浅谈人工智能的下一个十年

On the next decade of artificial intelligence

唐杰
（清华大学 计算机系）

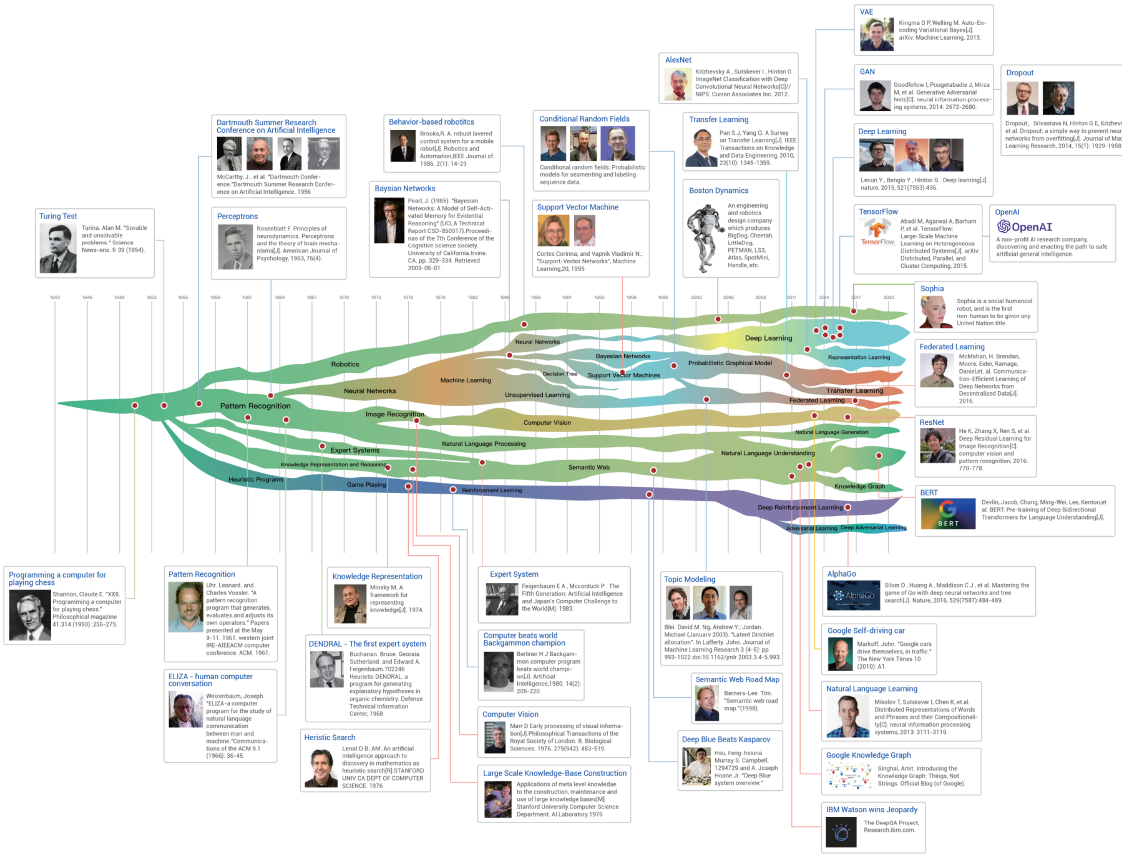
近年来，人工智能掀起了第三次浪潮，各个国家纷纷制定了人工智能的发展战略。在我国，人工智能已上升为国家战略，2016 年国务院发布《“十三五”国家科技创新规划》，明确将人工智能作为发展新一代信息技术的主要方向；2017 年 7 月，国务院颁布《新一代人工智能发展规划》；2017 年 10 月，人工智能被写入“十九大报告”；2020 年，人工智能又作为“新基建”七大领域之一被列为重点发展领域。

美国于 2016 年先后发布了《为人工智能的未来做好准备》和《国家人工智能研究与发展战略规划》两份报告，将人工智能提升到了国家战略的层面；2018 年，白宫举办人工智能峰会，邀请业界、学术界和政府代表参与，并成立了人工智能特别委员会。日本、德国等多个国家也发布了相关的战略、计划，大力推进人工智能的发展。

在这个时代背景下，本文浅谈人工智能历史并展望未来十年。首先，让我们从人工智能的发展历史中寻找灵感。

1 AI 的发展历史

下图给人工智能的历史和发展做了一个简单的梳理。



人工智能发展简史 (源图：https://www.aminer.cn/ai-history)

人工智能的起源可以追溯到阿兰·图灵 (Alan Turing)1936 年发表的《论可计算数及其在判定问题中的应用》，后来随着克劳德·香农 (Claude Shannon) 在 1950 年提出计算机博弈，以及阿

兰·图灵在 1954 年提出“图灵测试”，让机器产生智能这一想法开始进入人们的视野。1956 年达特茅斯学院召开了一个研讨会，约翰·麦卡锡 (John McCarthy)、马文·明斯基 (Marvin Minsky)、纳撒尼尔·罗切斯特 (Nathaniel Rochester) 以及克劳德·香农 (Claude Shannon) 等正式提出“人工智能”这一概念。算法方面，1957 年，弗兰克·罗森布拉特 (Frank Rosenblatt) 提出感知机算法 Perceptron，这不仅开启了机器学习的浪潮，也成为后来神经网络的基础，如果追溯的话，神经网络研究可以追溯到 1943 年神经生理学家麦卡洛克 (W. S. McCulloch) 和皮茨 (W. Pitts) 的神经元模型。

到了 20 世纪 60 年代，人工智能出现了第一次高潮，发展出了符号逻辑，解决了若干通用问题，自然语言处理和人机对话技术开始萌芽。其中的代表性事件是丹尼尔·博布罗 (Daniel Bobrow) 在 1964 年发表了 Natural Language Input for a Computer Problem Solving System，以及约瑟夫·维森鲍姆 (Joseph Weizenbaum) 在 1966 年发表了 ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine。早期的人工智能更多地侧重描述逻辑和通用问题求解，到了 60 年代末，爱德华·费根鲍姆 (Edward Feigenbaum) 提出首个专家系统 DENDRAL，并对知识库给出了初步的定义，这也孕育了后来的第二次人工智能浪潮。这个时期人们对人工智能的热情逐渐褪去，人工智能的发展也进入了一轮跨度将近 10 年的“寒冬”[①]。

20 世纪 70 年代末、80 年代初，人工智能进入了第二次浪潮，其中代表性的工作是 1976 年兰德尔·戴维斯 (Randall Davis) 构建和维护的大规模的知识库，1980 年德鲁·麦狄蒙 (Drew McDermott) 和乔恩·多伊尔 (Jon Doyle) 提出的非单调逻辑，以及后期出现的机器人系统。1980 年，汉斯·贝利纳 (Hans Berliner) 打造的计算机战胜双陆棋世界冠军成为标志性事件。随后，基于行为的机器人学在罗德尼·布鲁克斯 (Rodney Brooks) 和萨顿 (R. Sutton) 等的推动下快速发展，成为人工智能一个重要的发展分支。其中格瑞·特索罗 (Gerry Tesauro) 等打造的自我学习双陆棋程序又为后来的增强学习的发展奠定了基础。机器学习算法方面，这个时期可谓是百花齐放、百家争鸣。杰弗里·辛顿 (Geoffrey Hinton) 等提出的多层感知机，解决了 Perceptron 存在的不能做非线性分类的问题；朱迪亚·珀尔 boer(Judea Pearl) 倡导的概率方法和贝叶斯网络为后来的因果推断奠定了基础；以及机器学习方法在机器视觉等方向取得快速发展。

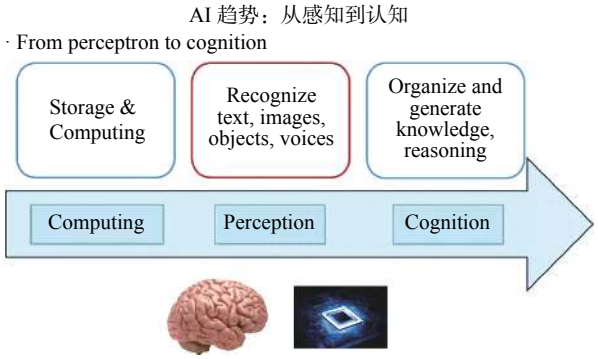
20 世纪 90 年代，AI 出现了两个很重要的发展：一方面是蒂姆·伯纳斯·李 (Tim Berners-Lee) 在 1998 年提出的语义网，即以语义为基础的知识网或知识表示。后来又出现了 OWL 语言和其他一些相关知识描述语言，这为知识库的两个核心——问题知识表达和开放知识实体给出了一个可能的解决方案 (尽管这一思路在后来一直没有得到广泛认可，直到 2012 年谷歌提出知识图谱的概念，才让这一方向有了明确的发展思路)。另一个重要的发展是统计机器学习理论，包括瓦普尼克·弗拉基米尔 (Vapnik Vladimir) 等提出的支持向量机、约翰·拉弗蒂 (John Lafferty) 等的条件随机场以及大卫·布雷 (David Blei) 和迈克尔·乔丹 (Michael Jordan) 等的话题模型 LDA。总的来讲，这一时期的主旋律是 AI 平稳发展，人工智能相关的各个领域都取得了长足进步。

第三次人工智能浪潮兴起的标志可能要数 2006 年，Hinton 等提出的深度学习，或者说 Hinton 等吹响了这次浪潮的号角。与之前最大的不同在于这次引领浪潮冲锋的是企业：塞巴斯蒂安·特龙 (Sebastian Thrun) 在谷歌领导了自动驾驶汽车项目；IBM 的沃森 (Watson) 于 2011 年在《危险边缘》(Jeopardy) 中战胜人类、获得冠军；苹果在 2011 年推出了自然语言问答工具 Siri 等；2016 年谷歌旗下 DeepMind 公司推出的阿尔法围棋 (AlphaGo) 战胜围棋世界冠军李世石等。可以说这次人工智能浪潮的影响是前所未有的，其中具体的进步与发展将在下文展开介绍。

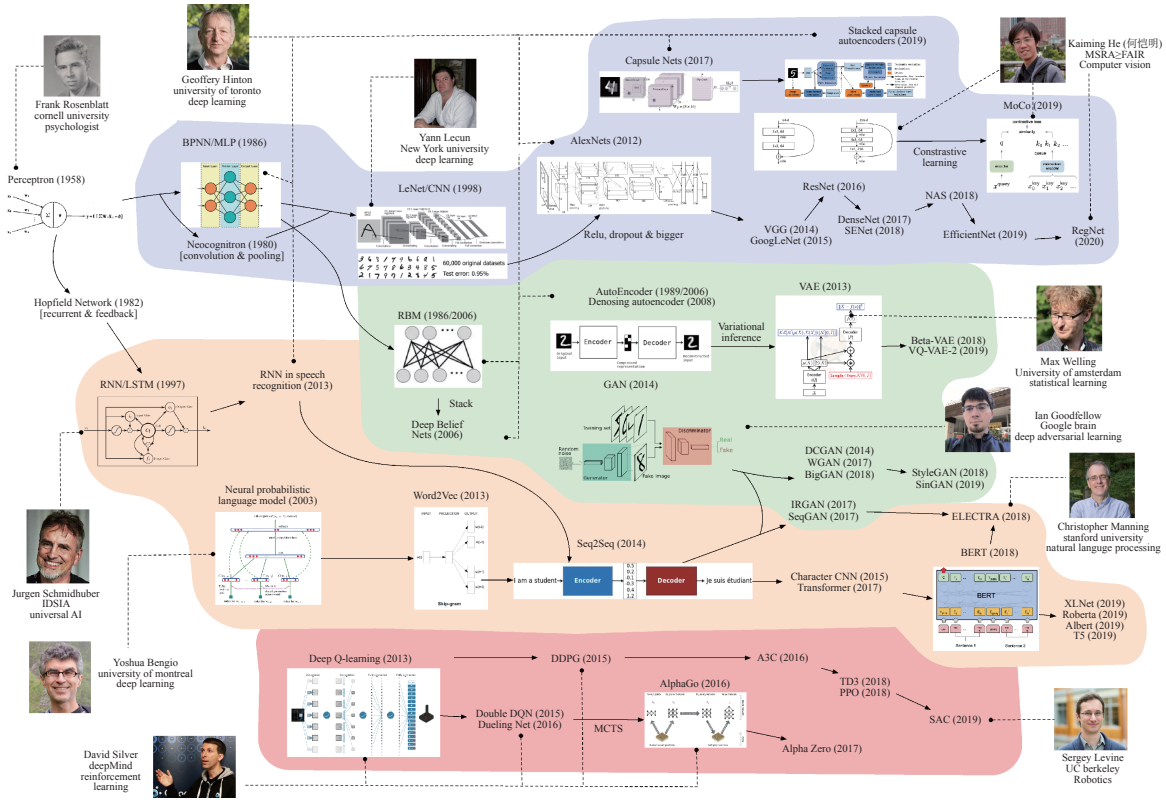
2 AI 近十年的发展

深入分析 AI 近十年的发展，会逐渐看到一个重要的现象：人工智能在感知方面取得了重要进展。在语音处理、文本处理、视频处理等多个方面，人工智能处理的效率和效果都已经超越了人类。可以说人工智能在感知方面已经逐渐接近人类的水平。人工智能也从感知开始逐渐走向认知，如下图所示：

① 很难说什么是真正的寒冬，不过在这个时期大家对人工智能的期望降低了。



首先来看看 AI 在感知方面做了哪些事情。在感知方面，AlphaGo、无人驾驶、文本和图片之间的跨媒体计算等取得了快速发展。总体来看，算法是感知时代最重要、最具统治力的内容。深度学习是近 10 年机器学习领域发展最快的一个分支，由于其重要性，3 位教授 (Geoffrey Hinton、Yann Lecun、Yoshua Bengio) 因此同获图灵奖。如果把最近十年的深度学习相关的重要算法进行梳理归类，可以得到下图所示的发展脉络。



总体来说，主要有 4 条发展脉络。

第 1 条发展脉络 (浅紫色区域) 以计算机视觉和卷积网络为主。这个脉络的进展可以追溯到在 1979 年福岛邦彦 (Kunihiko Fukushima) 提出的 Neocognitron。该研究给出了卷积和池化的思想。1986 年杰弗里·辛顿提出了反向传播训练 MLP(之前也有几个类似的研究)，该研究解决了感知机不能处理非线性学习的问题。1998 年，以扬·勒丘恩 (Yann LeCun) 为首的研究人员实现了一个 7 层的卷积神经网络 LeNet-5 以识别手写数字。现在普遍把 Yann LeCun 的这个研究作为卷积网络的源头，但其实在当时由于 SVM 的迅速崛起，这些神经网络的方法还没有引起广泛关注。真正使得卷积神经网络荣耀登场的事件是，2012 年 Hinton 组的 AlexNet(一个设计精巧的 CNN) 在 ImageNet 上以巨大优势夺冠，这引发了深度学习的热潮。AlexNet 在传统 CNN 的基础上加上了 ReLU、Dropout 等技巧，并且网络规模更大。这些技巧后来被证明非常有用，成为卷积神经网络的标配，被广泛发展。顺着 AlexNet 的思想，LeCun 组在 2013 年提出了一个 DropConnect，把 error rate 降低到了 11%。而新加坡国立大学 (NUS) 的颜水成组则提出了一个重要的 Network in Network(NIN) 方法，NIN 的思想是在原来的 CNN 结构中加入了一个 1×1 conv 层，NIN 的应用 2014 年也实现了 Imagine 另一个突破——图像检测的冠军。Network in Network 更加引发了人们对 CNN 结构改变的大胆创新。因此，两个新的架构 Inception 和 VGG 在 2014 年把网络加深到了 20 层左右，图像识别的 error rate(越小越好) 也大幅降低到 6.7%，接近人类错误率的 5.1%。2015 年，微软亚洲研究院 (MSRA) 的任少卿、何恺明、

孙剑等, 尝试把 Identity 加入到卷积神经网络中并提出 ResNet。最简单的 Identity 却出人意料的有效, 直接使 CNN 能够深化到 152 层、1202 层等, error rate 也降到了 3.6%。后来, ResNeXt、Residual-Attention、DenseNet、SENet 等也各有贡献, 各自引入了 Group convolutio、Attention、Dense connection、Channelwise-attention 等, 最终 ImageNet 将 error rate 降到了 2.2%, 远远低于人类的错误率。现在, 即使手机上的神经网络, 也能达到超过人类的水平。而另一个突破——在图像检测中, 任少卿、何恺明、孙剑等优化了原先的 R-CNN、fast R-CNN 等通过其他方法提出 region proposal, 然后用 CNN 去判断是否是 object 的方法, 提出了 faster R-CNN。Faster R-CNN 的主要贡献是使用和图像识别相同的 CNN feature, 发现 feature 不仅可以识别图片内容, 还可以用来识别图片的位置。也就是说, CNN 的 feature 非常有用, 包含了大量的信息, 可以同时用来做不同的任务。这个创新立刻把图像检测的 MAP 也翻倍了。在短短的 4 年中, ImageNet 图像检测的 MAP(越大越好) 从最初的 0.22 达到了 0.73。何恺明后来还提出了 Mask R-CNN, 即给 faster R-CNN 又加了一个 Mask Head, 发现即使只在训练中使用 Mask Head, 其信息可以传递回原先的 CNN feature 中, 获得了更精细的信息。由此, Mask R-CNN 得到了更好的结果。何恺明在 2009 年就以一个简单有效的去雾算法得到了 CVPR Best Paper, 在计算机视觉领域声名鹊起。后来更是提出了 ResNet 和 Faster R-CNN 两大创新, 直接颠覆了整个计算机视觉/机器学习领域。

另一方面, CNN 结构变得越来越复杂, 很多结构都很难通过直觉来解释和设计。2017 年, Hinton 认为反向传播和传统神经网络还存在一定缺陷, 因此提出 Capsule Net, 该模型增强了可解释性, 但目前在 CIFAR 等数据集上效果一般, 这个思路还需要继续验证和发展。谷歌提出了自动架构学习方法 NasNet(neural architecture search network) 来自动用 Reinforcement Learning 去搜索一个最优的神经网络结构。Nas 是目前 CV 界一个主流的方向, 可以自动寻找出最好的结构, 以及给定参数数量/运算量下最好的结构 (这样就可以应用于手机), 这是目前图像识别的一个重要发展方向。2019 年 4 月何恺明发表了一篇论文, 表示即使 Random 生成的网络连接结构 (只要按某些比较好的 Random 方法), 都会取得非常好的效果, 甚至比标准的好很多。Random 和 Nas 哪个是真的正确的道路, 这有待进一步的研究。最近, 恺明等提出了动量对比度 (MoCo) 用于无监督的视觉表示学习。MoCo 可以胜过在 PASCAL VOC、COCO 和其他数据集上进行监督的预训练对等任务中的检测/细分任务, 有时会大大超过它。这表明在许多视觉任务中, 无监督和有监督的表征学习之间的鸿沟已被大大消除。

第 2 条发展脉络 (浅绿色区域) 以生成模型为主。传统的生成模型是要预测联合概率分布 $P(x, y)$。机器学习方法中生成模型一直占据着非常重要的地位, 但基于神经网络的生成模型一直没有引起广泛关注。Hinton 在 2006 年的时候基于受限玻尔兹曼机 (RBM, 一个 20 世纪 80 年代左右提出的基于无向图模型的能量物理模型) 设计了一个机器学习的生成模型, 并且将其堆叠成为 Deep Belief Network, 使用逐层贪婪或者 wake-sleep 的方法训练, 当时模型的效果其实并没有那么好。但值得关注的是, 正是基于 RBM 模型, Hinton 等开始设计深度框架, 因此这也可以看做是深度学习的一个开端。Auto-Encoder 也是 20 世纪 80 年代 Hinton 提出的模型, 后来随着计算能力的进步也重新登上舞台。约书亚·本吉奥 (Yoshua Bengio) 等又提出了 Denoise Auto-Encoder, 主要针对数据中可能存在的噪音问题。麦克斯·威林 (Max Welling, 也是变分和概率图模型的高手) 等后来使用神经网络训练一个有一层隐变量的图模型, 由于使用了变分推断, 并且和 Auto-Encoder 有点像, 被称为 Variational Auto-Encoder。此模型中可以通过隐变量的分布采样, 经过后面的 Decoder 网络直接生成样本。生成对抗模型 GAN(generative adversarial network) 是 2014 年提出的非常受关注的模型, 它是一个通过判别器和生成器进行对抗训练的生成模型, 这个思路很有特色, 模型直接使用神经网络 G 隐式建模样本整体的概率分布, 每次运行相当于从分布中采样。随之而来引发了大量的研究, 包括: DCGAN 是一个相当好的卷积神经网络实现, WGAN 是通过维尔斯特拉斯距离替换原来的 JS 散度来度量分布之间的相似性的工作, 使得训练稳定。PGGAN 逐层增大网络, 生成逼真的人脸。

第 3 条发展脉络 (橙黄色区域) 是序列模型。序列模型不是因为深度学习才有的, 而是很早以前就有相关研究, 例如有向图模型中的隐马尔可夫 HMM 以及无向图模型中的条件随机场模型 CRF 都是非常成功的序列模型。即使在神经网络模型中, 1982 年就提出了 Hopfield Network, 即在

神经网络中加入了递归网络的思想。1997 年于尔根·施密德胡伯 (Jürgen Schmidhuber) 发明了长短期记忆模型 LSTM(long-short term memory), 这是一个里程碑式的工作。当然, 真正让序列神经网络模型得到广泛关注的还是 2013 年 Hinton 组使用 RNN 做语音识别的工作, 识别效果比传统方法显著提高。在文本分析方面, 另一个图灵奖获得者 Yoshua Bengio 在 SVM 很受关注的时期提出了一种基于神经网络的语言模型 (当然当时机器学习还是 SVM 和 CRF 的天下), 后来 Google 提出的 word2vec (2013) 也有一些反向传播的思想, 最重要的是给出了一个非常高效的实现, 从而引发了这方面研究的热潮。后来, 在机器翻译等任务上逐渐出现了以 RNN 为基础的 seq2seq 模型, 通过一个 Encoder 把一句话的语义信息压缩成向量再通过 Decoder 转换输出得到这句话的翻译结果, 后来该方法被扩展到和注意力机制 (Attention) 相结合, 也大大扩展了模型的表示能力和实际效果。再后来, 大家发现使用以字符为单位的 CNN 模型在很多语言任务也有不俗的表现, 而且时空消耗更少。Self-attention 实际上就是采取一种结构去同时考虑同一序列局部和全局的信息, Google 有一篇很有名的文章“Attention is All You Need”把基于 Attention 的序列神经模型推向高潮。当然 2019 年 ACL 上同样有另一篇文章给这一研究稍微降了降温。2018 年底 Google 提出 BERT 模型, 将 GPT 中的单向语言模型拓展为双向语言模型 (masked language model), 并在预训练中引入了 sentence prediction 任务。BERT 模型在 11 个任务中取得了最好的效果, 是深度学习在 NLP 领域又一个里程碑式的工作。BERT 自从在 arXiv 上发表以来获得了研究界和工业界的极大关注, 仿佛打开了深度学习在 NLP 应用的潘多拉魔盒。随后涌现了一大批类似于“BERT”的预训练 (pre-trained) 模型, 有引入 BERT 中双向上下文信息的广义自回归模型 XLNet, 也有改进 BERT 训练方式和目标的 RoBERTa 和 SpanBERT, 还有结合多任务以及知识蒸馏 (knowledge distillation) 强化 BERT 的 MT-DNN 等, 这些被大家称为 BERTology。

第 4 条发展脉络 (粉色区域) 是增强学习。这个领域最出名的当属 Deep Mind, 图中标出的大卫·席尔瓦 (David Silver) 博士是一直研究 RL 的高管。Q-learning 是很有名的传统 RL 算法, Deep Q-learning 将原来的 Q 值表用神经网络代替, 做了一个打砖块的任务。后来又被应用在许多游戏场景中, 其成果发表在 Nature 上。Double Dueling 对这个思路进行了一些扩展, 主要是 Q-Learning 的权重更新时序上。DeepMind 的其他工作如 DDPG、A3C 也非常有名, 它们是基于 Policy Gradient 和神经网络结合的变种。大家都熟知的 AlphaGo, 里面其实既用了 RL 的方法也有传统的蒙特卡洛搜索技巧。Deep Mind 后来提出了一个用 AlphaGo 框架、但通过主学习来玩不同 (棋类) 游戏的新算法 Alpha Zero。

总体来看, 在这个深度学习算法引领的人工智能浪潮中, 以神经网络为核心的机器学习算法取得了快速的进展。那么未来十年, AI 将何去何从?

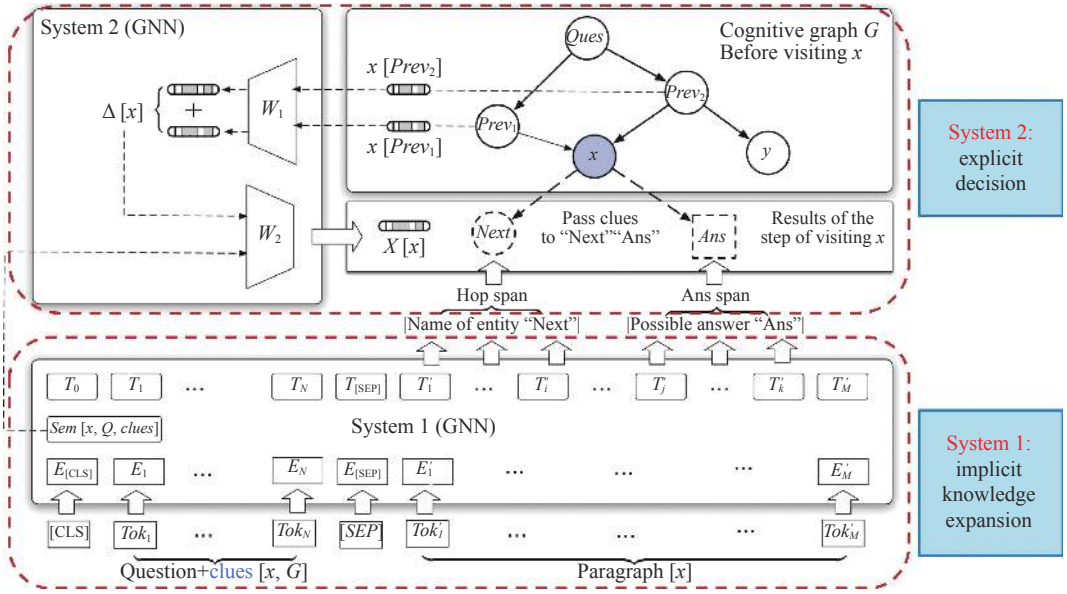
3　展望未来十年

这里, 我想首先引用张钹院士提出来的第三代人工智能的理论体系。2015 年, 张钹院士提出第三代人工智能体系的雏形。2017 年, 美国国防高级研究计划局 (DARPA) 发起 XAI 项目, 核心思想是从可解释的机器学习系统、人机交互技术以及可解释的心理学理论 3 个方面, 全面开展可解释性 AI 系统的研究。2018 年底, 张钹院士正式公开提出第三代人工智能的理论框架体系, 核心思想为: 1) 建立可解释、鲁棒性的人工智能理论和方法; 2) 发展安全、可靠、可信及可扩展的人工智能技术; 3) 推动人工智能创新应用。其中具体实施的路线图包括: 1) 与脑科学融合, 发展脑启发的人工智能理论; 2) 数据与知识融合的人工智能理论与方法。这标志着人工智能从感知时代逐渐进入认知时代。

Yoshua Bengio 在 NeuIPS 2019 上的报告“From System a1 Deep Learning to System 2 Deep Learning”讨论了深度学习发展的方向, 进一步肯定这一畅想。Bengio 肯定了人工智能已经在“听、说、看”等感知智能领域达到甚至超越人类水准, 但在需要外部知识、逻辑推理以及领域迁移的认知领域还处于初级阶段。认知智能将从认知心理学、脑科学中汲取灵感, 并结合知识图谱、因果推理等技术, 建立知识表示、推理的有效机制, 实现从感知智能到认知智能的关键突破。Bengio 介绍了人的认知系统包含两个子系统 (这是认知理论中大家共识的观点): System 1(子系统 1) 是直觉系统, 主要负责快速、无意识、非语言的认知, 比如当人被问到一个问题的时候, 可能下意识地或者说习惯性地回答, 这就属于 System 1 的范畴。Bengio 认为目前深度学习

主要就在做 System 1 的事情；System 2(子系统 2) 是逻辑分析系统，是有意识的、带逻辑、规划、推理以及可以语言表达的系统。人在通过 System 2 处理问题的时候，往往要收集相关数据、进行逻辑分析和推理，最终做出决策。目前的绝大多数人工智能系统都还没能实现 System 2，Bengio 提出这正是未来深度学习需要着重考虑的。当然 Bengio 也提到多智能体角度来实现 AI、以及从计算机角度需要考虑的问题，比如更好的模型和知识搜索。对于如何用深度学习来实现 System 2，Bengio 提到对于计算机来说，最关键就是处理数据分布中的变化。对于 System 2 来说，基本的要素包括：注意力和意识。注意力 (attention) 的实现在深度学习模型中已经有大量的研究和探讨，比如 GAT(图注意力机制) 等，意识这部分是比较难的。

笔者有幸在同一时期和 Bengio 课题组并行做了类似的认知工作，我们从 2018 年初开始研究认知计算，我们给他取了一个名字，叫做认知图谱 (cognitive graph)，下图展示了我们提出的基于双通道处理理论的认知系统框架。System 1 我们采用了 BERT 来实现，通过预训练可以得到每个实体的表示，在表示的基础上可以实现知识扩展；System 2 则采用图神经网络，这是因为 System 1 扩展的信息都传递给 System 2，使得 System 2 可以基于多方面的信息做决策。这个方法在推理方面还有所欠缺，但在多跳问题回答任务上取得了不错的结果，后续在推理方面可能还可以做很多有意思的扩展。相关论文发表在 ACL 2019 上。



这是一个总体的思路，要真正实现知识和推理，其实还需要万亿级的常识知识库支持，来支撑深度学习的计算，这样才能真正实现未来的人工智能。这一次人工智能浪潮也许到终点还是没有推理能力，没有可解释能力。而下一波人工智能浪潮的兴起，就是实现具有推理、具有可解释性、具有认知的人工智能，这是人工智能下一个 10 年要发展、也一定会发展的一个重要方向。

作者简介：

唐杰，教授，担任 IEEE T. on Big Data、AI OPEN 主编以及 WWW'21、CIKM'16、WSDM'15 的 PC Chair，主要研究方向为认知图谱、数据挖掘、社交网络和机器学习。主持研发了研究者社会网络挖掘系统 AMiner，杰出青年基金获得者，获北京市科技进步一等奖、人工智能学会一等奖、KDD 杰出贡献奖。发表学术论文 300 余篇，引用 15 000 余次。

中文引用格式：唐杰. 浅谈人工智能的下一个十年 [J]. 智能系统学报, 2020, 15(1): 187–192.

英文引用格式：TANG jie. On the next decade of artificial intelligence[J]. CAAI transactions on intelligent systems, 2020, 15(1): 187–192.