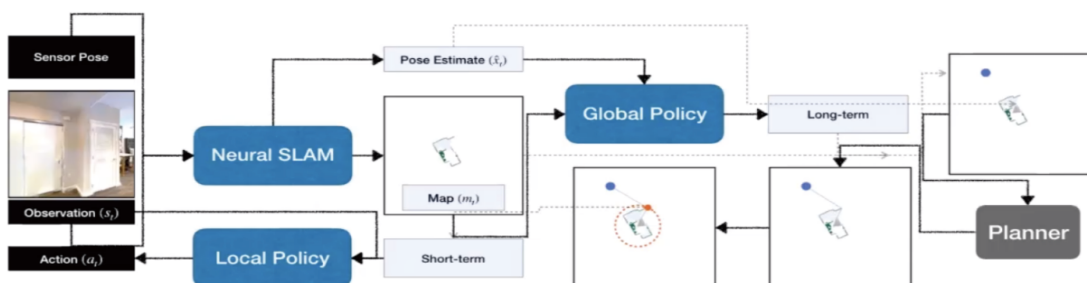


图 8：Active Neural SLAM 算法整体框架图

如图 8，在整个算法框架中我们看到，起关键作用的为蓝色的三个方框，分别为全局策略、局部策略及神经 SLAM。下面我们再具体介绍一下这三个“硬核”算法模块：

- **局部策略**，学习局部策略的即学习制定一个遵循规划者给出的计划的确定性策略。该策略旨在适应映射中的小错误。例如：神经 SLAM 模块有时会错误地预测地毯是障碍物。在这种情况下，智能代理者会计划绕过地毯。但是，如果短期目标超出后，神经 SLAM 模块仍然预测到地毯，那么算法根据观察后会反馈该信息，局部策略根据该信息理解地毯不是障碍，并学会克服它。
- **全局策略**，学习一个全局策略来抽样长期目标的另一种选择是使用一种称为基于边界的探索 (FBE) 的经典算法 (Yamauchi, 1997)。边界被定义为已探索的自由空间和未探索的空间之间的边界。基于边界的探索本质上是在这个边界上取样点作为探索空间的目标。在研究中 Ruslan 发现，在最靠近智能代理者的前沿点进行取样会产生更好的结果。定性地说，Ruslan 认为，基于前沿的探索花费了大量时间探索某些物品后面的角落或小区域。相比之下，训练有素的全局策略忽略了狭小的空间，选择了遥远的长期目标，从而提高了覆盖率。
- **神经 SLAM**，ANS 与基线的区别在于 ANS 使用额外的监督来训练姿态估计器。为了了解性能增益是否来自于这种额外的监督，Ruslan 从 ANS 中移除了位姿估计器，只使用输入传感器读数作为我们的姿势估计，结果表明，即使没有姿态估计器，ANS 仍然优于基线。同时，Ruslan 还观察到，在没有姿态估计器的情况下，小场景下的性能仅下降约 1%，而在大型场景中则下降约 10%。这是意料之中的，因为更大的场景需要更长的时间来探索，并且姿势错误会随着时间的推移而累积，从而导致偏移。将真实姿态作为输入传递给基线而不是传感器读数并没有提高它们的性能。

最后，Ruslan 总结到，ANS 提出了一个模块化的导航模型，它利用了传统导航方法和基于学习的导航方法的优点，该模型在探索和点目标任务上都优于先前的方法，并且在域、目标和任务之间表现出很强的泛化能力。未来，该模型可以扩展到复杂的语义任务，如语义目标导航和具体的问答，使用语义神经 SLAM 模块创建一个捕捉环境中对象语义属性的多通道地图。该模型还可以与先前的本地化工作相结合，以便在先前创建的地图中重新定位，以便在后续事件中高效导航。

三、结语

本次讲座中，Ruslan 为我们分享了两大挑战主题：1) 如何将领域知识应用于深度学习模型，从而提高开放式问答系统的搜索精度；2) 如何用深度强化学习，结合传统的导航方法，提高导航系统的精度及泛化能力。由于 Ruslan 所学甚广，因此在分享的过程中，他常常不局限于当下的问题及探索领域，旁征博引，为我们分享了许多经典的案例，使得听众的理解过程更加生动简洁。

哥伦比亚大学教授周以真：数据科学的使命与应用

整理：智源社区 韩鹏飞

在本届智源大会“全体大会”中，哥伦比亚大学周以真教授做了主题为：“Data for Good: Ensuring the Responsible Use of Data to Benefit Society”的演讲。在演讲中，周以真阐述了数据科学研究的使命，并介绍了哥伦比亚大学以此为基础的一些最新研究。

周以真 (Jeannette M. Wing)，哥伦比亚大学计算机科学学院教授，数据科学研究所所长，曾任微软全球资深副总裁。美国国家自然科学基金会计算与信息科学工程部助理部长。研究方向包括：可信人工智能、可信计算、隐私、安全、软件规范和验证、分布式和并发系统、编程语言、编程方法、软件工程。发表学术论文 300+，Google Scholar 他引 27000+ 次，其中论文《Computational Thinking》单篇引用 6000+。现为美国计算机学会 (ACM) 和国际电气电子工程师学会 (IEEE) 会士。个人主页：<http://www.cs.cmu.edu/~wing/>，实验室主页：<https://www.datascience.columbia.edu/columbia-data-science>

一、数据科学的使命

首先，什么是数据科学？在这里，周以真给出了一个数据生命周期示意图（如下图 1），其包含了数据生成、数据收集、数据处理、数据分析、数据可视化等过程。而数据生命周期中最令人感兴趣，也是在人工智能、机器学习中发挥重要作用的便是数据分析过程。而数据科学作为从数据中发掘价值、能够推进新知识的研究，其中的关键词就是价值。价值取决于终端用户的解释：对于科学家来说，价值就是发现新知识；对于天文学家来说，仰望星空的价值就是为了发现新的行星；对于企业家来说，价值就是利润。之前所提到的人工智能、机器学习、深度学习等都是孕育价值的工具。

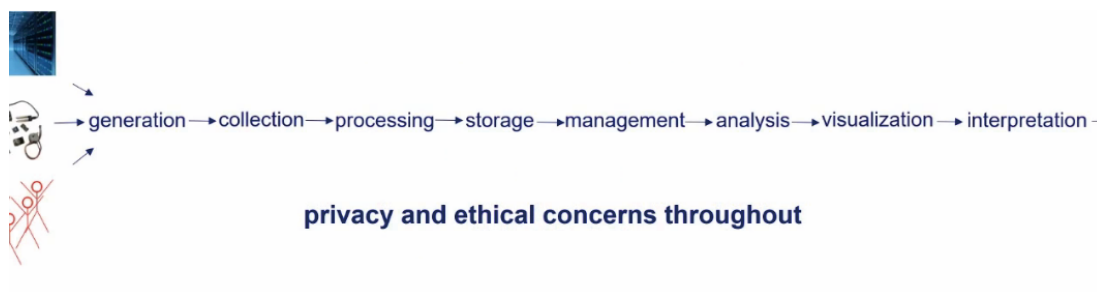


图 1：数据生命周期图

接下来，周以真为我们介绍了哥伦比亚大学数据科学研究所的主要情况。哥伦比亚大学数据科学研究所旨在培训下一代数据科学家，并开发创新技术服务于社会。该研究所与 17 个院所进行合作，拥有 350 多个附属学科，涉及多个交叉学科，致力于促进先进技术的合作，以收集和解释数据并解决社会面临的紧迫问题。该研究所还与业界紧密合作，将有前途的想法推向市场。

哥伦比亚大学数据科学研究所数据科学领域的研究，主要由数据驱动科学计算系统中心、网络安全中心、数据、媒体与社会中心、财务和业务分析中心、数据科学基础中心、健康分析中心、感应，收集和移动数据中心

和智慧城市中心八个研究中心，与计算社会科学、材料发现分析、教育三个工作小组构成。哥伦比亚大学数据科学学院是一个多学科、多专业交叉合作的学院，其研究方向主要包含健康分析、金融分析、智慧城市和计算社会科学等，数据科学的研究基石是计算机科学，统计学和运筹学。哥伦比亚大学在工程学院有很强的运筹学在商学院有很强的运筹学团队，研究方向涉及：金融、人工智能，信息安全，生物医学、天气气候、人机交互等众多新兴热点领域。部分项目合作如下图 2 所示：

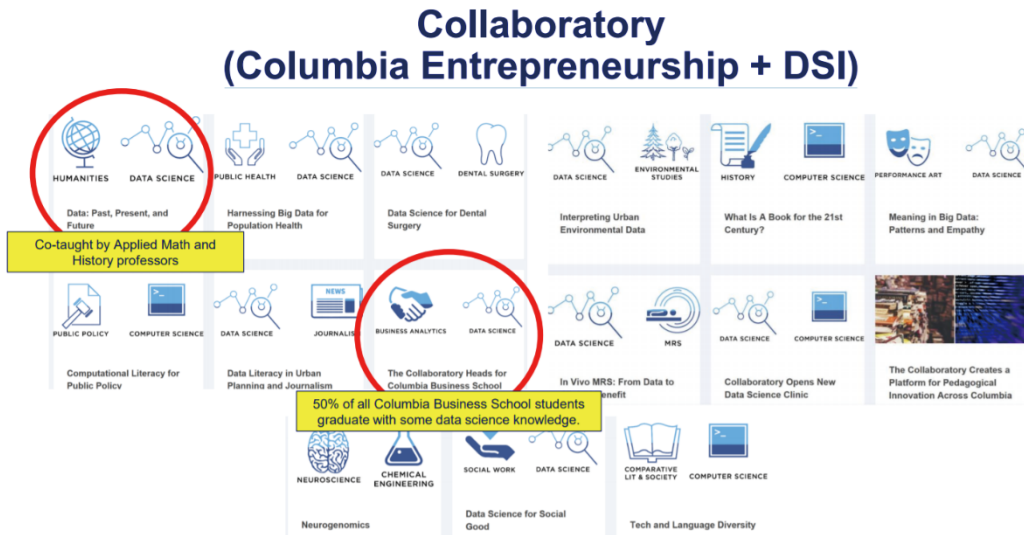


图 2：哥伦比亚大学数据科学研究所交叉合作项目示意图

哥伦比亚大学数据研究所在注重科学理论研究的同时，也着重于应用研究落地，其工业应用领域涉及制药、零售、金融等众多领域，并与阿里巴巴，百度、微软及 IBM 等众多国内外公司合作。以哥伦比亚大学数据科学研究所与 IBM 的合作为例，涉及的领域如下图 3 所示。

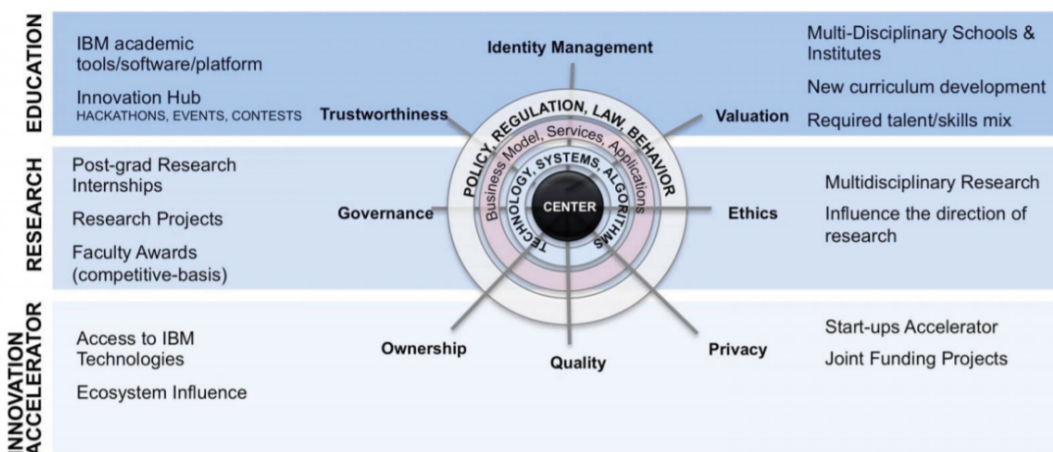


图 3：哥伦比亚 -IBM 区块链中心和数据透明度图

二、如何使用数据来造福人类社会

接下来，周以真基于建立哥伦比亚大学数据科学研究院的使命——如何使用数据来造福人类社会，介绍了未来研究工作的着重点：(1) 提高数据科学的研究水平；(2) 通过应用数据科学改变所有的领域、专业及部门；(3) 确保以负责任的方式使用数据造福社会，比如考虑人工智能的安全性和鲁棒性等。

2.1 提高数据科学的研究水平

周以真介绍，数据科学研究来自一个融合了统计学、计算机科学和社会科学的全新领域——因果推理。这对计算机科学、数据科学还有人工智能都有很重要的意义。我们想要解决的问题叫做多因果推理^[1] (Multiple Causal Inference)。传统的因果推理是单向的，一个原因带来一个结果，而在这里我们关注的是多个因果的联系。举例来说，假设导演想要拍电影，在挑选演员时想先预测一下用哪个演员能带来多少收入。我们现在有一个数据库，包含电影名、演员还有票房等数据，我们想通过统计的方式知道每个演员能转化成多少票房。这个工作的难点在于，这里面有不少混杂因子，既影响因，又影响果。比如说，电影的类别就是一个混杂因子，因为大部分时候动作片就比艺术片票房要高；还有电影的叙事方式也对票房有影响。这些混杂因子会影响建模。

所谓因果性与相关性不同，相关性指的是如果我们观测到了一个变量 X 的分布，就能推断出另一个变量 Y 的分布的话，那么说明 X 和 Y 是有相关性的。而因果性则强调，如果我们操作了某个变量 X，而这种操作 (Manipulate) 引起了 Y 变量的变化的话，那么我们才能说明 X 是 Y 的 Cause，而 Y 是 X 的 Effect。这是因果学习的基本出发点，我们要找的是这样的因果关系，而不是简单的相关关系。如图 4 所示，展示了一个经典的因果推理过程实例。

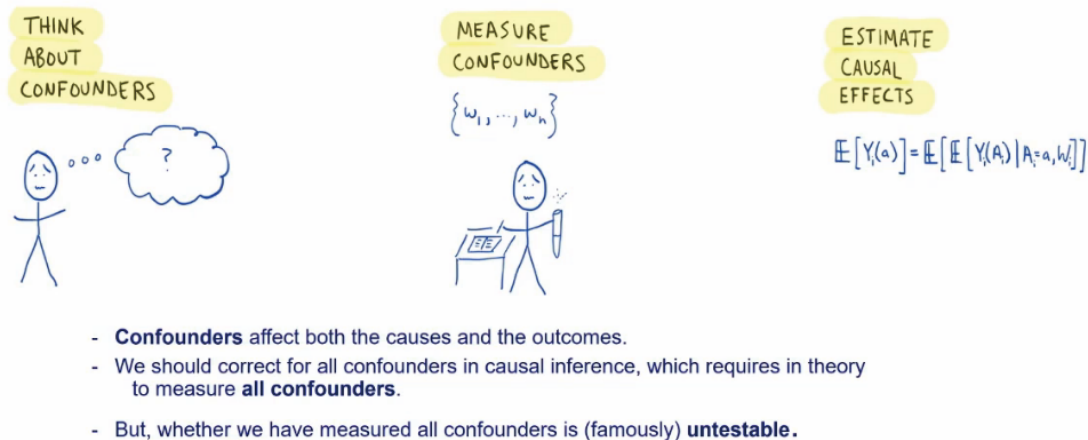


图 4：经典因果推理

周以真介绍说，传统的方法可能会先列出所有的影响因子，计算一下可能结果，但是否已经测量了所有的混杂因子是不可证伪的，于是周以真团队提出了这么一个想法——去除混杂因子 (Deconfounder)。这是一种结合了无监督机器学习和预测模型检测的算法，推断潜在变量将其作为未观察到的混杂因子的替代，然后用这个替代来执行因果推断，估计“接近真实”的因果效应。Deconfounder 有三个特点：比经典因果推断更弱的假设；对混杂因子的替代效果是可以检测的；无偏推断。

回到电影选角的例子，周以真团队通过 Deconfounder 的方法测量了 007 系列电影中演员的票房影响力，发现经过去除混杂因子后，演员 Sean Connery (James Bond) 的价值比没有去除混杂因子前提高了不少，而另外两个演员 Bernard Lee (M) 和 Desmond Llewelyn (Q) 的价值却下降了。事实上，多因果推断在现实生活中有很多应用，比如基因分析、挑选运动员还有商品定价等，有助于解决很多现实问题。如图 5 所示：

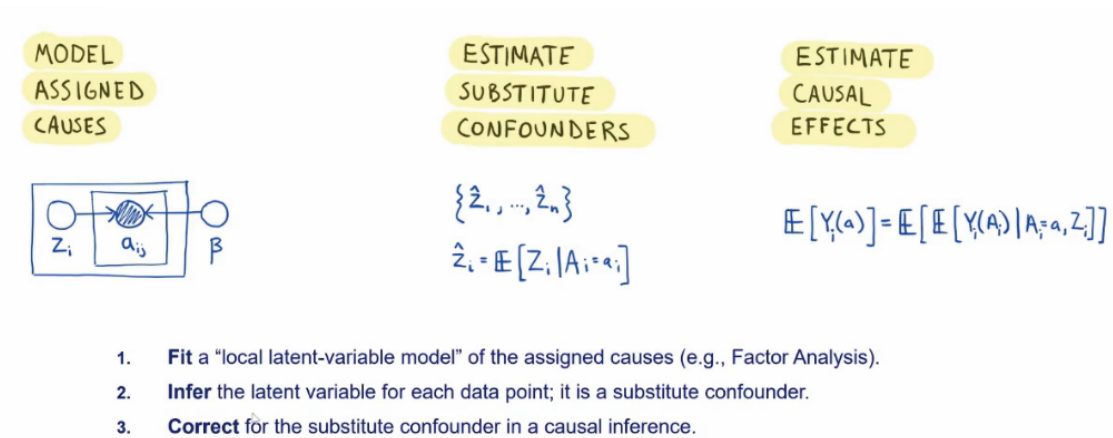


图 5：新思想——解构过程

2.2 通过应用数据科学改变所有的领域、专业及部门

数据科学的第二个使命就是可以将其应用到各行各业，及不同的研究领域，譬如金融学、生物信息学、天文天体学、信息科学等学科，以推动各行各业的转型与发展。接下来，周以真介绍了几个学科应用的例子。

比如有研究员研究了围绕胰腺癌细胞的微型炸弹的基因组^[2]。他们发现他们的微生物离子抵消了治疗肿瘤的化疗的效果。但是他们进一步发现如果你给肿瘤注射抗生素，那么这将对微型生物圈产生反作用，从而使化疗有效。

比如天文学家喜欢在任何领域使用最先进的技术来解决他们的大数据问题^[3]。他们从弱引力透镜获得的图像数据。结果用卷积神经网络来估计大爆炸模型的系数。比经典的统计方法精确得多。

哥伦比亚大学数据科学研究所和微软的一项合作研究，则是通过将机器学习算法与经济学进行结合，研究人工智能众包平台 Amazon Mechanical Turk 这样的劳动力市场^[4]，是否是一种买方垄断（即只有一个买方而有多个卖方，此时买方具有垄断性，可以付出较低的价格），是否是一个公平的劳动力市场等。

还有一个示例是将强化学习应用到金融工程中去。在当今社会中，有愈来愈多的人在使用机器人投资顾问，利用强化学习可以花费较短的时间来学习到用户的风险偏好，而且现在与机器人的交流也变得越来越人性化。经济学意义^[5]上，非常关注因果推理。不仅仅是相关性，而机器学习给了很多分析大数据的方法。经济学、大数据和机器学习算法的结合如何才能真正推动这两个领域的发展，

除此之外，周以真还介绍了历史文化方面的研究应用。哥伦比亚大学历史系已经开始利用机器学习模型、舆情分析等方法去分析历史文本^[6]。例如，每一年哥大历史系教授都会在美国政府发布的文件中搜集文本，比如搜集上世纪 70 年代外交官相互发的外交传电，通过分析和可视化这些文件，来了解 70 年代所发生的历史事件。如图 6 所示。



图 6：数据科学在历史方面的应用

2.3 Data for Good：负责任地使用数据

那么我们该如何以负责任的方式使用数据，并以数据为方式应对社会挑战？

微软曾经提出的的人工智能六大原则，周以真通过重新组合，就变成了一个缩写“FATES”，分别是公平 (Fairness)，负责任 (Accountability)，透明 (Transparency)，道德 (Ethics)，安全、保障与隐私 (Safety, Security, and Privacy)。

在谈论安全时，周以真提到了关于隐私的鲁棒性。如图 7 所示，这是一个测试深度学习系统的系统 DeepXplore^[7]，用以寻找深度学习系统可能犯的具体错误。

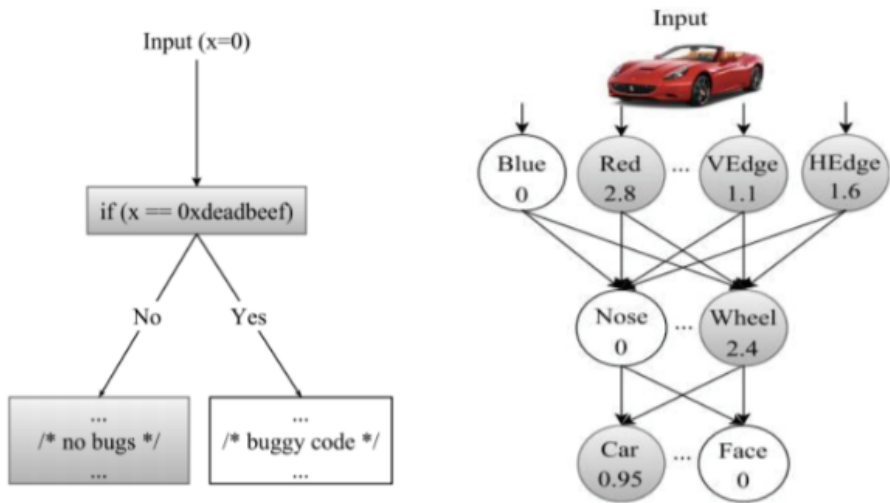


图 7：DeepXplore 算法框图

这个系统受到 2 种软件工程技术的启发，一种叫做神经元覆盖 (Neuron Coverage)，另一种是差异测试 (Differential Testing)。什么是神经元覆盖？在软件工程中，测试程序会用到代码覆盖，为软件每一个路径创建一个测试，神经元覆盖的思路与之类似，比如创建一些输入的事件实例去覆盖每一个神经元，然后我们就会发现很多错误。该系统使用这两种技术定义一种算法，基本上比较两种不同的 DNN，并展示如何修改输入，从而对 DNN 缺失进行分类，即变化图像。如图 8 所示，如果你看到左边的图像，这是自动驾驶汽车上的典型图像，汽车会按照正确的引导，向左行驶；但如果你稍微将图像变暗，那么 DNN 将不幸地作出错误的分类，并告诉汽车向右行驶——结果导致穿过护栏、越过悬崖以及人员伤亡。所以重点是系统能够找出导致 DNN 做错事的干扰输入。通过这个系统，他们揭示了 15 种当前最佳的深度学习模型的数千个不同的不正确的极端情况。这项工作获得了 SOSP 最佳论文奖。

DeepXplore

<https://github.com/peikexin9/deepxplore>

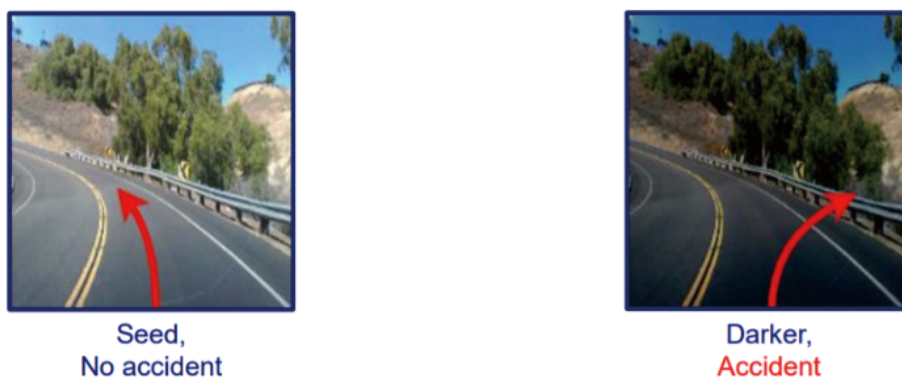


图 8: DeepXplore 算法效果图

周以真介绍的另一个项目是 PixelDP^[6]，缘起是现在有很多研究都揭示出 DNN 的脆弱性，连非常简单的涂鸦都会改变 DNN 的分类结果，而在自动驾驶时如果认错了交通标识，可能会导致非常严重的后果。这项研究工作受到了差分隐私 (Differential Privacy) 的启发，该理念来自于密码学，现在被嫁接到了机器学习中，试图使 DNN 更加强大，以应对图像中的污染等情况。在该项研究中，DNN 被增加了一个噪声层，可以确保将输入中污染的影响控制在一定范围内，让分类器不会认错。

Problem

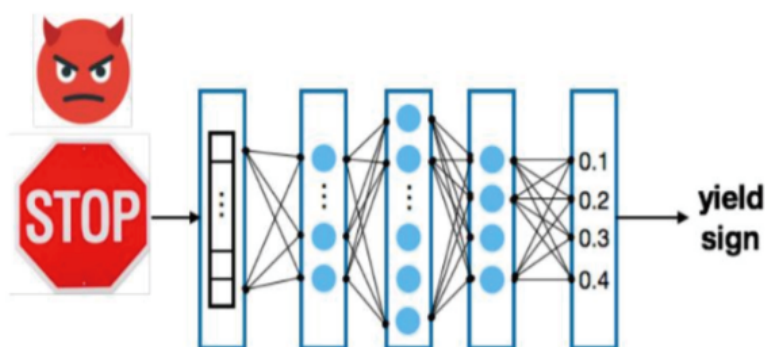


图 9: PixelDP 算法框图

对于来自现实社会的挑战，这些研究还处于起步阶段，但是周以真团队希望可以通过数据科学来帮助人们应对这些社会挑战。周以真介绍，目前哥伦比亚大学数据科学家们正在研究解决的几个挑战问题是：

(1) 如何利用大数据及机器学习方式，来更好的帮助农民们选择适合播种的农作物^[9]，来减少由于气候对农产品产生的影响？比如研究者们发现，如果印度的作物生产继续向稻米趋同，该国的粮食供应，可能更容易受到其后的影响（如干旱，酷热）。增加粗粮（如小麦和高粱）的产量份额，可以提高印度粮食生产抵御气候变化的能力，特别是在粗粮产量已与稻米产量相当的地方。在重要的农业地区实现农作物结构的多样化可帮助缓解某些气候变化（如干旱、酷热），因此通过研究对印度的作物种植多元化给出了一些建议，有助于抵御气候变化的不利影响。

(2) 流行病方面存在的挑战。周以真介绍，哥伦比亚大学已经建立了一个联邦数据集^[10]，包含 10 亿个病人的病例图，有 600 万个病人记录，500 万个独特的病人记录，它们来自 25 个不同的国家，基本上是从 80 个不同的数据库中提取，并且都是用同样的格式对数据进行了收集整理。周以真认为这些数据令人惊叹，可以使我们深入了解仅靠临床了解不到的医疗情况。



图 10: 哥伦比亚大学流行病数据库介绍

糖尿病，高血压和抑郁症三种疾病可视化结果如图 11 所示。

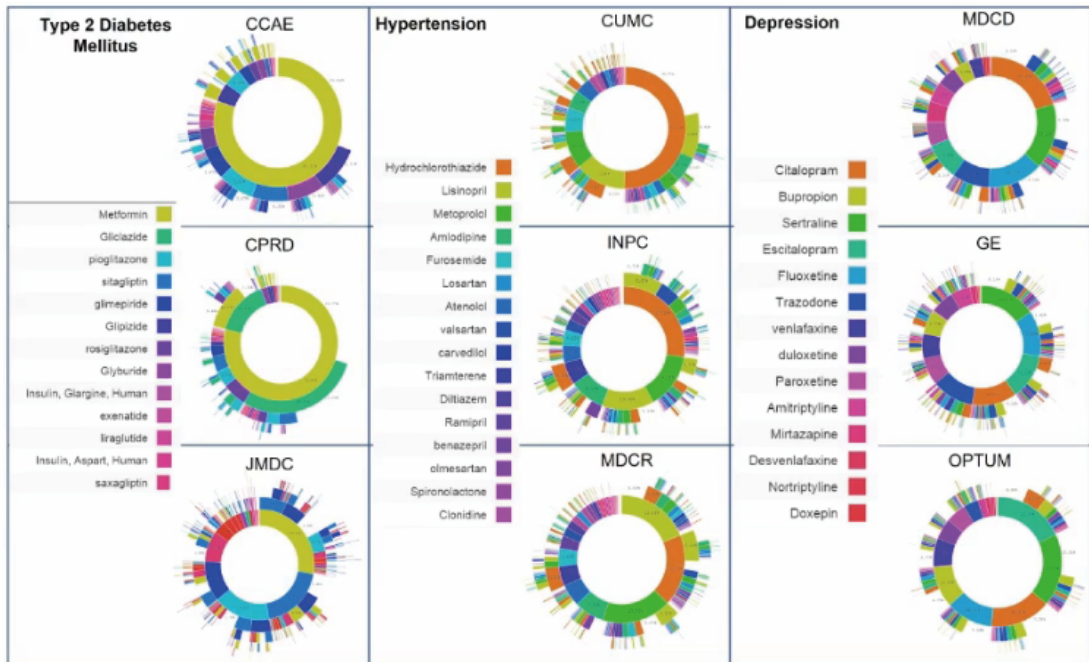


图 11：疾病可视化结果图

周以真指出，当我们有一个很大的数据集时，我们可以进行因果推理。所以我们可以想象构建电子健康档案^[11]是为了使用我前面描述的这个特定数据库的技术，其推理结果如图 12 所示。

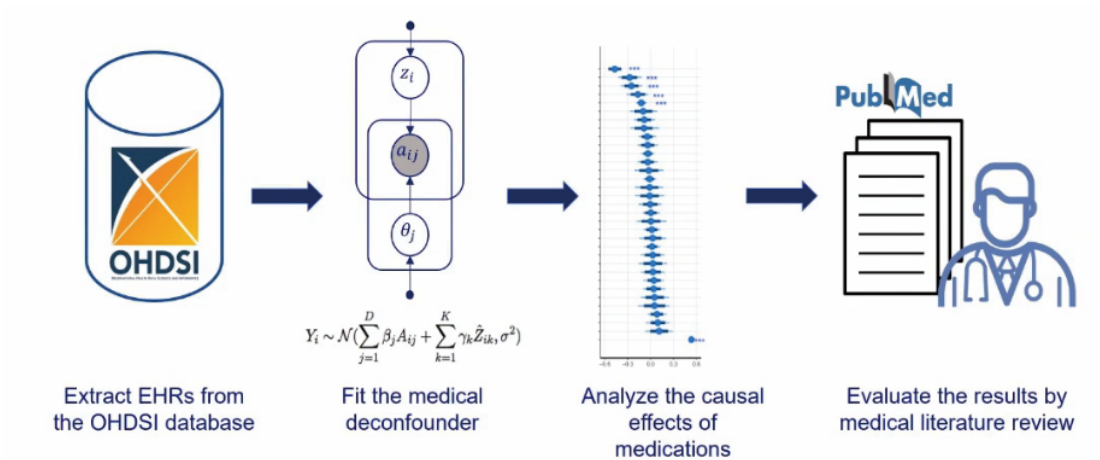


图 12：因果推理算法框图推理

通过该推理过程，我们可以看出，通过将深度学习与医学相结合，去除混杂因子降低了假阳性率和假阴性率。从因果关系到非因果关系的推理，帮助找到了更接近真实的治疗效果评估，它还确定了更符合医学文献中发现的有效药物。

参考文献:

- [1] Yixin Wang and David M. Blei, “The Blessings of Multiple Causes,” arXiv:1805.06826v2 [stat.ML], June 19, 2018
- [2] Geller L T, Barzily-Rokni M, Danino T, et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine[J]. *Science*, 2017, 357(6356): 1156–1160
- [3] Arushi Gupta, José Manuel Zorrilla Matilla, Daniel Hsu, Zoltán Haiman, “Non-Gaussian information from weak lensing data via deep learning,” *Physical Review D*, in press (accepted April 30, 2018), E-print available at <https://arxiv.org/abs/1802.01212>
- [4] Arindrajit Dube, Jeff Jacobs, Suresh Naidu, and Siddharth Suri, “Monopsony in Online Labor Markets,” forthcoming, *American Economic Review: Insights*, August 2018.
- [5] Agostino Capponi, Octavio Ruiz Lacedelli, and Matt Stern, “Robo-Advising as a Human-Machine Interaction System”, August 2018, preprint
- [6] Allison J. B. Chaney, Hanna Wallach, Matthew Connelly, and David M. Blei, Detecting and characterizing Events, in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, November 2016.
- [7] Kexin Pei, Yinzhi Cao, Junfeng Yang, and Suman Jana, “Deep Xplore: Automated Whitebox Testing of Deep Learning Systems, *Proceedings of the 26th ACM Symposium on Operating Systems Principles*, October 2017, Best Paper Award.
- [8] Mathias Lecuyer, Baggelis Atlidakis, Roxana Geambasu, Daniel Hsu, and Suman Jana, “Certified Robustness to Adversarial Examples with Differential Privacy, arXiv:1802.03471v2 , June 26, 2018, to appear *IEEE Security and Privacy (“Oakland”)* 2019.
- [9] Kyle F. Davis, Ashwini Chhtre, Narasimha D. Rao, Deepti Singh, Ruth DeFries, *Environmental Research Letters*, Volume: 14, Article number: 064013, <https://doi.org/10.1088/1748-9326/ab22db>
- [10] George Hripcsak, Patrick B. Ryan, Jon D. Duke, Nigam H. Shah, Rae Woong Park, Vojtech Huser, Marc A. Suchard, Martijn J. Schuemie, Frank J. DeFalco, Adler Perotte, Juan M. Banda, Christian G. Reich, Lisa M. Schilling, Michael E. Matheny, Daniella Meeker, Nicole Pratt, and David Madigan, “Characterizing treatment pathways at scale using the OHDSI network,” *PNAS Early Edition*, April 2016.
- [11] Linying Zhang, Yixin Wang, Anna Ostropolets, Jami J. Mulgrave, David M. Blei, George Hripcsak, “The Medical Deconfounder: Assessing Treatment Effect with Electronic Health Records (EHRs),” arXiv:1904.02098v1, April 2019.

图灵奖得主 Judea Pearl: 新因果科学与数据科学、人工智能的思考

整理：智源社区 龚鹤扬 高亦斌

在本次智源大会中，图灵奖得主 Judea Pearl 做了主题为《The New Science of Cause and Effect with reflections on data science and artificial intelligence》的演讲。

Judea Pearl (朱迪亚·珀尔)，加州大学洛杉矶分校 (UCLA) 教授，图灵奖 (Turing Award) 得主。他在 20 世纪 80 年代开发并倡导了 AI 的概率方法，被称为贝叶斯网络之父。然而为了强人工智能的愿景，Pearl 脱离主流 AI 研究社区，提出了一套因果的数学语言和理论，引领了正在席卷各个学科的“因果革命”。他自己最引以为傲的工作是“The fundamental law of counterfactuals。”^[4]

Judea Pearl 个人 Slogan: [To Build Truly Intelligent Machines, Teach Them Cause and Effect.](#)

在演讲中，Judea Pearl 站在整个数据科学的视角，简单回顾了过去的“大数据革命”，指出数据科学正在从当前以数据为中心的范式向以科学为中心的范式偏移，现在正在发生一场席卷各个研究领域的“因果革命”。Pearl 解释了什么是因果科学以及相关新逻辑和推理引擎的思想脉络，包括介绍了被称之为“Double-Helix”的两个因果推理的基本定理，并以其作为出发点推演出因果图模型框架，最后概述了该框架下因果推理的七大工具。另外，为了帮助读者们更加透彻地理解 Pearl 的因果推理思想，我们结合 Judea Pearl 近年来的论文、访谈和报告，以及其学生 Elias Bareinbion、马普智能所 Bernhard Scholkopf 团队等的相关研究工作，对本报告进行了一定的补充说明。

一、因果革命：改变数据科学的新革命

在报告中，Pearl 首先介绍了一场正在改变数据科学的新革命——“因果革命”。因果革命和以数据为中心的第一次数据科学革命，也就是大数据革命（涉及机器学习，深度学习机器应用，例如 Alpha-Go、语音识别、机器翻译、自动驾驶等等）的不同之处在于，它以科学为中心，涉及从数据到政策、可解释性、机制的泛化，再到一些社会科学中的基础概念信用、责任和公平性，甚至哲学中的创造性和自由意志。可以说，因果革命彻底改变了科学家处理因果问题的方式。

- **Data science is a two-body problem, connecting data and reality, including the forces behind the data.**
- **Data science is the art of interpreting reality in the light of data, not a mirror through which data sees itself from different angles.**
- **The ladder of causation is the double helix of causal thinking, defining what can and cannot be learned about actions and about worlds that could have been.**

图 1：Pearl 关于数据科学本质的洞见，见文献^[3]

因果革命中，数据科学的任务被重新分成了三类：预测，描述和反事实预测^[6]。关于它具体如何席卷各个学科，详情可参见：

- Bernhard Scholkopf 最引以为傲的论文之一《Causality for Machine Learning》，它概述了信息革命时代下因果和机器学习的融合的基本原理和深刻思考^[5]，
- 《Causal Inference and Data-Fusion in Econometrics》是 Elias Eareinboim(Pearl 学生) 关于因果结合经济学领域的最新综述^[7]。
- 因果也影响了社会科学，医疗健康科学，计算机和统计学等，见资料^[1,8]。

哈佛大学教授 Gary King (2014) 盛赞了这场因果革命，它指出“过去三十年关于因果理论的进展超过了人类前面积累的总和。” Pearl 继续解释说，成百上千过去认为不可能解决的问题，现在可以通过简单的数学和可计算的算法解决。在介绍完“因果革命”之后，Pearl 接着介绍了本次报告的大纲：

- 什么是因果科学，为什么它需要新的逻辑和推断引擎
- 如何让机器获得因果推理的能力（因果推理引擎的结构）
- 因果推理的两个基本定律
- 因果智慧的七个工具

二、什么是因果科学？

当前曲线拟合的机器学习和深度学习取得了巨大的成功，为什么需要研究因果^[8]？ Pearl 在去年接受 Lex Fridman 访谈^[4]时提到“Everything starts with the question: What is the research question?”。

而 Pearl 在这次报告中，则用了几个统计学中的经典例子。第一个问题是：“锻炼身体是否有利于健康？”见下图，x 轴表示运动时间，y 轴表示胆固醇水平。

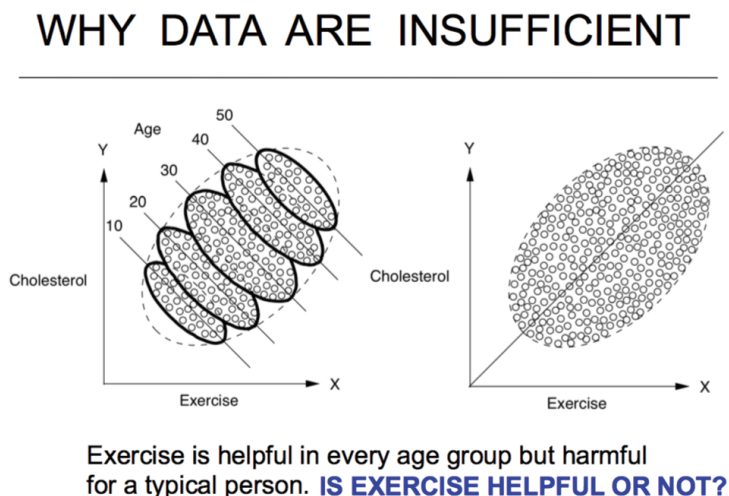


图 2：锻炼是否有利于健康？

一方面，在图 2 (左) 中，可以看大每个年龄组中都出现了向下的趋势，表明运动可能的确有降低人体胆固醇水平的效果；另一方面，在图 2 (右) 中，同样的散点图并不依据年龄对数据进行分层，那么我们会看到一个明显向上的趋势，这表明运动得越多，人体胆固醇水平就越高，这种矛盾在统计学中被成为辛普森悖论。Pearl 介绍的另外两个例子，一个是关于“药物”、“性别”、“死亡率”的研究问题：“药物有效果吗？”另外一个是关于“疫苗”、“天花”、“死亡率”的研究问题：“疫苗有效果吗？”

这几个例子共同说明了数据可能对你讲出两个不同的故事。如果信息发生了一些变化，得到的结论就可能是不一样的。更加准确地说，这几个例子本质上是要回答因果问题，仅有数据信息而没有先验因果关系信息的时候，就可能得出与关注研究的问题相互矛盾的答案。**回答因果问题需要因果信息。**

Pearl 在这次报告中指出，因果科学始于因果问题，因果科学研究如何回答因果问题。

什么是因果问题呢？他给了几个简单的例子：

1. 给定的治疗方法在预防疾病方面效果如何？
2. 是新的减税政策导致销售额上升吗？还是我们的营销活动？
3. 肥胖引起的年度医疗保健费用是多少？
4. 雇佣记录可以证明雇主犯了性别歧视吗？
5. 我即将辞职，我会后悔吗？

他解释到上面这五个因果问题，因为这些问题都包含着不对称信息，所以无法用现在标准的科学语言，也就是具备对称性的数学公式来描述。相对于“=”表示对称信息，他用箭头 \rightarrow 表示非对称信息，见下图：

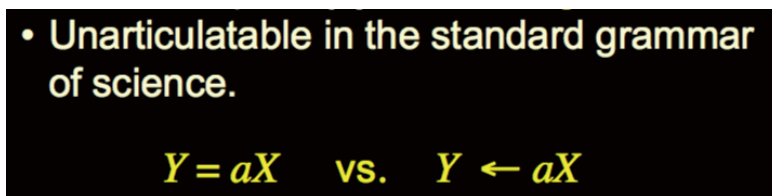


图 3：对称 VS 非对称

在过去的 30 年中，Pearl 和他的同事找到了非对称性的表达工具。他认为因果科学是回答因果问题的逻辑和工具，也就是推理引擎。通俗来说它有三个输入，包括我们想知道什么、我们已经知道什么和可用数据，以及作为输出的两类关注问题的答案：a) 现在某个行动会有什么结果？ b) 过去换个选择会有什么不同的结果？

因果推理是人类思想中不可或缺的组成部分，应该对其进行形式化和算法化处理，以实现人类水平的机器智能^[3]。

Pearl 描述了一个因果推理的三级结构，把因果信息按其能够回答的类型进行分类。该分类形成了一个三层的层级结构，某层的问题，只有在获取不低于该层信息时，才能够被回答。

a) 三个因果层级

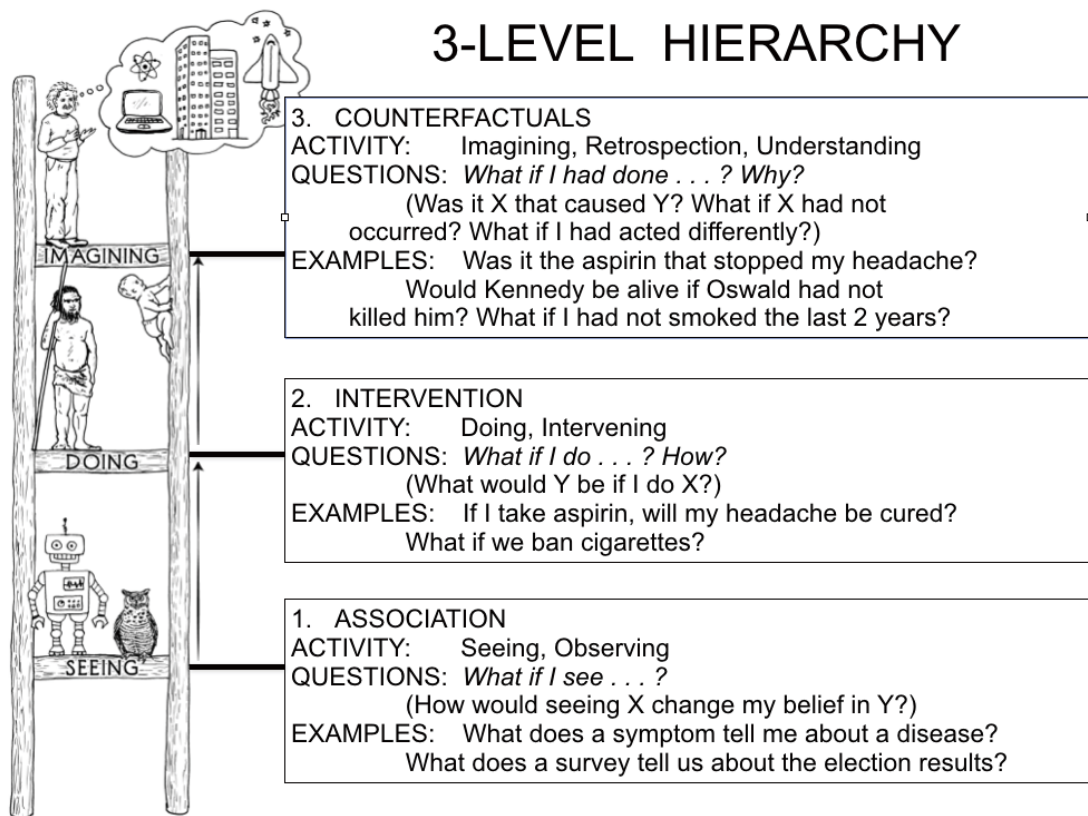


图 4：三个因果层级，参见书籍^[2]

第一层是关联 (Association)，它涉及由数据定义的统计相关性。大多数机器学习系统运行在这一层上。

第二层是干预 (Intervention)，不仅涉及能看到什么，还涉及一个干预或行动将会导致什么结果。作为例子，Pearl 提了一个问题：“如果我们把价格翻倍，将会发生什么？”

第三层是反事实 (Counterfactual)，是对以前发生的事情的反思和溯因，解决的是“如果过去作出不一样的行为，现在的结果会有何不同？”的问题。

顶层也就是反事实层是功能最强大的层次，如果我们有一个可以回答反事实问题的模型，那么我们也可以回答有关干预和观察的问题。例如，干预问题：What will happen if we double the price? 可以通过反事实问题来回答：What would happen had the price been twice its current value? 同样，一旦我们回答了干预问题，就可以回答关联问题。我们只是忽略了干预动作部分，而是让观测取代了。但是在相反的方向上，干预问题不能仅凭观测信息 (也就是统计相关性) 回答，涉及反思和溯因的反事实问题也不能仅用从随机对照实验中获得的干预信息来回答。

反事实是科学思维以及法律和道德推理的基础。举个例子，在法庭判定被告是否应该负法律责任的时候，判定有罪的一个依据是——“若非”被告的行为，损失就很可能不会发生 (For example, in civil court, a defendant is considered responsible for an injury if, **but for** the defendant's action, it is more likely than not the

injury would not have occurred)。“若非”的计算含义要求将现实世界与被告未发生某行为反事实世界进行比较。用个更通俗的例子来讲，已知的事实是“一个人吃了药死了”，对应的一个反事实问题是“如果此人没有吃药，不死的概率是多少？”

Pearl 在报告中指出，理解**因果推理**需要抓住一个窍门，那就是区分 **seeing** 和 **doing** 的不同，一个简单例子就是某个便利店中“观测到某商品的价格翻倍”和“店主强制让某商品价格翻倍”存在区别。Pearl 发明了 do 算子来数字化表示干预或行为，有了它我们能数学公式区分 seeing 和 doing:

$$P(y|x) \neq P(y|do(x))$$

这里我们借用文献^[5]中一个表格补充说明，表格的行是不同类型的问题，而列是不同类型模型:

表 1: 从统计模型到因果模型，再到物理模型^[5]

Table 1: A simple taxonomy of models. The most detailed model (top) is a mechanistic or physical one, usually in terms of differential equations. At the other end of the spectrum (bottom), we have a purely statistical model; this can be learned from data, but it often provides little insight beyond modeling associations between epiphenomena. Causal models can be seen as descriptions that lie in between, abstracting away from physical realism while retaining the power to answer certain interventional or counterfactual questions. See also [Mooij et al. \(2013\)](#) for a formal link between physical models and structural causal models.

Model	Predict in IID setting	Predict under distr. shift/intervention	Answer counter-factual questions	Obtain physical insight	Learn from data
Mechanistic/physical	yes	yes	yes	yes	?
Structural causal	yes	yes	yes	?	?
Causal graphical	yes	yes	no	?	?
Statistical	yes	no	no	no	yes

可以看到统计模型只有关联层的信息，所以只能回答相关性问题的，而不能回答干预问题和反事实问题。基于图的因果贝叶斯网络因果只有干预层的信息，所以只能回答干预和关联层的问题，而不能回答反事实问题。最后基于结构的因果模型，它的能力最接近物理模型，三个层级的问题都能够回答。事实上，朱松纯教授^[10]在报告的最后向 Pearl 提出了因果建模的定位问题，他说他喜欢**因果建模的思想**“From data to science”，从上面的表格中可以看到因果模型是从统计模型走向物理模型的过渡。

b) 因果推理引擎

Pearl 提出了一套基于结构的关于因果的数学语言和理论，作为因果科学是回答因果问题的推理引擎，该引擎的特点是“Knowledge in, Knowledge out, Data in between”，而基本出发点是因果推理的两大基本定律:

THE TWO FUNDAMENTAL LAWS OF CAUSAL INFERENCE

1. The Law of Counterfactuals (and Interventions)

$$Y_x(u) = Y_{M_x}(u)$$

(Y_x is equal to Y in a mutilated model M_x .)

2. The Law of Conditional Independence (d -separation)

$$(X \text{ sep } Y | Z)_{G(M)} \Rightarrow (X \perp\!\!\!\perp Y | Z) =$$

(Separation^P in the model \Rightarrow independence in the distribution.)

图 5：因果推理的两大基本定律

他指出第一个定律是关于反事实的信息，可以推演出需要使用函数来刻画变量之间的因果关系，而第二个定律刻画了因果图结构，因果图上每个每条缺失的边都意味着在给定某些变量下的条件独立性，可用它做模型检验，结构学习和因果问题的符号演算。他以这两大基本定律为出发点，发展出了被称为结构因果模型 (SCM) 的数学框架，该框架能够回答三个层级的因果问题。在如下的例子中，模型用函数关系表示，而因果图上的缺失的边 CW 和 SR 都意味着给定某些变量之下的条件独立性。

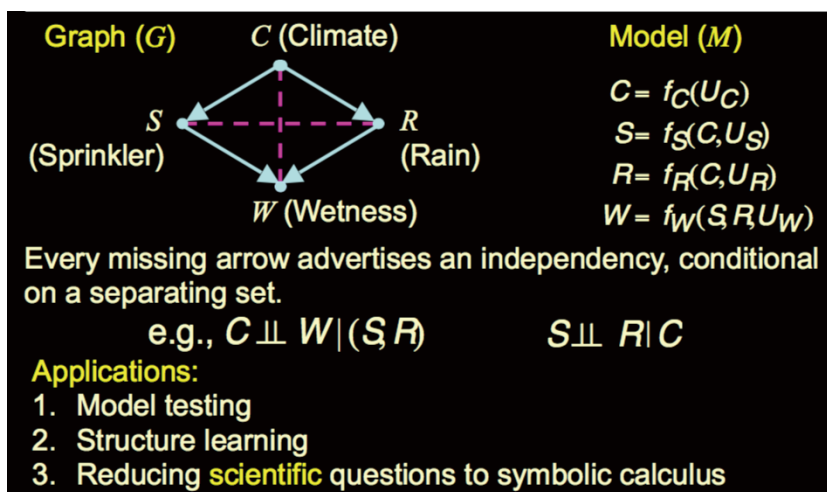


图 6：一个简单结构因果模型

现代因果建模工具的发展已对所有数据密集型科学（尤其是社会科学和流行病学）产生了变革性的影响，其中因果图已成为它们的第二语言。在这些学科中，因果图模型帮助科学家从观测数据中提取因果关系，并解构了困扰研究人员数十年的悖论。

我们根据 Pearl 的论文^[3]补充说明因果科学的推理引擎，它由三个部分组成：图模型，结构方程以及反事实和干预逻辑。图模型是一种语言，用于表示 Agent 对世界的了解。反事实帮助他们阐明他们想知道的事情。结构方程将两者以扎实的语义联系在一起。该推理引擎将假设（以图模型的形式）、数据和 Query 作为输入。

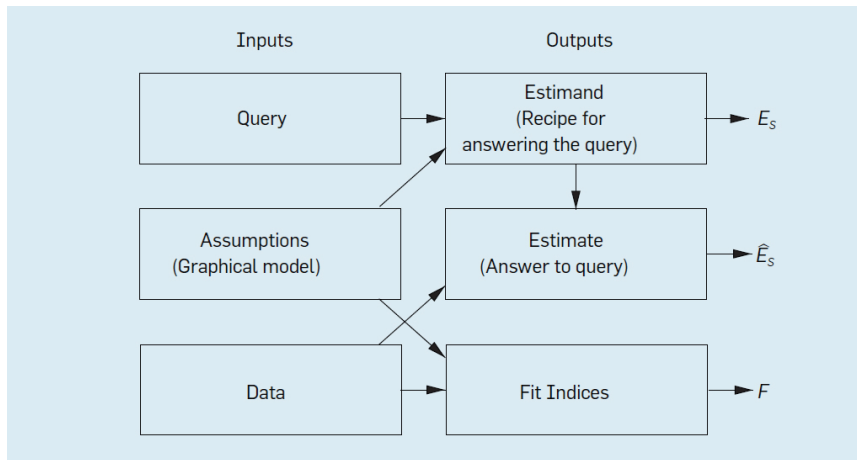


图 7: SCM 推理引擎如何结合数据和因果模型回答因果问题，见文献^[3]

SCM 推理引擎存在三个输出:

- 估计式 (Estimand) 是关注的查询 (Query) 的某个概率表达式，表示在已有模型假定下计算 Query 的一种方法;
- Estimate 是用某种统计方法和已有数据对 Estimand 概率表达式的估计;
- 一组拟合指标 (Fit Indices) 用于衡量数据与假设的兼容程度。

第一个输出是较难理解的，如果已有模型假定下某 Query 无法回答，也就是没有对应的 Estimand，则称该 Query 为“不可识别”，Pearl 的 do-calculus 就是判断 Query 是否可识别的一个完备的演算工具。Pearl 仅用抽象的方式描述了这个推理引擎，关于该引擎如何回答因果问题，如何解决数据科学中的混杂偏差，选择偏差和迁移学习问题的细节可以参考 Elia Bareinboim 的论文^[7]。

三、因果推理七个工具

接着 Pearl 概述了通过因果科学的推理引擎完成七个任务以及每个任务中使用的工具，并讨论了每个工具对自动推理技术的独特贡献。下面是因果智慧的七大工具:

- Tool 1. Encoding causal assumptions in transparent and testable way.
- Tool 2. Predicting the effects of actions and policies.
- Tool 3. Computing counterfactuals and finding causes of effects (attribution, explanation, susceptibility).
- Tool 4. Computing direct and indirect effects (Mediation) (discrimination, inequities, fairness)
- Tool 5. Integrating data from diverse sources (external validity and selection bias).
- Tool 6. Recovering from missing data.
- Tool 7. Discovering causal relations from data.

第一个工具就是使用因果图透明的编码了因果知识，使用 d- 分离图准则下的条件独立性来检验因果结构先验假设。Pearl 给出了 Shrier 和 Platt 于 2008 年提出的运动医学领域的简单例子（热身对于运动损伤的影响），在这个例子中变量之间相互如何相互影响被透明的表示了出来，变量之间因果关系的确定可以基于合理的事实，它能用 d- 分离图准则下的条件独立性来检验。

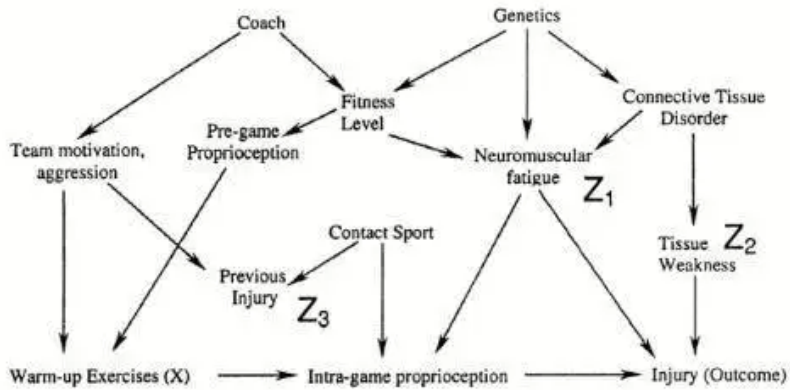


图 8：一个因果图模型实例

第二个工具就是使用 Do-Calculus 定义和回答干预层的因果问题，这些问题包括计算某个行为，干预或策略产生的因果效应。继续看热身对于运动损伤的因果效应例子，我们需要选择哪些变量作为控制变量呢？是不是控制变量越多越好呢？Pearl 的回答是，一定不要控制变量 Z3（也就是 Previous Injury）否则下图的路径会被联通，造成混杂偏差，不能正确估计回答该因果效应的问题，正确的做法是控制变量 Z1, Z2。具体 Do-Calculus 解决混杂偏差建议参考文献^[7]。

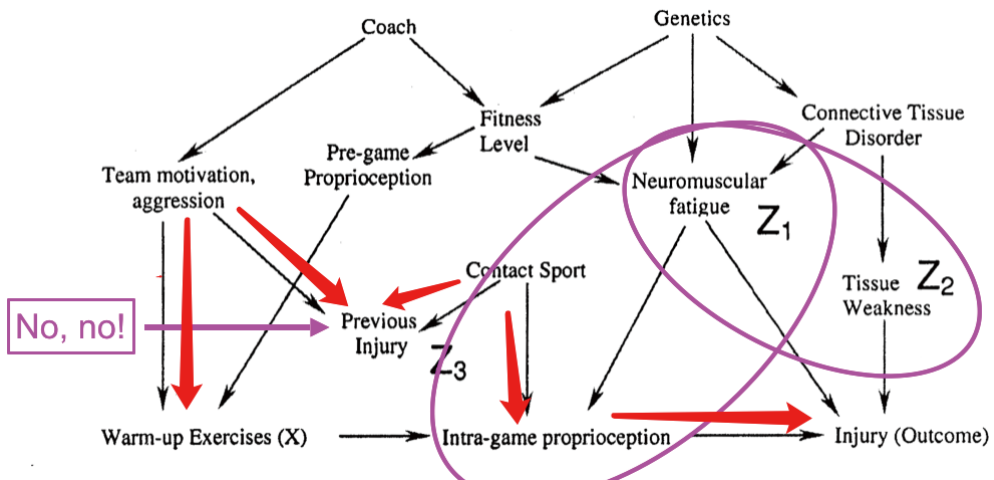


图 9：Do-Calculus 确定控制变量

第三个工具是计算反事实，找出某个结果的原因。我们要去找到事件结果的原因，我们要进行归因，进行解释。反事实是科学思维以及法律和道德推理的基础，例如，反事实问题“一个人吃了药死了，如果此人没有吃药，不死的概率是多少？”的答案是法律责任判定中的重要依据。

ATTRIBUTION

- Your Honor! My client (Mr. A) died **BECAUSE** he used this drug.



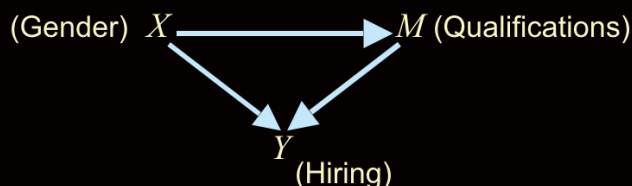
- Court to decide if it is **MORE PROBABLE THAN NOT** that Mr. A would be alive **BUT FOR** the drug!
- $PN = P(\text{alive}_{\text{no drugs}} \mid \text{dead, drug}) \geq 0.50$

图 10：反事实概率回答归因问题

第四个工具是中介分析。直接（间接）因果效应是指一个变量对另一变量的既定影响在多大程度上是直接的（间接的），这是许多不同学科都关注的一类重要因果问题，它们可用 Pearl 的推理引擎解决。这一工具现在也被用于判断「歧视」、「不公正的做法」、「不公平现象」，我们要讨论人工智能和机器学习领域中的公平性问题时，我们必须考虑中介效应，Pearl 给出了一个雇主是否有性别歧视的例子，也就是回答如何用数据判断雇主是否用性别决定录用员工与否的问题。

LEGAL DEFINITION OF DISCRIMINATION

Can data prove an employer guilty of hiring discrimination?



The Legal Definition:

Find the probability that “the employer **would** have acted differently **had** the employee been of different sex and qualification **had been** the same.”

图 11：中介分析回答性别歧视与否问题

第五个工具是泛化和数据融合。它的基本问题是关于如何融合分布不同的实验性研究或观测性研究，Pearl 及其学生 Elias Bareinboim 提出一个解决此任务的框架。Pearl 在文献^[1]中，解释了因果建模工具能够帮助解决 AI 应用中鲁棒性或适应性的原因。Pearl 在报告中用一个例子简单的讲解了一下基本思路，首先把数据源的数据类型（观测 / RCT）和其他特点用图表（上）列出来，然后转化称对应的因果图（下），其中图中 S 节点标记了数据源的 S 指向变量有差异。这样就可以用因果建模工具综合不同数据源，回答关注总体的感兴趣因果问题，其详情见文献^[7]

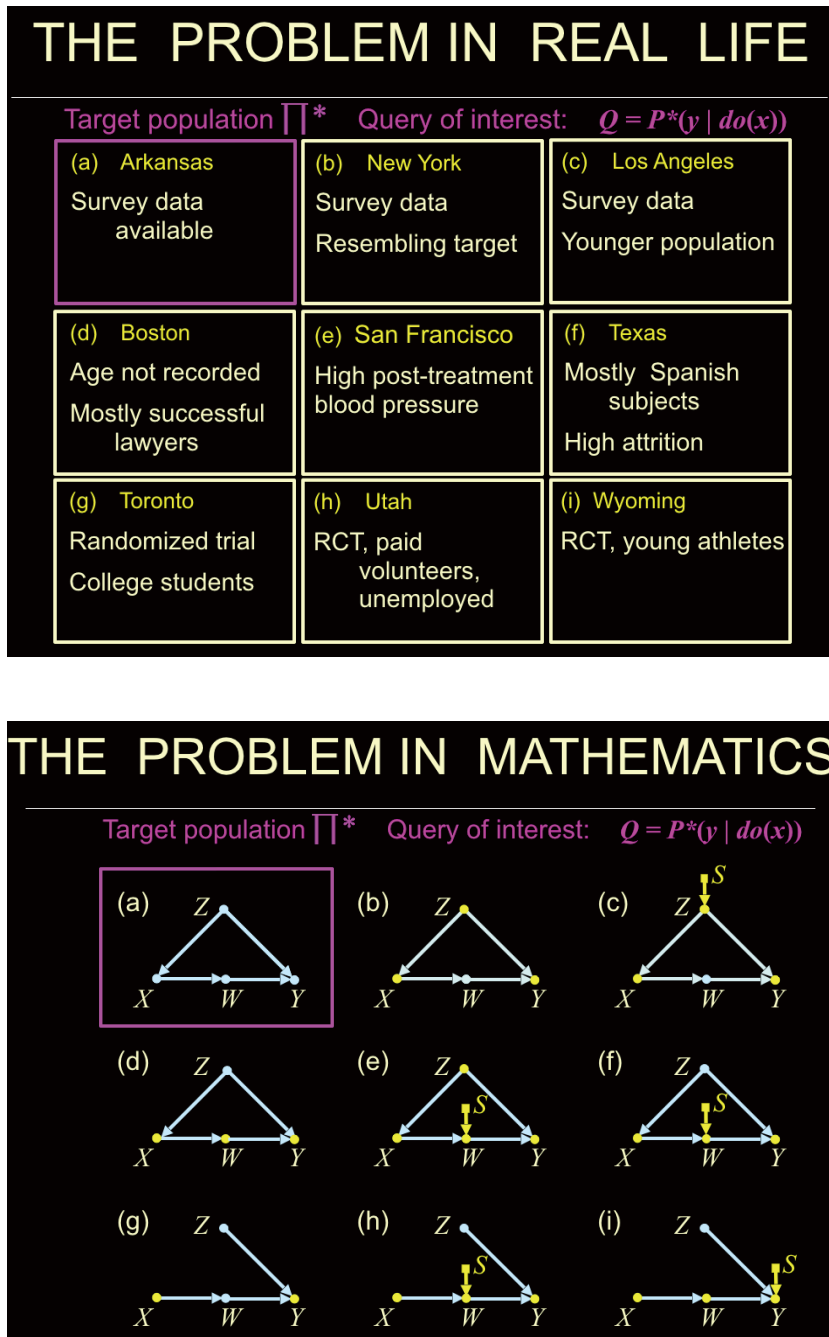
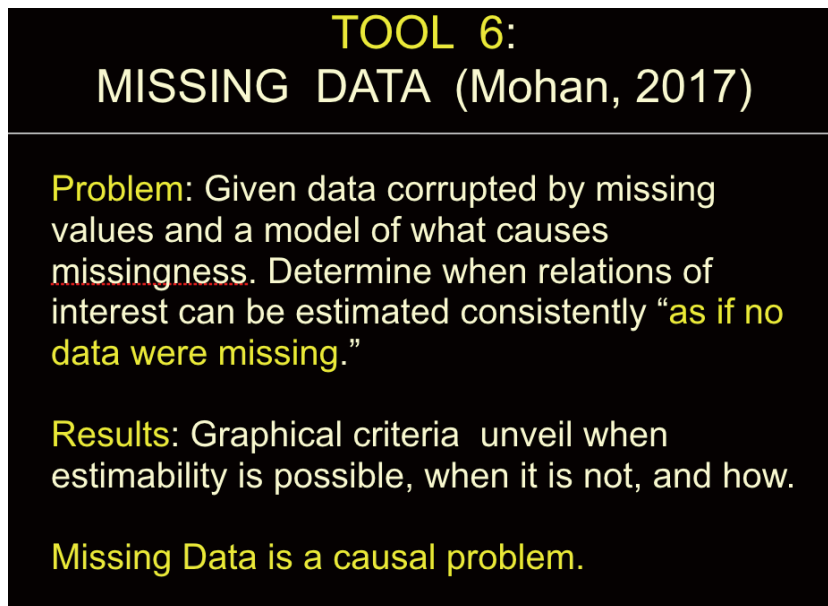


图 12：因果推理解决多源数据集融合问题实例

第六个工具是从缺失数据恢复。由于缺少数据而导致的问题困扰着实验科学的每个分支，所有的数据都会有一个缺失值的问题，它本质是一个因果问题。实际上建立一个缺失数据因果模型之后，可以用因果推理引擎判断和实现从缺失数据下的所关注问题的推断。



TOOL 6:
MISSING DATA (Mohan, 2017)

Problem: Given data corrupted by missing values and a model of what causes missingness. Determine when relations of interest can be estimated consistently “as if no data were missing.”

Results: Graphical criteria unveil when estimability is possible, when it is not, and how.

Missing Data is a causal problem.

图 13：缺失数据问题本质是因果问题

第七个工具是因果发现。因果发现就是要去寻找一系列的模式或者图结构，能够与数据相兼容的，同时能够进行简洁的表示。当前从数据中得到因果结构的因果结构学习，不仅要学习因果结构，还要学习从数据中直接得出哪些变量是因果变量的因果表示学习，这已经成为了当前一个热点研究课题，Pearl 在文献^[2]中提出了为 AI 提出小图灵测试作：

*How can machines **represent causal knowledge** in a way that would enable them to access the necessary information swiftly, answer questions correctly, and do it with ease, as a human can?*

因此因果表示学习是一个十分重要课题，向更多了解相关内容建议参考资料^[1, 2, 5, 9]。

四、结论

最后，Pearl 指出，因果革命不仅仅席卷了各个研究领域，也开始影响教育和实际应用，将数据科学从当前以数据为中心的范式往科学为中心的范式偏移。这个革命的高潮还没有到来，但我们将会看到一个非常声势浩大的革命，尤其是从社会智能的角度来观察。我们将能够与机器智能进行紧密的交流和互动，无论用户是谁，其可能是另外一个机器或者电力服务系统，这在社会智能领域未来的十年中就会发生。另外一个，是与伦理学相关的，即同理心，最终机器甚至会有意识和自由意志、同情心以及伦理道德。所以，我们每一个人都应更多地关注因果科学。

For me, consciousness is having a blueprint of your software. ---Pearl, 2019 Dec^[4]

精选参考资料

特别注意，我们在文章中引用论文的方式并不严格，应该当成一个参考资料来理解。事实上，这一部分给出了笔者认为对参与这场“因果革命”至关重要，精心挑选之后的资料。希望在教会机器因果思维，构建能够通过小图灵测试的 Causal AI 之路上，这些资料可以帮助读者快速理解和把握因果科学全局。

下面是这些资料及其简介：

[1] “Foundations and new horizons for causal inference” 研讨会，2019，

https://opc.mfo.de/show_workshop?id=3271

因果推断始于经济和生物统计等学科，它刚刚开始成为人工智能的一个重要工具，数学基础依旧很零碎，该研讨会聚集了来自人工智能，生物统计学，计算机科学，经济学，流行病学，机器学习，数学和统计学的顶尖研究人员，研讨会上的报告和讨论将有助于在未来几年内塑造和改变这一领域的发展。

[2] “The Book of Why: the new science of cause and effect,” Judea Pearl and Dana MacKenzie, (Basic Books, 2018). <http://bayes.cs.ucla.edu/WHY/>

[3] J. Pearl, “The Seven Tools of Causal Inference with Reflections on Machine Learning,” Communications of ACM, 62(3): 54–60, March 2019

[4] Judea Pearl: Causal Reasoning, Counterfactuals, Bayesian Networks, and the Path to AGI | AI Podcast

<https://youtu.be/pEBIOvF45ic>

这三个材料是 Pearl 的因果理论的最新综述，较为通俗易懂介绍了因果关系的新科学，描绘了一条教会机器因果思维的强人工智能之路。

[5] Causality for Machine Learning, Bernhard Schölkopf, 2019

这是一篇刚刚挂 arxiv 就被 Pearl 亲自 twitter 点赞的论文，是马普智能所所长 Bernhard Scholkopf 最引以为傲的论文之一，他将被 Pearl 点赞这件事情写在其个人主页自我介绍的第一段中。Scholkopf 及其团队在因果结合机器学习方面做了最多的工作，此文总结和升华了提出了信息革命时代下因果结合机器学习的一般理论和深刻思考。

[6] A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks, Miguel A. Hernán, John Hsu&Brian Healy, 2019

来自哈佛教授 Miguel A. Hernan 对当前数据科学的深刻反思，澄清了数据科学任务如何分类的基本问题：prediction, description and counterfactual prediction.

[7] Causal Inference and Data-Fusion in Econometrics, P. Hünermund, E. Bareinboim. Dec, 2019.

该论文是因果革命，Pearl 的因果图模型框架如何影响某一个特定领域 -- 计量经济学的范例。

[8] Beyond Curve Fitting: Causation, Counterfactuals, and Imagination-based AI, AAAI Spring Symposium, March 25–27, 2019 <https://why19.causalai.net/>

这是 Judea Pearl 和其学生 Elias Eareinboim 组织的一次研讨会，探索了 Causality + X: X 代表了某个学科。

[9] NeurIPS' 2019, From System 1 Deep Learning to System 2 Deep Learning, Yoshua Bengio. Video with synchronized slides

<https://slideslive.com/38921750/from-system-1-deep-learning-to-system-2-deep-learning>.

Bengio 描绘了一个具备因果推理能力的 AI 系统。

[10] 加州大学洛杉矶分校 UCLA 朱松纯教授，浅谈人工智能：现状、任务、构架与统一 | 正本清源