BAAI

青源会2022年人工智能重要方向进展与未来展望报告

涉及方向：

- 自然语言处理和知识图谱
- 信息检索与知识挖掘
- 人工智能数学和基础理论
- 计算机视觉
- 智能体系结构与芯片
- AI+Science
- 人工智能伦理与治理

青源会

2022年7月

目录

前言

青源会（智源人工智能青年科学家俱乐部）成立于 2021 年 6 月，旨在增进青年学者交叉方向交流合作，孕育有引领意义的创新成果；帮助青年学者克服早期生涯压力，提供展示才华与风貌的平台；构建开放、包容的新兴研究社区，发挥青年学者间的协同效应；鼓励探索面向学科重大问题与挑战的新思想、新方法、新途径。自成立以来，青源会为海内外人工智能青年科研和技术人员建立了宽松、活跃的交流平台，促进 AI 青年科研人员开心探索智能本质。

2022 年 5 月至 6 月，青源会举办了青源学术年会及一系列学术研讨活动，邀请自然语言处理、智能信息检索与挖掘、计算机视觉、智能体系结构与芯片、机器学习数理和基础理论、AI+科学，以及人工智能伦理治理等七个方向近百位研究者共同参与讨论。研究者对领域内的热点研究、重大挑战，以及未来发展方向提出了专业意见。

本文对青源会近期一系列活动中所探讨的重点内容和观点进行了整理汇总，形成《青源会 2022 年人工智能重要方向进展与未来展望报告》。我们期待，这份内容能够为人工智能领域青年学者、学生的学习和研究，提供参考。

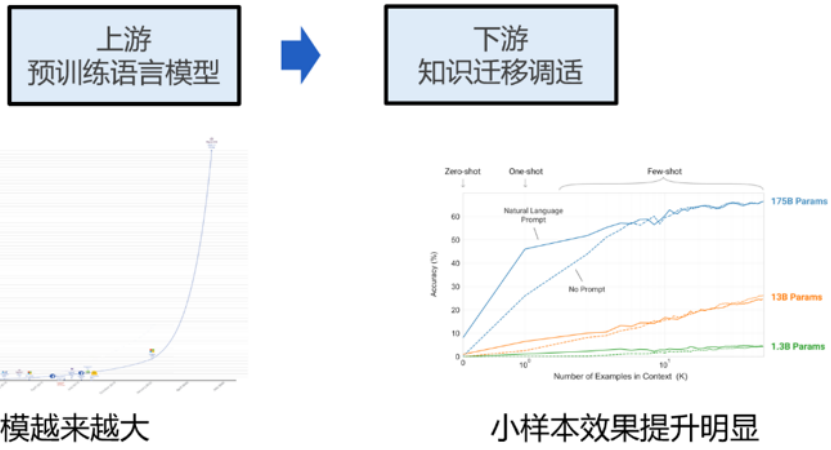
青源会

2022 年 7 月

第一章 自然语言处理和知识图谱领域进展及未来展望

青源会自然语言处理与知识图谱方向学者

目前，预训练语言模型已成为自然语言处理领域的基础模型。研究者可以通过训练大规模语料构建预训练模型，再将其迁移到一系列下游任务上，并实现性能的显著提升。同时 Transformer 已成为自然语言处理领域最主流网络架构。随着预训练语言模型的规模越来越大，在小样本学习领域的性能显著提升，研究者越来越关注这个领域的发展和变化。



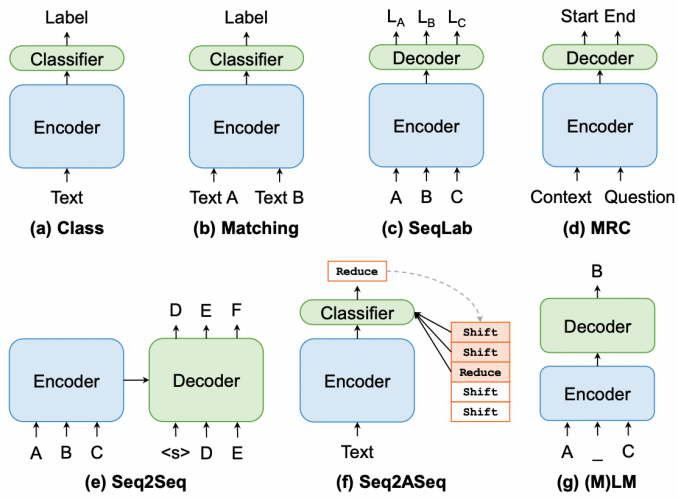
规模越来越大　　小样本效果提升明显

这些现象是否预示着自然语言处理领域和知识图谱领域更为深刻的范式转变？在大模型成为主流的情况下，有哪些新兴研究热点和机会？本章中，青源会自然语言处理和知识图谱方向的学者探讨了这一领域的范式转变情况，并对未来研究方向的发展提出了见解。

一、领域进展

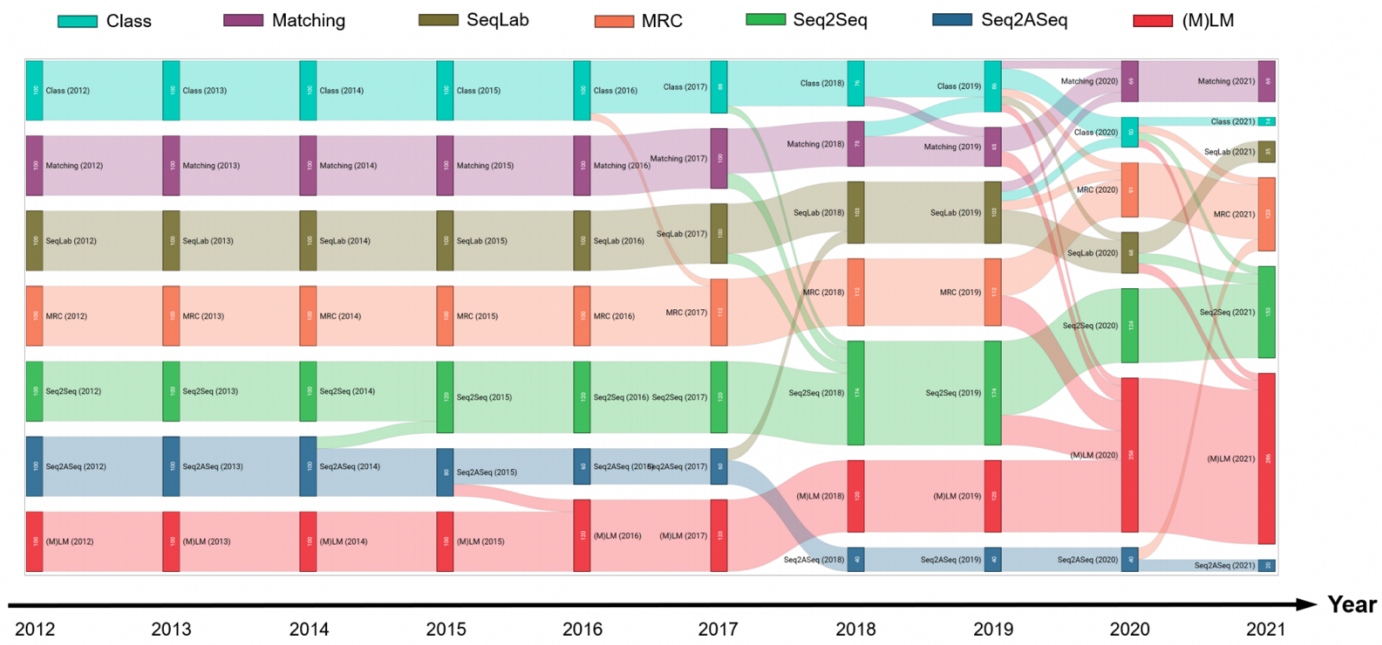
1.范式统一

自然语言处理中，有 NLU（自然语言理解）和 NLG（自然语言生成）两大类任务。但细分而言，具体可分为 7 种主流范式，包括分类、匹配、序列标注、阅读理解、序列到序列、序列到动作序列以及语言模型等。



来源： https://arxiv.org/pdf/2109.12575.pdf

不同的自然语言处理任务都有对应的解决范式。在自然语言处理发展的早期，每个任务所对应的范式是单一的。比如信息抽取任务通常使用序列标注范式来处理。近年来，可以观察到发生范式迁移（Paradigm Shift）的趋势——每种任务都在由传统范式迁移到更为统一的范式上，其中有几类范式逐渐显示出可以统一多种任务的潜力。一是 MRC 机器阅读理解范式，能够将所有自然语言处理的问题都变成机器阅读理解的形式。给定模型输入和查询的情况下，模型能够自动地提供输出；二是序列到序列方式，已经解决了包括分类、匹配在内的多项自然语言处理任务；三是目前流行的预训练语言模型范式，通过加入对应的提示（Prompt）将分类、翻译等任务改造成语言模型任务，如输入下一个词或者文本填空的训练，从而更好地完成这些任务。



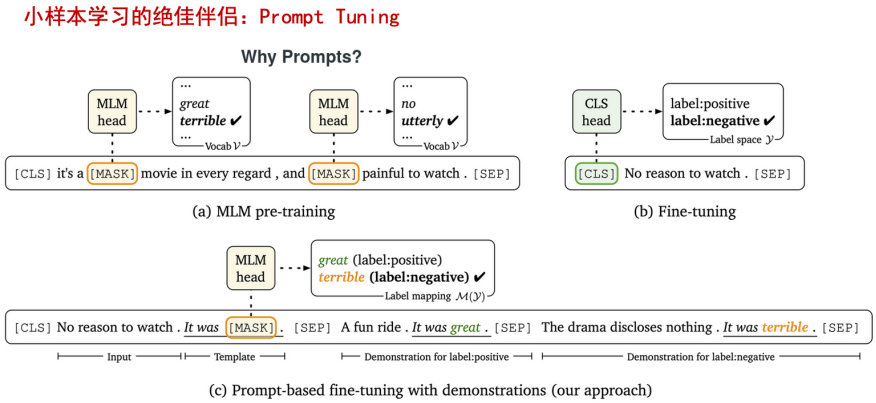
来源： https://arxiv.org/pdf/2109.12575.pdf

预训练模型将会加速自然语言的范式统一。在预训练模型性能足够强的情况下，研究者们为了使用模型能力，会主动将其需要解决的任务转变为适合预训练模型处理的形式。因此可以看到，当前 NLP 中的多种任务都在向预训练模型擅长的范式方向发生转变。

2.小样本学习成为热点

随着下游任务性能的显著提升，研究者开始关注零样本及小样本学习问题，并成为近年来非常重要的研究热点。

除了得益于预训练模型的支持，小样本学习的快速发展也归功于其绝佳的训练方式——提示调适（Prompt Tuning）。在传统的精调（Fine Tuning）阶段，研究者需要调整模型的参数，对应大模型来说效率相对不高，并需要较多的标注数据。现在，研究者可以构造提示及模板提供给大模型，充分利用大模型中学习到的知识，使其更好地完成下游任务。该种模式非常适合小样本学习，因为其非常依赖模型的先验知识。只要充分挖掘出模型中的隐藏知识，就可以直接提升小样本学习的性能。

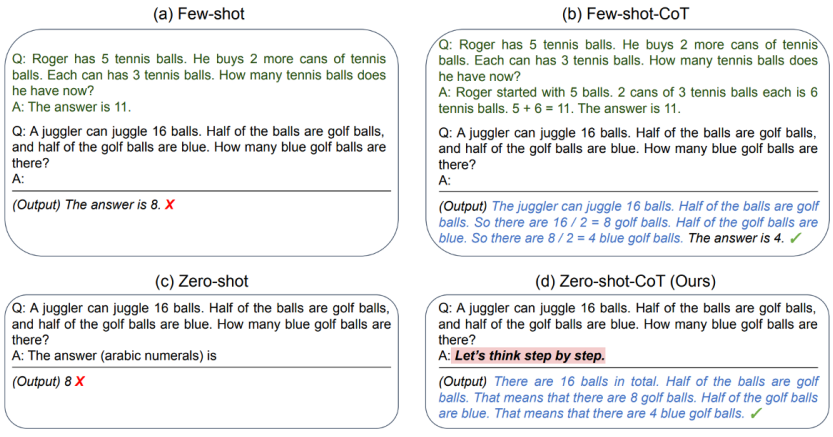


Pengfei Liu, et al. " Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing."
Tianyu Gao, Adam Fisch, Danqi Chen, **Making Pre-trained Language Models Better Few-shot Learners**, ACL 2021

3.充分挖掘大模型能力

尽管预训练大模型已成为一种主流的基础模型，其能力还有被进一步挖掘的空间。预训练模型或许可具备理解和常识推理的能力。

传统上，如果向预训练模型（如 GPT）询问理解、常识等方面的问题，模型可能会回答错误。但也有研究指出，如果将人类的推理过程作为示例展示给模型，模型会跟随这种推理模式，输出一个正确的答案。最近还有研究者提出了一种更简单的做法，只需要提供一句提示，如“Let's Think Step by Step”，模型就能够按照某种推理链条，自动找到答案。以上案例说明，预训练大模型中的潜力仍有待挖掘。

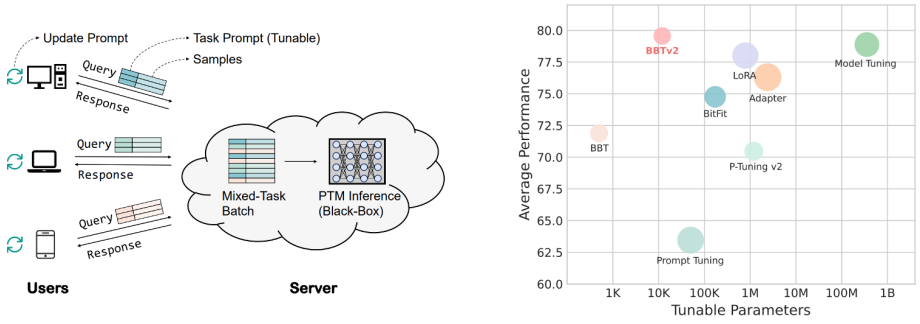


Takeshi, Kojima, et al. "**Large Language Models are Zero-Shot Reasoners**." arXiv preprint arXiv:2205.11916 (2022).

4.语言模型即服务

当前，语言模型的规模越来越大，其面向下游任务的调适成本也非常高。因此，一种更现实的利用大规模语言模型的方式是“语言模型即服务”（Language Model As A Service，LMAAS），即将大规模语言模型视为一种服务。当用户提供输入时，模型自动提供输出。

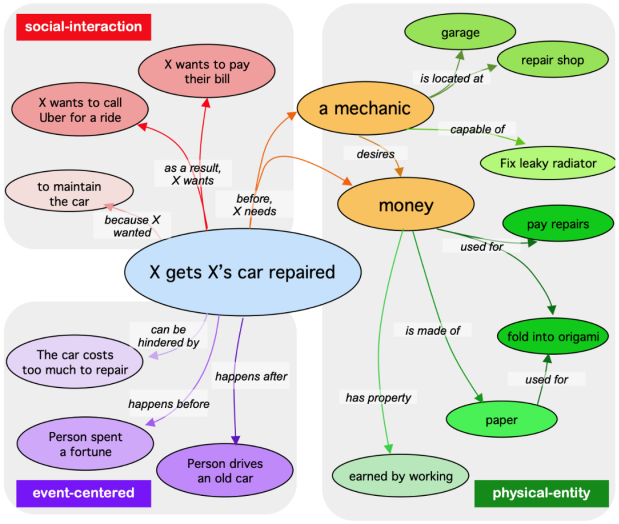
针对“语言模型即服务”的下游任务调适就变得非常重要。一种潜在的方法是黑箱调适（Black-box Tuning），调参过程仅依赖输入和输出的交互，无需计算大模型的梯度，因此调参过程也十分高效。最近的研究工作将这种方法应用于小样本学习，可以实现与梯度调参接近的水平。



Black-Box Tuning for Language-Model-as-a-Service, ICML 2022
BBTv2: **Pure Black-Box Optimization Can Be Comparable to Gradient Descent for Few-Shot Learning**, arXiv:2205.11200

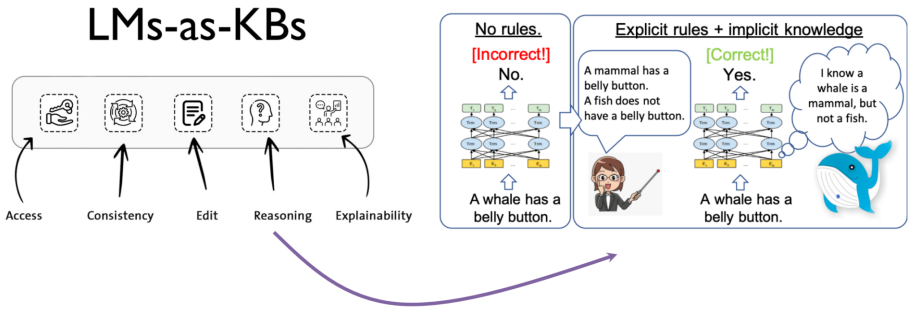
5.常识知识图谱

常识知识图谱是知识图谱面向更为复杂任务的发展方向。传统的知识图谱是静态的、仅表示概念的知识图谱。要推动知识图谱的进一步发展，需要让其能够表示更加复杂的事件，如动态事实。目前，常识知识图谱的发展态势较好，已经能够根据事件的描述，构造出更为复杂的事件关系，如时间关系、因果关系、条件关系、组成关系等，这样的知识图谱在常识性下游任务中具有较好的性能。



(Comet-) Atomic 2020: On Symbolic and Neural Commonsense Knowledge Graphs, AAAI 2021
Sap M, Le Bras R, Allaway E, et al. Atomic: An atlas of machine commonsense for if-then reasoning, AAAI 2019

此外，知识图谱的另一个发展方向是“LMs-as-KBs”（语言模型即知识库），可以通过类似于三元组查询的形式，由人类构造语言模型模板，让语言模型通过知识库查询的方式，从预训练的数据中返回答案，相比传统的知识检索具有优势。



A Review on Language Models as Knowledge Bases, arXiv:2204.06031
Leap-Of-Thought: Teaching Pre-Trained Models to Systematically Reason Over Implicit Knowledge, NIPS 2020

值得一提的是，如果以知识库的观点来看语言模型，可能面临五个方面的问题。一是**知识读取**，包括如何构造查询来获取模型中知识的方法；二是**知识一致性**，预训练模型中可能包含相互矛盾的知识，需要保持一致；三是**知识编辑**，当语言模型训练完成后，知识以模型参数的形式固定下来。如果知识本身发生了变化，需要更新模型，则需要重新训练；四是**知识推理**，语言模型本身推理能力较弱，较难完成与知识库类似的推理；五是**可解释性**，知识库本身是结构化的，具有可解释性，但语言模型的可解释性仍需加强。

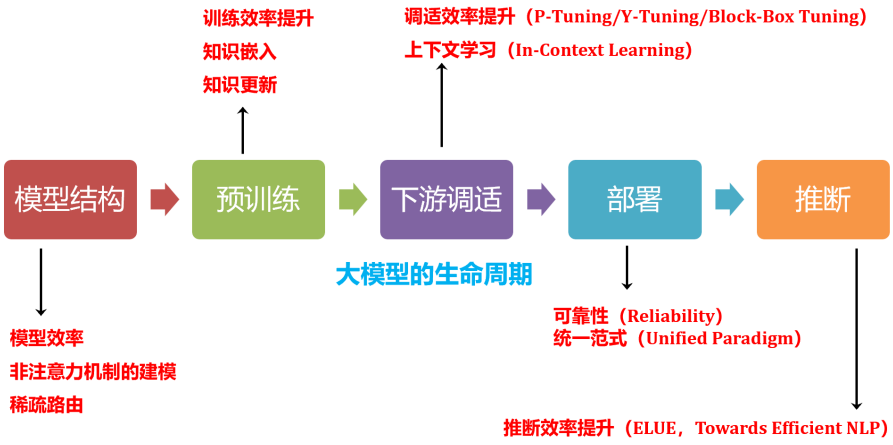
领域进展的总结如下：

二、当前研究中存在的问题

在受益于预训练大模型的同时，自然语言处理领域也面临一定的问题。一是当前主流工作同质化严重，采用的预训练模型架构也多为 BERT、T5 等。二是重回特征工程，当基础模型趋同后，研究者们比拼的更多是特征工程能力，而不是对模型原理的思考和方法上的改进。三是预训练模型研究能耗高、不环保，需要进一步探索更加高效的训练以及部署策略。

三、未来展望

面向未来，我们可以看到预训练大模型给我们带来了更多的研究方向和机会。从大模型的生命周期来看，模型研究需要经过架构设计、预训练、下游调适、部署、推断等几个步骤。在这些环节中，都有可以进一步发展的空间。



在**模型结构**方面，可以关注模型效率、与非注意力机制的交互方式，采用稀疏路由结构等新兴领域；在**预训练**方面，可关注如何设计预训练任务，提升预训练效率，寻找知识嵌入和更新模型的方法；在**下游调试**时，可进一步探究下游任务微调的效率问题，探索更加灵活的上下

文学习机制等；在**模型部署**阶段，可以研究模型的可靠性以及统一的任务范式，追求用单一模型支撑多种自然语言处理服务；在**模型推断**时，可进一步采用模型压缩、剪枝和动态路由等方法提高计算效率，为模型进行加速等。

同时，训练预训练模型也离不开语料的质量。预训练语言模型不是“语言的模型”，掌握语言能力本身并不需要如此庞大规模的训练预料。预训练模型实际上是一种“以语言承载的世界知识”的另一种表现形式。如果只学习语言，本质上不需要使用维基百科等知识密集的语料，但要训练一个高质量的预训练模型，维基百科是必不可少的训练数据，因为其中包含大量的知识。可以把预训练模型当成一种隐式“知识库”，以及参数化的“百科全书”。大模型不会去创新知识，它通过训练的方式，作为知识的一种呈现。因此，构造高质量的数据也是重要的研究方向。

除上述几个方向之外，自然语言处理领域值得关注的发展方向还包括：高质量中文数据资源建设、多语言多模态预训练模型、非自回归生成技术、开放环境中的自然语言学习、复杂知识推理、流式语音翻译、知识驱动的预训练模型、零/小样本学习、具身学习、神经-符号融合架构、对话与文本生成质量的评价问题，以及可信 NLP 技术等。

第二章 信息检索与知识挖掘领域进展及未来展望

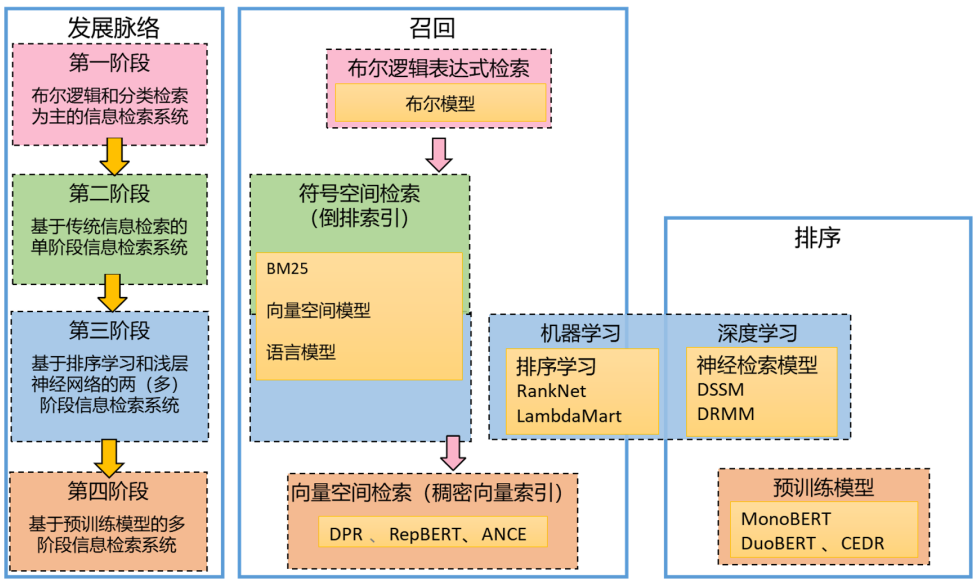
青源会信息检索与知识挖掘方向学者

信息检索与挖掘涵盖检索、推荐、知识挖掘、图神经网络、数据管理等多个领域，与经济社会紧密相关。随着近年来预训练模型和知识图谱的快速发展，以及针对计算加速的专用芯片诞生，信息检索与挖掘领域迎来了更为广阔的发展前景。本章中，青源会信息检索与挖掘方向研究者探讨了这一领域下的前沿进展，并对领域的研究热点和未来进行展望。

一、信息检索

在信息检索中，最重要的问题之一是研究 Query（查询）和 Document（文件），或者也可以将 Q&A 任务视为是问题-答案的匹配。从信息检索研究者视角来看，人们更关注的 Query 和 Document 的匹配程度或相关程度的问题。

根据信息检索任务的特点，可将其分成召回、排序两个过程。召回过程指的是从大规模的 Web Data 中搜索到与 Query 非常相关的 Document，然后再对这些 Document 进行精准排序。



（一）领域进展

自信息检索学科诞生伊始，已出现了多种研究方法，最初的研究基于布尔逻辑方法，之后有研究者提出了传统的信息检索模型，其中也包括传统的语言模型等。随后，检索模型的规模不断增大。随着机器学习的方法在信息检索领域的发展，逐渐成为检索领域中的一种主流方法。目前，因其具有通用任务能力，大规模预训练模型也在信息检索任务中发挥了重要作用。

1.基于预训练模型的查询和文件表示

自 2019 年开始，信息检索领域的研究重点是基于预训练模型（采用单塔或者多塔架构），对信息检索中需要匹配的 Query 和 Document 分别进行表示。此外，也有模型可以对二者之间的交互程度进行表示和建模的方法。

近来，具有代表性的工作是谷歌于 2021 年发布的论文“Transformer Memory as a Differentiable Search Index”，该研究在信息检索领域开启了新的里程碑。研究者将召回和排序两个过程融合为一个阶段，并提升了在召回时采用大规模预训练模型的计算效率。用户在使用过程中，可以很自然地输入查询，模型能够直接生成或提取用户需要的结果。

2.多模态信息提取研究

另一个检索领域的重要进展是对于不同模态信息的提取研究。过去，研究者普遍关注文本的检索和提取工作。随着近年来多种模态数据量的增长，对于图文相互检索的研究，甚至是跨模态数据的检索，已经成为了新的热点研究话题。典型案例有 CLIP、DALL·E、DALL·E2 等。

- CLIP: Connecting Text and Image
- DALL•E+DALL•E 2: Creating Images from Text

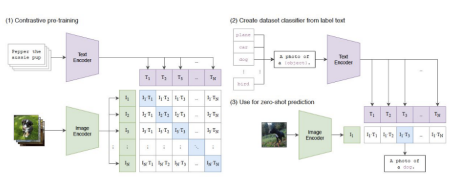
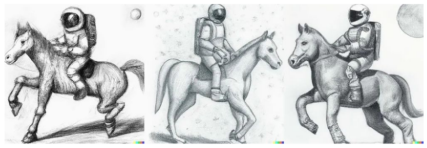


Figure 1. Summary of our approach. While standard image models jointly train an image feature extractor and a linear classifier to predict some label, CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples. At test time the learned text encoder synthesizes a zero-shot linear classifier by embedding the names or descriptions of the target dataset's classes.

An astronaut + riding a horse + in a photorealistic style

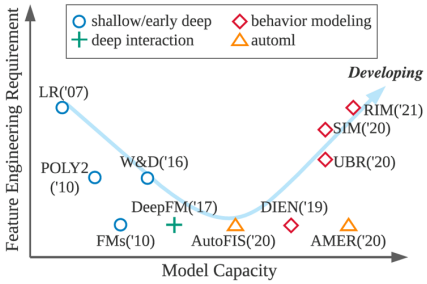


An astronaut + riding a horse + as a pencil drawing

3.基于样本交互的 CTR 预估模型

CTR 预估是搜索和推荐中的重要问题，开始研究者主要使用逻辑回归之类的模型，自 2015 年开始，深度学习减少了特征工程的工作量，通过挖掘特征之间的交互模式进行高阶特征的自动设计，2020 年开始为了进一步提升模型的能力，很多数据处理和特征工程方面的工作又回来了，近年来提出 UBR，DIEN 等工作更多的挖掘数据样例之间的关系，这也是一个新趋势的到来。

- Past developments
  - 2000 – 2015: from linear models to bi-linear models
  - 2015 – 2019: deep models on feature interaction mining



- Current trend
  - 2020 – now: sample interaction paradigm

4.可信的信息抽取模型

可信的信息抽取模型也成为研究者关注的话题。研究者希望，在为用户提供提供检索结果同时，还能够确保模型公平性。在这一领域主要分为两个方向，分别是外在信任和内在信任。外在信任主要是对关键维度的质量进行保证，即对结果进行评估；内在的信任主要是希望推理的过程能够符合人类的期望，为结果提供更好的解释。

（二）未来展望

在未来，信息检索领域可能会看到三个方面的趋势。一是真正能打通召回和排序两个阶段，构建端到端的、高效信息抽取框架。二是引入知识，解决理解文本时遇到的常识推理等问题。三是持续关注可信人工智能研究，保证检索系统的稳定性，并探究解决人工智能的伦理问题。

二、推荐系统

推荐系统是一个工程科学问题，其科学目标可以定义为通过理解用户的隐式需求，对信息进行过滤。构建推荐系统，要解决的核心问题是：怎样能够从用户历史行为中去理解用户的隐式需求。

（一）领域进展

在推荐系统领域，主要的进展包括：一是信息去噪。研究者希望即使是缺乏监控信号等问题的情况下，推荐系统能够识别并删除用户和商品信息中存在的噪音。二是数据去偏。尽管用户在使用过程中留下了许多用户痕迹，但这些数据包含了用户与系统的各种偏见，会影响推荐系统的性能，需要研究去除这些偏见的方法。三是细粒度建模。推荐系统需要在用户行为建模的过程中，更多地去考虑更为细粒度的场景信息建模，提供更精准的推荐结果。

（二）未来展望

在当前的推荐系统的研究中，仍存在一些问题，包括商业化程度较重、可能存在信息茧房等，需要开展交叉学科相关研究。同时，推荐系统还需要让渡给用户主导权，提供给用户更具有自主可控的机制，让推荐结果满足可信的需求，提供更多结果的可解释性。此外，推荐系统的学术性研究和工业落地之间还存在鸿沟。学术界关注的偏向方法层面，如图方法、强化学习方法等。但工业界关注的偏向场景层面，如对 Cross Domain、Large Scale、Long Tail 等问题的处理等。这导致从业者在关注不同的问题和技术解决方案上存在一定差异。

三、图神经网络

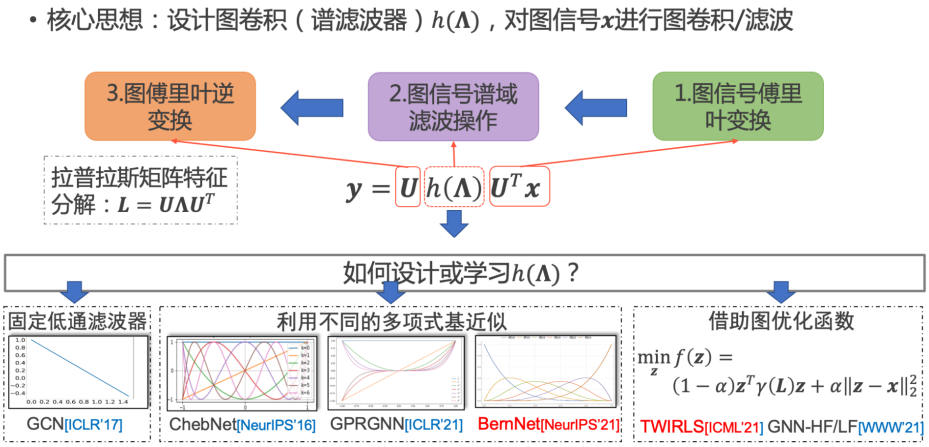
图神经网络存在于很多应用场景中。比如，在社交网络中，人与人都是图网络中的节点，用户和用户的连线就构成了图中的边。图神经网络在数据结构上，和图像、文本等不同，图结

构数据是一种新的数据类型，可专用于图机器学习。使用图神经网络，已经可以完成一些预测和推荐任务。

（一）领域进展

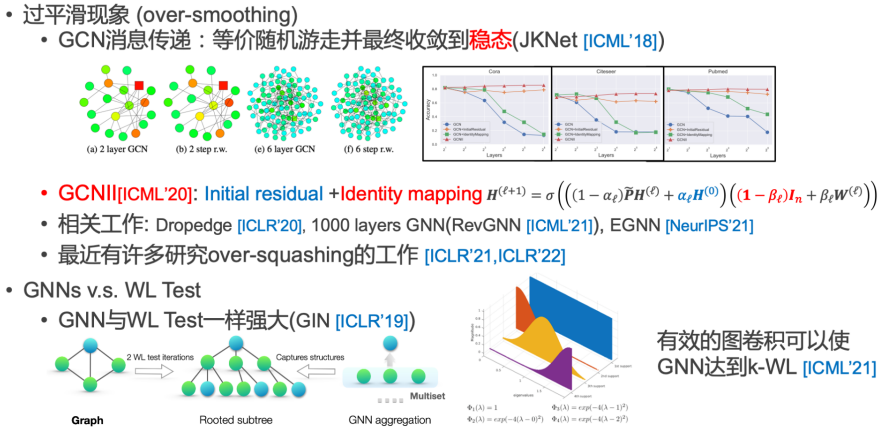
1.谱域问题

谱域问题的核心思想是设计在图上的卷积，通过一个谱域上的滤波器进行实现，核心问题是设计或学习谱域滤波。近年来已发展出一些不同的方法，包括固定低通滤波器方法、利用不同的多项式基近似方法，以及借助图优化函数进行设计和学习不同的模式。



2.空域问题

由于消息的传递产生了不同的机制，可以总结在这一范式下，对不同的机制进行定义。其中的核心问题在于改进这种消息传递的机制的方式。同时，不同的消息传递方法在图神经网络的表达能力上到底会产生什么样的结果，也是领域中重要的研究问题。



3.图神经网络的高效计算和可扩展性

在这一方面，有分层采样方法、图采样方法、线性模型方法、历史嵌入方法等，能够去解决高效计算问题。

4.大规模的图预训练研究

将预训练和图网络进行整合，能够在图神经网络的各种场景下得到有效应用。这其中包括了图监督训练方法、图自编码器方法、自回归方法、上下文预测、图对比学习等。

5.图数据的对抗攻击和防御问题

在攻击方面，根据攻击者所知被攻击模型的信息，可以分为白盒方法、灰盒方法、黑盒攻击模式等。在防御方面，目前已经出现了对抗训练方法、预处理方法、注意力机制方法、可验证鲁棒性方法等。

6.图结构的组合优化

通过多种方法改进模型结构，进一步提升模型的性能和表现，如基于强化学习的方法等。同时，可基于启发性方法，通过在神经网络的结构当中引入一些经典的方法，来进行近似模拟。还有混合方法——把两种模式结合在一起进行组合的优化。

（二）未来展望

图神经网络领域，目前还存在很多问题有待解决。一是在机器学习的理论核心假设方面还存在探索的空间。如在不同的数据条件下研究机器学习的理论核心假设。二是研究图层面和节点层面的任务之间的区别。研究者在定义不同的任务时，可能是依据不同的层面来进行研究。例如，在 Drug Discovery 问题中，较为重要的是研究图本身的表示，但是在节点分类或推荐系统中，关注的重点则是图中每个节点的表示。因此，在不同的场景和不同问题研究中，图神经网络的研究重点也应当区别对待。三是在大图场景下，对图神经网络进行高效计算，目前仍然还是一个未解决的问题。

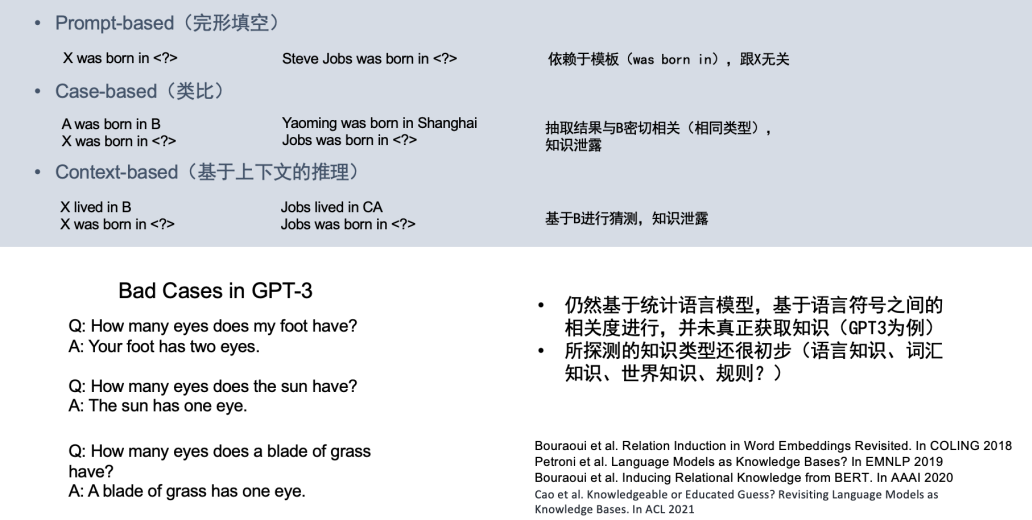
四、知识挖掘

（一）领域进展

1.知识学习和表示学习

在知识学习和表示学习层面，主要的进展有：知识图谱嵌入的新方法。研究发现，将引入了层次关系结构的知识图谱嵌入模型，能够有效刻画不同的粒度、不同层面的实体。此外，还有研究者提出，将基于隐含语义单元的知识图谱嵌入模型，通过隐式的方式，解决知识图谱当中的数据不平衡和数据稀疏的问题。还有通过输入共享语义单元的模式，实现信息互通和共享。此外，有研究者提出了基于多任务强化学习的鲁棒性知识图谱嵌入模型，可以有效、鲁棒地解决自动化抽取过程中带来的噪声问题。

在知识学习和表示学习层面，还出现了基于预训练模型的知识探测方法。研究者将大规模预训练模型视为规模较大的知识库，从而实现应用。主要的方法有基于 Prompt-based 的方法、Case-based 类比方法，基于上下文推理的方法等。



2.知识检索与问答

在知识图谱应用上，有研究者提出有基于语义的解析方法、基于信息检索的方法、多策略问答方法（如如利用知识进行增强，实现较精准的阅读理解，或通过检索模式挖掘）等。

对话式知识问答问题，是交互式研究中的热点。这一方向的挑战是模型会遇到省略的情况。此外，模型还面临有效学习语义解析器的问题。最后，Transformer 模型和逻辑语法模型也在这一领域得到了有效地应用。

3.多模态问答

多模态检索成为近期关注的问题，当前的方法包括单流方法、多流方法，以及不同模态之间对齐的技术，已经在知识挖掘领域开展应用。

图像问答可以分为三个层次：

1. 初级: 从图像识别的结果中直接得到答案
2. 中等：答案需要简单事实的支持
3. 高级：对于复杂的问题，答案不在图像中，可能涉及常识、具体话题和百科知识来进行推理

多模态问答的挑战：

- 多模态数据具有异构性
- 多模态数据的关联表示难度较大
- 多模态知识融合困难
- 多模态问答大多只能处理简单的问题
- 多模态知识问答推理能力弱
- 多模态知识问答可解释性差

（二）面临的挑战

知识挖掘方面面临的挑战主要包括：

1.知识资源的覆盖度问题

尽管研究者已经构建了包含不同的信息的图，但是在特定领域，依然面临领域知识缺乏的问题，在解决实际问题方面，知识的覆盖程度远远不够。

2.知识表示问题

当前，大多数知识资源是基于三元组的知识表示模型，但对于复杂结构的知识，三元组可能无法进行有效的表示。复杂架构的知识包括但不限于：事件、流程、计划、常识，以及过程类知识、非事实类知识等，这些都具有挑战性。

（三）未来展望

未来，在知识挖掘领域，可以看到如下趋势：一是神经网络与符号系统的融合，利用大规模统计模型，采用数据驱动的方法学习的知识，和通过人为定义、非常精确的知识结合，进行表示和推断，目前仍然缺乏良好的模式。二是面向预训练模型的知识探索，目前仍缺乏成熟的模式，存在知识覆盖度较低，表示方法不够准确等问题。基于多模态、多任务的预训练大模型热潮还将持续，无论是不同模态之间的转换，还是近期提出的通用模型，还将在知识挖掘领域持续发挥重要的作用。

五、图数据管理

图数据有很多的应用模式，因此需要去建立针对图数据的数据库和计算系统，对这些数据进行统一的管理和计算。

- 图数据库：输入查询语句，输出结果

- 图计算系统：编写程序调用编程接口，运行得到结果

（一）领域进展

图数据管理，特别是图数据库的建立和管理，以及计算底层逻辑方式上，近年来出现新的研究成果。在图匹配方面，出现了基于现有大数据平台、基于新型计算平台等的新方法。图计算系统方面，出现了以子图为核心，采用子图匹配、子图挖掘等图为核心的运算策略。此外，有以机器学习为核心，以图上机器学习任务为优化目标的研究工作。最后，还有基于 CPU、超算等新型计算平台上产生的新型图计算系统。

（二）面临的挑战

数据管理领域比较有挑战性的问题包括：

1.大规模的图查询：出现了采用路径关注、路径查询、子图查询、可视化查询等不同的方法。

2.大规模图挖掘：包括子图计数、社区挖掘、密集子图挖掘等问题。一些研究者提出了在图结构上运算模式、社区搜索算法，以及基于图的拓扑结构产生的图挖掘方法等，产生了较好的结果。

3.图数据流：由于序列的无限性，导致图数据流的研究往往会涉及到时间窗口的引入，为这一方向带来挑战。在这个方面的进展主要是关注持续性查询模式，在给定查询的前提下，随着图数据流实时更新，通过多种方法，实现实时维护查询对应结果。

4.面向图数据的硬件加速：将计算量较大的工作分配给专门的硬件进行处理。这里存在的挑战包括：图数据的不规则性、额外编程门槛成本以及顾此失彼带来的问题。

为了解决图数据硬件加速方面的挑战，近日在硬件领域，出现了基于 GPU 和 FPGA 等硬件的加速方案。一是采用 GPU，打造面向类型高度统一的、相互无依赖的大规模数据、不易被打断的计算环境，适合高并发计算需求，实现计算加速，降低开发成本，便于编译优化。二是基于 FPGA，根据设计人员的需求定制电路结构，避免低效访存带来的内存带宽的闲置，定制性地解决图的不规则性，避免大量内存所带来的原子操作。

（三）未来展望

未来，数据管理领域的发展方向包括：

1.构建分布式图数据库和图计算系统，但目前缺乏对查询语言完全的支持。

2.面向新型应用场景下的图数据库与图计算系统，如区块链和边缘计算等新场景下计算模式。

3.基于语义的图查询和挖掘方法。

4.图计算加速的核心方法，建立图数据的规整表示，并建立统一、专用、覆盖广的图算法库。

注：本章节观点整理自青源会 2022 年 5 月 10 日及 6 月 1 日组织的研讨会。研讨会召集人为：清华大学兰艳艳。参加研讨的嘉宾有：上海交通大学张伟楠、山东大学任昭春、中国科学技术大学何向南、武汉大学李晨亮、中国人民大学赵鑫、中国人民大学魏哲巍、中国科学院计算技术研究所沈华伟、浙江大学杨洋、同济大学王昊奋、中国科学院自动化研究所刘康、北京航空航天大学庄福振、复旦大学郑卫国、北京大学邹磊。

第三章 人工智能数学和理论基础领域进展及未来展望

青源会数学和理论基础方向学者

当前，针对人工神经网络的数学和理论研究方兴未艾，研究者从模型的表示能力、泛化能力、优化求解等方面开展研究。通过抽象的视角研究神经网络，可以帮助人们理解其本身的机制机理，有助于推动神经网络在多环境、多任务等方面的应用，并提升模型的效率和泛化能力，发现其能力边界。本章中，青源会数学和理论基础方向研究者探讨了近期的研究进展，并对领域内面临的挑战和未来发展进行了展望。

一、领域进展

1.深度学习的表示理论

在多年前，神经网络的表示理论已经取得了一些成功。例如，在 30 多年前，已经有研究说明双层神经网络具有通用近似（Universal Approximation）的特性。近几年，由于神经网络在训练推理方面的性能超过了传统模型，研究者开始研究多层神经网络的表示能力。

在表示理论方面，目前有三个方向值得关注。一是多层神经网络/深度神经网络在表示层面的效果研究。既然两层神经网络已经可以做到万能表示，多层或深度网络表达的效果相比双层神经网络是否会更好？在这一领域已经有一些结论，例如多层神经网络使用的神经元更少，能够更精简地表达一些函数。在这种情况下，如果考虑非参估计的设置，神经网络表示高阶光滑函数时，不存在维度灾难，且相比很多模型在数量级上有一定优势。

二是不同的神经网络结构具有的优势。目前研究有考虑残差神经网络。研究者主要关注残差网络相比普通的全连接前向网络，更适合表示的函数类型，以及残差网络在学习哪些问题上更具有优势。

三是研究利用神经网络的表达能力获取好的泛化能力，这需要偏差和方差平衡。传统的统计学习理论认为，一个模型想要学好数据，需要平衡偏差和方差。偏差是指模型不一定能够完全地表示数据分布的情况，其表示与实际可能有偏离。如果要提升模型的泛化能力，则偏差要变得更小，这需要在表示上具有更多优势。

2.深度学习的泛化理论

传统机器学习认为，泛化主要是在偏差和方差之前取得平衡。但是在深度学习领域，理论和实践还存在一些差距。对深度学习来说，样本量往往远小于训练参数量。在这种情况下，可能不需要对模型加入一些正则。如果不对模型加入正则，训练误差较低，因此导致模型过拟合。从传统的统计理论解释，过拟合的模型泛化性能不高。但是从实践来看，即使是训练误差较低的神经网络模型，其测试性能也较好，出现了名为良性过拟合（Benign Overfitting）的情况。这种现象值得探究。

此外，隐式正则领域也是泛化领域的一个热点。这里举两个例子，一是在频率空间，一些研究发现神经网络在训练中具有偏好低频的频率原则，一些工作为克服高频的学习和泛化困难提出了新的网络结构，比如多尺度网络和 NeRF。二是深度学习往往会使用低阶带有随机性的算法求解，如随机梯度下降法或者 Dropout 等，随机算法会使得模型求得的参数具有特殊性，相当于对这个模型加入了一个正则。隐式正则方向是研究过参数化网络能找到泛化好的解的一个重要方向。

3.深度学习的优化理论

神经网络是非凸模型，前一段时间的研究热点是鞍点逃离问题。有些神经网络满足严格鞍点的条件。在一般的非凸优化中，梯度下降不能保证模型达到全局最优。而严格鞍点条件能保证随机梯度下降法达到最优。另一个研究方向是研究极值点问题，探究多种极值点对应的性质，比如研究不同宽度、深度的网络之间的极值点之间的嵌入层次关系。此外，近年来研究较多的

是过参数化的神经网络。在实践中，神经网络的参数量往往会非常大，但过参数化可以使学习问题的变得简单。

近年来，在优化领域出现了对神经网络的相图分析，发现不同初始化下，神经网络会有性质不同的动力学行为，比如一类名为神经正切核的理论证明，在一定的初始值条件下，神经网络会约等于关于初始值的一阶展开、一阶逼近。而当神经网络在无穷宽的时候，网络和 Kernel 回归模型是等价的。 同时，还有一种名为神经网络的平均场理论，给出了神经网络更一般性的描述。神经平均场理论试图通过权重分布等视角来刻画神经网络运转的状态，其优势在于，当神经网络趋于无穷时，用一个分布来看待整个体系对每个神经元粒子的平均作用。由于研究多个神经元间的互相作用非常复杂，采用这种整体体系，能够简化神经元之间的相互作用。更一般地，对于两层无穷宽的网络，其相图可以分为线性区域（包括神经正切核区域）、临界区域（包括平均场区域）和非线性区域（参数有明显的凝聚现象的非线性区域）。

4.深度学习抽象模型

理解神经网络的原理，一方面需要从更高的角度进行分析。例如，使用可解析的模型来理解神经网络的优化和学习过程，包括前文提到的神经正切核理论和平均场理论。另一方面，一些研究者认为，完全通过数学层面的解析方法，并不一定能够解明神经网络的机制，可以采用微观观察的方式，理解神经网络的运转机制。例如，有一种名为局部弹性的方法，认为神经网络与线性模型的不同之处在于，神经网络面对决策面的分类是在缓慢的、局部的弹性变化，而不像线性模型一样是直线，会有较大的变化幅度。还有一种方法是层玻璃模型通过对神经网络逐层简化，进行分析，研究神经网络应该呈现的状态。

5.神经网络在其他应用方面的理论

在面向应用的研究中，一个重点方向是关于优化神经网络在多环境条件下的训练情况。该领域认为，神经网络在多环境中，如果环境会不断变化，理论上可以发现直接的回归模型不可能学到最优的回归法。有研究者探究使网络能够学到的根本特征的方法。

另一个方向是研究强化学习中的神经网络，意图将神经网络相关的理论应用在强化学习方面。例如，假设模型能够完成好一个回归问题，可以依据其理论，帮助强化学习模型学到更好的策略。

还有一个方向是在 AI for Science 中，发展神经网络相关的算法和理论用来解微分方程，主要是克服高维和刚性等问题，提升仿真模拟的效率。神经网络算法与科学问题结合也产生了新的理论问题，比如算法的适定性和收敛性等问题。

二、面临的挑战

1.神经网络算法研究和实际应用的模型存在鸿沟

目前在神经网络算法和模型之间还存在鸿沟，即研究者设计的算法并不跟模型完全挂钩，而是普遍使用简单的梯度法。同时，当前建立的很多算法理论，其假设可能与实际应用存在差距。例如，研究神经网络往往需要其达到无穷宽的状态。而样本的训练步长非常小，几乎趋于0，但在实践中，人们不会这样设置网络超参数。

2.传统模型的研究结果无法应用在神经网络上

同时，传统模型上得到的结果，并不一定能够用在神经网络上。当前的核心的问题是，如何能够将把神经网络中各种配置和性能指标之间的 Trade-off 分析清楚。正如前文提到，即使神经网络在理论上可以推导出良好的表达能力或泛化能力，但是在实际中还不能解决优化问题。将表达能力和泛化能力之间的关系分析清楚，保证优化可以解决，依然是一个挑战。此外，还有一些科学问题有待解决，如神经网络的收敛条件、收敛点的质量、隐式正则可以到达的点，以及收敛的速度等。

3.数理理论在其他方向的应用较少

最后，要将神经网络的数理理论普及推广到其他方向，目前的成果相对较少。在多环境应用中，以及数据并非完全独立或同分布的假设条件下，都有待研究相应理论。其他需要关心的问题还包括过参数化模型的必要性研究等。

三、未来展望

本章最后总结深度学习数理基础领域的五个重点研究问题，未来这一领域的研究可能会朝这些方向努力。

1.建立深度学习数理基础需要的数学工具：目前主要数学工具是依照动力系统、随机过程、计算复杂度、统计学习等发展的理论。是否需要引入代数或者几何的工具与思想，建立更好的理论基础。

2.数理基础研究的目标与方向：应树立数理基础研究在实践中达到的目标，了解实践中理论能够帮助解决的问题。

3.短期内，数理基础的发展方向：在目前神经网络及其理论发展不成熟的阶段，应明确其短期的发展方向，如类似过去的实验物理研究——先对物理现象给出一些解释，在寻找更为核心的理论支撑。

4.面向大规模预训练模型的统一描述：大规模预训练模型参数庞大，结构复杂，对其进行理论层面的描述是一个挑战。

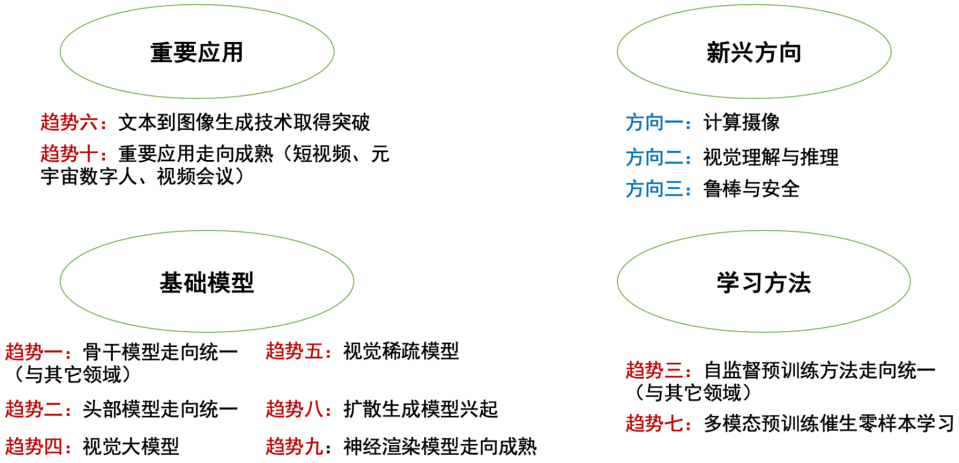
5.吸引纯数学家投入深度学习或机器学习的理论研究：吸纳纯数学领域的前沿理论和方法，有助于推动机器学习数理基础的研究工作。

注：本章节观点整理自青源会 2022 年 5 月 16 日及 6 月 1 日组织的研讨会。研讨会召集人为：美国宾夕法尼亚大学苏炜杰。参与研讨的嘉宾有：北京大学方聪、上海交通大学许志钦、上海交通大学周栋焯、北京大学邵嗣烘、中国科学院数学与系统科学研究院于海军、西安交通大学林绍波。

第四章 计算机视觉领域进展及未来展望

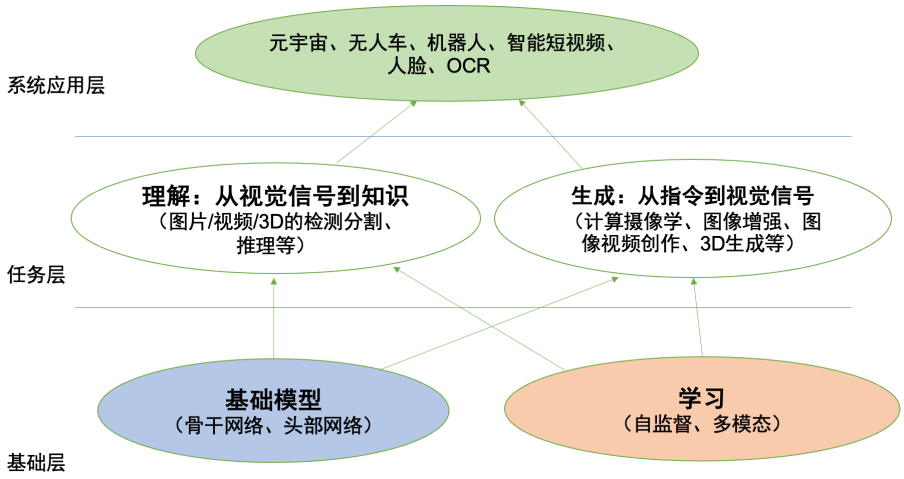
青源会计算机视觉方向学者

计算机视觉是一门处理和分析视觉信号的学科。随着视觉信号采集的设备越来越普及，以及视觉技术走向成熟，视觉应用已无处不在。视觉通常分为两大类问题，一类是“理解问题”——对采集的视觉信号进行分析和理解，视觉信号包括图像、视频或 3D 点云等。另一类是“生成任务”，给定输入（如文字描述），可以让模型生成图像、视频等。本章中，青源会计算机视觉领域的研究者针对计算机视觉的进展和未来发展进行探讨。



一、领域进展

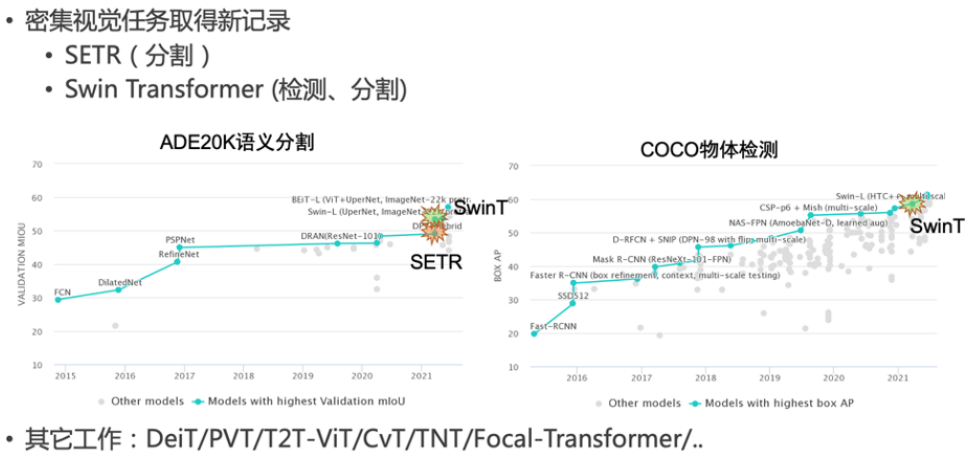
2021 年是计算机视觉迎来发展变革的一年，出现了四个主旋律，分别是：更统一的建模和学习模式、更大更稀疏的视觉模型、视觉领域解锁新技能和新模型，以及重要的模型或应用走向成熟。2021 年的视觉发展变革，几乎涉及到视觉领域所有的研究问题。下文将详细介绍四个主旋律下的重要进展。



（一）更统一的建模和学习模式

1.视觉 Transformer 成为主流骨干模型

Transformer 原本是面向 NLP 任务的架构。最近一年多以来，Transformer 逐渐成为视觉领域主流的骨干模型。它在视觉中应用的先驱工作是 2020 年谷歌提出 ViT 模型，在图像分类问题上取得令人印象的结果。到 2021 年初，基于 Transformer 架构的模型在密集视觉任务取得新的记录，如 SETR 和 Swin Transformer 等，它们分别在 ADE20K 语义分割和 COCO 物体检测上首次超越 CNN 架构。过去一年来，基于 Transformer 架构的模型不断涌现，包括 DeiT、PVT、T2T、CvT、TNT、Focal-Transformer 等。



同时，Transformer 也很快推广到其他视觉信号和其它视觉问题上，如视频方面的 ViViT；点云方面的 Point Transformer 等；底层视觉方面的 Image Processing Transformer 等。

在生成领域，Transformer 也开始取得新进展，如 2021 年初的 TransGAN 最早应用 Transformer 解决生成问题，此后年底的 StyleSwin 在性能上也首次超越 CNN。

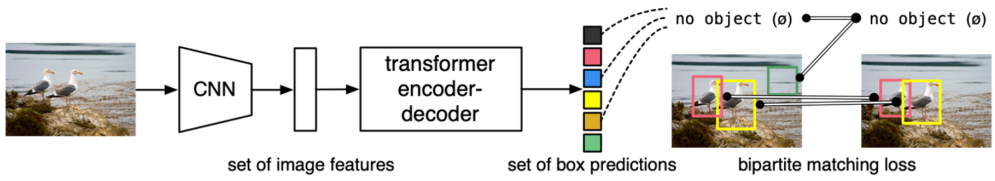
除了在视觉任务上广泛提升性能，在生态层面，企业开始建立围绕 Transformer 为核心的软硬件生态。例如，2022 年 Nvidia 推出全新的 GPU H100，其中专门为 Transformer 定制专用计算的引擎，相比上一代芯片，性能提高 6 倍。模型轻量化方面，一些研究者将 Transformer 应用在轻量模型的研发上，如 MobileFormer 等。

Transformer 还启发了新的算法架构设计和研究。例如，MLP-Mixer 将视觉骨干网络变为全 MLP 架构，也能够取得较好的效果。类似的研究包括 RepMLP、CylelMLP、PoolFormer 等。受 Transformer 的启发，卷积神经网络架构也迎来“第二次增长”，典型例子包括 D-DW Conv、ConvNeXt、RepLKNet 等。

2.下游感知任务建模走向统一

2020 年 DETR 的提出，启发了研究者探索统一下游感知任务的建模方法。DETR 将物体检测问题刻画为翻译问题，利用解码器将图像“翻译”成物体。尽管在过去的研究中，已有研究者将注意力机制用于建模物体和图像像素之间的关系，但将 Transformer 解码器整体应用于建模，尚属首次，这为研究者们提供了对感知任务建模的全新理解。

- 缘起：DETR（ECCV 2020）
  - 物体检测：利用Transformer解码器**将图像翻译成物体**



- 重要意义：提供了一种统一各种下游感知任务的建模解法

过去一年中，DETR 深化发展，并逐渐拓展到了其他的视觉领域，如 Deformable DETR，大幅提升了 DETR 性能、速度等指标。同时，有研究者将 DETR 推广到语义分割问题上，如

Mask Former 和 Mask2Former 等。也有应用到 3D 检测领域案例，如 DETR3D、group-free-3D。同时，DETR 的变体模型不断涌现，如全编码器的 YOLOS 模型等。

另一个重要的工作是 2022 年 3 月份提出的 DINO，该工作在 Swin-L 骨干网络下，取得 COCO 物体检测评测集上的新纪录：63.3 mAP，首次在榜单上超越了所有基于传统检测头部的方法。这表明，DETR 不仅有望统一下游建模任务，其在性能上相比于传统的检测头部网络也开始取得优势。

3.掩码图像建模（MIM）兴起

在自然语言处理领域，GPT 和 BERT 是两大类语言建模方法，重塑了整个领域的学习范式。其中，BERT 依赖于掩码语言建模技术。早先，视觉领域对该方法并不看好。OpenAI 于 2020 年首次进行了尝试，发布了名为 ImageGPT 的视觉预训练模型，和主流方法相比仍有较大差距。直到 2021 年 6 月 BEiT 的出现，掩码图像建模才取得了突破。该方法的核心是采用 VQ-VAE 架构的模型，在处理图像的过程中，将其分割为 Patch，进行一定比例的遮盖后，转化为序列 Token，让模型使用类似 BERT 的方法对被遮盖部分进行还原，从而实现视觉建模的目的。

此后，掩码图像建模技术中具有代表性工作是 MAE、SimMIM。MAE 和 SimMIM 研究提出，只需要让模型预测遮盖的像素，就能够实现简单但有效的建模。目前 MIM 领域发展快速，在图像领域有 data2vec、PeCo、CAE 等，在视频领域有 MaskFeat、BEVT、Video 等，下游任务则有 ViTDet、MIMDet 等，研究非常活跃。

- 兴起高潮（2021.10-11）：MAE (Meta) / SimMIM (MSRA) / iBOT (字节)

- 百花齐放（2021年12月-）
  - 图像：data2vec (Meta) / PeCo (MSRA) / CAE (百度)
  - 视频：MaskFeat (Meta) / BEVT (复旦 & Microsoft) / Video MAE (南大)
  - 下游任务：ViTDet (Meta) / MIMDet (华科)

当然，在 MIM 快速发展的同时，近期有研究者开始探索 MIM 能够有效的原因，也有一些工作，例如 Feature Distillation 这一工作发现对比学习经过特征蒸馏以后也能够取得相同的效果。

（二）更大更稀疏的视觉模型

1.视觉大模型

预训练大模型是过去几年 AI 发展主旋律，但是主要集中于自然语言处理领域。视觉领域在 2021 年开始迎来进展。谷歌构建了一个扩展的 ViT 模型，拥有 18 亿参数，并使用 30 亿的标注图像进行训练，在 ImageNet 上取得了新的记录（90.45%）。这一工作还表明，在视觉领域上，模型同样符合 Scaling Law。即：模型越大、性能越好。

- Scaling ViT (谷歌CVPR 2022)
  - 18亿参数（30亿标注图像）取得ImageNet-1K分类任务新纪录90.45%
  - 视觉Transformer同样符合Scaling Law（扩展定律）

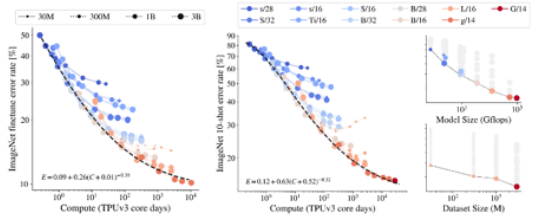
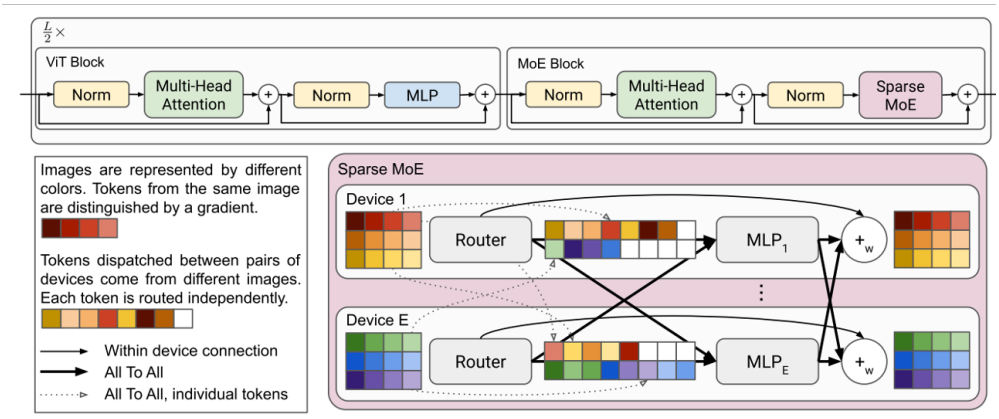


Figure 1: **Left/Center**: Representation quality, measured as ImageNet finetune and linear 10-shot error rate, as a function of total training compute. A saturating power-law approximates the Pareto frontier fairly accurately. Note that smaller models (blue shading), or models trained on fewer images (smaller markers), saturate and fall off the frontier when trained for longer. **Top right**: Representation quality when bottlenecked by model size. For each model size, a large dataset and amount of compute is used, so model capacity is the main bottleneck. Faintly-shaded markers depict sub-optimal runs of each model. **Bottom Right**: Representation quality by datasets size. For each dataset size, the model with an optimal size and amount of compute is highlighted, so dataset size is the main bottleneck.

Swin V2 则进一步证明了视觉大模型（30 亿参数）在广泛视觉问题上的有效性，其在图像分类、物体检测、语义分割和视频分类等任务上均达到了 SoTA 性能。这一工作也验证了自监督学习对于驱动大模型训练的有效性，基于 SimMIM 方法，Swin V2 用相比谷歌小 40 倍的标注数据（7000 万）达成了十亿级视觉模型的训练。

2.视觉稀疏动态模型

人脑作为一种稀疏动态模型，具有低功耗、部分激活、参数规模大、性能好等优势。研究者因此也在考虑建立视觉稀疏动态模型。MoE 是目前能够实现稀疏动态大模型的一种架构。在视觉领域中，已有研究者提出了 150 亿参数的 ViT-MoE。此外，也有研究者将 Swin Transformer 进行了 MoE 扩展，并将其应用于更广泛的视觉任务中，例如物体检测。



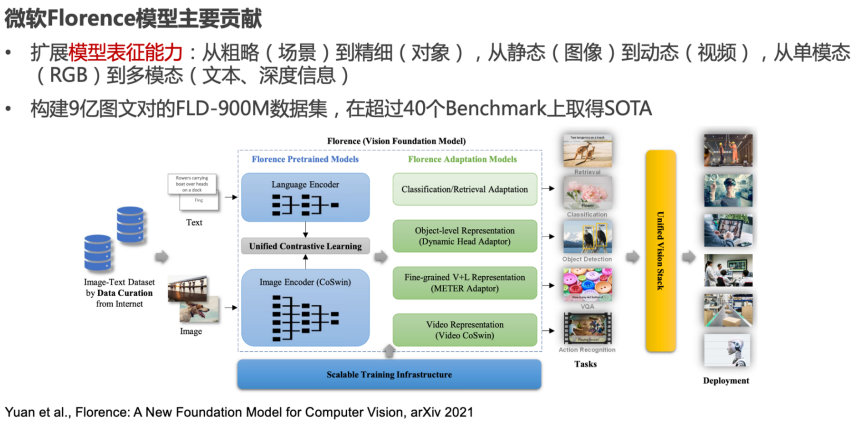
（三）视觉领域解锁新技能新模型

1.任意文本到图像生成

近年来，文本到图像生成的研究发展迅速。2021 年年初，OpenAI 提出 DALL·E 模型，可以根据用户输入的文本生成对应的图像，在研究领域引起轰动。2022 年年初，DALL·E 升级到第二代，提升了图像分辨率、文本匹配程度等指标。同时，谷歌也在近期发布了 Imagen 模型，效果同样出色。国内相关工作包括清华和智源共同完成的 CogView、微软亚洲研究院提出的 VQ-Diffusion 模型以及 NUWA-infinity 等。

2.多模态和零样本识别上的突破

通过提升视觉与语言之间的对齐，能够支持模型在更多多模态任务上取得应用，并降低标注数据的依赖情况。2021 年，OpenAI 发布 CLIP，在 ImageNet 的零样本分类任务上取得突破。之后的一年中，面向通用图文表征任务的模型发展迅速，典型案例包括微软的 Florence 模型，谷歌的 CoCa 模型，以及 DeepMind 的 Gato 模型和 Flamingo 模型等，分别属于面向不同下游任务的多模态模型。

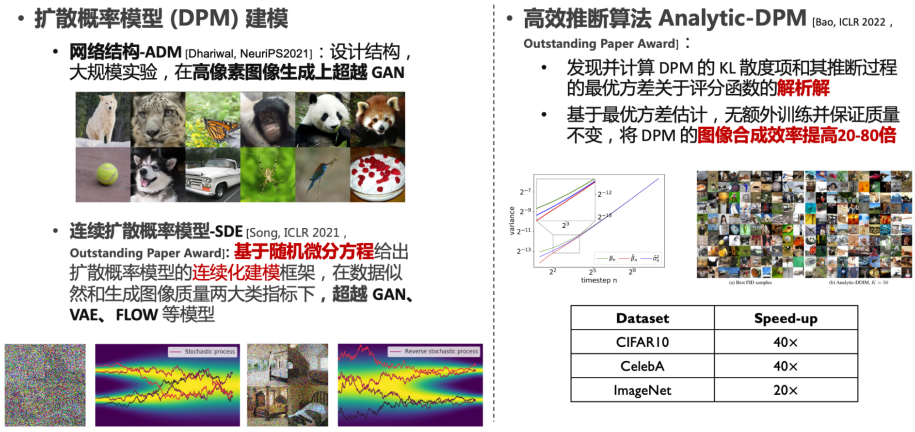


Yuan et al., Florence: A New Foundation Model for Computer Vision, arXiv 2021

3.扩散生成模型性能超越 GAN 和自回归方法

在图像生成方面，扩散概率模型（DPM），通过逆转扩散过程，迭代去噪的方式生成图像，在多个性能指标上已经超越了 GAN 和自回归模型。扩散模型的代表性工作包括 OpenAI 提出的网络结构 ADM，在高像素图像的生成效果超过生成对抗网络；斯坦福大学和谷歌大脑提出的随机微分方程（SDE）建模 DPM，在多种指标下超越 GAN、VAE、FLOW 等经典模型；清

华大学和中国人民大学合作的解析式 DPM 模型（Analytic-DPM），将 DPM 的图像合成效率提升了数十倍。SDE 和 Analytic-DPM 这两个代表性方法分别获得了 ICLR2021、2022 的 Outstanding Paper Award 奖项，说明该类模型在研究领域已得到广泛关注。



（四）重要模型或应用走向成熟

1.神经渲染技术 NeRF 走向成熟

来自 UCB、谷歌和 UCSD 的研究人员合作在 2020 年提出神经辐射场（NeRF）方法，可用于对场景进行建模，利用场景稀疏采样视角的图片集，可以合成场景新视角下的图像。该方法在近年来快速发展，已经从最初单一物体的多视角图片序列，扩展到面向互联网图片集，可以围绕一个景观（如不同的游客拍的不同类型的图片），通过对所有的图片进行合成，获得场景的神经辐射场。这一研究可以帮助企业构建大规模场景中的 NeRF 模型，建立整体的图像风貌（如城市景观）。

同时，NeRF 也扩展到了 3D 视觉任务上，如从图片中恢复高质量几何信息等，相关案例包括英伟达提出的快速 NeRF 重建方法，以及加州大学伯克利分校提出的，针对不同场景进行泛化的 PixelNeRF 等。这些技术在精度和效率上都取得了进步，推动 NeRF 技术走向应用。

此外，在应用对象方面，NeRF 已经从静态图像扩展到动态场景的表示和建模。此外，还有利用神经辐射场实现 3D 内容生成、可控的 2D 图片生成等研究，成果已应用在人体建模上，如国内研究者提出的 NeuralBody 工作等。

2.重要应用方向趋向成熟

计算机视觉领域的研究成果逐渐走向实际应用，在一些重要应用领域已经趋于成熟。今年视觉应用的主要进展包括：垂直任务上的性能提升，扩展更为丰富的应用场景，以及构建面向更复杂任务的数据集等。

例如，在场景文字检测方面，来自华中科大的研究人员提出的 DBNet 模型，在五个标准基准测试中取得当前最优的检测精度和效率。

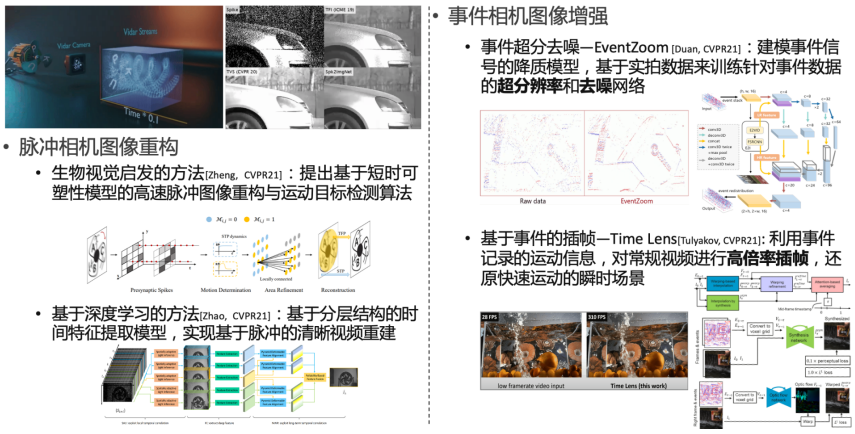
在扩展丰富的应用场景方面，计算机视觉已经在视频分类、视频解析和物体分割、视频内容创作、高分辨率实时抠图、视频会议——特别是元宇宙中的虚拟会议等实现应用。视频与图像最大区别是，前者引入了时序连续化的特征，需要研究者解决视频庞大计算量和时空维度解耦与冗余消除的问题。Transformer 也从图像走向了视频任务，在模型架构、分类精度等方面取得新进展。

在数据集构建方面，实际场景应用相关的数据集不断涌现。例如，在人脸识别领域，清华大学团队发布了超大规模 WebFace260M 数据集，清洗后的人脸图像数据量达到 42M。在视频分割领域，浙江大学构建了视频场景解析数据集 VSPW 和视频全景分割数据集 VIPSeg 等。这些大规模的数据集对领域发展起到了重要的推动作用。

二、新兴方向

1.计算摄像等视觉信号获取端研究的进一步发展

计算摄像研究的主要目的是提升视觉信号获取端的性能，相关的研究已有十多年的历史。近年来，具有代表性的工作是北京大学团队在脉冲相机方向的研究，团队研制的脉冲相机在图像重构、运动目标检测、视频重建等任务上都取得了良好的效果。



针对事件相机的研究也值得关注，典型案例包括 EventZoom、Time Lens 等。Time Lens 由华为提出，能够对常规视频进行高倍率的插帧，还原快速运动的瞬时场景。

2.视觉感知走向认知和推理

在高层视觉任务层面，视觉感知正逐步走向认知和推理。研究者试图面向视觉-语言任务，提出通用的视觉模型，以期实现认知层面的理解和推理工作。例如，亚马逊 AWS 在 2022 年的一项研究中提出了 X-DETR，其能够实现实例级的视觉和语言对齐，研究者希望通过学习通用概念表示，辅助更泛化的下游视觉-语言任务。另一项研究是 AI2 研究院和 UIUC 合作提出的迈向通用任务视觉系统的 GPV 模型，采用了问答模式统一各种底层、中层和高层的视觉任务，其中涵盖了物体检测识别、图像描述问答等，针对不同任务的自然语言问题输入，模型的答案输出既包含自然语言，又包含图像中相关的物体区域。

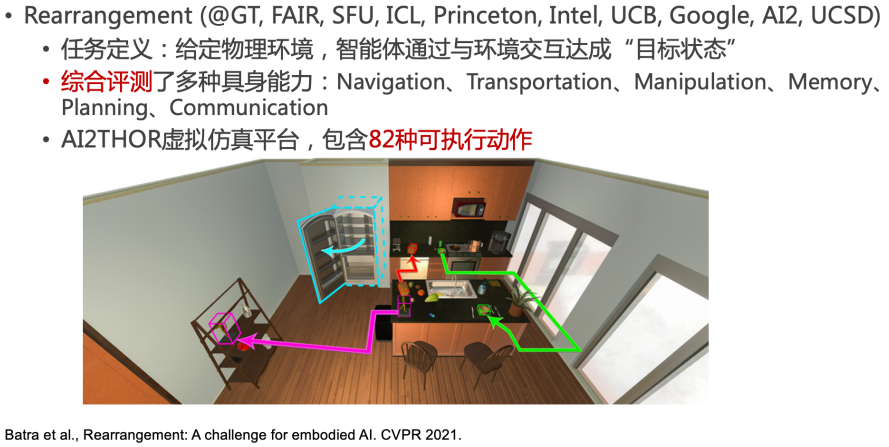
3.与概念结合辅助 3D 场景理解和下游任务

在 3D 场景理解方面，主要趋势可分为物体级和场景级两部分。物体级侧重于从视觉的认知与物理、几何等知识结合，通过概念组合的方式，让模型更好地学习物体 3D 的场景表示，从而辅助于下游的图文生成、视觉问答任务，并提升模型的泛化性。这些方法并非完全由数据驱动，而是结合神经和符号系统，将视觉感知、物理属性、动态预测和符号执行等纳入统一的框架中。针对场景级 3D 理解，研究者提出通过代理任务评测模型的 3D 场景理解能力，涵盖物体属性、空间关系推理等，代表性工作包括港中文深圳分校提出的 CLEVR 3D 工作，以及京都大学团队的 ScanQA 等。

4.模块化网络和具身智能兴起

在方法层面，值得关注是模块化网络，其具有可解释性强、泛化能力强的特点。在 MIT 的一项研究中，研究者模拟了人脑中的系统 2 运行机理，结合其中的逻辑推理规则，提高了神经序列模型的协调性和一致性。

在实际应用方面，视觉逐渐和语言、机器人等应用结合，促进了具身智能领域相关方法和任务的研究。去年，来自 FAIR、AI2、Google 等十余家科研机构的研究者提出了名为 Rearrangement 的具身智能评测任务，并组织了学术竞赛等活动，其中综合评测了导航、传输、操控、记忆、规划、通信等任务，采用 AI2THOR 仿真平台进行任务仿真与算法评测。



Batra et al., Rearrangement: A challenge for embodied AI. CVPR 2021.

5.视觉模型的鲁棒性与安全性受到关注

视觉模型的鲁棒性和安全性也受到研究者的关注，其内涵和外延不断扩展，如对抗样本的概念扩展，以及鲁棒性研究范式的统一等。此外，针对模型鲁棒性与安全性的评测基准也在逐渐完善，涵盖从卷积神经网络到 Transformer 等不同的模型架构，并将评测的训练范式拓展到了当前流行的自监督学习等新型的训练范式。

注：本章节观点整理自青源会 5 月 13 日及 6 月 1 日举行的研讨会。研讨会召集人为：清华大学黄高。参与研讨的嘉宾有：微软亚洲研究院胡瀚、中国科学院计算技术研究所王瑞平、旷视科技张祥雨、北京航空航天大学刘偲、商汤科技代季峰、华中科技大学王兴刚、北京智源人工智能研究院王鑫龙、北京交通大学魏云超、清华大学段岳圻、北京大学施柏鑫、浙江大学周晓巍、清华大学苏航、中国人民大学李崇轩、哈尔滨工业大学左旺孟。

第五章 智能体系结构与芯片领域进展及未来展望

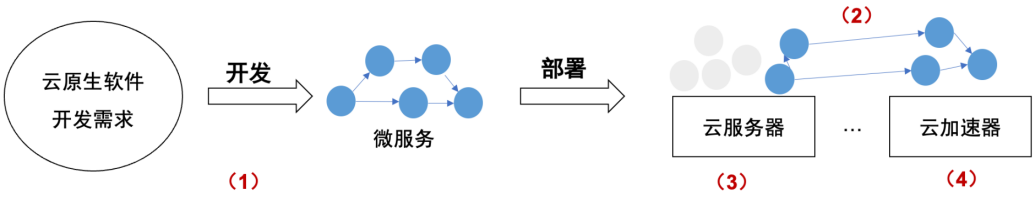
青源会智能体系结构与芯片方向学者

体系结构和芯片是人工智能基础层的重要部分，对 AI 计算的性能、效率、能耗等方面带来重要影响。在本章中，青源会体系结构与芯片专家就领域近年的发展热点和趋势进行了探讨。章节分为上层应用、中间系统软件、EDA 设计软件，以及底层芯片四个部分，分别探讨了这些领域值得关注的重点问题、挑战和发展趋势。

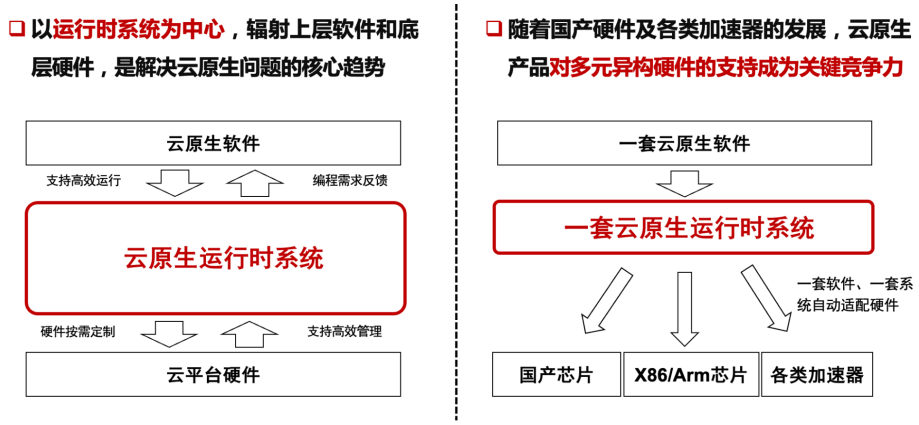
一、上层应用

1.云原生系统

云原生系统可分为软件开发、微服务部署、云服务器和云加速器等方面，主要面临以下挑战。一是云原生软件的编写开发问题，其涉及到复杂依赖微服务的高效表示。二是最小化流程处理时间的问题，涉及到对细粒度微服务的优化映射。三是进一步充分利用服务器硬件性能，将其用好、用满，涉及到高密度、高并发的混合部署技术。四是随着异构硬件数量增加，将这些硬件纳入云原生系统存在挑战，需要是吸纳异构硬件的共享。



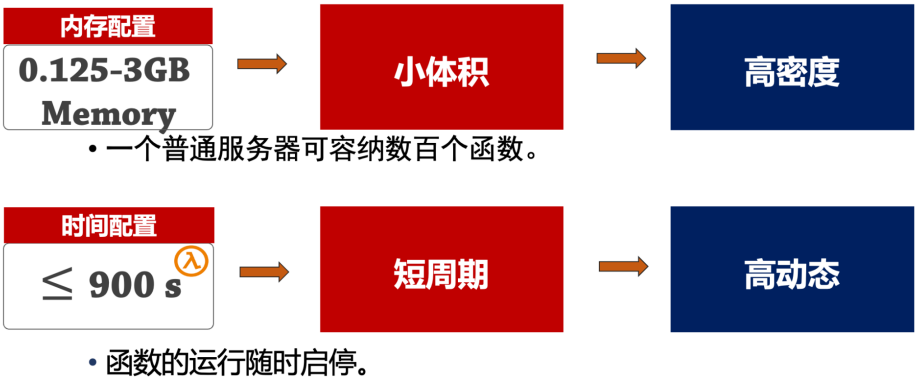
为了解决上述问题，以运行时系统为中心，辐射上层软件和底层硬件的思路，是解决云原生问题的核心发展方向。随着国产硬件（如加速器等）的发展，云原生产品对多元异构硬件的支持将会成为关键的问题。



2.Serverless 计算

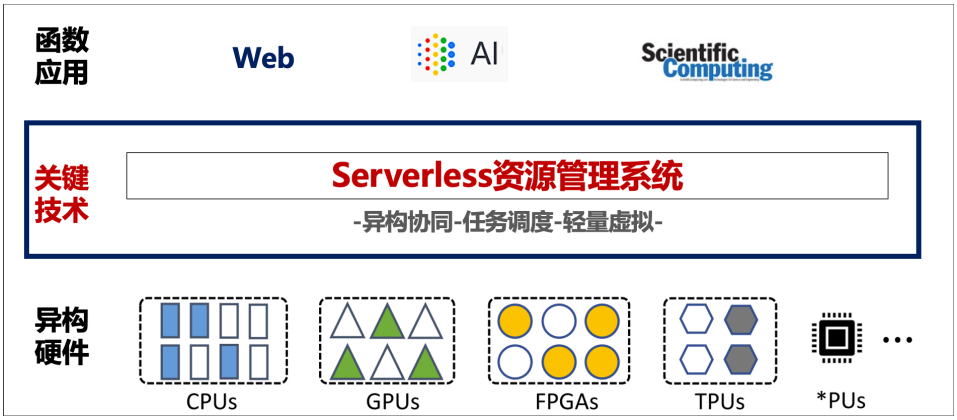
Serverless 计算近年发展速度加快，目前已经在 IoT 数据处理、小程序、地图等领域产生大量实际应用，同时也在大数据、在线网站、科学计算、增强学习训练等出现了探索性的应用落地。

Serverless 计算系统下的函数应用呈现两个特征：一是体积小。每个函数的内存容量通常只有几十兆，一个数百 GB 内存服务器同时可以容纳数百个应用，造成了高密度的部署。二是运行时间短，函数生命周期一般只有几秒时间，短周期特征导致计算机系统的动态性需求很高。



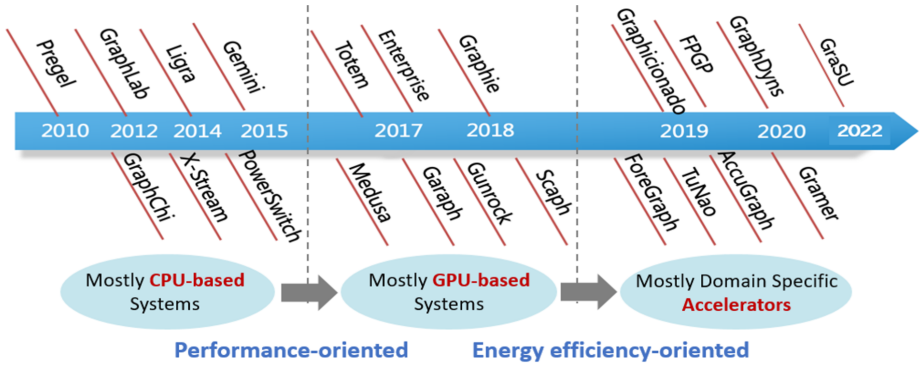
Serverless 函数的小体积和短周期特点，为服务器内存的资源管理带来了极大的挑战。将现有的 Web、AI 和大数据等负载部署到 Serverless 环境，将面临高扩展、高吞吐和高利用的研发需求。

同时，随着当前硬件体系结构快速变化，各种领域的专业硬件层出不穷，极大增加了硬件的异构性。要实现异构硬件下的高吞吐 Serverless 计算系统，需要研发新型的 Serverless 资源管理系统软件，同时在异构协同、任务调度、轻量虚拟等层面开展大量创新。

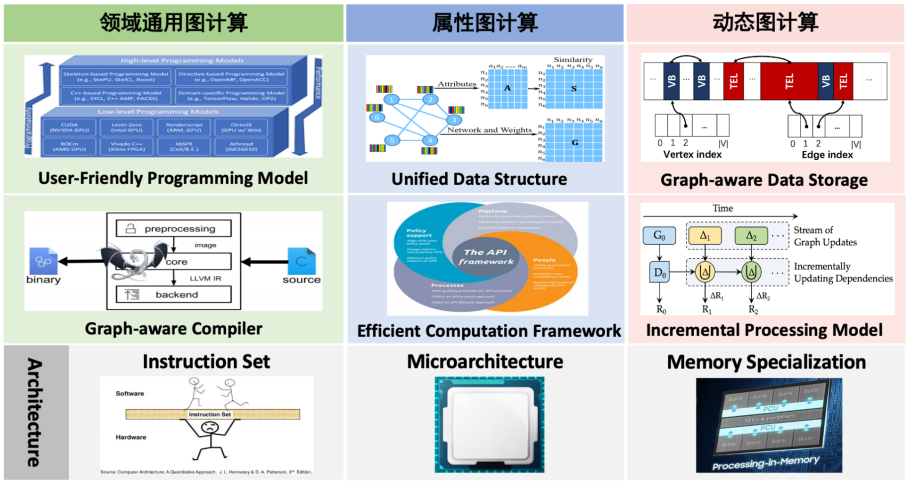


3.图计算

近年来，复杂图计算系统兴起，出现了超越二元常规图计算模式的“高阶多维图计算”特征。同时，传统上追求极致性能的图计算系统，正在向高可用方向快速转移。同时，图计算领域涌现出大量复杂图计算系统和加速器，如图挖掘类、图学习类、推荐系统类等。最后，相比传统图遍历类计算，图计算更加倾向于实用性应用，例如阿里达摩院的 GraphScope 等。



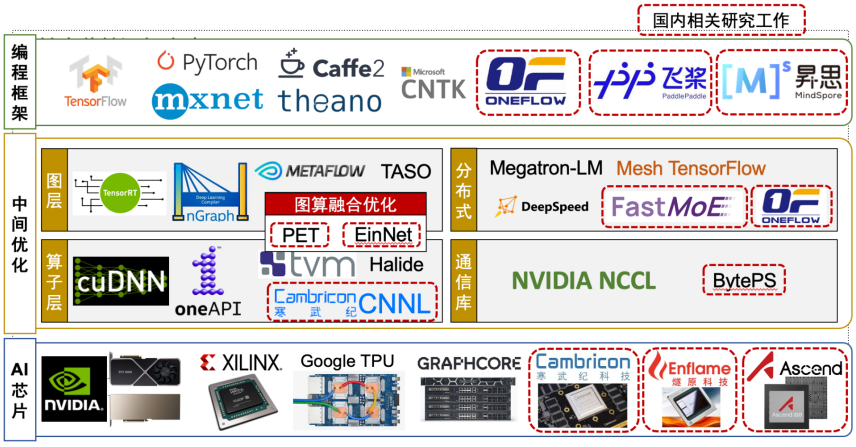
展望未来，实现复杂图计算系统，需要建立面向图计算的专业技术生态，其中包括完整的技术链、体系结构、系统软件构建、算法工具开发等。具体有以下几种演进路线：一是从分支通用到领域通用的图计算；二是研发动态图计算的专用加速器，三是进行复杂图模型的软硬件协同设计；四是推动属性图计算从实验室样品走向企业应用。



二、中间系统软硬件

1.深度学习编程框架

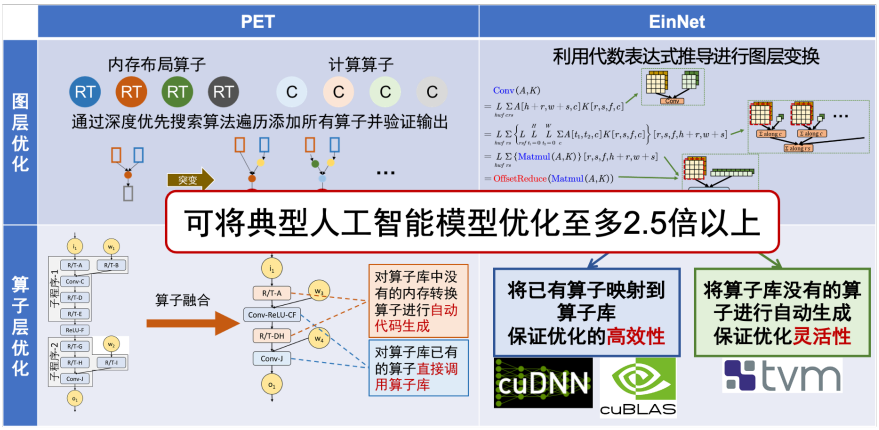
在深度学习兴起的同时，为了降低编程的复杂度，提高算法研发和模型构建的效率，深度学习编程框架应运而生。典型案例包括国际上的 TensorFlow、PyTorch，以及国内的 PaddlePaddle、MindSpore、OneFlow 等。



为了使深度学习应用能够更高效地部署到芯片上，许多研究关注人工智能模型到 AI 芯片的中间层优化。现有的优化框架，大体上将优化过程分为图层优化与算子层优化。其中，图层优化对人工智能模型的计算流图进行等价变换，以提高其性能，如 TensorRT、TensorFlow 等。算子层优化则通过数学库或代码生成的方式，提供可运行于人工智能芯片上的高效算子，如寒武纪的 CNNL 等。特别针对大规模训练，也有大量研究工作，对其分布式并行策略、调度策略、节点间通信等进行优化，如 FastMoE，NVIDIA NCCL 等。

此外，已出现图层和算子层共同优化的方法。例如，清华大学研究者最近提出了 PET 和 EinNet 两个研究工作，通过结合图层和算子层信息，对深度学习应用进行图算融合优化。

PET 在图层上将算子分为内存布局算子和计算算子，并通过深度优先搜索的方法遍历所有的优化组合，在代码生成阶段，对算子库中已有的算子直接调用算子库，对算子库中缺少的复杂内存布局算子进行代码的自动生成，并进行删冗、融合等优化，以进一步提升其运行效率。

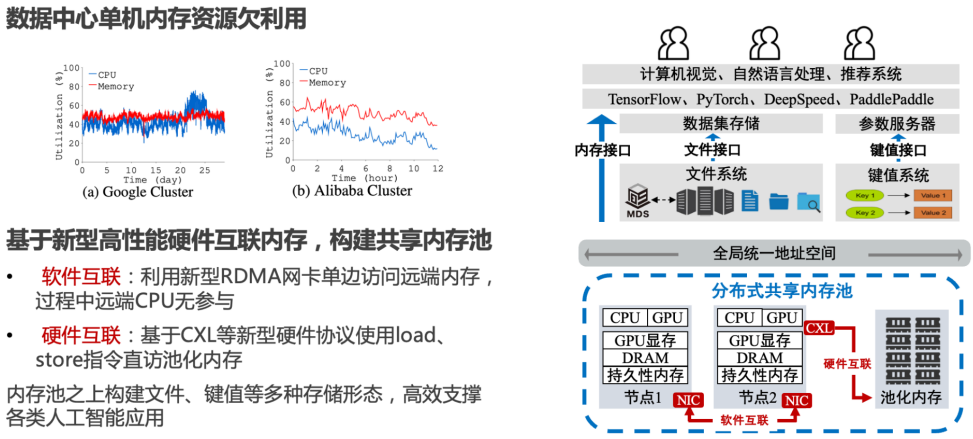


EinNet 通过代数表达式来对模型的计算图进行描述，并通过基于代数运算规则的自动推导进行变化，生成高效的计算图。在算子层分别将表达式映射到算子库或进行代码生成，以便同时保证优化的高效性和灵活性。

通过图算融合的优化方法，PET 和 EinNet 可将典型的人工智能模型优化至多 3 倍以上。

2.网络互联

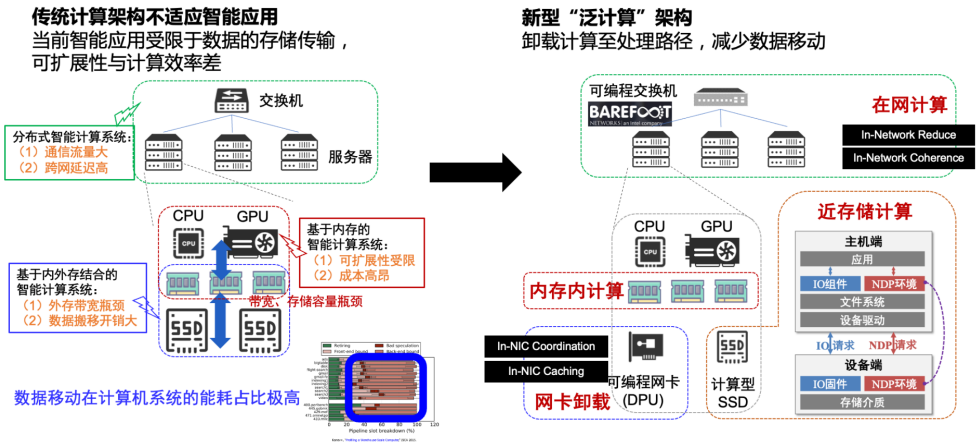
在通信网络方面，出现了网络互联形成大内存的趋势。分析指出，谷歌与阿里巴巴数据中心的服务器内存资源利用率均未达到 60%。为了提升内存资源利用率，近期的热点趋势是通过使用新型高性能硬件，将各个服务器的内存互联，构建共享内存池，提升内存利用效率。从软件互联的角度，则通过研究利用 RDMA 网卡单边访问远端内存，在这一过程中，远端 CPU 无需参与，减少额外的计算量。从硬件互联角度，则基于 CXL 等新型硬件协议，使用 load、store 指令直访池化内存。而在内存之上，构建文件、键值等多种存储形态，高效支撑各类人工智能应用。



[1] Qing Wang, Youyou Lu, and Jiwu Shu. "Sherman: A Write-Optimized Distributed B+ Tree Index on Disaggregated Memory." [SIGMOD'22]
[2] Minhui Xie, Youyou Lu et al. "Fleche: an efficient GPU embedding cache for personalized recommendations." [EuroSys' 22]

3.DPU

使用可编程的智能硬件（如 DPU）等进行近似数据处理，也是近期的热点趋势。传统的计算架构完全依赖 CPU/GPU 进行计算，需要先将数据从外存或跨网络拷贝至本机内存。这些数据拷贝受限于外存或者网络带宽瓶颈，可能影响系统的整体性能。可编程交换机、可编程网卡、计算型 SSD 等新型可编程智能硬件则为用户提供了近数据处理的机会。在这种新型的泛计算架构下，能够减少数据移动，提升行动性能。



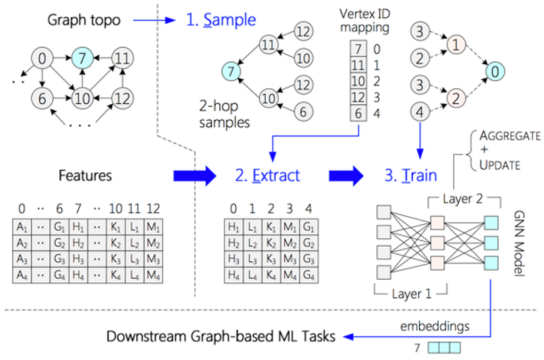
[1] Qing Wang, Youyou Lu, et al. "Concordia: Distributed Shared Memory with In-Network Cache Coherence." [FAST'21]
[2] Junru Li, Youyou Lu et al. "AINiCo: SmartNIC-accelerated Contention-aware Request Scheduling for Transaction Processing." [ATC' 22]

4.面向深度学习的软硬件协同优化

在推荐系统的存储和架构方面，基于深度学习的推荐系统，目前已经广泛应用在广告、电商、社交网络等。由于模型内存访问密集、计算稀疏，加上具有独特的不规则内存访问模式，使得在传统存储和体系架构下容易出现内存瓶颈和带宽瓶颈，限制应用的性能。而近存处理、存内计算等结构能够缓解推荐系统应用中对内存的压力。

在图神经网络方面，图神经网络在 AI 应用中非常流行，广泛应用于社交网络、知识图谱、推荐系统中。以训练流程为例，采样阶段占据训练过程总时间的 24%，抽取阶段，需要抽取出必要的数据，从 CPU 内存转移到 GPU 内存，则占据了过程的 54%。这也是由于内存随机访问、访问密集、计算稀疏等特征，造成在现有存储和体系架构下模型的训练效率不高。因此，可以通过近数据处理、加速采样阶段和数据移动等方式，来加速 GNN 的训练效率。对大规模的 GNN 训练，对分布式场景下，GNN 的通讯优化也是值得关注的问题。

➢ **面向图神经网络的存储和计算架构：**



5.硬件与压缩数据的融合

将硬件与压缩数据直接处理技术相结合，是未来体系结构与高性能计算的重要方向。通过基于规则压缩数据直接处理技术是一种新的计算范式，通过基于上下文无关文法的压缩表示，可以直接对压缩数据进行通用的数据处理，并和计算机并行体系结构相结合。相关成果已经能够做到对于大数据、非结构化数据以及数据库中的管理与分析，也包括在 GPU 等加速卡上压缩数据直接访问等。

6.CPU 和加速器集成于单一芯片设计

将 CPU 和加速器相结合，并放置在一个芯片上的设计，是未来计算机系统结构的重要研究方向。这种设计可以避免 CPU 和加速器之间的数据传输，使得两者有着更好的细粒度交互。同时，CPU 和 GPU 之间也可以进行细粒度的调度，表现出强劲的性能。

7.分布式系统

分布式系统方面，近来的研究进展是可编程网卡与分布式系统的协同设计。随着近年来 DPU 等可编程网络设备逐步出现，为传统的网络设备增加了计算能力，这些计算能力与分布式应用进行耦合开发，为加速分布式系统带来了新机遇。基于可编程网络的计算，具备与网络线速处理数据流的能力，可以极大提升分布式系统的效率。近年来，ATP 等系统的研发即为可编程网络与分布式系统耦合开发的用例。

智能网卡则在传统网卡上添加了计算芯片，将 DPU 部署在服务器端，提供对数据流的在网计算，可以将服务器的 CPU 从部分网络管理任务中解放出来，使 CPU 专注处理应用计算，有利于应用程序的部署和调度。

在分布式应用和可编程网络耦合开发背景下，设计传输计算一体化的全生命周期过程，将是下一步的研究重点，其中包括：设计编程范式，使传算一体能够适配异构的应用，便于程序员编程；设计编译技术，使通用编程范式适配异构的平台；设计运维方法，使异构平台上的处理能力、协同工作发挥最大的计算能力。

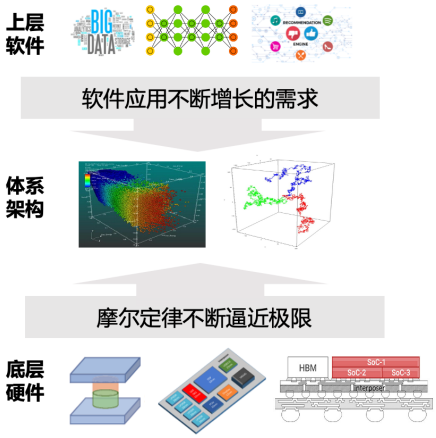
当然，设计传算一体的开发流程也面临挑战，主要在于用户熟悉的编程范式增量演进到传算一体，这一演化过程要求新的编程范式需要兼容已有的编程范式，提供新的语义或者更强的性能，同时需要对用户透明，方便用户使用。

三、EDA 设计软件

领域专用芯片的快速发展，对设计自动化工具提出迫切需求。传统 EDA 工具在系统层、RTL 层、架构层方面，无法对领域专用芯片的特殊问题进行建模和优化。因此，在这些层次上无法胜任领域专用芯片的设计。例如，传统的 EDA 无法处理神经网络加速器的数据流架构设计。目前领域专用芯片的上层设计多基于人工和经验开发，或者开发一些私有工具。

给定领域内的应用，自动设计某些指标最优化的专用架构，其中的关键问题是构建领域内的算法描述模型、硬件架构描述模型、设计空间搜索方法等。模型是 EDA 的关键之一，建立领域内通用的算法模型和架构模型，基于这些模型开发架构的设计空间搜索方法，使得专用领域的自动化设计成为可能。

- 应用算法进展驱动：
  - Transformer、图计算、推荐系统等
- 底层硬件进展驱动：
  - 2.5D/3D、chiplet、 新器件等
- 跨层次设计挑战：
  - 设计空间规模巨大、构成复杂
  - 架构设计的仿真验证耗时
  - 软件工具链显著影响硬件效率
- 技术路线：自动化DSA架构设计方法
  - 架构空间构建方法、架构表示方法、仿真平台/评估方法、空间探索方法等



定制计算架构的演进受到上层软件和底层硬件的两方面驱动。一方面，应用算法的发展对计算能力、存储容量、访问带宽等不断提出新需求。另一方面，摩尔定律逐渐放缓，新型底层器件、集成方式和封装工艺的出现，也为定制架构提供了新的机遇和挑战。

综合考虑这两方面的因素，EDA 需要实现在应用、架构、电路等方面的跨层次设计，这会导致架构设计空间不断增大，架构设计复杂度验证开销也迅速增加。为了克服这个挑战，可以针对领域应用的特点，合理设计架构表示方法，构建架构评估平台，并利用机器学习等搜索方法，高效地构建架构设计空间，并自动化地进行探索。

四、底层芯片

随着专用场景的日益发展，催生了许多专用加速硬件模块，来进行高效的处理和计算。但是随着专用计算算法发展加速，各个场景的碎片化倾向明显，这对芯片迭代和差异化设计有显著需求。

1.芯片设计成本优化

摩尔定律发展趋缓，芯片制造成本难以优化，且常规的整芯片设计开发周期较长。在这种背景下，催生了基于 Chiplet 的技术，名为 Hub+Side Die 模式。其中，Hub Die 为各场景所共有的一些功能模块，例如内存接口、主控 CPU 等。Side Die 则为各场景独有的计算加速模块，如 AI 加密等，它们之间通过 Die to Die 接口实现互联。通过这种方式，可以针对不同场景快速更换迭代专用计算 Chiplet，并选择高性价比工艺 Hub Die，也可以实现对不同场景的 IP 费用与设计开销摊分。

但 Hub+Side Die 模式主要面临四大核心难题：一是拆。这是项目级的考虑，即通过建立成本模型，根据项目所需的芯片面积、性能等需求，评估出合适的系统拆分机制和封装模式等。二是拼。这是架构层的考虑，主要包括 NoC 和 DFT 的设计，在多 Chiplet 的条件下，实现可靠的系统通信与测试。三是连。针对 Die to Die 的接口开发，基于并口或者串口方，实现超短距互联的高性能接口，实现国产化的类 UCIE 机制。四是封。基于国产本土封装技术，实现 MCM 到 2.5D 级别的封装模式。

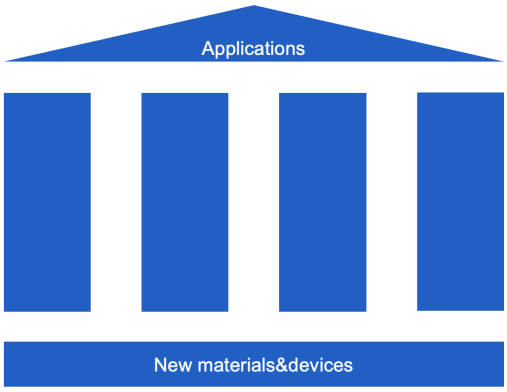
2.晶圆级集成

基于 Chiplet 的 Hub+Side Die 模式主打灵活的特点，在多项目上能够降低投入，而晶圆级的集成则突破现有的单 SoC 集成，实现大规模算力系统，以满足与日俱增的算力需求。传统的 HPC 领域通过多节点拓展算力，单节点计算能力不高，集群的计算密度低、通信开销大。而通过晶圆级集成，可以实现显著提高每个节点的算力大小。

晶圆级的集成除了上文提到的“拆、拼、连、封”四个问题外，还有以下方向值得关注。一是在架构设计阶段，每个 Chiplet 粒度探索值得研究。在 Chiplet 上，互联网络相较传统的 NoC，有更大的容错与拥塞控制要求。二是在实现阶段，需要针对集成的各个阶段研究相应的测试方案，并且由于大面积与大功耗特性，时钟、供电及散热问题不容小觑。三是在使用阶段，新型集成拓扑，给编译等中间件带来了新的研究问题。

3.智能芯片

智能芯片可以从两个方面进行阐释。一是 Chips for AI，即为 AI 为代表的新型应用，提供更好的芯片。AI 芯片体系结构作为整个系统的支柱，向上需要承接应用，向下要兼容新的器件。从上层应用发展看来，AI 驱动的新科学计算、新工业计算等新应用逐渐兴起，特别是以图神经网络为代表的图数据类应用，对体系结构提出了新的挑战。从底层材料和器件来看，针对计算方面有新型量子器件，针对存储有新型的存储器件，针对通信方面有光通信芯片。这些不同的新器件具有很大的发展潜力，但同样受到性能、功耗等方面的强约束和限制，这使得未来体系结构的发展需要考虑这些新型器件对特定领域应用的影响。

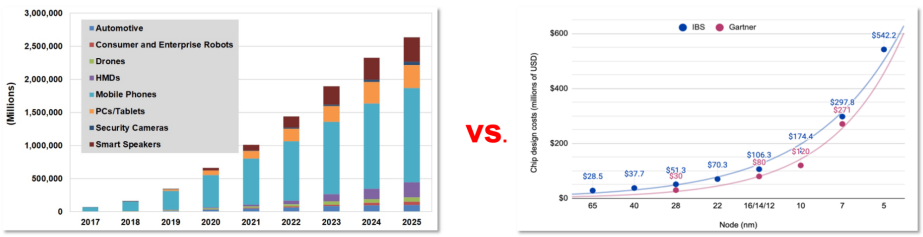


应用：
- 新科学计算，新工业计算
- 图数据

体系结构：
- 生态，包括开发效率
- 稀疏数据处理与计算
- 存算一体
- 多源异构SoC

新器件
- 存储：MRAM，RRAM
- 计算：量子
- 通信：光

二是 AI for Chips，即用人工智能做更好的芯片、更好地做出芯片。随着智能时代的到来和摩尔定律的大幅放缓，专用体系结构的芯片数量和种类大幅增加，体系结构将进入新的发展阶段。



- 黄金时代：更多种类、规模的芯片
- 更多、更专业设计人员+更长设计时间

芯片设计本身是一个代价很高的过程，即使经过了 40 多年的发展，也集成了越来越多的先进算法，芯片设计仍然周期长、过程复杂、对设计人员专业要求度高。特别是随着节点工艺的进入，芯片需要更专业的设计人员，花费越来越多的时间和金钱成本。解决芯片设计需求多和芯片设计代价高之间的矛盾，面临较大的挑战。而用人工智能方法，辅助参与芯片架构的设计和测试，将是有趣且有希望的尝试。

近日，谷歌的相关研究已经发表在《Nature》上——这项工作中，谷歌将芯片设计问题视为序列决策问题，利用强化学习的方法来解决，在不到 6 小时的时间内，该方法可以生成媲美或者超过人工的现代加速器往表上的布局。目前该方法已辅助设计了谷歌的 TPU 系列芯片，实现了落地应用。

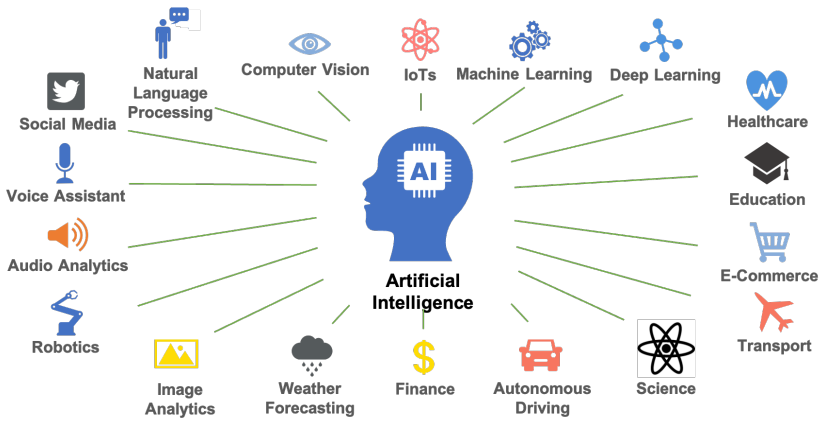
4.体系结构

在体系结构层面，有四个方向可以关注。一是让智能处理器更好用，包括生态构建，以及提升在不同规模、不同类别芯片上的应用开发效率等。二是稀疏计算带来的问题。稀疏作为一个很重要的特性，在过去几年被广泛关注，但是对于稀疏数据的处理仍然持续研究需要。三是不同芯片间，CPU 和芯片间的紧耦合需求。过去，研究者将各种功能从 CPU 上拿出来，做成单独的芯片。现在，为了满足在不同芯片间，CPU 和芯片间的紧耦合需求，研究者开始探索将多种芯片合到一起。例如，存算一体的研究就把数据和计算拉得更近，而多元异构 SoC 则把专用芯片之间拉得更近。

第六章 AI+Science 领域进展及未来展望

青源会 AI+Science 方向学者

随着人工智能从研究领域走向应用，正在推动各行各业快速发展。近来，科学研究已经成为 AI 应用的主战场之一，已形成名为 AI for Science 的领域成为新兴研究领域。本章中，多位青源会学者探讨了 AI+Science 的进展，主要分为以下几大场景：AI+生命科学、AI+材料科学、AI+大气科学、AI+神经科学、AI+应用数学/应用物理、AI+其他领域等。此外，与会者也探讨了 AI 可解释性在这一领域的进展和未来趋势。



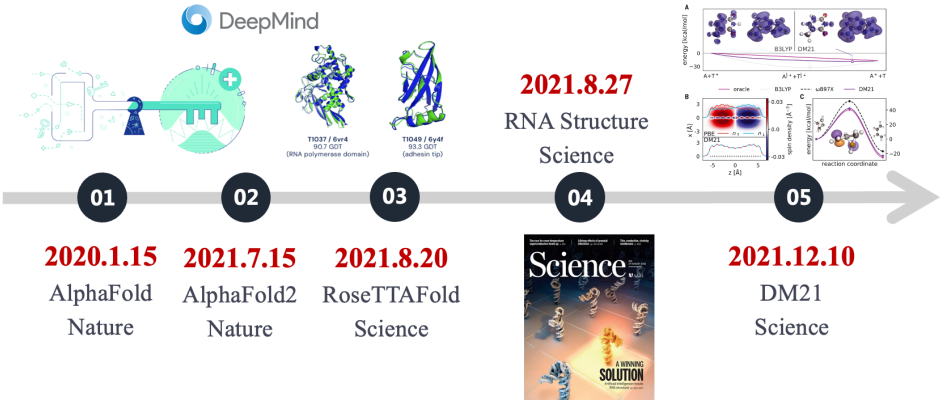
一、AI+生命科学

作为一种新型的智能科学计算模式，健康计算是以人工智能和数据驱动为核心的第四研究范式，它能够极大地助力人类探索生命科学的重大问题，在未来帮助人类更好地解决生命与健康问题。

（一）领域进展

在 AI+生命科学方向，近年来已有多项突破性技术。在蛋白质结构预测方面，具有代表性的工作是 2020 年 DeepMind 团队提出的 AlphaFold 和 2021 年新一代 AlphaFlod2，以及加州大学伯克利分校团队提出的 RoseTTAFold。此外，AI 在 RNA 预测，以及密度泛函估计等

领域均有论文发表，推动解决生命科学领域的一些重大问题。例如，2021 年 DeepMind 在《Science》上发表文章，提出了名为 DM21 的模型，用于量子计算中泛函密度估计，通过将 AI 的方法与物理模型相结合，有望解决生命科学领域的一些重大问题。



（二）发展方向

未来在 AI+生命科学领域，有四个方向具有发展潜力。

1.蛋白质结构预测

AlphaFold 在单链预测上已经产生了非常好的结果，但是在复合体的预测、相互作用预测、特殊蛋白质及其抗体结构的预测上，还存在挑战。

2.蛋白质设计

尽管以 Baker 为代表的研究人员，在基于领域的知识设计基础上，已获得了比较好的蛋白质设计的结果，论文也在今年早些时候发表在《Nature》上，但基于深度学习的蛋白预测依然存在精度不足等挑战，而基于结构的反向折叠预测目前仍处于初步发展阶段。未来在蛋白设计上面，AI 可能会持续产生突破。

3.药物设计

AI 在小分子药物设计方面已有成果，包括分子预训练模型、二维分子生成、分子性质预测、化学反应预测等。未来，在三维分子生成方面，基于靶结构的三维分子的生成，和大分子的药物设计上面，AI 仍有发展的潜力。

4.分子动力学模拟

基于 AI 促进分子动力学的模拟，将进一步推动药物和蛋白质的优化设计。分子动力学模拟是非常重要的研究领域，在小分子药物和蛋白质构向研究上，分子动力学都扮演着重要的角色。

（三）面临的挑战

AI+生命科学是一个交叉学科，其学科门槛相对较高，有多个方面的问题仍待解决。一是寻找到适合 AI 解决的问题，相比于提升 AI 解决特定问题的能力相对更为关键。例如，AlphaFold 选择了 AI 容易切入的蛋白质结构预测领域，才取得了轰动性的成果。二是在数据方面，生命科学中的数据存在稀疏、噪声大、数据分散等特点，还可能有隐私问题，这对 AI 算法的构建和训练带来挑战。三是在建模方面，生命科学数据类型可能有着更高的建模复杂度。四是 AI+生命科学研究对算力有非常大的需求，一般的研究实验室难以承受。最后，AI+生命科学对研究者和工程团队都有很高的要求，是一个系统性的学科。这需要 AI 科学家、生命科学家和工程团队三者有机结合，相互配合，共同完成具有挑战性的课题。

二、AI+材料

AI+材料的研究源自 2016 年 5 月《Nature》刊登一篇名为《人工智能将要重塑材料科学》文章。近年来，《Nature》、《Science》及其子刊相继报道了该领域的突破性进展。AI+材料总体可分为两个方向，一是新材料合成，二是材料预测建模。

新材料合成方面，机器学习模型预测的准确性已经超过了具有丰富经验的化学家。例如，2021 年 5 月 4 号的《Nature》的封面文章中介绍，研究者基于机器学习算法，采用合成失败的实验数据训练 AI，使其能够预测新材料合成的结果，使发明新材料的可能性大幅度提高。

主要思想：利用机器学习算法，用失败的实验数据预测了新材料的合成。

- 机器学习模型预测的准确率超过了经验丰富的化学家。
- 通过人工智能技术发明新材料的可能性大幅提高

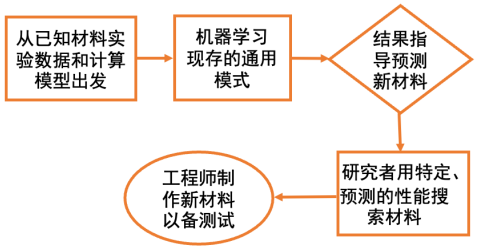
Nature， 2021.05.04

Published: 04 May 2016

Machine-learning-assisted materials discovery using failed experiments

Paul Raccuglia, Katherine C. Elbert, Philip D. F. Adler, Casey Falk, Malia B. Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A. Friedler, Joshua Schrier & Alexander J. Norquist

Nature **533**, 73–76 (2016) | Cite this article



在材料预测建模方面，AI 能够为实验条件下的材料性能预测开辟新途径。2021 年 1 月 6 日，《Nature》以封面文章报道，研究者在 10 纳米尺度上，机器学习模型来描述 10 万个原子多态转变的过程，从而预测材料的结构、稳定性、以及电子的特性等。

主要思想：利用原子机器学习模型描述10万个原子(10nm尺度)多态转变过程，预测结构、稳定性和电子性质。

- 观察到结构坍塌成明显的超高密度的非晶相（VHDA）
- 基于电子态密度的机器学习模型证实VHDA形成
- 为实验条件下的材料预测建模开辟了新途径

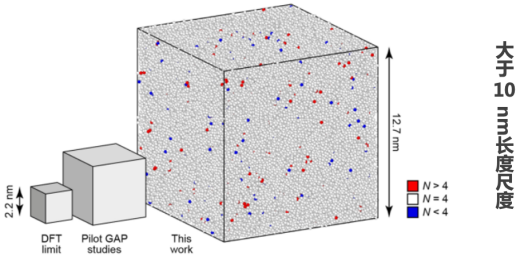
Nature， 2021.01.06

Article | Published: 06 January 2021

Origins of structural and electronic transitions in disordered silicon

Volker L. Deringer, Noam Bernstein, Gábor Csányi, Chiheb Ben Mahmoud, Michele Ceriotti, Mark Wilson, David A. Drabold & Stephen R. Elliott

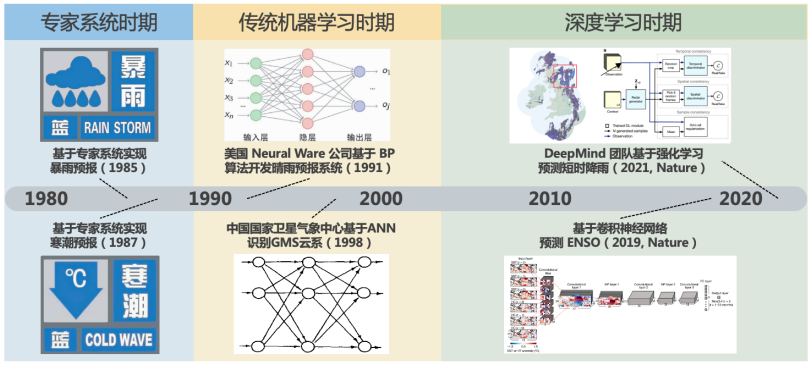
Nature **589**, 59–64(2021) | Cite this article



描述10 nm尺度体系的液态-非晶和非晶-非晶转变，预测结构、稳定性和电子性质

三、AI+大气科学

作为经典的数据分析的应用学科，大气科学和人工智能的渊源由来已久。2006 年以来，随着深度学习的发展，AI 在气象领域的应用越来越多，加速了大气科学的发展变革。

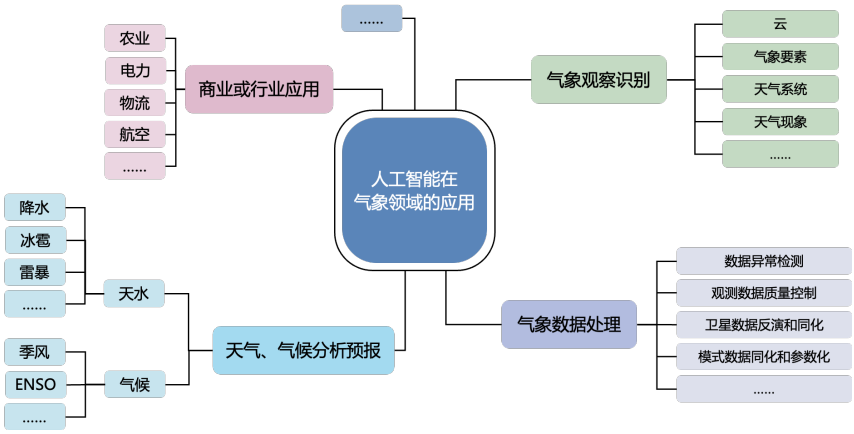


（一）领域进展

AI+大气科学方面，主要代表性工作包括利用神经网络预测气候现象和降水量等工作，如基于卷积神经网络预测厄尔尼诺现象，以及 2021 年 DeepMind 发表的文章——利用深度学习方法预测降水的研究等。

降水预测是非常复杂的问题，传统方法是解已知的、基于物理模型的微分方程组，效率较低，很难实现短时间内的准确预报。深度学习方法可以直接基于雷达测量数据预测降水情况，具备了短时间内降雨的预测能力。

此外，人工智能在气象领域应用也包括气象观察识别、气象数据处理、天气及气候分析预报、商业或行业应用场景等。



（二）发展方向

未来 AI+大气科学将在以下三大方向进一步发展。

1.在天气、气候预报模式的研发和方法上的改进：包括基于卫星资料数据的深度学习建模、超并发处理、多模式集成预报等。

2.AI 对气象探测技术的改进：如气象探测设备的卫星化、智能化和网络化发展，以及基于非气象观测设备的气象信息反演技术。

3.利用 AI 改进人工影响天气的作业方法：主要是基于 AI 和模拟技术实现“人影作业”的自动化技术。“人影”指的是用人为的方法去影响降雨。另外，还可以通过学习生雹天气系统的孕育期特征，用于消雹作业等。

（三）面临的挑战

AI+大气科学面临的挑战主要有四个方面，分别是：数据采集、数据规模、数据建模、研究团队。

1.数据采集

在气象研究中，采集到数据是异构的，非常丰富，且性质不同。例如，卫星测量数据主以图像为主，与雷达测量的数据，以及地面站测量的数据，彼此之间都有着不同的特性。其几何形状、时空分辨率、代表的物理含义都可能不相同。开发 AI 模型的过程中，需要对这些异构数据进行处理和融合。

2.数据规模

气象卫星、雷达等传感器时刻收集和测量气象数据，系统处理的数据规模在 TB 级别以上，且均为高精度数据。利用 AI 的方法处理气象数据，对算力、对计算资源需求较高。

3.数据建模

气象模型需要做到高可解释性，并保持物理一致性等要求。

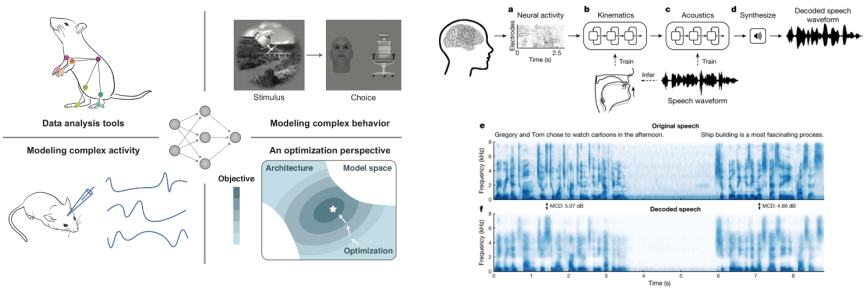
4.研究团队

研究团队方面，国内尚未完全建立 AI+大气科学的研究团队，国内学者相比其他国际学者研究影响力有待提升。

四、AI+神经科学

随着深度学习模型广泛应用于各个学科，神经科学家也在不断尝试使用深度神经网络作为工具，研究真实大脑中的神经元的活动机制，对神经元的活动进行建模，并构建更精准的脑机接口的模型。

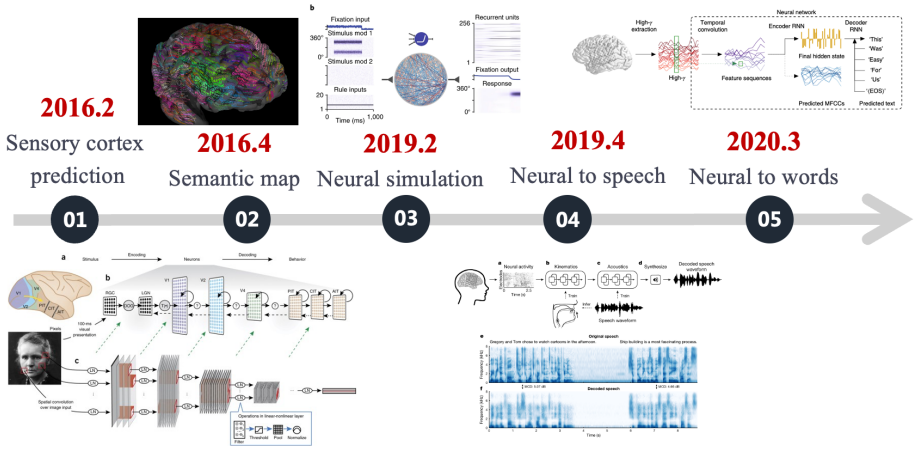
- 随着深度学习模型的快速发展，AI正在越来越多的用于神经科学的研究



Yang, G. R., & Wang, X. J. (2020). Artificial neural networks for neuroscientists: A primer. *Neuron, 107*(6), 1048-1070.
Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. Nature, 568(7753), 493-498.

（一）领域进展

从 2016 年至今，AI+神经科学已产生了一系列重大研究成果，已有多篇文章发表在《Nature》及子刊上。例如，2016 年 MIT 团队发现 CNN 模型可以预测人的视觉系统中不同皮层的神经活动。之后，加州大学伯克利分校团队发布了用磁向量方法结合神经影像数据，绘制出大脑中的语义地图的成果。近期，加州大学研究团队提出了结合神经网络模型的方法，能够对脑机接口技术进行改进，在神经信号中直接解码自然语言，使准确率大幅度提升。



（二）发展方向

AI+神经科学方面，有三个潜在的研究方向。一是利用深度学习的表征能力，研究对应的大脑表征，如感知觉和语言的神经表征预测等。二是启发建立新的认知机制假设，如基于神经科学研究去设计新型计算模型。三是构建更加实用的脑机接口模型，如借助预训练的语言模型从神经影像数据中解码连续的语言信号，提升脑机接口的解码性能等。

（三）面临的挑战

AI+神经科学也是一个交叉学科，具有较高的领域知识门槛。在特定场景和应用上，AI+神经科学研究的建模复杂度相对比较高，对算力的需求较大。

五、AI+应用数学/应用物理学

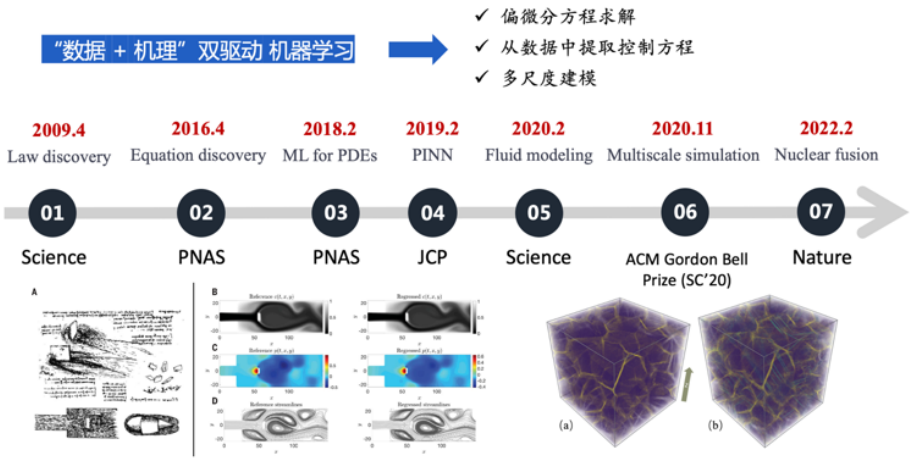
在应用数学和应用物理学研究和工程应用中，标注数据往往比较少，研究者所掌握的先验知识并不是特别全面。将 AI 和先验知识有效结合，是研究者所关注的重要问题。

（一）领域进展

目前在这一领域，AI 已实现有效的应用。一是将先验知识嵌入到 AI 模型中，减少模型本身的数据需求，提升模型的泛化能力和可解释性。此外，可利用标注数据构建特定的 AI 模型，在融合物理先验知识的同时，探索未知的物理规律，发现新知识。

在过去十几年中，AI+应用数学/应用物理领域已经取得许多进展，主要集中在“数据+机理”双驱动的机器学习方法和应用方面，其中包括偏微分方程求解、从数据中提取控制方程，以及复杂系统的多尺度建模等。

在偏微分方程求解方面，美国布朗大学研究者提出物理启发深度学习方法，可以用来对复杂的偏微分方程进行正计算和反问题的求解。几个月前，DeepMind 团队利用强化学习的策略设计了一个新的方法，可以解决核反应堆的控制问题。而机器学习方法在解决偏微分方程的同时，也可以加速大规模的原子和分子体系进行多尺度建模等。



（二）发展方向

当前在 AI+应用数学/应用物理学方面，有四个重要的发展方向。一是设计知识嵌入的合理机制。研究者依然需要重视先验知识，探索将已知知识、机理、定律有效嵌入到模型中的方法。二是探索从数据中发现未知知识的方法，需要将连接主义和符号主义的算法进行统一，在数据中发现新的知识、机理和规律。三是构建“数据+机理”融合的双驱动模型，对解决复杂系统的

多尺度建模仿真问题具有重要意义。四是推动上述方法在各类跨学科复杂场景和系统中的应用，比如湍流模拟、材料、电磁、生物化学、多物理场研究等。

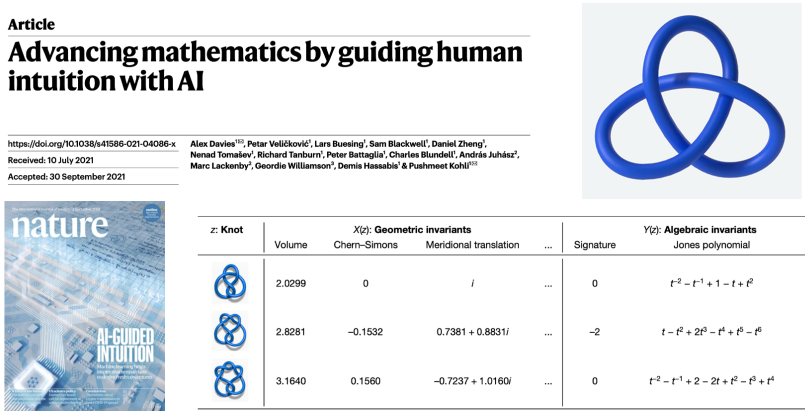
（三）面临的挑战

AI+应用数学/应用物理学同样面临的挑战包括：学科高门槛的特点、对待解决问题的定义、特殊的数据等。此外，将知识嵌入与知识发现有机统一，实现数据和已知的机理有效融合，是需要研究者进一步探索的难点。

六、AI 在其他领域的进展

1.AI+理论数学

理论数学主要是通过发现某种特定的数学模式，并利用这些模式来提出和证明数学猜想，形成相应的定理。DeepMind 去年在《Nature》发表一篇文章，利用机器学习归因技术，在纽结理论和表示论这两个领域，协助数学家发现了全新的猜想和定理，在拓扑几何界产生了轰动性的影响。



Article

Advancing mathematics by guiding human intuition with AI

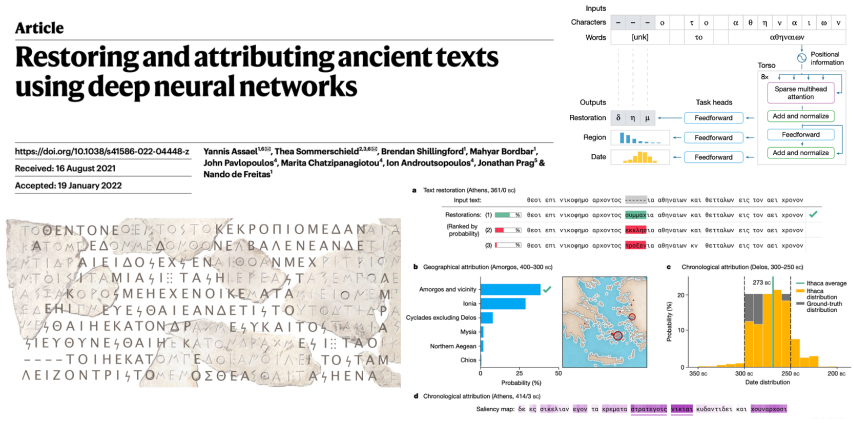
https://doi.org/10.1038/s41586-021-04086-x
Received: 10 July 2021
Accepted: 30 September 2021

Alex Davies[1✉], Petar Veličković[1], Lars Buesing[1], Sam Blackwell[1], Daniel Zheng[1], Nenad Tomašev[1], Richard Tanburn[1], Peter Battaglia[1], Charles Blundell[1], András Juhász[2], Marc Lackenby[2], Geordie Williamson[3], Demis Hassabis[1] & Pushmeet Kohli[1✉]

z: Knot	$X(z)$: Geometric invariants				$Y(z)$: Algebraic invariants		
	Volume	Chern–Simons	Meridional translation	...	Signature	Jones polynomial	...
	2.0299	0	i	...	0	$t^{-2} - t^{-1} + 1 - t + t^2$	...
	2.8281	−0.1532	$0.7381 + 0.8831i$	...	−2	$t - t^2 + 2t^3 - t^4 + t^5 - t^6$	...
	3.1640	0.1560	$-0.7237 + 1.0160i$	...	0	$t^{-2} - t^{-1} + 2 - 2t + t^2 - t^3 + t^4$	...

2.AI+考古学

机器学习在考古学方面已实现部分应用。2022 年早些时候，DeepMind 在《Nature》上发文，创建一种名为 Ithaca 的模型，可以帮助恢复古希腊铭文中缺失的文字，还可以为文字写下的时间和可能的地理来源提供建议，体现了 AI 在人文社科方面成功应用。



七、AI 的可解释性

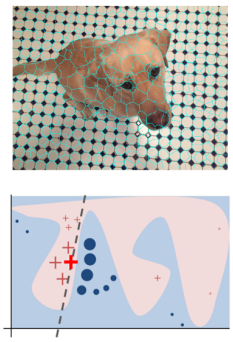
在 AI+Science 领域，模型的可解释性至关重要。理解模型推理和决策的依据，有助于人们信任其产生的结果，并为破解自然科学规律提供重要参考。

（一）领域进展

AI 可解释研究包括研究启发式算法、探索公理化体系等。启发式算法方面，可以用经典算法作为判断 AI 可解释性的准则，如 Lime 等算法，可以找到局部信息，通过线性函数拟合的方式实现。

由于各种启发式算法层出不穷，研究者很难判断算法的好坏，研究者进一步提出了可解释性的公理化理论体系，如果算法满足特定的公理（如 Efficiency、Symmetry、Linearity 等不同的性质），就可以认为算法的可解释性较好。

- 对于涉及重要决策的交叉领域，AI模型的可解释性至关重要
  - 医疗、自动驾驶、法律、教育……
- 深度神经网络作为非线性黑箱，如何让人类理解其决策？
- 经典算法：Lime (Ribeiro et al 2016)
  - 纯启发式：找到局部信息，线性函数拟合
- 公理化体系
  - 可解释性算法百花齐放，如何比较？
  - 假设公理：
  - Efficiency, Symmetry, Linearity, Null player, Consistency, etc.



Lime算法

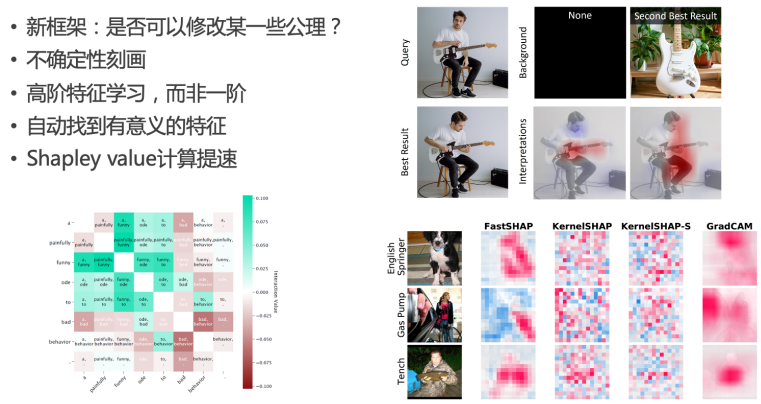
公理方法分为两种模式，分别是离散变量公理系统和连续变量公理系统。基于离散变量公理系统需要满足一些特征，比如 Local Accuracy、Missingness、Consistency 一致性等。其中，Shapley Value 是唯一的解决方案。基于这一思路，研究者提出了很多算法，包括 Sensitivity、Implementation Invariance 等的一系列公理。对于连续变量公理系统，主要是基于 Sensitivity、Implementation Invariance 等特征提出来一系列的解决方案，如基于 Integrated Gradients 方案等。

（二）面临的挑战和发展方向

未来在 AI 可解释方面，未来探索的重要问题是：通过提出新的框架，去修改某一些公理，或提出完全不同的、更加实用的算法。此外，AI 可解释性研究在实际应用中，需要对模型不确定性进行刻画。

此外，研究者也希望模型能够自动寻找到一些有特定意义的特征。例如，在 AI+Science 的特定场景，将模型找到的特征，和领域知识等实现统一，能够加快相关领域发现新的规律。

总体而言，AI 可解释性的理论和方法也是 AI+Science 研究的基础组成部分，与 AI+Science 进展发展相辅相成，相互促进。



注：本章节观点整理自青源会 2022 年 5 月 9 日及 6 月 1 日组织的研讨会。研讨会召集人为：中国科学院自动化研究所张家俊。参与研讨的嘉宾有：中国人民大学孙浩、清华大学兰艳艳、清华大学李国齐、北京智源人工智能研究院黄文灏、清华大学袁洋、中国科学院自动化研究所王少楠。

第七章 人工智能伦理与治理领域进展及未来展望

青源会人工智能伦理治理方向学者

当前，智能算法已经在人类社会经济和生活中许多领域得到广泛应用，如金融科技、智能制造、社会治理、医疗、智慧城市等。过去，算法是规则化、确定的、决策稳定、结果可解释的，几乎不存在安全问题。然而，数据驱动的人工智能算法（以下简称“智能算法”），模型本身高复杂、自适应演化等特性，使得智能算法黑箱难解释、决策结果不确定，安全问题凸显。为此，面向人工智能的技术规制和伦理问题逐渐进入研究人员和公众的视野。青源会人工智能与伦理治理领域的专家学者就智能算法的技术和治理方面存在的问题、挑战和发展方向进行了探讨。

一、领域进展

智能算法的安全问题总体可分为三大类，包括：算法技术缺陷带来的信任风险、广泛应用带来的社会影响，以及将算法作为工具用于对抗所带来的影响不可控等问题。解决智能算法带来的安全问题，主要可以从两个层面着手：技术层面和社会层面。技术层面主要针对算法本身的缺陷和漏洞，社会层面则重点关注算法在使用过程中产生的社会问题。

1.国际国内出台多项人工智能伦理规范

AI 伦理方面，人类社会对人工智能的伦理规范已经上升到了法律层面。2021 年 11 月 25 日，联合国教科文组织发布了《人工智能伦理建议书》，开始着手对人工智能伦理制定相应的规范框架。国内，2021 年 9 月 25 日发布了《新一代人工智能伦理规范》，将人工智能视为一项新型技术进行规范管理。

2.我国相关部门出台人工智能技术规制

在技术规制方面，我国在过去 1-2 年进展较快。例如，2021 年 9 月 17 日国家互联网信息办公室等九部委制定了《关于加强互联网信息服务算法综合治理的指导意见》，意见指出，要用三年左右时间，逐步建立治理机制健全、监管体系完善、算法生态规范的算法安全综合治理格局。此外，我国相继出台了两个法律性文件，分别是《互联网信息服务算法推荐管理规定》和《互联网信息服务深度合成管理规定》，前者重点关注互联网信息服务领域的推荐等五大类算法，重点规范信息呈现给互联网用户过程中的技术，后者重点关注深度学习和图像合成技术。这两个法律文件的出台，标志着我国开始实施对人工智能的技术规制。

3.针对智能算法的安全评估的研究启动

了解算法中存在的问题及其严重程度，并对算法进行评估，是 AI 伦理研究领域的未来发展方向。目前，算法安全评估研究已纳入国家重点研发计划中的“网络空间安全治理”专项里。在 2022 年的《国家重点研发计划申报指南》中包括了两项相关内容，一是“2.8 智能算法模型安全评估与风险监测技术”，二是“2.19 互联网信息推荐算法安全评估理论与方法”。在这一研究方向上，中国电子技术标准化研究院提交了《机器学习算法安全评估规范》，中国信息通信研究院与京东探索研究院也联合推出了《可信人工智能白皮书》，AI 伦理领域开始将算法评估的标准和规范上升到了更高的层次。

4.学界提出面向人工智能算法决策的审计框架

当智能算法代替人类进行程序化决策，就需要对过程进行审计。2020 年，一些国际研究者发表了题为《如何规制算法决策》的文章，其中提出了一个审计框架，包括四个不同的审计过程：

1）过程审计：要求算法决策过程是可解释的、透明的；

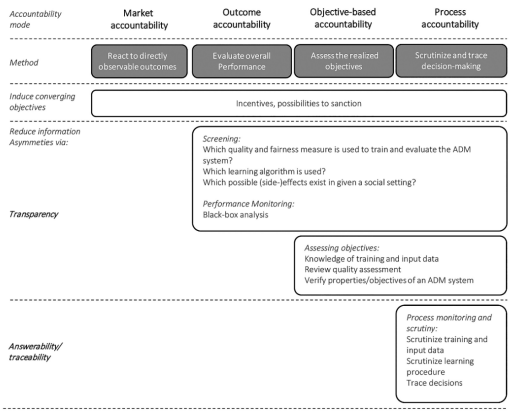
2）目标审计：不论决策过程，只关注算法或模型的训练目标；

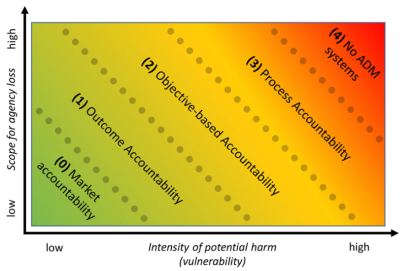
3）结果审计：不论目标，只判断结果的好坏；

4）市场审计：在决策过程、目标和结果之外，只根据市场对算法的认可程度进行评价，即使算法提供的服务在过程、目标和结果上存在问题，只要市场认可，算法即可存在。

国际研究者提出的这项审计框架，为国家人工智能伦理监测提供了可遵循的依据。

算法决策（ADM: Algorithmic Decision-Making）的审计：过程审计、目标审计、结果审计、市场审计





T. D. Krafft, K. A. Zweig, & P. D. König. How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. *Regulation & Governance*, *16*(1), 119-136, 2020.

5.呼吁建立横跨监管部门和学术界的统一体系治理方案

目前，学术界所认为的算法安全治理和监管部门提出的治理方案之间缺乏有效的沟通。监管部门主要从生态角度，关注治理问题，而学界研究更多注重模型的可解释性、公平性、鲁棒性、泛化性等技术层面。这需要一个统一的体系，让治理方案变为可以由技术支撑的实际方案。由上海交通大学张拳石等提出了“学界广泛接受的治理方案——统一体系”的概念，并发布了相关的论文。方案指出了 AI 伦理领域，技术和治理方案应当关注的重点，以及二者之间的对照。

6.可验证的算法鲁棒性方法持续发展

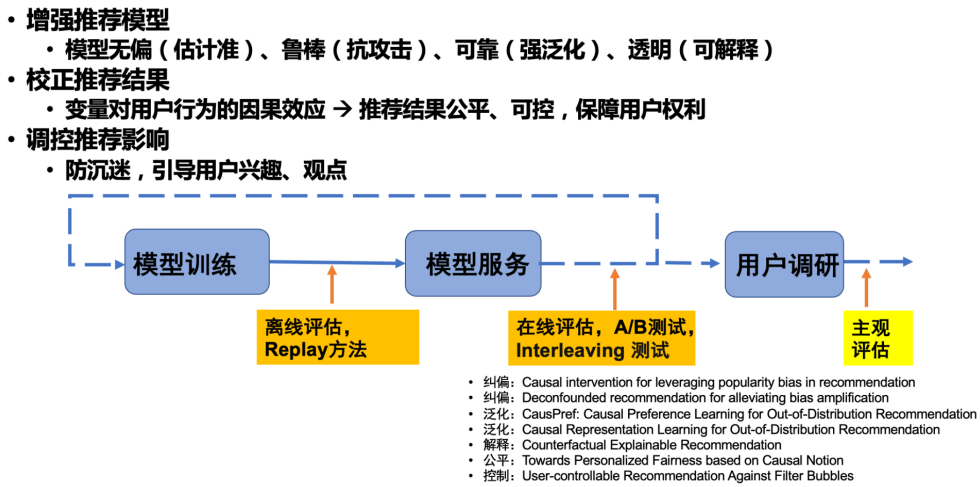
鲁棒性问题是 AI 领域一直倍受关注的问题，算法的安全研究从早期在封闭数据上的鲁棒性或者经验性鲁棒性研究，逐渐过渡到可以从理论上保证的、具有一定范围内预测结果的可靠性研究，因此被称为“可验证的鲁棒性”。可验证的鲁棒性方法在近几年得到了长足的发展，经历了从小数据小模型到大数据大模型的限定扰动研究，到现在大数据大模型的非限定扰动等几个阶段的发展。

7.人工智能的遗忘权获得重视

当模型训练之后就会“记住”训练数据。让模型能够定向地忘掉一些数据，被称为“Learning to Unlearn”，在数据隐私、安全等一系列的场景下有广泛的需求。目前，在这一领域主要有两种思路，一是基于分而治之的思路——将数据切分训练，之后再进行模型的集成，让数据和模型之间的耦合减弱，从而可以通过删除特定模型的方法，让系统定向地遗忘某些数据；二是基于梯度的模型微调，在数据训练后，对模型进行梯度等方面的处理，使其可以遗忘一些数据，属于后置的遗忘工作。

8.推荐系统合规引起社会关注

当前，应用在互联网信息服务中的推荐算法已经成为社会关注的重点。推荐算法作为一项应用广泛的技术，应当确保信息的呈现合规。一些研究提出使用增强推荐模型，提升模型的无偏、鲁棒、可靠、透明等方面的性能。此外，还有研究探索校正推荐结果的方法，降低推荐系统对用户带来的负面影响。最后，有研究从生态角度切入，通过建立良好的推荐系统发展生态，让算法良性发展。



二、面临的挑战

当前在 AI 伦理领域，重点有四个关键问题有待解决。

1.智能算法能力边界的判定

由于智能算法本身存在不确定和复杂的特性，需要对其能力进行判定，明确算法适用的范围和性能边界。由于智能算法中的模型结构非常复杂，难以兼顾其精准性和稳定性，需要明确二者应当达到的水平。而且，应用场景的变化，可能使过去算法基于的先验假设无法继续成立，因此需要提前获知算法的能力边界。

2.算法性能与可信约束之间的矛盾

在实际应用中，使用者需要对算法进行社会属性的约束，确保其公平无偏。但在技术上，使用者希望尽可能提高算法的性能。这两者之间可能会产生矛盾，平衡智能算法的性能和可信约束，是未来重要的研究问题。

3.算法黑箱与透明监测之间的鸿沟

算法黑箱与算法的规范和监测之间存在矛盾。算法本身是黑箱的，其内在的机理和外显行为之间缺少有效的对应。构造有效的反射互操作，帮助监管方跨越智能算法外显行为和内在机理之间的鸿沟，实现透明化监测。

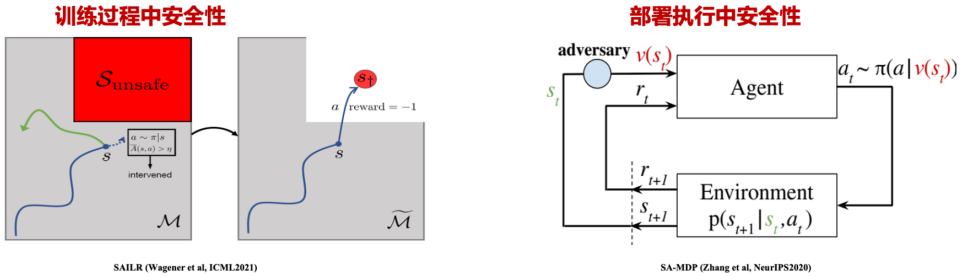
4.人机混合的复杂系统管理难度大

未来，算法将长期与人类并存，人机之间的边界日益模糊，未来将产生人机共存的复杂系统，这类系统的管理将面临新的挑战。

三、未来展望

1.从感知算法的安全性到决策算法的安全性

随着强化学习等决策算法开始大规模应用，需要判断决策过程中存在的安全问题及规避方法。

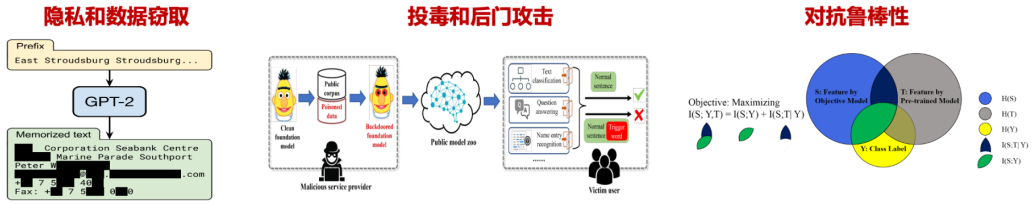


SAILR (Wagener et al, ICML2021)　　SA-MDP (Zhang et al, NeurIPS2020)

- Wagener, Nolan C., Byron Boots, and Ching-An Cheng. "Safe reinforcement learning using advantage-based intervention." In International Conference on Machine Learning (ICML), 2021.
- Huan Zhang, Hongge Chen, Chaowei Xiao, Bo Li, Mingyan Liu, Duane Boning, Cho-Jui Hsieh. Robust Deep Reinforcement Learning against Adversarial Perturbations on State Observations. (NeurIPS 2020)
- ChengYang Ying, Xinning Zhou, Hang Su, Dong Yan, Ning Chen, Jun Zhu. Towards Safe Reinforcement Learning via Constraining Conditional Value-at-Risk, To Appear in proc. of International Joint Conference on Artificial Intelligence (IJCAI), 2022.

2.预训练大模型的安全性引起关注

随着大模型自然语言和图像多媒体中得到应用，其带来的隐私问题、表面投毒问题、对抗鲁棒性等问题更加凸显。



Privacy and Data Hacking (Carlini et al, USENIX Security 21)　　Backdoor Attack (Chen et al, NeurIPS2021)　　Adversarial Robustness (Dong et al, NeurIPS2021)

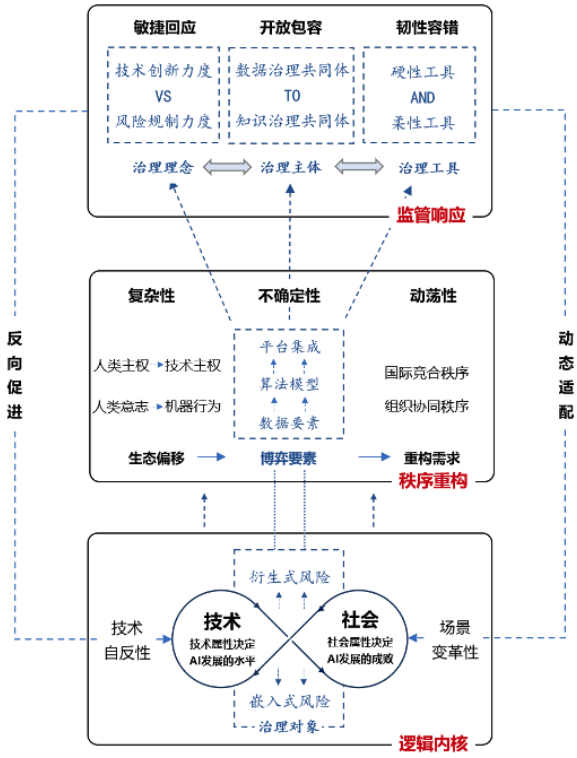
- Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, U. and Oprea, A., 2021. Extracting training data from large language models. In 30th USENIX Security Symposium (USENIX Security 21) (pp. 2633-2650).（谷歌学术引用226）
- Chen, Kangjie, Yuxian Meng, Xiaofei Sun, Shangwei Guo, Tianwei Zhang, Jiwei Li, and Chun Fan. "Badpre: Task-agnostic backdoor attacks to pre-trained nlp foundation models." International Conference on Learning Representations (ICLR), 2022
- Dong, Xinshuai, Anh Tuan Luu, Min Lin, Shuicheng Yan, and Hanwang Zhang. "How Should Pre-Trained Language Models Be Fine-Tuned Towards Adversarial Robustness?." Advances in Neural Information Processing Systems (NeurIPS), 2021

3.让算法治理拥抱“技治”主义

当前的算法治理基本上依靠法律条文、国家规范执行，应针对算法的特性发展相应的治理技术，实现算法的“以技治技”发展。

4.建立我国新一代人工智能治理工作框架

目前国内外均提出了人工智能治理的相关规范，但目前我国仍缺乏新一代人工智能治理的具体工作框架。应整合社会各界对 AI 社会技术复合体的离散性认知，实现对 AI 系统生态主权调适，突破 AI 包容审慎实践的探索，建立我国的统一治理框架。



建立基于“逻辑-秩序-监管”的人工智能治理工作框架

（姜李丹，薛澜. 我国新一代人工智能治理的时代挑战与范式变革. 公共管理学报，2022）

5.其他趋势

此外，部分专家还提到了 AI 伦理领域的一些发展趋势，包括：神经网络的交互式修复（对神经网络进行 Debug）、验证黑盒神经网络的正确性、AI 可持续发展、AI 技术全链条风险监测技术和监管机制建设等。

注：本章节观点整理自青源会 5 月 16 日及 6 月 1 日举行的研讨会。研讨会召集人为：中国科学院计算技术研究所沈华伟。参与研讨的嘉宾有：电子科技大学贾开、北京邮电大学姜李丹、清华大学庞祯敬、北京师范大学余振、清华大学曾雄、清华大学张辉、中国科学院计算技术研究所曹婧、中国科学院计算技术研究所陈薇、清华大学崔鹏、中国科学技术大学冯福利、清华大学苏航、阿里巴巴达摩院孙飞、上海交通大学张拳石。

本报告所含内容为一般性信息[1]，不构成任何专业业务的判断依据。同时，本报告的信息来源于已公开的信息内容，我们对这些信息的准确性、完整性或可靠性作尽可能的追求，但无法做任何保证和承诺，本文所列个人及其所在单位不对任何因此报告导致的直接或间接损失或损害承担任何责任。

青源会联系人：廖璐[2] | 智源社区项目与国际合作经理

邮　箱：liaolu@baai.ac.cn

报告内容联系人：戴一鸣[3] | 智源社区分析师

邮　箱：ymdai@baai.ac.cn

[1] 本报告文字整理自青源学术年会演讲内容。除有特殊说明的外，所用图片皆来自青源学术年会研讨的幻灯片。如涉及您的工作，请及时联系，我们将更新到后续的版本中。

[2] 负责青源会学术研讨活动的组织和协调。

[3] 负责本报告内容的整理编辑汇总。

■ 联系我们

电话：010-5095 5974

邮箱：press@baai.ac.cn

网站：hub.baai.ac.cn

智源社区微信公众号