

Report on
Artificial Intelligence Development

人工智能 发展报告 2011-2020

清华-中国工程院知识智能联合研究中心
清华大学人工智能研究院知识智能研究中心
中国人工智能学会

核心发现

- 1.** 科技情报大数据挖掘与服务系统平台 **AMiner** 评选出过去十年**十大 AI 研究热点**，分别为：深度神经网络、特征抽取、图像分类、目标检测、语义分割、表示学习、生成对抗网络、语义网络、协同过滤和机器翻译。
- 2.** 过去十年有 5 位人工智能领域学者获得图灵奖殊荣，分别是在计算理论做出重要贡献的 Leslie Valiant，通过概率和因果推理促进人工智能发展的 Judea Pearl，以及在深度神经网络概念和工程上做出重大突破的 Yoshua Bengio、Geoffrey Hinton 和 Yann LeCun。
- 3.** 过去十年人工智能国际顶刊顶会的最佳论文奖项较多授予的领域是**计算理论、安全与隐私和机器学习**。
- 4.** 人工智能不同子领域论文的最高引用量量级跨度很大。最高引用量前十的论文研究主题以机器学习为首，其次是计算机视觉领域研究，二者的论文引用率均达到 25 万次以上。
- 5.** 中国在自然语言处理、芯片技术、机器学习、信息检索与挖掘等 10 多个 AI 子领域的科研产出水平都紧随美国之后，居于世界

前列；在多媒体与物联网领域的论文产出量超过美国，居于全球第一；而在人机交互、知识工程、机器人、计算机图形、计算理论领域，中国还需努力追赶。

6. 全球范围内，美国 AI 高层次学者的数量最多，占比 62.2%，是第二位国家（中国）AI 高层次学者数量的 6 倍以上。

7. **清华大学**是唯一入选全球人工智能领域高层次学者数量 TOP10 的中国机构。

8. 国内人工智能领域高层次人才主要分布在京津冀、长三角和珠三角地区。四成以上的 AI 高层次人才集聚在北京。

9. 过去十年全球人工智能专利申请量 521264。中国专利申请量为 389571，位居世界第一，占全球总量的 74.7%，是排名第二的美国专利申请量的 8.2 倍。

10. 人工智能未来重点发展的技术方向包括：强化学习、神经形态硬件、知识图谱、智能机器人、可解释性 AI、数字伦理、知识指导的自然语言处理等。

目录

1	前言.....	1
2	报告相关说明.....	3
	2.1 研究方法.....	3
	2.2 研究领域划分.....	4
	2.3 研究范围：国际顶级期刊和会议.....	6
3	人工智能发展历程.....	7
	3.1 人工智能的三次发展浪潮.....	7
	3.2 人工智能理论三个阶段.....	10
	3.3 人工智能技术发展脉络.....	12
	3.4 人工智能技术成熟度.....	14
4	人工智能过去十年研究进展.....	21
	4.1 全球高水平论文发表情况.....	21
	4.2 获得图灵奖的人工智能技术.....	23
	4.3 媒体评选出的重大人工智能技术.....	28
	4.4 国际顶会顶刊最佳论文授予领域分析.....	32
	4.5 国际顶会顶刊领域高影响力论文分析.....	36
	4.6 AMiner：过去十年十大人工智能研究热点.....	43
	1. 深度神经网络.....	44
	2. 特征抽取.....	45
	3. 图像分类.....	45
	4. 目标检测.....	46
	5. 语义分割.....	46
	6. 表示学习.....	47
	7. 生成对抗网络.....	47
	8. 语义网络.....	48
	9. 协同过滤.....	48
	10. 机器翻译.....	49
5	人工智能子领域研究现状与趋势.....	50

5.1 机器学习.....	50
5.1.1 机器学习概念.....	50
5.1.2 机器学习过去十年主要进展.....	52
5.1.3 机器学习经典算法.....	52
5.1.4 深度学习主要模型.....	67
5.1.5 机器学习技术研究趋势.....	83
5.2 自然语言处理.....	84
5.2.1 自然语言处理概念.....	84
5.2.2 自然语言处理过去十年主要进展.....	86
5.2.3 自然语言处理技术研究趋势.....	89
5.3 知识工程.....	91
5.3.1 知识工程概念.....	91
5.3.2 知识工程过去十年主要进展.....	92
5.3.3 知识工程领域技术研究趋势.....	96
5.4 信息检索与推荐.....	98
5.4.1 信息检索与推荐概念.....	98
5.4.2 信息检索与推荐技术过去十年主要进展.....	99
5.4.3 信息检索与推荐技术研究趋势.....	102
5.5 计算机视觉.....	104
5.5.1 计算机视觉概念.....	104
5.5.2 计算机视觉过去十年主要进展.....	105
5.5.3 计算机视觉研究趋势.....	107
5.6 语音识别.....	109
5.6.1 语音识别概念.....	109
5.6.2 语音识别过去十年主要进展.....	110
5.6.3 语音识别研究趋势.....	112
5.7 机器人.....	114
5.7.1 机器人概念.....	114
5.7.2 机器人过去十年主要进展.....	115

5.7.3 机器人研究趋势.....	117
5.8 数据挖掘.....	119
5.8.1 数据挖掘概念.....	119
5.8.2 数据挖掘过去十年主要进展.....	119
5.8.3 数据挖掘研究趋势.....	121
5.9 人机交互.....	122
5.9.1 人机交互概念.....	122
5.9.2 人机交互过去十年主要进展.....	122
5.9.3 人机交互研究趋势.....	125
5.10 可视化.....	127
5.10.1 可视化技术概念.....	127
5.10.2 可视化技术过去十年主要进展.....	128
5.10.3 可视化技术研究趋势.....	131
5.11 其他人工智能外延技术.....	133
5.11.1 经典 AI	133
5.11.2 安全与隐私.....	134
5.11.3 芯片技术.....	135
5.11.4 数据库.....	138
5.11.5 计算机图形.....	143
5.11.6 多媒体.....	148
5.11.7 计算理论.....	154
5.11.8 计算机网络.....	155
5.11.9 计算机系统.....	156
5.11.10 物联网.....	158
6 人工智能领域高层次人才现状.....	160
6.1 全球 AI 领域高层次人才分析	160
6.1.1 国家地区分布.....	161
6.1.1 高层次学者数量 TOP10 国家	162
6.1.3 所在机构分布.....	162

6.2	中国 AI 领域高层次人才分析	163
6.2.1	省市地区分布.....	163
6.2.2	高层次学者数量 TOP10 国内城市	164
6.2.3	所在机构分布.....	165
6.3	中国 AI 高层次人才的国际合作论文分析	166
6.4	代表性高层次 AI 学者	167
6.5	中国 AI 高层次人才培养	170
7	人工智能专利分析.....	172
7.1	全球 AI 专利分析	172
7.1.1	专利申请趋势.....	172
7.1.2	专利申请国家/地区排名	173
7.1.3	申请人排名.....	173
7.2	中国 AI 专利分析	175
7.2.1	国内 AI 专利申请趋势	175
7.2.2	专利申请量省市排名.....	175
7.2.3	申请人排名.....	176
8	人工智能技术应用：赋能其他行业发展.....	177
8.1	智慧医疗.....	177
	（1）医疗影像智能诊断.....	178
	（2）新药研发.....	180
	（3）基因测序.....	181
8.2	智慧金融.....	183
	（1）智能风控.....	184
	（2）智慧银行.....	184
	（3）智慧投顾.....	186
8.3	智慧城市.....	188
	（1）智能政务.....	189
	（2）智能基础设施系统.....	191
	（3）智能交通.....	192

8.4 智慧教育.....	196
(1) 智慧校园.....	197
(2) 智慧课堂.....	197
(3) 智适应教学.....	199
8.5 智能制造.....	200
(1) 智能工厂.....	201
(2) 智能物流.....	204
(3) 智能系统.....	205
9 人工智能发展的机遇和挑战.....	207
9.1 人工智能未来发展机遇.....	207
9.1.1 全球主要经济体 AI 发展支持政策	207
9.1.2 中国 AI 发展支持政策	212
9.2 人工智能未来技术研究方向.....	214
9.3 人工智能发展面临的问题.....	218
9.3.1 安全与伦理挑战.....	218
9.3.2 国家之间技术限制.....	219
附录 1 人工智能领域国际顶级期刊和会议.....	223
附录 2 全国 70 所高校人工智能学院、研究院、研究所名单.....	224

1 前言

人工智能在过去十年中从实验室走向产业化生产，重塑传统行业模式、引领未来的价值已经凸显，并为全球经济和社会活动做出了不容忽视的贡献。

当前，人工智能已经迎来其发展史上的第三次浪潮。人工智能理论和技术取得了飞速发展，在语音识别、文本识别、视频识别等感知领域取得了突破，达到或超过人类水准，成为引领新一轮科技革命和产业变革的战略性技术。人工智能的应用领域也快速向多方向发展，出现在与人们日常生活息息相关的越来越多的场景中。

全球经济体都高度重视人工智能领域的科学研究和产业发展，不仅意识到人工智能是各国新一轮经济竞争的核心驱动力，而且也都意识到能够推动技术突破和创造性应用的高端人才对人工智能发展的至关重要作用。在中国，人工智能相关的学院、研究院、企业组织在 2013、2014、2015 年左右如雨后春笋般接连出现。政府、高校、企业各方都积极投入到人工智能的发展和应用中。

对于人工智能的未来，人类有很大期待，也有很多争论。有人认为今后十年人工智能将进入人类生活中每一个领域，为各行业赋能，将深刻改变世界的生产和生活方式，带来更多商业利益和社会效益；也有人对人工智能的未来发展表示担忧，认为人工智能将是人类生存所面对的最大威胁。

未来十年，人工智能技术将实现从感知智能到认知智能的新突破。中国科学院张钹院士、中国工程院李德毅院士和王恩东院士等业界人士指出，人工智能已经历计算智能、感知智能，将迈入认知智能的技术阶段。中科院院士、清华大学人工智能研究院院长、计算机系教授张钹多次指出，第三代人工智能是实现可解释的、鲁棒的、可信安全的智能系统，依靠知识、数据、算法和算力四个要素，将实现从不带认知的人工智能转变为带认知的人工智能。美国国防部 DARPA 提出第三代人工智能是能够实现情景自适应的并且具有感知、推理、规划和学习能力的智能系统^[1]。数据驱动的方法如何利用知识表示和推理获得智能系统的可解释和鲁棒性是当前人工智能的发展趋势，也是面临的重要挑战。图灵奖获得者

^[1] Gunning, D., & Yeh, P. Z. (2016). Introduction to the Special Issue on Innovative Applications of Artificial Intelligence 2015. *AI Magazine*, 37(2), 5-6.
<https://doi.org/10.1609/aimag.v37i2.2624>

Bengio^[2]也指出未来深度学习的研究将发展到系统 2（有意识地完成）的深度学习，实现系统 2 的逻辑推理功能。

本报告旨在展现人工智能发展所经历的重大科研进展、成果产出、技术影响力、人才储备，并且从人工智能技术发展的角度，阐述人工智能取得的重要成果，尤其大数据时代人工智能最具代表性的内容；介绍了全球主要国家人工智能战略支持政策和人才储备，尤其是各子领域高层次人才现状；最后深入探讨了人工智能的未来发展蓝图，理论、技术和应用方面的重大变化与挑战，以及如何赋能其他产业发展等议题。

^[2] Yoshua Bengio. (2019). Towards Compositional Understanding of the World by Agent-Based Deep Learning [Invited talk at the NeurIPS'2019 workshop on Context and Compositionality in Biological and Artificial Neural Networks], Vancouver, BC.

2 报告相关说明

2.1 研究方法

本报告综合采用了定量和定性相结合的研究方法。报告中的相关科技情报技术和人才分析，主要借助科技情报大数据挖掘与服务系统平台 AMiner，利用大数据分析和挖掘技术，采用了 2011-2020 年期间人工智能领域的顶级期刊和会议（详见附录，共计 44 个）所收录的全部论文数据和专利数据。

1.人才分析法

通过挖掘和分析论文数据，获取论文学者信息，通过命名排歧和信息抽取等大数据分析挖掘和人工智能技术，对人工智能领域的专家学者进行学者画像，构建人工智能领域人才智库，然后根据学者画像信息进行领域人才相关分析。此外，还抽取论文作者的供职机构和国家信息，对机构的研究者和论文数量进行统计。

本报告中，对人工智能的人才分析包括作者所在机构，所属国家和城市分析等多个维度。AMiner 按照发表在人工智能领域或子领域的论文数量进行排序，得到全领域和子领域中排名 TOP10 的国家、机构以及国内城市名单。

此外，本报告还针对高层次学者进行了专门分析。人工智能领域**高层次学者**的判定原则是指被评选进入 AI2000 人工智能全球最具影响力学者榜单（以下称为 AI 2000）的学者。AI 2000 涵盖人工智能学科 20 个子领域，具体评选规则与遴选方法是为每个子领域每年选出 10 名获奖者；每年遴选时，参考过去 10 年该领域最有影响力的期刊和会议^[3]所发表论文的引用情况，排名前 10 的学者当选该领域当年“AI2000 最具影响力学者奖”，排名前 100 的其他学者获“AI2000 最具影响力学者提名奖”。榜单通过 AMiner 系统中所收录的数据采用智能算法自动化生成，确保了榜单的客观、公平、公正、公开。榜单采用的引用数据来源于 Google Scholar。有关 AI 各子领域更多的影响力学者信息，请参见详情页 <https://www.aminer.cn/ai2000>。

2.技术趋势分析法

借助 AMiner 平台，通过挖掘和分析期刊会议的论文数据，获取人工智能 20

^[3] 具体期刊和会议的名称列表请参见附录。

个子领域相关的论文中关键词，统计这些关键词的起止年份，再按照窗格统计词频，进行词频分析和对比；同时，构建该领域的知识图谱，再综合运用知识图谱、自然语言处理、可视化、文献计量学等技术手段，基于论文和学者数据，分析得到人工智能及其子领域的技术研究热点和发展趋势方向。

3. 技术研究热点榜单研究方法

人工智能技术研究热点榜单旨在通过 AMiner 学术数据库在全球研究范围内遴选每年度 10 项人工智能学科最有影响力的技术。

具体方法为参考人工智能领域在顶级期刊和会议过去 10 年所发表论文，从标题和摘要信息中抽取论文技术研究主题和所在子领域，并按照该技术主题研究的论文发表数量、论文引用量和该技术主题进入每个会议高引论文前十名的次数（即该技术的领域引用特征）进行综合评测排序，通过 AI 算法计算出该研究主题的 AMiner 影响力指数，以此获得人工智能领域研究热点的总榜单，评选出当年的“十大 AI 研究热点”。其中，领域的国际顶级期刊和会议由技术委员会专家确定，共计包括 44 个人工智能领域顶级会议，具体名称详见附录 1。

有关 AI 领域更多的榜单信息，请参见详情页 <https://aiopenindex.aminer.cn/>。

2.2 研究领域划分

人工智能既是计算机科学的一个分支，又是一个融合了多种学科的交叉学科，加上其最近几年的高速发展，内涵和外延也在不断的变化，新兴的子领域不断涌现，工业界和学术界并没有一个对人工智能的明确定义。

本报告综合参考了计算机领域较为公认的权威性机构关于学科的分类方法，这些机构包括：国际计算机学会（Association for Computing Machinery, ACM），中国计算机学会（China Computer Federation, CCF）和电气与电子工程师协会（Institute of Electrical and Electronics Engineers, IEEE），同时又融合了国内外专家学者的建议，选择了自然语言处理、芯片技术、机器学习、信息检索与推荐、人机交互、知识工程、语音识别、计算机视觉、机器人、数据挖掘、经典 AI、数据库、计算机图形、多媒体、可视化、安全与隐私、计算机网络、计算机系统、计算理论和物联网作为 20 个人工智能子领域进行研究，如图 2-1 所示。

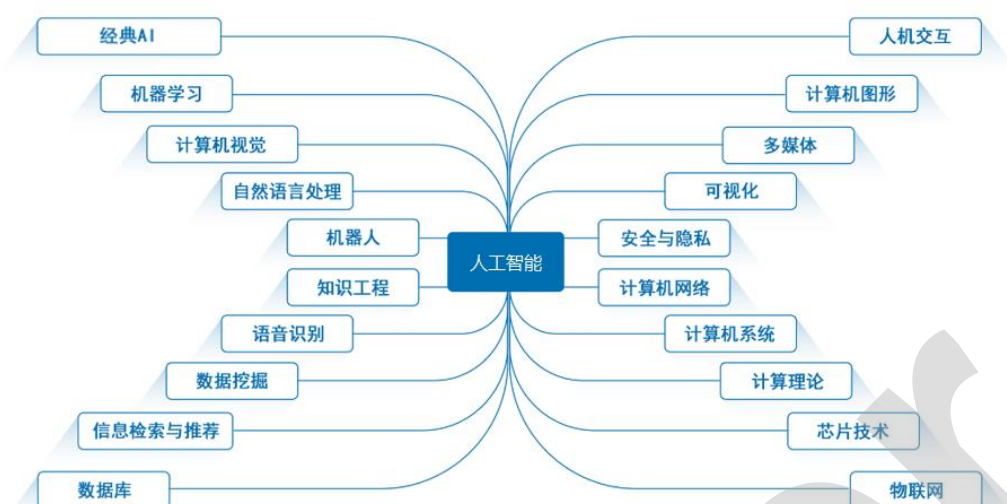


图 2-1 人工智能各子领域导图

根据技术的独立性特征，人工智能 20 个子领域可以进一步分为以机器学习、自然语言处理、知识工程、信息检索与推荐、计算机视觉、语音识别和机器人等为核心的七个人工智能关键技术，以及人机交互、数据挖掘、芯片技术、经典 AI、数据库、计算机图形、多媒体、可视化、安全与隐私、计算机网络、计算机系统、计算理论和物联网等十三个人工智能外延技术，如图 2-2 所示。

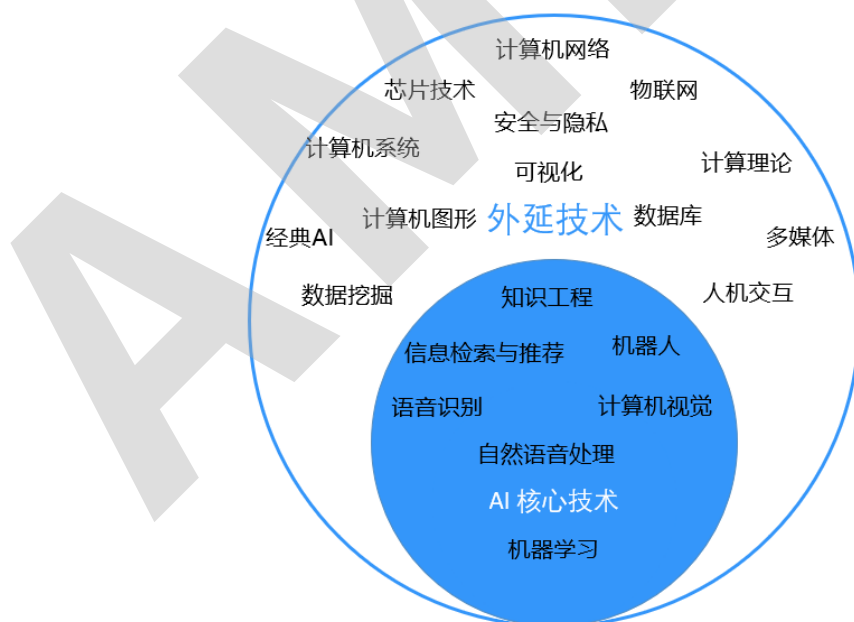


图 2-2 人工智能 20 个子领域的核心和外延划分

2.3 研究范围：国际顶级期刊和会议

本报告所采用的人工智能领域顶级期刊和会议，是根据《CCF 推荐国际学术期刊和会议目录》和 ACM 计算机学分类系统相关子领域的 A 类期刊和会议作为数据来源，并征求相关专家和团体意见，补充了新涌现学科的顶级期刊和会议，详细名单如附录 1 所示。

AMiner

3 人工智能发展历程

人工智能是研究人类智能行为规律（如学习、计算、推理、思考、规划等），构造具有一定智慧能力的人工系统，以完成往常需要人的智慧才能胜任的工作⁴。这是人工智能领域的先驱、麻省理工学院计算机科学家帕特里克·亨利·温斯顿（Patrick Henry Winston）教授在《人工智能》一书中对人工智能的定义。从其历史发展看，人工智能经历了三次发展浪潮；从技术脉络看，其经历了数据、算法、算力及应用场景；从理论发展看，人工智能经历了计算、感知和认知三个阶段。

3.1 人工智能的三次发展浪潮

从 1956 年“人工智能”概念在达特茅斯会议上首次被提出至今，人工智能发展已经历经 60 余年，经历了三次发展浪潮。当前全球人工智能正处于第三次发展浪潮之中。

第一次浪潮（1956–1974 年）：1956 年达特茅斯会议上，“人工智能”的概念提出，同时出现了最初的成就和最早的一批研究者，被广泛认为是 AI 诞生的标志，随之掀起了人工智能的第一次发展浪潮。该时期的核心是让机器具备逻辑推理能力，开发出计算机可以解决代数应用题、证明几何定理、学习和使用英语的程序，并且研发出第一款感知神经网络软件和聊天软件。**第二次浪潮（1980–1987 年）：**这一时期内，解决特定领域问题的“专家系统”AI 程序开始为全世界的公司所采纳，AI 变得实用起来，知识库系统和知识工程成为了 80 年代 AI 研究的主要方向。Hopfield 神经网络和反向传播（BP）算法被提出。**第三次浪潮（1993–2011 年）：**这一时期，计算性能上的基础性障碍已被逐渐克服，里程碑事件有 1997 年的深蓝战胜国际象棋世界冠军、2016 年和 2017 年的 AlphaGo 击败围棋职业选手等。同时，“智能代理”范式被广泛接受，AI 技术发展超越了研究人类智能的范畴；AI 与数学、经济学等其他学科展开更高层次的合作。2006

⁴ Owen T. Artificial Intelligence by Patrick Henry Winston (second edition) Addison-Wesley Publishing Company, Massachusetts, USA, July 1984.

年以来,深度学习理论的突破更是带动了人工智能第三次浪潮。这一阶段互联网、云计算、大数据、芯片等新兴技术为人工智能各项技术的发展提供了充足的数据支持和算力支撑,以“人工智能+”为代表的业务创新模式也随着人工智能技术和产业的发展日趋成熟,这将极大优化社会的生产力,并对现有的产业结构产生深远的影响。

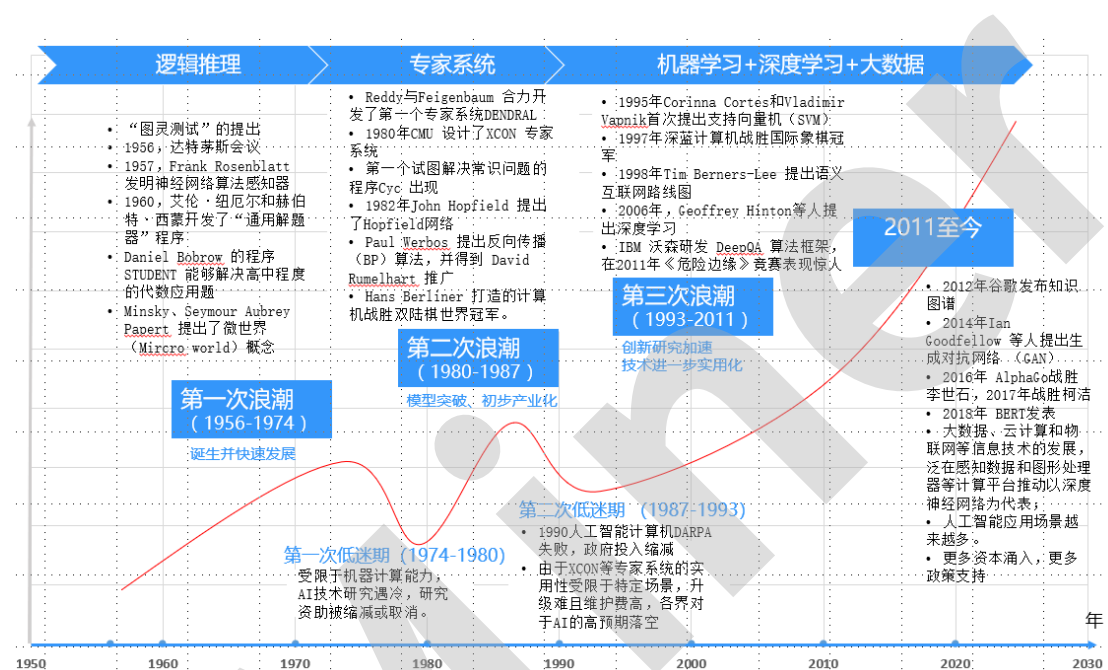


图 3-1 人工智能的三次发展浪潮

具体而言,随着克劳德·香农(Claude Shannon)在1950年提出“计算机博弈”(其中主要思路在“深蓝”和AlphaGo中还能看到),以及阿兰·图灵(Alan Turing)在1954年提出“图灵测试”,人工智能这一概念开始进入人们的视野。1956年夏季,在美国达特茅斯会议上,以约翰·麦卡锡(John McCarthy)、马文·闵斯基(Marvin Minsky)、内森·罗切斯特(Nathan Rochester)和克劳德·香农(Claude Shannon)等为首的年轻科学家汇聚一堂,共同研究和探讨用机器模拟智能的一系列有关问题,并首次提出了“人工智能”这一术语,它标志着“人工智能”这门新兴学科的正式诞生。

到了20世纪60年代,人工智能出现第一波高潮,发展出了自然语言处理和人机对话技术。其中的代表性事件是丹尼尔·博布罗(Daniel Bobrow)在1964年发表的“Natural language input for a computer problem solving system”,以及约

瑟夫·维森鲍姆(Joseph Weizenbaum)在 1966 年发表的“ELIZA-a computer program for the study of natural language communication between man and machine”。艾伦·纽厄尔和赫伯特·西蒙试图通过通用解题器 (General Problem Solver) 程序, 将搜索式推理推广到一般情形。赫伯特·吉宁特 (Herbert Gelernter) 的几何定理证明机 (1958) 也是基于搜索算法证明几何与代数问题的程序。60 年代后期, 麻省理工大学 AI 实验室的马文·闵斯基和西摩尔·派普特提出“微世界”思想, 指出在成熟的学科中往往使用简化模型帮助基本原则的理解。人工智能发展初期的突破性进展大大提升了人们对人工智能的期望, 很多人因此将人工智能神话, 认为它能够解决已有科技无法解决的许多问题, 但接二连三的失败和预期目标的落空使人工智能的发展走入低谷。

1968 年, 爱德华·费根鲍姆(Edward Feigenbaum)提出首个专家系统 DENDRAL 的时候对知识库给出了初步的定义, 其中隐含了第二波人工智能浪潮兴起的契机。经过一代人的努力之后, 20 世纪 70 年代出现的专家系统模拟人类专家的知识和经验解决特定领域的问题, 成效显著, 推动人工智能走入应用发展的新高潮。经过实践应用, 人们发现专家系统存在应用领域狭窄、缺乏常识性知识、知识获取困难、推理方法单一、缺乏分布式功能和难以与现有数据库兼容等问题, 因此一度将人工智能打入冷宫, 相关科研经费大幅度缩减, 人工智能相关研究也步入了长达 10 年的低迷期。

20 世纪 80 年代, 人工智能进入了第二波浪潮。1980 年德鲁·麦狄蒙 (Drew McDermott) 和乔恩·多伊尔 (Jon Doyle) 提出非单调逻辑, 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) 的推动下快速发展, 成为人工智能一个重要的发展分支。这其中格瑞·特索罗 (Gerry Tesauro) 等人打造的自我学习双陆棋程序为后来强化学习的发展奠定了基础。1986 年戴维·鲁梅哈特 (David Rumelhart), 杰弗里·辛顿 (Geoffrey Hinton) 等人推广了由保罗·韦尔博斯 (Paul Werbos) 提出反向传播算法 (BP 算法), 使得大规模神经网络训练成为可能。

20 世纪 90 年代至 2010 年, 互联网技术的发展和高性能计算机的出现, 加速了人工智能的创新研究, 人们渐渐使用人工智能算法来解决数据采集和处理中

的很多问题，促使人工智能技术进一步走向实用化。这一阶段，AI 出现了两个很重要的发展：一是蒂姆·伯纳斯·李（Tim Berners-Lee）在 1998 年提出的语义互联网路线图，即以语义为基础的知识网或知识表达。后来又出现了 OWL 语言和其他一些相关知识描述语言。二是杰弗里·辛顿（Geoffrey Hinton）等人在 2006 年提出的深度学习，这标志着第三次人工智能浪潮的兴起。在这次浪潮中，很多企业参与其中，如塞巴斯蒂安·特（Sebastian Thrun）在谷歌主导下推出的自动驾驶汽车、Google DeepMind 开发的 AlphaGo 战胜人类围棋职业选手、IBM 的沃森（Watson）在《危险边缘》（Jeopardy）中获得冠军、苹果推出的自然语言问答工具 Siri 等。

2011 年至今，大数据、云计算、互联网和物联网等信息技术的发展，泛在感知数据和图形处理器等计算平台推动以深度神经网络为代表的人工智能技术飞速发展，使得人工智能出现在越来越多的场景中，成为了与人们日常生活息息相关的一项技术。2013 年 4 月，《麻省理工学院技术评论》杂志将深度学习列为 2013 年十大突破性技术之首。深度学习使机器具有能够模仿视听和思考等人类活动，解决了很多复杂的模式识别难题，在搜索技术、数据挖掘、机器学习、机器翻译、自然语言处理、多媒体学习、语音、推荐和个性化技术以及其他相关领域都取得了很多成果，使得人工智能相关技术取得了巨大进步。

3.2 人工智能理论三个阶段

关于人工智能理论阶段，存在两种划分方法。

第一种划分方法是计算、感知、认知三个阶段。中国科学院张钹院士、中国工程院李德毅院士和王恩东院士等业界人士认为，人工智能从技术阶段上主要分为**运算智能、感知智能和认知智能三个层次**。运算智能，即快速计算和记忆存储能力；感知智能，即视觉、听觉、触觉等感知能力；认知智能就是具有推理、可解释性的能力，也是人工智能的高级阶段。

科大讯飞高级副总裁、研究院院长胡国平在 AI WORLD 2018 峰会上提到：计算智能、感知智能和认知智能，是探索人工智能道路上的三大台阶。计算智能是让机器能存会算；感知智能是让机器能听会说、能看会认；认知智能是解决机

器能理解会思考的问题^[5]。

张钹院士曾在 2015 年提出第三代人工智能体系的雏形，并于 2018 年底正式公开了第三代人工智能的理论框架体系，其核心思想为：（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。也有学者称符号主义为第一代 AI、称连接主义为第二代 AI，将要发展的 AI 称为第三代 AI^[6]。张钹院士强调^[7]，发展第三代人工智能，需要依靠知识、数据、算法和算力四个要素，其中的关键问题是算法。

第二种划分方法是感知、认知、决策三个阶段。中国科学院大学人工智能学院徐波院长，根据人工智能解决问题的不同阶段，将其归纳为：感知智能、认知智能和决策智能三个阶段。其中，感知智能和认知智能的观点与第一种划分方法相同。徐波院长认为，随着机器认知水平越来越高，其“自主性”进一步增强，具备了应对更多样态势和功能的能力，人类无可避免地会将越来越多的问题交给机器决策。为了提高机器决策的准确度，需要加强复杂问题下，提升人机信任度，增强人类与智能系统交互协作智能的研究，即决策智能^[8]。

人工智能的发展经历了表示计算到感知智能两个阶段，下一个阶段的核心是认知。以上两种划分法对此存在共识，即目前发展处于第三代人工智能。

在感知计算方面，AlphaGo、无人驾驶、类人机器人、文本和图片之间的跨媒体计算等取得了快速发展。AlphaGo 是第一个击败人类职业围棋选手、第一个战胜围棋世界冠军的人工智能机器人，应用了如神经网络学习、深度学习、蒙特卡洛树搜索法等人工智能新技术，使其实力有了实质性的飞跃。无人驾驶汽车主要依靠车内的以计算机系统为主的智能驾驶仪和车载传感器，可以实现感知道路环境，自动规划行车路线并控制车辆到达预定目标。人工智能在语音识别、文本识别、视频识别等方面已经超越了人类，可以说 AI 在感知方面已经逐渐接近人

^[5] 搜狐 [EB/OL] https://www.sohu.com/a/256323105_100024677

^[6] 张钹，朱军，苏航，迈向第三代人工智能，中国科学：信息科学 50, 1281 (2020); doi: 10.1360/SSI-2020-0204

^[7] 张钹院士：基础科研不光要允许失败，还要经得起失败，“馊主意”也比没主意好，量子位公众号，2020 年 10 月 25 日，<https://mp.weixin.qq.com/s/W1Zel5E984q9-NUIz2MeYw>

^[8] 徐波：新一代人工智能正在崛起，2020-05-06，中国科学院 [EB/OL] http://www.ia.cas.cn/xwzx/ttxw/202005/t20200507_5576178.html

类的水平。

人工智能依靠机器学习和深度学习取得了快速进展,但存在依赖大规模标注数据进行监督训练的问题,要实现真正的类人智能(强人工智能),机器还需要掌握大量的常识性知识,以人的思维模式和知识结构来进行语言理解、视觉场景解析和决策分析。谷歌在 2012 年提出了知识图谱(Knowledge Graph)概念,提供了一种更好的组织、管理和理解互联网海量信息的能力,将互联网的信息表达成更接近于人类认知世界的形式。知识图谱的构建过程本质是让机器具备认知能力,可以理解世界和各种行业领域,是实现强人工智能的基石。

阿里巴巴达摩院发布“2020 十大科技趋势”报告中提到,人工智能已经在“听、说、看”等感知智能领域达到或超越了人类水准,但在需要外部知识、逻辑推理或者领域迁移的认知智能领域还处于初级阶段。

图灵奖获得者 Manuel Blum 夫妇在 2020 世界人工智能大会上提出**意识 AI**(意识智能)的思想,这是一个既经典又全新的概念和思路。核心理念就是构造一个新型的可用数学建模、可计算的机器认知/意识模型,如图 3-2 所示。

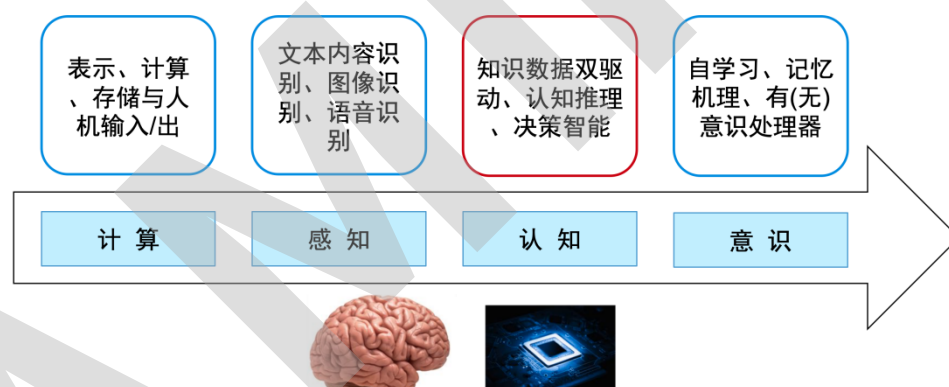


图 3-2 人工智能发展的几个阶段

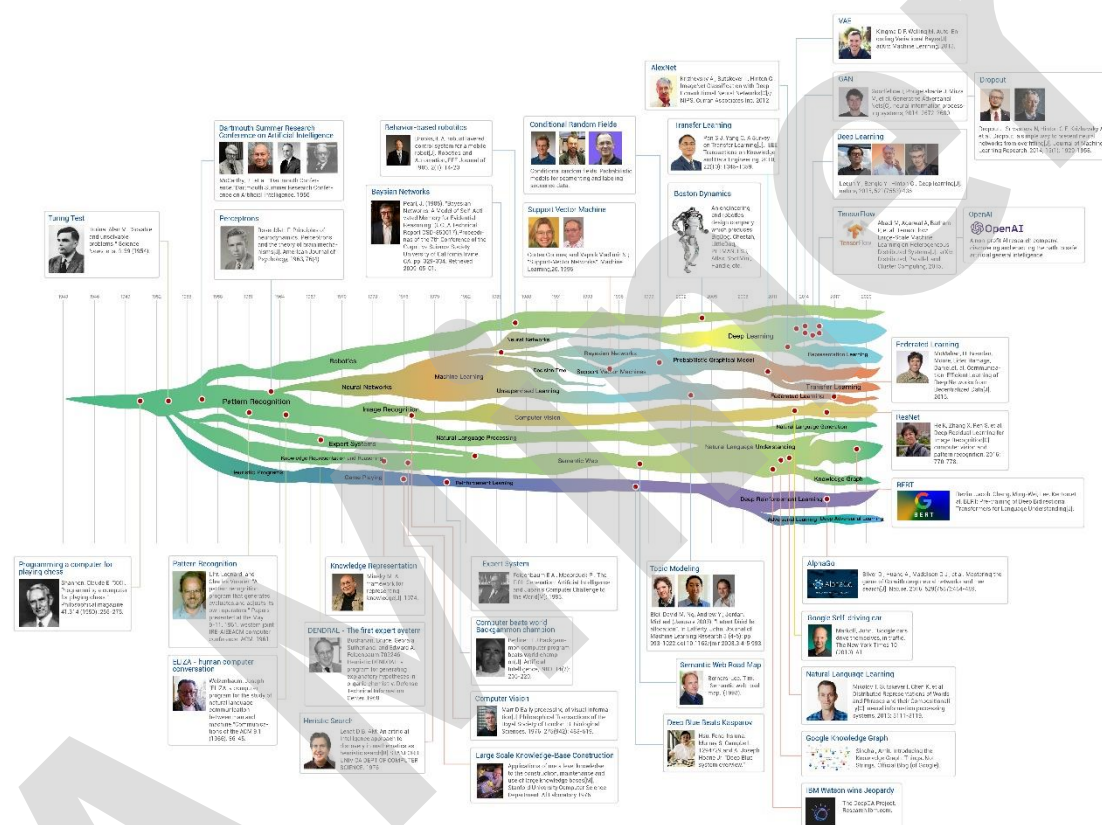
3.3 人工智能技术发展脉络

人工智能涉及计算机技术、控制论、信息论、语言学、神经生理学、心理学、数学、哲学等多学科领域的交叉与融合,其概念与内涵随着相关学科和应用领域的发展而持续变换。

中国人工智能学会理事长李德毅院士在 2018 中新人工智能高峰论坛上发表演讲认为,AI 的内涵包括四个方面,分别是脑认知基础、机器感知与模式识别、

自然语言处理与理解、知识工程。在这个核心之外，AI 的外延还包括机器人与智能系统。直面解决现实问题是新一轮人工智能的起点和落脚点，未来的人工智能工具需要体现人的认知力、创造力，成为人类认识世界、改造世界新的切入点，成为先进社会重要经济来源^[9]。

图 3-3 展示了人工智能自诞生以来在 60 多年的发展历史中所取得的一些标志性成果和技术^[10]，例如自然语言处理和人机对话技术、知识库、非单调逻辑、机器人系统、语义互联网路线图、深度学习等。



来源: <https://www.aminer.cn/ai-history>

图 3-3 人工智能技术发展脉络

中国新一代人工智能发展战略研究院执行院长龚克认为，人工智能有 4 个要素：算法、算力、数据、应用场景^[11]。

^[9] 院士李德毅：人工智能将会改变这四个重要行业，2018-06-01 来源：网易智能[EB/OL] <http://tech.163.com/18/0601/11/DJ7BMV9700098IEO.html>

^[10] 清华教授唐杰：人工智能的下个十年，2020-04-25 搜狐[EB/OL].

https://www.sohu.com/a/391117885_468714

^[11] 人工智能科技[EB/OL]. <http://www.aistudyblog.com/aimassage/20181110/15175.html>

人工智能领域应用最广泛的算法是机器学习和深度学习。从宏观来看，算法是人工智能的重要组成部分，而深度学习是近年来发展最快速的机器学习算法，因其在计算机视觉、自然语言处理等领域中的优异表现，大幅加快人工智能应用落地速度，催生了很多相关工具和平台，如百度飞桨深度学习开源框架、阿里巴巴的深度学习框架 X-Deep Learning、旷视人工智能计算平台 Brain++ 等。然而，机器学习和深度学习算法虽然在人工智能领域取得了显著成绩，但是受限于底层算法，使得人工智能技术目前已经触及天花板。张钹院士建议，人工智能迫切需要推动到新的阶段，有赖于与数学、脑科学等结合实现底层理论的突破^[12]。

人工智能的技术和产业发展只靠算法是不够的，需要加强以算力为核心的基础能力建设，比如智能体系架构和芯片。尤其是深度学习计算所需数据量巨大，对算力要求很高，在已经固化的硬件加速器上无法得到很好的支持，需要解决性能和灵活度之间的平衡问题。

人工智能的发展需要有大数据支撑来学习大量的知识和经验。大数据具有体量大、多样性、价值密度低、速度快等特点，在快速获得有价值信息方面，自然语言处理技术既带来了很大驱动力又面临着很多挑战。前微软公司全球执行副总裁沈向洋先生在公开演讲时说：“懂语言者得天下……下一个十年，人工智能的突破在自然语言的理解……人工智能对人类影响最为深刻的就是自然语言方面”^[13]。自然语言处理被视为解决人工智能完备（AI-complete）的核心问题之一。

人工智能的一个重要应用场景是智能信息检索与挖掘，通过打造具备自然语言对话、高精度知识融合、场景感知、互联网数据与个人数据高效整合等能力的智能信息助手，可以有效解决用户精准多样的信息需求与庞大互联网数据规模之间的鸿沟问题。

3.4 人工智能技术成熟度

世界上从事技术预测的较知名机构有 Gartner、美国兰德公司 RAND、日本科技政策所、亚太经合组织 APEC 技术预测中心等机构。技术成熟度（Technology

^[12] CSDN [EB/OL]

https://blog.csdn.net/cf2SudS8x8F0v/article/details/90555753?utm_medium=distribute.pc_relevant.none-task-blog-baidujs-8

^[13] 电子发烧友 [EB/OL] <http://www.elecfans.com/rengongzhineng/1209079.html>

Readiness Levels, TRLs) 是对技术成熟程度进行量化评价的一套系统化标准、方法和工具,它最早由美国航空航天局(NASA)提出。在这些技术预测机构之中,Gartner 发布的技术成熟度曲线(Hype Cycle)因模型较成熟,广泛被用来评估新科技的可见度,目前已成为是科技产业界技术预测的风向标。本报告将基于 Gartner 近年发布的关于人工智能的技术成熟度曲线,来分析人工智能领域新兴技术的成熟度发展变化情况。

Gartner 公司每年会根据其国际分析师预测把各种新科技的发展阶段及要达到成熟所需的时间绘制成一条“Gartner 新兴技术成熟度曲线”(The Gartner Hype Cycle for Emerging Technologies),有助于市场了解当下热点及未来趋势。在 Gartner 技术成熟度曲线中,技术预计成熟时间分为“不到 2 年”、“2 至 5 年”、“5 至 10 年”、“10 年以后”和“之前已被淘汰(Obsolete before Plateau)”五个时段;技术所处发展状态分为“萌芽期”(Innovation Trigger)、“期望膨胀期”(Peak of Inflated Expectations)、“幻灭期”(Trough of Disillusionment)、“复苏期”(Slope of Enlightenment)和“生产成熟期”(Plateau of Productivity)五个阶段。

基于 Gartner 技术成熟度曲线,人工智能领域新兴热点技术过去五年的发展状态变化如表 3-1 所示。由表可见,机器学习相关技术自 2015 年以来一直被寄予较高期望,2015 至 2019 年期间机器学习技术每年发展均一直维持在“期望膨胀期”、直到 2020 年才离开高期望的峰值并且开始进入“幻灭期”。2020 年通用人工智能和增强智能技术还处于“萌芽期”,而可解释的 AI 则度过“萌芽期”并开始提供了许多实际的实现方式,用于解释涉及各种问题和数据的复杂模型。此外,2020 年处于“期望膨胀期”的技术包括知识图谱、深度神经网络、决策职能、智能机器人和数字伦理;处于“幻灭期”的技术包括机器学习、认知计算、自然语言处理、计算机视觉、无人驾驶汽车。

表 3-1 人工智能子领域过去五年重点技术发展状态一览

AI 子领域	技术标签	2015	2016	2017	2018	2019	2020 年
机器学习	机器学习	∧	∧	∧	∧	∧	U
	深度学习	⊙	-	∧	-	-	-
	深度强化学习	-	⊙	⊙	-	-	-
	深度神经网络(深度学习)	-	∧	-	-	∧	∧

	深度神经网络 ASICs	-	-	⊙	∧	∧	∧
	强化学习	-	-	-	-	⊙	-
	量子计算	-	-	-	-	∧	-
	认知计算	⊙	⊙	∧	U	U	U
自然语言	自然语言处理	⊙	∧	∧	∧	U	U
	自然语言生成	⊙	⊙	⊙	⊙	-	-
	自然语言问答	∧	-	-	-	-	-
知识图谱	知识图谱	-	-	-	⊙	⊙	∧
计算机视觉	计算机视觉	-	∧	∧	U	U	U
语音识别	语音识别	⊕	⊕	⊕	⊕	⊕	-
其他智能	决策智能	-	-	-	-	⊙	∧
	可解释的 AI	-	-	-	-	⊙	-
	通用机器智能	⊙	⊙	-	-	-	-
	通用人工智能	-	-	⊙	⊙	⊙	⊙
	增强智能	-	-	-	-	⊙	⊙
	智能机器人	⊙	∧	-	∧	⊙	∧
	无人驾驶汽车	∧	∧	∧	U	U	U
	增强现实技术	U	U	U	U	-	-
	数字伦理	-	⊙	⊙	∧	∧	∧

注：⊙——萌芽期；∧——期望膨胀期；U——幻灭期；⊕——生产成熟期；——未被提及。资料来源：Gartner 新兴技术成熟度曲线。

此外，基于 Gartner 技术成熟度曲线，还可以得到人工智能领域新兴热点技术预计成熟时间的历年演变，具体信息如表 3-2 所示。值得注意的是，深度强化学习在 2018 年被归入深度神经网络（深度学习）¹⁴；而 Gartner 在不同年份对此技术称呼不同，先后使用了深度学习、深度神经网络、深度神经网络（深度学习）等名称。分析表 3-2 发现，这些人工智能领域的新兴技术具有以下发展特征。

1. 目前已成熟应用的 AI 技术主要是语音识别。

语音识别技术自 2015 年以来一直处于“生产成熟期”，预计实现时间“小于 2 年”。该技术目前主要有两个应用方向：一个方向是大词汇量连续语音识别系统，通常在计算机平台上实现的例如听写机、语音信息查询服务等应用；另外一个方向是使用专用芯片或专门硬件系统而实现的语音产品，如手机语音拨号、汽车设备的语音控制、智能玩具、个人数字助理、家电遥控、工业控制及医疗领域的语音命令和语音应答等。语音到文本的应用受到了算法和硬件进步的双重推动。

¹⁴ Gartner&中国信通院 2018AI 发展白皮书 p.43

<http://www.caict.ac.cn/kxyj/qwfb/bps/201809/P020180918696200669434.pdf>

表 3-2 人工智能子领域过去五年重点技术预期成熟时间（单位：年）

AI 子领域	技术标签	2015	2016	2017	2018	2019	2020 年
机器学习	机器学习	2-5	2-5	2-5	2-5	2-5	2-5
	深度学习	5-10	-	2-5	-	-	-
	深度强化学习	-	5-10	5-10	-	-	-
	深度神经网络（深度学习）	-	5-10	-	2-5	2-5	2-5
	深度神经网络 ASICs	-	-	2-5	2-5	2-5	2-5
	强化学习	-	-	-	-	5-10	-
	量子计算	-	-	-	-	> 10	-
	认知计算	5-10	5-10	5-10	5-10	5-10	之前已被淘汰
自然语言处理	自然语言处理	5-10	5-10	5-10	5-10	5-10	5-10
	自然语言生成	5-10	2-5	2-5	2-5	-	-
	自然语言问答	5-10	-	-	-	-	-
知识图谱	知识图谱	-	-	-	5-10	5-10	5-10
计算机视觉	计算机视觉	-	5-10	5-10	2-5	2-5	2-5
语音识别	语音识别	< 2	< 2	< 2	< 2	< 2	-
其他智能	决策智能	-	-	-	-	5-10	2-5
	可解释的 AI	-	-	-	-	5-10	-
	通用机器智能	> 10	> 10	-	-	-	-
	通用人工智能	-	-	> 10	> 10	> 10	> 10
	增强智能	-	-	-	-	2-5	2-5
	智能机器人	5-10	5-10	-	5-10	5-10	5-10
	无人驾驶汽车	5-10	> 10	> 10	> 10	> 10	> 10
	增强现实技术	5-10	5-10	5-10	5-10	-	-
	数字伦理	-	5-10	5-10	5-10	5-10	5-10

资料来源：Gartner 新兴技术成熟度曲线。

2. 下一代即将步入生产成熟的 AI 技术是机器学习、计算机视觉、深度神经网络及其专用芯片、决策智能和增强智能。这些技术的预计成熟时间是“2-5 年”。

机器学习技术从 2015 至 2020 年一直被认为是距离达到生产力成熟期是“2-5 年”。深度学习技术在 2015 年被认为处于萌芽期，2017 年被寄予厚望而跃升到“期望膨胀期”，其预计成熟时间从“5-10 年”缩短到“2-5 年”。

计算机视觉技术从 2016 年首次出现就以“期望膨胀期”状态出现在 Gartner 人工智能技术成熟度曲线中，预计成熟时间为“5-10 年”。在两年之后的 2018 年，计算机视觉的技术发展区间就转变为“幻灭期”、预计成熟时间缩短至“2-5 年”。

深度神经网络使计算机能够处理比以前复杂得多的数据，如视频、图像、语音和文本数据，从而支持人工智能的最新进展。基于神经网络的深度学习主要应用于图像、文本、语音等领域。互联网巨头在各自的业务范围内部署基于 DNN 的系统，例如亚马逊 Alexa 的语音到文本功能、Google 的搜索能力、图像识别和自动驾驶汽车以及 Facebook 的脸识别技术。深度神经网络 ASICs 则是一种特殊用途的处理器，加速了系统的计算速度。该技术预计成熟时间从 2018 年起降至“2-5 年”。

决策智能和增强智能是 AI 的新兴方向，由于相关技术的中间过程不确定性、机器决策和行为背后原因的不可解释、系统的鲁棒性，以及隐私保护等一系列问题尚未解决，其技术的大规模应用落地还需一定时间。

3. 强化学习、自然语言处理、知识图谱、智能机器人、数字伦理等 AI 技术还在研究发展中，距离生产成熟至少均还需“5-10 年”。

强化学习让机器有了自我学习、自我思考的能力。该技术于 2019 年首次在 Gartner 技术成熟度曲线中出现并处于“萌芽期”，目前主要应用在游戏 AI 领域。

自然语言处理（NLP）技术近五年来的预计成熟时间一直维持在“5-10 年”不变，但是该技术发展状态变化迅速，从 2015 年的“萌芽期”、转到 2016 年的“期望膨胀期”并维持了三年，之后的 2019 年便跌入“幻灭期”。自然语言生成（NLG）是 NLP 的重要组成部分，旨在动态地识别数据中最相关的含义，将非语言格式的数据转换成人类可以理解的语言格式，降低人类和机器之间的沟通鸿沟。其典型应用包括自动写新闻、聊天机器人、数据解读和报告生成。目前 NLG 仍处于初步采用阶段。

自从谷歌于 2012 年推出知识图谱的概念之后，该技术首次出现在 Gartner 人工智能技术成熟度曲线中是在 2018 年。知识图谱以结构化的形式描述客观世界中概念、实体及其之间的关系，将互联网的信息表达成更接近人类认知世界的形式，提供了一种更好地组织、管理和理解互联网海量信息的能力。使用自然语言处理和相关的文本分析技术，知识图谱非常适合存储从非结构化资源分析中提取的数据。它们还能够存储结构化数据，包括隐式提供结构和内容的元数据，编码支持各种用例的处理信息。知识图谱技术发展迅速，经过两年的“萌芽期”，在 2020 年就进入“期望膨胀期”发展区间，但是，知识图谱在工业界还没有形

成大规模的应用，很多仍处于调研阶段。无论是通用知识图谱还是行业知识图谱，在实际应用中都面对很多挑战，例如，数据的可访问性、数据的深层关系、知识抽取困难、领域知识的集成、知识表达的复杂度和粒度不同、策略的生成、图谱构建自动化程度低等问题。该技术的预计成熟时间一直是“5-10 年”。

智能机器人是在物理世界中自主工作的机电主体，在短时间内从人工指导的训练和演示中学习，或通过其在工作上的监督经验来学习解决问题。智能机器人可以用语音、语言与人类互动，由于先进的感官能力，甚至可能与人类一起工作。与工业机器人相比，智能机器人迄今被采用的数量要少得多。

数字伦理包括安全、网络犯罪、隐私、社会互动、治理、自由意志等经济社会的方方面面。随着人们对数据隐私保护等法规意识的增强，数字伦理跃升到通胀预期的顶峰。但是，构建能够让社会公众信任人工智能等新技术的规制体系，让技术接受价值引导还需时日。

4. 距离生产成熟还很遥远的技术是通用人工智能、无人驾驶汽车。

通用人工智能也被称为“强人工智能”。在未来 10 年内，通用人工智能不太可能出现。当它最终出现时，很可能是许多特殊用途人工智能技术结合在一起的结果。无人驾驶汽车又称轮式移动机器人，是智能汽车的最高级形态，目前虽然已经处在公开路测阶段，但是该技术的成熟应用受制于传感器技术成本高、计算能力难以应对复杂多变的现实路况等多因素，在未来 10 年内，也不太可能出现。

5. 通用人工智能兴起与认知计算在高峰之前被淘汰密切相关。

认知计算是认知科学的核心技术子领域之一，是人工智能的重要组成部分，是模拟人脑认知过程的计算机系统。认知计算旨在放大人类智能，帮助人类更好地思考，核心在于辅助人类决策。

虽然认知计算包括部分人工智能领域的元素，但是它涉及的范围更广。认知计算更偏向于技术体系，人工智能更偏向于最终的应用形态。认知计算迅速攀升至通胀预期的顶峰，这是因为在最新一代人工智能市场上，目前的认知计算系统只是模仿、扩展人类的认知能力，而系统供应商和销售商则普遍过度推广和炒作这一术语。此外，随着以深度学习为代表的通用人工智能技术在众多领域取得了巨大进展，已成为市场上的热点。而认知计算的主要目标并不是构建能够思考的

机器，而是创造能够分析海量数据、自我学习且自主运转的计算机系统，从而增强人类认知能力，辅助人类工作。

Gartner 认为认知计算是一个被市场供应商过度夸耀的词语，Gartner 不相信这些系统有真正的认知，并认为这些系统仅能模仿或扩展人类的认知能力^[15]。因此，认知计算在各行业被广泛应用与渗透之前，在 Gartner 人工智能技术成熟度曲线中处于“高峰前被淘汰”的状态。

6. 自然语言生成等前沿技术缺乏突破而消失在成熟度曲线中。

自然语言生成是人工智能和计算语言学的分支，是生成可理解文本的计算机系统，于 2018 年的预计成熟期是 2-5 年；自然语言问答是能够接受用自然语言描述的问题，通过语义相似度的计算，自动地返回相关答案的系统，于 2018 年的预计成熟期是 5-10 年。之后，这两项技术虽然也得到了很大的发展，但是，总体上实现起来还有很多困难。目前的问答系统并不具备任何思维和推论能力，还不能像人类一样能自如地回答用户提出的各种问题，所能回答的问题受限于文档库，而且准确率还较低。自然语言生成也需要涉及许多自然语言处理技术，例如句法分析技术、文本生成技术等。目前这些技术还不够成熟，仅停留于研究阶段，在 Gartner 人工智能技术成熟度曲线中没有再出现，可能被归入到自然语言处理和自然语言生成之中。

¹⁵ Gartner&中国信通院 2018AI 发展白皮书 p.65
<http://www.caict.ac.cn/kxyj/qwfb/bps/201809/P020180918696200669434.pdf>

4 人工智能过去十年研究进展

4.1 全球高水平论文发表情况

人工智能过去十年发展快速，从学术研究走向商业化。本报告专注于通过分析在国际顶级期刊和会议上的人工智能领域科研论文发表情况来研究该领域的成果产出。由图 4-1 可见，从 2011 年以来人工智能领域高水平论文发表量整体上呈现稳步增长态势，取得了很多科研成果。这些科研成果涵盖 R-CNN 算法、神经机器翻译的新方法等。

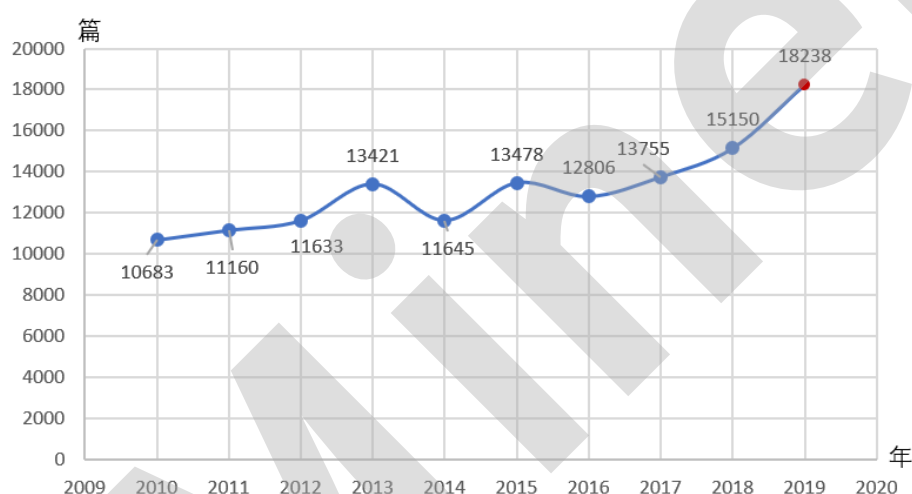


图 4-1 过去十年人工智能领域国际顶级期刊会议论文数量趋势

从高水平科研论文的国家分布来看，人工智能领域论文发表量居于前十的国家依次是美国、中国、德国、英国、日本、加拿大、法国、韩国、意大利和澳大利亚，如图 4-2 所示。美国和中国的高水平论文发表量明显高于其他国家，分别位居第一、二名，中国的论文量紧随美国之后。

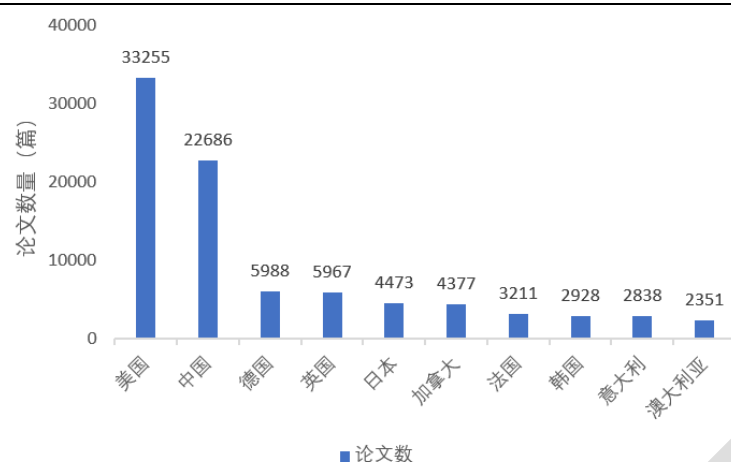


图 4-2 过去十年人工智能领域高水平论文发表量前十国家

研究发现，各个国家的人工智能领域高水平科研论文发布具有以下特征。

1. 开展跨国科研合作较多的国家是美国和中国

从论文的国际合作网络看，美国和中国的 AI 高水平论文发表均存在较多的跨国合作现象，如图 4-3 所示。其中，AI 技术实力领先的美国所参与的高水平论文跨国合作最多，是各国的主要合作国家，过去十年，美国的 33255 篇 AI 高水平论文之中，出现过中国、英国、加拿大、德国、印度等 30 多个合作国家，合作国家数量最多；中国的跨国科研合作国家数量位居第二，在其 22686 篇 AI 高水平论文之中，出现了美国、加拿大、新加坡、英国、日本等 20 多个合作国家；英国和德国的 AI 高水平论文跨国合作国家数量均为 18 个。其余国家在 AI 高水平论文方面开展的跨国合作数量较少。

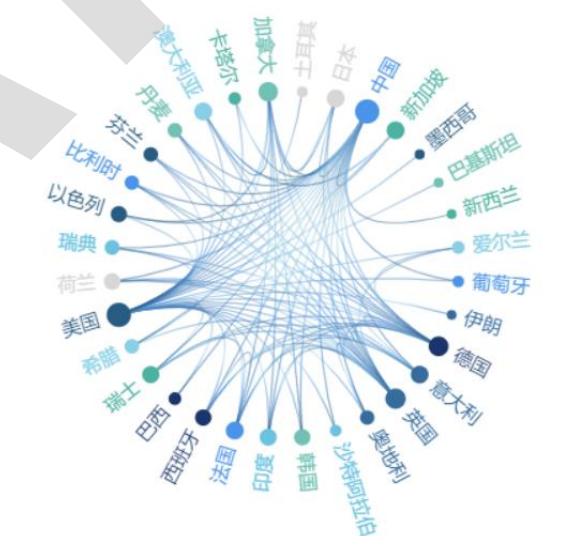


图 4-3 过去十年人工智能领域高水平论文发表国际合作国家分布

2. 中美两国是对方 AI 领域的重要科研合作伙伴

在 AI 高水平论文发表方面，美国和中国均是对方的重要科研合作伙伴。过去十年，美国在其 30 多个合作国家之中，与中国合作的 AI 高水平论文数量占比最多，为 18.53%。同时，中国在其 20 多个合作国家之中，与美国合作的论文数量最多，占比 27.16%。可见，开展国际科研合作已成为中美两国 AI 研究成果产出的重要方式。

3. 跨国科研合作可以提高合作本国 AI 研究成果的影响力

分析发现，美国的 AI 领域高水平论文平均引用率为 44.99，中国的 AI 领域高水平论文平均引用率为 31.88。相比而言，中国和美国合作论文的平均引用率达 51.2，其影响力明显高于中国和美国各自论文的平均引用水平，这表明跨国合作的科研成果在世界人工智能研发领域的展示和交流几率大大增加。

4.2 获得图灵奖的人工智能技术

图灵奖（ACM A.M. Turing Award）是计算机界最负盛名、最崇高的一个奖项，有“计算机界的诺贝尔奖”之称。图灵奖是计算机协会（ACM）于 1966 年设立的奖项，专门奖励对计算机事业做出重要贡献的个人。其名称取自世界计算机科学的先驱、英国科学家、曼彻斯特大学教授艾伦·图灵（A.M. Turing）。图灵奖获奖者必须是在计算机领域具有持久而重大的先进性的技术贡献，大多数获奖者是计算机科学家。第一位图灵得主是卡耐基梅隆大学的 Alan Perlis（1966 年），第一位女性获奖者是 IBM 的 Frances E. Allen（2006 年）。

通过 AMiner 智能引擎，可以自动收集历年来图灵奖获得者及其学者画像信息（基本信息、研究兴趣等），以及该学者的论文和专著等信息。由于每年度的图灵获奖者一般在次年 3 月下旬由美国计算机协会（ACM）官方颁发，因此本报告统计了截至 2020 年颁发的近十年（2010-2019 年）图灵奖得主数据。分析发现，图灵奖近十年授予领域具有如下特征。

1. 十年中三次正式颁奖给人工智能领域，推动该领域发展

图灵奖颁发的领域，在一定程度上反映了计算机科学技术发展方向的缩影。数据显示，过去十年图灵奖分别授予给了计算理论、概率和因果推理、密码学、分布式和并发系统、数据库系统、万维网、计算机系统、深度神经网络和 3D 计

计算机图形学九个领域，具体如图 4-4 所示。从获奖内容、创新角度、研究领域等维度来看，图灵奖注重原始理论创新和学科交叉，具有科研优势积累现象。

过去十年的图灵奖有三次正式授予给人工智能领域。第一次是 2010 年，Leslie Valiant 因对计算理论的贡献（PAC、枚举复杂性、代数计算和并行分布式计算）获得图灵奖，该成果是人工智能领域快速发展的数学基础之一。第二次是 2011 年，因 Judea Pearl 通过概率和因果推理对人工智能做出贡献而颁奖；第三次是 2018 年，深度学习领域三位大神 Yoshua Bengio、Geoffrey Hinton 和 Yann LeCun 因为在概念和工程上的重大突破推动了神经网络成为计算机领域关键技术而荣获图灵奖。Hinton 的反向传播（BP）算法、LeCun 对卷积神经网络（CNN）的推动以及 Bengio 对循环神经网络（RNN）的贡献是目前图像识别、语音识别、自然语言处理等获得跳跃式发展的基础。中国科学院张钹院士在《迈向第三代人工智能》^[16]一文中也提到这 5 位图灵奖得主在创建第二次 AI 中所做出的重大贡献。

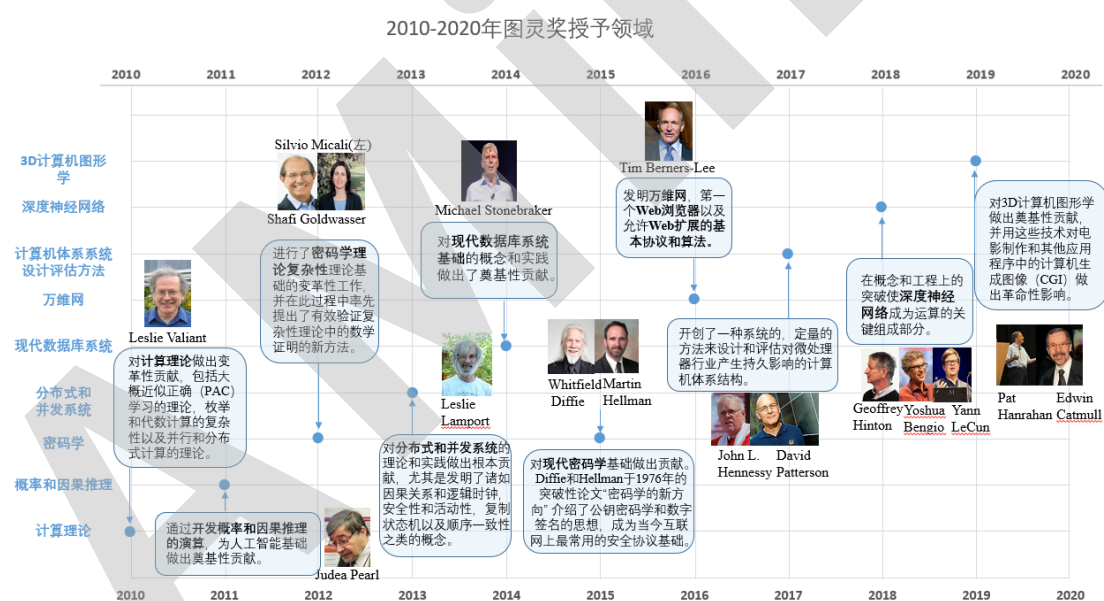


图 4-4 2010-2020 年图灵奖授予的计算机领域

2. 人工智能领域获奖人数占据四分之一

由图 4-4 可见，过去十年，共有 16 位学者获得图灵奖。其中，包括 5 位人工智能领域学者获此殊荣，占比 31%，这反映出人工智能在计算机学科中的地位

¹⁶ 张钹,朱军,苏航. 迈向第三代人工智能[J]. 中国科学:信息科学, 2020, v.50(09):7-28.

已不容忽视。同时，人工智能领域图灵奖从初期的单独获奖者到近年来的共同获奖者，越来越呈现出高层次学者强强联合的研究趋势。

3. 美国培养并拥有八成以上的图灵奖得主

过去十年的 16 位图灵奖获得者之中，有 13 位来自美国、2 位来自英国、1 位来自加拿大，如表 4-1 所示。在美国的 13 位图灵奖得主之中，有 10 位是在美国本国接受的全部高等教育、2 位拥有美国和其他国家教育背景、仅 1 位没有美国教育背景。其中，2011 年获奖者 Judea Pearl 拥有以色列本科教育和美国纽约大学博士教育背景；2012 年图灵奖得主 Silvio Micali 拥有意大利本科教育和美国加州伯克利大学博士教育背景。唯一没有美国教育背景的是 2018 年图灵奖得主 Yann LeCun，他仅有法国教育背景。八成以上图灵奖得主具有美国教育或工作背景的事实，反映出美国人工智能高层次人才培养的强势竞争力。

表 4-1 2010-2020 年图灵奖得主所在国家及教育背景情况

获奖时所在国家	教育背景	人数
美国	美国	10
	法国	1
	以色列、美国	1
	意大利、美国	1
加拿大	加拿大	1
英国	英国	2

4. 欧洲培养的高层次人才中有三位被吸引到美国学习或工作并获图灵奖

从这些图灵获奖者的教育背景来看，美国与欧洲国家的学术交流非常多。Yann LeCun 拥有法国教育背景后到美国任职并获图灵奖；Judea Pearl 和 Silvio Micali 分别从以色列和意大利被吸引到美国继续深造而拥有跨国双重教育背景，后来均到美国任职并在美国获得图灵奖。此外，英国的这两位图灵奖得主虽然均是在本国接受的高等教育并且获奖时都在本国，但都有过一些美国任职经历。Geoffrey Hinton 博士毕业于英国爱丁堡大学，后来陆续在谷歌、卡内基梅隆大学、加州大学圣地亚哥分校等美国机构任职；Tim Berners-Lee 在英国牛津大学本科毕业后，也有过在麻省理工学院任职的经历。可见，美国的高等教育体系不仅培养了自己本国的 AI 领域高端人才，而且从欧洲国家吸引和留住了多位领域精英。

5. 仅加拿大图灵奖得主没有任何美国教育和任职经历

在这 16 位图灵奖获得者之中，仅有加拿大的 Yoshua Bengio 在本国的麦吉尔大学接受了高等教育并在本国蒙特利尔大学任职，并于 2018 年因在神经网络概念和工程上的突破而获奖。这在一定程度上反映出加拿大在人工智能领域高层次人才培养和质量上较为成功。

6. 图灵得主们在领域相关论文发表后需要平均等待 37.1 年之后才获奖

通过 AMiner 人才画像数据获取这些图灵奖得主所发表的第一篇与获奖理由相关主题的论文，计算得出该论文发表年份距离作者获得图灵奖时间，从而得到这些图灵得主的获奖时间长短，如图 4-5 所示。结果发现，图灵奖得主获奖时一般距离其首次发表获奖领域相关论文至少已经三十年以上，平均为 37.1 年。其中，Edwin Catmull 于 2019 年获得图灵奖，距离其在计算机图形学领域发表的最早论文 A system for computer generated movies 已经过去了 47 年，等待获奖时间最久。而获奖等待时间最短的是 Tim Berners-Lee，他于 1990 年发表 WorldWideWeb: Proposal for a Hypertext Project 论文，仅在 26 年后的 2016 年就因发明万维网、Web 浏览器以及允许 Web 扩展的基本协议和算法获得图灵奖。

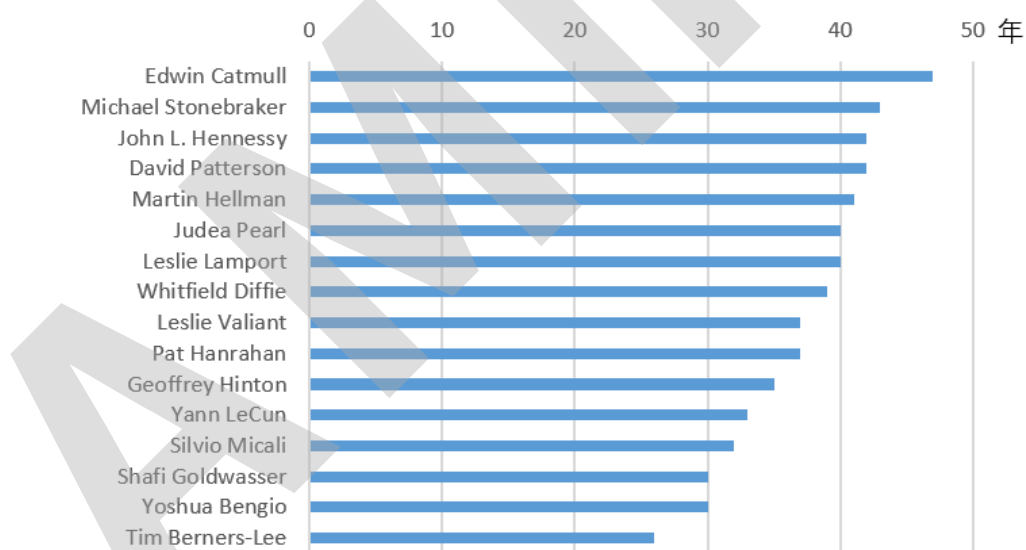


图 4-5 图灵奖得主的首篇领域论文发表距离获奖年份的时长

7. 八成以上图灵得主获奖时已经度过了其科研论文高峰产出期

基于 AMiner 平台上各位图灵奖得主的论文数据和人物画像，分析发现，图灵奖得主一般会在获奖后保持原来的研究方向，但是他们的论文发表量却减少了。有 80.1% 的图灵得主在获图灵奖后的论文年均产出量低于其获奖前的年均论文

产出量，如图 4-6 所示，这反映出他们在获得图灵奖时已经普遍过了其学术产出高峰时期。

值得一提的是，Yoshua Bengio, Judea Pearl 和 Michael Stonebraker 三位学者是例外，他们在获得图灵奖之后的年均论文产出量较其获奖前均有不同程度的增加，分别增加了 328.07%、49.53%和 12.32%。其中，Yoshua Bengio 在 2018 年获得图灵奖后论文年均发表量激增特别明显，并在 2019 年发表 96 篇论文，达到其论文产出峰值，他的 AMiner 学术画像及年度论文发表量如图 4-7 所示。

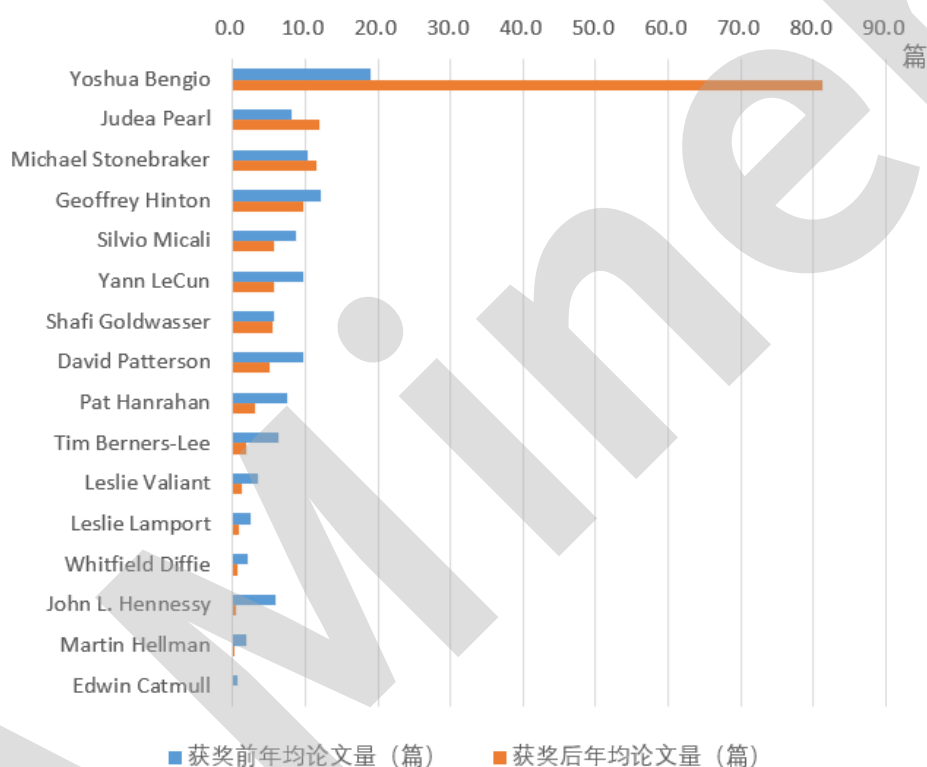


图 4-6 2010-2020 年图灵奖得主获奖前后的年均论文发表量



图 4-7 2018 年图灵奖获得者 Yoshua Bengio 的 AMiner 学术画像

需要指出的是，部分图灵得主在获奖后的论文产出量减少除了其学术产出减少之外，还存在以下两个原因。一是他们可能已不再全力进行学术研究，而是转向参与社会事业等，从而导致其获奖后的论文发表量减少。例如，Whitfield Diffie 在 2015 年因密码学的贡献获得图灵奖后，转向致力于促进信息安全和隐私权的保护；Martin Hellman 因密码学的贡献获得同年的图灵奖后，转向致力于研究国际安全与核武器削减。二是也有少量图灵得主在发表相关论文后早已投身于工业界，例如，计算机图形学先驱 Edwin Catmull 在研究生毕业后就在卢卡斯、皮克斯等公司就职，早已离开学术界，曾担任 Pixar 动画和 Disney 动画的总裁，四次获电影 Oscar 奖（1993、1996、2001 和 2008），其 AMiner 学术画像及年度论文发表量如图 4-8 所示。

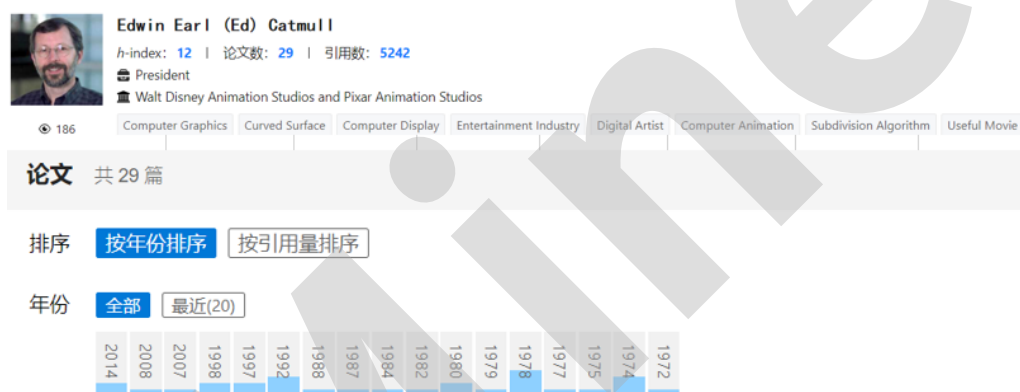


图 4-8 2019 年图灵奖获得者 Edwin Catmull 的 AMiner 学术画像

4.3 媒体评选出的重大人工智能技术

自 2001 年起，《麻省理工科技评论》每年都会评选出当年的“十大突破性技术”。所评选出的技术榜单曾精准预测了脑机接口、智能手表、癌症基因疗法、深度学习等诸多热门技术的崛起，在全球科技领域具有举足轻重的影响力。

分析发现，媒体评出的人类突破性技术之中近 30%与人工智能相关。通过分析挖掘 2010-2020 年《麻省理工科技评论》评选出当年的“十大突破性技术”，发现有 32 项人工智能相关技术入选榜单，占比近 30%，其中包括：2013 年的深度学习、2014 年的神经形态芯片、2016 年的语音接口与知识分享型机器人、2017 年的自动驾驶卡车与强化学习，2018 年的流利对话的 AI 助手、给所有人的人工智能和对抗性神经网络、2019 年的灵巧机器人，以及 2020 年的微型人工智能和

人工智能发现分子。具体名单如表 4-2 所示。

表 4-2 2010-2020 年入选《麻省理工科技评论》“十大突破性技术”榜单的人工智能相关技术

年份	技术名称-中文	技术名称-英文	简介
2020	人工智能发现分子	AI-discovered Molecules	利用深度学习和生成模型相关的技术，成功确定了具有理想特性的新分子进行药物合成和测试。
	微型人工智能	Tiny AI	在不丧失能力的情况下，缩小现有深度学习模型，设备不需要与云端交互就能实现智能化操作。
2019	灵巧机器人	Robot Dexterity	能自我学习并学会应对人类现实世界。
	流利对话的 AI 助手	Smooth-talking AI Assistant	可以执行基于对话的任务。
2018	给所有人的人工智能	AI for Everyone	将机器学习工具搬上云端，有助于 AI 更广泛的传播
	对抗性神经网络	Dueling Neural Networks	两个 AI 系统可以通过相互对抗来创造超级真实的原创图像或声音，给机器带来一种类似想象力的能力
2017	强化学习	Reinforcement Learning	一种能使计算机在没有明确指导的情况下像人一样自主学习的人工智能方法
	自动驾驶货车	Self-Driving Trucks	可以在高速路上自动驾驶的长途货车。
	刷脸支付	Paying with Your Face	一种安全方便的支付方式，但存在隐私泄露问题。
	僵尸物联网	Botnets of Things	可以感染并控制摄像头、监视器以及其他消费电子产品的恶意软件，可造成大规模的网络瘫痪。
2016	语音接口	Conversational Interfaces	将语音识别和自然语言理解相结合
	知识分享型机器人	Robots That Teach Each Other	可以学习任务，并将知识传送到云端，以供其他机器人学习
	特斯拉自动驾驶仪	Tesla Autopilot	汽车可以在各种环境下安全自驾
2015	虚拟现实设备	Magic Leap	让虚拟物品出现在真实场景中的设备。

年份	技术名称-中文	技术名称-英文	简介
	苹果支付	Apple Pay	让手机成为钱包，能够方便地在日常场景下使用。
2014	虚拟现实头盔	Oculus Rift	视觉沉浸式的娱乐方式和交际手段。
	超私密智能手机	Ultraprivate Smartphones	传输最少个人信息的手机
	智能风能和太阳能	Smart Wind and Solar Power	极准确的风能和太阳能预报。
	神经形态芯片	Neuromorphic Chips	配置更像大脑，能使计算机更快更敏锐地感知周围环境。
	灵巧型行走机器人	Agile Robots	能平衡和灵巧地在不平的地面行走和奔跑
2013	深度学习	Deep Learning	借助大量算力，机器可以识别对象并实时翻译语音。
	蓝领机器人	Baxter: The Blue-Collar Robot Baxter	能够取代制造业的蓝领工人
	智能手表	Watches	简单和易用，从概念到量产。
2012	3D 晶体管	3-D Transistors	可提升芯片的计算速度、减少错误，还能极大地降低能耗，延长摩尔定律的寿命。
	脸书的时间线	Facebook's Timeline	将用户产生的大量数据组织起来
2011	手势界面	Gestural Interfaces	使用非触摸、动作感应手势控制界面。
	云端流媒体	Cloud Streaming	快速配置专业流，随时随地流向任何规模的受众。
	社会索引	Social Indexing	能挖掘用户的朋友的兴趣并从多个站点收集信息。
2010	实时搜索	Real-Time Search	用户的搜索结果是实时变化的。
	云编程	Cloud Programming	使开发人员能够用熟悉的编程方式快速构建任何一种基于云计算的应用。
	植入式芯片	Implantable Electronics	具有生物相容性、符合医用安全标准、能够适用于人体内部复杂环境的特殊芯片。
	移动 3D	Mobile 3D	移动设备上的 3D。

来源：根据公开资料整理。

入选过去十年媒体“突破性技术”榜单的这些人工智能技术具有如下特征。

1. 机器人智能越来越强大、取代人类工作的范围和可能性越来越大

机器人技术过去十年的发展进步很快，历经了 2013 年蓝领机器人、2014 年的灵巧型行走机器人、2016 年的知识分享型机器人直到 2019 年的灵巧机器人，其智能化特征越来越明显，在人类社会中的运用范围也越来越广泛。初期的**蓝领机器人**不需要休息与食物、能 24 小时工作且没有情绪、会绝对服从指令，旨在未来制造业中成为取代蓝领工人的重要生产力。后来的**灵巧型行走机器人**拥有非常出色的平衡性和灵巧性，能在崎岖不平的地面行走和奔跑，具备作为机器人在人类环境中安全、轻松活动的的能力，能够在人类环境中导航、参与紧急救援行动、从事日常杂活。**知识分享型机器人**则更加智能，它可以学习任务，在抓取各种工具的同时“锻炼”自己的控制软件，同时将知识传送到云端，以供其他机器人学习。近期的**灵巧机器人**更是基于 AI 系统，通过强化学习和自我学习学会应对人类现实世界以胜任更多的任务，甚至取代人类工作。

2. AI 技术使得人机通信的方式越来越像人类之间的自然交互

过去十年，人与计算机之间进行有效通信的技术方式发生了很大变化，变得越来越人性化。人机之间先进的通信技术在 2011 年是实时搜索和手势界面，2013 年换成了简单和易于使用的智能手表，到 2014 年变为视觉沉浸式界面的虚拟现实头盔与能够保护个人隐私的超私密智能手机。后来的人机通信方式更多利用了**人类的生物特征进行交互**，例如 2016 年的将语音识别和自然语言理解相结合的语音接口、2017 年的刷脸支付，以及 2018 年通过将机器学习工具搬上云端而实现了给所有人的人工智能、2019 流利对话的 AI 助手。可见，借助于越来越先进的人工智能技术，人机之间的通信方式逐渐在向人与人之间的自然交流方向转变。

3. AI 赋能的应用场景逐渐已渗透到更尖端、更微观领域

人工智能可以赋能不同领域，通过发挥展现 AI 技术的先进性、创新性，为用户带来更便捷、更高效的服务。过去十年，人工智能赋能的应用场景不仅覆盖了人类日常的工作、娱乐、购物和交通，而且**延伸到信息发现、用户兴趣挖掘、知识分享，以及探索自然周围环境等尖端领域，帮助人类学习和完成高难度任务**。同时，智能化应用场景还呈现出向微型和微观领域发展的态势。2020 年的**微型人工智能技术**就是保证在不丧失能力的情况下，缩小现有深度学习模型，设备不

需要与云端交互就能实现智能化操作。而人工智能发现分子技术，则是通过机器学习工具来探索包含已知分子及其特性的大型数据库，并利用这些信息实现以更快的速度、更低的成本发现新的候选药物的可能性，成为扩展药物发现领域化学家们激发想象力的新工具。例如 2019 年 9 月，香港的 Insilico Medicine 公司和多伦多大学的研究团队利用深度学习和生成模型相关的技术，成功确定了大约 30000 种具有理想特性的新分子，用于药物合成和测试。2020 年 11 月 30 日，DeepMind 研发的人工智能系统 AlphaFold 在第 14 届国际蛋白质结构预测竞赛（CASP）中摘得桂冠，被誉为是 2020 年 AI 的最大突破、是首个有望获得诺贝尔奖的人工智能成果。AlphaFold 依赖深度神经网络且主要是卷积神经网络技术，获得了从蛋白质的氨基酸序列确定蛋白质 3D 形状的能力，且产生的蛋白质 3D 模型比以往任何一种方法都精确得多，从而解决了困扰生物学界 50 年的“蛋白质折叠”难题。Nature 期刊以“*It will change everything*”^[17]（将改变一切）为标题，指出 DeepMind 在解决蛋白结构问题上做出“巨大飞跃”。

4.4 国际顶会顶刊最佳论文授予领域分析

人工智能领域顶级期刊和会议每年都会在众多学术论文之中，通过“双盲评审”，评选出最有新意和价值的研究论文作为最佳研究论文，并授予“Best Paper”奖项。每年大会的最佳论文奖一般分两类，一类是最佳研究论文（Research track），另一类是最佳应用论文（Applied track）。部分会议在每年选出多篇最佳论文（分列第一、二、三名），也有部分顶会每隔几年才会选出一篇最佳论文。

从过去十多年的经验来看，国际顶会历年的最佳研究论文都会对之后很多领域的研究有着开创性的影响。因此，不论是从阅读经典文献的角度，还是从学习最新研究成果的角度来说，分析和探讨每年的最佳研究论文都极具价值。

本部分收集整理了 2011-2020 年期间的人工智能领域国际顶级会议最佳论文奖项第一名的全部论文（对于不区分名次的最佳论文则全部收录），再对这些论文所属领域进行分析。统计发现，过去十年荣获“最佳论文”奖项的论文来自 34 个国际顶会、共计 440 篇，其中，research track 最佳论文 409 篇，占比 93%。各

¹⁷ Ewen Callaway, *It will change everything: DeepMind's AI makes gigantic leap in solving protein structures*, 30 November 2020, <https://www.nature.com/articles/d41586-020-03348-4>

会议最佳论文量的具体分布如表 4-3 所示，FOCS、IEEE VIS、ISSCC 等最佳论文数量较多主要是因为该会议每年颁发 3 篇杰出论文奖(Outstanding Paper Award)且不区分先后名次。ICASSP 等会议最佳论文量较少主要是由于存在缺失数据。

表 4-3 2011-2020 年人工智能领域国际顶级会议最佳论文授予量分布（单位：篇）

会议简称	所属子领域	最佳论文量	会议简称	所属子领域	最佳论文量
FOCS	计算理论	23	OSDI	计算机系统	12
ISSCC	芯片技术	22	SIGMOD	数据库	12
S&P	安全与隐私	22	SOSP	计算机系统	12
IEEE VIS	可视化	21	MobiCom	计算机网络	11
CCS	安全与隐私	20	WSDM	数据挖掘	11
ISWC	知识工程	20	WWW	信息检索与推荐	11
STOC	计算理论	20	ICLR	机器学习	10
ICML	机器学习	18	IJCAI	经典 AI	10
NeurIPS	机器学习	16	IROS	机器人	10
AAAI	经典 AI	14	RecSys	信息检索与推荐	10
ICRA	机器人	14	SIGIR	信息检索与推荐	10
SIGCOMM	计算机网络	14	KDD	数据挖掘	9
VLDB	数据库	14	FPGA	芯片技术	7
EMNLP	自然语言处理	13	ECCV	计算机视觉	6
ACL	自然语言处理	12	ICCV	计算机视觉	5
CVPR	计算机视觉	12	KR	知识工程	5
DAC	芯片技术	12	ICASSP	语音识别	2

从所属细分研究领域来看，这些最佳论文覆盖了机器学习、计算机视觉、自然语言处理、机器人、知识工程、语音识别、数据挖掘、信息检索与推荐、数据库、人机交互、计算机图形学、可视化、安全与隐私、计算机网络、计算机系统、计算理论、经典 AI、芯片技术等 18 个子领域。

1. 顶会最佳论文奖呈现出较多跨领域授予现象

总体而言，尽管这些国际顶级会议将大部分的最佳论文奖都授予给了本会议所属的 AI 子领域，但是授予非本会议领域的最佳论文数量占比较高，达 22.3%。其中，WSDM 是数据挖掘领域重要国际会议，在它授予的最佳论文奖之中，有 90.9%的最佳论文被授予给非数据挖掘领域的论文，在所有会议中占比最高。其次，KDD 会议将 88.9%最佳论文奖授予给非数据挖掘领域的论文，WWW 会议将 63.6%的最佳论文奖授予给非信息检索与推荐领域的论文。各个会议最佳论文授予情况具体如表 4-4 所示。

表 4-4 2011-2020 年人工智能领域国际顶级会议最佳论文授予其他 AI 领域比例

会议简称	会议所属领域	最佳论文量（篇）	授予其他子领域最佳论文占比
WSDM	数据挖掘	11	90.9%
KDD	数据挖掘	9	88.9%
WWW	信息检索与推荐	11	63.6%
IJCAI	经典 AI	10	60.0%
AAAI	经典 AI	14	57.1%
MobiCom	计算机网络	11	54.5%
ICASSP	语音识别	2	50.0%
OSDI	计算机系统	12	50.0%
SIGMOD	数据库	12	50.0%
VLDB	数据库	14	50.0%
FPGA	芯片技术	7	42.9%
KR	知识工程	5	40.0%
DAC	芯片技术	12	33.3%
ICML	机器学习	18	27.8%
NeurIPS	机器学习	16	25.0%
SIGCOMM	计算机网络	14	21.4%
ICCV	计算机视觉	5	20.0%
ISWC	知识工程	20	20.0%
IROS	机器人	10	10.0%
RecSys	信息检索与推荐	10	10.0%
SIGIR	信息检索与推荐	10	10.0%
ACL	自然语言处理	12	8.3%
SOSP	计算机系统	12	8.3%
IEEE VIS	可视化	21	4.8%
FOCS	计算理论	23	4.3%
CCS	安全与隐私	20	0.0%
CVPR	计算机视觉	12	
ECCV	计算机视觉	6	
EMNLP	自然语言处理	13	
ICLR	机器学习	10	
ICRA	机器人	14	
ISSCC	芯片技术	22	
S&P	安全与隐私	22	
STOC	计算理论	20	

2. 信息检索与推荐、机器学习和计算理论出现较多的跨领域授予最佳论文

从跨领域授予最佳论文奖的整体数量来看，信息检索与推荐、机器学习和计算理论是获得最佳论文奖项数量较多的三个领域，占比均超过 10%，详细情况如

图 4-9 所示。这反映出这三个子领域的跨领域研究成果所获的专业认可度较高，在一定程度上促进了相关技术在多个 AI 子领域的快速发展和进步。从来源会议来看，信息检索与推荐领域最佳论文除了被 RecSys、SIGIR 和 WWW 本领域会议最多授予之外，也较多被数据挖掘领域会议 WSDM、数据库领域的 SIGMOD 和 VLDB 授予最佳论文奖，如图 4-10 所示。

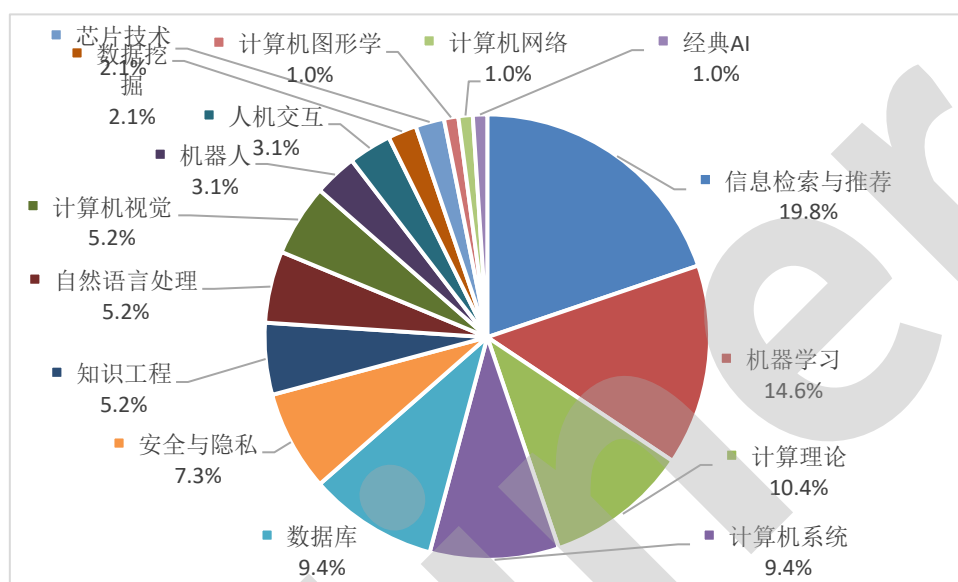


图 4-9 人工智能领域国际顶级会议最佳论文奖跨 AI 子领域授予分布图

WSDM 2019 [MAP: 0.0]		排名	被引用
BEST PAPER!	Slice - Scalable Linear Extreme Classifiers Trained on 100 Million Labels for Related Searches.	14	16
	Himanshu Jain, Venkatesh Balasubramanian, Bhanu Chunduri, Manik Varma		
SIGMOD 2018 [MAP: 0.0]		排名	被引用
BEST PAPER!	SuRF: Practical Range Query Filtering with Fast Succinct Tries.	11	40
	Huan Chen Zhang, Hyeontaek Lim, Viktor Leis, David G. Andersen, Michael Kaminsky, Kimberly Keeton, Andrew Pavlo		
VLDB 2017 [MAP: 0.6111]		排名	被引用
BEST PAPER!	Provenance for Natural Language Queries.	1	22
	Daniel Deutch, Nave Frost, Amir Gilad		

来源：AMiner 会议系统

图 4-10 信息检索与推荐领域最佳论文跨领域授予会议示例

3. 机器学习技术成果集中在 2016-2018 年获得较多奖项认可

过去十年顶会最佳论文授予技术领域的年度趋势如图 4-11 所示，其中，色块颜色代表该项技术在某项会议所被授予的最佳论文数量，色块颜色越深表示论文数量越多。

从最佳论文奖被授予领域的年度趋势来看，机器学习领域过去十年内最佳论文在 2016 年被授予最多，为 9 篇，其次是 2017 和 2018 年，相关最佳论文数量

均为 8 篇。从来源会议来看，共有 14 个顶会将最佳论文奖分别授予给机器学习领域。其中，最佳论文奖授予量较多的两大会议是 International Conference on Machine Learning（ICML）和 International Conference on Learning Representations（ICLR），分别为 13 和 12 篇。

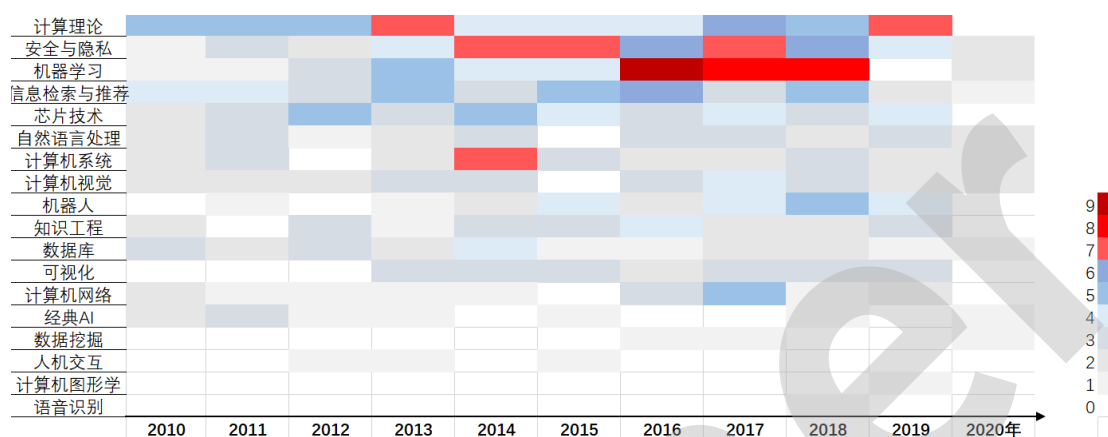


图 4-11 人工智能领域国际顶刊顶会最佳论文授予领域年度趋势

4. 安全与隐私领域最佳论文授予数量呈现增多趋势

随着人工智能快速发展和应用，许多领域开始注重技术的安全与隐私性。这体现在逐年增多的领域最佳论文数量上。尤其是 2014 年之后，安全与隐私领域年度最佳论文数量均超过 5 篇。

过去十年来，CCS、ICML、OSDI、S&P、SIGCOMM 和 WWW 等顶会均曾授予过安全与隐私领域的最佳论文奖。安全与隐私的最佳论文授予主要来自于 IEEE Symposium on Security and Privacy（S&P）和 ACM Conference on Computer and Communications Security（CCS）两大会议。

4.5 国际顶会顶刊领域高影响力论文分析

论文引用量是衡量一个科研文献被业界认可度的标志，也是该文献影响力的重要体现。本部分针对人工智能国际顶会顶刊 2011-2020 年期间的所发表论文的引用量特征及所属领域进行分析。结果分析发现，某一学术会议中引用量最高的论文未必是该会议授予最佳论文奖的论文，反之亦然。

分析还发现，人工智能不同子领域论文的最高引用量的量级跨度很大。如图 4-12 所示，2011-2020 年期间人工智能国际顶会顶刊最高引用量前十论文研究以

机器学习领域为首，其次是计算机视觉领域研究论文。机器学习和计算机视觉领域论文的引用量级均达到 25 万次以上，明显高于其他子领域最高引用论文的引用量。在所有子领域之中，知识工程领域论文的引用量级最少，不足于机器学习领域论文引用量的 2%。

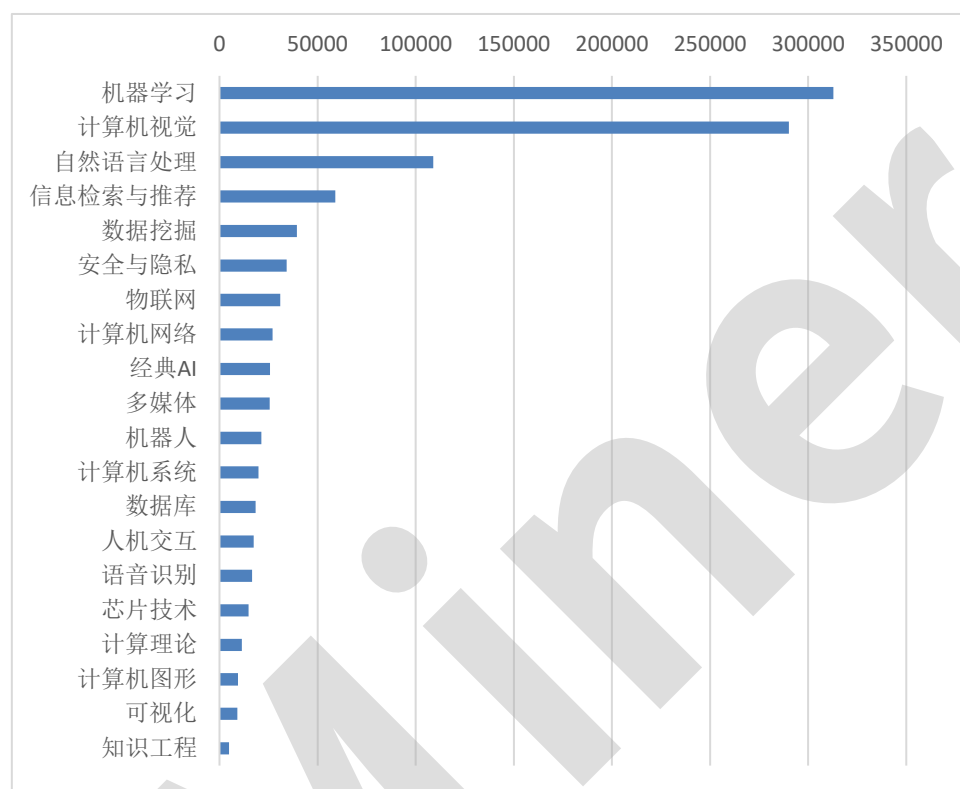


图 4-12 2011-2020 年人工智能国际顶会顶刊各子领域最高引用量前十论文的引用量分布

具体来看，人工智能各个子领域在过去十年中出现在国际顶级会议期刊上的最高影响力论文相关信息如表 4-5 所示。其中，计算机视觉领域最高影响力论文是 2016 年 CVPR 上、以 Facebook AI Research 何恺明为第一作者的“Deep Residual Learning for Image Recognition”文章，其引用量已超过 6 万。机器学习领域引用量最高的论文是发表在 2015 年 ICLR 会议上的“Adam: A Method for Stochastic Optimization”，该文是由 Google Brain 的 Diederik P. Kingma 和加拿大多伦多大学的助理教授 Jimmy Lei Ba 联合发表，目前引用量将近 6 万。

表 4-5 2011-2020 年人工智能子领域最高影响力论文

子领域	论文标题	引用量
计算机视觉	 Deep Residual Learning for Image Recognition Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the laye... CVPR, (2016)	66187
机器学习	 Adam: A Method for Stochastic Optimization Diederik P. Kingma, Jimmy Lei Ba We introduce Adam, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. The method is straightforward to implement, is computationally efficient, has little memory requirements, is invari... international conference on learning representations, (2015)	62517
自然语言处理	 Glove: Global Vectors for Word Representation Jeffrey Pennington, Richard Socher, Christopher D. Manning EMNLP, pp.1532-1543, (2014)	18865
多媒体	 Caffe: Convolutional Architecture for Fast Feature Embedding Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B. Girshick, Sergio Guadarrama, Trevor Darrell Caffe provides multimedia scientists and practitioners with a clean and modifiable framework for state-of-the-art deep learning algorithms and a collection of reference models. The framework is a BSD-licensed C++ library with Python and MATLAB bindings for training and deploying ... ACM Multimedia 2001, (2014): 675-678	14408
信息检索与推荐	 What is Twitter, a social network or a news media Haewoon Kwak, Changhyun Lee, Hosung Park, Sue Moon Twitter, a microblogging service less than three years old, commands more than 41 million users as of July 2009 and is growing fast. Twitter users tweet about any topic within the 140-character limit and follow others to receive their tweets. The goal of this paper is to study th... WWW, pp.591-600, (2010)	8207
数据挖掘	 XGBoost: A Scalable Tree Boosting System Tianqi Chen, Carlos Guestrin Tree boosting is a highly effective and widely used machine learning method. In this paper, we describe a scalable end-to-end tree boosting system called XGBoost, which is used widely by data scientists to achieve state-of-the-art results on many machine learning challenges. We p... KDD, (2016)	8092
语音识别	 Speech recognition with deep recurrent neural networks Alex Graves, Abdel-rahman Mohamed, Geoffrey E. Hinton Recurrent neural networks (RNNs) are a powerful model for sequential data. End-to-end training methods such as Connectionist Temporal Classification make it possible to train RNNs for sequence labelling problems where the input-output alignment is unknown. The combination of thes... Acoustics, Speech and Signal Processing, (2013): 6645-6649	6595
经典 AI	 Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke Very deep convolutional networks have been central to the largest advances in image recognition performance in recent years. One example is the Inception architecture that has been shown to achieve very good performance at relatively low computational cost. Recently, the introduc... AAAI, (2017): 4278-4284	5979

子领域	论文标题	引用量
物联网	Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas Thomas L. Marzetta A cellular base station serves a multiplicity of single-antenna terminals over the same time-frequency interval. Time-division duplex operation combined with reverse-link pilots enables the base station to estimate the reciprocal forward- and reverse-link channels. The conjugate... IEEE Transactions on Wireless Communications, no. 11 (2010): 3590-3600	5256
计算机网络	Fog computing and its role in the internet of things Flavio Bonomi, Rodolfo Milito, Jiang Zhu, Sateesh Addepalli Fog Computing extends the Cloud Computing paradigm to the edge of the network, thus enabling a new breed of applications and services. Defining characteristics of the Fog are: a) Low latency and location awareness; b) Wide-spread geographical distribution; c) Mobility; d) Very la... MCC@SIGCOMM, (2012)	4941
机器人	3D is here: Point Cloud Library (PCL) Radu Bogdan Rusu, Steve Cousins ICRA, (2011)	4138
数据库	Pregel: a system for large-scale graph processing Grzegorz Malewicz, Matthew H. Austern, Aart J. C. Bik, James C. Dehnert, Ilan Horn, Naty Leiser, Grzegorz Czajkowski Many practical computing problems concern large graphs. Standard examples include the Web graph and various social networks. The scale of these graphs - in some cases billions of vertices, trillions of edges - poses challenges to their efficient processing. In this paper we prese... international conference on management of data, pp.135-146, (2010)	4022
计算机系统	TaintDroid: An Information-Flow Tracking System for Realtime Privacy Monitoring on Smartphones William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung, Patrick McDaniel, Anmol N. Sheth Today's smartphone operating systems frequently fail to provide users with visibility into how third-party applications collect and share their private data. We address these shortcomings with TaintDroid, an efficient, system-wide dynamic taint tracking and analysis system capabl... ACM Trans. Comput. Syst., no. 2 (2014)	3921
可视化	D³: Data-Driven Documents Michael Bostock, Vadim Ogievetsky, Jeffrey Heer Data-Driven Documents (D3) is a novel representation-transparent approach to visualization for the web. Rather than hide the underlying scenegraph within a toolkit-specific abstraction, D3 enables direct inspection and manipulation of a native representation: the standard documen... IEEE Trans. Vis. Comput. Graph., no. 12 (2011): 2301-2309	2899
安全与隐私	Towards Evaluating the Robustness of Neural Networks Nicholas Carlini, David A. Wagner Neural networks provide state-of-the-art results for most machine learning tasks. Unfortunately, neural networks are vulnerable to adversarial examples: given an input x and any target classification t, it is possible to find a new input x' that is similar to x but classified... IEEE Symposium on Security and Privacy, (2017)	2834
人机交互	Gamification. using game-design elements in non-gaming contexts Sebastian Deterding, Miguel Sicart, Lennart Nacke, Kenton O'Hara, Dan Dixon "Gamification" is an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement. The recent introduction of 'gamified' applications to large audiences promises new additions to the existing rich and diverse ... CHI Extended Abstracts, pp.2425-2428, (2011)	2209

子领域	论文标题	引用量
芯片技术	Cyber-physical systems: the next computing revolution Ragunathan (Raj) Rajkumar, Insup Lee, Lui Sha, John Stankovic Cyber-physical systems (CPS) are physical and engineered systems whose operations are monitored, coordinated, controlled and integrated by a computing and communication core. Just as the internet transformed how humans interact with one another, cyber-physical systems will transf... Design Automation Conference, pp.731-736, (2010)	1738
计算理论	Efficient Fully Homomorphic Encryption from (Standard) LWE Zvika Brakerski, Vinod Vaikuntanathan We present a fully homomorphic encryption scheme that is based solely on the (standard) learning with errors (LWE) assumption. Applying known results on LWE, the security of our scheme is based on the worst-case hardness of "short vector problems" on arbitrary lattices. Our cons... Foundations of Computer Science, (2011): 97-106	1663
计算机图形	Image smoothing via L0 gradient minimization Li Xu, Cewu Lu, Yi Xu Q., Jiayia Jia We present a new image editing method, particularly effective for sharpening major edges by increasing the steepness of transition while eliminating a manageable degree of low-amplitude structures. The seemingly contradictory effect is achieved in an optimization framework making... international conference on computer graphics and interactive techniques, no. 6 (2011)	1168
知识工程	Semantic sentiment analysis of twitter Hassan Saif, Yulan He, Harith Alani Sentiment analysis over Twitter offer organisations a fast and effective way to monitor the publics' feelings towards their brand, business, directors, etc. A wide range of features and methods for training sentiment classifiers for Twitter datasets have been researched in recent... International Semantic Web Conference (1), pp.508-524, (2012)	580

机器学习领域，影响力排名前 10 论文的引用率都超过万次，且半数以上论文引用率超过 2 万次，如表 4-6 所示。从论文研究主题来看，这 10 篇最高影响力论文全部都是与深度学习相关的。从论文来源来看，这 10 篇最高影响力论文之中，有 5 篇来自 NeurIPS、3 篇来自 ICLR、2 篇来自 ICML。居于首位的是 2015 年 ICLR 会议上由 Google Brain 的 Diederik P. Kingma 和加拿大多伦多大学的助理教授 Jimmy Lei Ba 联合发表的“Adam: A Method for Stochastic Optimization”论文。影响力排名第二位的是 Apple 公司研究员 Ian J. Goodfellow 发表在 NeurIPS 2014 上的一篇论文“Generative Adversarial Nets”，该文也是 GANs 的开山之作。影响力排名第三位的论文是 Facebook 科学家 Tomas Mikolov 发表在 NeurIPS 2013 上的“Distributed Representations of Words and Phrases and their Compositionality”。

表 4-6 2011-2020 年机器学习领域最高影响力论文前十

排名	最高影响力论文的标题	引用量	发表年	作者	来源
1	Adam: A Method for Stochastic Optimization.	58807	2015	D.P. Kingma, Jimmy Lei Ba	ICLR
2	Generative Adversarial Nets	25039	2014	I. J. Goodfellow, Jean Pouget-	NeurIPS

排名	最高影响力论文的标题	引用量	发表年	作者	来源
				Abadie, et al	
3	Distributed Representations of Words and Phrases and their Compositionality	23990	2013	Tomas Mikolov, Ilya Sutskever, et al	NeurIPS
4	Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks	23747	2017	Shaoqing Ren, Kaiming He, Ross B. Girshick, et al	NeurIPS
5	Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.	22547	2015	Sergey Ioffe, Christian Szegedy	ICML
6	Efficient Estimation of Word Representations in Vector Space	18991	2013	T. Mikolov, Kai Chen, et al	ICLR
7	Neural Machine Translation by Jointly Learning to Align and Translate.	15326	2014	D Bahdanau, Kyunghyun Cho, Yoshua Bengio	ICLR
8	Attention Is All You Need	14780	2017	Ashish Vaswani, N. Shazeer, et al	NeurIPS
9	Sequence to Sequence Learning with Neural Networks.	13038	2014	Ilya Sutskever, Oriol Vinyals, Quoc V. Le	NeurIPS
10	Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation	12336	2014	K Cho, B Van Merriënboer, C Gulcehre	EMNLP

计算机视觉领域，最高影响力前十论文如表 4-7 所示。其中，最高引用的论文是发布于 2016 年 CVPR 上、以 Facebook AI Research 何恺明为第一作者的“Deep Residual Learning for Image Recognition”文章，其引用量已超过 6 万。这篇论文也荣获了当年 CVPR Best Paper 奖项，联合作者还包括旷视科技的研究员张祥雨、首席科学家与研究院院长孙剑，以及当时就职于 Momenta 任少卿。该论文是一篇非常经典的神经网络的论文，主要通过构建了一种新的网络结构来解决当网络层数过高之后更深层的网络的效果没有稍浅层网络好的问题，并且做出了适当解释以及用 ResNet 残差网络解决了问题。

表 4-7 2011-2020 年计算机视觉领域最高影响力论文前十

排名	最高影响力论文的标题	引用量	发表年	作者	来源
1	Deep Residual Learning for Image Recognition	61600	2016	Kaiming He, Xiangyu Zhang, et al	CVPR
2	Going Deeper with Convolutions.	25849	2015	Christian Szegedy, Wei Liu, Yangqing Jia, et al	CVPR
3	Fully convolutional networks	19356	2015	Jonathan Long,	CVPR

排名	最高影响力论文的标题	引用量	发表年	作者	来源
	for semantic segmentation			Evan Shelhamer, et al	
4	Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation	15239	2014	Ross B. Girshick, Jeff Donahue, et al	CVPR
5	Microsoft COCO: Common Objects in Context	12556	2014	Tsung-Yi Lin, Michael Maire, Serge Belongie, et al	ECCV
6	Densely Connected Convolutional Networks	12498	2017	Gao Huang, Zhuang Liu, Kilian Q. Weinberger	CVPR
7	You Only Look Once: Unified, Real-Time Object Detection	12295	2016	joseph redmon, Santosh Kumar Divvala, et al	CVPR
8	Fast R-CNN	12060	2015	Ross B. Girshick	ICCV
9	SSD: Single Shot MultiBox Detector	11362	2016	wei liu, dragomir anguelov, et al	ECCV
10	Visualizing and Understanding Convolutional Networks	10795	2014	Matthew D. Zeiler, Rob Fergus	ECCV

自然语言处理领域,过去十年中最高影响力前十论文的具体信息如表 4-8 所示。其中,最高引用量论文是 Google 研究员 Jeffrey Pennington 在 2014 年 EMNLP 会议发表的论文 “Glove: Global Vectors for Word Representation”, 这篇论文提出的单词表示模型是通过仅训练单词-单词共现矩阵中的非零元素, 而不是整个稀疏矩阵或大型语料库中的单个上下文窗口, 来有效地利用统计信息。该模型产生一个具有有意义子结构的向量空间, 在相似性任务和命名实体识别方面优于相关模型。

表 4-8 2011-2020 年自然语言处理领域最高影响力论文前十

排名	最高影响力论文的标题	引用量	发表年	作者	来源
1	Glove: Global Vectors for Word Representation.	17858	2014	J Pennington, R Socher, C. D. Manning	EMNLP
2	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.	12556	2018	J Devlin, M-W Chang, K Lee, K. Toutanova	NAACL
3	Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.	11665	2014	Kyunghyun Cho, Bart van Merriënboer, et al	EMNLP
4	Convolutional Neural Networks for Sentence Classification.	9177	2014	Yoon Kim	EMNLP
5	The Stanford CoreNLP Natural Language Processing Toolkit.	5785	2014	C. D. Manning, M Surdeanu, J	ACL

排名	最高影响力论文的标题	引用量	发表年	作者	来源
				Bauer, J Rose Finkel, et al	
6	Deep contextualized word representations.	5229	2018	Matthew E. Peters, Mark Neumann, et al	NAACL
7	Effective Approaches to Attention-based Neural Machine Translation	4781	2015	Thang Luong, Hieu Pham, Christopher D. Manning	EMNLP
8	Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank	4586	2013	R Socher, A Perelygin, J Wu, et al	EMNLP
9	Linguistic Regularities in Continuous Space Word Representations.	3207	2013	Tomas Mikolov, Wen-tau Yih, Geoffrey Zweig	NAACL
10	Neural Machine Translation of Rare Words with Subword Units	3042	2016	R Sennrich, B Haddow, A Birch	ACL

4.6 AMiner：过去十年十大人工智能研究热点

人工智能研究热点旨在通过 AMiner 全球学术数据库遴选过去十年人工智能领域最热门的研究技术主题。

过去十年人工智能研究热点 TOP10 遴选方法为参考人工智能领域在国际顶级期刊和会议过去 10 年所发表论文，从标题和摘要信息中抽取论文技术研究主题和所在 AI 子领域，并按照该技术主题研究的论文发表数量、论文引用量和该技术主题进入每个会议高引论文前十名的次数（即该技术的领域引用特征）进行综合评测排序，通过 AI 算法计算出该技术的 AMiner 影响力指数，以此获得人工智能领域研究热点的总榜单，评选出过去十年“十大 AI 研究热点”。其中，领域的国际顶级期刊和会议由技术委员会专家确定，共计包括 44 个人工智能领域顶级会议，具体名称详见附录 1。该技术的 AMiner 影响力指数计算公式如下：

$$Z = \sum (w_1 P, w_2 C, w_3 R)$$

其中，Z——该技术的 AMiner 影响力指数；

w_1, w_2, w_3 ——影响力系数；

P ——该技术主题研究的论文发表数量；

C ——该技术主题论文的引用量；

R ——该技术的领域引用特征值，即该技术主题进入每个会议高引论文前十名的次数；

本次评测结果显示，过去十年十大 AI 研究热点分别为：**深度神经网络**、**特征抽取**、**图像分类**、**目标检测**、**语义分割**、**表示学习**、**生成对抗网络**、**语义网络**、**协同过滤**和**机器翻译**。各研究热点名称、相关各项指标及其最终的 AMiner 指数结果如表 4-9 所示。

表 4-9 AMiner 评选出的近十年十大 AI 研究热点

排名	AI 热点名称	主题论文量	主题论文引用量	顶会顶刊高引前十出现次数	AMiner 指数
1	深度神经网络	5405	299729	125	98.16
2	特征抽取	1747	95205	8	21.51
3	图像分类	612	50309	16	14.14
4	目标检测	472	49602	13	12.73
5	语义分割	275	27893	23	12.01
6	表示学习	711	49892	8	11.88
7	生成对抗网络	362	24536	22	11.44
8	语义网络	1192	44897	2	10.60
9	协同过滤	289	36681	12	9.98
10	机器翻译	389	23119	14	8.84

这十个 AI 研究热点的入选理由分别如下。

1. 深度神经网络

深度神经网络是深度学习的基础，又被称为深度前馈网络（DFN）、多层感知机（Multi-Layer perceptron, MLP），可以理解为是有很多隐藏层的神经网络。深度神经网络可以在没有大量标记数据的情况下解决问题。代表算法包括卷积神经网络、循环神经网络、递归神经网络等。深度神经网络的被引用量保持了较长时间的稳定平稳增长；深度神经网络技术则于 2014 年开始获得更多引用。目前，深度神经网络（DNN）是许多人工智能应用的基础，从自动驾驶汽车、癌症检测到大型游戏等。在这许多领域中，DNN 实现了超越人类的准确率。

入选理由

AMiner 数据显示，过去十年中，有 5405 篇以深度神经网络为研究主题的论文在人工智能国际顶会顶刊论文中发表，其总引用量达 299729，并且在这些顶会顶刊论文引用量排名前十的论文中出现过 125 次。该技术的最终指数评分为 98.16，位列过去十年最热门 AI 研究主题之首。

2. 特征抽取

特征抽取 (Feature Extraction) 热门是信息检索与推荐中的一项技术, 专指使用计算机提取一组测量值中属于特征性的信息的方法及过程, 并将所抽取出的有效实体信息进行结构化存储。目前特征抽取已引入机器学习、深度学习、神经网络技术, 其中, 神经网络计算已应用于图片特征抽取。针对某个特定图片, 通过卷积神经网络对图片进行特征抽取得到表征图片的特征, 利用度量学习方法如欧式距离对图片特征进行计算, 对图片距离进行排序, 得到初级检索结果, 再根据图片数据的上下文信息和流形结构对图像检索结果进行重排序, 从而提高图像检索准确率, 得到最终的检索结果。

入选理由

AMiner 数据显示, 过去十年中, 有 1747 篇以特征抽取为研究主题的论文在人工智能国际顶会顶刊论文中发表, 其总引用量达 95205, 并且在这些顶会顶刊论文引用量排名前十的论文中出现过 8 次。该技术的最终评分为 **21.51**, 位列过去十年 AI 研究热点亚军。

3. 图像分类

图像分类 (Image Classification) 是指计算机利用算法从给定的分类集合中给某个特定图像正确分配一个标签的任务, 其目标是将不同的图像划分到不同的类别中, 并实现最小的分类误差, 较多应用于计算机视觉、信息检索与推荐领域。2012 年, 加拿大认知心理学家和计算机科学家 Geoffrey Everest Hinton 的博士生 Alex Krizhevsky 在 ILSVRC 将深度学习用于大规模图像分类中并提出了 CNN 模型, 其计算效果大幅度超越传统方法, 获得了 ILSVRC2012 冠军, 该模型被称作 AlexNet。从 AlexNet 之后, 涌现了一系列 CNN 模型, 不断地在 ImageNet 上刷新成绩。目前的深度学习模型的识别能力已经超过了人眼。

入选理由

AMiner 数据显示, 过去十年中, 有 612 篇以图像分类为研究主题的论文在人工智能国际顶会顶刊论文中发表, 其总引用量达 50309, 并且在这些顶会顶刊论文引用量排名前十的论文中出现过 16 次。该技术的最终评分为 **14.14**, 位列过

去十年最热门 AI 研究主题第三名。

4. 目标检测

目标检测（Object Detection）作为计算机视觉和图像处理领域一个分支，是指利用图像处理与模式识别等领域的理论和方法，检测出数字图像和视频中存在的特定类别的目标对象，确定这些目标对象的语义类别，并标定出目标对象在图像中的位置。对象检测是对象识别的前提，具有很大发展潜力。对象检测已经有许多有用有趣的实际应用，如人脸识别、行人检测、视觉搜索引擎、计数、航拍图像分析等。深度学习模型在图像分类任务中碾压了其他传统方法。很多对象检测的新方法和新应用推动了深度学习最前沿的科技发展。

目前流行的目标检测算法有 One-stage 和 Two-stage 两种。Two-stage 指的是检测算法需要分两步完成，即先获取候选区域、再进行分类与回归，比如 R-CNN 系列算法，准确度较高但是速度慢；One-stage 检测不需要单独寻找候选区域，典型的有 SSD/YOLO，其速度快但是准确性较低。对象检测在过去 10 年内出现了由基于区域的卷积神经网络（R-CNN）、YOLO 分治网络、SSD 等多种方法。这些新技术在速度和精确度上都较先有了很大提升。目前，计算机进行对象检测的能力在很多方面已经超越人类。例如，AI 通过深度学习进行对象检测和识别，已经可以更准确地检测乳腺癌。

入选理由

过去十年中，有 472 篇以目标检测为研究主题的论文在人工智能国际顶会顶刊论文中发表，其总引用量达 49602 次，并且在这些顶会顶刊论文引用量排名前十的论文中出现过 13 次。该技术的最终评分为 12.73，位列最热门 AI 研究主题第四名。

5. 语义分割

语义分割（Semantic Segmentation）是让计算机根据图像的语义进行分割，判断图像中哪些像素属于哪个目标。近年来，许多语义分割问题正在采用深度学习技术来解决，最常见的是卷积神经网络，在精度上大大超过了其他方法以及效率。

目前语义分割的应用领域主要有：地理信息系统、无人车驾驶、医疗影像分析和机器人等领域。

入选理由

过去十年中，有 275 篇以语义分割为研究主题的论文在人工智能国际顶会顶刊论文中发表，其总引用量达 27893 次，并且在这些顶会顶刊论文引用量排名前十的论文中出现过 23 次。该技术的最终评分为 12.01，位列最热门 AI 研究主题第五名。

6. 表示学习

表示学习（Representation Learning），是指将原始数据转换成能够被机器学习的一种深度学习技术。它能够从复杂的原始数据中提炼有效特征，剔除无效或者冗余信息，形成可用的数据表示。在知识表示学习中，词嵌入（Word Embedding）是自然语言处理的重要突破之一，它可以将词表示为实数域向量，进而为机器学习和深度学习提供模型训练的基础。近些年很多专家和学者利用词嵌入的表示学习原理进行相关领域的研究，主要的表示方法包括 Word2Vec、One-hot、词共现等。这些方法已经成为当下人工智能技术应用的基础，为机器学习提供了高效的表示能力。

入选理由

过去十年中，有 711 篇以表示学习为研究主题的论文在人工智能国际顶会顶刊论文中发表，其总引用量达 49892 次，并且在这些顶会顶刊论文引用量排名前十的论文中出现过 8 次。该技术的最终评分为 11.88，位列最具影响力 AI 技术第六名。

7. 生成对抗网络

生成对抗网络（Generative Adversarial Networks, GAN）是用于无监督学习的机器学习模型，由 Ian Goodfellow 等人在 2014 年提出。由神经网络构成判别器和生成器构成，通过一种互相竞争的机制组成的一种学习框架。GAN 功能强大，学习性质是无监督的，也不需要标记数据。传统的生成模型最早要追溯到 80 年

代的 RBM，以及后来逐渐使用深度神经网络进行包装的 AutoEncoder，然后就是现在的生成模型 GAN。GAN 具有大量的实际用例，如图像生成、艺术品生成、音乐生成和视频生成。此外，它还可以提高图像质量，并且完成图像风格化或着色、面部生成以及其他更多有趣的任务。

入选理由

过去十年在人工智能国际顶会顶刊论文中发表有 362 篇以生成对抗网络为研究主题的论文，其总引用量达 24536 次，并且在这些顶会顶刊论文引用量排名前十的论文中出现过 22 次。该技术的最终评分为 11.44，位列最热门 AI 研究主题第七名。

8 语义网络

语义网络 (Semantic Network) 是一种以网络格式表达人类知识构造的形式，是人工智能程序运用的表示方式之一，相关研究主要集中在信息检索与推荐、知识工程领域。语义网络是一种面向语义的结构，它们一般使用一组推理规则，规则是为了正确处理出现在网络中的特种弧而专门设计的。语义网络可以深层次地表示知识，包括实体结构、层次及实体间的因果关系；无推理规律可循；知识表达的自然性可以直接从语言语句强化而来。

入选理由

过去十年在人工智能国际顶会顶刊论文中有 1192 篇以语义网络为研究主题的论文发表，总引用量达 44897 次，并且在这些顶会顶刊论文引用量排名前十的论文中出现过 2 次。语义网络技术的最终评分为 10.60，位列最热门 AI 研究主题第八名。

9. 协同过滤

协同过滤 (CF) 是推荐系统使用的一种技术，通过收集来自多个用户的偏好、兴趣、评价标准等用户行为数据来过滤信息，并自动预测 (过滤) 用户兴趣的方法，为用户提供有针对性的推荐及其所需信息。大多数协同过滤系统都应用基于相似度索引的技术。协同过滤是解决信息超载问题的一个有效办法。无论是基于

用户-用户的协同过滤，还是项目-项目的协同过滤，都有效地提高了用户信息的使用效率。

入选理由

过去十年在人工智能国际顶会顶刊论文中有 289 篇以协同过滤为研究主题的论文，其总引用量达 36681 次，并且在这些顶会顶刊论文引用量排名前十的论文中出现过 12 次。该技术的最终评分为 9.98，位列最热门 AI 研究主题第九名。

10. 机器翻译

机器翻译（Machine Translation），又称为自动翻译，是利用计算机把一种自然源语言转变为另一种自然目标语言的过程，通常指自然语言之间句子和全文的翻译。它是自然语言处理（Natural Language Processing）的一个分支，与计算语言学（Computational Linguistics）、自然语言理解（Natural Language Understanding）之间存在密不可分的关系。机器翻译是人工智能的终极目标之一，其核心语言理解和语言生成是自然语言处理的两大基本问题。近几年来，随着深度学习技术的发展，神经机器翻译取得了巨大进展，其生成的译文接近自然语句，成为了主流语言学习模型。

入选理由

过去十年在人工智能国际顶会顶刊论文中发表有 389 篇以机器翻译为研究主题的论文，其总引用量达 23119 次，并且在这些顶会顶刊论文引用量排名前十的论文中出现过 14 次。该技术的最终评分为 8.84，成为 AI 研究热点第十名。

5 人工智能子领域研究现状与趋势

5.1 机器学习

机器学习是利用计算机模拟人的学习能力,从样本数据中学习得到知识和经验,然后用于实际的推断和决策。机器学习是一门多学科交叉专业,涵盖概率论知识、统计学知识、近似理论知识和复杂算法知识,使用计算机作为工具并致力于真实实时的模拟和实现人类学习方式,以获取新的知识或技能,并将现有内容进行知识结构划分来有效提高学习效率。机器学习推动人工智能快速发展,是第三次人工智能发展浪潮的重要推动因素。

5.1.1 机器学习概念

机器学习是现阶段解决很多人工智能问题的主流方法,是现代人工智能的本质^[18],目前正处于高速发展之中。它从概念诞生到技术的普遍应用经过了漫长的过程。机器学习领域诞生了众多的经典理论(如 PAC 学习理论、决策树、支持向量机 SVM、Adaboost、循环神经网络 RNN 和 LSTM、流形学习、随机森林 Random Forest 等),并走向实用。

在机器学习发展过程中,众多优秀学者为推动机器学习的发展做出了巨大的贡献。1950 年图灵在关于图灵测试的文章中就已提及机器学习的概念。1952 年,IBM 的亚瑟·塞缪尔(Arthur Samuel,被誉为“机器学习之父”)设计了一款可以学习的西洋跳棋程序,并在 1956 年正式提出了“机器学习”这一概念,认为“机器学习是在不直接针对问题进行编程的情况下,赋予计算机学习能力的一个研究领域”。有着“全球机器学习教父”之称的 Tom Mitchell 则将机器学习定义为:对于某类任务 T 和性能度量 P,如果计算机程序在 T 上以 P 衡量的性能随着经验 E 而自我完善,就称这个计算机程序从经验 E 学习。以上这些定义都比较简单抽象。随着时间的变迁,机器学习的内涵和外延不断变化。

如今普遍认为,机器学习(Machine Learning,简称 ML)的处理系统和算法

[18] 张志华,机器学习的发展历程及启示,中国计算机学会通讯,第 12 卷,第 11 期,2016 年 11 月

是主要通过找出数据里隐藏的模式进而做出预测的识别模式，它是人工智能（Artificial Intelligence，简称 AI）的一个重要子领域，而人工智能又与更广泛的数据挖掘（Data Mining，简称 DM）和知识发现（Knowledge Discovery in Database，简称 KDD）领域相交叉。为了更好地理解和区分人工智能、机器学习、数据挖掘、模式识别（Pattern Recognition）、统计（Statistics）、神经计算（Neuro Computing）、数据库（Databases）、知识发现等概念，特绘制其交叉关系如图 5-1 所示。

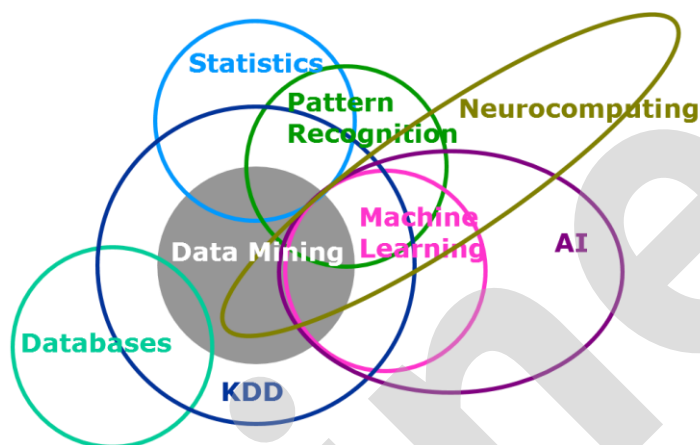


图 5-1 机器学习相关概念的辨识

机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。机器学习过程可以参见图 5-2 的表示。

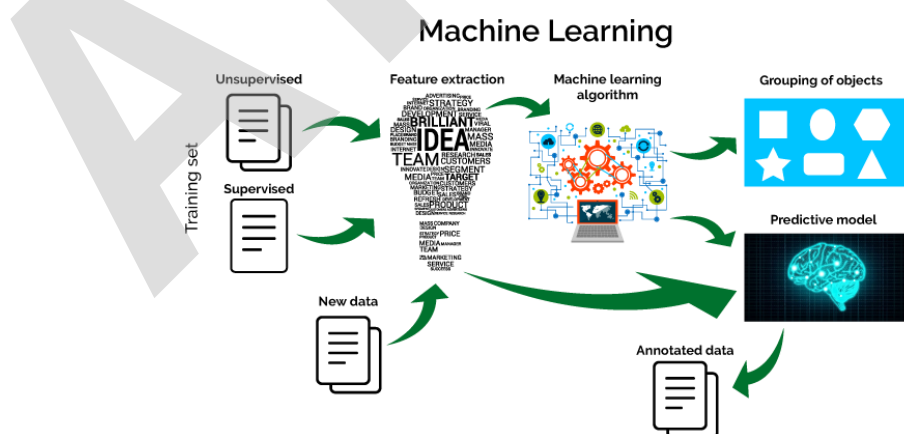


图 5-2 机器学习基本过程

5.1.2 机器学习过去十年主要进展

从上世纪 50 年代的图灵测试提出、塞缪尔开发的西洋跳棋程序，60 年代中到 70 年代末的发展几乎停滞，到 80 年代使用神经网络反向传播（BP）算法训练的多参数线性规划（MLP）理念以及昆兰在 1986 年提出的“决策树”（ID3 算法），再到后来的支持向量机（SVM）算法以及 2006 年辛顿（Hinton）提出深度学习（Deep Learning），机器学习研究已走过低迷进入蓬勃发展期。

近十年来，机器学习研究热度高涨，主要集中在深度学习。特别是在 2012 年之后，随着计算能力的进步和海量训练样本的支持，深度学习^{[19],[20]}（Deep Learning）技术走向前台并急速发展，较好地解决了人工智能领域的一些重点问题如图像分析和语音识别等，并带动了产业界的广泛应用。2015 年，为纪念人工智能概念提出 60 周年，学者 LeCun、Bengio 和 Hinton 推出了深度学习的联合综述。深度学习可以让那些拥有多个处理层的计算模型来学习具有多层次抽象的数据的表示，这些方法在许多方面都带来了显著的改善。深度学习的出现，让图像、语音等感知类问题取得了真正意义上的突破，离实际应用已如此之近，将人工智能推进到一个新时代。

5.1.3 机器学习经典算法

机器学习算法可以按照不同的标准来进行分类。比如按函数 $f(x, \theta)$ 的不同，机器学习算法可以分为线性模型和非线性模型；按照学习准则的不同，机器学习算法也可以分为统计方法和非统计方法。

按照训练样本提供的信息以及反馈方式的不同，机器学习算法可以分为以下几类。

- 监督学习（Supervised Learning）

监督学习中的数据集是有标签的，对于给出的样本是有答案的。如果机器学

[19] G.E.Hinton, et al. Reducing the Dimensionality of Data with Neural Networks, Science 313, 504(2006)

[20] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet Classification with Deep Convolutional Neural Networks[C]. Neural information processing systems, 2012: 1097-1105.

习的目标是通过建模样本的特征 x 和标签 y 之间的关系： $f(x, \theta)$ 或 $p(y|x, \theta)$ ，并且训练集中每个样本都有标签，那么这类机器学习称为监督学习。根据标签类型的不同，又可以将其分为分类问题和回归问题两类。前者是预测某一样东西所属的类别（离散的），比如给定一个人的身高、年龄、体重等信息，然后判断性别、是否健康等；后者则是预测某一样本所对应的实数输出（连续的），比如预测某一地区人的平均身高。大部分模型都是属于监督学习，包括线性分类器、支持向量机等。常见的监督学习算法有： k -近邻算法（ k -Nearest Neighbors, kNN）、决策树（Decision Trees）、朴素贝叶斯（Naive Bayesian）等。监督学习的基本流程如图 5-3 所示。

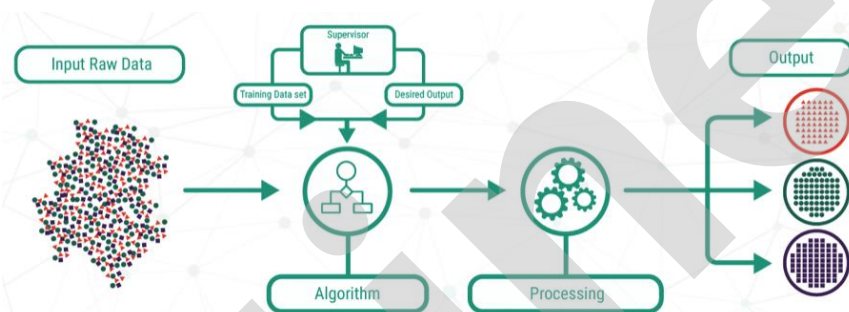


图 5-3 监督学习的基本流程

● 无监督学习 (Unsupervised Learning, UL)

跟监督学习相反，无监督学习中数据集是完全没有标签的，依据相似样本在数据空间中一般距离较近这一假设，将样本分类。

常见的无监督学习算法包括：稀疏自编码 (Sparse Auto-Encoder)、主成分分析 (Principal Component Analysis, PCA)、K-Means 算法 (K 均值算法)、DBSCAN 算法 (Density-Based Spatial Clustering of Applications with Noise)、最大期望算法 (Expectation-Maximization algorithm, EM) 等。

利用无监督学习可以解决的问题可以分为关联分析、聚类问题和维度约减。

关联分析是指发现不同事物之间同时出现的概率。它被广泛地应用在购物篮分析中。如果发现买面包的客户有百分之八十的概率买鸡蛋，那么商家就会把鸡蛋和面包放在相邻的货架上。

聚类问题是指将相似的样本划分为一个簇 (Cluster)。与分类问题不同，聚

类问题预先并不知道类别，自然训练数据也没有类别的标签。

维度约减是指减少数据维度的同时保证不丢失有意义的信息。利用特征提取方法和特征选择方法，可以达到维度约减的效果。特征选择是指选择原始变量的子集。特征提取是将数据从高维度转换到低维度。广为人知的主成分分析算法就是特征提取的方法。

非监督学习的基本处理流程如图 5-4 所示。

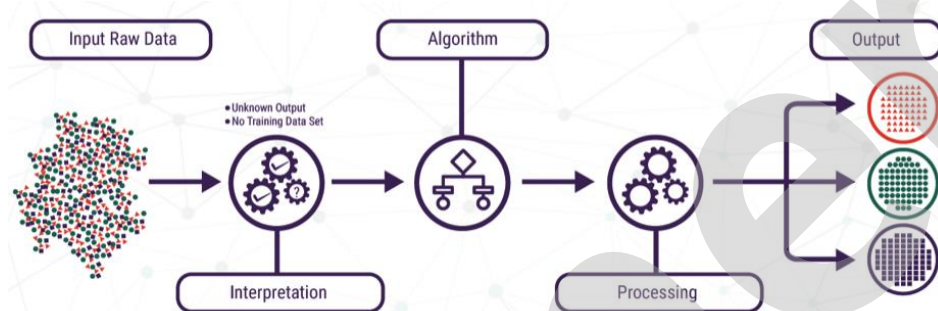


图 5-4 非监督学习的基本流程

可以看到相对于监督学习，非监督学习的过程中没有监督者（Supervisor）的干预。图 5-5 是一个典型的监督学习和非监督学习的对比，左侧是对一群有标签数据的分类，而右侧是对一群无标签数据的聚类。

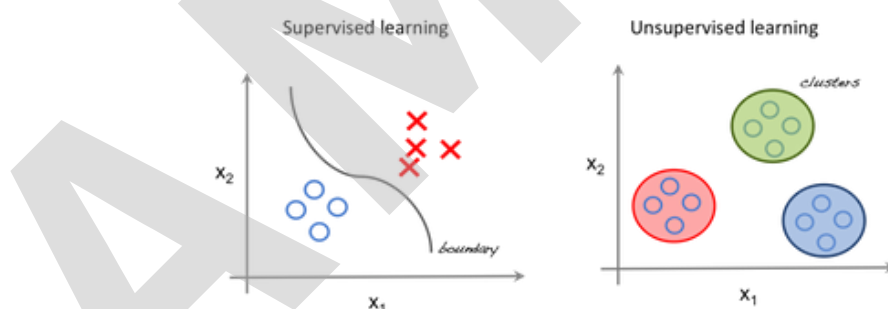


图 5-5 一个典型的监督学习和非监督学习对比

无监督学习在过去十年间的发展主要集中在生成模型（Generative Model）和表征学习（Representation Learning）两个方面。对于生成模型，变分自编码器（Variational Autoencoder, VAE）和生成式对抗网络（Generative Adversarial Nets, GAN）相继成为无监督复杂概率分布学习的主流方法，并由此衍生和发展出大量的无监督生成模型。对于表征学习，以双向生成式对抗网络（BiGAN）、深度聚

集（DeepCluster）、逆约束自编码器插值（ACAI）和动量对比（MoCo）为代表的方法充分结合了深度神经网络强大的特征表示能力，利用无监督数据进行表征学习，为后续更复杂多样的监督学习任务打下了坚实的数据表征基础。

此外，自监督学习（Self-supervised Learning）近年来也逐渐成为无监督表征学习一种新思路，它试图通过根据数据本身的结构或特性人为地构造标签，从而实现监督式的模型学习。

● 半监督学习（Semi-Supervised Learning）

半监督学习是监督学习与无监督学习相结合的一种学习方法。半监督学习一般针对的问题是数据量大，但是有标签数据少或者说标签数据的获取很难很贵的情况，训练的时候有一部分是有标签的，而有一部分是没有的。与使用所有标签数据的模型相比，使用训练集的训练模型在训练时可以更为准确，而且训练成本更低。常见的两种半监督的学习方式是直推学习（Transductive learning）和归纳学习（Inductive learning）。

直推学习（Transductive learning）：没有标记的数据是测试数据，可以用测试的数据进行训练。需要注意，这里只是用了测试数据中的特征（Feature）而没有用标签（Label），所以并不是一种欺骗的方法。

归纳学习（Inductive learning）：没有标签的数据不是测试集。

半监督学习的基本流程如图 5-6 所示：

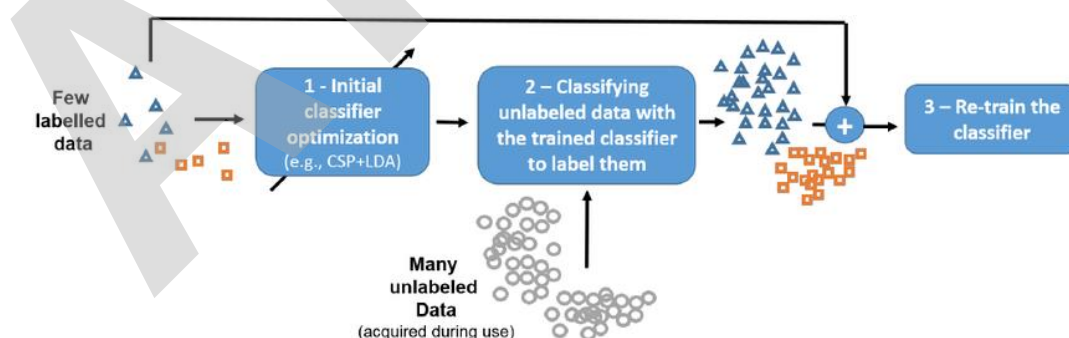


图 5-6 半监督学习的基本流程

监督学习、半监督学习和非监督学习之间的区别如图 5-7 所示。

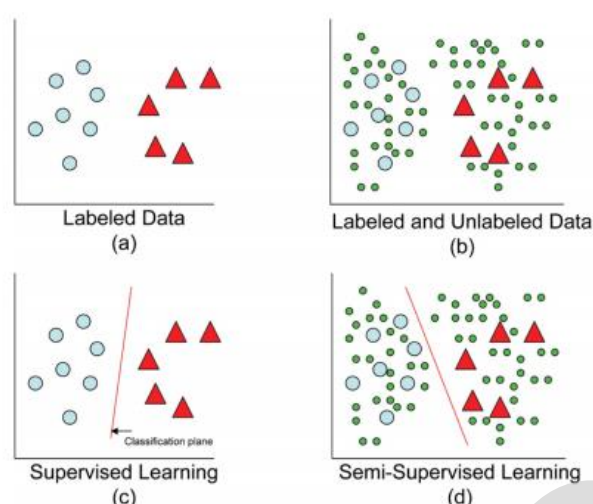


图 5-7 监督学习、半监督学习和非监督学习的简单对比

可以看到，图 5-7（a）中，红色三角形数据和蓝色圆点数据为标注数据；图 5-7（b）中，绿色的小圆点为非标注数据。图 5-7（c）显示监督学习将有标签的数据进行分类；而半监督学习如图 5-7（d）中部分是有标签的，部分是没有标签的，一般而言，半监督学习侧重于在有监督的分类算法中加入无标记样本来实现半监督分类。

近年来，在半监督学习方面，以自训练(Self-training)、协同训练(Co-training)、基于图的标签传播(Graph-based Label Propagation)为代表的经典半监督学习思想与深度学习模型紧密融合，形成了诸多具有代表性的深度半监督学习模型或方法。对于归纳式学习(Inductive Learning)，提出了伪标签(Pseudo-Label)方法对深度神经网络进行自训练，设计了经典的端到端半监督深度模型 LadderNet，提出了自集成(Self-Ensembling)和平均教师(Mean Teacher)半监督深度模型来进一步提高伪标签的质量。对于直推式学习(Transductive Learning)，图卷积网络(Graph Convolutional Networks, GCN)的提出将传统的基于图的半监督学习研究往前大大推进了一步，而最近的图学习卷积网络 GLCN 和图过滤卷积网络 IGCN 则对 GCN 进行了优化，进一步提升了半监督学习的性能。

● 强化学习(Reinforcement Learning, RL)

强化学习从动物学习、参数扰动自适应控制等理论发展而来，基本原理是：

如果 Agent 的某个行为策略导致环境正的奖赏（强化信号），那么 Agent 以后产生这个行为策略的趋势便会加强。Agent 的目标是在每个离散状态发现最优策略以使期望的折扣奖赏和最大。

强化学习在机器人学科中被广泛应用。在与障碍物碰撞后，机器人通过传感器收到负面的反馈从而学会去避免冲突。在视频游戏中，可以通过反复试验采用一定的动作，获得更高的分数。Agent 能利用回报去理解玩家最优的状态和当前应该采取的动作。

图 5-8 采用一只老鼠来模拟强化学习中的 Agent，其任务是走出迷宫，每走一步都有一个方法来衡量其走的好与坏，基本学习过程是当其走得好的时候就给予一定的奖励（如一块蛋糕）。通过这种方式，Agent 在行动评价的环境中获得知识，改进行动方案以适应环境。

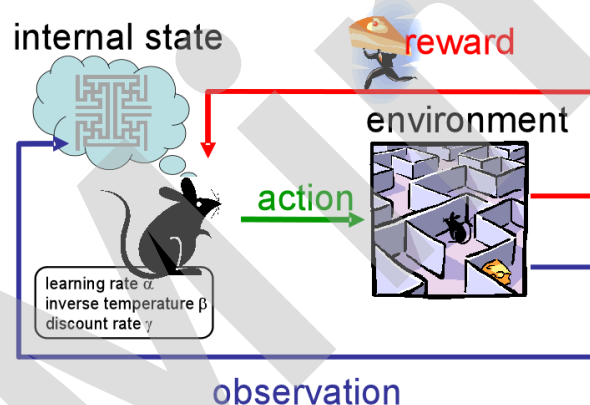
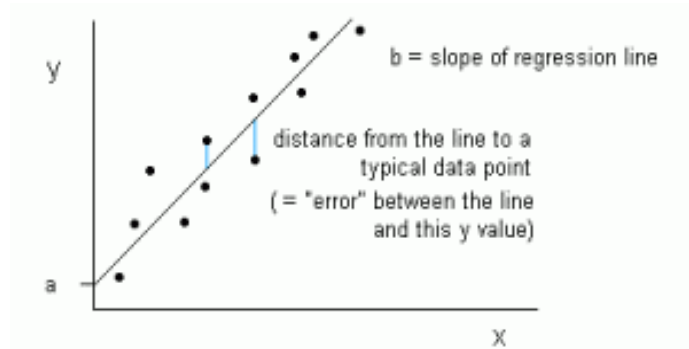


图 5-8 强化学习的基本学习流程

部分机器学习代表算法介绍如下。

- 线性回归

在机器学习中，有一组输入变量 (x) 用于确定输出变量 (y)。输入变量和输出变量之间存在某种关系，机器学习的目标是量化这种关系。

图 5-9 数据集的绘制 x 和 y 值

在线性回归中，输入变量 (x) 和输出变量 (y) 之间的关系表示为 $y = ax + b$ 的方程。因此，线性回归的目标是找出系数 a 和 b 的值。这里， a 是直线的斜率， b 是直线的截距。图 5-9 显示了数据集的 x 和 y 值，线性回归的目标是拟合最接近大部分点的线。

- 分类与回归树 (CART)

CART 是决策树的一个实现方式，由 ID3, C4.5 演化而来，是许多基于树的 bagging、boosting 模型的基础。CART 可用于分类与回归。

CART 是在给定输入随机变量 x 条件下输出随机变量 y 的条件概率分布，与 ID3 和 C4.5 的决策树所不同的是，ID3 和 C4.5 生成的决策树可以是多叉的，每个节点下的叉数由该节点特征的取值种类而定，比如特征年龄分为(青年, 中年, 老年)，那么该节点下可分为 3 叉。而 CART 为假设决策树为二叉树，内部结点特征取值为“是”和“否”。左分支取值为“是”，右分支取值为“否”。这样的决策树等价于递归地二分每一个特征，将输入空间划分为有限个单元，并在这些单元上预测概率分布，也就是在输入给定的条件下输出条件概率分布。

- 随机森林 (Random Forest)

随机森林指的是利用多棵决策树对样本进行训练并预测的一种分类器。它包含多个决策树的分类器，并且其输出的类别是由个别树输出的类别的众数而定。随机森林是一种灵活且易于使用的机器学习算法，即便没有超参数调优，也可以在大多数情况下得到很好的结果。随机森林也是最常用的算法之一，因为它很简易，既可用于分类也能用于回归。

其基本的构建算法过程如下：

1. 用 N 来表示训练用例（样本）的个数， M 表示特征数目。
2. 输入特征数目 m ，用于确定决策树上一个节点的决策结果；其中 m 应远小于 M 。
3. 从 N 个训练用例（样本）中以有放回抽样的方式，取样 N 次，形成一个训练集（即 **bootstrap** 取样），并用未抽到的用例（样本）作预测，评估其误差。
4. 对于每一个节点，随机选择 m 个特征，决策树上每个节点的决策都是基于这些特征确定的。根据这 m 个特征，计算其最佳的分裂方式。
5. 每棵树都会完整成长而不会剪枝，这有可能在建完一棵正常树状分类器后被采用）。

一个简单的随机森林算法示意如下：

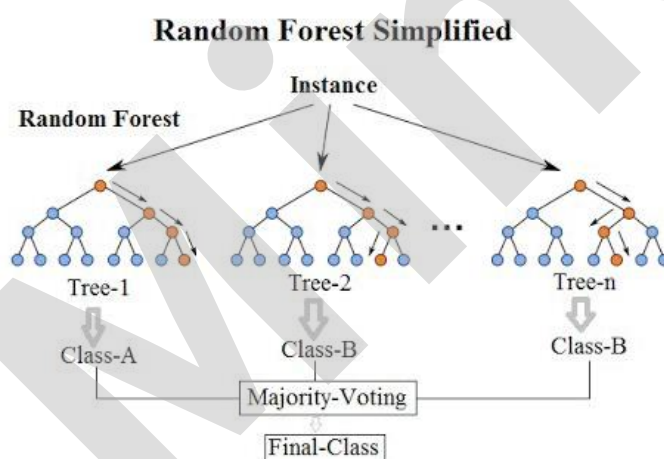


图 5-10 一个简单的随机森林算法示意

随机森林集成了所有的分类投票结果，将投票次数最多的类别指定为最终的输出，这就是一种最简单的 **Bagging** 思想。

● 逻辑回归

逻辑回归最适合二进制分类（ $y = 0$ 或 1 的数据集，其中 1 表示默认类）例如：在预测事件是否发生时，发生的事件被分类为 1 。在预测人会生病或不生病，生病的实例记为 1 ）。它是以其中使用的变换函数命名的，称为逻辑函数 $h(x) = 1 / (1 + e^{-x})$ ，它是一个 S 形曲线。

在逻辑回归中，输出是以缺省类别的概率形式出现的。因为这是一个概率，所以输出在 0-1 的范围内。输出（ y 值）通过对数转换 x 值，使用对数函数 $h(x) = 1 / (1 + e^{-x})$ 来生成。然后应用一个阈值来强制这个概率进入二元分类。

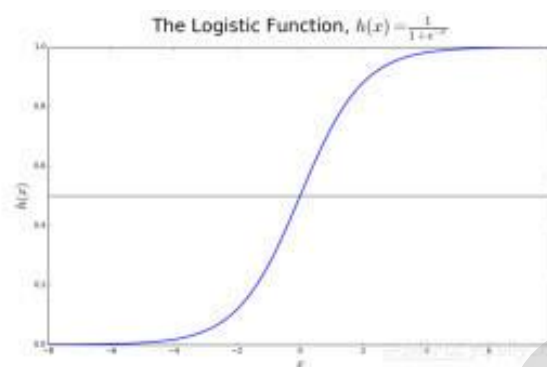


图 5-11 逻辑函数曲线图

图 5-11 判断了肿瘤是恶性还是良性。默认变量是 $y = 1$ （肿瘤=恶性）； x 变量可以是肿瘤的信息，例如肿瘤的尺寸。如图所示，逻辑函数将数据集的各种实例的 x 值转换成 0 到 1 的范围。如果概率超过阈值 0.5（由水平线示出），则将肿瘤分类为恶性。

$$p(x) = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

$$\log\left(\frac{p(x)}{1-p(x)}\right) = b_0 + b_1 x$$

逻辑回归的目标是使用训练数据来找到系数 b_0 和 b_1 的值，以使预测结果与实际结果之间的误差最小化。这些系数是使用最大似然估计来计算的。

● 朴素贝叶斯（Naive Bayesian）

朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法。朴素贝叶斯分类器基于一个简单的假定：给定目标值时属性之间相互条件独立。

通过以上定理和“朴素”的假定，可知：

$$P(\text{Category} | \text{Document}) = P(\text{Document} | \text{Category}) * P(\text{Category}) / P(\text{Document})$$

朴素贝叶斯的基本方法：在统计数据的基础上，依据条件概率公式，计算当前特征的样本属于某个分类的概率，选择最大的概率分类。

对于给出的待分类项，求解在此项出现的条件下各个类别出现的概率，哪个最大，就认为此待分类项属于哪个类别。其计算流程表述如下：

(1) $x = \{a_1, a_2, \dots, a_m\}$ 为待分类项，每个 a_i 为 x 的一个特征属性

(2) 有类别集合 $C = \{y_1, y_2, \dots, y_n\}$

(3) 计算 $P(y_1|x), P(y_2|x), \dots, P(y_n|x)$

(4) 如果 $P(y_k|x) = \max\{P(y_1|x)\}$

● k 最近邻 (kNN)

kNN (k -Nearest Neighbor) 的核心思想是如果一个样本在特征空间中的 k 个最相邻的样本中的大多数属于某一个类别，则该样本也属于这个类别，并具有这个类别上样本的特性。该方法在确定分类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别。kNN 方法在做类别决策时，只与极少量的相邻样本有关。由于 kNN 方法主要靠周围有限的邻近的样本，而不是靠判别类域的方法来确定所属类别的，因此对于类域的交叉或重叠较多的待分样本集来说，kNN 方法较其他方法更为适合。

kNN 算法不仅可以用于分类，还可以用于回归。通过找出一个样本的 k 个最近邻居，将这些邻居的属性的平均值赋给该样本，就可以得到该样本的属性。如图 5-12 是 kNN 算法中， k 等于不同值时的算法分类结果：

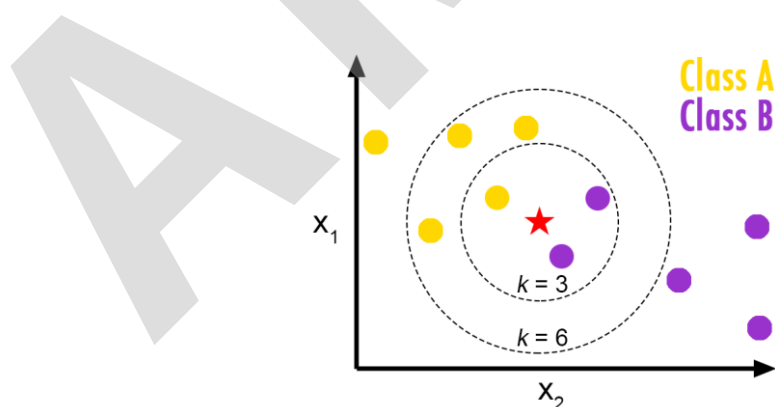


图 5-12 kNN 算法简单示例

简单来说，kNN 可以看成：有一堆已经知道分类的数据，然后当一个新数据进入的时候，就开始跟训练数据里的每个点求距离，然后挑离这个训练数据最近

的 k 个点，看看这几个点属于什么类型，然后用少数服从多数的原则，给新数据归类。

● AdaBoost

Adaptive Boosting 或称为 AdaBoost，是多种学习算法的融合。它是一种迭代算法，其核心思想是针对同一个训练集训练不同的分类器(弱分类器)，然后把把这些弱分类器集合起来，构成一个更强的最终分类器(强分类器)。其算法本身是通过改变数据分布来实现的，它根据每次训练集之中每个样本的分类是否正确，以及上次的总体分类的准确率，来确定每个样本的权值。将修改过权值的新数据集送给下层分类器进行训练，然后将每次训练得到的分类器融合起来，作为最终的决策分类器。

AdaBoost 是最常用的算法。它可用于回归或者分类算法。相比其他机器学习算法，它克服了过拟合的问题，通常对异常值和噪声数据敏感。为了创建一个强大的复合学习器，AdaBoost 使用了多次迭代。因此，它又被称为“Adaptive Boosting”。通过迭代添加弱学习器，AdaBoost 创建了一个强学习器。一个新的弱学习器加到实体上，并且调整加权向量，作为对前一轮中错误分类的样例的回应。得到的结果，是一个比弱分类器有更高准确性的分类器。

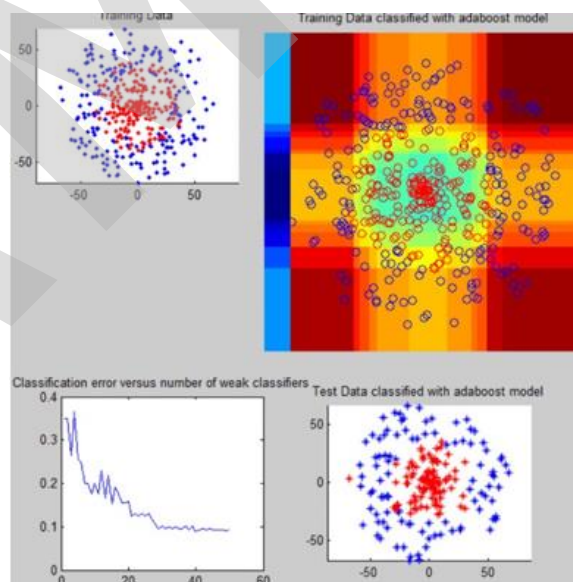


图 5-13 AdaBoost 执行

AdaBoost 有助于将弱阈值的分类器提升为强分类器。图 5-13 描述了

AdaBoost 的执行，只用了简单易于理解的代码在一个文件中就实现了。这个函数包含一个弱分类器和 boosting 组件。弱分类器在一维的数据中尝试去寻找最理想的阈值来将数据分离为两类。boosting 组件迭代调用分类器，经过每一步分类，它改变了错误分类示例的权重。因此，创建了一个级联的弱分类器，它的行为就像一个强分类器。

目前，对 Adaboost 算法的研究以及应用大多集中于分类问题，同时近年也出现了一些在回归问题上的应用。Adaboost 系列主要解决了：两类问题、多类单标签问题、多类多标签问题、大类单标签问题和回归问题。它用全部的训练样本进行学习。

● K-均值算法（K-Means）

K-均值是著名聚类算法，它找出代表聚类结构的 k 个质心。如果有一个点到某一质心的距离比其他质心都近，这个点则指派到这个最近的质心所代表的簇。依次，利用当前已聚类的数据点找出一个新质心，再利用质心给新的数据指派一个簇。

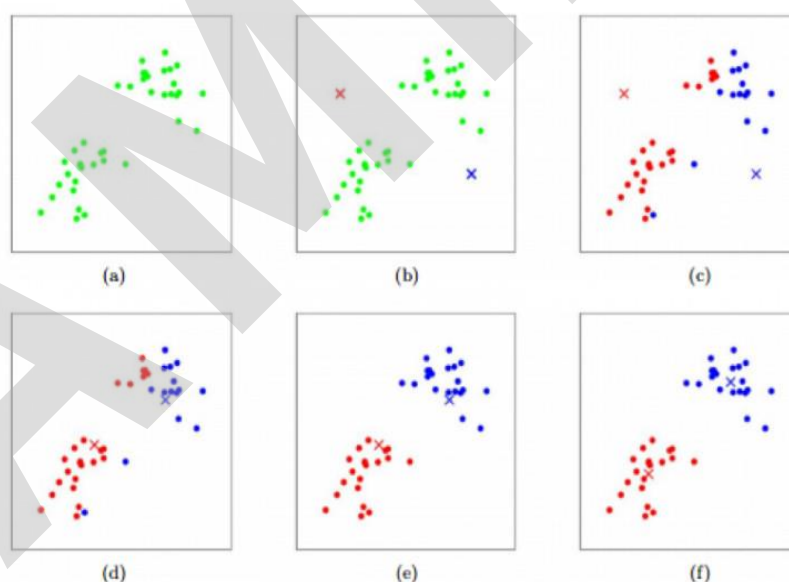


图 5-14 K-均值算法图示

注：图 5-14 中用“x”表示聚类质心，用点表示训练样本：（a）原始数据集；（b）随机初始化聚类质心；（c-f）k-均值迭代 2 次的示意图。

在每次迭代中每个训练样例都被指派到一个最近的聚类质心,每个聚类质心被移动到分配给它的点的平均值的位罝。

- 支持向量机 (SVM)

支持向量机 (Support Vector Machine, SVM) 是一类按监督学习 (Supervised Learning) 方式对数据进行二元分类 (Binary Classification) 的广义线性分类器 (Generalized Linear Classifier), 其决策边界是对学习样本求解的最大边距超平面 (Maximum-Margin Hyperplane)。基本思想是找到集合边缘上的若干数据 (称为支持向量 (Support Vector)), 用这些点找出一个平面 (称为决策面), 使得支持向量到该平面的距离最大。

由简至繁的 SVM 模型包括:

- ◆ 当训练样本线性可分时, 通过硬间隔最大化, 学习一个线性可分支持向量机;
- ◆ 当训练样本近似线性可分时, 通过软间隔最大化, 学习一个线性支持向量机;
- ◆ 当训练样本线性不可分时, 通过核技巧和软间隔最大化, 学习一个非线性支持向量机;

在分类问题中, 很多时候有多个解, 如图 5-15 左侧所示, 在理想的线性可分的情况下其决策平面会有多个。而 SVM 的基本模型是在特征空间上找到最佳的分离超平面使得训练集上正负样本间隔最大, SVM 算法计算出来的分界会保留对类别最大的间距, 即有足够的余量, 如图 5-15 右侧所示。

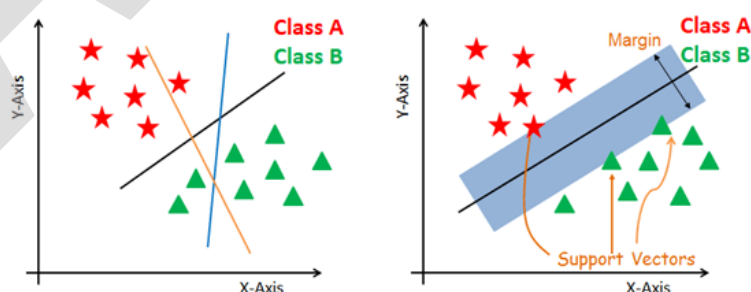


图 5-15 SVM 的决策平面

在解决线性不可分问题时，它可以通过引入核函数，巧妙地解决了在高维空间中的内积运算，从而很好地解决了非线性分类问题。如图 5-16 所示，通过核函数的引入，将线性不可分的数据映射到一个高维的特征空间内，使得数据在特征空间内是可分的。

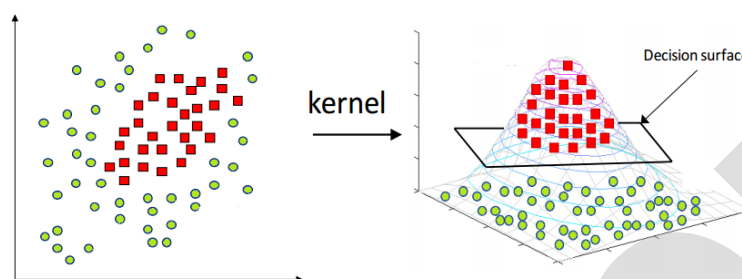


图 5-16 SVM 的核函数

● 人工神经网络 ANN (Artificial Neural Network)

人工神经网络 ANN (Artificial Neural Network) 是由大量处理单元互联组成的非线性、自适应信息处理系统。它是一种模仿动物神经网络行为特征，进行分布式并行信息处理的算法数学模型。其基本过程可以概述如下：外部刺激通过神经末梢，转化为电信号，传导到神经细胞（又叫神经元）；无数神经元构成神经中枢；神经中枢综合各种信号，做出判断；人体根据神经中枢的指令，对外部刺激做出反应。其过程表述如图 5-17 所示：

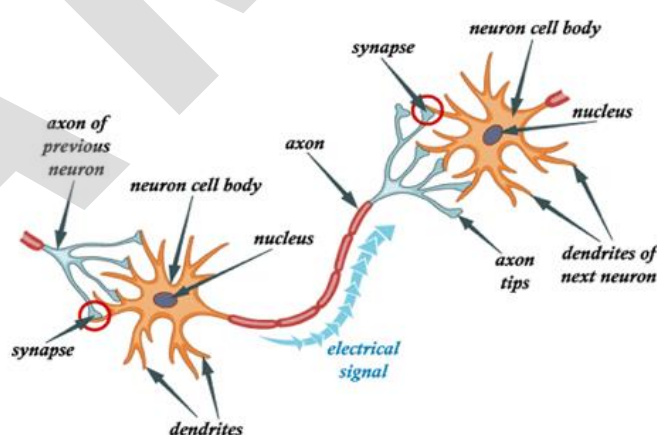


图 5-17 神经网络的传导过程

人工神经网络经历了漫长的发展阶段。最早是上个世纪六十年代提出的“人

造神经元”模型，叫做“感知器”（Perceptron）。感知机模型是机器学习二分类问题中的一个非常简单的模型。它的基本结构如图 5-18 所示。

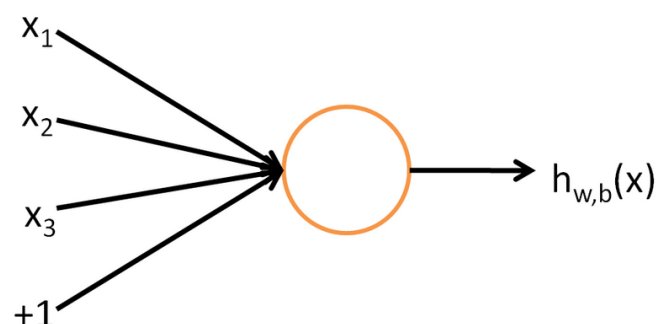


图 5-18 感知机

随着反向传播算法、最大池化（Max-Pooling）等技术的发明，神经网络进入了飞速发展的阶段。神经网络就是将许多个单一“神经元”联结在一起，这样，一个“神经元”的输出就可以是另一个“神经元”的输入。典型的人工神经网络具有以下三个部分：

结构（Architecture）指定了网络中的变量和它们的拓扑关系。

激励函数（Activity Rule）大部分神经网络模型具有一个短时间尺度的动力学规则，来定义神经元如何根据其他神经元的活动来改变自己的激励值。

学习规则（Learning Rule）指定了网络中的权重如何随着时间推进而调整。

一个典型的人工神经网络结构如图 5-19 所示：

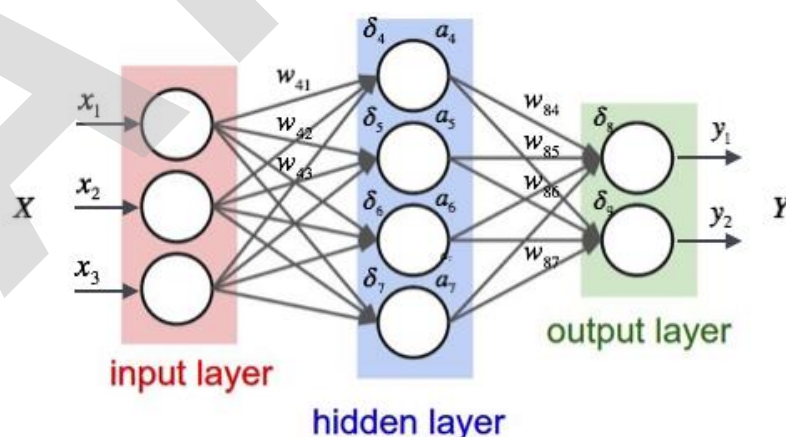


图 5-19 典型的人工神经网络结构

人工神经网络具有四个基本特征：非线性、非局限性、非常定性和非凸性。

人工神经网络的特点和优越性，主要表现在三个方面：具有自学习功能、具有联想存储功能和具有高速寻找最优解的能力。

5.1.4 深度学习主要模型

过去几年里，机器学习里最大放异彩的领域莫过于深度学习。深度学习的基础模型，承担了将原始数据映射到特征空间的任务，克服了传统机器学习依赖人工特征的缺陷，是实现端到端学习的关键。

深度学习模型的发展可以追溯到 1958 年研究认知的心理学家 Frank 发明的感知机（Perceptron）。这在当时掀起一股热潮。后来 Marvin Minsky（人工智能大师）和 Seymour Papert 发现感知机的缺陷：不能处理异或回路等非线性问题，以及当时存在计算能力不足以处理大型神经网络的问题，于是整个神经网络的研究进入停滞期。

神经网络研究领域领军者辛顿（Hinton）在 2006 年提出了神经网络深度学习（Deep Learning）算法，使神经网络的能力大大提高。深度学习可以让那些拥有多个处理层的计算模型来学习具有多层次抽象的数据的表示，这些方法在许多方面都带来了显著改善。深度学习的出现，让图像、语音等感知类问题取得了真正意义上的突破，将人工智能推进到一个新时代。

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，三位教授（Geoffrey Hinton、Yann Lecun、Yoshua Bengio）因此同获图灵奖。

总体来说，深度学习主要经历了 4 条发展脉络：以前向网络为代表的深度学习、以自学习自编码为代表的学习、自循环神经网络算法和强化学习算法。

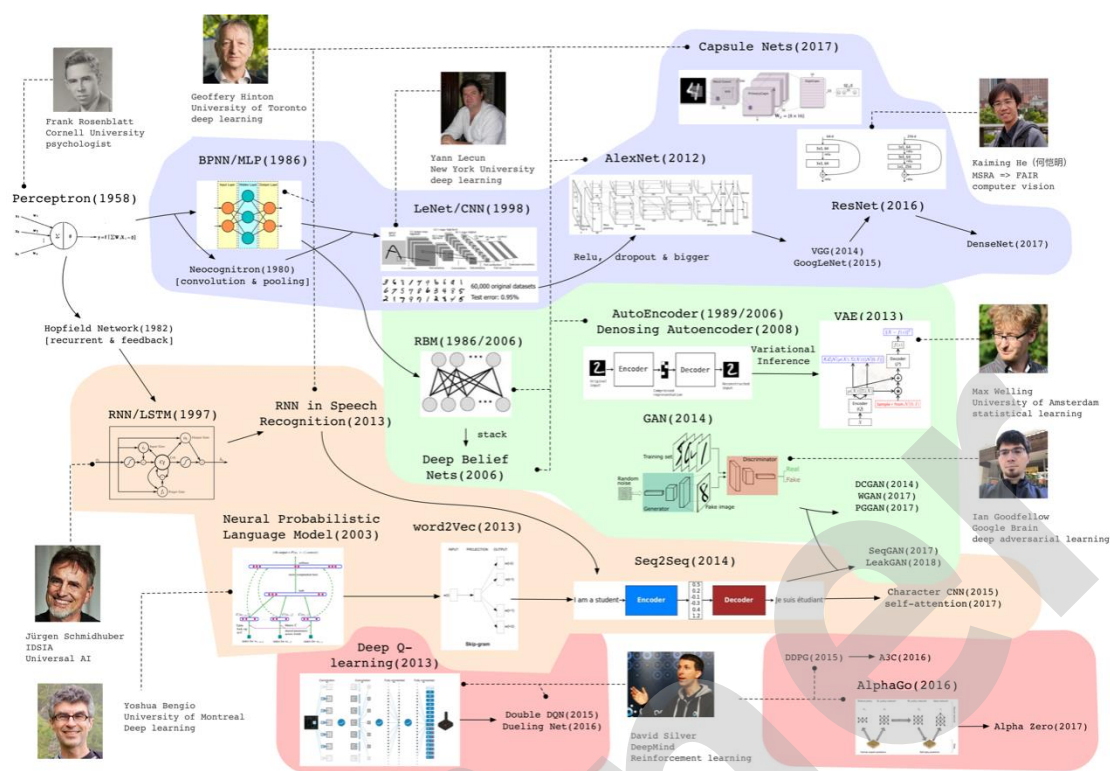


图 5-20 深度学习算法的 4 个发展脉络

第一个发展脉络（图 5-20 中的浅紫色区域）以计算机视觉和卷积网络为主，内容是以前向网络为代表的深度学习算法。这个脉络的进展可以追溯到 1979 年，日本学者邦彦（Kunihiko Fukushima）提出的神经认知机（Neocognitron）。该研究给出了卷积和池化的思想。1986 年，杰弗里·辛顿（Geoffrey Hinton）提出了反向传播训练 MLP 解决了感知机不能处理非线性学习的问题，引发了神经网络研究的热潮。1998 年，以“卷积网络之父”美国学者杨立昆（Yann LeCun）为首的研究人员实现了一个七层的卷积神经网络 LeNet-5 以识别手写数字。2012 年，辛顿组提出的 AlexNet 在 ImageNet 上以巨大优势夺冠，引发了研究深度学习的热潮。AlexNet 在传统 CNN 的基础上加上了 ReLU、Dropout 等技巧，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，随后出现了 VGG、GoogLeNet 等新模型。2016 年，青年计算机视觉科学家何恺明提出残差网络(ResNet)，极大增加了网络深度，效果有很大提升，成为图像识别、目标检测网络中的骨干架构。近年的 CVPR 会议 Best Paper 的获得者黄高提出的密集卷积网络（DenseNet），将这个思路继续发展下去。2017 年，辛顿认为反向传播和传统神经网络还存在一定缺陷，因此提出叫胶囊网络（Capsule Net），

该模型增强了可解释性，减少了神经网络参数，但目前在 CIFAR-10 等数据集上效果一般，这个思路还需要继续验证和发展。

第二个发展脉络（图 5-20 中的浅绿色区域）以**生成模型**为主，是一个以**自学习、自编码为代表的学习**。机器学习方法中**生成模型**一直占据着一个非常重要的地位，但基于神经网络的生成模型一直没有引起广泛关注。辛顿在 2006 年的时候基于受限玻尔兹曼机（RBM，一个 19 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，并且将其堆栈成为深度信念网络(Deep Belief Network)，使用逐层贪婪或者 wake-sleep 的方法训练，当时模型的效果不大理想。但值得关注的是，正是基于 RBM 模型，辛顿等人开始设计深度框架，因此这也可以看做深度学习的一个开端。自编码器（Auto-Encoder）也是上个世纪 80 年代辛顿组提出的模型，后来随着计算能力的进步重新登上舞台。图灵奖获得者本吉奥（Yoshua Bengio）等人又提出了去噪自编码器（Denoise Auto-Encoder），主要针对资料中可能存在的噪音问题。麦克斯·威灵（Max Welling）等人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，并且最后长得跟自编码器（Auto-Encoder）有点像，被称为变分自编码器（Variational Auto-Encoder）。此模型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 年提出的非常火的模型，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。后来引起大量跟随的研究，包括：DCGAN（一个相当好的卷积神经网络实现）、WGAN（通过维尔斯特拉距离替换原来的 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）。

第三个发展脉络（图 5-20 中的橙黄色区域）是**序列模型，自循环神经网络（概率图模型的发展）的算法**。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科夫 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 神经网络，即在神经网络中加入了递归网络的思想。1997 年

施米德胡贝(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，真正让神经网络模型得到广泛关注的还是 2013 年辛顿组使用 RNN 做语音识别的工作，比传统方法高出一大截。在文本分析方面，本吉奥在 SVM 很火的时期提出了一种基于神经网络的语言模型（当时机器学习还是 SVM 和 CRF 的天下），后来 Google 在 2013 年提出的 word2vec 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模型（Attention）相结合，也大大增强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都有不俗的表现，而且时空消耗更少。自注意模型（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”把基于注意力机制的序列神经模型推向高潮。

第四个发展脉络（图 5-20 中的粉色区域）是**增强学习**，以**强化学习**为代表。这个领域最出名的当属谷歌的 DeepMind 公司，图中标出的戴维·席尔瓦（David Silver）博士是一直研究强化学习的高管。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。后来又应用在许多游戏场景中，并将其成果发表在 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，它们是基于策略梯度（Policy Gradient）和神经网络结合的变种。大家都熟知的 AlphaGo，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，但通过自主学习来玩不同（棋类）游戏的新算法 Alpha Zero。

下面对深度学习的不同方面进行简单解读。

5.1.4.1 卷积神经网络

卷积神经网络（CNN）作为最重要的基础模型之一，以其强大的层次化特征建模能力，非常适合处理图片、视频等类型的数据，引起了学术界和工业界的广

泛研究和应用。

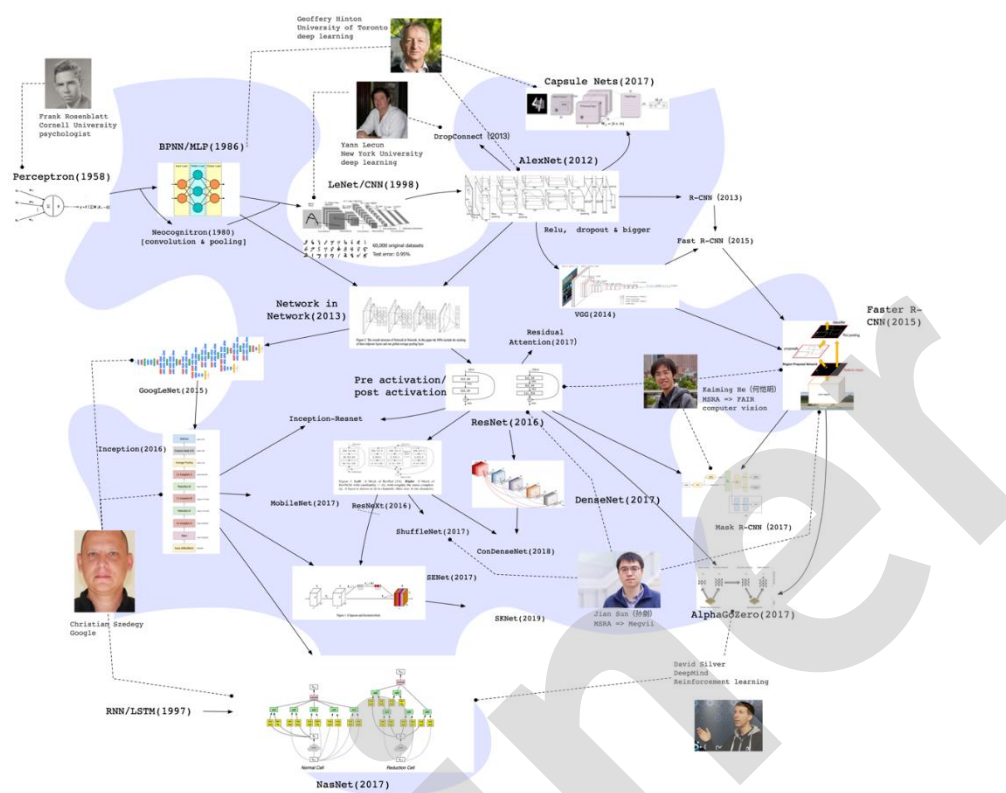


图 5-21 卷积神经网络的重要进展

卷积神经网络的起源最早可以追溯到 1962 年 Hubel 和 Wiesel 对猫大脑中的视觉系统研究。1980 年，一个日本科学家福岛邦彦（Kunihiko Fukushima）提出了一个包含卷积层、池化层的神经网络结构。在这个基础上，Yann Lecun 将 BP 算法应用到这个神经网络结构的训练上，就形成了当代卷积神经网络的雏形。

最初的 CNN 效果并不算好，而且训练也非常困难。虽然在阅读支票、识别数字之类的任务上有一定的效果，但由于在一般的实际任务中表现不如 SVM、Boosting 等算法，因此一直处于学术界的边缘地位。直到 2012 年，ImageNet 图像识别大赛中，Hinton 组的 AlexNet 引入了全新的深层结构和 Dropout 方法，一下子把 error rate 从 25% 降低到了 15%，在大规模图像分类数据集 ImageNet 上实现了领先传统算法 10 个点以上的突破，这颠覆了图像识别领域，卷积神经网络的威力开始初现。AlexNet 有很多创新，尽管都不是很难的方法。其最主要的结果是让人们意识到原来那个福岛邦彦提出的、Yann LeCun 优化的 LeNet 结构原来是有很大大改进空间的：只要通过一些方法能够加深这个网络到 8 层左右，让网

络表达能力提升，就能得到出人意料的好结果。

顺着 AlexNet 的思想，LeCun 组 2013 年提出一个 **DropConnect**，把 Error Rate 降低到了 11%。而 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 的应用也得到了 2014 年 *Imagine* 另一个挑战——图像检测的冠军。Network in Network 更加引发了大家对 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，MSRA 的任少卿、何恺明、孙剑等人，尝试把 identity 加入到卷积神经网络中提出 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 convolution, Attention, Dense connection, Channelwise-Attention 等，最终 ImageNet 上 Error Rate 降到了 2.2%，大大超过人类的错误率。而另一个挑战——图像检测中，也是任少卿、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。这个创新一下子把图像检测的 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 达到了最终的 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。

CNN 结构变得越来越复杂，很多结构很难解释和设计。于是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。NasNet 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），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 NasNet 哪个是真的正确的道路，有待进一步研究。

卷积神经网络 CNN 的发展引发了其他领域的很多变革。比如：利用 CNN，AlphaGo 战胜了李世石，攻破了围棋（基础版本的 AlphaGo 其实和人类高手比起来是有胜有负的）。后来利用了 ResNet 和 Faster-RCNN 的思想，一年后的 Master 则完全战胜了所有人类围棋高手。后来又有很多复现的开源围棋 AI，每一个都能用不大的计算量超过所有的人类高手。以至于现在人们讲棋的时候，都是按着 AI 的胜率来讲了。

5.1.4.2 AutoEncoder

AutoEncoder 的基本思想是利用神经网络来做无监督学习，就是把样本的输入同时作为神经网络的输入和输出。本质上是希望学习到输入样本的表示（Encoding）。早期 AutoEncoder 的研究主要是数据过于稀疏、数据高维导致计算复杂度高。比较早用神经网络做 AutoEncoder 的可以追溯到 80 年代的 BPNN 和 MLP 以及当时 Hinton 推崇的 RBM。后来到了 2000 年以后还坚持在做的只剩下 Hinton 的 RBM 了。从 2000 年以后，随着神经网络的快速兴起，AutoEncoder 也得到快速发展，基本上有几条线：稀疏 AutoEncoder、噪音容忍 AutoEncoder、卷积 AutoEncoder、变分 AutoEncoder。最新的进展是结合对抗思想的对抗 AutoEncoder。

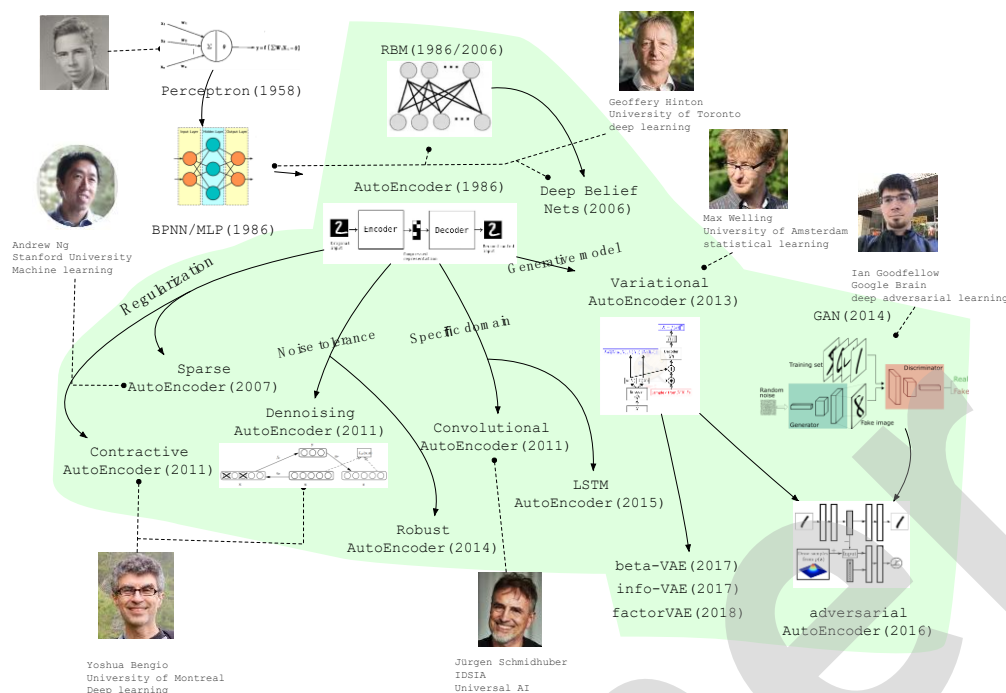


图 5-22 Auto-Encoder 的重要进展

稀疏 AutoEncoder 在学习输入样本表示时可以学习到相对比较稀疏的表示结果，这在 Overcomplete AutoEncoder（就是学习得到高维表示）方法中尤为重要。代表性人物包括斯坦福大学的 Andrew Ng 和蒙特利尔的 Yoshua Bengio 教授。具体方法是在原来的损失函数中加一个控制稀疏化的正则化项，通过控制优化过程来实现。

Denoising AutoEncoder 的核心思想是提高 Encoder 的鲁棒性，本质上就是避免可能的 overfitting。一个办法是在输入中加入随机噪音（比如随机置 0 一些输入，或者随机把部分输入变为 marked），这些思想后来在 BERT 等模型中也有广泛使用；另一个办法就是结合正则化的思想，比如在目标函数中加上 eEncoder 的 Jacobian 范数。Jacobian 范数可以让学习到的特征表示更具有差异性。

著名研究者 Jurgen Schmidhuber 提出了基于卷积网络的 AutoEncoder 以及后来的 LSTM AutoEncoder。Max Welling 基于变分思想提出变分 AutoEncoder 方法 VAE，这也是一个里程碑式的研究成果。后面很多研究者在这个基础上进行了扩展，包括 info-VAE、beta-VAE 和 factorVAE 等。最近还有人借鉴 Ian Goodfellow 等人提出的对抗建模思想提出 Adversarial AutoEncoder，也取得了很好的效果。这和之前的噪音容忍的 AE 学习也有一定呼应。除了上面的思想，就是可以把上

面的各种方法 stacking 起来。

5.1.4.2 循环神经网络 RNN

循环神经网络（RNN）是处理序列数据的重要模型之一，经典工作包括 LSTM^[21]、GRU^[22]等，在机器翻译等序列-序列转换、序列建模等问题中有着广泛的应用。

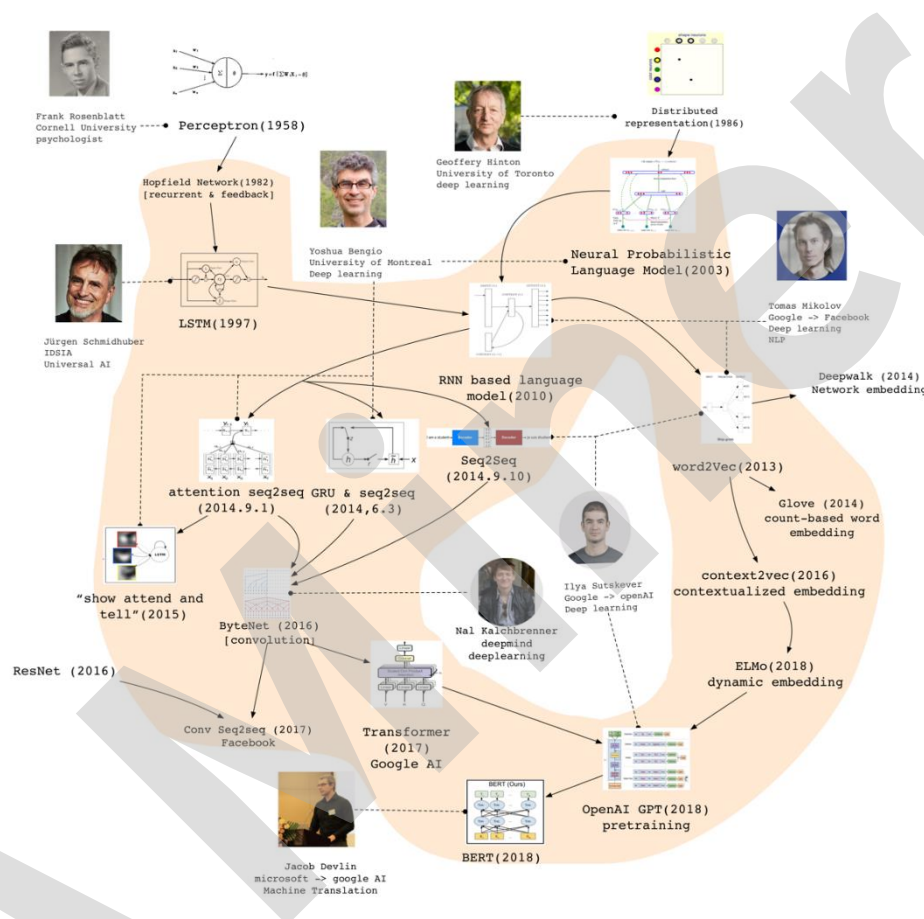


图 5-23 循环神经网络 RNN 的重要进展

1982 年，美国加州理工学院物理学家 John Hopfield 发明了一种单层反馈神经网络 Hopfield Network，用来解决组合优化问题。这是最早的 RNN 的雏形。1986 年，另一位机器学习的泰斗 Michael I. Jordan 定义了 Recurrent 的概念，提出 Jordan Network。1990 年，美国认知科学家 Jeffrey L. Elman 对 Jordan Network

[21] S. Hochreiter, J. Schmidhuber. Long short-term memory. Neural computation, 9(8): 1735-1780, 1997.

[22] Chung J, Gulcehre C, Cho K H, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. Eprint Arxiv, 2014.

进行了简化,并采用 BP 算法进行训练,便有了如今最简单的包含单个自连接节点的 RNN 模型。但此时 RNN 由于梯度消失 (Gradient Vanishing) 及梯度爆炸 (Gradient Exploding) 的问题,训练非常困难,应用非常受限。直到 1997 年,瑞士人工智能研究所的主任 Jurgen Schmidhuber 提出长短期记忆(LSTM),LSTM 使用门控单元及记忆机制大大缓解了早期 RNN 训练的问题。同样在 1997 年, Mike Schuster 提出双向 RNN 模型 (Bidirectional RNN)。这两种模型大大改进了早期 RNN 结构,拓宽了 RNN 的应用范围,为后续序列建模的发展奠定了基础。此时 RNN 虽然在一些序列建模任务上取得了不错的效果,但由于计算资源消耗大,后续几年一直没有太大的进展。

2010 年, Tomas Mikolov 对 Bengio 等人提出的 Feedforward Neural Network Language Model(NNLM)进行了改进,提出了基于 RNN 的语言模型(RNN LM),并将其用在语音识别任务中,大幅提升了识别精度。在此基础上 Tomas Mikolov 于 2013 年提出了大名鼎鼎的 word2vec,与 NNLM 及 RNNLM 不同, word2vec 的目标不再专注于建模语言模型,而是如何利用语言模型学习每个单词的语义化向量 (Distributed Representation),当然 Distributed Representation 概念最早要来源于 Hinton 1986 年的工作。word2vec 引发了深度学习在自然语言处理领域的浪潮,此外还启发了 Knowledge Representation, Network Representation 等新领域。

另一方面,2014 年 Bengio 团队与 Google 几乎同时提出了 seq2seq 架构,将 RNN 用于机器翻译。没过多久, Bengio 团队又提出注意力 Attention 机制,对 seq2seq 架构进行改进。自此机器翻译全面进入到神经机器翻译(NMT)的时代, NMT 不仅过程简单,而且效果要远超统计机器翻译的效果。目前主流的机器翻译系统几乎都采用了神经机器翻译的技术。除此之外, Attention 机制也被广泛用于基于深度学习的各种任务中。

近两年,相关领域仍有一些突破性进展,2017 年, Facebook 人工智能实验室提出基于卷积神经网络的 seq2seq 架构,将 RNN 替换为带有门控单元的 CNN,提升效果的同时大幅加快了模型训练速度。此后不久, Google 提出 Transformer 架构,使用 Self-Attention 代替原有的 RNN 及 CNN,更进一步降低了模型复杂度。在词表示学习方面, Allen 人工智能研究所 2018 年提出上下文相关的表示学

习方法 ELMo，利用双向 LSTM 语言模型对不同语境下的单词，学习不同的向量表示，在 6 个 NLP 任务上取得了提升。OpenAI 团队在此基础上提出预训练模型 GPT，把 LSTM 替换为 Transformer 来训练语言模型，在应用到具体任务时，与之前学习词向量当作特征的方式不同，GPT 直接在预训练得到的语言模型最后一层接上 Softmax 作为任务输出层，然后再对模型进行微调，在多项任务上 GPT 取得了更好的效果。

不久之后，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。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。这些被称为 BERTology。

5.1.4.5 网络表示学习与图神经网络（GNN）

这个方面的研究可以追溯到 Hinton 当年 1986 的 Distributed Representation，后来 Stanford 的 Andrew Ng 实验室做了个 Neural Tensor Network，本质就是把知识之间的关系和表示学习一起放到 tensor 里面来做。后来 Facebook 的 Antonie Bordes 提出了 TransE，把知识网络的三元组融合到了表示学习中，这是 NLP 和知识图谱中的一个非常重要的研究，后面延续了一系列的工作，包括 TransH、TransR、TransA、TransG。

从表示学习本身来看，Neural Language Model 对于单词和文本的表示，是对原来向量模型的一个自然扩展，本质上类似一个隐含语义分析，只是这里用的是神经网络来做学习。RNN Based Language Model 是利用 RNN 进行表示学习，更好地保持了语言模型的连续性。但这个阶段的研究当时大部分都没有火起来，一是当时深度学习还没火起来，二是这些算法都还比较慢。2013 年 Tomas Mikolov 和 Jeff Dean 等人做 word2vec，占据了“天时、地利、人和”：深度学习开始发热、算法简单有效。现在 word2vec 已经轻松超过 1 万多引用了。后面的扩展也

很多, 如 pharagraph2vec、doc2vec, context2vec。以至于后面有一段时间, “2vec” 成了流行取名字的方法。最近的进展是 ELMo、OpenAI 的 GPT 和谷歌的 BERT。

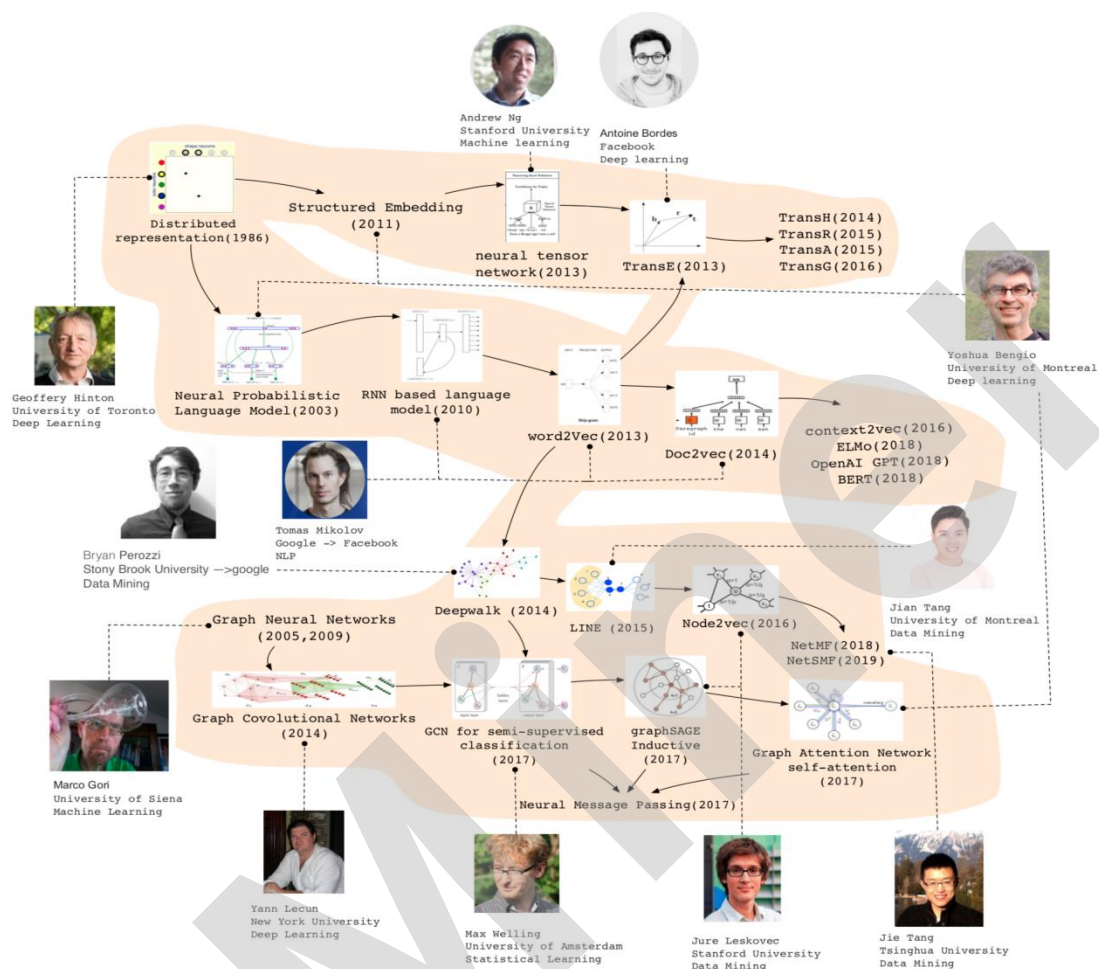


图 5-24 网络表示学习与图神经网络的重要进展

表示学习的另一个脉络就是扩展到网络数据上, 在 NLP 领域的 Structured Embedding、TransE 等模型更多的是语言中的局部结构信息, 而网络中还有更加复杂的拓扑结果。原 Stony Brook 大学的 Bryan 提出 DeepWalk, 这个算最早把 word2vec 稍微扩展了一下, 应用于网络数据, 这篇文章获得了当年 KDD 的最佳论文和后来 KDD 的最佳博士论文。很快这个工作吸引了大量关注, Jian Tang (原北大、微软, 现在 MILA) 等人做了两阶扩展, 斯坦福的 Jure Leskovec 做了面向社交网络的“三阶”扩展 node2vec, 后来清华也给出了一个理论证明, 证明这些不同方法本质上都在做一个矩阵分解, 并基于此提出了一个 NetMF 的算法以及其适用于大规模网络的实现 NetSMF。ProNE 是另一个清华作品, 其主要特点是

高效和高精度。该方法非常简单，本质上是在原来的表示学习上引入了一个类似卷积但又不是卷积的操作，大大提高了精度。

最近的网络表示学习更多的是用卷积网络直接对图做，大方向是 Graph Neural Network，最早是 Siena 大学的 Marco 等人在 2005 和 2009 年提出的，但当时没引起太大关注。后来 Yann Lecun 提出的 Graph Convolutional Networks，还有 Kipf & Welling 等人提出的 Semi-Supervised 的 GCN。这一系列的研究本质上就是 Neural Message Passing，在最近引起大量关注。斯坦福的 Jure 也提出了 GraphSage，利用 NMP 简化了卷积，提高了速度，并且支持 Inductive Learning，再后来 Yoshua 团队又提出了 Graph Attention Network，进一步提高了图卷积精度。最近网络表示学习非常热，前前后后都能看到三大巨头 Hinton、Yoshua 和 Yann 的影子。在未来若干年还会继续是个研究热点。

5.1.4.6 增强学习

Deep Mind 是一家英国人工智能公司，创立于 2010 年，2014 年被 Google 收购。创始人哈萨比斯出身于伦敦，母亲为新加坡华裔，13 岁便已经获得国际象棋大师的头衔，19 岁开始学习围棋，当前是围棋业余初段。DeepMind 于 2014 年开始开发 AlphaGO。2015 年 10 月，AlphaGO 5:1 樊麾；2016 年 3 月，AlphaGo 4:1 李世石；2017 年 5 月，AlphaGO 3:0 柯洁；2017 年 10 月 19 日，AlphaGo Zero 发表在 Nature，其思路是从零开始，自我对弈，40 天超过所有版本。2018 年 12 月 7 日，AlphaZero 再次发表于 Science，AlphaZero 使用与 AlphaGo Zero 类似但更一般性的算法，在不做太多改变的前提下，并将算法从围棋延伸到将棋与国际象棋上。2018 年 12 月，Deep Mind 公司推出 AlphaFold，可以根据基因序列预测蛋白质结构。2019 年 1 月 25 日，Deep Mind 公司 AlphaStar，在《星海争霸 II》以 10:1 战胜人类职业玩家。另一条在美国的战线，可能最著名的是 Open AI 公司，这是 Hinton 的高徒 Ilya Sutskever（AlexNet 发明人）创立的公司。2019 年 4 月，Open AI 推出 five dota2，2-0 战胜 Dota2 的 TI8 冠军战队 OG。

在研究方法上 Deep Q-Network(DQN)利用神经网络对 Q 值进行函数近似，并利用了 Experience Replay 和 Fixed Target Network 的策略让 DQN 可以收敛，在 Atari 的不少游戏上都超过了人类水平。Double DQN 是深度学习版本的 double

Q-learning，它通过微小的修改就成功减小了 DQN 中 max 操作带来的 bias。再后来，Dueling DQN 将 Q-network 分成了 action-dependent 和 action-independent 两个部分，从而提高了 DQN。DQN 是为 Value 的期望建模，greedy 的时候也是最大化期望的形式，Categorical DQN 的想法是直接为 Value 的分布进行建模。Noise DQN 在网络中添加了噪声，从而达到 exploration 的效果。DQN 还有非常多的提升版本，Rainbow 整合了多种 DQN 版本。Ape-X 从 Rainbow 的工作中发现 Replay 的优先级对于性能影响是最大的，故扩大 Prioritised Replay Buffer，并使用 360 个 actor 做分布式训练，比 Rainbow 更快也更好。

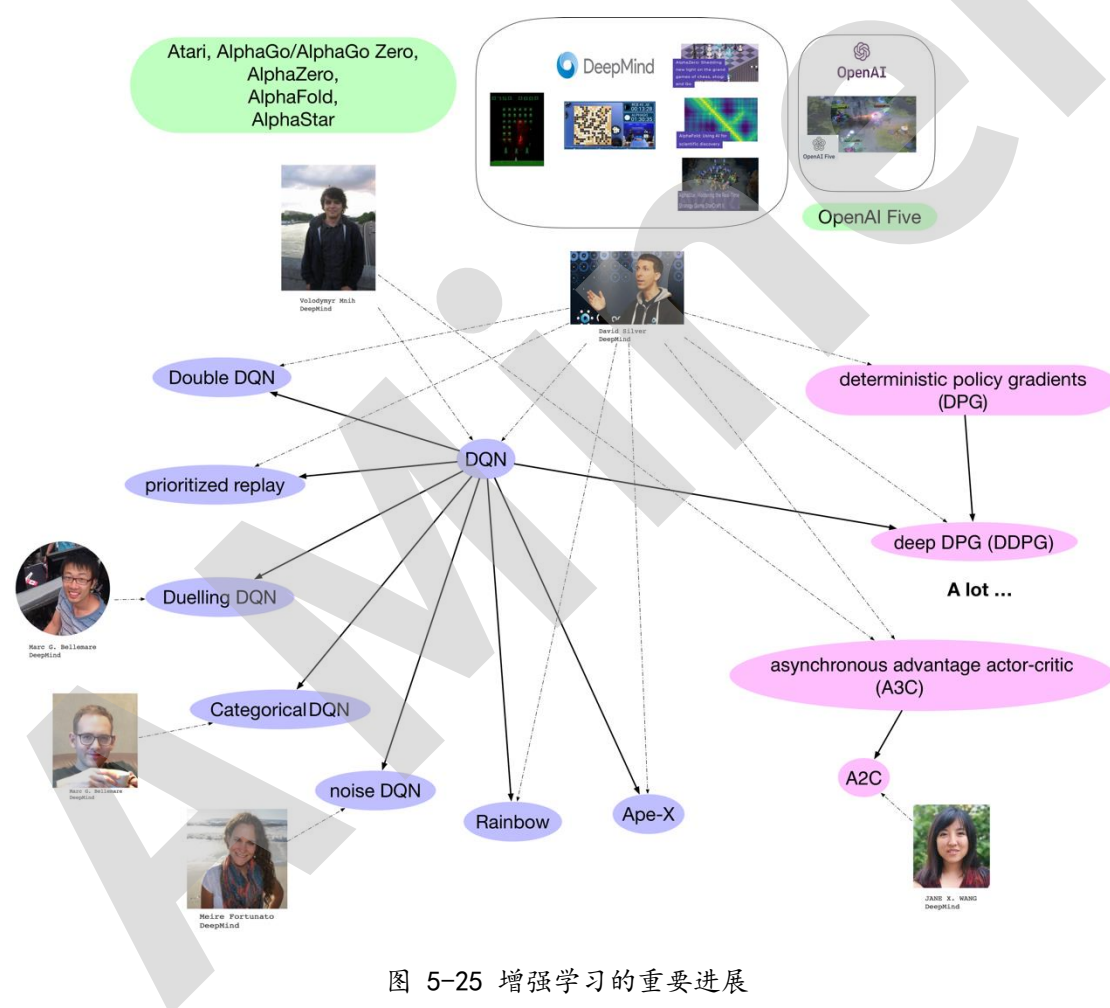


图 5-25 增强学习的重要进展

Deterministic policy gradients (DPG) 将 policy gradients 方法中的随机 policy 推广为确定性 policy。Deep DPG 使用了神经网络表示高维 state，是结合了 DQN 和 DPG 的 actor-critic 算法。A3C 是经典的 policy gradient 方法，可以并行 multiple agent 的训练，并异步更新参数。A2C 是 A3C 的同步、确定性 policy 版本，同步

的梯度更新，可以让并行训练更快收敛。

5.1.4.7 生成对抗网络

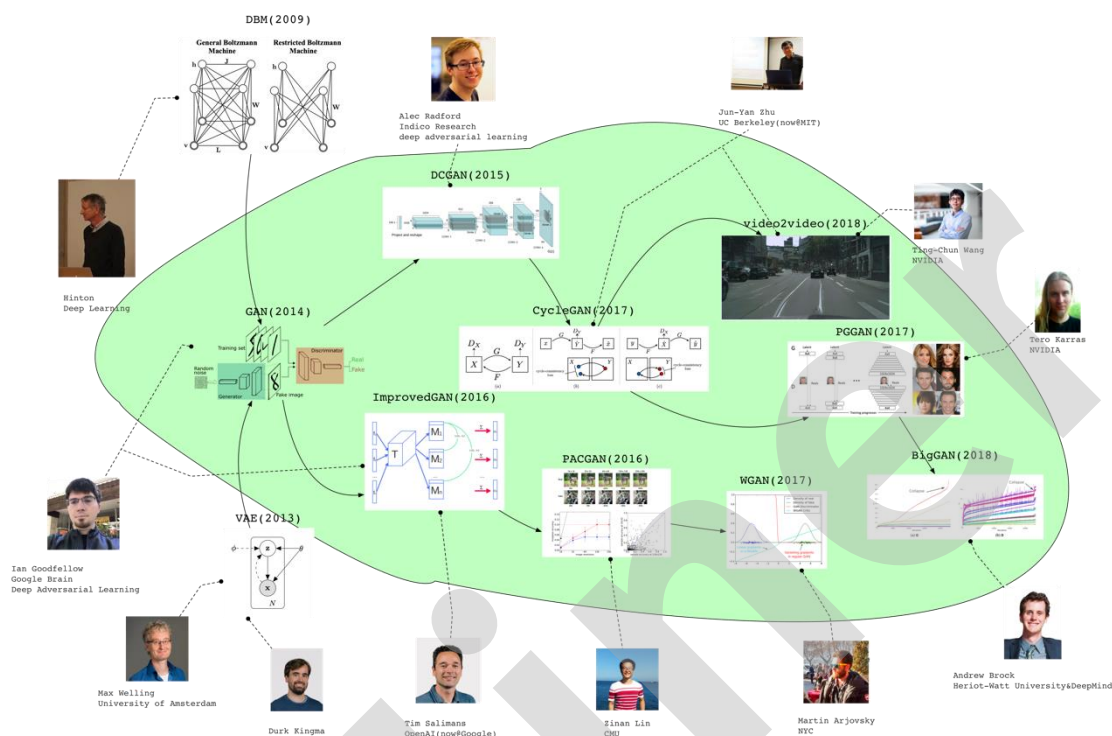


图 5-26 生成对抗网络的重要进展

GAN 最近几年发展非常快，这也是 Yoshua Bengio 获得图灵奖的贡献之一。传统的生成模型是要预测联合概率分布 $P(x,y)$ 。首先玻尔兹曼机（RBM）这个模型其实是一个基于能量的模型，1986 年时就有，Hinton 在 2006 年重新拿出来作为一个生成模型，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。

AutoEncoder 也是上个世纪 80 年代 Hinton 就提出的模型，此时由于计算能力的进步也重新登上舞台。Bengio 等人又研发了 Denoise AutoEncoder。Max Welling 等人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，并且最后长得跟 AutoEncoder 有点像，被称为 Variational AutoEncoder。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。

在生成模型方面，最近一个最重要的进展就是对抗生成网络（GAN）。2014 年 Ian Goodfellow 在 NIPS 上发表了最初的 GAN 文章，到现在已经有近九千引

用。这个模型引起如此关注的原因在于：一是这个模型理论上非常优雅，易于理解；二是效果确实好。GAN 的发明人 Ian Goodfellow 本科在斯坦福，硕士在 Andrew Ng 手下，博士师从蒙特利尔 Yoshua Bengio 和 Aaron Courville。另外还有几个有名的 GAN 的扩展，包括：cycleGAN 和 vid2vid。

5.1.4.8 老虎机

老虎机也是机器学习的一个重要分支，和深度学习有着或多或少的联系。老虎机实际上是个赌博机器。老虎机模型是个数学模型，追溯基本认为是一个病理学家 Thompson 在 1933 年提出的。他当时觉得验证新药的医学的随机双盲实验有些残酷的地方，对于被分到药效较差的新药的那一组病人并不公平。为了在实验中途就验证药物药效从而避免给病人带来痛苦，他提出了一个序列决策模型。但是，实际使用还是有很多问题，比如中途效果不好评价。美国 FDA 对在医学随机双盲实验中使用这种自适应调整的多臂老虎机方法，目前仍然只是建议使用。现在，老虎机模型在搜索和推荐方面的应用很多。

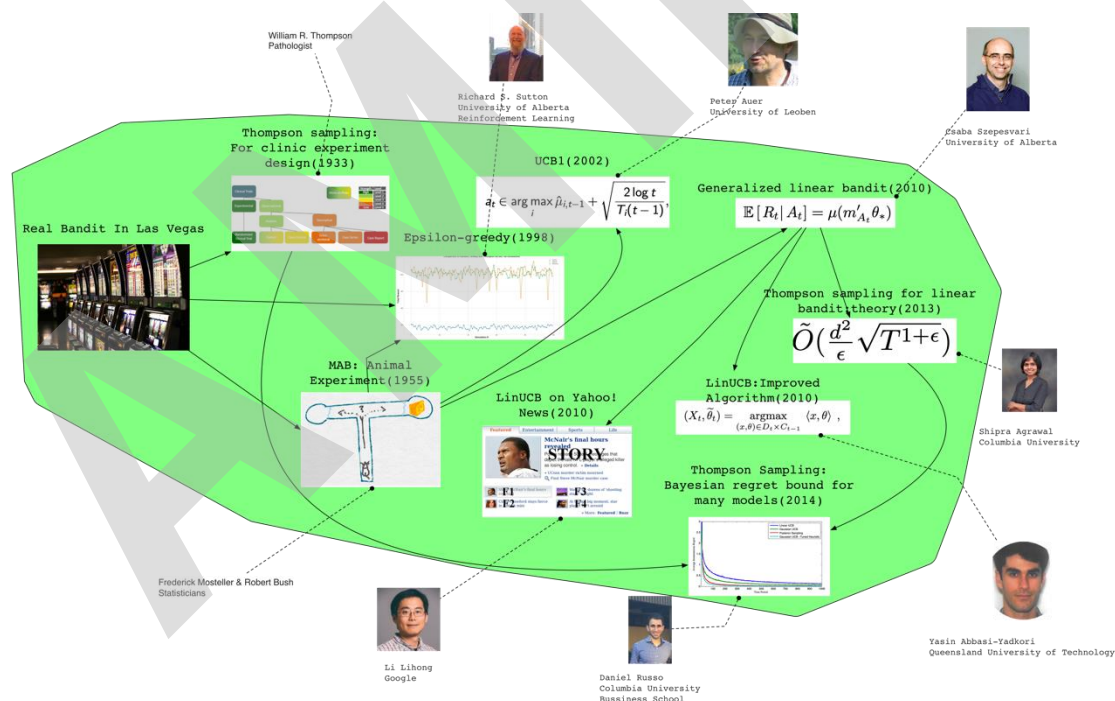


图 5-27 老虎机的重要进展

Epsilon-greedy 是种预留一点点机会去尝试的思想。Sutton 是强化学习方面的大佬，编写的教材 *Reinforcement Learning* 引用五位数，里边讲解了这个算法。

Peter Auer 不仅分析了 UCB 算法的理论性质，还分析了 Epsilon-greedy 的理论性质。这篇纯理论文章的引用量也达到了两千多。Frederick Mosteller 是哈佛统计系奠基人，20 世纪统计学界的超级牛人。他们当时做老虎机模型，主要是想给真实的动物或者人的序列决策建模，想抽象一个框架出来。所以他们做了一个老鼠找蛋糕的实验，也做了关于人玩赌博用的老虎机的实验。Li Lihong 在 Yahoo! news 上的 LinUCB 的工作发表在 WWW 上，这篇应用文章获得了大量关注，引用上千。他 1933 年用 Thompson sampling 这个很古旧的方法，作了一些系统性的实验，从实验结果的角度说明 Thompson sampling 效果很好。这篇文章发在 NIPS2011 上，也获得了大量关注。后来大批理论学者跟进在 2010 年后把 Thompson sampling 在线性模型上的理论基础建立起来。

5.1.5 机器学习技术研究趋势

技术研究发展趋势分析描述了技术的出现、变迁和消亡的全过程，可以帮助研究人员理解领域的研究历史和现状，快速识别研究的前沿热点问题。本部分借助 AMiner 平台，从人工智能 20 个子领域相关的论文中抽取关键词，然后统计这些关键词的起止年份，划分时间窗格，再按照窗格统计词频，最终生成待研究领域的技术研究发展趋势图。

技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序。

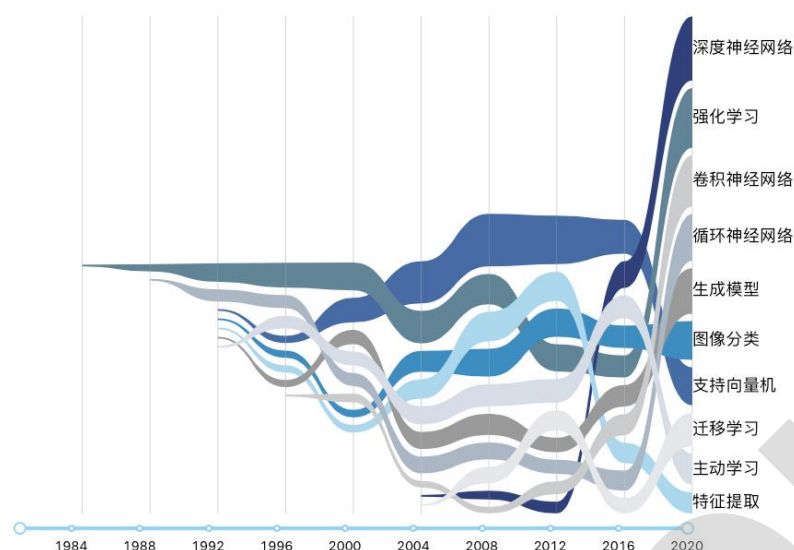


图 5-28 机器学习领域技术研究发展趋势

机器学习领域研究热度较高的 10 个技术话题，包括：强化学习（Reinforcement Learning）、深度学习（Deep Neural Network）、卷积神经网络（Convolutional Neural Network）、循环神经网络（Recurrent Neural Network）、生成模型（Generative Model）、图像分类（Image Classification）、支持向量机（Support Vector Machine）、迁移学习（Transfer Learning）、主动学习（Active Learning）、特征提取（Feature Extraction），如图 5-28 所示。近十年内，以深度学习、强化学习为代表的深度学习相关的技术话题研究热度上升很快，这些现象与深度学习算法具有自适应、自学习的强大并行信息处理能力不无联系。

5.2 自然语言处理

5.2.1 自然语言处理概念

自然语言是指汉语、英语、法语等人们日常使用的语言，是人类社会发展演变而来的语言，它是人类学习生活的重要工具。概括说来，自然语言是指人类社会约定俗成的，区别于如程序设计的语言的人工语言。

自然语言处理，是指用计算机对自然语言的形、音、义等信息进行处理，即对字、词、句、篇章的输入、输出、识别、分析、理解、生成等的操作和加工。自然语言处理是关于计算机科学与人工智能的一个重要学科，是语言学与数学、

计算机科学、人工智能之间互相作用的领域，主要研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法，以实现人机间的信息交流，例如能有效地实现自然语言通信的计算机软件系统。

自然语言处理的具体表现形式包括机器翻译、文本摘要、文本分类、文本校对、信息抽取、语音合成、语音识别等。自然语言处理机制涉及两个流程，包括自然语言理解和自然语言生成。自然语言理解是指计算机能够理解自然语言文本的意义，自然语言生成则是指能以自然语言文本来表达给定的意图。

自然语言的理解和分析是一个层次化的过程，许多语言学家把这一过程分为五个层次，可以更好地体现语言本身的构成，五个层次分别是语音分析、词法分析、句法分析、语义分析和语用分析，如图 5-29 所示。



图 5-29 自然语言理解层次

- **语音分析**是要根据音位规则，从语音流中区分出一个个独立的音素，再根据音位形态规则找出音节及其对应的词素或词。
- **词法分析**是找出词汇的各个词素，从中获得语言学的信息。
- **句法分析**是对句子和短语的结构进行分析，目的是要找出词、短语等的相互关系以及各自在句中的作用。
- **语义分析**是找出词义、结构意义及其结合意义，从而确定语言所表达的真正含义或概念。
- **语用分析**是研究语言所存在的外界环境对语言使用者所产生的影响。

在人工智能领域或者是语音信息处理领域中，学者们普遍认为采用图灵试验可以判断计算机是否理解了某种自然语言，具体的判别标准如下。

1. **问答**，机器人能正确回答输入文本中的有关问题；
2. **文摘生成**，机器有能力生成输入文本的摘要；
3. **释义**，机器能用不同的词语和句型来复述其输入的文本；

4. 翻译，机器具有把一种语言翻译成另一种语言的能力。

5.2.2 自然语言处理过去十年主要进展

自然语言处理自 20 世纪 50 年代末六十年代初兴起。“图灵测试”提出一般被认为是自然语言处理思想的开端，20 世纪 50 年代到 70 年代自然语言处理主要采用基于规则的方法，具有语句覆盖不全、方法对开发者的计算机和语言学背景要求极高的等缺点，实用化较差。后来发展过程中的代表性事件是丹尼尔·博布罗（Daniel Bobrow）在 1964 年发表的“Natural language input for a computer problem solving system”，以及约瑟夫·维森鲍姆（Joseph Weizenbaum）在 1966 年发表的“ELIZA-a computer program for the study of natural language communication between man and machine”。70 年代以后随着互联网的高速发展，丰富的语料库成为现实以及硬件不断更新完善，自然语言处理基于统计的方法逐渐代替了基于规则的方法。贾里尼克和他领导的 IBM 华生实验室是推动这一转变的关键，他们采用基于统计的方法，将当时的语音识别率从 70%提升到 90%。在这一阶段，自然语言处理基于数学模型和统计的方法取得了实质性的突破，从实验室走向实际应用。90 年代中期以来，计算机的速度和存储量大幅增加，为自然语言处理改善了物质基础，使得语音和语言处理的商品化开发成为可能；同时期 Internet 商业化和网络技术的发展使得基于自然语言的信息检索和信息抽取的需求变得更加突出。

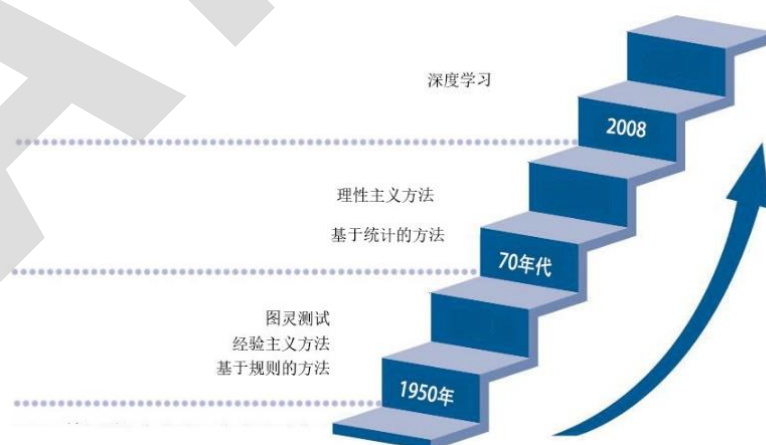


图 5-30 自然语言发展历程

21 世纪，自然语言处理发展出神经网络语言模型、多任务学习、预训练语言模型等典型技术。从 2008 年到现在，在图像识别和语音识别领域的成果激励下，人们也逐渐开始引入深度学习来做自然语言处理研究，由最初的词向量到 2013 年的 word2vec，将深度学习与自然语言处理的结合推向了高潮，并在机器翻译、问答系统、阅读理解等领域取得了一定成功。深度学习是一个多层的神经网络，从输入层开始经过逐层非线性的变化得到输出。从输入到输出做端到端的训练。把输入到输出对的数据准备好，设计并训练一个神经网络，即可执行预想的任务。RNN 已经是自然语言处理最常用的方法之一，GRU、LSTM 等模型相继引发了一轮又一轮的热潮。2014 年提出的 seq2seq 模型，被广泛应用到在机器翻译、文本摘要和图像字幕等任务中并取得很大成功，在实现人机自然通信方面发挥了关键性作用。

近年来，自然语言处理在词向量（Word Embedding）表示、文本的（编码）encoder 和 decoder（反编码）技术以及大规模预训练模型（pre-trained）上的方法极大地促进了自然语言处理的研究。

尤其，预训练语言模型在自然语言处理领域有了重要进展。预训练模型指的是首先在大规模无监督的语料上进行长时间的无监督或者是自监督的预先训练（pre-training），获得通用的语言建模和表示能力。之后在应用到实际任务时对模型不需要做大的改动，只需要在原有语言表示模型上增加针对特定任务获得输出结果的输出层，并使用任务语料对模型进行少许训练即可，这一步骤被称作微调（Fine-tuning）。

自 ELMo、GPT、BERT 等一系列预训练语言表示模型（Pre-trained Language Representation Model）出现以来，预训练模型在绝大多数自然语言处理任务上都展现出了远远超过传统模型的效果，受到越来越多的关注，是自然语言处理领域近年来最大的突破之一。

BERT（Bidirectional Encoder Representation from Transformer）是 Google AI 于 NAACL（The North American Chapter of the Association for Computational Linguistics）2019 提出的一个预训练语言模型。BERT 的创新点是提出了有效的无监督预训练任务，从而使得模型能够从无标注语料中获得通用的语言建模能力。

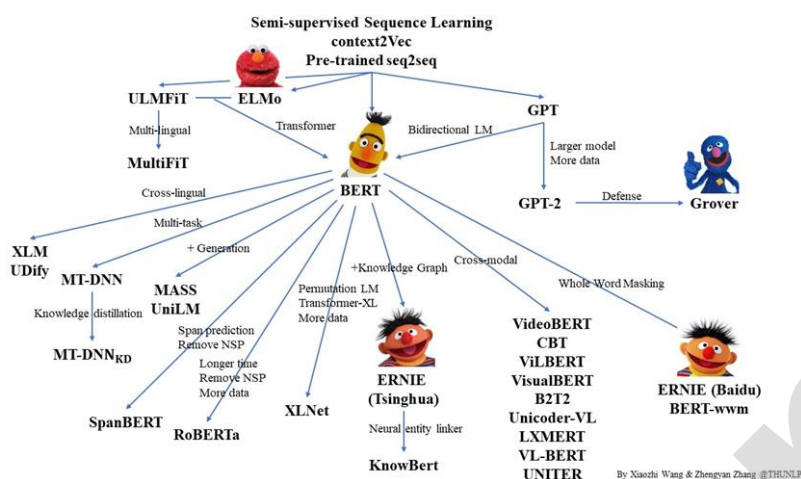


图 5-31 几个 BERT 扩展模型

BERT 之后涌现了许多对其进行扩展的模型（如图 5-31 所示），包括：跨语言预训练的 XLM 和 UDify、跨模态预训练的模型、融合知识图谱的 ERNIE、将 seq2seq 等语言生成任务整合入 BERT 类模型的 MASS 和 UniLM 等。其中几个重要的进展包括：

（1）XLNet 使用 Transformer-XL 替代了 Transformer 作为基础模型，拥有编码超长序列的能力。XLNet 提出了一个新的预训练语言任务：Permutation Language Modeling（排列语言模型），模型将句子内的词语打乱顺序，从而使得预测当前词语时可以利用双向信息。XLNet 相对 BERT 也使用了更多的语料。

（2）RoBERTa 采用了与 BERT 具有相同的模型结构，同样采用了屏蔽语言模型任务进行预训练，但舍弃了 BERT 中下句预测模型。此外，RoBERTa 采用了更大规模的数据和更鲁棒的优化方法，从而取得了更好的表现。

（3）ALBERT 模型针对 BERT 参数量过大难以训练的问题做了优化，一是对词向量矩阵做分解，二是在层与层之间共享参数。此外，ALBERT 将下句预测模型替换为句序预测任务，即给定一些句子预测它们的排列顺序。

GPT（Generative Pre-training Transformer）是由 Alec Radford 编写的语言模型，2018 年由 OpenAI 发布，目前是第 3 代。根据 OpenAI 的官方解释，GPT 是一种通过无人监督预训练和监督微调实现文本的自动补全工具。GPT 采用了生成式语言模型（两个神经网络通过竞争相互完善），通过对不同的书

面材料集与长篇连载文本的预训练，能够获取知识并处理长程依赖关系。它是一套人工智能的生成网络模型，可以用文字、图片、音乐、一段程序、一个功能或者数据分析结果来生成新的内容。

GPT-3 最初出现在 2020 年 5 月底公开发布的一篇论文“*Language Models are Few-Shot Learners*”(小样本学习的语言模型)²³。GPT-3 主要聚焦于更通用的 NLP 模型，解决当前 BERT 类模型的两个缺点：对领域内有标签数据的过分依赖、对于领域数据分布的过拟合。相比前二代，**GPT-3** 是一个超级大的自然语言处理模型，区别主要在于密度的改变和稀疏的自注意力模型。GPT-3 基于 1,750 亿个参数，它将学习能力转移到同一领域的多个相关任务中，既能做组词造句，又能做阅读理解。

5.2.3 自然语言处理技术研究趋势

借助 AMiner 平台，从自然语言处理领域相关的论文中抽取关键词，然后统计这些关键词的起止年份，划分时间窗格，再按照窗格统计词频，最终生成待研究领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-32 所示。

²³共计 31 位作者 Submitted on 28 May 2020 (v1), last revised 22 Jul 2020 (this version, v4), paper 链接: <https://arxiv.org/abs/2005.14165>, github 链接: <https://github.com/openai/gpt-3>

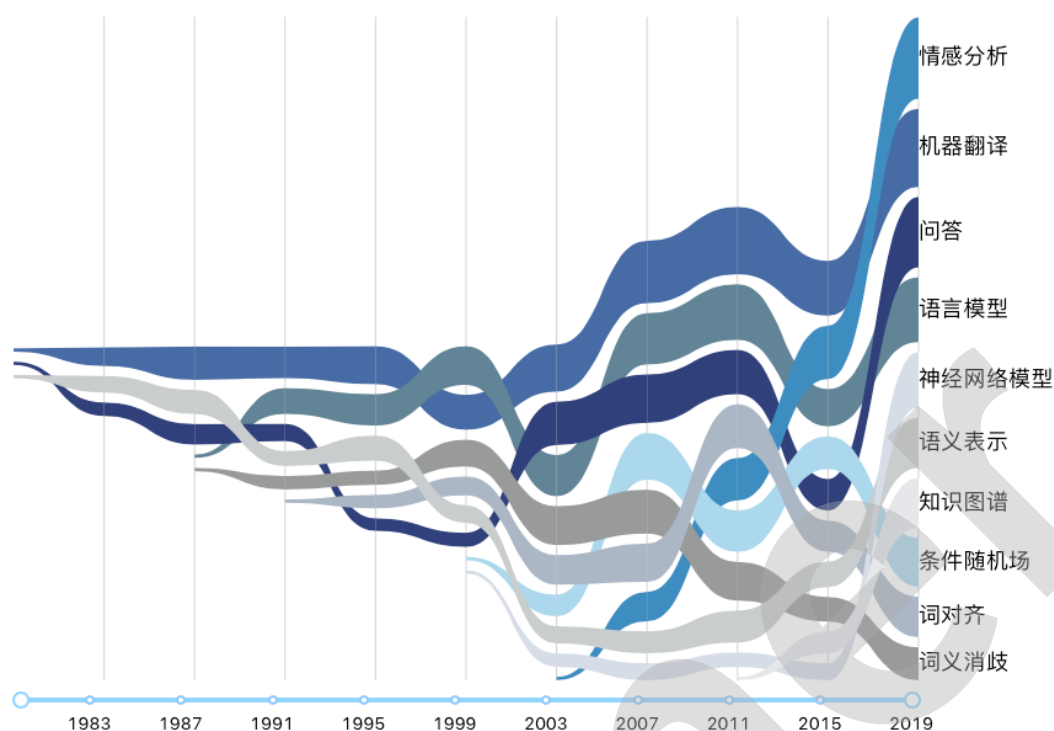


图 5-32 自然语言处理领域技术研究发展趋势

自然语言处理领域研究热度较高的 10 个技术话题，包括：情感分析（Sentiment Analysis）、机器翻译（Machine Translation）、问答（Question answering）、语言模型（Language Model）、神经网络模型（Neural Network Model）、语义表示（Semantic Representation）、知识图谱（Knowledge Graph）、词对齐（Word Alignment）、条件随机场（Conditional Random Field）和词义消歧（Word Sense Disambiguation），如上图所示。2003 年以后，情感分析的研究热度增长迅速，其基于自然语言处理的数据挖掘过程和技术被应用于从用户生成内容中提取和分析主观信息，随着社交网络的发展成为自然语言处理最有吸引力的用例之一；机器翻译、问答等技术话题研究热度也一直保持上升态势，其中机器翻译的研究热度长期维持在较高阶段，与神经网络模型的引入并取得不错成绩具有很大关联。未来，自然语言处理技术将与人工智能的发展更多相互融合。

5.3 知识工程

5.3.1 知识工程概念

知识工程是一门运用现代科技手段高效率、大容量地获得知识、信息的工程技术学科。目的是最大限度地提高人的才智和创造力，掌握知识和技能，为智力开发服务，主要研究如何组成由电子计算机和现代通讯技术结合而成的新的通讯、教育、控制系统。知识工程包括知识获取、知识验证、知识表示、推论、解释和理由五个活动过程。知识工程是人工智能的原理和方法，它为电子计算机的进一步智能化提供了条件，是关系人工智能发展的关键性学科。

知识工程的概念是 1977 年由美国斯坦福大学计算机科学家、1994 年图灵奖获得者、知识工程的建立者爱德华·费根鲍姆（Edward Feigenbaum）在第五届国际人工智能会议上提出的，后来在 1980 年的一个项目报告 *Knowledge Engineering: The Applied Side of Artificial Intelligence* 中再次提出，从此确立了知识工程在人工智能中的核心地位。费根鲍姆认为知识工程是将知识集成到计算机系统从而完成只有特定领域专家才能完成的复杂任务。

知识工程的目的是开发一个实现自动化知识转移和利用的技术和方法²⁴。知识工程主要包括：知识获取、知识表示和知识应用。近年来热度火爆的知识图谱，就是新一代的知识工程技术。

在大数据时代，知识工程是从大数据中自动或半自动获取知识，建立基于知识的系统，以提供互联网智能知识服务。大数据对智能服务的需求，已经从单纯的搜集获取信息，转变为自动化的知识服务。我们需要利用知识工程为大数据添加语义/知识，使数据产生智慧（Smart Data），完成从数据到信息到知识，最终到智能应用的转变过程，从而实现对大数据的洞察、提供用户关心问题的答案、为决策提供支持、改进用户体验等目标。

知识工程是一门为了传递、利用、扩展知识而涉及到计划、设计、构造、知

²⁴ Julius T. Tou. Knowledge engineering[J]. International Journal of Computer & Information Sciences, 1980, 9(4).

识库或知识库系统的学科。知识工程是以知识为基础的系统，就是通过智能软件而建立的专家系统。知识工程可以看成是人工智能在知识信息处理方面的发展，研究如何由计算机表示知识，进行问题的自动求解。知识工程的研究使人工智能的研究从理论转向应用，从基于推理的模型转向基于知识的模型，包括了整个知识信息处理的研究。

5.3.2 知识工程过去十年主要进展

回顾知识工程四十年来发展历程，可以将知识工程的演进过程和技术进展，分成前知识工程、专家系统、万维网 1.0，群体智能以及知识图谱五个标志性的阶段，如图 5-33 所示。

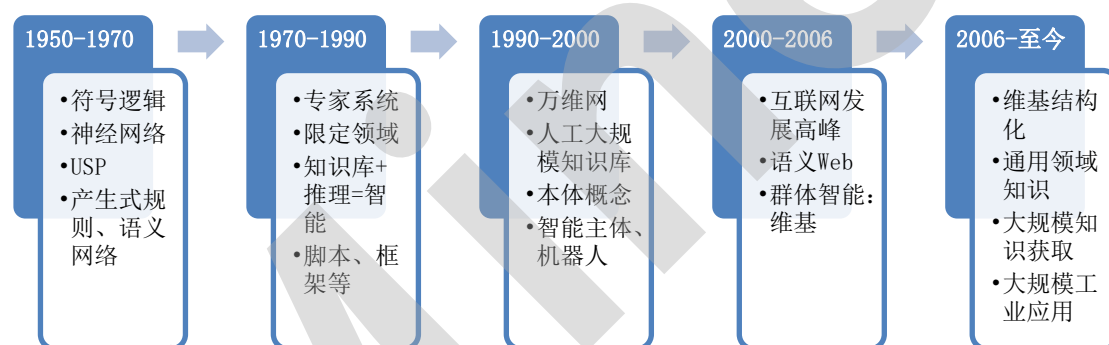


图 5-33 知识工程发展历程

在前知识工程时期（1950-1970 年），知识表示方法主要有逻辑知识表示、产生式规则、语义网络等，具有代表性的工作是通用问题求解程序（GPS）。在专家系统时期（1970-1990 年），知识表示方法有新的演进，包括框架和脚本等，人工智能开始转向建立基于知识的系统，通过“知识库+推理机”实现机器智能，这一时期涌现出很多成功的限定领域专家系统，如 MYCIN 医疗诊断专家系统、识别分子结构的 DENRAL 专家系统以及计算机故障诊断 XCON 专家系统等，帮助将专家的领域知识转变成计算机可以处理的知识。在万维网 1.0 时期（1990-2000 年），研究中提出了本体的知识表示方法，万维网为人们提供了一个可以共享信息的开放平台，出现了很多人工构建大规模知识库，包括广泛应用的英文 WordNet，采用一阶谓词逻辑知识表示的 Cyc 常识知识库，以及中文的 HowNet。在群体智能时期（2000-2006 年），语义 Web 的概念被提出，旨在对互联网内容

进行结构化语义表示，利用本体描述互联网内容的语义结构，通过对网页进行语义标识得到网页语义信息，从而获得网页内容的语义信息，使人和机器能够更好地协同工作。W3C 进一步提出万维网上语义标识语言 RDF（资源描述框架）和 OWL（万维网本体表述语言）等描述万维网内容语义的知识描述规范。

万维网的出现使得知识从封闭知识走向开放知识，从集中构建知识成为分布群体智能知识。原来专家系统是系统内部定义的知识，现在可以实现知识源之间相互链接，可以通过关联来产生更多的知识而非完全由固定人生产。这个过程中出现了群体智能，典型代表是维基百科，体现了互联网大众用户对知识的贡献，成为今天大规模结构化知识图谱的重要基础。

从 2006 年至今，知识工程正处于知识图谱的新发展时期。随着大规模维基百科类富结构知识资源的出现和网络规模信息提取方法的进步，大规模知识获取方法取得了巨大进展。与 Cyc、WordNet 和 HowNet 等手工研制的知识库和本体的开创性项目不同，这一时期知识获取是自动化的，并且在网络规模下运行。

当前自动构建的知识库已成为语义搜索、大数据分析、智能推荐和数据集成的强大资产，在大型行业和领域中正在得到广泛使用。典型的例子是谷歌收购 Freebase 后在 2012 年推出的知识图谱（Knowledge Graph），Facebook 的图谱搜索，Microsoft Satori 以及商业、金融、生命科学等领域特定的知识库。最具代表性大规模网络知识获取的工作包括 DBpedia、Freebase、KnowItAll、WikiTaxonomy 和 YAGO，以及 BabelNet、ConceptNet、DeepDive、NELL、Probase、Wikidata、XLORE、Zhishi.me、CNDBpedia 等。这些知识图谱遵循 RDF 数据模型，包含千万级或者亿级规模的实体，以及数十亿或百亿事实（即属性值和与其他实体的关系），并且这些实体被组织在成千上万的由语义体现的客观世界的概念结构中。

目前知识图谱的发展和应用状况，除了通用的大规模知识图谱，各行业也在建立行业和领域的知识图谱，当前知识图谱的应用包括语义搜索、问答系统与聊天、大数据语义分析以及智能知识服务等，在智能客服、商业智能等真实场景体现出广泛的应用价值，而更多知识图谱的创新应用还有待开发。知识图谱目前的主要应用如下。

- **知识融合：**当前互联网大数据具有分布异构的特点，通过知识图谱可以

对这些数据资源进行语义标注和链接，建立以知识为中心的资源语义集成服务；

- **语义搜索和推荐：**知识图谱可以将用户搜索输入的关键词，映射为知识图谱中客观世界的概念和实体，搜索结果直接显示出满足用户需求的结构化信息内容，而不是互联网网页；
- **问答和对话系统：**基于知识的问答系统将知识图谱看成一个大规模知识库，通过理解将用户的问题转化为对知识图谱的查询，直接得到用户关心问题的答案；
- **大数据分析决策：**知识图谱通过语义链接可以帮助理解大数据，获得对大数据的洞察，提供决策支持。

近年来，在知识工程领域，知识获取、推理和应用研究取得了显著的进展，主要表现在如下几个方面。

- **资源匮乏情况下的知识获取**

知识图谱的构建始终是知识图谱领域的核心问题之一，近年来除了传统的有监督的实体、关系、事件知识获取的研究外，也涌现了一批在弱资源情况下的知识获取方法。例如：在集合扩展（实体集扩展）研究中，Learning to Bootstrap for Entity Set Expansion 使用蒙特卡洛树搜索策略的 bootstrap 方法有效地提升了实体集扩展方法的稳定性，尤其是在与分类体系相关任务的同时优化上。HiExpan: Task-Guided Taxonomy Construction by Hierarchical Tree Expansion 提出一个知识分类体系的扩展框架，模型利用弱监督关系抽取模型，从一个小型的上下位关系树出发，抽取扩展的节点并扩展成一个更加丰富的上下位体系。FewRel 2.0: Towards More Challenging Few-Shot Relation Classification 提出了少次学习任务，通过设计少次学习机制，能够利用从过往数据中学到的泛化知识，结合新类型数据的少量训练样本，实现快速迁移学习。COMET: Commonsense Transformers for Automatic Knowledge Graph Construction 提出常识 Transformer 架构，将 GPT-2 等语言模型与种子知识图谱相结合，学习其结构和关系，根据图表征形成语言模型，从而生成新的知识并将它们添加到种子图中。

- **知识图谱的知识补全和可解释推理**

传统的表示学习缺乏可解释性，知识图谱推理越来越受到关注，其中既有使用强化学习方法寻找路径的方法，也有使用实体邻居和注意力权重做可解释性推理方法。*Multi-Hop Knowledge Graph Reasoning with Reward Shaping*²⁵是基于多跳推理的知识库问答方法，基于强化学习扩展在知识图谱的推理路径，以获得问题的正确答案。*Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs*²⁶提出一种基于注意力机制的特征嵌入方法，获取实体邻近范围内的实体和关系特征，引入关系聚类和多跳关系，有效提升了基于多跳推理的知识图谱补全的效果。*Iteratively Learning Embeddings and Rules for Knowledge Graph Reasoning*²⁷研究如何迭代地进行知识表示学习和规则学习，提出的 IterE 模型可以利用学习的规则改进稀疏实体的表示学习，进而提升规则学习和链接预测效果。

- 基于知识图谱的推荐和对话问答

将知识图谱作为辅助信息引入到推荐系统中，可以有效地解决传统推荐系统存在的稀疏性和冷启动问题，近几年吸引大量研究人员进行相关工作。随着图卷积神经网络，图注意力机制等技术的逐渐兴起，基于图表示学习的推荐模型达到了更高的表现效果，并为推荐系统的可解释性提供了帮助。*KGAT: Knowledge Graph Attention Network for Recommendation*²⁸利用知识图谱中商品之间的关系，训练了一个端到端的含注意力机制的模型，用于提高推荐系统的能力。*AKUPM: Attention-Enhanced Knowledge-Aware User Preference Model for Recommendation*²⁹使用注意力模型，利用知识图谱对用户进行建模，显著提升了推荐系统的效果。

²⁵ Lin, X. V., Socher, R., & Xiong, C. (2018). Multi-hop knowledge graph reasoning with reward shaping. arXiv preprint arXiv:1808.10568.

²⁶ Nathani, D., Chauhan, J., Sharma, C., & Kaul, M. (2019). Learning attention-based embeddings for relation prediction in knowledge graphs. arXiv preprint arXiv:1906.01195.

²⁷ Zhang, W., Paudel, B., Wang, L., Chen, J., Zhu, H., Zhang, W., ... & Chen, H. (2019, May). Iteratively learning embeddings and rules for knowledge graph reasoning. In The World Wide Web Conference (pp. 2366-2377).

²⁸ Wang, X., He, X., Cao, Y., Liu, M., & Chua, T. S. (2019, July). Kgat: Knowledge graph attention network for recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 950-958).

²⁹ Tang, X., Wang, T., Yang, H., & Song, H. (2019, July). AKUPM: Attention-enhanced knowledge-aware user preference model for recommendation. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 1891-1899).

*Reinforcement Knowledge Graph Reasoning for Explainable Recommendation*³⁰结合强化学习的框架和知识图谱推理来提供对推荐结果的解释。在对话问答方面，以前对话生成的信息源是文本与对话记录，但如果遇到词表之外的（Out-of-Vocabulary）的词，模型往往难以生成合适的、有信息量的回复，而会产生一些低质量的、模棱两可的回复。*Commonsense Knowledge Aware Conversation Generation with Graph*³¹提出一种基于常识知识图谱的对话模型 CCM 来理解对话，产生信息丰富且合适的回复。

5.3.3 知识工程领域技术研究趋势

借助 AMiner 平台，从知识工程领域相关的论文中抽取关键词，然后统计这些关键词的起止年份，划分时间窗格，再按照窗格统计词频，最终生成待研究领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-34 所示。

³⁰ Xian, Y., Fu, Z., Muthukrishnan, S., De Melo, G., & Zhang, Y. (2019, July). Reinforcement knowledge graph reasoning for explainable recommendation. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 285-294).

³¹ Zhou, H., Young, T., Huang, M., Zhao, H., Xu, J., & Zhu, X. (2018, July). Commonsense Knowledge Aware Conversation Generation with Graph Attention. In IJCAI (pp. 4623-4629).

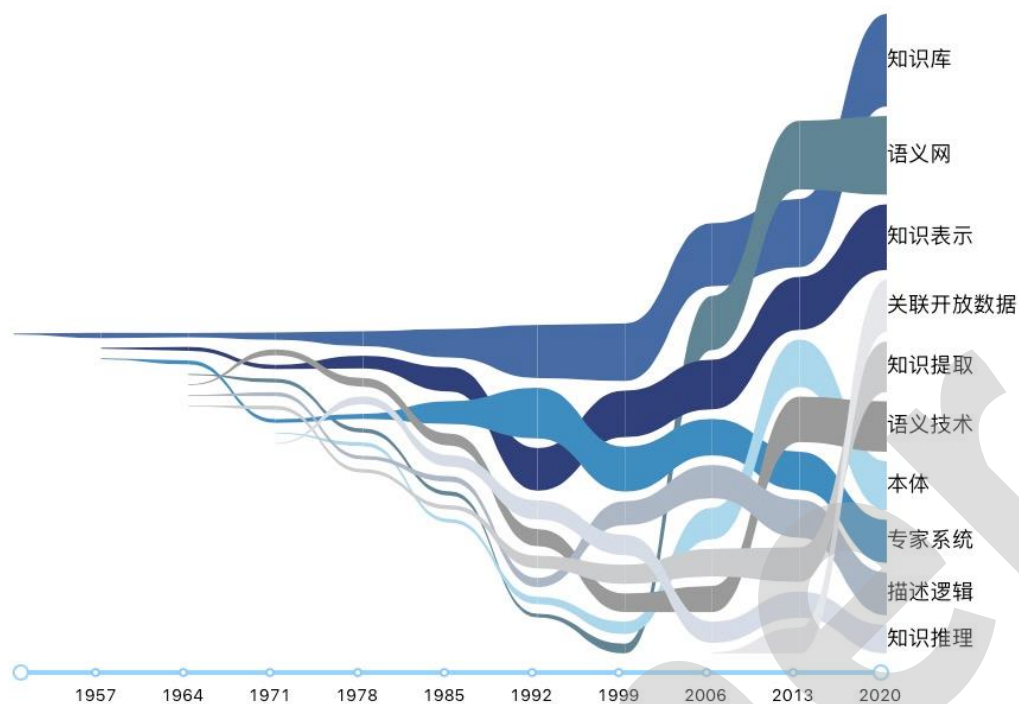


图 5-34 知识工程领域技术研究发展趋势

知识工程领域研究热度较高的 10 个技术话题，包括：知识库（Knowledge Base）、语义网（Semantic Web）、知识表示（Knowledge Representation）、关联开放数据（Linked Open Data）、知识提取（Knowledge Extraction）、语义技术（Semantic Technology）、本体（Ontology）、描述逻辑（Description Logic）、专家系统（Expert System）、知识推理（Knowledge Reasoning），如图 5-34 所示。从中可以看出，近年来知识库、语义网和知识表示等技术话题的研究热度上升较快，与知识图谱（一种语义网络结构的知识库）的热门研究密切相关。

5.4 信息检索与推荐

5.4.1 信息检索与推荐概念

信息检索是信息按一定的方式进行加工、整理、组织并存储起来，再根据信息用户特定的需要将相关信息准确的查找出来的过程，涉及信息的表示、存储、组织和访问，其主要目的是获取与需求匹配的信息。信息推荐是指系统向用户推荐用户可能感兴趣但又没有获取的有效信息，其实现主要依靠推荐系统。

信息检索（Information Retrieval，IR）是计算机科学的一大领域，主要研究如何为用户访问他们感兴趣的信息提供各种便利的手段，即：信息检索涉及对文档、网页、联机目录、结构化和半结构化记录及多媒体对象等信息的表示、存储、组织和访问，信息的表示和组织必须便于用户访问他们感兴趣的信息。“好”的搜索算法是需要让用户获取信息的效率更高、停留时间更短。如今，信息检索研究包括用户建模、Web 搜索、文本分析、系统构架、用户界面、数据可视化、过滤和语言处理等技术。

信息检索的主要环节包括信息内容分析与编码、组成有序的信息集合以及用户提问处理和检索输出。其中信息提问与信息集合的匹配、选择是整个环节中的重要部分。信息检索框架如图 5-35 所示。

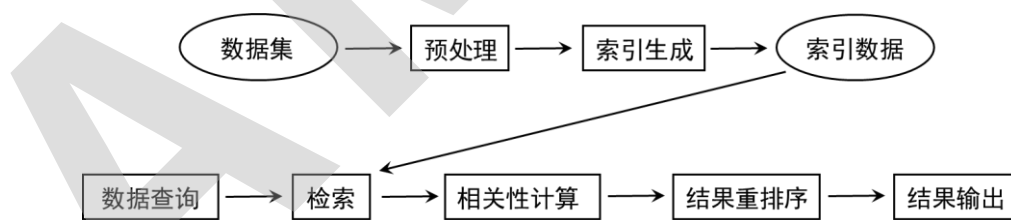


图 5-35 信息检索系统流程

推荐系统（Recommendation System，RS）是指信息过滤技术，从海量项目（项目是推荐系统所推荐内容的统称，包括商品、新闻、微博、音乐等产品及服务）中找到用户感兴趣的部分并将其推荐给用户，这在用户没有明确需求或者项目数量过于巨大、凌乱时，能很好地为用户服务，解决信息过载问题^{错误!未找到引用源。}。由于推荐系统在个性化方面的运作空间大，推荐系统又被称为“个性化推荐”“智

能推荐”。

推荐系统模型流程通常由用户特征收集、用户行为建模与分析、推荐与排序三个重要模块组成，如图 5-36 所示。

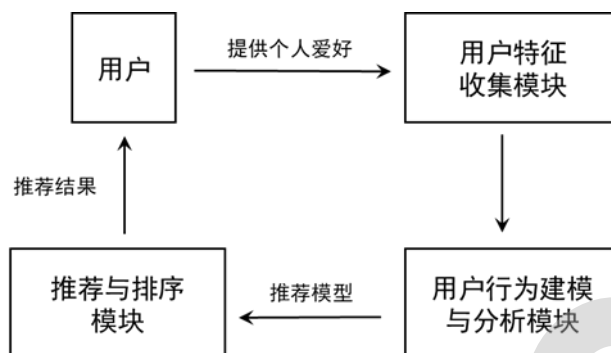


图 5-36 推荐系统模型流程

信息的检索与推荐都是用户获取有用信息的手段，这两种方式是互补并存的关系。与搜索引擎不同的，推荐系统不需要用户提供明确的需求，而是通过分析用户的历史行为给用户的兴趣建模，从而主动给用户推荐能够满足他们兴趣和需求的信息。因此，搜索引擎满足了用户有明确目的的主动查找需求，而推荐系统能够在用户没有明确目的的时候帮助他们发现感兴趣的新内容。此外，推荐算法和被推荐的内容往往是紧密结合在一起的，用户获取推荐结果的过程可以是持续的、长期的，衡量推荐系统是否足够好，往往要依据是否能让用户停留更多的时间，对用户兴趣的挖掘越深入，越“懂”用户，那么推荐的成功率越高，用户粘性也越高。

5.4.2 信息检索与推荐技术过去十年主要进展

信息检索发展经历了三个阶段：（1）数字图书馆 / 文档电子化时代（20 世纪 50-60 年代）：信息检索技术主要应用于封闭数据集合、单机模式或专网内的主机-终点模式，在商业应用方面，则是提供软件/解决方案，专网内的查询服务。

（2）早期互联网时代（20 世纪 70-90 年代），信息检索的应用形态特征是开放的、大规模的、实时的、多媒体的，尤其巨型搜索引擎采集到的公开数据和用户访问日志等非公开数据深刻地影响着这一时期信息检索领域的创新模式。谷歌开

发出了著名的 PageRank 搜索引擎算法，将互联网中大多数的网页通过基于链接来计算网页质量的方式进行排名，为搜索引擎用户提供较好的基于链接查询的搜索结果。（3）Web 2.0 时代（2000 年以后），信息搜索的发展更加关注用户需求，以实现内容与行为的精准 Web 搜索；此外，信息检索实现了内容数据与社会各侧面的电子化数据（万维网、社交网、物联网、地理信息等）的全面融合；尤其是对社交网络数据的采集和大数据处理技术出现了社会化趋势。

信息推荐系统的研究始于 1994 年明尼苏达大学 Group Lens 研究组推出的 Group Lens 系统与协同过滤的思想，之后，先后经历了卡耐基·梅隆大学 Robert Armstrong 等人提出的个性化导航系统 Web Watcher；斯坦福大学 Marko Balabanovic 等人推出的个性化推荐系统 LIRA；1997 年，AT&T 实验室提出了基于协同过滤的个性化推荐系统 PHOAKS 和 Referral Web；2000 年，NEC 研究院的 Kurt 等人为搜索引擎 Cite Seer 增加了个性化推荐功能；2003 年，Google 推出的 AdWords 盈利模式，通过用户搜索的关键词来提供精准的广告。

信息推荐系统的演变始终伴随着网络的发展，第一代信息推荐系统使用传统网站从以下三个来源收集信息：来自购买或使用过的产品的基础内容数据；用户记录中收集的人口统计数据；以及从用户的项目偏好中收集的基于记忆的数据。第二代推荐系统通过收集社交信息，例如朋友、关注者、粉丝等。第三代推荐系统使用网上集成设备提供的信息。

蕴含着智能推荐系统和强大的搜索引擎的在线商店、在线音乐、在线视频和图片库等已成为人们快速寻找信息的主要方式。此类系统的流行程度和有用性在于它们能够便捷地显示几乎无限的物品信息。比如，Amazon、Netflix 等推荐系统寻找用户感兴趣商品的核心机制都是用户兴趣与商品匹配的机制。

近年来，为了提高信息检索与推荐系统中算法模型的准确性和可解释性，研究人员主要关注无偏的在线排序学习模型，以及利用知识信息增强推荐系统的表现和可解释性等方面的研究。其中，无偏的在线排序学习模型是指自动利用大规模用户点击数据训练搜索结果的排序模型。用户点击数据是现代搜索引擎的重要数据来源，具有成本低廉，并且对以用户为中心的检索应用程序（如搜索排名）特别有用等优点。为了充分利用用户点击数据开发一个无偏的学习排名系统，研究人员试图消除用户偏见对排名模型训练的影响。

此外，一种基于反事实学习和图形模型的无偏学习排名框架引起了人们的广泛关注。该框架侧重于使用反事实学习直接训练带有偏倚点击数据的排名模型。这个无偏的学习排名框架对待点击偏差作为一个反事实的影响和去偏用户反馈加权每点击与他们的反向倾向加权。它使用倾向性模型来量化点击的偏差，并没有明确地估计查询文档与培训数据的相关性。研究人员从理论上证明，在正确的偏差估计下，在该框架下使用点击数据训练的排序模型将收敛于使用真实相关信号训练的排序模型。

信息检索与推荐系统可以为用户推荐其感兴趣的内容并给出个性化的建议。而现在的推荐系统大都着眼于被推荐对象的序列建模，而忽略了它们细粒度的特征。为了解决以上问题，研究人员提出了多任务可解释推荐模型（Multi-Task Explainable Recommendation, MTER）和知识增强的序列推荐模型（Knowledge-enhanced Sequential Recommender, KSR）。其中，MTER 模型是一个用于可解释推荐任务的多任务学习方法，通过联合张量分解将用户、产品、特征和观点短语映射到同一向量空间，来从用户评论中提取产品细粒度的个性化特征。KSR 模型提出了利用结合知识库的记忆网络来增强推荐系统的特征捕获能力与解释性，解决序列化推荐系统不具有解释性，且无法获取用户细粒度特征的不足。MTER 和 KSR 模型通过对推荐结果的解释，分析被推荐对象的特征，让用户可以对使用哪些推荐结果做出更明智，更准确的决策，从而提高他们的满意度。

2016 年，为了协助信息检索与推荐领域的算法模型的训练和优化，微软公司提供了一个大规模支持机器阅读理解和问答系统等多种领域研究的数据集 MS MACRO (Microsoft Machine Reading Comprehension)。该数据集从必应 (bing) 的搜索查询记录中取样，每个问题都有人工生成的答案和完全人工重写的答案。此外，数据集包含从通过 bing 检索的 web 文档中提取的百万个密码，这些密码提供了管理自然语言答案所需的信息。

近年来，信息检索与推荐领域比较流行的开源平台主要包括基于深度学习的检索模型 (MatchZoo)、基于 tensorflow 的 learning to rank 模型 (TF-Ranking) 和 microsoft recommenders。

2017 年，中国科学院计算技术研究所网络数据科学与技术重点实验室发布了深度文本匹配开源项目 MatchZoo。MatchZoo 是一个 Python 环境下基于

TensorFlow 开发的开源文本匹配工具，使用了 Keras 中的神经网络层，并有数据预处理，模型构建，训练与评测三大模块组成，旨在让大家更加直观地了解深度文本匹配模型的设计、更加便利地比较不同模型的性能差异、更加快捷地开发新型的深度匹配模型。MatchZoo 提供了基准数据集（TREC MQ 系列数据、WiKiQA 数据等）进行开发与测试，整合了当前最流行的深度文本匹配的方法（包括 DRMM, MatchPyramid, DUET, MVLSTM, aNMM, ARC-I, ARC-II, DSSM, CDSSM 等算法的统一实现），旨在为信息检索、数据挖掘、自然语言处理、机器学习等领域内的研究与开发人员提供便利，可以应用到的任务场景包括文本检索、自动问答、复述问题、对话系统等。

TF-Ranking 是一个可扩展的基于 tensorflow 的用于排序的库，由 google 于 2018 年提出。TF-Ranking 提供了一个统一的框架，其中包括一套最先进的学习排序算法，并支持成对或列表损失函数、多项评分、排序度量优化和无偏学习排序。TF-Ranking 速度很快并且易于使用，可以创建高质量的排序模型。统一的框架使机器学习的研究人员、实践者和爱好者能够在一个库中评估和选择一系列不同的排序模型。此外，这个开源库不仅提供了合理的默认模型，还可以让用户能够开发自己的定制模型，且提供了灵活的 API，用户可以在其中定义和插入自己定制的损失函数、评分函数和指标。Microsoft Recommenders 是微软云计算和人工智能开发团队与微软亚洲研究院团队深度合作，基于多年来各类大型企业级客户的项目经验以及最新学术研究成果，搭建的完整推荐系统的最新实操技巧开源项目。该项目有效解决了定制和搭建企业级推荐系统中的几个难点，包括如何将学术研究成果或开源社区提供的范例适用于企业级应用、如何集成信息检索与推荐领域的学习指导资源协助从业人员深入理解并实际搭建完整推荐系统、如何选择最优算法以应对具体应用场景等。

5.4.3 信息检索与推荐技术研究趋势

获取高质量的信息永远是人类的核心需求。信息检索和推荐过去十年经历了蓬勃发展。以搜索引擎为代表的一系列信息检索与推荐工具成为了人们日常生活中不可或缺的工具之一。随着互联网的高速发展和移动设备的普及，用户的信息获取方式和使用场景也逐渐发生变化。未来几年，信息检索、挖掘与推荐将朝着

更加智能化、准确化、个性化、专业化的方向发展，传统的信息检索、问答、对话和推荐技术将更加紧密地结合，使得信息检索与挖掘领域飞速发展。

借助 AMiner 平台，生成的信息检索与挖掘领域的技术研究发展趋势图结果如图 5-37 所示。

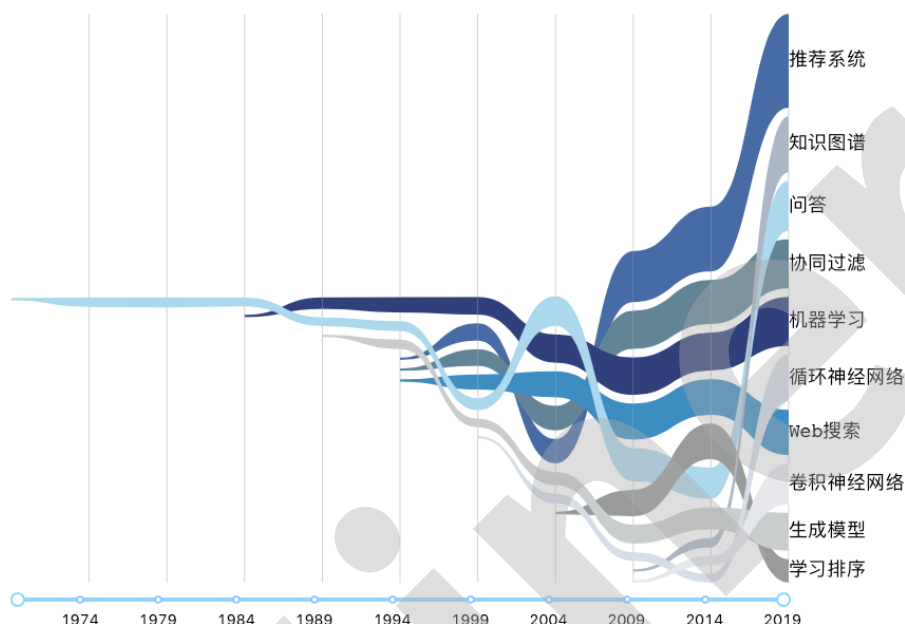


图 5-37 信息检索与推荐领域技术研究发展趋势

信息检索与推荐领域研究热度较高的 10 个技术话题，包括：推荐系统（Recommender System）、知识图谱（Knowledge Graph）、问答（Question answering）、协同过滤（Collaborative Filtering）、机器学习（Machine Learning）、循环神经网络（Recurrent Neural Network）、Web 搜索（Web Search Engine）、卷积神经网络（Convolutional Neural Network）、生成模型（Generative Model）、学习排序 Learning To Rank），如图 5-37 所示。近年来，推荐系统、知识图谱和问答等技术话题的研究热度上升速度较快，逐渐发展为高热度研究话题。知识图谱由谷歌在 2012 年首次提出，初衷是提升谷歌搜索引擎的质量。知识图谱因为其强大的信息组织和存储能力，在信息检索与推荐领域应用广泛。

5.5 计算机视觉

5.5.1 计算机视觉概念

计算机视觉（Computer Vision）技术利用摄像机以及电脑替代人眼对目标进行识别、跟踪、测量和判别决策等，并进一步做图形处理，使电脑处理的信息成为更适合人眼观察或传送给仪器检测的图像。计算机视觉技术是一门包括计算机科学与工程、神经生理学、物理学、信号处理、认知科学、应用数学与统计等多学科的综合性科学技术，是人工智能的一个重要分支，目前在智能安防、自动驾驶汽车、医疗保健、生成制造等领域具有重要的应用价值。

计算机视觉的研究目标是使计算机具备人类的视觉能力，能看懂图像内容、理解动态场景，期望计算机能自动提取图像、视频等视觉数据中蕴含的层次化语义概念及多语义概念间的时空关联等。

计算机视觉的研究内容，大体可以分为物体视觉（Object Vision）和空间视觉（Spatial Vision）两大部分。物体视觉在于对物体进行精细分类和鉴别，而空间视觉在于确定物体的位置和形状，为“动作（action）”服务。正像著名的认知心理学家 J.J.Gibson 所言，视觉的主要功能在于“适应外界环境，控制自身运动”。适应外界环境和控制自身运动，是生物生存的需求，这些功能的实现需要靠物体视觉和空间视觉协调完成。

计算机视觉包括了诸多不同的研究方向，比较基础和热门的方向有物体识别和检测（Object Detection）、语义分割（Semantic Segmentation）、运动和跟踪（Motion & Tracking）、视觉问答（Visual Question & Answering）等^{错误!未找到引用源。}。

物体识别和检测，是指对于给定的一张输入图片，算法能够自动找出图片中的常见物体，并将其所属类别及位置输出出来。衍生出了诸如人脸检测（Face Detection）、车辆检测（Vehicle Detection）等细分类的检测算法。

语义分割可以看作是一种特殊的分类，即将输入图像的每一个像素点进行归类，用一张图就可以很清晰地描述出来。物体检测和识别通常是将物体在原图像上框出，而语义分割是从每一个像素上进行分类，图像中的每一个像素都有属于

自己的类别。

跟踪的评判标准主要是在一段给定的视频中，在第一帧给出被跟踪物体的位置及尺度大小，在后续的视频当中，跟踪算法需要从视频中去寻找到被跟踪物体的位置，并适应各类光照变换、运动模糊以及表观的变化等。跟踪研究近年来由过去的非深度算法跨越向了深度学习算法，精度也越来越高，不过实时的深度学习跟踪算法精度一直难以提升，而精度非常高的跟踪算法的速度又十分之慢，因此在实际应用中也很难派上用场。目前而言，跟踪的很多算法都是改进自检测或识别算法。

视觉问答 (Visual Question Answering, VQA) 的研究目的旨在根据输入图像，由用户进行提问，而算法自动根据提问内容进行回答。除了问答以外，还有一种算法被称为标题生成算法 (Caption Generation)，即计算机根据图像自动生成一段描述该图像的文本，而不进行问答。对于这类跨越两种数据形态（如文本和图像）的算法，有时候也可以称之为多模态或跨模态问题。

5.5.2 计算机视觉过去十年主要进展

计算机视觉研究兴起于 20 世纪 60 年代初。1982 年马尔 (David Marr)《视觉》一书的问世，可以看成计算机视觉成为了一门独立学科的标志。在计算机视觉 40 多年的发展中，总体上经历了三个主要历程：马尔计算视觉 (Computational Vision)、多视几何与分层三维重建和基于学习的视觉。过去十年来，巨量数据的不断涌现与计算能力的快速提升，给以非结构化视觉数据为研究对象的计算机视觉带来了巨大的发展机遇与挑战性难题；计算机视觉的部分研究成果已实际应用，催生出人脸识别、物体识别与分类、智能视频监控等多个极具显示度的商业化应用。

过去十年来，多视几何理论已基本完善。多视几何理论下的分层三维重建能有效提高三维重建的鲁棒性和精度。2000 年美国 GE 研究院的 R.Hartley 和英国牛津大学的 A.Zisserman 合作的 *Multiple View Geometry in Computer Vision* 著作 (Hartley & Zisserman, 2000) 对这方面的内容给出了比较系统的总结。而后这方面的工作主要集中在如何提高“大数据下鲁棒性重建的计算效率”。如何在保

证鲁棒性的前提下快速进行大场景的三维重建是后期研究的重点。

基于机器学习的视觉研究可以分为两个阶段：本世纪初的以流形学习为代表的子空间法和目前以深度学习为代表的视觉方法。流形学习是从图像表达学习其“内在流形”（Intrinsic Manifold）表达的过程，这种学习过程一般是一种非线性优化过程。目前深度学习方法被引入到计算机视觉中。

2012 年 6 月，谷歌研究人员 Jeff Dean 和吴恩达从 YouTube 视频中提取了 1000 万个未标记的图像，训练一个由 16000 个电脑处理器组成的庞大神经网络。在没有给出任何识别信息的情况下，人工智能通过深度学习算法准确地从中识别出了猫科动物的照片。

随着以卷积神经网络（CNN）的深度学习技术的发展，以图像去噪、图像超分辨和图像去模糊等为代表的图像复原与增强以及以视觉生成与合成为代表的图像编辑问题获得了较多关注。2014 年，Schmidt 等学者^[32]针对图像复原问题，提出了基于逐次迭代学习的判别学习模型，即基于半二次分裂算法的级联收缩场（Cascade of Shrinkage Fields, CSF），该方法通过将预测过程展开为迭代学习算法，从训练数据中学习阶段模型参数。2015 年，Chen 等学者^[33]从反应扩散方程的角度出发，提出了非线性反应扩散（Trainable Nonlinear Reaction Diffusion, TNRD）模型，对每次迭代的滤波器和响应函数进行学习，并从递归神经网络的角度对模型进行了解释。受 CSF 和 TNRD 启发，一些学者^[34]设计了一种基于卷积神经网络的深度去噪网络 DnCNN（Denoising Convolutional Neural Network），该模型通过端到端的残差学习，从函数回归角度用卷积神经网络将噪声从噪声图像中分离出来，取得了显著优于其他方法的去噪结果。

32 Schmidt U, Roth S. Shrinkage Fields for Effective Image Restoration[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 2774-2781.

33 Chen Y, Pock T. Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1256-1272.

34 Yuan Q, Zhang Q, Li J, et al. Hyperspectral Image Denoising Employing a Spatial-spectral Deep Residual Convolutional Neural Network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(2): 1205-1218.

5.5.3 计算机视觉研究趋势

借助 AMiner 平台，从计算机视觉领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-38 所示。

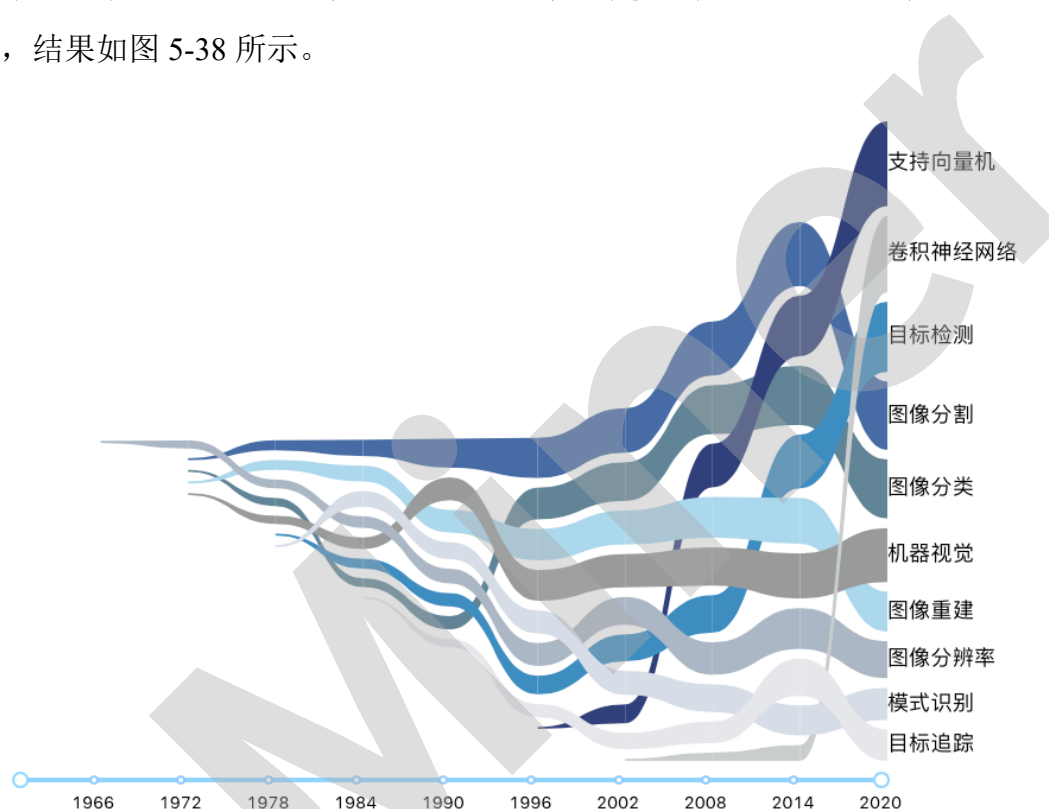


图 5-38 计算机视觉领域技术研究发展趋势

计算机视觉领域研究热度较高的十个技术话题，包括：支持向量机（Support Vector Machine）、卷积神经网络（Convolutional Neural Network）、目标检测（Object Detection）、图像分割（Image Segmentation）、图像分类（Image Classification）、机器视觉（Machine Vision）、图像重建（Image Reconstruction）、图像分辨率（Image Resolution）、模式识别（Pattern Recognition）、目标追踪（Object Tracking），如上图所示。其中，支持向量机和卷积神经网络在图像分类、目标检测、图像分割等多个计算机视觉应用领域具有显著的效果，该技术研究热度持续攀升。

随着计算机视觉研究的不断推进，研究人员开始挑战更加困难的计算机视觉问题，例如，图像描述、事件推理、场景理解等。为了解决更加复杂的图像理解

任务，一个重要的研究趋势是多学科融合，例如，融合自然语言处理领域的技术来完成图像描述的任务。图像描述是一个融合计算机视觉、自然语言处理和机器学习的综合问题，其目标是翻译一幅图片为一段描述文字。目前主流框架为基于递归神经网络的编码器解码器结构，其核心思想类似于自然语言机器翻译。但是，由于递归网络不易提取输入图像和文本的空间以及层次化约束关系，层次化的卷积神经网络以及启发自认知模型的注意力机制受到关注。如何进一步从认知等多学科汲取知识，构建多模态多层次的描述模型是当前图像描述问题研究的重点。

事件推理目标是识别复杂视频中的事件类别并对其因果关系进行合理的推理和预测。与一般视频分析相比，其难点在于事件视频更加复杂，更加多样化，而最终目标也更具挑战性。不同于大规模图像识别任务，事件推理任务受限于训练数据的规模，还无法构建端到端的事件推理系统。目前主要使用图像深度网络作为视频的特征提取器，利用多模态特征融合模型，并利用记忆网络的推理能力，实现对事件的识别和推理认知。如何利用视频数据丰富的时空关系以及事件之间的语义相关性，将是今后的关注重点。

场景理解的目的是计算机视觉系统通过分析处理自身所配置的传感器采集的环境感知数据，获得周围场景的几何/拓扑结构、组成要素（人、车及物体等）及其时空变化，并进行语义推理，形成行为决策与运动控制的时间、空间约束。近年来，场景理解已经从一个初期难以实现的目标成为目前几乎所有先进计算机视觉系统正在不断寻求新突破的重要研究方向。利用社会-长短记忆网络（Social-LSTM）实现多个行人之间的状态联系建模，结合各自运动历史状态，决策出未来时间内的运动走向。视觉技术的发展需要设计新的模型，它们需要能考虑到空间和时间信息；需要高质量的人类检测和视频对象检测数据集；需要结合文本和声音的跨模态集成；也需要在与世界的交互中学习。

5.6 语音识别

5.6.1 语音识别概念

语音识别是让机器识别和理解说话人语音信号内容的技术。其目的是将语音信号转变为计算机可读的文本字符或者命令的智能技术,利用计算机理解讲话人的语义内容,使其听懂人类的语音,从而判断说话人的意图。语音识别技术主要包括特征提取技术、模式匹配准则及模型训练技术三个方面。

语音识别是一种非常自然和有效的人机交流方式,与很多学科紧密相连,比如语言学、声学、信息理论、信号处理、计算机科学、心理和生理学等。

语音识别包含两个阶段:第一个阶段是学习和训练,即提取语音库中语音样本的特征参数作为训练数据,合理设置模型参数的初始值,对模型各个参数进行重估,使识别系统具有最佳的识别效果;第二个阶段就是识别,将待识别语音信号的特征根据一定的准则与训练好的模板库进行比较,最后通过一定的识别算法得出识别结果。识别结果的好坏与模板库是否准确、模型参数的好坏以及特征参数的选择都有直接的关系。

语音识别的基本结构如图 5-39 所示。语音识别首先要对采集的语音信号进行预处理,然后利用相关的语音信号处理方法计算语音的声学参数,提取相应的特征参数,最后根据提取的特征参数进行语音识别。其中,预处理主要是对输入语音信号进行预加重和分段加窗等处理,并滤除其中的不重要信息及背景噪声等,然后进行端点检测,以确定有效的语音段。特征参数提取是将反映信号特征的关键信息提取出来,以此降低维数减小计算量,用于后续处理,这相当于一种信息压缩。之后进行特征参数提取,用于语音训练和识别。常用的特征参数有基于时域的幅度、过零率、能量以及基于频域的线性预测倒谱系数、Mel 倒谱系数等。

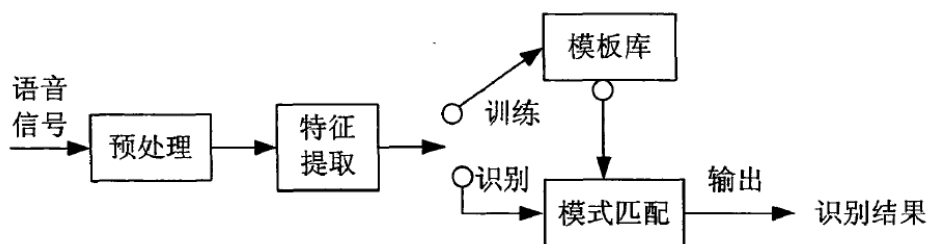


图 5-39 语音识别系统框架

5.6.2 语音识别过去十年主要进展

语音识别的研究工作可以追溯到 20 世纪 50 年代。1952 年，AT & T 贝尔研究所的 Davis, Biddulph 和 Balashek 研究成功了世界上第一个语音识别系统 Audry 系统，可以识别 10 个英文数字发音。60 年代，计算机的应用推动了语音识别技术的发展，动态规划线性预测分析技术理论被提出，相关研究成果随之出现。70 年代，伴随着自然语言理解的研究以及微电子技术的发展，语音识别领域取得了突破性进展。这一时期的语音识别方法基本上是采用传统的模式识别策略，还提出了矢量量化和隐马尔可夫模型理论。80 年代，语音识别研究进一步走向深入，隐马尔可夫模型（HMM）技术的成熟和不断完善，并最终成为语音识别的主流方法；以知识为基础的语音识别的研究日益受到重视；人工神经网络（ANN）在语音识别中的应用研究兴起。90 年代，语音识别技术逐渐走向实用化，在建立模型、提取和优化特征参数方面取得了突破性的进展，使系统具有更好的自适应性。许多具有代表性的产品问世。比如 IBM 公司研发的汉语 ViaVoice 系统，以及 Dragon 工司研发的 DragonDictate 系统，都具有说话人自适应能力，能在用户使用过程中不断提高识别率。

21 世纪之后，深度学习技术极大地促进了语音识别技术的进步和应用的广泛发展，大大提高了识别精度。语音识别技术在手机、家电、游戏机等嵌入式设备中得到了大量应用，并主要应用于语音的控制以及文本内容的输入中。2009 年，Hinton 将深度神经网络（DNN）应用于语音的声学建模，在 TIMIT 上获得了当时的最好结果。

过去十年，随着人工智能机器学习领域深度学习研究的发展，以及大数据语料的积累，语音识别技术得到突飞猛进的发展进步。随着计算机小型化的发展趋

势出现,语音识别成为人机接口的关键一步。语音识别技术已逐渐被应用于工业、通信、商务、家电、医疗、汽车电子以及家庭服务等各个领域,逐渐走入人们的日常生活。例如,现今流行的手机语音助手(美国苹果公司的 Siri 语音助手、智能 360 语音助手、百度语音助手等),就是将语音识别技术应用到智能手机中,能够实现人与手机的智能对话功能。

2011 年底,微软研究院的俞栋、邓力又把 DNN 技术应用在了大词汇量连续语音识别任务上,大大降低了语音识别错误率。从此语音识别进入 DNN-HMM 时代。DNN 带来的好处是不再需要对语音数据分布进行假设,将相邻的语音帧拼接又包含了语音的时序结构信息,使得对于状态的分类概率有了明显提升。同时 DNN 还具有强大环境学习能力,可以提升对噪声和口音的鲁棒性。

语音识别经历了从 2012 年最开始的 DNN 的引入时的 Hybrid HMM 结构,再到 2015 年开始吸引大家研究兴趣的 CTC 算法,而后到 2018 年的 Attention 相关结构的研究热点。Attention 相关算法在语音识别或者说话人识别研究的文章中出现频率极高。从最开始 Attention,到 Listen-Attend-Spell,再到 Self-Attention (或者 Transformer),在不同的文章被作者多次介绍和分析,频繁出现在了相关文章的 Introduction 环节中。在 Attention 结构下,依然还有很多内容需要研究者们进一步地探索:例如在一些情况下 Hybrid 结构依然能够得到 State-of-the-art 的结果,以及语音数据库规模和 Attention 模型性能之间的关系。

近年来智能语音进入了快速增长期,语音识别作为语音领域的重要分支获得了广泛的关注,如何提高声学建模能力和如何进行端到端的联合优化是语音识别领域中的重要课题。

在近两年的研究中,端到端语音识别仍然是 ASR (Automatic Speech Recognition) 研究的一大热点。基于 Attention 机制的识别系统已经成为了语音技术研究主流。同时,随着端到端语音识别框架日益完善,研究者们对端到端模型的训练和设计更加的关注。远场语音识别(far-field ASR)、模型结构(ASR network architecture)、模型训练(model training for ASR)、跨语种或者多语种语音识别(cross-lingual and multi-lingual ASR)以及一些端到端语音识别(end-to-end ASR)成为研究热点。

在语音合成方面，高音质语音生成算法及 Voice conversion 是近两年研究者关注的两大热点，Voice Conversion 方向的研究重点主要集中在基于 GAN 的方法上。在语言模型方面（Language Model）的研究热点主要包括 NLP 模型的迁移，低频单词的表示，以及深层 Transformer 等。

在说话人识别方面，说话人信息，特别是说话人识别及切分，正被越来越多的研究者所重视。目前 Attention 在说话人方面更类似一种 Time Pooling，比 Average Pooling 及 Stats Pooling 更能捕捉对说话人信息更重要的信息，从而带来性能提升。说话人识别技术经历深度学习带来的性能飞跃后，在模型结构、损失函数等方面的探讨已经较为成熟，以 TDNN、ResNet 加上 LMCL、ArcFace 的主流模型开始不断刷新各数据集的性能上限。模型以外的因素逐渐成为制约说话人系统的瓶颈。说话人技术目前也逐渐暴露出与人脸识别同样的易受攻击的问题。因此，ASVspoof 这样的 Challenge 从 2015 年起就开始关注声纹反作弊问题。相信随着此类研究的不断深入，结合声纹系统的性能提升，声纹将有望变成我们的“声音身份证”。

2020 年，Facebook 推出了自动语音识别（ASR）领域的 wave2vec 算法模型，并且公开了该算法细节。Facebook 基于 wav2vec 的模型实现了 2.43% 的单词错误率，准确率高于 Deep Speech 2、监督迁移学习（Supervised Transfer Learning）等主流算法。

5.6.3 语音识别研究趋势

随着人工智能的迅速发展，语音识别的技术越来越成为国内外研究机构的焦点。人们致力于使机器能够听懂人类的话语指令，并希望通过语音实现对机器的控制。作为一项人机交互的关键技术，语音识别在过去的几十年里取得了飞速的发展，在研究和探索过程中针对语音识别的各部流程进行了各种各样的尝试和改造，以期发现更好的方法来完成语音识别流程中的各个步骤，以此来促进在不同环境下语音识别的效率和准确率。研究人员从最简单的非常小词汇量的阅读式的语音识别问题开始，逐渐转向越来越复杂的问题。在未来，语音检索、命令控制、自动客户服务、机器自动翻译等人工智能相关领域成为语音识别的应用前景。

借助 AMiner 平台，从语音识别领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-40 所示。

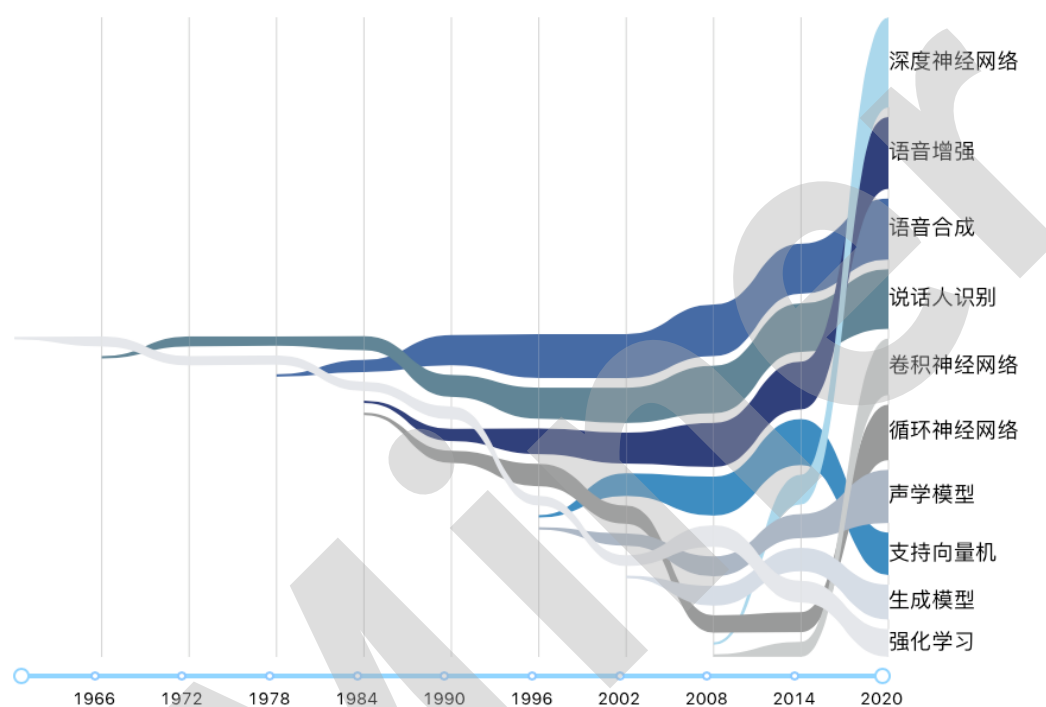


图 5-40 语音识别领域技术研究发展趋势

研究显示，语音识别领域研究热度较高的 10 个技术话题，包括：深度神经网络（Deep Neural Network）、语音增强（Speech Enhancement）、语音合成（Speech Synthesis）、说话人识别（Speaker Recognition）、卷积神经网络（Convolutional Neural Network）、循环神经网络（Recurrent Neural Network）、支持向量机（Support Vector Machine）、主成分分析（Principal Component Analysis）、生成模型（Generative Model）、强化学习（Reinforcement Learning），如图 5-40 所示。近年来，深度神经网络在语音识别领域的研究热度上升很快，它的加入使得语音识别效果有了极大的提高。

5.7 机器人

5.7.1 机器人概念

机器人广义上包括一切模拟人类行为或思想以及模拟其他生物的机械(如机器狗, 机器猫等)。狭义的机器人定义还有很多分类法及争议, 有些电脑程序甚至也被称为机器人(例如爬虫机器人)。目前国际上对于机器人的定义纷繁复杂, RIA、JIRA、NBS、ISO 等组织都有各自的定义, 迄今为止, 尚没有一个统一的机器人定义。联合国标准化组织采纳了美国机器人协会给机器人下的定义: “一种可编程和多功能的操作机; 或是为了执行不同的任务而具有可用电脑改变和可编程动作的专门系统。一般由执行机构、驱动装置、检测装置和控制系统和复杂机械等组成”。

借由模仿逼真的外观及自动化的动作, 机器人是指能够半自主或全自主工作的智能机器, 具有感知、决策、执行能力, 可以辅助甚至替代人类完成危险、繁重、复杂的工作。机器人是自动执行工作的机器装置, 既可以接受人类指挥, 又可以运行预先编排的程序, 也可以根据以人工智能技术制定的原则纲领行动。机器人具有提高工作效率与质量, 服务人类生活, 扩大或延伸人的活动及能力范围等作用, 是人类历史上非常重要的发明成果。

机器人的范围很广, 可以是自主或是半自主的, 从本田技研工业的 ASIMO 或是 TOSY 的 TOPIO 等拟人机器人到工业机器人, 也包括多台一起动作的群机器人, 甚至是纳米机器人。

机器人学(Robotics)的研究目标是以智能计算机为基础的机器人的基本组织和操作, 它包括基础研究和应用研究两方面内容, 研究课题包括机械手设计、机器人动力和控制、轨迹设计与规划、传感器、机器人视觉、机器人控制语言、装置与系统结构和机械智能等。机器人学综合了力学、机械学、电子学、生物学、控制论、传感器、计算机、人工智能、系统工程、材料学和仿生学等多种学科领域的知识。

过去几十年, 机器人技术的研究与应用, 大大推动了人类的工业化和现代化

进程，并逐步形成了机器人的产业链，使机器人的应用范围也日趋广泛。机器人技术最早应用于工业领域。在当代工业中，机器人指能自动执行任务的人造机器装置，用以取代或协助人类工作，一般会是机电装置，由计算机程序或电子电路控制。机器人可以作一些重复性高或是危险、人类不愿意从事的工作，也可以做一些因为尺寸限制，人类无法完成的工作，甚至是像外太空或是深海中，不适人类生存的环境。机器人在越来越多方面可以取代人类，或是在外貌、行为或认知，甚至情感上取代人类。

随着机器人技术的发展和各行业需求的提升，近年来，在计算机技术、网络技术、MEMS 技术等新技术发展的推动下，机器人技术正从传统的工业制造领域向医疗服务、教育娱乐、勘探勘测、生物工程、救灾救援等领域迅速扩展，适应不同领域需求的机器人系统被深入研究和开发。

5.7.2 机器人过去十年主要进展

机器人技术研究可以追溯到二十世纪 40-50 年代，当时美国许多国家实验室进行了机器人方面的初步探索，主要研制了遥控操纵器、遥控式机械手等。1954 年，美国的 George C. Devol 设计并制作了世界上第一台机器人实验装置，发表了《适用于重复作业的通用性工业机器人》一文，并获得了专利。这种机器人的主要技术功能就是“可编程”以及“示教再现”。1961 年，美国麻省理工学院 Lincoln 实验室把一个配有接触传感器的遥控操纵器的从动部分与计算机连结在一起，可以凭触觉决定物体的状态。随后，用电视摄像头作为输入的计算机图像处理、物体辨识的研究工作也陆续取得成果。1968 年，美国斯坦福人工智能实验室（SAIL）的 J. McCarthy 等人研究了带有手、眼、耳的计算机系统。70 年代后，机器人已经拥有类似人的感觉，被称为第二代机器人（感觉型）。机器人产业蓬勃兴起，机器人技术发展为专门的学科。随后的大规模集成电路技术的飞跃发展及微型计算机的普遍应用，则使机器人的控制性能大幅度地得到提高、成本不断降低。80 年代后，随着计算机、传感器技术的发展，机器人已经具备了初步的感知、反馈能力，在日本、德国、美国等国家的工业生产中开始逐步大量应用。90 年代第三代机器人（智能机器人）问世，具有感觉、运动和思考能力，能够进行自主决策。

机器人朝着智能化的方向发展，这种机器人带有多种传感器，能够将多种传

感器得到的信息进行融合，能够有效地适应变化的环境，具有很强的自适应能力、学习能力和自治功能。智能机器人的发展主要经历了三个阶段，分别是可编程试教、再现型机器人，有感知能力和自适应能力的智能机器人。其中所涉及到的关键技术有多传感器信息融合、导航与定位、路径规划、机器人视觉智能控制和人机接口技术等。

21 世纪以来，随着劳动力成本的不断提高、技术的不断进步，各国陆续进行制造业的转型与升级，出现了机器人替代人的热潮。同时，人工智能发展日新月异，服务机器人也开始走进普通家庭的生活。世界上许多机器人科技公司都在大力发展机器人技术，机器人的特质与有机生命越来越接近。

过去十年来，随着 AI 兴起，机器人拥有了一种新型的学习方式：深度强化学习。这一新方式借助通用化的神经网络表示，处理复杂的传感器输入，来让机器人从自己的经验活动中直接学习行为。

波士顿动力公司在机器人领域的产品机器狗 Spot（2005 年研发）和双足人形机器人 Atlas（2013 年）都曾成为焦点。2016 年由谷歌（Google）旗下 DeepMind 公司研发推出的 AlphaGo 成为第一个击败人类职业围棋选手、第一个战胜围棋世界冠军的人工智能机器人，它应用了如神经网络学习、深度学习、蒙特卡洛树搜索法等人工智能技术，实力有了实质性的飞跃。

机器人的学习可分为三个模式，即：强化学习（智能体和环境交互，最大化获得回报）、模仿学习（智能体观测专家和环境交互过程，学习专家行为逻辑）、偏好学习（专家观测并评估智能体和环境交互过程，智能体学习最大化专家评估的策略）。

原来的强化学习算法都还不能够适用于有复杂系统的机器人，不足以支撑机器人在短时间内就学习到行为，另外在安全性上也难以保障。针对这种困境，2019 年初，谷歌 AI 与 UC 伯克利大学合作研发了一种新的强化学习算法 SAC（Soft Actor-Critic）。SAC 非常适应真实世界中的机器人技能学习，可以在几个小时内学会解决真实世界的机器人问题，而且它的一套超参数能够在多种不同的环境中工作，效率十分之高。SAC 的开发基于最大熵强化学习这个框架。此框架尝试让预期回报最大化，同时让策略的熵最大化。一般而言，熵更高的策略具有更高的

随机性。从直觉上看，这意味着，最大熵强化学习能取得高回报策略中具有最高随机性的那个策略。SAC 学习一个随机策略，这个策略会把状态映射到动作，也映射到一个能够估计当前策略目标价值的 Q 函数，这个 Q 函数还能通过逼近动态编程来优化它们。SAC 通过这样的方式，来让经过熵强化的回报最大化。此过程中，目标会被看作一个绝对真的方法，来导出更好的强化学习算法，它们有足够高的样本效率，且表现稳定，完全可以应用到真实世界的机器人学习中去。

最新的机器人产品包括：2019 年 6 月，亚马逊在 MARS 人工智能大会上发布了仓库机器人 Pegasus（一种新型包裹分拣机器人）、大型模组化运输机器人 Xanthus。2020 年，斯坦福大学的研究人员通过借鉴传统机器人技术的特性，开发了一种新型的软机器人，它既安全，又保留了可以移动和改变形状的能力。康奈尔大学研究人员创造的首个含半导体元件的行走微型四脚机器人，该机器人尺寸小于 0.1mm（约为人头发宽度），可通过激光控制其腿部行走。意大利理工学院 IIT 实验室在 Science Robotics《科学机器人》杂志中展示了一款仿生手 Hannes，不仅外观与人手非常相似，而且可使截肢患者恢复 90% 以上功能，其设计理念也被授予 Compasso d'Oro 国际工业设计奖。

5.7.3 机器人研究趋势

目前，将人工智能技术应用到机器人的研究已经取得了显著的成绩。在现代社会，智能机器人发挥的作用越来越重要，智能机器人已成为衡量一国工业化水平的重要标志。

借助 AMiner 平台，从机器人领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-41 所示。

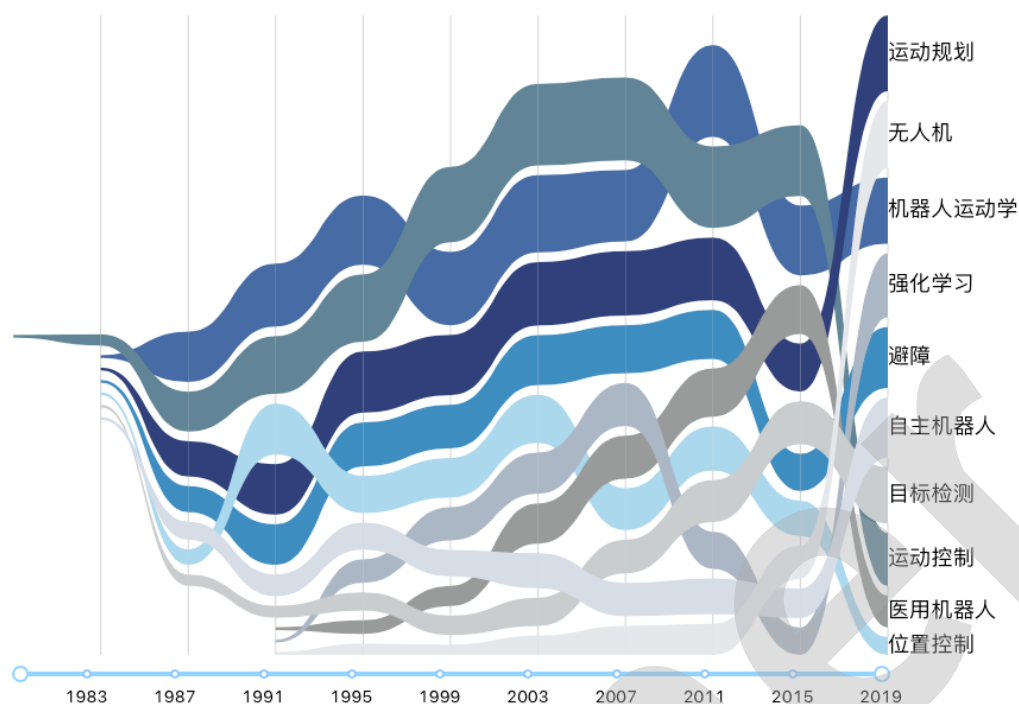


图 5-41 机器人领域技术研究发展趋势

机器人领域研究热度较高的 10 个技术话题，包括：运动规划（Motion Planning）、无人机（Unmanned Aerial Vehicle）、机器人运动学（Robot Kinematics）、强化学习（Reinforcement Learning）、避障（Obstacle Avoidance）、自主机器人（Autonomous Robot）、目标检测（Object Detection）、动作控制（Motion Control）、位置控制（Position Control）、医用机器人（Medical Robotics），如上图所示。近五年，运动规划、无人机和强化学习等技术话题近年来研究热度显著增长，机器人运动学的研究热度保持较为稳定的高水平。

5.8 数据挖掘

5.8.1 数据挖掘概念

数据挖掘（Data Mining），是指从大量的数据中自动搜索隐藏于其中的有着特殊关系性的数据和信息，并将其转化为计算机可处理的结构化表示，是知识发现的一个关键步骤。数据挖掘是一门综合的技术，涉及统计学、数据库技术和人工智能技术的综合，价值在于用数据挖掘技术改善预测模型。

数据挖掘的对象可以是任何类型的数据源，包括数据库数据、数据仓库、事物数据，以及文本、多媒体数据、空间数据、时序数据、web 数据、数据流、图或网络数据等。

数据挖掘涉及的常见任务有：（1）数据表征：是对目标类数据的一般特征或特征的总结。（2）异常检测：数据库可能包含的不符合数据一般行为或模型的数据对象，这些数据对象即被成为异常值。（3）关联规则学习：搜索变量之间的关系。（4）聚类：是在未知数据的结构下，发现数据的类别与结构，通常使用基于最大化类内相似性和最小化类间相似性的原则对对象进行聚类或分组。（5）分类：是查找描述和区分数据类别或概念的模型（或函数）的过程，目的是为了能够使用模型来预测类别标签未知的对象的类别。（6）回归：试图找到能够以最小误差对该数据建模的函数。（7）数据演化分析：描述并建模其行为随时间变化的对象的规则或趋势。分析特征包括时间序列数据分析、序列或周期性模式匹配以及基于相似性的数据分析。

5.8.2 数据挖掘过去十年主要进展

数据挖掘在 20 世纪的发展经历了数据搜集阶段（20 世纪 60 年代）、数据存储阶段（70 年代）、数据分析访问阶段（80 年代）、数据仓库决策与支持阶段（90 年代）。

21 世纪以来，信息化的发展非常快，数据也在不断更新。对这些大数据进行

处理，并从中提取出有价值的信息，相应的数据挖掘技术越来越成熟，已经被广泛地应用于各类实际问题，包括金融数据分析、推荐系统等。数据挖掘技术结合了数据库、人工智能、模式识别、神经网络、机器学习、统计、高性能计算、数据可视化、空间数据分析和信息检索等多方面知识。

过去十年来，迎来了大数据时代，每天都在产生数以亿计的数据。自动地从大规模的数据中挖掘出有价值、有意义的知识或者模式，变得越来越迫切。数据挖掘相关研究结合实际问题，注重与机器学习、统计学科等的交叉，从大数据中挖掘出有价值的信息。在此期间，数据挖掘领域研究主要归为两大类：复杂数据挖掘与分布式数据挖掘。

复杂数据包括序列数据、图数据等。在序列数据挖掘中，基于注意力（Attention）机制的 Transformer 模型表现出了巨大的潜力，在机器翻译等任务上取得了非常好的效果。随后，BERT 模型使用双向 Transformer 通过预训练方式在各种自然语言处理的任务上都达到了当时最好的结果。在图数据挖掘研究中，网络表示学习仍然是近年来非常热门的话题。从 DeepWalk 算法开始，基于随机游走的算法在无监督的表示学习任务中表现良好。NetMF 算法将几种基于随机游走的算法统一写成了矩阵分解的形式，给网络表示学习算法提供了理论基础。图卷积神经网络是另一种处理图数据的有效方法，借鉴了图谱论中的图卷积并使用图的拉普拉斯矩阵，在半监督的节点分类任务和图分类任务中都表现出很好的效果。除此之外，异构网络的表示与挖掘也逐渐被大家所关注。

分布式数据挖掘已成为数据挖掘领域非常有前途的方向。随着数据挖掘计算成本的增加和数据隐私保护的问题，分布式数据挖掘开始备受关注。分布式数据挖掘遵循“全局分布、局部集中”的挖掘原则，利用分布式计算方式对分布式资源进行挖掘，通过整合局部知识来获得全局知识，以此降低计算成本并增强数据保密性。由于分布式数据挖掘采用了不同的计算方式，传统的数据挖掘技术很难直接应用于分布式数据挖掘。目前，数据安全与数据隐私开始被大家所关注。2018年5月，通用数据保护条例（GDPR）在欧盟正式生效，这也使得基于隐私保护的分布式数据挖掘方法逐渐被研究者所重视。

5.8.3 数据挖掘研究趋势

数据挖掘是人工智能领域研究的热点问题，致力于在大规模、不完全、有噪声、模糊随机的数据集中提取隐含其中、潜在有用的未知知识。

借助 AMiner 平台，从数据挖掘领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-42 所示。

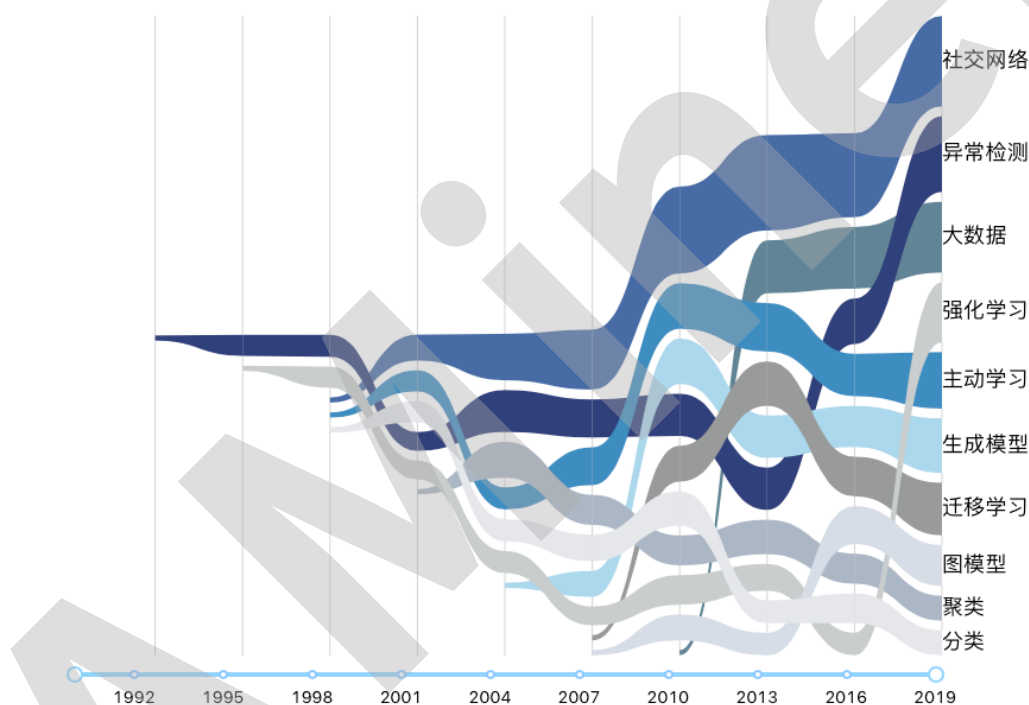


图 5-42 数据挖掘领域技术研究发展趋势

数据挖掘领域研究热度较高的 10 个技术话题，包括：社交网络（Social Network）、异常检测（Anomaly Detection）、大数据（Big Data）、强化学习（Reinforcement Learning）、生成模型（Generative Model）、迁移学习（Transfer Learning）、图模型（Graphical Model）、聚类（Clustering）、分类（Classification），如图 5-42 所示。近十年内，社交网络 and 大数据等技术话题的研究热度一直保持在较高水平。这与移动互联网的发展，以及社交网络平台的不断涌现有关。近五年内，异常检测和强化学习相关话题的研究热度也明显上升。

5.9 人机交互

5.9.1 人机交互概念

人机交互（Human-Computer Interaction, HCI），也称人机互动，是人与计算机之间为完成某项任务所进行的信息交换过程，是一门研究系统与用户之间的交互关系的学问。系统可以是各种各样的机器，也可以是计算机化的系统和软件。人机交互界面通常是指用户的可见部分，用户通过人机交互界面与系统交流，并进行操作。人机交互技术是计算机用户界面设计中的重要内容之一，与认知学、人机工程学、心理学等学科领域有密切的联系。

人机交互是现代信息技术、人工智能技术研究的热门方向。

目前关于人机交互的定义主要有三种：一是 ACM（Association for Computing Machinery）的定义：有关交互计算机系统设计、评估、实现以及与之相关现象的学科；二是伯明翰大学教授 Alan Dix 的定义：人机交互是研究人、计算机以及他们之间相互作用方式的学科，学习人机交互的目的是使计算机技术更好地为人类服务；三是宾夕法尼亚州立大学 John M. Carroll 的定义：人机交互指的是有关可用性的学习和实践，是关于理解和构建用户乐于使用的软件和技术，并能在使用时发现产品有效性的学科。

5.9.2 人机交互过去十年主要进展

人机交互技术的发展是一个从人适应计算机到计算机不断地适应人的发展历程。交互的信息由精确的输入输出信息逐渐变成非精确的输入输出信息。交互方式也从早期的由设计者进行手工作业，逐渐转变为由计算机程序员可采用批处理作业语言或交互命令语言的方式和计算机打交道，到由普通用户使用图形用户界面（GUI）进行交互以及网络用户界面。

以鼠标发明为标志的图形用户界面（Graphical User Interface, GUI）的产生，一改规范命令与计算机交互的命令行界面模式（Command Line Interface, CLI），GUI 提供了普通人与计算机便捷交互的工具和方法，让计算机从实验室走进办公室、走入家庭，十多年前，触屏技术成为产品技术，GUI 中的鼠标被人的天然指

点（pointing）工具——手指所取代，计算机又变身出手机，成为更多人方便使用的随身掌上工具。

随着网络的普及性发展和无线通讯技术的发展，人机交互领域面临着巨大的挑战和机遇，传统的图形界面交互已经产生了本质的变化，人们的需求不再局限于界面美学形式的创新，用户更多的希望在使用多媒体终端时，有着更便捷、更符合他们使用习惯，同时又有着比较美观的操作界面。利用人的多种感觉通道和动作通道（如语音、手写、姿势、视线、表情等输入），以并行、非精确的方式与（可见或不可见的）计算机环境进行交互，使人们从传统的交互方式的束缚解脱出来，使人们进入自然和谐的人机交互时期。这一时期的主要研究内容包括：多通道交互、情感计算、自然语言理解、虚拟现实、智能用户界面等方面。

多通道交互（Multi Modal Interaction, MMI）是指使用多种通道与计算机通信的人机交互方式。它既适应了“以人为中心”的自然交互准则，也推动了互联网时代信息产业（包括移动计算、移动通信、网络服务器等）的快速发展。通道（modality）涵盖了用户表达意图、执行动作或感知反馈信息的各种通信方法，如言语、眼神、脸部表情、唇动、手动、手势、头动、肢体姿势、触觉、嗅觉或味觉等。采用这种方式的计算机用户界面称为“多通道用户界面”。目前，人类最长使用的多通道交互技术包括手写识别、笔式交互、语音识别、语音合成、数字墨水、视线跟踪技术、触觉通道的力反馈装置、生物特征识别技术和人脸表情识别技术等方面。

情感人机交互、让计算机具有情感能力首先是由美国 MIT 大学 Minsky 教授（人工智能创始人之一）提出的。他在 1985 年的专著“*The Society of Mind*”中指出，问题不在于智能机器能否有任何情感，而在于机器实现智能时怎么能够没有情感。在赋予计算机情感能力并让计算机能够理解和表达情感的研究领域，美国 MIT 媒体实验室 Picard 教授团队于 1997 年在专著“*Affective Computing*（情感计算）”中首次提出情感计算一词，即是关于情感、情感产生以及影响情感方面的计算。MIT 对情感计算进行全方位研究，正在开发研究情感机器人，最终有可能人机融合。IBM 公司的“蓝眼计划”，可使计算机知道人想干什么。CMU 主要研究可穿戴计算机。

虚拟现实（Virtual Reality, VR）是以计算机技术为核心，结合相关科学技术，生成与一定范围真实环境在视、听、触感等方面高度近似的数字化环境，用户借助必要的装备与数字化环境中的对象进行交互作用、相互影响，可以产生亲临对应真实环境的感受和体验。虚拟现实是人类在探索自然、认识自然过程中创造产生，逐步形成的一种用于认识自然、模拟自然，进而更好地适应和利用自然的科学方法和科学技术。虚拟现实技术广泛应用于军事演练、虚拟医学手术训练、远程会诊、手术规划及导航、工业产品论证设计、装配、数字博物馆/科学馆、大型活动开闭幕式彩排仿真、沉浸式互动游戏等方面。

智能用户界面（Intelligent User Interface, IUI）是致力于改善人机交互的高效率、有效性和自然性的人机界面。它通过表达、推理，按照用户模型、领域模型、任务模型、谈话模型和媒体模型来实现人机交互。智能用户界面主要使用人工智能技术去实现人机通信，提高了人机交互的可用性：如知识表示技术支持基于模型的用户界面生成，规划识别和生成支持用户界面的对话管理，而语言、手势和图像理解支持多通道输入的分析，用户建模则实现了对自适应交互的支持等。当然，智能用户界面也离不开认知心理学、人机工程学的支持。

智能体代理（Agents）是一个能够感知外界环境并具有自主行为能力的以实现其设计目标的自治系统。智能的 Agent 系统可以根据用户的喜好和需要配置具有个性化特点的应用程序。基于此技术，可以实现自适应用户系统、用户建模和自适应脑界面。

过去十年，是人机交互向自然交互蓬勃发展的十年。已经有越来越多的搜索引擎宣布支持自然语言搜索特性，自然语言人机交互界面在智能短信服务、情报检索、人机对话等方面也具有广阔的发展前景和极高的应用价值，并有一些阶段性成果出现在商业运用中。

更少依赖操控工具，发展学习和使用成本更小的自然交互技术，一直是人机交互研究的价值追求。随着感知和计算技术的进步，自然交互技术创新层出不穷，并能迅速成为新型产品技术。这些人机交互突破技术可分为 3 大类：支持自然动作的感知技术、面向穿戴的新型终端和基于语音识别的对话交互。

人体动作蕴含丰富的语义，动作交互技术一方面需要感知技术的进步，另一

方面需要发现或设计有明确交互语义的动作（gesture，姿态、手势）。如今，二维表面上，多指触摸动作在触屏上已普遍可用，三维空间中，嵌入了深度摄像头的手持和固定设备，能比较准确识别人的姿态和动作，做出响应。不同于人脸识别等目标明确的视觉识别任务，动作交互不仅要求视觉识别的准确度，更需要研究基于交互任务的动作表达的自然性与一致性，所以，除了动作语义很直白的动作游戏（body game），三维动作交互尚缺少普遍认知和接受的交互动作语义。手势的不可见性，是动作交互的主要难题。

穿戴（wearable）设备中，手环设备基本只有健康和活动检测功能，智能手表可以算做创新终端，但作为缩小版的手机，由于交互界面的缩小和操作方式的限制（通常是小界面上双手参与操作），其承载功能也较手机缩减很多。VR / AR（虚拟现实 / 增强现实）的一个理想载体是头戴式设备，最近几年，多款智能眼镜产品面世，较之前笨重的头盔轻便了许多，逼真的虚拟场景和准确的现实对象识别信息都可以清晰呈现在眼前，并在特定领域开拓着增强体验的应用；然而，智能眼镜尚缺少与其三维真实显示匹配的准确的自然输入技术，以及从眼手绑定在手机上转变到眼手分离的眼镜设备上时，尚未建立起相应的交互模式。

自然语言对话式交互得益于大数据和智能技术的进步，多语言的自然语音识别技术在用户终端上都达到了很高的可用水平，并且，语音识别超越文本输入方式，成为智能软件助理的使能技术，近两年，更是有基于语音接口的家居产品如雨后天晴般出现，VUI（Voice User Interface，语音用户界面）已经成为交互术语。然而，VUI 的局限也是显而易见的，相对并行模式的视觉通道，串行模式的语音通道的带宽显然窄的多，出声的使用方式在很多场合是不合适的，但作为一种可用的自然交互技术，有效提升了用户体验。

5.9.3 人机交互研究趋势

人机交互是计算机系统的重要组成部分，是影响人工智能产品易用性和友善性的重要条件。现阶段，智能人机交互成为人机交互的研究热点问题。

借助 AMiner 平台，从人机交互领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色

带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-43 所示。

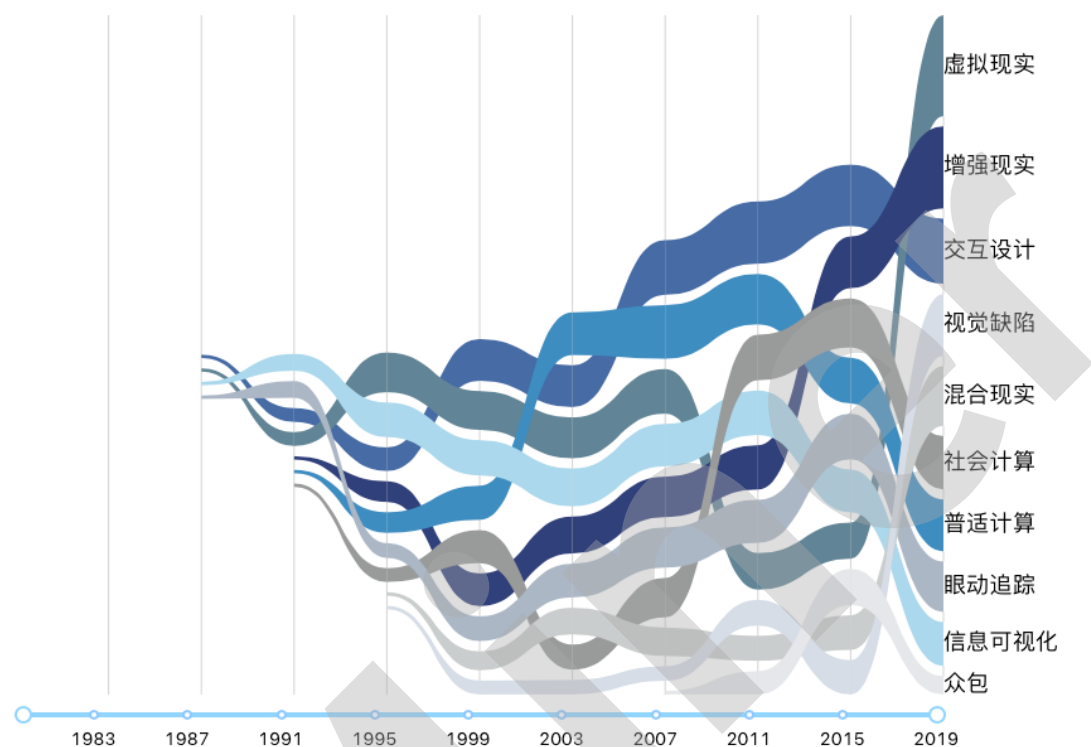


图 5-43 人机交互领域技术研究发展趋势

人机交互领域研究热度较高的 10 个技术话题，包括：虚拟现实（Virtual Reality）、增强现实（Augmented Reality）、交互设计（Interaction Design）、视觉缺陷（Visual Impairment）、混合现实（Mixed Reality）、社会计算（Social Computing）、普适计算（Ubiquitous Computing）、眼动追踪（Eye Tracking）、信息可视化（Information Visualization）、众包（Crowdsourcing），如图 5-43 所示。近几年，虚拟现实、增强现实等研究发展迅猛，科学家们对这两种技术将要在各行各业所产生的作用给予极高的预期值。

5.10 可视化

5.10.1 可视化技术概念

可视化技术是把各种不同类型的数据转化为可视的表示形式,并获得对数据更深层次认识的过程。可视化将复杂的信息以图像的形式呈现出来,让这些信息更容易、快速地被理解,因此,它也是一种放大人类感知的图形化表示方法。

可视化技术充分利用计算机图形学、图像处理、用户界面、人机交互等技术,以人们惯于接受的表格、图形、图像等形式,并辅以信息处理技术(例如:数据挖掘、机器学习等)将复杂的客观事物进行图形化展现,使其便于人们的记忆和理解。可视化为人类与计算机这两个信息处理系统之间提供了一个接口,对于信息的处理和表达方式有其独有的优势,其特点可总结为可视性、交互性和多维性。

目前,数据可视化针对不同的数据类型及研究方向,可以进一步划分为科学数据可视化、信息可视化,以及可视分析学三个子领域。这三个领域既紧密相关又分别专注于不同类型的数据及可视化问题。具体而言,科学可视化是针对科学数据的可视化展现技术。科学数据,例如,医疗过程中由CT扫描生成的影像数据、风洞实验而产生的流体数据、以及分子的化学结构等,是对物理世界的客观描述,往往是通过科学仪器而测量得到的数据。这类数据的可视化主要关注于如何以清晰直观的方式展现数据所刻画的真实物理状态。因此,科学可视化往往呈现的是三维场景下的时空信息。信息可视化注重于如何以图形的方式直观展现抽象数据,它涉及到了对人类图形认知系统的研究。在这里,抽象数据(例如:图形数据、多维度数据、文本数据等)往往是对各应用领域所产生数据的高层次概括,记录的是抽象化的信息。针对这样的数据,信息可视化着眼于多维度信息的可视编码技术,即如何以低维度(2D)的图形符号来直观展现并揭示抽象数据中所隐藏的潜在规律与模式;可视分析学是多领域技术结合的产物,旨在结合并利用信息可视化、人机交互、以及数据挖掘领域的相关技术,将人的判断与反馈作为数据分析中重要的一环,从而达到精准数据分析、推理及判断的目的。

可视化技术的重要性在于,通过提供对数据和知识的展现,建立用户与数据系统交互的良好沟通渠道,利用人类对图形信息与生俱来的模式识别能力,通过

以直观的图像化方式展现数据，从而帮助用户快速发觉数据中的潜在规律，并借助分析人员的领域知识与经验，对模式进行精准分析、判断、推理，从而达到辅助决策的目的。

可视化和可视分析在欧美等国家较为普遍，大多数高校都有相关学科，在我国起步较晚，目前只有清华、北大、浙大等十几所高校有相关研究生专业。

目前可视化技术在各行各业中均得到了广泛的应用。其中，可视化技术在信息安全、智慧医疗、电子商务、机器学习、智慧城市、文化体育、数字新闻、气象预报、地质勘测、军事、反恐等诸多领域发挥着越来越重要的应用。

5.10.2 可视化技术过去十年主要进展

可视化技术最早可追溯到 19 世纪，统计学理论的建立促进并加速了可视化领域发展。后来经过一段时间停滞，20 世纪后期随着计算机的出现，人类处理数据的能力有了跨越式的提升，在现代统计学与计算机计算能力的共同推动下，可视化迎来了复苏与更快速的发展。20 世纪 70 年代到 80 年代，人们主要尝试使用多维定量数据的静态图来表现静态数据，80 年代中期出现了动态统计图，最终在上世纪末两种方式开始合并，致力于实现动态、可交互的数据可视化，动态交互式的数据可视化方式成为新的发展主题。这一时期，数据可视化的潜力来自动态图形方法的发展，允许对图形对象和相关统计特性进行即时和直接的操纵。同时，出现交互系统，通过调整控制来选择参考分布的形状参数和功率变换。这可以看作动态交互式可视化发展的起源，并推动了数据可视化的发展。80 年代末，可视化技术提出并最早应用于计算机科学领域，形成了可视化技术的一个重要分支——科学计算可视化。

过去十年是数据驱动的时代，基于大数据的可视化蓬勃发展。步入 21 世纪互联网数据量猛增，人们逐渐开始对大数据的处理进行了重点关注。之后全球每天的新增数据量就已经开始以指数倍膨胀，用户对于数据的使用效率也在日益提升，数据的服务商开始需要从多个维度向用户提供服务，人们对数据可视化技术的依赖程度也不断加深。大数据时代的到来对数据可视化的发展有着冲击性的影响，继续以传统展现形式来表达庞大的数据量中的信息是不可能的，大规模的动

态化数据要依靠更有效的处理算法和表达形式才能够传达出有价值的信息，因此大数据可视化的研究成为新的时代命题。在应对大数据时，不但要考虑快速增加的数据量，还需要考虑到数据类型变化，这种数据扩展性的问题需要更深入的研究才能解决；互联网的加入增加了数据更新的频率和获取的渠道，而实时数据的巨大价值只有通过有效的可视化处理才可以体现，于是，动态交互的技术向交互式实时数据可视化发展。2010年后数据可视化工具基本以表格、图形(chart)、地图等可视化元素为主，数据可进行过滤、联动、跳转、高亮等分析手段做动态分析。如何建立一种有效的、可交互式的大数据可视化方案来表达大规模、不同类型的实时数据，成为数据可视化学科的主要研究方向。

随着大数据时代的到来，可视化已经成为一个必不可少的工具。现有的可视化软件及工具，可用于设计可视化应用程序和构建可视化分析系统，有助于可视化的广泛使用。为了减少制作可视化的技术负担，一些可视化工具提供了声明性语法，其中包括了 Vega-Lite 和 P5。声明性语法可以将可视化设计与执行细节分离，这使分析人员可以专注于特定于应用程序的设计决策。Vega-Lite^[35]是 2017 年 InfoVis 的最佳论文，它基于底层可视化语法 Vega 进行上层封装，提出了一套能够快速构建交互式可视化的高阶语法。同时 Vega-Lite 和 P5 都提供了易于使用的编程接口。Vega-Lite 是一套能够快速构建交互式可视化的高阶语法，它是基于 Vega 和 D3 等底层可视化语法的上层封装。相比于其他比较底层可视化语法，Vega-Lite 可以通过几行 JSON 配置代码即可完成一些通用的图表创建，而相反地，想要用 D3 等去构建一个基础的统计图表则可能需要编写多行代码，如果涉及到交互的话代码量更是会大大增加。P5（便携渐进式并行处理流水线，Portable Progressive Parallel Processing Pipelines）是一个基于 Web 的可视化工具包，它集成了 GPU 计算与渐进式处理，并且提供了带有声明性语法的 API，可用于指定渐进式数据转换和可视化操作，从而帮助分析人员构建融合了高性能计算和渐进式分析工作流的可视化系统。

故事叙述是可视化研究的一个重要且新兴的方向。与传统的、强调数据分析的可视化思路不同，故事叙述强调数据的传达与沟通，强调数据与人（且通常是

³⁵ Satyanarayan A, Moritz D, Wongsuphasawat K, et al. Vega-lite: A grammar of interactive graphics[J]. IEEE transactions on visualization and computer graphics, 2016, 23(1): 341-350.

普通人)的连结。在此思路下,研究者们致力于探究:何种数据呈现与讲述技巧,可以使数据具备吸引力、记忆度;数据故事的创作流程是怎样的,存在哪些需求和痛点;以及如何自动生成数据故事等等。用讲故事的方式来呈现可视化,本质上是体现了一种人本导向,即以人的需求出发,提取和分析数据,并以对人友好的方式,将数据中的信息传达出去。随着我们的社会越来越依赖数据赋能,更好地构建数据与人的关系,将成为一条必经之路。叙述可视化的应用,不仅在于那些以“叙述”为主业的领域,如新闻媒体、广告宣传,更在于需要用数据来影响人、说服人、打动人的各行各业。对于研究者来说,相关的研究方向则包括可视化设计、人机交互、认知与感知、智能生成与推荐等。

数据可视化领域中大多数的可视化生成系统往往是基于数据的交互式探索,也包括商业领域的知名的可视化工具 Tableau 和 PowerBI。近些年来,为了避免繁杂的数据分析步骤并提升用户效率,可视化的自动生成逐渐成为行业领域中的研究热点。一系列基于规则和机器学习的推荐方法层出不穷,在自动生成可视化的最新研究中,研究者希望在保证准确表现数据的同时,也能将视觉设计的因素考虑在内,确保可视化的美观性和数据的表现力。例如, DataShot 和 Text-to-Viz, 分别从数据和自然语言两个角度去自动生成富有设计感的数据可视化,前者直接从表格数据生成信息简报,后者根据用户的自然语言输入生成对应的信息图。制作一个有效且美观的数据可视化往往需要跨专业领域的技能,尤其是需要同时具备数据分析能力和平面设计能力,而这对于一个没有专业训练的普通用户来说是比较困难的。DataShot 和 Text-to-Viz 等前沿的技术研究均通过自动化的方法从数据洞察和设计美学两个方面帮助用户生成可视化,降低用户制作可视化的门槛,并有效提高生产效率。

LSTMVis^[36]是一个递归神经网络的可视化分析工具,它着重于对 RNNs 中的隐藏特征进行可视化分析。LSTMVis 结合了一个基于时间序列的选择界面和一个交互式的匹配工具来搜索大型数据集中相似的隐藏状态模式。系统的主要功能是理解模型中动态变化的隐藏状态。该系统允许用户选择一个假设的输入范围

³⁶ Strobel, H., Gehrmann, S., Pfister, H., & Rush, A. . (2017). LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks., IEEE Transactions on Visualization and Computer Graphics, no. 1: 667-676

来关注局部的改变，将这些状态改变与大型数据集中类似的模式进行匹配，并将这些选择出来的模式进行对齐分析。RNNs 在序列建模方面有着重要的作用，但是模型中的隐藏层含义很难被解释清楚。对于一个完成训练的 RNN 模型，分析人员并不清楚这个模型是如何理解序列中不同节点之间的关系的。LSTMVis 能够帮助用户交互式地探索 RNN 模型复杂的网络结构，并将模型中抽象表示的隐藏层信息与人类可理解的原始输入进行关联。

过去十年，大数据可视化广泛应用于各个领域。例如，社交媒体可视化、信息扩散可视化，社会竞争与合作可视化，人员流动性可视化、体育比赛实况与趋势可视化，以及主要开发用来展现用户的个人健康信息、汇总并展现公众健康信息、分析并展现临床电子病历记录中的规律与模式的医疗数据可视化等等。

5.10.3 可视化技术研究趋势

可视化是利用计算机图形学和图像处理技术，将数据转换成图形或图像在屏幕上显示出来，再进行交互处理的理论、方法和技术。可视化发展的主要研究趋势包括：海量、异构、时变、多维数据的可视化展示方案；可视化在可解释性深度学习领域的应用；自动可视化生成技术的研究；基于形式概念分析理论的知识可视化方法；可视化模式识别；整体可视与局部详细可视相结合的新方法研究等。同时，可视化发展面临着以下的挑战：如何进一步深入挖掘人类对于图形、动画、以及交互的感知及认知模式，从而进一步完善可视化的相关理论；如何打破“手工作坊”式的针对每一个问题，单独定制数据可视化设计方案的传统模式，大规模批量创造生成风格化的可视展现；以及，如何根据用户的数据分析任务与需求自动推荐合适的可视化展现方式。

借助 AMiner 平台，从可视化领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-44 所示。

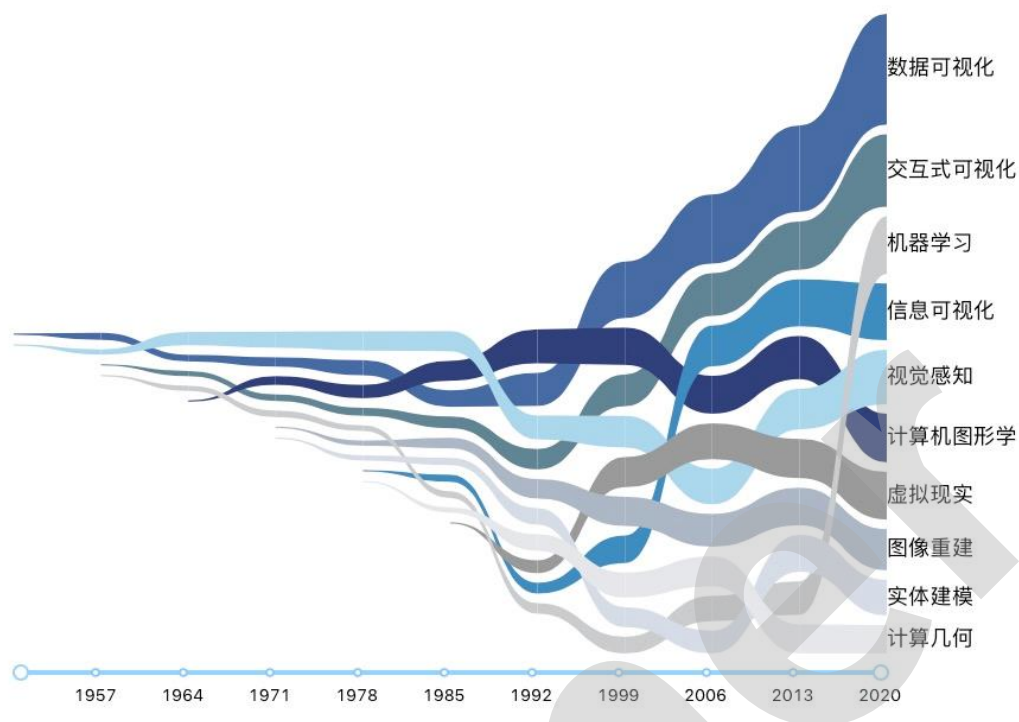


图 5-44 可视化领域技术研究发展趋势

可视化领域研究热度较高的 10 个技术话题，包括：计算机图形学（Computer Graphics）、计算几何（Computational Geometry）、实体建模（Solid Modeling）、图像重建（Image Reconstruction）、机器学习（Machine Learning）、深度学习（Deep Learning）、视觉感知（Visual Perception）、信息可视化（Information Visualization）、交互式可视化（Interactive Visualization）、数据可视化（Data Visualization），如图 5-44 所示。从中可以看出数据可视化、交互式可视化等技术话题的研究热度呈上升趋势且始终保持较高的研究热度。机器学习从 2013 年以来在可视化领域得到飞速发展，这一现象与机器学习中直方图、密度图、箱线图、相关系数矩阵图、散点矩阵图等可以很好地可视化数据和信息不无关系。

5.11 其他人工智能外延技术

5.11.1 经典 AI

经典 AI 领域主要是根据 AAAI Conference on Artificial Intelligence (AAAI) 和 International Joint Conference on Artificial Intelligence (IJCAI) 两大会议议题而设置。这两个会议是人工智能领域的顶级会议，分别在 1980 年和 1969 年成立第一次会议，创办历史悠久，领域覆盖面广，涵盖知识工程、自然语言处理、机器学习、深度学习、数据挖掘、计算机视觉、信息检索与推荐等多种领域。

借助 AMiner 平台，从经典 AI 领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-45 所示。

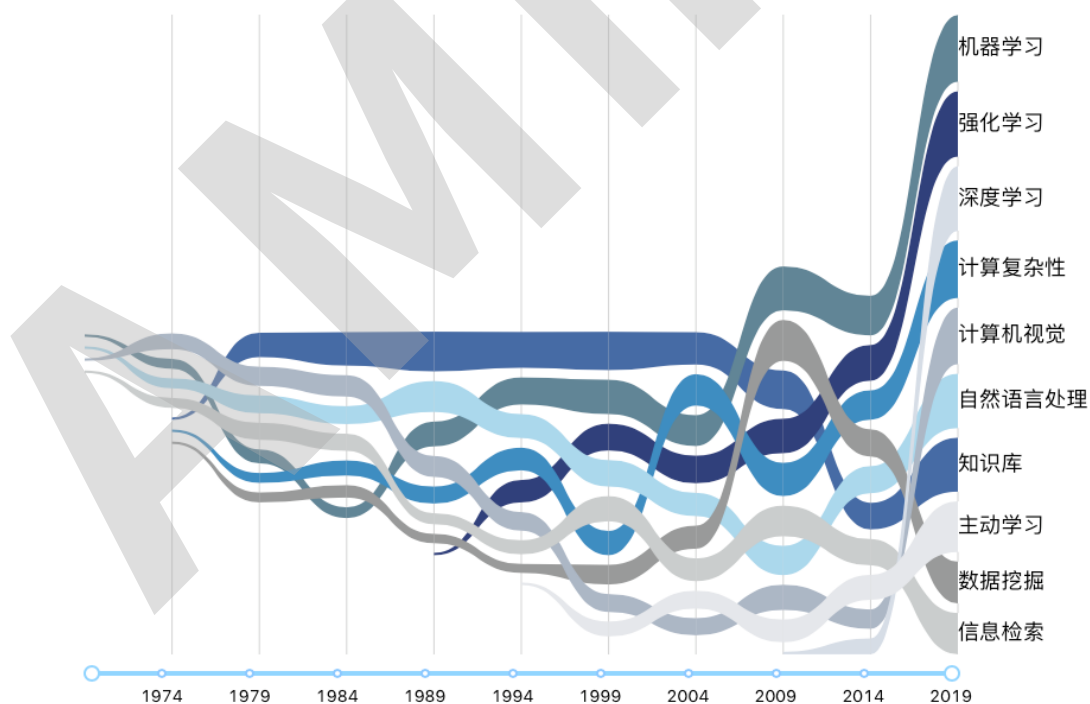


图 5-45 经典 AI 领域技术研究发展趋势

人工智能经典 AI 领域研究热度较高的 10 个技术话题，包括：机器学习 (Machine Learning)、强化学习 (Reinforcement Learning)、深度学习 (Deep

Learning)、知识库(Knowledge Base)、自然语言处理(Natural Language Processing)、计算机视觉(Computer Vision)、主动学习(Active Learning)、数据挖掘(Data Mining)、信息检索(Information Retrieval)、计算复杂性(Computational Complexity), 利用 AMiner 平台技术趋势分析技术, 展示了它们历年的研究热度变化情况, 如图 5-45 所示。

从趋势图中可以看出, 随着时间推进, 这些技术关键词的研究热度逐年增加。尤其是机器学习、强化学习、深度学习的研究热度在近 10 年来快速增长。其中, 深度学习具有学习能力强、适应性好、可移植性好等优点, 在图像识别、语音识别、自然语言处理等领域表现尤为突出, 因此在 10 年间就已经成为经典 AI 领域讨论最多的技术之一。

5.11.2 安全与隐私

人工智能安全包含两个层面的含义: 一是基于人工智能的信息安全, 用人工智能支撑安全, 助力信息安全防御; 二是人工智能自身的安全问题, 人工智能自身存在着安全脆弱性。随着人工智能技术的快速发展, 大数据作为基础性资源, 给社会进步、经济发展带来强大的驱动力。但同时也带来了很大的安全与隐私泄露问题。为了获得大规模、高质量的数据资源, 通过各种手段获得政府、企业或者个人的私密数据, 大数据将网络大众带入到开放透明的裸奔时代。人工智能伦理将是未来智能社会的发展基石, 明确人工智能在安全、隐私、公平等方面的伦理原则, 推进人工智能行业健康发展。

借助 AMiner 平台, 从安全与隐私研究领域相关的论文中抽取关键词, 然后按照时间窗格统计词频, 最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词, 其宽度表示该技术关键词在当年的研究热度, 与当年该技术关键词的论文数量呈正相关, 每一年份中按照其热度由高到低进行排序, 结果如图 5-46 所示。

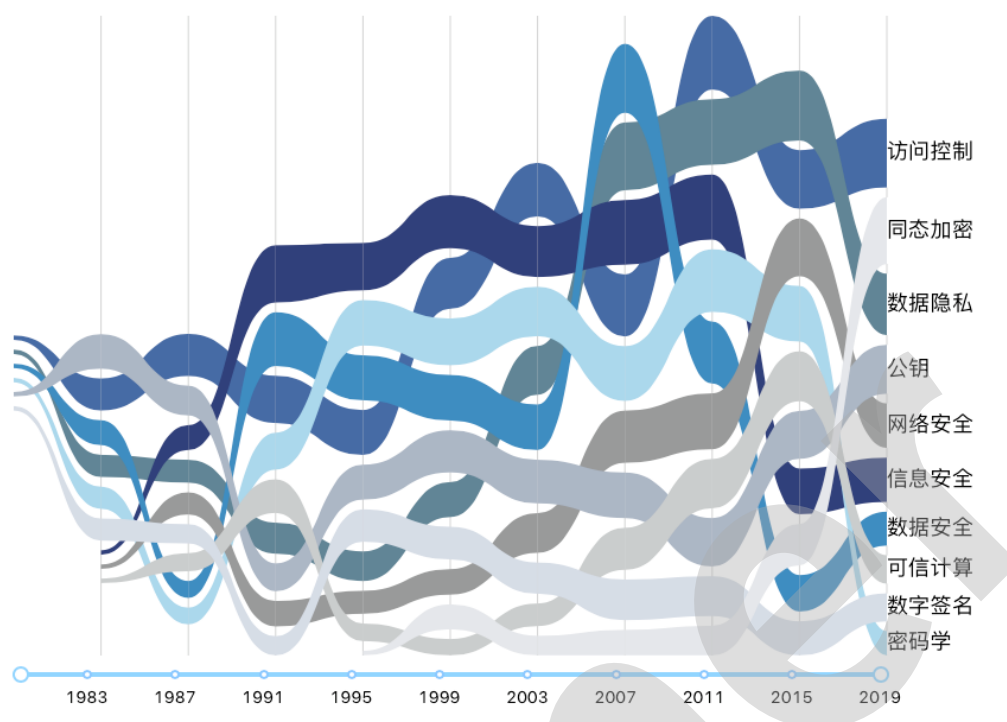


图 5-46 安全与隐私领域技术研究发展趋势

安全与隐私领域研究热度较高的 10 个技术话题，包括：密码学（Cryptography）、访问控制（Access Control）、网络安全（Network Security）、信息安全（Information Security）、可信计算（Trusted Computing）、公钥（Public Key Encryption）、数据安全（Data Security）、同态加密（Homomorphic Encryption）、数字签名（Digital Signature）、数据隐私（Data Privacy），如上图所示。访问控制、同态加密以及数据隐私是近年来研究热度较高的话题，尤其是同态加密近五年的研究热度呈现明显的上升趋势。

5.11.3 芯片技术

人工智能芯片技术也被称为 AI 加速器或计算卡，即为人工智能提供算力资源的专用计算硬件，是能够影响人工智能发展的重要因素。

人工智能芯片按照技术架构分类，包括通用芯片、半定制化芯片、全定制化芯片、类脑芯片。

人工智能的发展与硬件系统密不可分。1985 年起，贝尔实验室开发了一系列神经网络芯片。1989 年，Intel 发布了包含 64 模拟神经元和 10,240 模拟突触的芯片 ETANN。英伟达公司在 2006 年的时候推出了统一计算设备架构（Compute

Unified Device Architecture, CUDA) 以及对应的 G80 平台, 第一次让 GPU 具有可编程性, 让 GPU 的核心流式处理器既具有处理像素、顶点、图形等渲染能力, 又同时具备通用的单精度浮点处理能力。

近十年来, 随着人工智能算法的广泛应用, 对硬件的处理能力和能效也提出了更高的要求。2012 年, 吴恩达发现 GPU 具有加速深度神经网络的能力。此后, GPU 的规模应用推动了以人工神经网络为基础的深度学习技术的发展。2015 年, 微软研究院用 GPU 加速的深度神经网络, 在 ImageNet 比赛中获得了多项击败人的辨识准确度, 这是第一次机器视觉的识别率打败了人眼的识别率(错误率 5%)。2016 年, 谷歌旗下 Deepmind 团队研发的机器人 AlphaGo 以 4 比 1 战胜世界围棋冠军职业九段棋手李世石(AlphaGo 的神经网络训练用了 50 片 GPU, 走棋网络用了 174 片 GPU), 引发了围棋界的轩然大波。

国内学术界和工业界在人工智能芯片行业取得了不菲的成绩, 包括云端全功能人工智能芯片“百度昆仑”、异构融合类脑计算芯片“面向通用人工智能的异构融合天机芯片”、支持多种运算精度的“思元 270 芯片”、具有生态开放、超强算力、高速互联等优势“云燧 T10 芯片”、超低功耗的智能语音 SoC 芯片 TX210、多模态智能计算芯片 TX510、奥比中光研发的国产 3D 视觉感知芯片等。在未来, 智能芯片作为智能创新的物质载体, 应用方向面向高级复杂算法和通用性人工智能平台, 适用于各种具体的行业领域, 具有十分广泛的应用前景。

在科研领域, 清华微电子所魏少军教授团队在可重构深度神经网络计算芯片设计上取得了重要突破^[37], 形成了 Thinker 系列人工智能芯片, 赢得国内外关注。清华微电子所魏少军、尹首一教授团队的论文《面向神经网络处理器的非规则网络结构高效调度技术》(“Efficient Scheduling of Irregular Network Structures on CNN Accelerators”) 获得第 16 届 ACM/IEEE 国际嵌入式系统会议最佳论文奖。

借助 AMiner 平台, 从可视化领域相关的论文中抽取关键词, 然后按照时间窗格统计词频, 最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词, 其宽度表示该技术关键词在当年的研究热度, 与当年该技术关键词的论文数量呈正相关, 每一年份中按照其热度由高到低进行排序,

³⁷ 尹首一, 欧阳鹏, 唐士斌, 等. Thinker: 可重构混合神经网络计算芯片[J]. 人工智能, 2018, 000(002):P.34-45.

结果如图 5-47 所示。

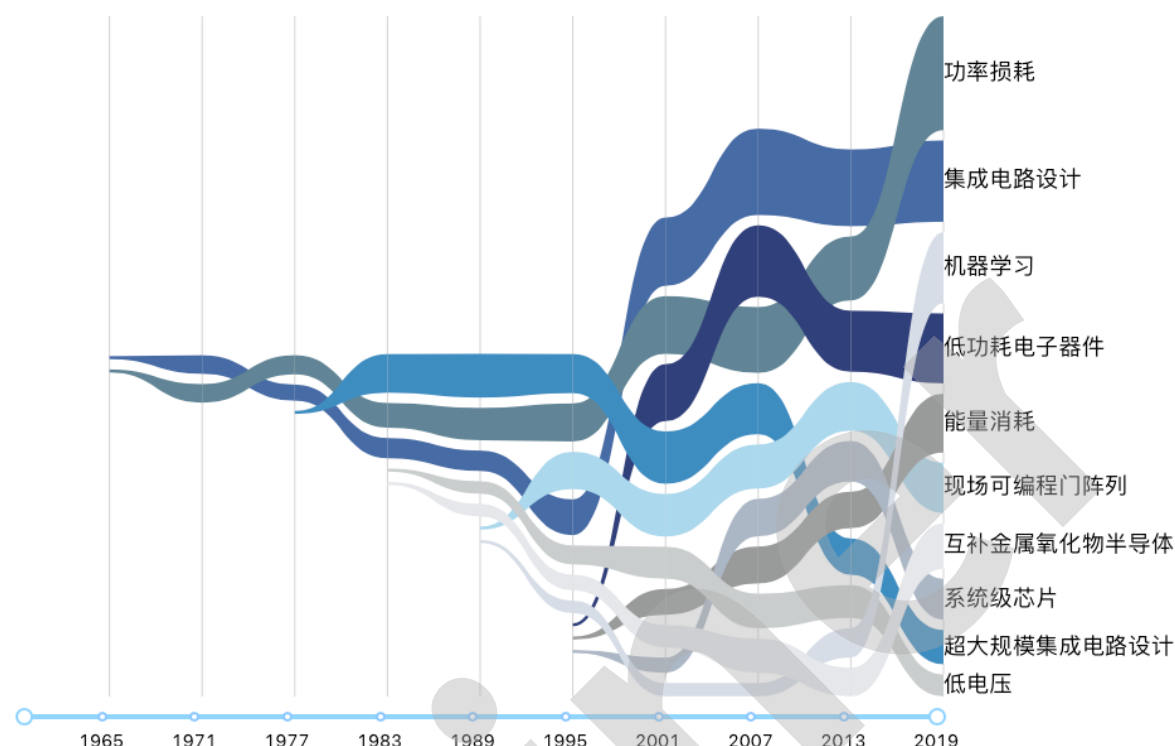


图 5-47 芯片技术领域技术研究发展趋势

芯片技术领域研究热度较高的 10 个技术话题，包括：能量消耗（Power Consumption）、集成电路设计（Integrated Circuit Design）、机器学习（Machine Learning）、低功耗电子器件（Low Power Electronic）、现场可编程逻辑门阵列（Field Programmable Gate Array）、互补金属氧化物半导体（Complementary Metal Oxide Semiconductor）、功率损耗（Power Dissipation）、系统级芯片（System on Chip）、超大规模集成电路设计（Very Large Scale Integration Circuit Design）、低电压（Low Voltage）。从图 5-47 中可以看出，功率损耗和集成电路设计的研究热度一直保持在较高水平；从 2013 年起，机器学习的相关研究也快速升温，成为芯片技术领域研究新热点。

机器学习芯片架构发展的源动力来自于两方面：算法应用和底层器件。机器学习芯片重点关注着深度神经网络和卷积神经网络，近年来的相关研究逐步拓展到对新的智能算法的加速上，如：面向递归神经网络的加速器、面向记忆增强神经网络（MANNs）的加速器以及面向图神经网络的加速器等。此外，与基于传统器件不同，新的器件和工艺极有可能带来机器学习芯片性能和能效的飞跃。典

型的研究包括：基于近似存储的 DNN 推理和基于多芯片组件（MCM）的加速器架构等。

为了帮助理解大脑工作机制并高效解决智能任务，目前已有大量针对脉冲神经网络的神经拟态芯片的工作，其中的典型代表是基于混合模拟信号电路的 Neurogrid 和数字电路的 TrueNorth。2019 年，支持机器学习和类脑计算的异构融合类脑计算芯片“天机芯”作为封面文章发表于 Nature，并在无人驾驶自行车上进行了实验。

未来的一个重要研究方向是高效率的通用智能芯片。高效率的通用智能芯片可以集成多个异构智能处理单元（Intelligent Processing Unit, IPU），分别处理学习、感知、记忆、抽象、推理、联想等不同的智能基本任务。多个 IPU 也可以协同起来完成更复杂的智能任务。

总之，智能体系架构与芯片方向的研究将持续在新兴算法/场景和新型器件/工艺的双重驱动下，在降低编程门槛的同时不断提升对广泛智能任务的处理性能和能效，为探索通用人工智能的机理和应用提供强大算力支撑。

5.11.4 数据库

随着计算机技术与网络通信技术的快速发展，数据库技术已经成为当今信息社会中对大量数据进行组织与管理的重要手段，是网络信息化管理系统的基础。数据库是按一定的结构和规则组织起来的相关数据的集合，是综合各用户数据形成的数据集合，是存放数据的仓库^[38]。数据库是以一定方式储存在一起、能与多个用户共享、具有尽可能小的冗余度、与应用程序彼此独立的数据集合，可视为电子化的文件柜，用户可以对文件中的数据进行新增、查询、更新、删除等操作。

数据库技术的发展可以根据数据模型或结构划分为三个时代：导航数据库时代、SQL/关系时代、后关系时代。

基于导航的数据库管理系统（DBMS）主要发展于 20 世纪 60 年代。各种 DBMS 软件都是基于某种数据模型的。通常也按照数据模型的特点将传统数据

³⁸ 陈黎. 我国数据库的发展现状与趋势[J]. 现代情报, 2006(11):140-142.

库系统分成网状数据库（Network database）、层次数据库（Hierarchical database）和关系数据库（Relational database）三类。

关系数据库的发展可追溯至 20 世纪 70 年代初 Edgar Frank Codd 撰写的关于数据库建设新方法以及实现大型共享数据库的开创性关联数据模型的相关论文。关系模型既可以表达层次结构模型，也可以表示导航模型，也可以表示其原生表格模型，从而根据应用程序的需要对这三种模型进行纯粹的或组合的建模。同时期，IBM 开始基于 Codd 的概念开发原型系统 SystemR。第一个版本于 1974 年 5 月发布，随后的多用户版本在 1978 年和 1979 年由客户进行了测试，并添加了一种标准化的查询语言 SQL（Structured Query Language）。结构化查询语言是高级的非过程化编程语言，允许用户在高层数据结构上工作。它不要求用户指定对数据的存放方法，也不需要用户了解具体的数据存放方式，所以具有完全不同底层结构的不同数据库系统，可以使用相同的结构化查询语言作为数据输入与管理的接口。结构化查询语言语句可以嵌套，这使它具有极大的灵活性和强大的功能。

后关系包含了面向对象型数据库及新型数据库技术的发展。面向对象数据库系统（OODBS）发展于 20 世纪 90 年代。面向对象数据库系统支持定义和操作 OODB，应满足两个标准：首先它是数据库系统，其次它也是面向对象系统。第一个标准即作为数据库系统应具备的能力（持久性、事务管理、并发控制、恢复、查询、版本管理、完整性、安全性）。第二个标准就是要求面向对象数据库充分支持完整的面向对象（OO）概念和控制机制。面向对象数据库简写为：面向对象数据库=面向对象系统+数据库能力。

21 世纪初，NoSQL 与 NewSQL 技术的逐渐发展。NoSQL，泛指非关系型的数据库。NoSQL 数据库的产生是为了解决大规模数据集合多重数据种类带来的挑战，尤其是大数据应用难题。NoSQL 早期就有人提出，发展至 2009 年趋势越发高涨。NoSQL 的拥护者们提倡运用非关系型的数据存储，相对于铺天盖地的关系型数据库运用，这一概念无疑是一种全新的思维的注入。NewSQL 一词是由 451Group 的分析师 Matthew 在研究论文中提出的，是对各种新的可扩展/高性能数据库的简称，这类数据库不仅具有 NoSQL 对海量数据的存储管理能力，还保持了传统数据库支持 ACID 和 SQL 等特性。

图数据库也是非关系型的数据库的一种，是一个使用图结构进行语义查询的数据库，它使用节点、边和属性来表示和存储数据。该系统的关键概念是图，它直接将存储中的数据项，与数据节点和节点间表示关系的边的集合相关联。这些关系允许直接将存储区中的数据链接在一起，并且在许多情况下，可以通过一个操作进行检索。图数据库将数据之间的关系作为优先级。查询图数据库中的关系很快，因为它们永久存储在数据库本身中。可以使用图数据库直观地显示关系，使其对于高度互连的数据非常有用。

NewSQL 属于分布式数据库，分布式数据库是指利用高速计算机网络，将物理上分散的多个数据存储单元，连接起来组成一个逻辑上统一的数据库。分布式数据库的基本思想是，将原来集中式数据库中的数据，分散存储到多个通过网络连接的数据存储节点上，以获取更大的存储容量和更高的并发访问量。

过去十年，新一代数据库系统不仅保持和继承了传统数据库系统的各项功能，支持知识管理、数据管理和对象管理，而且还对其它应用系统开放，在网络上支持标准网络协议，具有良好的可连接性、可移植性、可互操作性和可扩展性。具有此类特征的新型数据库系统包括如分布式数据库系统、知识库系统和主动数据库系统等。

分布式数据库系统由一组分布在网络中的不同计算机上的数据组成。系统中每台服务器有自己的数据库系统及若干台客户机，3 台服务器之间通过网络相连。网络中的每个节点具有独立处理的能力，可以执行局部应用，同时每个节点也能通过网络子系统执行全局应用。用户通过客户机可以对本地服务器中的数据库执行某些应用，也可以对 2 个或 2 个以上节点中的数据库执行某些应用。

知识库系统是数据库和人工智能两种技术相结合的产物。它是在数据库技术中引入人工智能技术，把数据库看作一个人工智能系统，利用人工智能技术来提高 DBMS 的表达、推理和查询能力。其功能主要体现在数据库系统推理能力的扩充、语义知识的引入、知识的获取、知识和数据的有效组织及管理等方面；而效率则体现在数据库对用户查询的快速响应和查询优化上。

主动数据库是相对于传统数据库的被动性而言的，其主要任务是提供对紧急情况及时反应能力，同时提高数据库管理系统的模块化程度。主动数据库通常

采用的方法是在传统数据库系统中嵌入事件、条件、动作规则，在某一事件发生时引发数据库管理系统去检测数据库的当前状态；判断是否满足设定的条件以便触发规定动作的执行。

大数据时代的到来和云计算的发展也带来了数据库领域的革新，数据库管理系统以及数据库相关技术的发展成为促进计算机科学和人工智能发展的重要力量。数据库的智能化可以借鉴 AI 技术来实现数据库管理系统的表达、推理和查询能力；智能化的数据库表现为扩充现有数据库系统的功能，使其具备一定的演绎、推理功能，提高系统的智能化程度。

数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等互相渗透和结合，是当前数据库技术应用的主要特征。

近年来，以机器学习为代表的人工智能技术因其强大的学习和适应能力，在多个领域都大放异彩。同样的，在数据管理领域，传统机器学习和深度学习等技术也有着巨大的潜力和广阔的应用前景。例如，数据库系统所积累的海量历史查询记录可以为基于学习的数据库智能优化技术提供数据支撑。一方面，我们可以构建包含查询、视图或数据库状态的有标签数据，比如，在视图选择问题中，这个标签是指每个候选视图是否被选中。另一方面，在缺乏标签数据的时候，我们可以利用（深度）强化学习技术探索性地（从选择结果的反馈中学习）选择最优的候选视图。此外，人工智能技术让自治数据库的自动决策管理、自动调优和自动组装等需求成为可能。在以深度学习为代表的人工智能技术的加持下，让数据库朝着更加智能的方向发展，数据管理技术也随之智能化。近些年涌现的自治数据库和人工智能原生数据库（如 SageDB, XuanyuanDB），通过融合人工智能技术到数据库系统的各个模块（优化器、执行器和存储引擎等）和数据管理的生命周期，可以大幅度提升数据库各方面的性能，为下一代数据库和人工智能技术的发展指明了一个方向。

在另外一方面，数据管理技术也能以基础设施的身份来支持人工智能的发展。目前的人工智能在落地过程中还面临着一些挑战性。例如，人工智能算法训练效率较低，现有人工智能系统缺少执行优化技术（如大规模缓存、数据分块分区、索引等），不仅会导致大量的计算、存储资源浪费，而且会提高程序异常的发生

率（如内存溢出、进程阻塞等），严重影响单个任务的执行效率。其次，人工智能技术往往依赖高质量的训练数据，现实中的训练数据往往是包含很多缺失值、异常值和别名等类型的错误，这些错误通常会影响训练效率，对模型的质量造成干扰。面向人工智能的数据管理技术可以为解决上述挑战做出贡献。

借助 AMiner 平台，从数据库技术领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-48 所示。

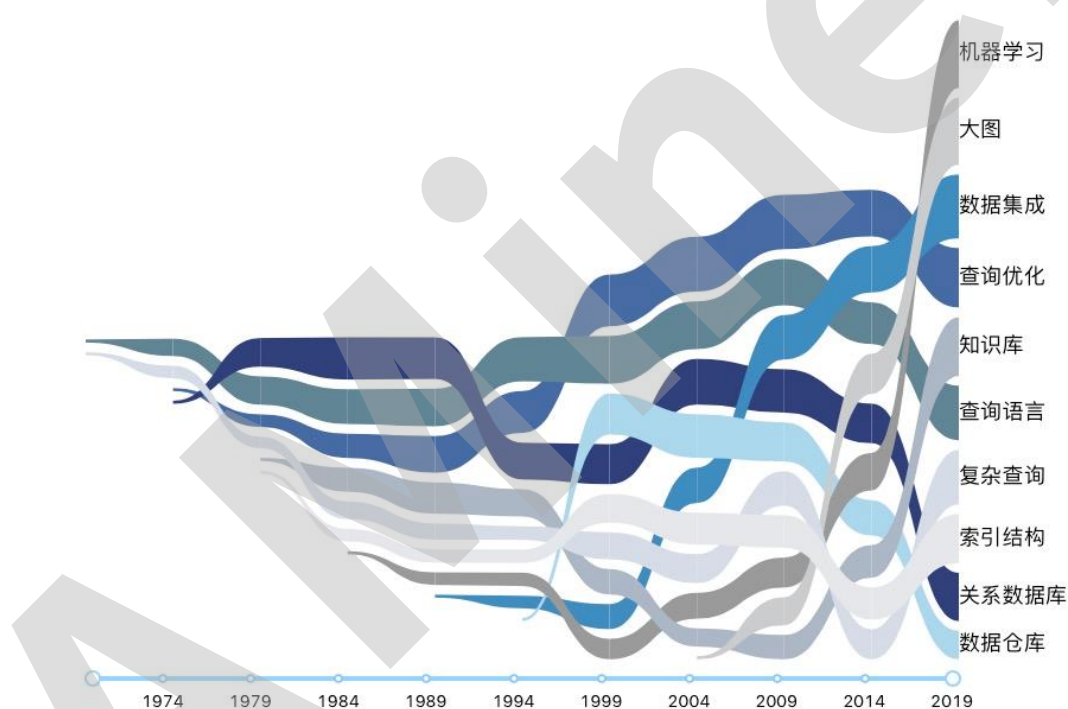


图 5-48 数据库领域技术研究发展趋势

数据库领域研究热度较高的 10 个技术话题，包括：数据集成（Data Integration）、查询优化（Query Optimization）、机器学习（Machine Learning）、查询语言（Query Language）、关系数据库（Relational Database）、知识库（Knowledge Base）、复杂查询（Complex Query）、数据仓库（Data Warehouse）、索引结构（Index Structure）、大图（Large Graph），如上图所示。从中可以看出，机器学习、数据集成和知识库在近年来的研究热度上升很快，而查询优化技术一直保持着较高的热度。

5.11.5 计算机图形

国际标准化组织 ISO 将计算机图形学定义为：计算机图形学是一门研究通过计算机将数据转换成图形，并在专门显示设备上显示的原理方法和技术的学科。它是建立在传统的图形学理论、应用数学及计算机科学基础上的一门边缘学科。这里的图形是指三维图形的处理。简单来讲，它的主要研究内容是研究如何在计算机中表示图形，以及利用计算机进行图形的计算处理和显示的相关原理和算法。

计算机图形学的开创之初，主要解决的问题是在计算机中表示三维结合图形以及如何利用计算机进行图形的生成处理和显示的相关原理和算法，目的是产生令人赏心悦目的真实感图像。随着近些年的发展，计算机图形学研究内容变得非常广泛，包括图形硬件、图形标准、图形交互技术、栅格图形生成算法、曲线曲面造型、实体造型、真实版图形的计算、显示算法、科学计算可视化、计算机动画、虚拟现实、自然景物仿真等。

计算机图形学的总体框架可以包括以下几个部分：数学和算法基础、建模、渲染以及人机交互技术。计算机图形学需要一些基本的数学算法，例如向量和几何的变化、几何建模式的三维空间变化、三维到二维的图形变换等等。建模是进行图形描述和计算，由于在多维空间中有各种组合模型，有一些是解析式表达的简单形体，也有一些隐函数表达的复杂曲线，因此需要进行复杂的建模工作。渲染也叫绘制，指的是模型的视觉实现过程，例如对光照纹理等理论和算法进行处理，其中也需要大量的计算。交互技术可以说是图形学交互的重要工具，是计算机图形学的重要应用。

计算机图形学可追溯到 20 世纪 50 年代。1950 年，美国 MIT 的旋风一号（whirlwind I）计算机配备了世界上第一台显示器——阴极射线管（CRT）来显示一些简单的图形，使得计算机摆脱了纯数值计算的单一用途，能够进行简单的图形显示，但是还不能对图形进行交互操作。50 年代末期，MIT 的林肯实验室在“旋风”计算机上开发 SAGE（Semi-Automatic Ground Environment System）空中防御体系。SAGE 于 1957 年投入试运行，已经能够将雷达信号转换为显示器上的图形并具有简单的人机交互功能，操作者使用光笔点击屏幕上的目标即可获得敌机的飞行信息，这是人类第一次使用光笔在屏幕上选取图形。1959 年，麻

省理工学院林肯实验室第一次使用了具有指挥和控制功能的 CRT，“被动式”图形学开始迈向交互式计算机图形学。

1962 年美国 MIT 林肯实验室的 IvanE.Sutherland 发表了一篇题为“sketchpad: 一个人机交互通信的图形系统”的博士论文，首次使用了“Computer Graphics”这一概念，证明了交互式计算机图形学是一个可行的、有应用价值的研究领域，从而确立了计算机图形学正式成为一个独立学科的分支。1968 年 IvanE.Sutherland 又发表了《头戴式三维显示器》的论文，在头盔的封闭环境下，利用计算机成像的左右视图匹配，生成立体场景，使人置身于虚拟现实。IvanE.Sutherland 为计算机图形学技术做出了巨大的贡献，被称作计算机图形学的开山鼻祖，1988 年 IvanE.Sutherland 被授予 A.M 图灵奖。并且这一时期，光栅图形学算法开始萌芽。

计算机图形学在 20 世纪 70 年代进入了兴盛期。这一时期，光栅图形学算法迅速发展，区域填充、裁剪、消隐等概念及其相应算法纷纷被提出，实用的 CAD 图形系统也开始出现。除此之外，真实感图形学和实体造型技术的产生也是 70 年代计算机图形学的两个重要进展。1970 年 J.Bouknight 在 ACM 上发表论文，提出了第一个光反射模型，指出物体表面朝向是确定物体表面上一点光强的主要因素，并用 Lambert 漫反射定律计算物体表面上各多边形的光强，对光照射不到的地方用环境光代替。1971 年 Henri Gouraud 在 IEEE Trans.Computer 上提出被称为 Gouraud 明暗处理的“漫反射模型+插值”思想，对多面体模型，用漫反射模型计算多边形顶点的光亮度，再用增量法插值计算多边形的其他内部点。1975 年 Phong 在 ACM 上发表论文提出了著名的简单光照模型“Phong 模型”，Phone 模型虽然只是一个经验模型，但其真实度已经达到了较好的显示效果。这些都是真实感图形学最早的开创性工作。从 1973 年开始，相继出现了英国剑桥大学 CAD 小组的 Build 系统、美国罗彻斯特大学的 PADL-1 系统等实体造型系统，这些都为 CAD 领域的发展做出了重要贡献。

70 年代图形软件标准化程度提高。1974 年，ACMSIGGRAPH “与机器无关的图形技术”的工作会议的召开，提出了图形软件标准化问题，ACM 成立图形标准化委员会，制定“核心图形系统”（core graphics system），ISO 发布 CGI、CGM、GKS、PHIGS 一系列的图形标准，其中 1977 年的 CKS 是 ISO 批准的第

一个图形软件标准软件，是一个二维图形软件标准，1986 年，ISO 公布了程序员级的分层结构交互图形系统 PHIGS，这是一些非官方的图形软件，广泛应用于工业界并成为事实上的标准，PHIGS 是对 CKS 的扩充，增加的功能有对象建模、彩色设定、表面绘制和图形管理等。伺候 PHIGS 的扩充成为 PHIGS+；1988 年的 CKS3D，是 ISO 批准的第二个图形软件标准软件，是一个三维图形软件标准。

20 世纪 80 年代以后，出现了带有光栅扫描显示器的微型计算机和图形工作站，极大地推动了计算机图形学的发展，如 Machintosh、IBM 公司的 PC 及其兼容机，Apollp、Sun 工作站等。随着奔腾 III 和奔腾 IV 系列 CPU 的出现，计算机图形软件功能开始部分地由硬件实现。高性能显卡和液晶显示屏的使用，高传输率大容量硬盘的出现，特别是 Internet 的普及使得微型计算机和图形工作站在运算速度、图形显示细节上的差距越来越小，这些都为图形学的飞速发展奠定了物质基础。1980 年 Turner Whitted 提出了光透视模型，并第一次给出光线跟踪算法的范例，实现了 Whitted 模型；1984 年美国 Cornell 大学和日本广岛大学的学者分别将热辐射工程中的辐射度算法引入到计算机图形学中，用辐射度的方法成功地模拟了理想漫反射表面间的多重漫反射效果。以上二者的提出，标志着真实感图形的显示算法已逐渐成熟。80 年代中期以后，超大规模集成电路的发展，计算机运算能力的提高，图形处理速度的加快，促进了图形学各个研究方向的充分发展和广泛应用。

20 世纪 90 年代以后，微机和软件系统的普及使得图形学的应用领域日益广泛，计算机图形学朝着标准化、集成化和智能化的方向发展，多媒体、人工智能、计算机可视化、虚拟现实等分支蓬勃发展，三维造型也获得了长足发展。ISO 公布的图形标准越来越精细，更加成熟。这时存在着一些事实上的标准，如 SGI 公司开发的 OpenGL 开放式三维图形标准，微软公司为 PC 游戏开发的应用程序接口标准 DirectX 等，Adobe 公司 Postscript 等，均朝着开放、高效的方向发展。

过去十年，随着数字化技术和互联网的发展，计算机图形学在许多领域都已经得到了广泛的应用，如遥感图像分析、多媒体通信、医疗诊断、机器人视觉等。当前计算机图形学的研究逐渐向多学科交叉融合方向发展，即有与认知计算、计算器学习、人机交互的融合，也有与大数据分析、可视化的融合；不仅针对三维数字模型，而且涵盖了图像视频，与计算机视觉深度交叉。计算机图形学的发展

的一个潜在趋势是不再有明确清晰的主体，更多地体现出方法和技术上的创新。

近两年来，计算机图形学重要期刊会议的相关论文的热点研究内容主要集中在自监督学习（Self-Supervised Learning）、全景分割（Panoptic Segmentation）、网络结构搜索（Neural Architecture Search）和生成式对抗网络（Generative Adversarial Networks）等方面。

自监督学习研究早期主要集中在代理任务的设计和选取上，怎样的代理任务才能最好地提取出有益于下游任务的特征以及为何这些代理任务能够有效，这些是理论层面上自监督学习仍需要解决的问题。随着大量围绕着实例判别代理任务的相关工作的提出，一些工作将其中的核心思想进行展开并提出了对比学习的概念。通过将原来两个图片实例特征间的对比延伸到任意两个模态间特征的对比，使得模型学习不同模态间一致的特征表达并用最大化互信息作为新的衡量准则。

在已有工作中，比较典型的代理任务有将图片分块然后预测不同分块间的相对位置或者将分块打乱后重排得到原图，以及基于图片的上下文信息进行补全和图片不同颜色通道间的相互预测等。目前在图像与图形学领域，取得性能突破的方法主要仍局限在监督学习的框架之下，随着无标记数据的不断爆增和模型性能进一步提升的需求，无监督学习将会越来越受到学术界和工业界的重视。作为目前无监督学习中的一支，自监督学习因其良好的特征判别能力和对大规模数据扩展能力，也将受到更广泛的关注。

全景分割作为一个统一的任务在 2018 年被提出，它的目标是为图像中的所有像素点都分配一个语义类别和一个实例编号。从另一个角度来说，全景分割算法需要预测出图像中每一个像素点的所属类别和所属实例。在全景分割任务的基础上，近期的进展主要体现在三个方面：（1）从图像整体的角度考虑全景分割，共享网络主干（backbone）形成设计整体网络结构；（2）考虑图像中不同元素之间的交互，建模物体与语义概念之间的关系；（3）提出可学习模块，解决预测结果层面的冲突。

目前深度学习的方法在各类图像与图形分析任务中取得了非常大的成功，伴随这一成功而来的是对网络结构设计要求的不断提高。自动化网络设计自然而然地成为了自动化机器学习的下一个目标。早期的相关工作证明了使用强化学习算

法可以发现好的网络架构，但是这些方法在计算过程中需要消耗大量计算资源，因此后续的工作都集中在如何减少计算负担上。搜索空间的设计也是一项重要研究热点，同时，研究人员又拓宽了神经结构搜索的视野，将多种优化目标考虑在内，而不仅仅是减少搜索时间和提高网络精度。

在图像合成方面，近期最引人关注的工作是生成对抗网络。生成对抗网络由一个生成网络 G 和一个判别网络 D 组成。生成网络 G 和判别网络 D 在训练阶段使用对抗的方式进行学习，生成网络 G 的目标是生成尽可能真实的图片使得判别网络认为这是一张真实的图片；而判别网络 D 的任务则是判别合成的图像是真实的还是生成的。在这种两者对抗的学习过程中，生成 G 学会如何生成真实的图片。目前，生成对抗网络的主要研究方向包括条件生成对抗网络、损失函数的改进、模型结构的改进及训练方法的改进等。

借助 AMiner 平台，从计算机图形研究领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-49 所示。

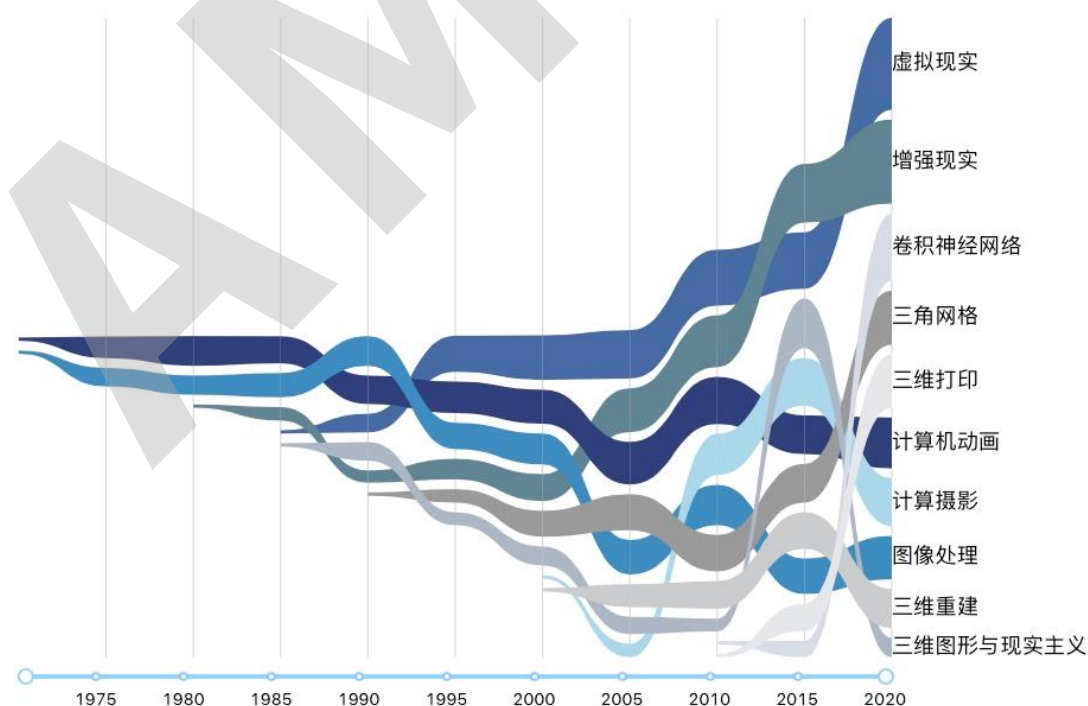


图 5-49 计算机图形领域技术研究发展趋势

计算机图形领域研究热度较高的 10 个技术话题，包括：虚拟现实（Virtual Reality）、增强现实（Augmented Reality）、计算摄影（Computational Photography）、计算机动画（Computer Animation）、三维图形与现实主义（Three-Dimensional Graphics And Realism）、三角网格（Triangle Mesh）、图像处理（Image Processing）、卷积神经网络（Convolutional Neural Network）、三维打印（3D Printing）、三维重建（3D Reconstruction），如上图所示。其中虚拟现实和增强现实等技术话题的研究热度一直维持在较为领先的位置，卷积神经网络与三维打印近几年来也呈现出快速上升的发展态势。

5.11.6 多媒体

多媒体技术是融计算机、声音、文本、图像、动画、视频和通信等多种功能于一体的技术，它借助日益普及的高速信息网，可实现计算机的全球联网和信息资源共享，并且它给传统的计算机系统、音频和视频设备带来了方向性的变革，对大众传媒产生深远的影响。

在计算机领域，多媒体（Multimedia）技术中的“媒体”有下列五大类^[39]：

（1）感觉媒体（Perception medium）指的是能使人产生直接感觉的媒体。如声音、动画、文本等；（2）表示媒体（Representation medium）指的是为了传送感觉媒体而人为研究出来的媒体。如语言编码、电报码、条形码等；（3）显示媒体（Presentation medium）指的是用于通信中使电信号和感觉媒体之间产生转换用的媒体。如键盘、鼠标器、打印机等；（4）存储媒体（Storage medium）指的是用于存放某种媒体的媒体。如纸张、磁带、磁盘、光盘等；（5）传输媒体（Transmission medium）指的是用于传输某些媒体的媒体。如电话线、电缆、光纤等。

多媒体是多种媒体的综合，一般包括文本，声音和图像等多种媒体形式，在计算机系统中，多媒体指组合两种或两种以上媒体的一种人机交互式信息交流和传播媒体。多媒体技术是通过计算机对语言文字、数据、音频、视频等各种信息进行存储和管理，使用户能够通过多种感官跟计算机进行实时信息交流的技术。多媒体技术所展示、承载的内容实际上都是计算机技术的产物。多媒体技术是信

³⁹ 王昕. 关于媒体和多媒体的概念[J]. 现代电信科技, 1998(10): 44.

息时代的产物，已经发展成为现代信息交流的关键方式。

多媒体技术的一些概念和方法，起源于 20 世纪 60 年代，实现于 20 世纪 80 年代中期。自 20 世纪 90 年代以来，多媒体技术逐渐成熟，从以研究开发为重心转移到以应用为重心。在近几年的研究中，多媒体技术呈现出与计算机体系结构、计算机网络、人机交互、信息安全、社会网络等多学科交叉融合的发展趋势。

多媒体分析以文本、图像、声音、视频等多种不同类型媒体的数据为研究对象，主要的研究目的一方面是使计算机具备人类的多媒体（如视、听）理解能力，另一方面是从多媒体数据中挖掘信息和知识、帮助人类更好地理解世界。

多媒体技术研究领域包括多媒体信息处理、多媒体数据压缩编码、多媒体内容分析与检索技术、多媒体交互与集成、多媒体通信与网络、多媒体内容安全、多媒体系统与虚拟现实等。具体涉及的技术内容^[40]包括：

- 多媒体数据压缩：多模态转换、压缩编码；
- 多媒体处理：音频信息处理，如音乐合成、语音识别、文字与语音相互转换；图像处理，虚拟现实；
- 多媒体数据存储：多媒体数据库；
- 多媒体数据检索：基于内容的图像检索，视频检索；
- 多媒体著作工具：多媒体同步、超媒体和超文本；
- 多媒体通信与分布式多媒体：CSCW、会议系统、VOD 和系统设计；
- 多媒体专用设备技术：多媒体专用芯片技术，多媒体专用输入输出技术；
- 多媒体应用：CAI 与远程教学，GIS 与数字地球、多媒体远程监控等。

多媒体技术的一些概念和方法，起源于 20 世纪 60 年代。1965 年，纳尔逊（Ted Nelson）为计算机处理文本文件提出了一种把文本中遇到的相关文本组织在一起的方法，称之为“hypertext（超文本）”。与传统的方式不同，超文本以

⁴⁰ 多媒体技术；

<https://wiki.mbalib.com/wiki/%E5%A4%9A%E5%AA%92%E4%BD%93%E6%8A%80%E6%9C%AF>.

非线性方式组织文本，使计算机能够响应人的思维以及能够方便地获取所需要的信息。万维网（WWW）上的多媒体信息正是采用了超文本思想与技术，组成了全球范围的超媒体空间。

多媒体技术实现于 20 世纪 80 年代中期。1984 年美国 Apple 公司在研制 Macintosh 计算机时，为了增加图形处理功能，改善人机交互界面，创造性地使用了位映射（bitmap）、窗口（window）、图符（icon）等技术。这一系列改进所带来的图形用户界面（GUI）深受用户的欢迎，加上引入鼠标作为交互设备，配合 GUI 使用，大大方便了用户的操作。Apple 公司在 1987 年又引入了“超级卡”（Hypercard），使 Macintosh 机成为更容易使用、易学习并且能处理多媒体信息的机器，受到计算机用户的一致赞誉。

自 20 世纪 90 年代以来，多媒体技术逐渐成熟，进入标准化阶段。多媒体技术从以研究开发为重心转移到以应用为重心。

静态图像的一个标准，是国际电信联盟（ITU）的 T.81。静态图像的主要标准称为 JPEG 标准（ISO/IEC10918）。它是 ISO 和 IEC 联合成立的专家组 JPEG（Joint Photographic Experts Group）建立的适用于单色和彩色、多灰度连续色调静态图像国际标准。该标准在 1991 年通过，成为 ISO/IEC10918 标准，全称为“多灰度静态图像的数字压缩编码”。

视频/运动图像的主要标准是国际标准化组织（ISO）下属的一个专家组 MPEG（Moving Picture Experts Group）制定的 MPEG-1（ISO/IEC11172）、MPEG-2（ISO/IEC13818）和 MPEG-4（ISO/IEC14496）三个标准。与 MPEG-1、4 等效的国际电信联盟（ITU）标准，在运动图像方面有利于视频会议的 H.261（Px64）、用于可视电话的 H.263。

在多媒体数字通信方面（包括电视会议等）制定了一系列国际标准（表 01-03-2），称为 H 系列标准。这个系列标准分为两代。H.320、H.321 和 H.322 是第一代标准，都以 1990 年通过的 ISDN 网络上的 H.320 为基础。H.323、H.324 和 H.310 是第二代，使用新的 H.245 控制协议并且支持一系列改进的多媒体编、解码器。

更深层次的多媒体技术标准也开始推出或列入开发中。一个典型的标准是称

作“多媒体内容描述接口”的 MPEG-7 标准（ISO/IEC15938）。与已经推出的几个 MPEG 标准不同，MPEG-7 是一个关于表示音/视信息的标准。它的七个组成部分中，系统、描述定义语言（DDL）、视频、音频和多媒体描述方案等已经成为正式标准，参考软件和一致性测试则计划在 2002 年 9 月成为标准。

多媒体各种标准的制定和应用极大地推动了多媒体产业的发展。很多多媒体标准和实现方法（如 JPEG、MPEG 等）已被作到芯片级，并作为成熟的商品投入市场。与此同时，涉及到多媒体领域的各种软件系统及工具也层出不穷。代表之一是进一步发展多媒体芯片和处理器。1997 年 1 月美国 Intel 公司推出了具有 MMX 技术的奔腾处理器（Pentium processor with MMX），使它成为多媒体计算机的一个标准。奔腾处理器在体系结构上有三个主要的特点：（1）增加了新的指令，使计算机硬件本身就具有多媒体的处理功能（新添 57 个多媒体指令集），能更有效地处理视频、音频和图形数据。（2）单条指令多数据处理（SIMD, Single Instruction Multiple Dataprocess）减少了视频、音频、图形和动画处理中常有的耗时的多循环。（3）更大的片内高速缓存，减少了处理器不得不访问片外低速存储器的次数。奔腾处理器使多媒体的运行速度成倍增加，并已开始取代一些普通的功能卡板。

随着网络电脑（Internet PC、NC）及新一代消费性电子产品，如电视机顶盒（Set-Top Box）、DVD、视频电话（Video Phone）、视频会议（Video Conference）等观念的崛起，强调应用于影像及通讯处理上最佳的数字信号处理器（DSP），经过另一番的结构包装，可由软件驱动组态的方式，进入咨询及消费性的多媒体处理器市场。

过去十年来，随着数字化技术的发展，多媒体技术突飞猛进，多媒体技术呈现出与计算机体系结构、计算机网络、人机交互、信息安全、社会网络等多学科交叉融合的发展趋势。近两年多媒体领域研究热点主要集中在大规模图像视频分析、社交媒体研究、多模态人机交互、计算视觉、计算图像、实时视频流化等方面。音视频技术是当前最活跃、发展最迅速的高新技术领域之一。

现在多媒体技术及应用正在向更深层次发展。下一代用户界面、基于内容的多媒体信息检索、保证服务质量的多媒体全光通信网、基于高速互联网的新一代

分布式多媒体信息系统等，多媒体技术和应用正在迅速发展。

由于多媒体数据往往是多种信息的传递媒介（例如一段视频中往往会同时使得文字信息、视觉信息和听觉信息得到传播），多模态学习已逐渐发展为多媒体内容分析与理解的主要手段。监督学习是多媒体数据分析的长远目标。在计算图像方面，大规模数据集的构建仍是一个热点研究方向，尤其语义对象的像素级标注需求越来越强烈，能够人机交互标注的过程中不断学习的协同标注方法得到了广泛关注。

此外，自动机器学习（AutoML）和元学习（Meta Learning）的最新研究成果及其在多媒体上的应用也逐渐增多。在图像压缩处理方面，也有一些研究工作将深度学习用于图像或视频压缩后处理，并得到了一定的效果。然而，现有工作的一个主要问题是用于后处理的深度网络较为复杂，计算速度慢，不满足实际应用的需求。如何在处理效果和处理速度之间取得一个折中，是压缩后处理的一个主要挑战。

图神经网络（Graph Neural Network, GNN）在多媒体领域的应用是近两年的热点研究方向，应用场景包括：个性化推荐，如基于多模态图卷积网络（MMGCN）的多模态推荐方法；短视频推荐，如使用基于图的顺序网络进行建模；多视频摘要，如采用图卷积网络衡量每个视频的重要性和相关性；基于文本的行人搜索，如使用深度对抗图注意力卷积网络（A-GANet）利用文本和视觉场景图学习联合特征空间；视频关系监测，如使用转移图神经网络（DoT-GNN）解决图像外观变化的问题。

随着 Mask-RCNN 与 RetinaNet 的发展，物体检测研究日趋成熟，但当前的技术依然存在诸多缺陷，为此，针对现代目标检测的基本框架（backbone、head、scale、batchsize 与 post-processing），神经网络架构搜索（NAS）以及细粒度图像分析（FGIA）等 3 个方面的潜在难题成为主要研究内容，尤其是后两者，将成为未来视觉物体检测的两个重要研究维度。

借助 AMiner 平台，从多媒体研究领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与

当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-50 所示。

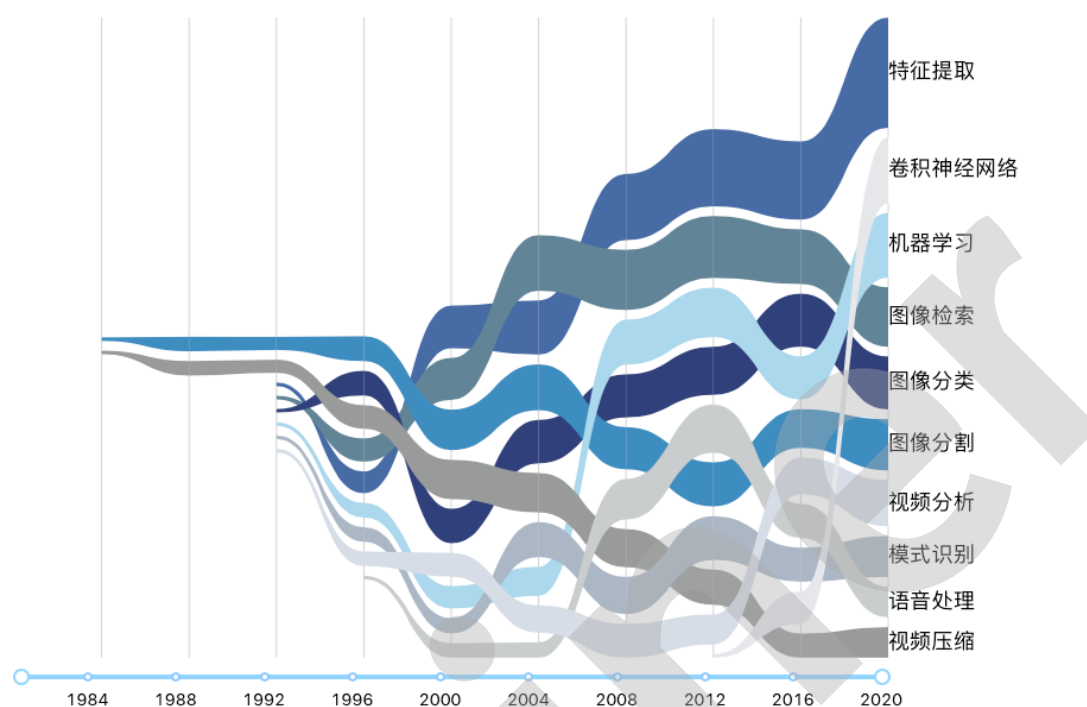


图 5-50 多媒体领域技术研究发展趋势

多媒体领域研究热度较高的 10 个技术话题，包括：图像检索（Image Retrieval）、机器学习（Machine Learning）、特征提取（Feature Extraction）、模式识别（Pattern Recognition）、卷积神经网络（Convolutional Neural Network）、语音处理（Speech Processing）、视频分析（Video Analysis）、视频压缩（Video Compression）、图像分类（Image Classification）、图像分割（Image Segmentation），如上图所示。近年来，特征提取的研究热度一直保持较高趋势，卷积神经网络的研究热度近五年来热度上升速度非常之快。

5.11.7 计算理论

计算理论是涉及理论计算机科学和数学等多学科综合的理论,是现代密码协议、计算机设计和许多应用领域的基础。它主要包括算法、算法学、计算复杂性理论、可计算性理论、自动机理论和形式语言理论等。计算理论于 19 世纪上半叶初具雏形,目前已广泛应用于应用程序设计编译等方面。随着科技的进步,计算理论将在更多领域展开应用。

借助 AMiner 平台,从计算理论领域相关的论文中抽取关键词,然后按照时间窗格统计词频,最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词,其宽度表示该技术关键词在当年的研究热度,与当年该技术关键词的论文数量呈正相关,每一年份中按照其热度由高到低进行排序,结果如图 5-51 所示。

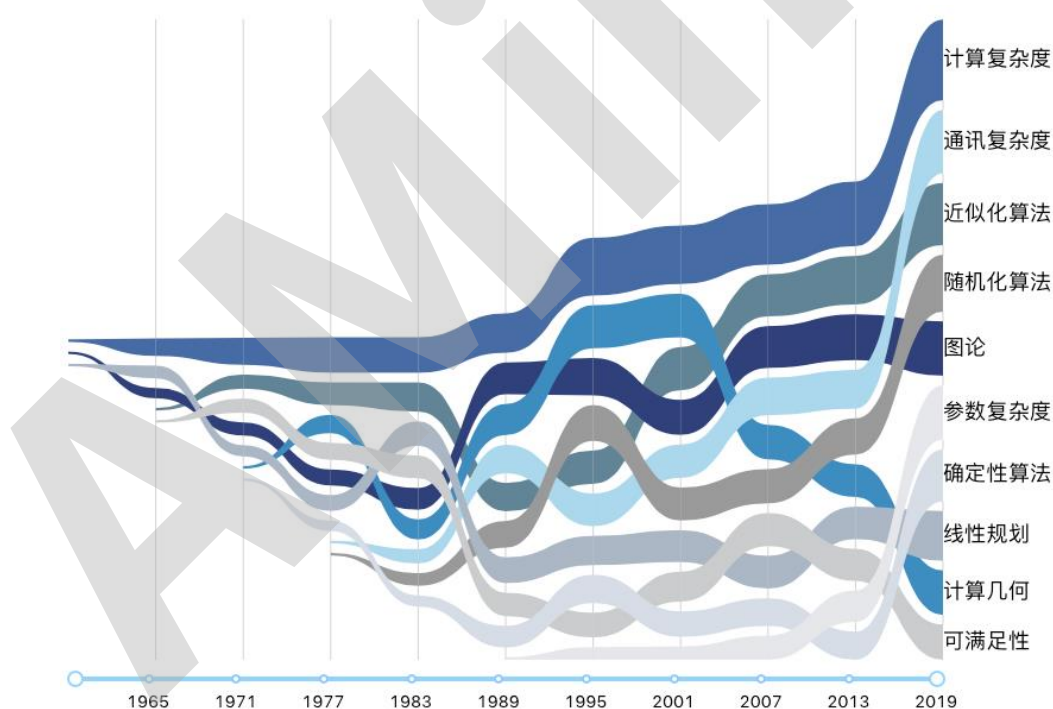


图 5-51 计算理论领域技术研究发展趋势

计算理论领域研究热度较高的 10 个技术话题,包括:计算复杂度 (Computational Complexity)、图论 (Graph Theory)、可满足性 (Satisfiability)、参数复杂度 (Parameterized Complexity)、计算几何 (Computational Geometry)、近似化算法 (Approximation Algorithm)、通讯复杂度 (Communication Complexity)、

线性规划（Linear Programming）、确定性算法（Deterministic Algorithm）、随机化算法（Randomized Algorithm），如图 5-51 所示。计算复杂度、通讯复杂度等技术话题的研究热度近年来快速上升，这主要是由于机器学习和深度学习在人工智能领域的广泛应用，为了提高算法效率，计算复杂度的研究热度一直居高不下。

5.11.8 计算机网络

计算机网络是指将地理位置不同的具有独立功能的多台计算机及其外部设备，通过通信线路连接起来，在网络操作系统、网络管理软件及网络通信协议的管理和协调下，实现资源共享和信息传递的计算机系统。最早的计算机网络诞生于 20 世纪 60 年代至 70 年代，20 世纪 70 年代计算机网络的基本概念初步形成。20 世纪 90 年代至今的、以因特网为代表的互联网成为最具有代表性的计算机网络。计算机网络是人工智能发展的重要依托平台，也是人工智能发展必不可少的条件。

借助 AMiner 平台，从计算机网络领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-52 所示。

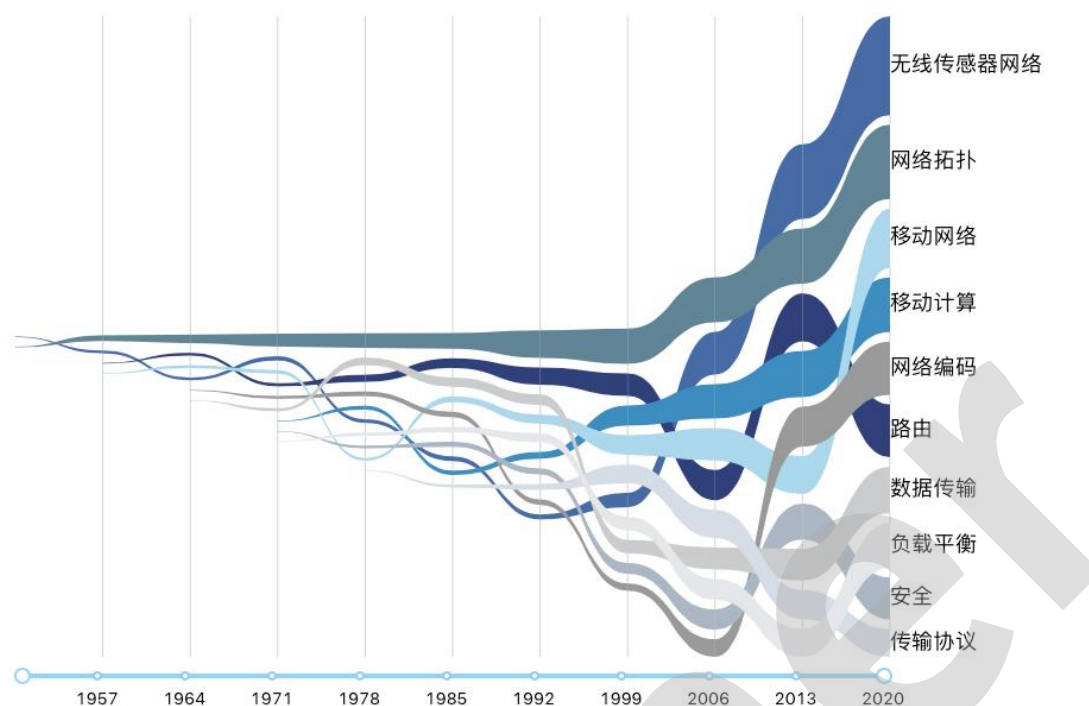


图 5-52 计算机网络领域技术研究发展趋势

计算机网络领域研究热度较高的 10 个技术话题，包括：无线传感器网络（Wireless Sensor Networks）、移动网络（Mobile Network）、网络拓扑（Network Topology）、路由（Routing）、数据传输（Data Transmission）、网络编码（Network Coding）、安全（Security）、移动计算（Mobile Computing）、负载均衡（Load Balancing）、传输协议（Transport Protocols），如图 5-52 所示。其中，无线传感器网络、网络拓扑等技术话题的研究热度在近十年内都呈现了比较明显的上升趋势。从有线网络到无线网络的变化加速了无线传感器网络的成长，通过无线传感器网络可接触到比有线网络更广泛、更弹性化以及可拓展的多元应用。随着近几年无线传感器网络技术的高速发展，越来越多的无线传感器网络产品开始投入使用，比如电气自动化、工业制造、导航定位等领域的无线传感装置。

5.11.9 计算机系统

计算机系统指用于数据库管理的计算机硬软件及网络系统。计算机系统由硬件系统和软件系统组成。硬件系统是借助电、磁、光、机械等原理构成的各种物理部件的有机组合，是系统赖以工作的实体。软件系统是各种程序和文件，用于

指挥全系统按指定的要求进行工作。自 20 世纪四十年代第一台计算机诞生发展至今,计算机系统在各个方面都有飞跃性的发展,其影响也深入到人工智能领域,为人工智能的发展提供了关键的技术支撑。

借助 AMiner 平台,从计算机系统领域相关的论文中抽取关键词,然后按照时间窗格统计词频,最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词,其宽度表示该技术关键词在当年的研究热度,与当年该技术关键词的论文数量呈正相关,每一年份中按照其热度由高到低进行排序,结果如图 5-53 所示。

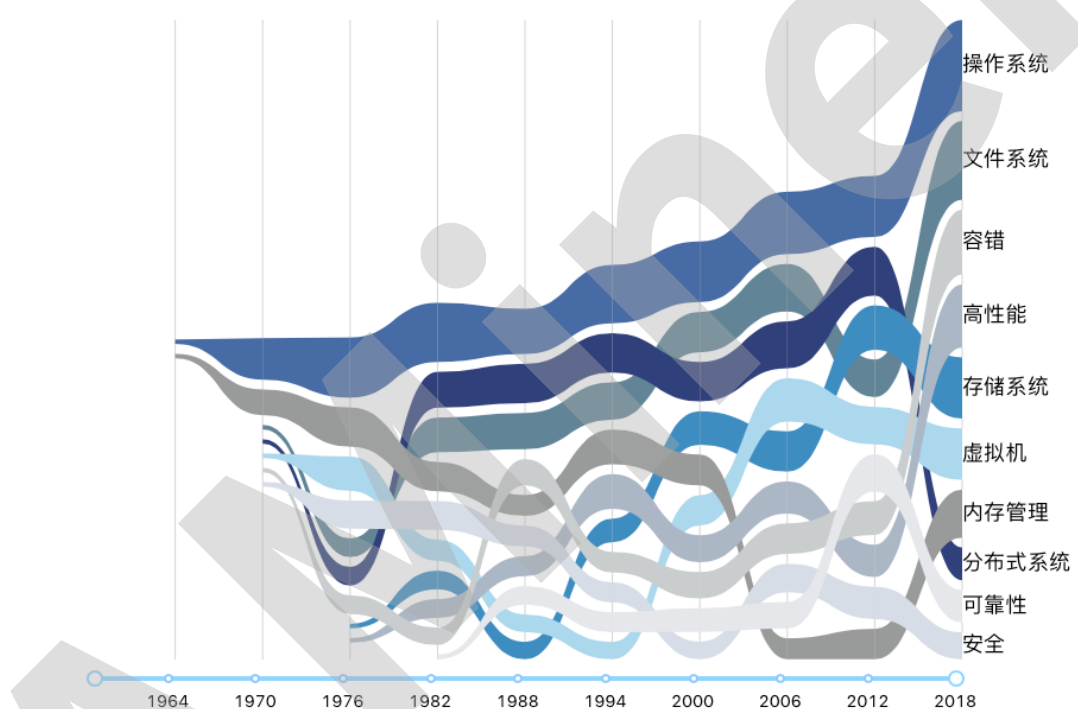


图 5-53 计算机系统领域技术研究发展趋势

计算机系统领域研究热度较高的 10 个技术话题,包括: 操作系统(Operating System)、文件系统(File System)、分布式系统(Distributed System)、虚拟内存(Virtual Memory)、存储系统(Storage System)、虚拟机(Virtual Machine)、内存管理(Memory Management)、高性能(High Performance)、容错(Fault Tolerance)、安全(Security), 如图 5-53 所示。从中可以看出操作系统、文件系统、容错和高性能等技术话题的研究热度上升很快,其中操作系统一直是领域内热度最高的技术话题。操作系统是计算机系统中最底层的软件,它控制所有计算机运行的程序并管理整个计算机的资源,是计算机裸机与应用程序及用户之间

的桥梁，因此它也一直受到计算机系统领域研究者的重视。

5.11.10 物联网

物联网是指通过信息传感器、射频识别技术、全球定位系统、红外感应器、激光扫描器等各种装置与技术，实时采集任何需要监控、连接、互动的物体或过程，采集其声、光、热、电、力学、化学、生物、位置等信息，通过网络接入，实现物与物、物与人的泛在连接，实现对物品和过程的智能化感知、识别和管理。20 世纪 90 年代物联网的概念已经被提及，21 世纪初物联网的概念正式提出，也标志着“物联网”通信时代的来临。物联网的核心是信息交互，其基本特征包括整体感知、可靠传输、智能处理三个方面，因此物联网在人工智能领域的应用十分广泛，其处理信息的功能在其中发挥至关重要的作用。

借助 AMiner 平台，从物联网研究领域相关的论文中抽取关键词，然后按照时间窗格统计词频，最终生成该领域的技术研究发展趋势图。技术趋势图中每条色带表示一个研究话题关键词，其宽度表示该技术关键词在当年的研究热度，与当年该技术关键词的论文数量呈正相关，每一年份中按照其热度由高到低进行排序，结果如图 5-54 所示。

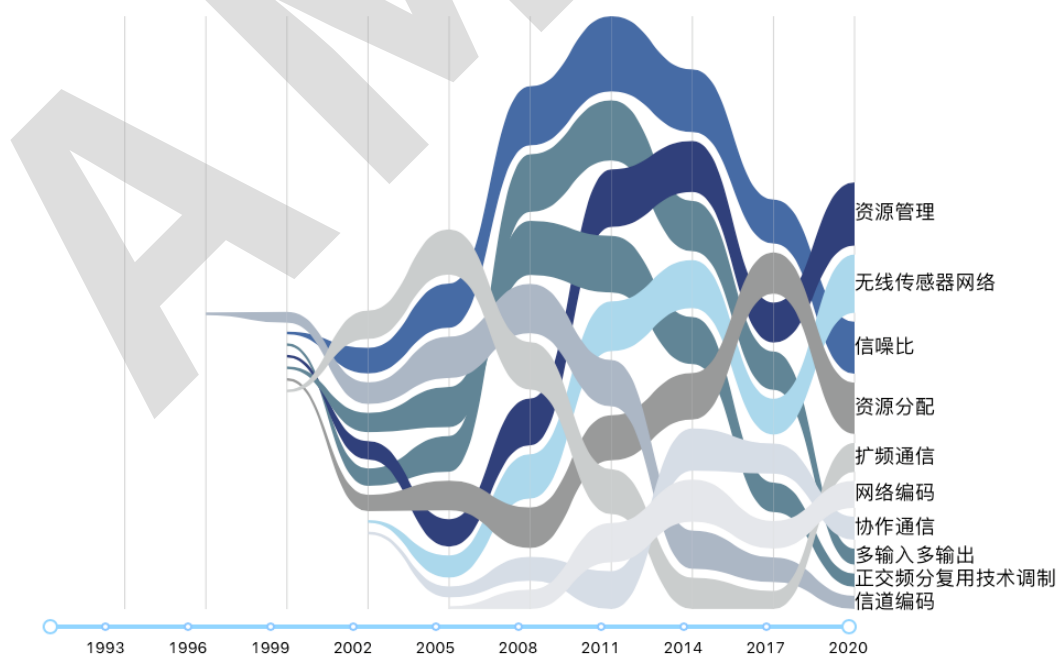


图 5-54 物联网领域技术研究发展趋势

物联网领域研究热度较高的 10 个技术话题，包括：信噪比（Signal To Noise Ratio）、资源管理（Resource Management）、资源分配（Resource Allocation）、多输入多输出（Multiple Input Multiple Output）、无线传感器网络（Wireless Sensor Network）、正交频分复用技术调制（Orthogonal Frequency Division Multiplexing Modulation）、信道编码（Channel Coding）、网络编码（Network Coding）、协作通信（Cooperative Communication）、扩频通信（Spread Spectrum Communication），如图 5-54 所示。其中资源管理和无线传感器网络等技术话题近年来的研究热度上升较快。

6 人工智能领域高层次人才现状

人才是经济社会发展的第一资源，更是国家科技创新经济发展的首要驱动力。高层次人才是知识创新和科技创新的关键力量，也是引领产业发展的核心要素。基于清华大学唐杰教授团队自主研发的“科技情报大数据挖掘与服务平台”（简称 AMiner）^[41]，本报告对人工智能领域科研人才进行分析发掘，分析得出人工智能领域人才现状。

在本报告中，人工智能领域**高层次学者**的判定原则是指被评选进入 AI 2000 人工智能全球最具影响力学者榜单（以下称为 AI 2000）的学者（全文同）。AI 2000 涵盖人工智能学科 20 个子领域，具体评选规则与遴选方法请参见详情页 <https://www.aminer.cn/ai2000>。

6.1 全球 AI 领域高层次人才分析

过去十年，全球人工智能发展迅速。中国、美国、欧盟、英国、德国等国家纷纷从战略上布局人工智能，加强顶层设计和人才培养。本报告数据显示，全球人工智能领域学者数量共计 155408 位，覆盖 120 多个国家，主要集中在北美洲、欧洲、东亚地区。

人工智能领域论文发表量反映了一个国家在领域的科研实力。在 AI 领域论文发表量 TOP10 的国家之中，美国、中国和德国的论文产出量分别位前三名，其余国家（英国、加拿大、日本、法国、澳大利亚、韩国和新加坡）领域论文产出量均在 2 万篇以下。其中，美国在 AI 领域的论文发表数量和人才数量都位于全球第一，有近四成的全球 AI 领域论文产出是来自美国，并且美国 AI 学者数量约占全球领域学者总量的 31.6%。中国在人工智能领域的论文发表数量（25418 篇）和人才数量（17368 位）仅低于美国，同时，大幅领先于其他国家。这反映出中国近年来发布的从产业发展、教育等各个方面支持人工智能发展的一系列支

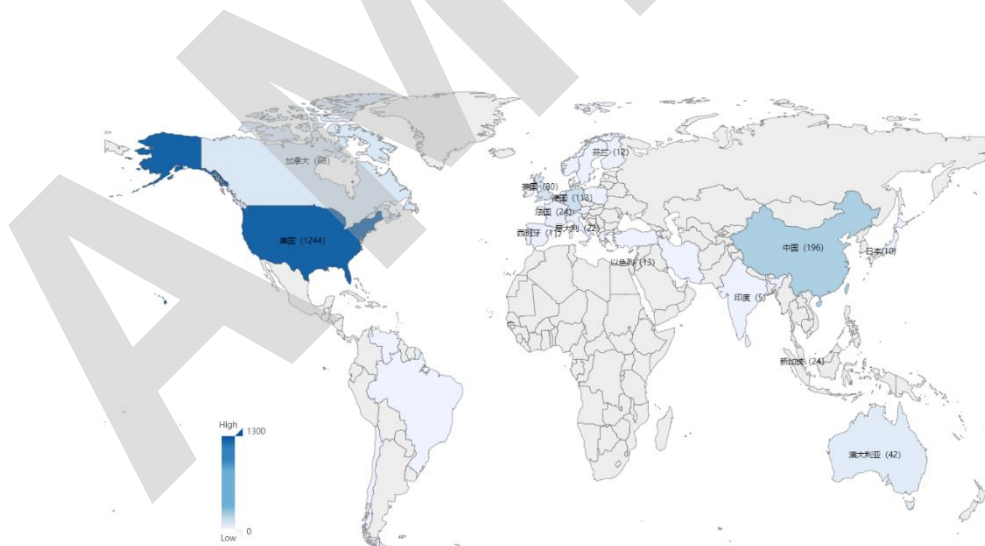
AMiner 平台（<https://www.aminer.cn/>），2006 年上线，经过十多年的建设发展，已建立全领域论文文献及专家学者库，收录论文文献超过 2.3 亿及论文引用关系 7.5 亿，收录学者 1.3 亿。

持政策，以及不断加强人才培养和补齐人才短板的行动已见成效。

从子领域的领先国家来看，美国在人工智能领域具有明显的科研产出优势，在几乎所有子领域的论文产出量均居于全球首位。中国的 AI 科研产出水平在自然语言处理、芯片技术、机器学习、信息检索与挖掘等 10 多个子领域都紧随美国之后，并且，在多媒体、物联网领域的论文产出量超过美国、居于全球第一；而在人机交互、知识工程、机器人、计算机图形、计算理论领域，中国还需努力追赶。

6.1.1 国家地区分布

AI 高层次学者是指入选 AI2000 榜单的 2000 位人才。由于存在同一学者入选不同领域的现象,经过去重处理后,AI 高层次学者共计 1833 位。从这些高层次学者分布看,如图 6-1 所示,AI 领域全球高层次学者覆盖全球 37 个国家,主要集中在北美洲的美国地区;欧洲中西部也有一定的高层次学者分布;亚洲的高层次人才主要分布于中国、新加坡及日韩等地区;其他诸如南美洲、非洲等地区的高层次学者数量稀少。

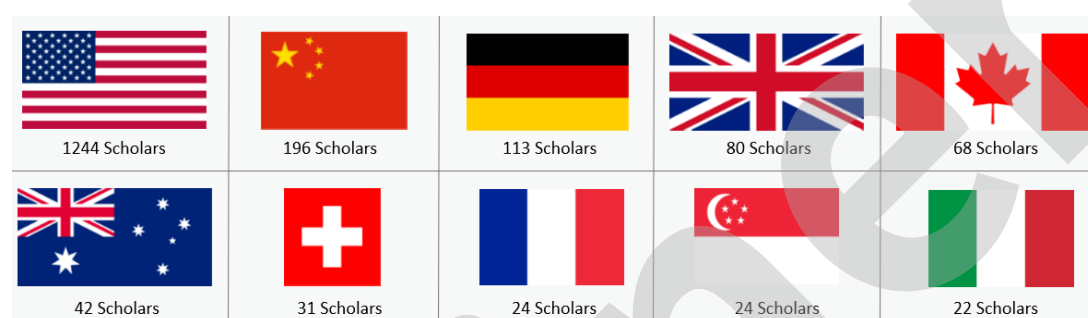


来源: AMiner

图 6-1 全球人工智能领域高层次学者分布

6.1.1 高层次学者数量 TOP10 国家

从国家角度看 AI 高层次学者分布，美国 AI 高层次学者的数量最多，有 1244 人次，占比 62.2%，超过总人数的一半以上，且是第二位国家数量的 6 倍以上。中国排在美国之后，位列第二，有 196 人次，占比 9.8%。德国位列第三，是欧洲学者数量最多的国家；其余国家的学者数量均在 100 人次以下。人工智能领域高层次学者人数 TOP10 的国家如图 6-2 所示。



来源：AMiner

图 6-2 人工智能领域高层次学者数量 TOP10 国家

6.1.3 所在机构分布

总体来看，全球范围内，美国和中国的机构在人工智能领域的论文产出和学者数量较多，占据了 AI 领域论文量排名前 10 机构的全部席位。

从 AI 高层次人才分布看，全球 AI 高层次人才隶属于各个国家的高等院校或高科技公司的科研部门。如图 6-3 所示，全球人工智能领域高层次学者量 TOP10 机构之中，位居首位的是美国的谷歌公司，拥有 185 人，也是唯一一家高层次学者数过百的机构；从国家分布来看，清华大学是唯一入选 TOP10 的中国机构，其余均为美国机构，且美国机构高层次学者总体人数遥遥领先。

 185 Scholars	 91 Scholars	 59 Scholars	 51 Scholars	 50 Scholars
 45 Scholars	 44 Scholars	 43 Scholars	 28 Scholars	 27 Scholars

来源：AMiner

图 6-3 全球人工智能领域高层次学者量 TOP10 机构

从子领域论文量来看，美国的大学和科技机构在 AI 各个细分方向上的发展较为均衡，且在自然语言处理、芯片技术、机器学习、信息检索与挖掘、人机交互等 10 多个子领域的发展居于全球领先席位。这反映出美国在人工智能领域的顶级实力。

中国的 AI 机构在语音识别、经典 AI、计算机网络、多媒体、可视化和物联网等领域实力较强，进入全球先进行列。这些机构主要是位于北京的清华大学、中科院和北京邮电大学，以及浙江大学。

6.2 中国 AI 领域高层次人才分析

过去十年，我国人工智能发展迅猛。2017 年，人工智能首次被写入全国政府工作报告，我国确定新一代人工智能发展三步走战略目标，并将人工智能上升为国家战略层面。本报告数据显示，我国人工智能领域学者数量共计 17368 位，覆盖 100 多个国内城市。从地域分布来看，我国 AI 人才主要集中在京津冀、长三角和珠三角地区。

6.2.1 省市地区分布

国内 AI 领域高层次人才也主要分布在京津冀、长三角和珠三角地区，如图 6-4 所示。其中，京津冀地区（主要是北京市）在 AI 领域的高层次人才数量最多。长三角地区也有较多的 AI 高层次人才分布。相比之下，内陆地区领域高层次人才较为缺乏。在学者分布地图中，颜色越深，表示学者越集中；颜色越浅，

表示学者越稀少。

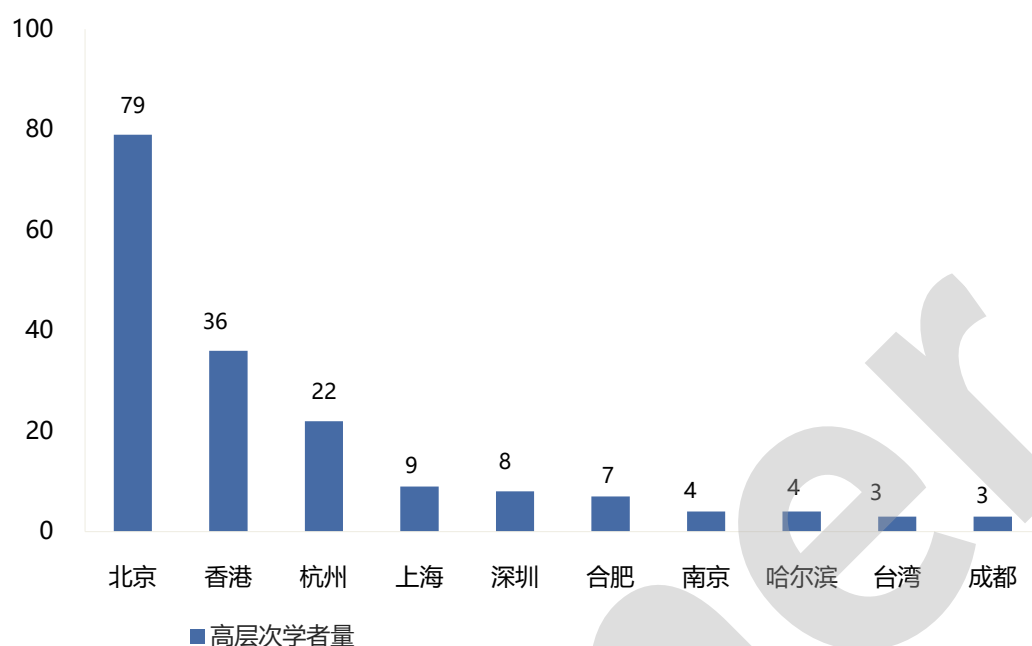


来源：AMiner

图 6-4 我国人工智能领域高层次人才的省市分布图

6.2.2 高层次学者数量 TOP10 国内城市

从 AI 高层次学者分布来看，北京仍是拥有 AI 高层次学者数量最多的国内城市，有 79 位，占比 45.4%，接近于国内 AI 高层次人才的一半，如图 6-5 所示。北京作为政治中心、文化中心、国际交往中心、科技创新中心具有先天优势，拥有数量众多的 AI 企业和多所知名高校和研究机构，北京的高水平 AI 论文发表量和高层次学者量明显领先于其他国内城市。同时，从子领域发展来看，北京在 AI 各个细分方向上的发展较为均衡，相关论文产出量均居于全国领先地位。



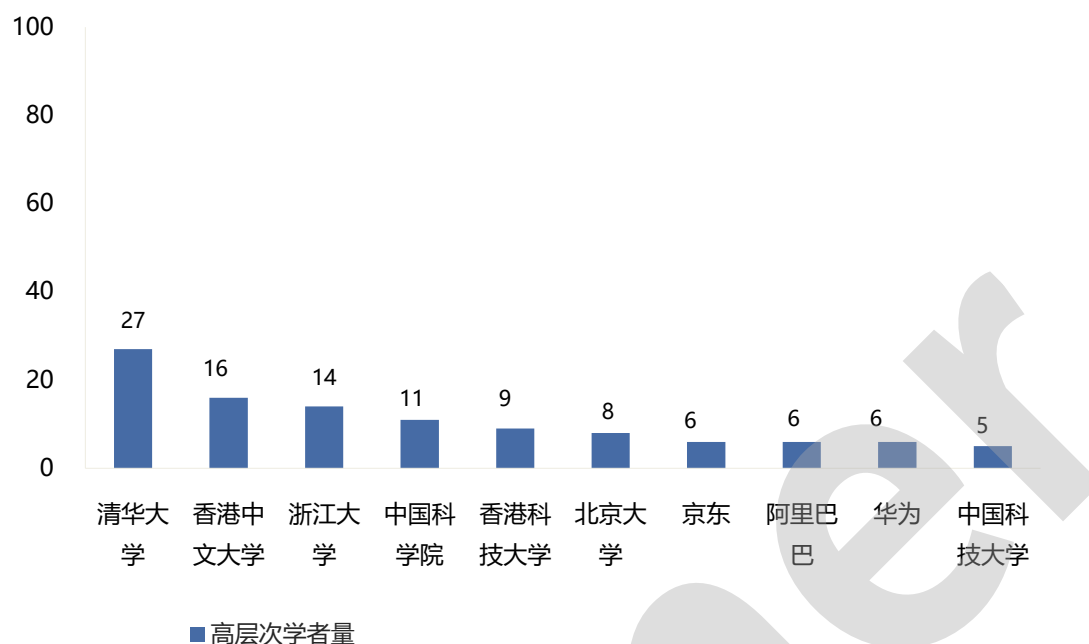
来源: AMiner

图 6-5 人工智能领域高层次学者数量 TOP10 的中国城市

6.2.3 所在机构分布

国内人工智能领域研究领先的机构主要以北京、香港、杭州、上海等地的高等院校为主。北京在人工智能领域的资源优势,该城市拥有清华大学、北京大学、中国科学院等知名高校。杭州和香港的机构也处于 AI 子领域研究前列,主要由于前者拥有阿里巴巴和浙江大学,后者则因其香港科技大学和香港中文大学等实力高校。

在国内机构之中,北京的清华大学不仅拥有 AI 领域学者数量最多,而且所拥有的领域高层次人才数量也居于国内首位,有 27 位。国内高层次 AI 人才基本都隶属于高校。香港中文大学、浙江大学和中国科学院在人工智能领域的高层次学者数量分别为 16、14 和 11 位。其他的国内机构所拥有的 AI 领域高层次人才数量均不足十位,如图 6-6 所示。

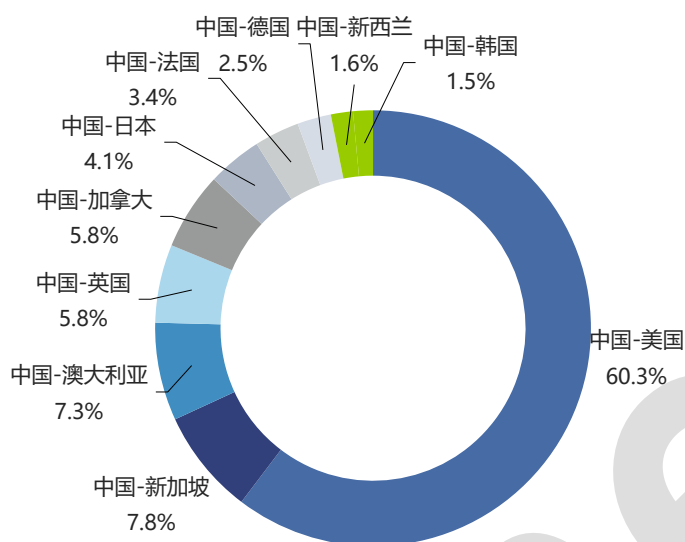


来源: AMiner

图 6-6 人工智能领域高层次学者数量 TOP10 的中国机构

6.3 中国 AI 高层次人才的跨国合作论文分析

中国 AI 领域学者积极展开跨国科研论文合作并发表领域高水平论文。与中国领域学者合作最多 10 个国家分别是美国、新加坡、澳大利亚、英国、加拿大、日本、法国、德国、新西兰和韩国,如图 6-7 所示。其中,美国是中国 AI 领域学者跨国合作最多的科研伙伴,两国领域学者合作论文量在中外论文合作之中占比过半。



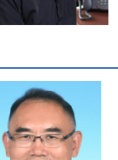
来源: AMiner

图 6-7 中国 AI 领域学者跨国合作论文量 TOP10 国家占比

中国在各个 AI 子领域的跨国合作论文产出数量差别较大。相对而言,在计算机视觉、经典 AI、语音识别、信息检索与挖掘、机器人、自然语言处理、机器学习和多媒体领域,中外 AI 合作论文较多。而在芯片技术、人机交互、知识工程、数据挖掘等领域的外中合作论文数量较少,未来需要进一步加强。

6.4 代表性高层次 AI 学者

AI 子领域	代表学者画像
机器学习	 Ilya Sutskever Chief Scientist OpenAI United States Neural Network Probabilistic Model Fixed Point Belief Network Boltzmann Machine Powerful Sequence Model Deep Learning 浏览量: 6303 h-index: 46 论文数: 82 引用数: 66388
自然语言处理	 Christopher D. Manning Professor Computer Science Department, Stanford University United States Information Retrieval Knowledge Management Computer Science Natural Language Processing Natural Language Unsupervised Learning 浏览量: 4611 h-index: 121 论文数: 464 引用数: 126766

信息检索与推荐	 Haewoon Kwak Researcher Qatar Computing Research Institute, Hamad Bin Khalifa University Qatar Social Network Case Study Greedy Algorithms Machine Learning Toxic Behavior Community Mining Complex Network Crowdsourcing 浏览量: 1001 h-index: 25 论文数: 110 引用数: 13558
知识工程	 Christian Bizer Professor University of Mannheim Germany Semantic Web Linked Data Web Of Data RDF Data Source Dbpedia Data Integration Semantic Web Technology 浏览量: 530 h-index: 52 论文数: 151 引用数: 25184
语音识别	 Geoffrey E. Hinton 教授 Department of Computer Science, University of Toronto Canada Neural Network Boltzmann Machine Speech Recognition Unsupervised Learning Neural Networks Hidden Markov Model 浏览量: 49623 h-index: 151 论文数: 447 引用数: 313000
计算机视觉	 何恺明 (Kaiming He) ★ 其他 Facebook AI Research United States Object Detection Computer Vision Feature Extraction Image Classification Vectors Proposals Encoding Image Matting 浏览量: 6813 h-index: 60 论文数: 83 引用数: 132732
机器人	 Wolfram Burgard Professor University of Freiburg, Department of Computer Science Germany Mobile Robots Mobile Robot Path Planning Navigation Simultaneous Localization And Mapping Robots Probability Robot Vision 浏览量: 2300 h-index: 110 论文数: 841 引用数: 80334
数据挖掘	 Jure Leskovec Associate Professor Computer Science Department, Stanford University United States Social Network Social Networks Machine Learning Social Media Data Mining Information Network Information Cascades Information Diffusion 浏览量: 6805 h-index: 106 论文数: 310 引用数: 70267
人机交互	 Meredith Ringel Morris Microsoft Research Lab/Paul G. Allen School of Computer Science & Engineering, University of Washington United States Social Network Accessibility User Interface Search Engine Solar System Social Media Collaborative Web Search Crowdsourcing 浏览量: 1455 h-index: 60 论文数: 212 引用数: 13093
可视化	 Jeffrey Michael Heer Professor Computer Science & Engineering, University of Washington/Paul G. Allen School of Computer Science & Engineering, University of Washington United States Visualization Information Visualization Data Visualisation Data Analysis Data Visualization Interactive Visualization Visual Analysis Social Data Analysis 浏览量: 1208 h-index: 59 论文数: 153 引用数: 18177
经典 AI	 杨强 (Qiang Yang) ★ 教授 Department of Computer Science and Engineering, Hong Kong University of Science and Technology/WeBank China Data Mining Transfer Learning Machine Learning Collaborative Filtering Activity Recognition Case Base Reasoning Information Retrieval Training Data 浏览量: 20716 h-index: 107 论文数: 702 引用数: 58262

安全与隐私	 David Wagner Professor Computer Science Division, University of California, Berkeley United States Block Cipher Privacy Security Property Static Analysis Cryptanalysis Provably Secure Voting Machine Security 浏览量: 1281 h-index: 91 论文数: 219 引用数: 47444
芯片技术	 Anantha P. Chandrakasan Dean, MIT School of Engineering Massachusetts Institute of Technology United States Low Power Electronics Cmos Integrated Circuits Integrated Circuit Design Energy Efficient Low-power Electronics Energy Efficiency CHIP 浏览量: 3220 h-index: 99 论文数: 606 引用数: 61642
数据库	 Michael J. Franklin Professor University of Chicago United States Data Management Query Processing Sensor Network Database System Client Server Query Optimization Databases Cloud Computing 浏览量: 3787 h-index: 98 论文数: 340 引用数: 58363
计算机图形	 Frédéric Durand Professor Computer Graphics Group, CSAIL EECS Massachusetts Institute of Technology United States Global Illumination Computational Photography Image Processing Visibility Complex Data Structure Bilateral Filtering Graphics Hardware 浏览量: 1357 h-index: 77 论文数: 244 引用数: 26380
多媒体	 Trevor Darrell Professor CS Division, University of California, Berkeley United States Computer Vision Object Recognition Motion Estimation Real Time Image Classification Feature Extraction Gesture Recognition Tracking 浏览量: 16930 h-index: 116 论文数: 513 引用数: 93690
计算理论	 Craig Gentry Scientist IBM Research United States Homomorphic Encryption Homomorphic Encryption Scheme Public Key Encryption Homomorphic Operation Encryption Scheme 浏览量: 537 h-index: 50 论文数: 103 引用数: 25521
计算机网络	 Sachin Katti Associate Professor Computer Science Department, Stanford University United States Wireless Network Network Coding Wireless Communication Full Duplex Interference Cancellation Wireless Mobile Device Wireless Mesh Network 浏览量: 671 h-index: 50 论文数: 154 引用数: 18441
计算机系统	 Michael Acheson Isard Research Fellow Magdalen College United Kingdom Computer Vision Probability Distribution Kalman Filter Programming Model Information Retrieval Visual Tracking Active Contour Belief Propagation 浏览量: 838 h-index: 61 论文数: 107 引用数: 55505
物联网	 Jeffrey Andrews Professor University of Texas at Austin United States Interference Signal To Noise Ratio Throughput Base Station Probability Cellular Radio Cellular Network Stochastic Geometry 浏览量: 586 h-index: 73 论文数: 358 引用数: 28793

6.5 中国 AI 高层次人才培养

中国 AI 领域高层次人才培养从 2018 年起开始重点发展，主要由高校通过成立 AI 学院研究院、设立 AI 专业的方式进行培养。教育部在《高等学校人工智能创新行动计划》（教技〔2018〕3 号）中提出，要加强人工智能领域专业建设，形成“人工智能+X”复合专业培养新模式。到 2020 年建设 100 个“人工智能+X”复合特色专业，建立 50 家人工智能学院、研究院或交叉研究中心，并引导高校通过增量支持和存量调整，加大人工智能领域人才培养力度。到 2020 年，高校要基本完成新一代人工智能发展的高校科技创新体系和学科体系的优化布局。到 2030 年，高校要成为建设世界主要人工智能创新中心的核心力量和引领新一代人工智能发展的人才高地。

教育部于 2019 年 3 月颁布《关于公布 2018 年度普通高等学校本科专业备案和审批结果的通知》，将人工智能专业列入新增审批本科专业名单，专业代码为 080717T（T 代表特设专业），学位授予门类为工学。在此之前，国内没有高校在本科阶段设置人工智能专业。

2020 年 2 月，教育部颁布《关于公布 2019 年度普通高等学校本科专业备案和审批结果的通知》。统计结果显示，人工智能方面，本次全国范围内获得人工智能专业首批建设资格的共有 180 所，相比 2018 年的 35 所，增加 414%，反映出人工智能专业的热度攀升。

截至目前，国内共有 215 所高校成立“人工智能”本科专业。这些高校之中，有 60 所双一流大学（占比 28%），其他 155 所为普通本科院校。

从地域分布看，2019 年度新增人工智能专业较多的省份依次是山东 14 所、江苏 13 所、北京 11 所、安徽 10 所、河南 10 所、四川 10 所，其余省份新增人工智能专业的高校数量均不足 10 所。但这些数字加起来占全国高校总量比例仍然较小，高校人工智能本科教育仍处于起步和发展阶段。

在 AI 人才紧缺，国家政策推动 AI 发展趋势等因素影响下，相比于建设人工智能专业，很多高校更愿意设立一个人工智能研究独立学院。截至 2019 年 6 月，至少有 38 所高校设立了独立人工智能学院，全面开展本科阶段、研究生阶段的

教育，并且在 2019 年开始以人工智能专业招收本科生。

据统计^[42]，截止到 2019 年年底，我国已经有 66 所高校成功建设人工智能学院、研究院、研究中心或研究所，超额完成了教育部在《高等学校人工智能创新行动计划》中强调的到 2020 年在全国高校中建立 50 家人工智能学院、研究院或交叉研究中心的目标。

总之，中国各大高校设立人工智能一级学科、建立人工智能学院，有助于精准布点人工智能相关专业以满足国家和区域的产业发展需求，有助于加快建设一流人才队伍和高水平创新团队、进一步推动国际学术交流与合作、专业和教材建设，提高人才培养质量，推动科技成果转化。

⁴² 2019 全国 66 所高校人工智能学院、人工智能研究院建设情况一览表，高校大数据与人工智能推进联盟，2019 年 12 月 18 日

7 人工智能专利分析

专利是创新成果的应用表现形式。本部分将通过人工智能领域专利分析，挖掘该技术的创新应用情况。以墨创全球专利数据库作为数据来源，使用行业专家和相关技术领域专利审查专家共同给出的人工智能领域关键词在标题和摘要中进行检索，搜索时间范围限定为 2011-2020 年。

专利数据分析发现，随着核心算法的突破、计算能力的迅速提升以及海量数据的支撑，过去十年的人工智能专利申请量呈现逐渐上升态势。

7.1 全球 AI 专利分析

7.1.1 专利申请趋势

全球范围内，过去十年人工智能领域的专利申请量 521264，总体上呈逐年上升趋势，如图 7-1 所示。

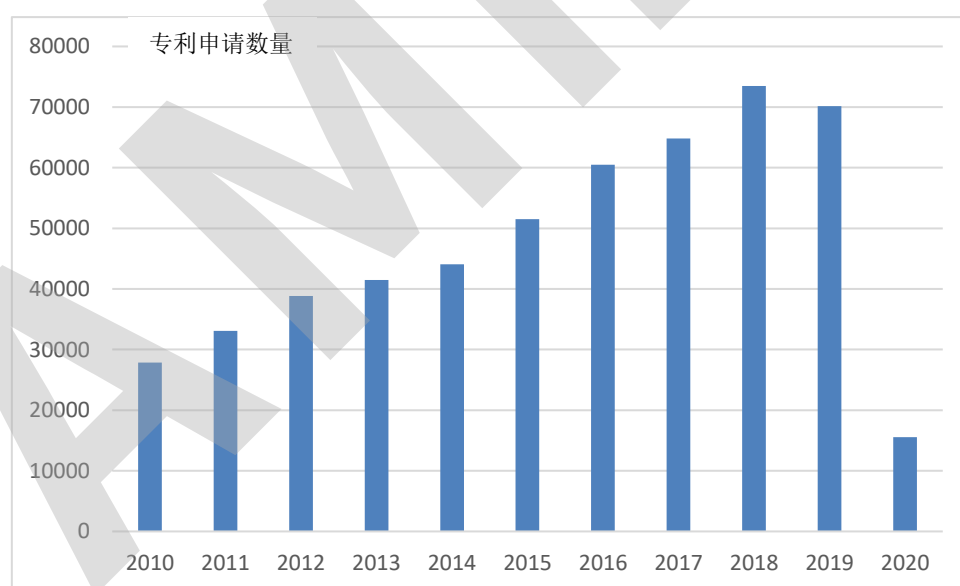


图 7-1 全球人工智能专利申请量年度变化趋势⁴³

⁴³ 受数据公开滞后影响，专利数据截止至 2020 年 8 月。

7.1.2 专利申请国家/地区排名

全球 AI 专利申请数量排名领先的国家/地区如图 7-2 所示。从图中可以看出，目前，全球人工智能专利申请集中在中国、美国、日本、韩国。其中，中国和美国处于领先地位，遥遥领先其他国家。中国专利申请量为 389571，位居世界第一，占全球总量的 74.7%，是排名第二的美国的 8.2 倍。

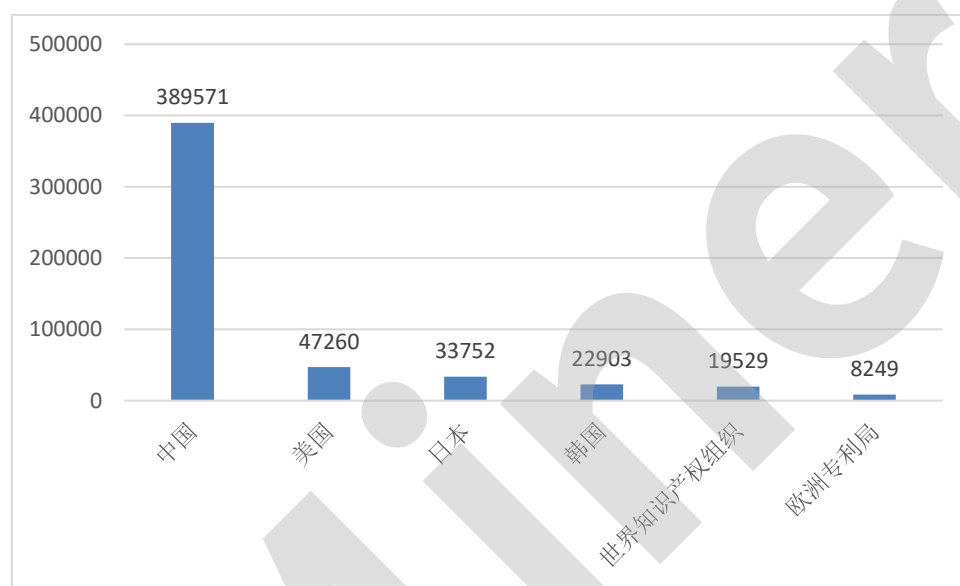


图 7-2 全球 AI 专利申请量 Top10 国家

7.1.3 申请人排名

过去十年，全球人工智能专利申请之中，将近一半的申请人是来自于企业。高校和研究所的相关申请量共计约两成。

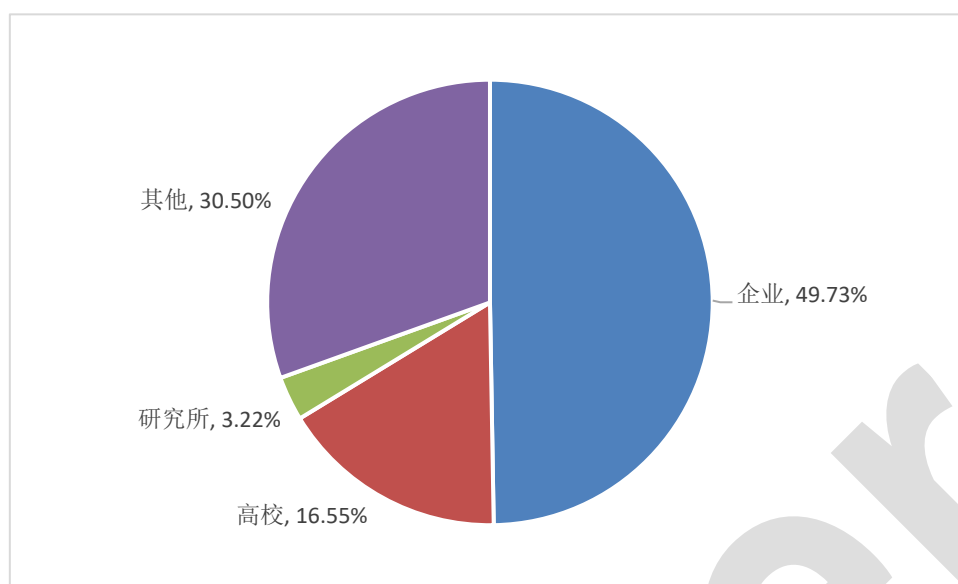


图 7-3 全球 AI 专利申请类型分布

人工智能专利申请量前十的机构集中在日本、中国、韩国和美国。其中，日本的佳能是一家致力于图像、光学和办公自动化产品的公司，该公司的绝大多数专利都与成像有关，申请量最高的人工智能功能应用类别是计算机视觉。美国 IBM 公司的专利申请很多都与 IBM 的自然语言处理和机器学习技术有关。中国的国家电网的专利申请多与电网控制、配电利用网络、风电场、绿色能源等领域的人工智能开发有关。

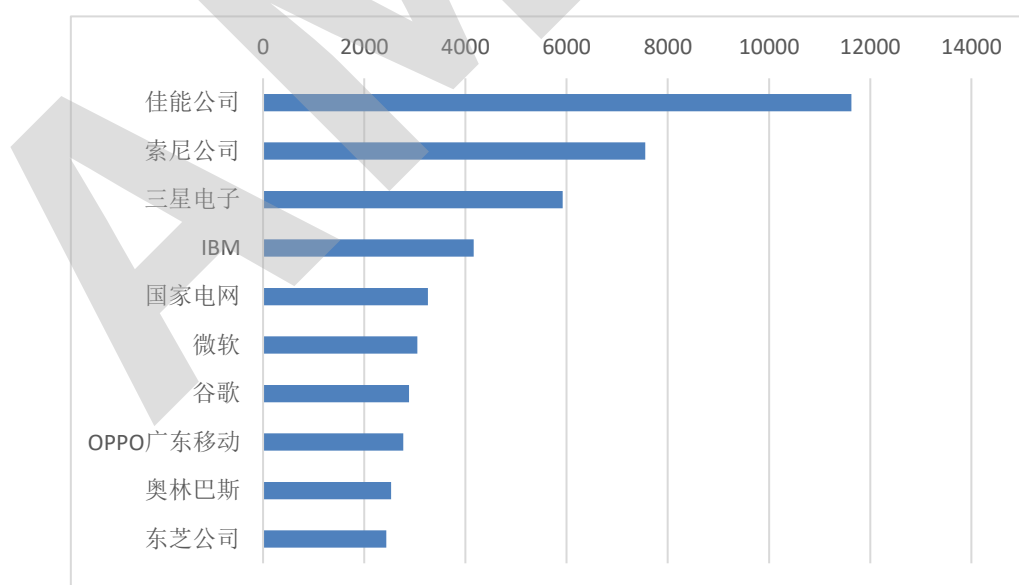


图 7-4 全球 AI 专利申请人排名 TOP10

7.2 中国 AI 专利分析

7.2.1 国内 AI 专利申请趋势

过去十年，中国人工智能领域的专利申请量 389571，约占全球申请量的 74.7%。总体上，国内的人工智能相关专利申请量呈逐年上升趋势，并且在 2015 年后增长速度明显加快，如图 7-5 所示。

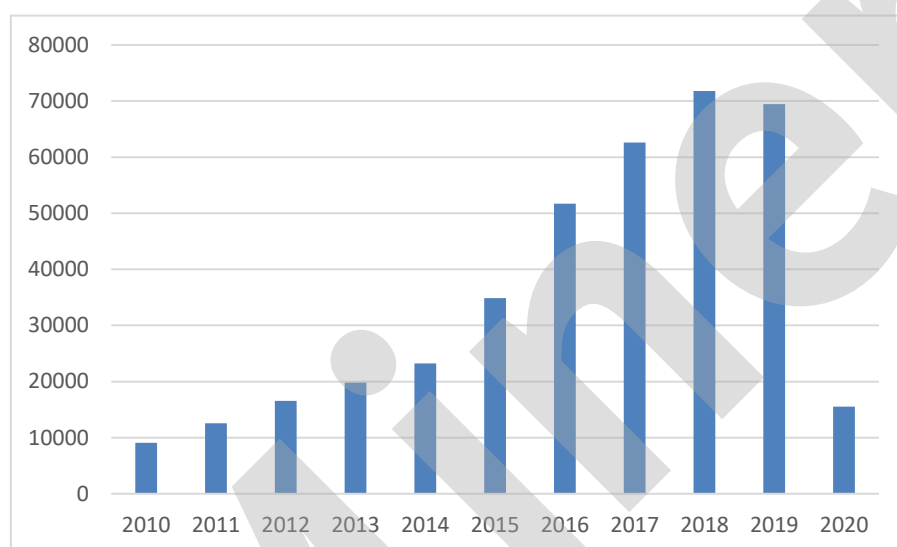


图 7-5 中国人工智能专利申请量年度趋势⁴⁴

7.2.2 专利申请量省市排名

中国各省市 AI 专利申请数量的分布情况如图 7-6 所示。从图中可以看出，广东省的 AI 专利申请量以 72737 位居第一，比排名第二的北京市（50906）多出 42.8%，具有突出优势。前十名的省份主要分布在东部、中部、西部等地区，分布较为均衡，但是以东部省市居多，江浙沪三省市均位居前五名。这与这些地区的经济水平、发展程度、科研投入及知识产权保护等因素密切相关。

⁴⁴ 受数据公开滞后影响，专利数据截止至 2020 年 8 月。

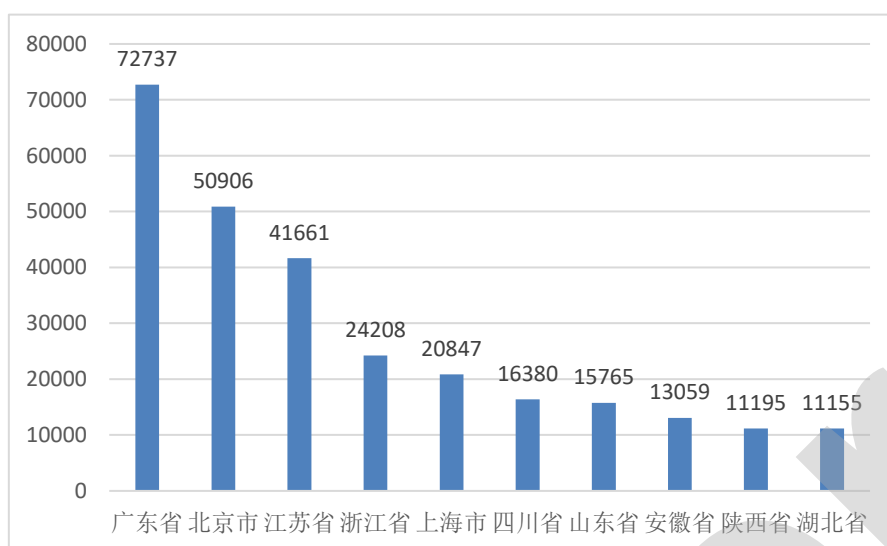


图 7-6 全国 AI 专利申请量 TOP 10 省份

7.2.3 申请人排名

中国 AI 专利申请数量排名前十的机构如图 7-7 所示，包括 5 家企业和 5 所高校，主要分布在广东、北京、浙江和四川。目前中国在 AI 专利领域的创新主要还是依靠高科技互联网企业和高校科研机构等方面的共同努力。国家电网专利申请量最多，其次是腾讯科技，体现出这两家企业在 AI 领域的创新能力比较突出，对相关技术领域的引领作用较强。

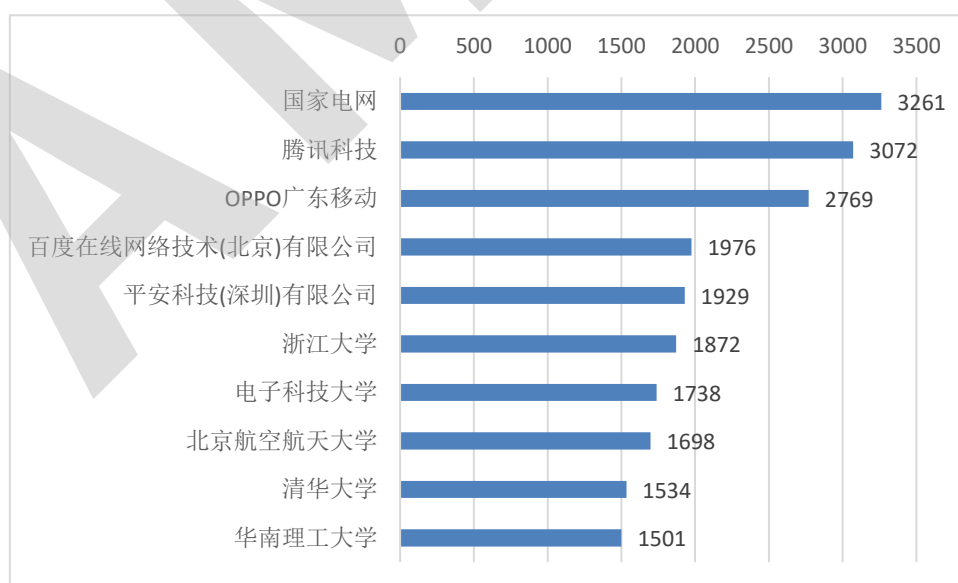


图 7-7 中国 AI 专利申请量 TOP 10 机构

8 人工智能技术应用：赋能其他行业发展

过去十年，人工智能技术与传统行业深度融合，广泛应用于交通、医疗、教育和工业等多个领域，在有效降低劳动成本、优化产品和服务、创造新市场和就业等方面为人类的生产和生活带来革命性的转变。本节将着重介绍人工智能如何赋能于医疗、交通、城市、金融和教育领域，以及该智能产业在未来所面临的机遇与挑战。

8.1 智慧医疗

医疗健康产业是经济系统中提供预防、诊断、治疗、康复和缓和性医疗商品和服务的部门的总称，通常包括医药工业、医药商业、医疗服务、保健品、健康保健服务等领域。

近年来，人工智能技术与医疗健康领域的融合不断加深，随着人工智能领域的自然语言处理、语音识别、计算机视觉等技术的逐渐成熟，人工智能的应用场景愈发丰富（如表 8-1 所示）。目前，智能医疗被广泛应用于电子病历、影像诊断、远程诊断、医疗机器人、新药研发和基因测序等场景，成为影响医疗行业发展，提升医疗服务水平的重要因素。

表 8-1 人工智能技术在医疗领域的应用

应用场景	部分应用的人工智能技术
电子病历	自然语言处理、语音识别
影像诊断	计算机视觉、计算机图形、深度学习
医疗机器人	机器人、人机交互
健康管理	数据库、信息检索与推荐
远程诊断	知识图谱、深度学习算法
新药研发	知识工程、数据挖掘
基因测序	机器学习、深度学习

（1）医疗影像智能诊断

医疗影像智能诊断技术利用计算机图形技术对图像数据进行深层次的分析，具体是通过 CT 扫描获得数字化的图像，然后将图像数据导入软件系统中进行分析，通过机器的“深度学习”建立模型。对建立的模型，使用一定数量的确诊病例来验证其是否可靠，然后用于检测其他的未知病例。

人工智能赋能的医疗影像智能诊断，一方面是通过计算机视觉技术对医学影像进行数据感知、智能化分析挖掘、快速阅片并获取有效信息和智能诊断，以辅助医师解读医学影像；另一方面，通过深度学习海量的影像数据和临床诊断数据，不断对模型进行训练，促使其掌握诊断能力，可以对医学影像实现诸如分类、检测、分割和配准等精准分析，从而帮助医生进行准确诊断。

这种人工智能和影像诊断相结合的技术大多应用于肿瘤诊断。乳腺癌是全球女性最普遍的癌症，也是女性第二大死亡原因。如果能够尽早发现和治疗可以改善患者的预后并大大改善存活率。DeepMind 和 Google Health 开发了一种能够帮助医生及早发现乳腺癌的 AI 系统。通过在美国和英国的女性患者的乳房 X 射线照片上进行算法训练，结果显示，该系统性能优于人类放射科医生。相关研究结果发表在《自然》杂志上。此外，清华大学知识工程研究室（KEG）、北京智源人工智能研究院、协和医院和首都医科大学的研究团队也联合推出一项名为 **Sono Breast** 的研究成果，提出了一个经过监督学习预训练的、基于图像块的卷积神经网络分类器，可以利用超声波图像进行乳腺癌诊断筛查。该工具包中包括一个恶性预测工具和一个分子亚型分析工具，可以进行乳腺癌的分子分型以及恶性肿瘤识别。该模型在乳腺癌分子分型的准确率达到 56.3%，F1 Score 是 45.8%。

人工智能驱动的医疗影像智能诊断还被应用在预测心脑血管方面。例如，北京富通东方科技开发的眼底荧光血管视频自动分析软件 Vessels，能够快速、自动发现动脉、静脉的变化，双眼臂循环时间异常等信息，借助知识图谱技术，能够预测心脑血管病变的可能性，为临床诊断提供依据。

在 2020 年的抗击 COVID-19 工作中，医疗影像智能诊断技术发挥了巨大的作用。在新冠疫情期间，AI 技术响应速度快、介入的力度大，帮助推出了 CT 辅

助诊断系统、个人自我诊断系统，并帮助医生快速熟悉了新冠的特点，了解和掌握了诊断和筛查标准，提升了诊断的信心。例如，北京推想科技公司开发的新型冠状病毒肺炎 AI 系统，因在世界权威医学杂志《柳叶刀》系列文章中被多次提及而受到关注^[45]。为了训练人工智能模型，该系统使用 NVIDIA V100 GPU 分析了来自中国早期疫情的 2000 多幅 CT 图像，帮助医生提高诊断速度，尽快向客户医院反馈诊断结果，降低疫情扩散的风险；为了构建 AI 驱动的医学成像和基因组学的应用程序框架，使用了 NVIDIA Clara；系统还可以帮助医生准确判断病人病情，为诊断和治疗提供依据，制定符合患者病症的精准诊疗方案，提高肺炎诊疗规范与质量。在遇到疑似新冠肺炎 CT 影像时，系统会立刻向医生预警。



图 8-1 新型冠状病毒肺炎 AI 系统^[46]

未来的医疗影像智能诊断发展面临着医疗数据共享、患者隐私等问题。与其他 AI 赋能行业的大数据相比，获取高质量的医疗影像数据相对困难。不同医疗机构的数据，目前还很少互通、共享。而单个医疗机构积累的数据往往不足以训练出有效的深度学习模型。此外，使用医疗影像数据进行人工智能算法的训练还涉及保护病人隐私等非技术问题。因此，在基于医疗影像的智能诊断方面，无论是数据还是算法，都有相当大的提高空间。

^[45] 柳叶刀 [EB/OL][https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(20\)30054-6/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30054-6/fulltext)

^[46] 北京推想科技公司官网 [EB/OL]<https://www.infervision.com/>

（2）新药研发

在新药研发过程中会有以下五个人工智能应用场景。（1）在制药工业中，应用人工智能技术访问大型数据集，系统地用于训练机器学习模型从而驱动数据集的预测属性，可以帮助研究者充分理解疾病机制，缩短靶点发现周期；（2）在药物发现中，临床候选分子必须满足一系列不同的标准，可应用机器学习技术，如支持向量机，随机森林或贝叶斯学习等进行化合物优化设计^[47]；（3）在临床前开发阶段，利用机器学习方法可以预测药物的分子特性、水溶性、毒性、口服吸收潜力等^[48]；（4）在临床研究阶段，将机器学习和认知计算等人工智能技术应用到研究设计、流程管理、数据统计分析等诸多方面，可全面提升临床试验的效率^[49]；（5）在审批与上市阶段，通过自然语言处理技术完成海量文献和大型数据集的信息综合和汇总，为新药研发人员持续提供药物研发情报的药物研发信息数据库^[50]。

在制药方面，全球十大制药公司（根据收入排名），瑞士诺华、瑞士罗氏、美国辉瑞、德国默克、瑞典阿斯利康、英国葛兰素史克、法国赛诺菲、美国艾伯维、美国百时美施贵宝和美国强生纷纷涉足人工智能领域，先后开展合作或收购项目，具体包括药物试验、药物发现和病人分析等。在药物研发数据库方面，汤森路透旗下的汤森路透知识产权与科技事业部开发的 Thomson Reuters Integrity, PJB Publications LTD 旗下的 Pharmaprojects, Venture Valuation VV AG (Swiss) 开发的 biotechgate 全球性生命科学数据库最为知名。药智、咸达、丁香园、米内、医药魔方、医药地理等国内企业建立的药品研发、生产检验、合理用药、市场信息综合数据库较为知名。

新药研发是一项漫长而且低效率的工作。制药网数据显示，所有进入临床试验阶段的药物，只有不到 12% 的药品最终能够上市销售，而且一款新药的平均研

^[47] Cyclica launches ligand express™, a disruptive cloud-based platform to revolutionize drug discovery [EB/OL]. <https://cyclicarx.com/news/cyclica-launches-ligand-express>.

^[48] Jing Y, Bian Y, Hu Z, et al. Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era[J]. The AAPS journal, 2018, 20(3): 58.

^[49] Mervin L H, Afzal A M, Drakakis G, et al. Target prediction utilising negative bioactivity data covering large chemical space[J]. Journal of cheminformatics, 2015, 7(1): 1-16.

^[50] 金小桃.健康医疗大数据[M].北京:人民卫生出版社, 2018:159-160.

发成本高达 26 亿美金^[51]。人工智能应用于新药研发可以大大缩短筛选候选药物分子的时间，节约研发成本。但是目前人工智能辅助新药开发的工作仍处于验证阶段，让人工智能技术驱动对接整个药物研发过程，实现真正的普适 AI 还有很长的一段路要走，其面临的挑战主要表现在研发结果不确定。人工智能模型基于数据学习，而非因果/规则推理。药物研发不确定性大、试错成本高、周期长，因此国内涉足企业较少。新药研发是一个复杂的系统工程。双重的不确定性导致人工智能在新药研发的各个阶段可能增加新药研发结果的不确定性。

人工智能应用于新药研发还存在算法监管问题。监管机构需要清楚地知道药物开发中使用的算法，从而理解人工智能主导的决策背后的逻辑。如果不对监管实现算法透明化，人工智能将会是一个无法进行严谨的科学评价及验证的“黑匣子”。这可能会导致在药物审批过程中出现种种无法预料的问题，比如监管对人工智能“发现”的生物标记物的接受度不明。为了避免上述问题，监管方应该和药企积极商讨、出台更加切实有效的监管途径。

（3）基因测序

基因组测序是一种实验室过程，用于测定一个人完整 DNA/RNA 序列中几乎所有的核苷酸，包括非编码序列^[52]。对 DNA/RNA 进行测序意味着确定构成 DNA/RNA 分子的四个化学结构单元（称为“碱基”）的顺序。基因测序所获得的序列，能够告诉科学家们特定 DNA/RNA 片段里所携带的遗传信息的种类。

在基因测序领域，第一代和第二代测序方法均需要算法聚类相似 DNA 片段并按重复片段重建。在这一点上，人工智能技术可以有效地提高聚类的精确度和速度。第三代测序方法则需要 AI 学习分析 DNA 单分子的电信号或者其他表现特征来重建核酸序列。

例如，Google—DeepVariant^[53] 作为一种深度学习技术，可以从多种不同的

^[51] 制药网 [EB/OL]<http://www.zyzhan.com/news/detail/63844.html>

^[52] NIH <https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/whole-genome-sequencing>

^[53] 谷歌 AI blog, 2017/12/04 <https://ai.googleblog.com/2017/12/deepvariant-highly-accurate-genomes.html>

高通量（HTS）测序获得的数据重建真实的基因组序列。即使测序质量不佳或者来自非人类样本，该测序技术结果的准确性明显高于以前的经典方法。DeepVariant 是一个分析管道，它使用深度神经网络从下一代 DNA 测序数据中调用遗传变异。DeepVariant 依赖 Nucleus，这是一个 Python 和 C++ 代码库，用于以通用基因组文件格式（例如 SAM 和 VCF）读写数据，旨在与 TensorFlow 机器学习框架轻松集成^[54]。

人工智能在基因检测与识别上，主要应用于异常基因发现。异常基因可能导致癌症或者遗传病和罕见病的出现。Google 和 Illumina 都推出了基于 AI 的异常基因发现平台。例如，Illumina—SpliceAI^{[55][56]}一种开源的、基于深度学习的工具，包含 32 层的深层神经网络，用于识别剪接异常基因。SpliceAI 可以使用 RNA 测序数据，在患有罕见遗传病的患者中发现以前被忽视的非编码突变，目前已经在实验中发现了自闭症和智力残疾患者以前被忽视的非编码突变。

在单细胞测序层面，人工智能技术已经深度参与了分子级别基因药物发现和设计，能够学习模拟细胞内化合物性质，在分子层面发现药物和精准设计针对异常基因的靶向药。例如，Deep genomics—Project Saturn^[57]是一个围绕分子生物学 AI 构建的药物设计系统，能够快速发现和开发基因药物。Project Saturn 能模拟和评估针对 100 万个目标的超过 690 亿个寡核苷酸分子，从而生成含有超过 1000 种化合物的库。这些化合物经过实验验证可按预期操纵细胞。

总之，未来 AI 医疗领域的目标是达到**一键诊断、居家诊疗**，但是其实现主要存在两方面的困难与挑战。一是**海量数据的处理存在困难**，需要 AI 技术人员与医生一起发现、抓取特征数据。二是**隐私和伦理问题**。AI+医疗不仅是一个技术问题，更是一个涉及人身安全的社会学问题。这不仅需要技术公司保障医疗数据安全、注意可解释性，而且需要第三方监管机构的监督。

⁵⁴ DeepVariant 的 GitHub 网页 <https://github.com/google/deepvariant>

⁵⁵ Illumina 的 SpliceAI 平台 GitHub 网页 <https://github.com/Illumina/SpliceAI>

⁵⁶ K. Jaganathan, 2019 [https://www.cell.com/cell/pdf/S0092-8674\(18\)31629-5.pdf](https://www.cell.com/cell/pdf/S0092-8674(18)31629-5.pdf)

⁵⁷ Deepgenomics 官网, 2020 <https://www.deepgenomics.com/project-saturn/>

8.2 智慧金融

智能金融是人工智能技术与金融业深度融合的新业态，是用机器代替和超越人类部分经营管理经验与能力的金融模式变革^[58]。人工智能正逐步成为决定金融业沟通客户、发现客户金融需求、辅助金融决策的重要因素。

人工智能技术在金融业中可以用于服务客户，支持授信、各类金融交易和金融分析中的决策，并用于风险防控和监督，将大幅改变金融现有格局，金融服务将会更加地个性化与智能化。智能金融对于金融机构的业务部门来说，可以帮助获客、精准服务客户和提高效率；对于金融机构的风控部门来说，可以提高风险控制，增加安全性；对于用户来说，可以实现资产优化配置，体验到金融机构更加完美的服务。

目前，人工智能在金融行业中的应用场景及相关的技术如表 8-2 所示，主要有**智能获客**，依托大数据以及对金融用户进行画像，通过需求响应模型，极大地提升获客效率；**身份识别**，以人工智能为内核，通过人脸识别、声纹识别、指静脉识别等生物识别手段，再加上各类票据、身份证、银行卡等证件票据的 OCR 识别等技术手段，对用户身份进行验证，大幅降低核验成本，有助于提高安全性；**大数据风控**，通过大数据、算力、算法的结合，搭建反欺诈、信用风险等模型，多维度控制金融机构的信用风险和操作风险，同时避免资产损失；**智能投顾**，基于大数据和算法能力，对用户与资产信息进行标签化，精准匹配用户与资产；**智能客服**，基于自然语言处理能力和语音识别能力，拓展客服领域的深度和广度，大幅降低服务成本，提升服务体验。

表 8-2 人工智能技术在金融领域的应用

应用场景	部分应用的人工智能技术
智能获客	机器学习、数据挖掘、信息检索与推荐
身份识别	计算机视觉、语音识别、安全与隐私
智能风控	机器学习、安全与隐私、数据挖掘、深度神经网络

^[58] 、中国金融新闻网[EB/OL].https://www.financialnews.com.cn/kj/jrcx/201912/t20191222_173871.html

应用场景	部分应用的人工智能技术
智能投顾	信息检索与推荐、数据挖掘
智能客服	人机交互、自然语言处理、语音识别
移动支付	安全与隐私、计算机视觉、语音识别、计算机网络、物联网、数据挖掘
业务流程优化	物联网、数据挖掘、机器学习

（1）智能风控

人工智能在金融行业的另一个重要应用是风险控制智能化,比如反欺诈交易。通过机器学习用户的行为模式,结合欺诈黑名单等数据支撑,反欺诈模型能够辨别特殊用户在贷款申请流程、审批流程中的反常行为,从而做出欺诈行为判断。此外,对于恶意套现、盗刷卡、垃圾注册、营销作弊、虚假交易等行为,也可以由机器从海量的交易数据中学习知识和规则、发现异常、及时干预阻断,打造安全、可靠的金融环境。

评判一个用户数据信息的传统方式是通过收入证明、社保、央行征信等金融强变量信息。人工智能赋能的智能风控则是通过若干个金融弱变量信息,包括地理位置信息、与人交往信息、阅读习惯信息、消费行为信息等做出模型,依此决定是否放贷。过去,一笔传统个人贷款业务可能需要 2~3 天审批,而一个基于人工智能模型的自动审批方案只需几秒钟就能完成。例如:某金融科技公司的智能风控平台拥有超过 7,500 个风控模型数据变量,可以对 98% 的订单全自动审核,秒级反馈结果;平台中的小微金融资产处理技术平台“虫洞”则可直连金融机构,根据用户属性、订单类型及资产属性,自动对资产进行分级、定价,再按照不同资金方提交的资产要求,将这些消费金融资产实时推荐给金融机构,由他们完成审核实时放款。

（2）智慧银行

智慧银行^[59]是传统银行、网络银行的高级阶段,是银行企业以智慧化手段和

⁵⁹ 工行加快推进智慧银行建设. 央广网.2018-04-27,

新的思维模式来审视自身需求，并利用创新科技塑造新服务、新产品、新的运营和业务模式，实现规模经济，提升效率和降低成本，达到有效的客户管理和高效的营销绩效的目的。

人工智能技术应用正在深度改变着银行的服务交互方式、数据处理方式、组织架构模式，以及金融行业的参与者构成。AI 构成的智慧银行将是传统银行转型的必由之路。

人工智能可以通过用户画像和大数据模型为银行等金融公司精准找到用户，实现精准营销。同时，通过智能客服机器人挖掘客户潜在需求，还可以提升销售转化率、客服效率和用户体验，极大地降低人力成本。银行通过将语音识别和自然语言理解技术集成于客服系统，通过机器人完成客服系统的语音导航、语音交互、语音咨询功能，实现了客服系统“自助+智能+人工”三层的服务模式，全天候提供服务，在提升客户体验的同时，也大大降低了银行的运营成本。

在银行的前端即客户端，人工智能接入成为智慧银行与客户交流交互带来全新的界面，降低银行与客户交互成本，提高交互效率。主要体现在以下五个方面。一是**利用信息检索和推荐技术进行智慧营销**。通过融入生活（生产）场景以及智能全渠道建设，实现一贯式线上线下智慧体验；银行可利用大数据平台，对用户信息进行分析，为客户进行客户定位和建议推荐。这个过程中，推荐引擎及用到的技术（图分析、协同过滤、社交图谱）将帮助银行利用多种信息数据，提高个性化推送能力。二是**贯穿售前、售中、售后整个服务周期的智慧客户体验分析**。三是**基于人脸识别、智能问答库等技术的智能自助信贷审核系统**。贷中审批环节，银行可以将金融科技、模型以及策略有机的相结合，标准化、智能化审批流程，形成智能审批能力，从而大幅提高审批的效率。其中，微表情识别功能构建了先进模型，而智能问答库和微表情识别功能匹配了应对策略。四是**基于情感分析、语音识别、问答系统、自然语言处理技术的客服系统**。针对售后服务、问题查询等交流性环节，通过智能客服、远程可视呼叫等功能提升沟通能力，增强交叉销售机会。五是**基于情感分析、语音识别等技术的智能催收**。在催收管理期，通过失联修复、催收评分、催收策略、催收队列安排以及催收机器人等实现智能催收，

可以大幅降低催收成本并有效提高催收效率和客户体验。

在银行后端，人工智能系统能够为银行从业人员提供信息支持和操作辅助，在不牺牲安全性的情况下提高协作效率和生产力。主要表现为：一是**客户画像/客户识别**，并将对客户进行细分，通过“感知客户”与“客户感知”体系重构银行与客户的关系。二是**智慧风控**。智慧风控运用于信贷风控、反欺诈、反洗钱、交易监控、催收等场景，已成为智慧银行转型发展及经营绩效提升的助推器。三是**智慧反欺诈**。通过知识图谱、机器学习、数据挖掘等技术，智慧银行可以在反欺诈规则上，构建反欺诈模型，使用图形图像识别、声纹识别等技术，并利用行内积累的非结构化数据提炼交叉验证规则，进一步提升反欺诈能力。知识图谱技术可以整合客户的各种大数据，形成关系网络。该技术可通过无监督学习算法，挖掘诈骗团伙的特点，通过逻辑不一致性识别、发现诈骗团伙。而推进设备指纹技术，辅助以位置验证，识别客户的身份并基于实时用户信息验证电子设备。四是**运营管理云平台**（流程自动化机器人）。按照各运营活动内在逻辑，在线上实现运营活动的虚拟集中。在智慧运营的体系下，通过云平台将运营资源进行整合和分配。此外，还可以借助流程自动化机器人对流程进行优化，打造足以支撑智能化前端业务的后台运营能力。

智慧银行的实现面临着两方面的挑战，分别为技术和运营^[60]。技术上，智慧银行需要拥有足够速度、精确和多面性的金融科技。然而，目前的银行数据系统（如支付和贷款系统）不得不维护大量遗留系统，这些系统缺乏 AI 所需的多样计算、数据处理和实时分析需要的容量和可塑性。许多银行的数据储存系统都是数据孤岛，并且分析工作多基于单个数据，没有中心化的数据库。

（3）智慧投顾

智能投顾，是指运用人工智能机器学习算法、大数据、以及量化金融模型，根据投资者的风险偏好及承受水平、预期收益目标及所处生命周期，基于投资组合优化等理论模型，运用一系列智能算法，为客户提供个人资产分析、风险偏好

⁶⁰ AI-bank of the future: Can banks meet the AI challenge? 麦肯锡，2020 年 9 月，
<https://www.mckinsey.com/industries/financial-services/our-insights/ai-bank-of-the-future-can-banks-meet-the-ai-challenge>

分析、智能化投资建议、资产配置、税务与财务管理及其它增值服务的在线投顾服务模式，同时还实时监测市场动态，对客户的资产配置组合进行自动跟踪和再平衡。

智能投顾的理论依据是基于以 Markowitz 均值-方差模型、Black-Litterman 模型为代表的现代投资组合理论模型。随着大数据、云计算、人工智能等金融科技创新发展的推动下，智能投顾随之产生并已获得了较广泛应用，未来前景看好。

“智能算法”是智能投顾的核心，也是智能投顾平台连接客户端和金融机构端的最重要工具。智能投顾通常通过大数据获得客户个性化风险偏好、投资目标及其变化规律；然后，根据客户服务的个性化风险偏好，通过智能算法模型，定制出合理的个性化资产配置方案；此外，利用互联网平台对客户的个性化资产配置方案进行动态实时的跟踪、调整与更新。

智能投顾是在人工智能时代产生的新型金融服务模式，需要人与机器的深度融合。相较于传统投顾更具客观性，智能投顾在一定程度上降低了因业绩而误导客户的风险。值得注意的是，智能算法技术存在恶意代码、操作系统风险等因素，会影响到投资者的账户资金安全，蕴含着多种新型交易风险，构成了对投资者的潜在威胁。目前，我国智能投顾尚无明确的行业标准和监管意见。

目前，“AI+金融”的发展正处于井喷爆发的前期，基础硬件、数据、算法、应用到人才等各个产业生态环节需要在逐步建立和完善的行业监督标准和政府监管政策的指导下，形成软硬一体化的通用构建方法和标准流程，加速“AI+金融”整个产业链及生态的覆盖，赋能普惠金融。

在今后的发展过程中，数据的可解释性、系统的风险性和隐私保护是智慧金融的三大核心问题。金融机构需要通过提升透明度等方式，建立一种平衡的知情信任来提升数据的解释性。为应对系统风险性问题，金融机构应通过强化关键市场信号的捕捉与报告、采用可交互式的人工智能决策引擎、采用熔断工具等方式，降低系统性风险。毕竟当金融机构均采用人工智能进行决策时，其市场信号解读可能趋同与不断强化，导致形成偏离正常市场规律的结果。而这些不正常的市场变化会成为人工智能的学习基础，将人工智能的决策逻辑进一步畸化，造成恶劣后果。

隐私保护也是智慧金融面临的严峻问题。近年来，每年发生金融隐私泄露事件大约以 35% 的速度在增长，有公开报道或记录 2016 年 1093 起，2017 年 1511 起，2018 年 1967 起，2019 年 2300 余起^[61]。加之近年来人工智能技术在金融行业的广泛应用，由此带来的金融隐私（银行数据、保险数据、其他平台金融数据、网贷业务及大数据）保护问题日益凸显，在央行获批个人征信牌照的机构仅有百行征信 1 家，远远满足不了民间网贷的需求，建立完善的金融隐私保护监管体系已刻不容缓。

8.3 智慧城市

智慧城市（Smart city）又称为数字城市，最早由 IBM 公司于 2010 年提出，即“城市由关系到城市主要功能的不同类型的网络、基础设施和环境六个核心系统组成：组织（人）、业务/政务、交通、通讯、水和能源。智慧的城市系统不是零散的，而是以一种协作的方式相互衔接所组成的宏观系统。”

随着人工智能等前沿技术的融入，城市基础设施得到了创新升级，全方位助力城市向智慧化方向发展，包括政府服务、地理系统、物流系统、循环系统和能源系统等各个领域。但智慧城市的首要任务还是优化城市功能，推动经济增长，同时利用智能技术和数据分析提高市民的生活质量。

运用大数据、云计算、区块链、人工智能等前沿技术推动城市管理手段、管理模式、管理理念创新，从数字化到智能化再到智慧化，让城市更聪明一些、更智慧一些，是推动城市治理体系和治理能力现代化的必由之路。典型应用场景及相关技术如表 8-3 所示。

表 8-3 人工智能技术在智慧城市领域的应用

应用场景	部分应用的人工智能技术
智能政务	自然语言处理、自然语言理解、机器学习、物联网、智能语音、图像识别
城市指挥中心	机器学习、数据挖掘、物联网、可视化

^[61] 金融隐私保护问题分析及对策，2020-05-25，万方科技[EB/OL].
<http://www.wfnetworks.cn/news1/shownews.php?id=482>

应用场景	部分应用的人工智能技术
城市公共安全	机器学习、数据挖掘、物联网、安全与隐私
智慧物流及建筑服务系统	机器学习、物联网、安全与隐私
智慧能源系统	机器学习、强化学习、深度学习、物联网、知识图谱、卷积神经网络
智能交通系统	机器学习、数据挖掘、计算机视觉、自然语言处理、自然语言理解、图像识别、智能芯片
城市环境管理	机器学习、物联网、可视化
智能家居	自然语言处理、自然语言理解、安全与隐私、人机交互、语音识别、计算机视觉
智能医疗系统	自然语言处理、自然语言理解、机器学习、安全与隐私、知识图谱
自动驾驶	机器学习、计算机视觉、人机交互

(1) 智能政务

智能政务是指利用云计算、物联网、互联网等先进技术，通过检测、分析、整合、智能响应，综合各职能部门，对现有各种资源进行信息高度整合，提高政府的业务办理和管理效率，同时加强监管，强化政务透明度，提供更好的服务、绿色的环境、和谐的社会，保证小镇可持续发展，为企业及公众建立一个良好的工作、生活和休闲环境。作为智慧城市建设生态系统中的重要组成部分，在人工智能技术的推动下，正朝着更具人性化与针对性的方向发展^[62]。一方面面向居民与企业的公共服务将更加符合人的习惯，而非现阶段单纯依靠在线接口；另一方面人工智能的决策将更加有效，精度大幅提升，处置方案更加灵活。

智慧政务的优势在于，能够以“互联网”的思维充分有效地利用政务大数据来提高应用成效。基于互联网，围绕政府管理、社会保障、公共安全、社会信用、市场监管、食品药品安全、医疗卫生、国民教育、劳动就业、养老服务、公共交通等应用服务，建设互联网政务服务平台。实现政府各部门的审批、核准、备案、证明等行政权力网上运行，企业和个人的在线申请、政府在线办理，全面实现网

^[62] Lv Z, Li X, Wang W, et al. Government affairs service platform for smart city[J]. Future Generation Computer Systems, 2018, 81: 443-451.

上政务公开和互动交流。

以移动客户端为接入口，搭建移动端的网上行政服务大厅，为市民提供信息查询、在线预约、在线办理等公共服务；为政府实现移动办公、舆情管理、在线审批、信息公开、应急预案、交通管理等。依托统一的智慧政务平台，不断创新各类便民事项的“互联网政务”应用。

政务服务是智能政务中最为核心也是推进速度最快的领域之一。中国各地政府通过建设一站式服务平台积极推进政务智慧化。“京京”是北京市人民政府门户网站（“首都之窗”）在线咨询服务智能机器人，服务内容覆盖政务咨询、建议、投诉、举报、新冠疫情防控等各个方面。交互形态上，涵盖文本和语音模式；渠道上，支持微信、App、小程序、PC 的极速接入，可为市民提供 7×24 小时的在线咨询服务（图 8-2）。此外，其他具有代表性的智能政府还包括深圳公安局将传统的窗口“面对面”排队向网上办理转变，“刷脸”就可以进行户政办理，同时基本建成全市统一的政务信息资源共享体系，汇集 29 家单位的 385 类信息资源、38 亿多条数据^[63]，为政务服务全面智能化提供数据支持。杭州构建一体化的智能电子政务管理体系，数字城管、规划系统、财政系统业务系统在电子政务外网得到整合，并提供一站式服务等。



^[63] 从深圳政务信息资源共享实践成效充分认识政务信息资源共享的基础性作用，2016-09-22，新华网[EB/OL].<http://www.0745news.cn/2016/0922/974615.shtml>

图 8-2 北京市政府在线咨询服务智能机器人“京京”^[64]

人工智能技术的出现，为智能政务业务协同发展过程中面临的政府管理、社会保障、公共安全、社会信用、市场监管等信息传递以及政务服务场所的问题提供了新的解决方案。然而，城市的条块化管理，使得政务服务部门难以从全局识别流程的各要素，给业务流程构建和优化带来了新的挑战。现阶段，面向智能政务的业务流程在协同方面主要包括城市服务流程角色分配问题、跨区级信息共享问题、信息安全问题等。

（2）智能基础设施系统

城市是人工智能应用场景最终落地的综合载体。随着 AI 等前沿技术的融入，城市基础设施得到了创新升级，将全方位助力城市向智慧化方向发展。

人工智能是建设未来智慧城市的重要技术之一，为了解决目前城市发展中遇到的人口密集、能源结构单一、资源配送效率低、交通物流风险大、垃圾回收利用率低、空气质量不佳等一系列痛点，引领城市实现健康舒适、碳排放不断减少、具备高安全性、生活高度便利化等美好愿景，可将人工智能技术应用于城市基础设施系统，通过在城市各大综合载体中先行构建，最终推广至整个城市层面，最终构建完善的智慧城市基础设施系统。人工智能在城市基础设施系统中的应用主要可分为城市绿地系统、交通系统、物流系统、循环系统、能源系统等创新系统。

绿地系统（水治理公园）是指通过打造城市智慧水资源管理体系，以实时应对全球气候变化对城市水环境造成的影响，最终实现节约循环、环境适应、安全使用的目标。在实施过程中，先行在城市综合载体开展海绵城市和雨水花园技术的应用，构建城市智慧综合管廊，通过结合形成更加成熟的海绵城市运营管理模式。同时，可将绿地系统与污水与废物垃圾处理系统、能源系统有机结合，贯彻实施循环经济的理念。远期可进一步推广至整个城市层面，通过各种措施，确保城市能够应对全球变暖带来的水环境危机。

^[64] 北京市人民政府门户网站[EB/OL]. <http://www.beijing.gov.cn/jj/>

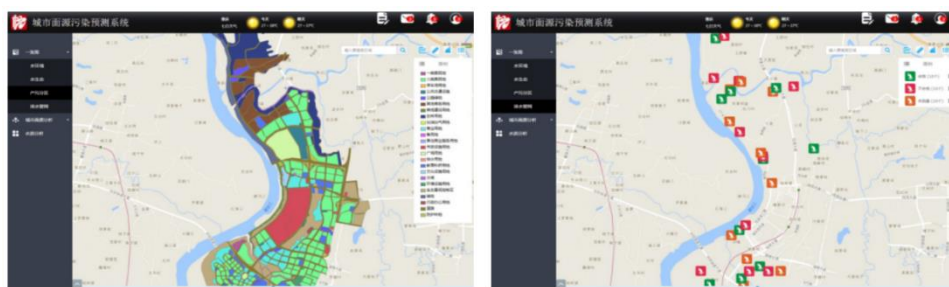


图 8-3 北京某公司推出的智能水环境管控平台^[65]

杭州是在发掘和探索智慧化城市建设的创新理念方面较为突出。城市大脑正在成为杭州城市治理新的基础设施，持续提升城市治理体系和治理能力现代化水平。杭州城市大脑由中枢、系统平台、数字驾驶舱、场景四要素组成，通过全社会的数据互通，数字化的全面协同，跨部门的流程再造，实现民生直达、惠企直达、基层治理直达。杭州城市大脑建设牢牢抓住便民服务和治理能力提升两大关键，经历了三个阶段，“数字治堵”、“数字治城”、“数字治疫”。

（3）智能交通

人工智能时代，交通系统和汽车相关的智能出行正在被重新定义，出行的三大元素“人”“车”“路”被赋予类人的决策行为，整个出行生态也会发生巨大的改变。随着人工智能技术在交通领域的应用朝着智能化、电动化和共享化的方向发展，以自动驾驶为核心的智能交通产业链正在逐步形成。

在智能交通系统方面，借助人工智能技术，将核心交通元素联通，实现信息互通与共享以及各交通元素的彼此协调、优化配置和高效使用，形成人、车和交通的一个高效协同环境，建立安全、高效、便捷和低碳的交通。例如交通信号灯智能适配、出行路径优化、公共交通工具调度决策等应用场景。

在自动驾驶方面，通过车上搭载传感器，感知周围环境，通过算法的模型识别和计算，辅助汽车电子控制单元或直接辅助驾驶员做出决策，从而让汽车行驶更加智能化，提升汽车驾驶的安全性和舒适性。例如驾驶辅助、自动驾驶、智能导航等应用场景。

智能交通系统（Intelligent Traffic System，ITS）是通信、信息和控制技术在

^[65] 智能水环境管控平台[EB/OL].http://www.sinfotek.com/solution_info.php?num=1000436

交通系统中集成应用的产物。ITS 借助例如通过交通信息采集系统采集道路中的车辆流量、行车速度等信息，信息分析处理系统处理后形成实时路况，决策系统据此调整道路红绿灯时长，调整可变车道或潮汐车道的通行方向等，通过信息发布系统将路况推送到导航软件和广播中，让人们合理规划行驶路线。通过不停车收费系统（ETC），实现对通过 ETC 入口站的车辆身份及信息自动采集、处理、收费和放行，有效提高通行能力、简化收费管理、降低环境污染。

随着我国经济的高速发展、城镇化进程的不断加快，城市交通压力不断增大，人工智能技术在缓解交通压力、交通信号灯智能化、路径优化、智能导航等方面都做出了突出贡献。从智能交通的商业潜力来看，人工智能有效地降低了人和物的运输成本。从智能交通的发展前景来看，掌握用户数据等于掌握智能交通的未来，其核心趋势是在于对数据和平台资源的整合。

在未来智慧城市中，人和城市的需求将作为出发点，积极响应未来人工智能技术的发展，搭建各类智能交通应用场景，创造更加高效、活力、可持续的智能交通出行体系，满足人性化出行。

在具体实施层面，从人工智能在交通领域的技术路径来看，近期可在综合载体中引入新型道路分级系统，改革路网密度和道路断面，同时引入车路协同技术、智慧停车技术构建智慧场景，小范围开展无人共享出行、无人出租车系统。



图 8-4 某地政府部门推出的智慧交通平台^[66]

在赋能智能城市应用、构建智能城市新生态领域，京东数科研发了智能城市

^[66] 智慧交通平台[EB/OL].<http://www.univalsoft.com/6643.html>

操作系统，包括城市全域数据感知引擎、城市时空数据引擎-JUST、城市时空智能引擎-FAST、联邦数字网关、城市数据治理引擎、京东城市视频大数据引擎和京东城市莫奈可视化引擎^[67]，如图 8-5 所示。该系统可以在数据不出库的情况下，实现知识共享，打通数据壁垒；根据经验归纳差异化应用背后的公共算法，支持第三方快速开发行业应用；与城市交通、能源、规划等领域的伙伴合作，提供智能化解决方案；利用六个时空数据模型收纳城市万千数据，扩展性强。大兴机场智能停车系统、雄安新区数据平台、南通雪亮工程和财政厅大数据融合中心目前都以该智能城市操作系统作为支撑。



图 8-5 京东“智能城市操作系统”综合解决方案^[68]

物流系统（自动化流转）：自动化流转的物流系统以低成本、高效化和安全化为目标，通过自动化、智能化手段解决物流“最后一公里”难题，提升物流配送效率，降低成本，并显著改善街道交通环境。

在具体操作层面，短期内可在城市综合载体内预留地下物流系统的管道，安装无线充电基础设施。待配套技术成熟后，做小范围试运行。中远期与现有的物流企业合作，开发机器人分发系统，利用现有技术资源，研究突破地下物流系统管网的技术，推广至整个城市。

[67] 京东数科[EB/OL]. <https://www.jddglobal.com/products/operating>

[68] 京东数科[EB/OL]. http://www.xinhuanet.com/tech/2019-03/22/c_1124268811.htm



图 8-6 某物流公司的物流机器人正在高效工作^[69]

自动驾驶汽车（Automated Vehicle），是一种通过计算机实现自动驾驶的智能汽车。根据智能化水平的不同，可由低到高细分为辅助驾驶（Level 1）、部分自动化（Level 2）、有条件自动化（Level 3）、高度自动化（Level 4）和完全自动化（Level 5）5 个级别^[70]。

自动驾驶的核心依靠感知探测一定范围内的障碍物，并依据已设置好的路线规划实施驾驶行为，各式车载雷达、传感器、辅助驾驶系统和高精地图可以实现驾驶、车和路的交互与融合。车载雷达可探测路肩、车辆、行人等的方位、距离及移动速度，视觉传感器用来识别车道线、停止线、交通信号灯、交通标志牌、行人及车辆等信息，定位传感器用来实时获取经纬度坐标、速度、加速度、航向角等高精度定位，车身传感器通过整车网络接口获取诸如车速、轮速档位等车辆本身的信息。高级辅助驾驶系统（ADAS）实时收集车内外的环境数据以及时察觉潜在危险，高精度地图实现地图匹配、辅助环境感知、路径规划的作用。

目前，自动驾驶发展面临一些挑战。一是该成熟度未达预期。由于技术尚未成熟、自动驾驶汽车规模量产还需时日，自动驾驶距离 Level 4、Level 5 级别的落地仍有非常长的路程要迈过。二是精确定位和导航精度有待提高。目前，我国自主研发的北斗导航系统在性能上与美国的 GPS 相比还存在一定的差距。高精度激光雷达、毫米波雷达等车载传感器作为智能交通的眼睛，国内近几年虽有快速的进步，但与国外先进水平相比仍有相当的差距。三是相关政策法规有待完善。

^[69] 物流网[EB/OL].<http://www.hhhtckd.com/wlzx/39847.html>

^[70] Standard S A E. J3016[J]. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems, 2014, 4: 593-598.

目前并没有哪些国家对自动驾驶汽车相关事故责任等问题专门制定完善的政策法规。四是数据可靠性有待加强。如何打造安全可靠的数据链路，不被黑客侵扰等也将是需要长期面临的困难与挑战。

8.4 智慧教育

智能教育（Artificial Intelligence in Education, AIED）又常被称为“AI+教育”，是指将人工智能和教育科学结合在一起，以发展适应性学习情境，它通过开发和使用具有灵活性、包容性、参与性和个性化的智能工具来赋能教育。

智能教育的建设涵盖校园 IT 基础建设、互动教学硬件设备、信息化平台及软件、线上内容资源，通过软件、硬件和内容的整合构成了智能教育行业——在线教育、智慧校园和智慧课题^[71]。

智能教育建立在与学生充分的交互和数据获取的基础上，并在海量的教育数据中，匹配用户的学习需求，然后进行统计分析和评估反馈，可应用于教学过程中的“教、学、评、测、练”五大环节。通过数据采集分析，智能教育能够实现教学从数字化到数据化，为老师减负增效、提供更丰富的教学资源；实现对学生个性化分析、以学定教、实时反馈和自动化辅导、提升学习的效率与质量；能够为教学管理提供大数据辅助决策与建议，为科学治理提供支撑。

人工智能赋能教育领域的技术主要包括语音语义识别、图像识别、知识图谱和深度学习等。语音语义识别、图像识别实现了规模化的自动批改和个性化反馈；知识图谱和深度学习技术搜集学生学习数据并完成自动化辅导和答疑，智能化推荐适合学生的学习内容，高效、显著地提升学习效果。

人工智能技术在智能教育中的应用场景包括：智适应学习、教育机器人、智慧校园、智能课堂等，如表 8-4 所示。相关应用可分为三大类：一是间接辅助型，即利用 AI 科技手段辅助教学提高学生成绩、效率及自主学习能力，对应智能教育中的“在线教育”和“智慧课堂”；二是直接学习型，即将科技项目作为教学内容；三是基础运营型，即利用科技来提升学校管理能力、教师教学质量^[71]，对应智能教育中的“智慧校园”。

⁷¹ 前瞻产业研究院，《中国智能教育行业发展前景预测与投资战略规划分析报告》

表 8-4 人工智能技术在教育领域的应用

应用场景	部分应用的人工智能技术
智适应学习	机器学习、计算机视觉、语音识别、人机交互
智慧校园	语音识别、自然语言处理、知识工程
智慧课堂	计算机视觉、计算机图形学、自然语言处理、语音识别、信息检索与推荐、人机交互、可视化、机器人
智能题库	计算机视觉、计算机图形学、信息检索与推荐
语音测评	语音识别、自然语言处理、知识工程
人机对话	机器学习、人机交互、语音识别、自然语言处理、机器人
教学辅助	自然语言处理、计算机视觉、语音识别

（1）智慧校园

智慧校园作为智能教育的基础，主要包括宽带校园、平安校园、智联校园和智慧教室四个层面。宽带校园旨在实现校园宽带网络全覆盖，为各类教学应用提供高速管道。平安校园主要是为了实现校园的全方位安防监控，提供一个安全和谐的物理环境。智联校园是指基于物联网、人工智能和大数据等技术，实现校园园区各智能终端的全联接，可以有效提高学校、教师的管理效率，便利师生学习与生活，及时保障学生安全，还通过融合物联网平台、校园一卡通等管理系统采集学生的学习和生活数据，可为后续大数据分析运营提供数据来源。智慧教室是智慧校园的重要应用场景，主要建设“智慧课堂”硬件设施配置、标准化考场配置。目前，宽带校园普及完成率相对较高，现阶段行业风口集中在平安校园、智联校园和智慧教室的建设上。

（2）智慧课堂

智慧课堂主要包括实体互动课堂、VR/AR 课堂和全息课堂等。

实体互动课堂是指在传统课堂基础上，利用视频系统、语音系统、智能终端更高效地进行互动教学。以学生为中心的互动课堂，能更好地激发学生兴趣、提

升教学质量，同时使得家校学习形成闭环，能更好地促进教与学的质量提升和教育目标的实现。例如，联想智慧课堂系统将信息智能化技术与教学融合，以软硬件一体化的教育平板为核心建设智能化教学环境，使用云端一体化的智慧课堂提供教学资源 and 教学工具，开展交互式教学和基于人工智能阅卷的智能化教学评价，构建了以学习者为中心的交互式智慧课堂教学新模式。此外，互动课堂可以为教育资源不足地区提供优质教育保障，促进教学公平和均衡发展。

VR/AR 课堂和全息课堂则是通过技术提供强大的“沉浸式体验”优势，实现更为直观的教学体验，提供显示世界中难以实现或不可能实现的课程和培训，可以有效提高用户的知识感知度和保留度，有效提高教育质量；但由于技术难度、成本高等多方面原因，目前在国内应用相对较少。其中，VR/AR 课堂需要配置 VR 眼镜、AR 终端等设置，主要应用于早教、K12、职业培训和高校教育等。全息课堂需要配置全息屏、全息眼镜、全息 3D 投影等设备，具备虚实交互和虚实一体能力，目前国内应用处于起步阶段。

针对课堂教学中，学生的习惯养成、教师的管理精力不足等问题，基于人脸识别、语音识别、动作识别、情绪识别、眼球识别等技术构建的课堂辅助系统，可以做到课程全覆盖，既能对学生的考勤、课堂表现及专注度进行监测分析，也能对教师的教学质量进行智能评估，即时生成课堂评估报告并快速反馈，教师可实时调整授课内容和方式，保证了高质量、全交互的在线课堂体验，从而形成了课上有行为、行为有识别、识别有产出的完整监课闭环。

例如，以英国 CenturyTech 公司、好未来的 GodEye 为代表的监课系统从学生课堂行为出发，通过捕捉多维度的课堂数据评估学生的专注度和接受度，精准定位学生课堂异常学习状态，并为每位学生生成详尽的个性化课堂报告，助力教师掌握学生现阶段学习状态，大量节省了教师学情评价的时间；接入教学系统后的 GodEye 已将原来不足 10% 的课堂教学评估覆盖率提升至 100%，质检效率提升 50 倍^[72]。针对教师授课情况，北京学而思网校开发的“AI 老师监课系统”深度融合了语音识别和表情识别两大技术，从课堂表现的“亲和力”“清晰度”“流

^[72] 科技部新一代人工智能发展研究中心&罗兰贝格管理咨询公司. 智能教育创新应用发展报告[R]. 北京：科技部新一代人工智能发展研究中心，2019.

畅度”“互动”“重点”等几个维度，对教师授课情况进行实时数据化、可视化评估，教师可根据评分实时调整授课内容和方式，保证了教学质量^[73]。

（3）智适应教学

智适应教育强调针对学生具体情况和个体需求来提供个性化解决方案，提倡所有学习进度和教学方法都要针对每个学习者的需求进行优化。智能教学系统（Intelligent Teaching System, ITS）是 AI 在教育领域研究最早和应用最广泛的技术方向之一。系统基于 OCR 识别、自然语言处理、数据挖掘等技术构建。ITS 可提供智能推荐服务、学情分析服务、错因诊断以及决策支持服务。

人工智能智慧教学系统，不但可实现基础的课程音视频录制、直播、教学互动等功能，更能通过教室中的摄像头收集视频，通过 5G 网络将视频上传至 MEC 边缘云，部署在边缘云的人工智能服务器对视频进行分析收集数据，5G 的大带宽、低时延的特性可以使 AI 助教系统仿佛部署在本地一样快速反应^[74]。

当前在国际上影响较大的 ITS 是由孟菲斯大学 AI 研究所历经十五年研究开发的 AutoTutor 系统，它通过自然语言与学生对话进而帮助他们学习物理和计算机知识，还可通过识别面部表情和身体姿势自动跟踪学习者的认知和情绪，并以自适应方式对认知失衡和认知混乱进行干预以促进深度学习^[75]。其他知名的智适应教学系统还包括 kidaptive、AdaptedMind、I-Ready、area9、LINGVIST、Knewton 等。国内智适应产品代表有专注于学前教育至高中教育（Kindergarten Through Twelfth Grade, K12）领域智能个性化辅导的松鼠 AI，深耕初高中 1 对 1 数学的学吧课堂，拥有庞大语音数据库的英语流利说，留学考试辅导机构智课网等。

人工智能在教育领域的深度广泛应用，将推动教育理念、教学方式以及管理模式创新，重塑教育体系，但智能教育的未来发展面临着一些挑战。

一是人工智能技术和教育体系还需待进一步融合。教育涉及“教、学、评、

^[73] 亿欧·学而思网校发布“AI 老师”系列新产品，用人工智能重新定义课堂[EB/OL]. <http://www.iyiou.com/p/115881.html>

^[74] 华为，2018 智慧教育场景白皮书 P.9

^[75] The University of Memphis. AutoTutor[EB/OL]. <https://www.memphis.edu/iis/projects/autotutor.php>

测、练”五大环节，其复杂性大大增加了人工智能技术与教育的融合难度；现有人工智能技术的应用多为语音识别、图像识别等识别领域的应用，但在理解层面和创造层面的评测环节中仍未实现实质性突破。例如当前自然语言处理技术仅能对语法结构、拼写正误等方面进行识别，而在篇章结构、语言逻辑、情感表达等语义理解层面依然十分欠缺。因此，人工智能技术还需要与脑科学、神经科学、认知科学等学科进一步交叉融合，加强智慧教育开放创新平台建设，推动开放平台向公众开放智能教育关键共性技术，打造智能教育生态。

二是智能教育从“教”到“育”的转变。对于“育人”方面来说，人工智能不可能替代教师，此时教师的职责则应该从“教”转变到“育”以及情感的沟通上。因此，当前智能教育面临着如何让教师与人工智能系统之间合理协作，以最大限度地发挥二者的协作优势，共同为学生提供更有个性的、更有效的教学体验。

三是智能教育中的伦理、社会及安全问题。智能教育中对图像识别技术、语音语义识别等均有广泛应用。在图像识别技术上，出于对当事人隐私或者某些信息的保护，电视和网络会将文字或脸部像素化。但随着人工智能的发展，这种保护隐私的方法已经不再可靠。因此，伦理道德和安全保障是教育人工智能也必须面对的挑战之一。

8.5 智能制造

智能制造（Intelligent Manufacturing, IM）是基于新一代信息通信技术与先进制造技术深度融合，贯穿于设计、生产、管理、服务等制造活动的各个环节，具有自感知、自学习、自决策、自执行、自适应等功能新型生产方式，其本质是新一代信息技术与制造业的深度融合。

国际上，智能制造通常是指一种由智能机器人和人类专家共同组成的人机一体智能系统，其技术包括自动化、信息化、互联网和智能化四个层次^[76]。

我国将智能制造作为《中国制造 2025》的主攻方向；并且在规划中，指出智能制造是未来制造业发展的重大趋势和核心内容，也是解决我国制造业由大变强

⁷⁶前瞻产业研究院，2019 年中国智能制造发展现状及趋势分析报告

的根本路径。智能制造正在引领制造企业全流程的价值最大化。

从智能制造业角度出发，人工智能技术正在深入改造制造行业。新一代人工智能技术与制造业实体经济的深度融合，成为应用市场一大亮点，催生了智能装备、智能工厂、智能服务等应用场景，创造出自动化的一些新需求、新产业、新业态。典型应用场景及相关人工智能技术应用如表 8-5 所示。

表 8-5 人工智能技术在智能制造领域的应用

应用场景	部分应用的人工智能技术
智能工厂	自然语言处理、机器学习、物联网、人机交互、可视化
工程设计、工程工艺设计	机器学习、计算机视觉、计算机图形学、可视化
生产制造、CIMS 等	机器学习、计算机视觉、机器人、芯片技术、人机交互、人工神经网络
生成调度	机器学习、物联网、计算机视觉
故障诊断	机器学习、物联网、人机交互、计算机网络、智能识别
智能物流	机器学习、物联网、计算机视觉
智能 MES 生产信息化管理系统	机器学习、数据挖掘、可视化

（1）智能工厂

智能工厂是实现智能制造的重要载体，其本质是以信息物理系统（CPS）和工业互联网为核心，利用信息技术和智能装备对生产工艺、组织流程、管理模式以及产品全生命周期进行数字化、网络化、智能化改造，加强设备、制造单元、生产线、车间、工厂的互联互通，实现人、机、法、料、环高度协同融合，推动企业纵向集成和横向集成，并基于工业大数据应用和工业云服务，为企业提供工厂级的端到端整体解决方案，实现提质增效和产业转型升级。

智能生产是指使用智能装备、传感器、过程控制、智能物流、制造执行系统、信息物理系统组成的人机一体化系统；按照工艺设计要求，实现整个生产制造过程的智能化生产、有限能力排产、物料自动配送、状态跟踪、优化控制、智能调度、设备运行状态监控、质量追溯和管理、车间绩效等；对生产、设备、质量的

异常做出正确的判断和处置；实现制造执行与运营管理、研发设计、智能装备的集成；实现设计制造一体化，管控一体化。

智能工厂有三种不同的建设模式。

第一种模式是从生产过程数字化到智能工厂。在石化、钢铁、冶金、建材、纺织、造纸、医药、食品等流程制造领域，企业发展智能制造的内在动力在于产品品质可控，侧重从生产数字化建设起步，基于品控需求从产品末端控制向全流程控制转变。因此其智能工厂建设模式为：一是推进生产过程数字化，在生产制造、过程管理等单个环节信息化系统建设的基础上，构建覆盖全流程的动态透明可追溯体系，基于统一的可视化平台实现产品生产全过程跨部门协同控制；二是推进生产管理一体化，搭建企业 CPS 系统，深化生产制造与运营管理、采购销售等核心业务系统集成，促进企业内部资源和信息的整合和共享；三是推进供应链协同化，基于原材料采购和配送需求，将 CPS 系统拓展至供应商和物流企业，横向集成供应商和物料配送协同资源和网络，实现外部原材料供应和内部生产配送的系统化、流程化，提高工厂内外供应链运行效率；四是整体打造大数据化智能工厂，推进端到端集成，开展个性化定制业务。

第二种模式是从智能制造生产单元（装备和产品）到智能工厂。在机械、汽车、航空、船舶、轻工、家用电器和电子信息等离散制造领域，企业发展智能制造的核心目的是拓展产品价值空间，侧重从单台设备自动化和产品智能化入手，基于生产效率和产品效能的提升实现价值增长。因此其智能工厂建设模式为：一是推进生产设备（生产线）智能化，通过引进各类符合生产所需的智能装备，建立基于 CPS 系统的车间级智能生产单元，提高精准制造、敏捷制造能力。二是拓展基于产品智能化的增值服务，利用产品的智能装置实现与 CPS 系统的互联互通，支持产品的远程故障诊断和实时诊断等服务。三是推进车间级与企业级系统集成，实现生产和经营的无缝集成和上下游企业间的信息共享，开展基于横向价值网络的协同创新。四是推进生产与服务的集成，基于智能工厂实现服务化转型，提高产业效率和核心竞争力。

例如，广州数控通过利用工业以太网将单元级的传感器、工业机器人、数控机床，以及各类机械设备与车间级的柔性生产线总控制台相连，利用以太网将总

控台与企业管理级的各类服务器相连,再通过互联网将企业管理系统与产业链上下游企业相连,打通了产品全生命周期各环节的数据通道,实现了生产过程的远程数据采集分析和故障监测诊断。三一重工的 18 号厂房是总装车间,有混凝土机械、路面机械、港口机械等多条装配线,通过在生产车间建立“部件工作中心岛”,即单元化生产,将每一类部件从生产到下线所有工艺集中在一个区域内,在保证结构件制造工艺不改变、生产人员不增加的情况下,实现了减少占地面积、提高生产效率、降低运行成本的目的。目前,三一重工已建成车间智能监控网络和刀具管理系统、公共制造资源定位与物料跟踪管理系统、计划、物流、质量管控系统、生产控制中心(PCC)中央控制系统等智能系统,还与其他单位共同研发了智能上下料机械手、基于 DNC 系统的车间设备智能监控网络、智能化立体仓库与 AGV 运输软硬件系统、基于 RFID 设备及无线传感网络的物料和资源跟踪定位系统、高级计划排程系统(APS)、制造执行系统(MES)、物流执行系统(LES)、在线质量检测系统(SPC)、生产控制中心管理决策系统等关键核心智能装置,实现了对制造资源跟踪、生产过程监控,计划、物流、质量集成化管控下的均衡化混流生产。

第三种模式是从个性化定制到互联工厂。在家电、服装、家居等消费品制造领域,企业发展智能制造的重点在于充分满足消费者多元化需求的同时实现规模经济生产,侧重通过互联网平台开展大规模个性化定制模式创新。因此其智能工厂建设模式为:一是推进个性化定制生产,引入柔性化生产线,搭建互联网平台,促进企业与用户深度交互、广泛征集需求,基于需求数据模型开展精益生产;二是推进设计虚拟化,依托互联网逆向整合设计环节,打通设计、生产、服务数据链,采用虚拟仿真技术优化生产工艺;三是推进制造网络协同化,变革传统垂直组织模式,以扁平化、虚拟化新型制造平台为纽带集聚产业链上下游资源,发展远程定制、异地设计、当地生产的网络协同制造新模式。

近年来,全球各主要经济体都在大力推进制造业的复兴。在工业 4.0、工业互联网、物联网、云计算等热潮下,全球众多优秀制造企业都开展了智能工厂建设实践。例如,西门子安贝格电子工厂实现了多品种工控机的混线生产;FANUC 公司实现了机器人和伺服电机生产过程的高 2 度自动化和智能化,并利用自动化

立体仓库在车间内的各个智能制造单元之间传递物料，实现了最高 720 小时无人值守；施耐德电气实现了电气开关制造和包装过程的全自动化；美国哈雷戴维森公司广泛利用以加工中心和机器人构成的智能制造单元，实现大批量定制；三菱电机名古屋制作所采用人机结合的新型机器人装配产线，实现从自动化到智能化的转变，显著提高了单位生产面积的产量；全球重卡巨头 MAN 公司搭建了完备的厂内物流体系，利用 AGV 装载进行装配的部件和整车，便于灵活调整装配线，并建立了物料超市，取得明显成效。

在 2020 年疫情期间，电力管理大厂伊顿(Eaton)凭借工业 4.0 解决方案抗疫，以 Microsoft HoloLens 2 AR 眼镜在实体厂区显示 3D 影像并支持远距连接，透过远程执行核心工作并保障员工安全。日本制造大厂日立(Hitachi)不仅以热像仪确保工作场所的员工体温，也结合视讯分析个人防护设备的使用情况。

（2）智能物流

建设智能仓储和物流配送系统是实现智能生产的重要组成部分。智能仓储和物流系统由仓储物流信息管理系统、自动控制系统、物流设施设备系统组成。

人工智能在物流业的落地应用主要指的是基于人工智能技术（机器学习、深度学习、计算机视觉、自动驾驶等）的软硬件产品及服务（无人卡车、无人机/无人车、智能调度系统等）在物流活动各环节（运输、仓储、配送、客服等）中的实际落地应用。

目前，在物流行业实现应用的人工智能技术主要以深度学习、计算机视觉、自动驾驶及自然语言理解为主。其中，深度学习在运输路径规划、运力资源优化、配送智能调度等场景中发挥至关重要的作用；计算机视觉技术在物流领域应用很广，智能仓储机器人、无人配送车、无人配送机等智能设备都以视觉技术为基础，此外，计算机视觉还能实现运单识别、体积测量、装载率测定、分拣行为检测等多项功能；自动驾驶技术是运输环节智能化的核心技术，目前尚未正式使用，但头部企业的无人卡车已经开始在特定路段进行实地路测和试运行；自然语言理解主要用于物流企业的智能客服系统，能够有效降低企业在客服环节的人工成本。

AI 可以助力于物流降本增效。人工智能技术产品的加入能够进一步推动物

流业向“智慧物流”发展，更大限度地降低人工成本、提升经营效率。智能物流系统将工厂内的智能识别设备（如 RFID、传感器等）、智能物流装备、信息控制系统（如 MES 等）有效串联，应用于从采购、生产、仓储到发货等全部作业环节。因此，智能物流系统是智能制造企业提高生产效率、订单交付能力、库存周转水平三大智能制造关键指标的重要支撑，也是保证产品品质、提升制造企业竞争力的核心。

目前，智能物流服务比较丰富。例如，百度大脑推出的 EasyDL 可以进行物体检测助力医药物流行业高效药盒分拣。EasyDL 专业版的物体检测模型 3D 定位精度在 1-2mm，姿态精度在 1 度左右，机械臂臂展较长，通常为 1.3 或 1.5 米，完成一次分拣任务的周期可低至 7-8 秒左右，机械臂支持 24 小时不间断工作，分拣效率相比人工大大提升。

运输是物流产业链条的核心环节，在物流成本构成中占比 50% 以上。人工智能在运输中的应用方向集中在无人卡车及车辆管理。无人卡车主要以自动驾驶技术为核心。车辆管理主要是基于计算机视觉与 AIoT 产品技术，为运输车辆管理系统提供实时感知功能。尽管人工智能赋能物流运输的最终形态将是由无人卡车替代人工驾驶卡车，但是由于运输环境及运输设备的复杂性，现阶段人工智能在物流运输中的应用尚处于起步阶段。

（3）智能系统

MES（Manufacturing Execution System），即制造执行系统，是一个面向制造企业车间执行层的生产信息化管理系统，主要负责车间生产管理和调度执行。MES 对于生产信息实施监控，将企业上层生产计划系统与车间下层的设备控制系统联系起来，打通工厂信息通道。MES 可以为企业提供制造数据管理、计划排程管理、生产调度管理、库存管理、质量管理、人力资源管理、工作及设备管理、工具工装管理、采购管理、成本管理、项目看板管理、生产过程控制、底层数据集成分析、上层数据集成分解等管理模块。目前，国内生产的 MES，大多数已实现生产调度、产品跟踪、质量控制、设备故障分析、网络报表等各项功能的统一集成，但在“系统安全性”上尚存在缺陷。

MES 是美国先进制造研究机构 AMR 于 20 世纪 90 年代提出的概念,将 MES 定义为“**位于上层的计划管理系统与底层的工业控制之间的、面向车间层的管理信息系统**”,它为操作人员/管理人员提供计划的执行、跟踪以及所有资源(人、设备、物料、客户需求等)的当前状态。

近年来, MES 软件成为智能工厂多个环节数据交换的核心。目前产业格局仍是欧美企业主导。全球工业软件产业主要由欧美企业主导,例如 SAP、Siemens、IBM、达索系统等。MES 软件具有较强的行业应用特性,与特定的行业关系紧密,需要大量行业领域知识的积累,因此形成不同 MES 厂家占据不同行业的局面。

制造执行系统协会(Manufacturing Execution System Association, MESA)对 MES 所下的定义是:**“MES 在工厂综合自动化系统中起着中间层的作用,在 ERP 系统产生的长期计划指导下, MES 根据底层控制系统采集的与生产有关的实时数据,对短期生产作业的计划调度、监控、资源配置和生产过程进行优化。”**

制造执行系统 MES 为企业创造的价值是:缩短在制品周转和等待时间,提高设备利用率和车间生产能力,提高现场异常情况的响应和处理能力,缩短计划编制周期,以及计划人员的人力成本,提高计划准确性,从计划的粗放式管理向细化到工序的详细计划转变,提高生产统计的准确性和及时性,降低库存水平和在制品数量,不断改善质量控制过程,提高产品质量。

人工智能对高端制造业赋能效果显著,是目前制造业的主要发展方向。目前在 MES 系统中应用的人工智能技术包括深度学习、计算机视觉等。结合人工智能的 MES,能更有效地进行产品和工艺管理、计划调度生产管理、物料物流管理、生产质量管理、车间设备维护管理。

9 人工智能发展的机遇和挑战

9.1 人工智能未来发展机遇

目前，全球已有美国、中国、欧盟、英国、日本、德国、加拿大等 10 余个国家和地区纷纷发布了人工智能相关国家发展战略或政策规划，用于支持 AI 未来发展。这些国家几乎都将人工智能视为引领未来、重塑传统行业结构的前沿性、战略性技术，积极推动人工智能发展及应用，注重人工智能人才培养。这是 AI 未来发展的重要历史机遇。

9.1.1 全球主要经济体 AI 发展支持政策

美国：指定 AI 研究为政府优先事项并调入更多支持资金和资源

美国高度重视人工智能全面发展，包括立法、研发投资、人才培养等多个方面纷纷给予支持。2016 年，美国国家科学技术委员会（NSTC）发布《国家人工智能研发战略计划》全面布局人工智能发展。与此同时，美国总统办公室发布报告《为未来人工智能做好准备》，以应对人工智能带来的潜在风险，以及《人工智能、自动化与经济报告》，强调人工智能驱动的自动化对经济发展的影响。

2018 年，美国白宫首次将人工智能指定为政府研发的优先事项，并且成立人工智能特别委员会，旨在协调联邦政府各机构之间人工智能研发优先事项，并向白宫提出行动建议，以确保美国人工智能技术的领导地位。美国国防部高级研究项目局宣布投资 20 亿美元开发下一代人工智能技术。美国国会两院讨论包括《人工智能未来法案》《人工智能就业法案》和《国家安全委员会人工智能法案》等法案。五角大楼成立了“联合人工智能中心”，确保国防部对人工智能相关数据信息的高效利用。

2019 年，美国白宫科学和技术政策办公室（OSTP）发布了由总统特朗普签署的《美国人工智能倡议》，将人工智能的重要性上升到美国经济安全和国家安全的层面，要求调配更多联邦资金和资源转向人工智能研究，并呼吁美国主导国际人工智能标准的制定，开展人工智能时代美国劳动力培养的研究。白宫还对《国

家人工智能研发战略规划》进行了更新，确定了联邦投资于人工智能研发的优先事项。美国国防部网站公布《2018 年国防部人工智能战略摘要——利用人工智能促进安全与繁荣》，并成立联合人工智能中心（JAIC），旨在加快人工智能快速赋能关键作战任务，统筹协调人工智能研发项目，积极维持美国在 AI 方面的战略地位。同年，国防授权法案批准设立人工智能国家安全委员会，该委员会旨在全面审查、分析人工智能技术及系统；商务部成立白宫劳动力委员会，以帮助美国储备人工智能等新兴科技发展所需的人才；国家科学基金会持续资助人工智能基础研究领域。

欧盟：重视并推动 AI 发展中的伦理和安全理念

欧盟在人工智能发展战略中坚持推行以人为本的理念，在 2018 年发布了《欧盟人工智能战略》，推动欧盟人工智能领域的技术研发、道德规范制定以及投资规划，并计划在 2020 年底至少投入 200 亿欧元。随后，欧盟宣布在“地平线 2020”研究与创新项目中对人工智能研发投入 15 亿欧元的专项资金，将资助创建欧洲人工智能生态系统的支撑平台。

欧盟专门设立了高级别人工智能专家组（AIHLEG），就人工智能的投资和政策制定提出建议，为人工智能的道德发展制定指导方针。该专家组制定了《可信赖的人工智能道德准则草案》，提出实现可信赖人工智能的道德准则和具体要求，包括数据保护和数据透明度等问题。该草案是欧盟为增加政府和私营部门人工智能领域合作的提出的三大战略之一，三大战略包括：增加政府和私营部门对人工智能的投资、为人工智能可能引发的社会和经济变革做好准备、建立适当的人工智能道德和法律框架。

此外，欧盟成员国还于 2018 年签署了《人工智能合作宣言》，就人工智能可能引发的社会、经济、伦理道德和法律等重要问题开展合作，确保欧洲在人工智能研发和应用上具有强大竞争力。随后，又发布《促进人工智能在欧洲发展和应用的协调行动计划》，提出设计伦理和设计安全两大关键原则，旨在使欧盟成为发展前沿、符合道德伦理、安全的人工智能技术的世界领先地位，强调将通过以人为本的方式促进人工智能技术发展。

2019 年，欧盟启动了 AI FOR EU 项目，建立人工智能需求平台、开放协作平台，整合汇聚 21 个成员国 79 家研发机构、中小企业和大型企业的数据、计算、算法和工具等人工智能资源，提供统一开放服务。此外还发布了《人工智能伦理准则》，以提升人们对人工智能产业的信任。

英国：不断加大政策、资金、人才和国际合作方面的布局力度

英国政府在 2017 年发布的《产业战略：建设适应未来的英国》中，确立了人工智能发展的四个优先领域：将英国建设为全球 AI 与数据创新中心；支持各行业利用 AI 和数据分析技术；在数据和人工智能的安全等方面保持世界领先；培养公民工作技能。随后，发布了《在英国发展人工智能》，建议建立人工智能和数据科学的艾伦·图灵研究所，旨在与其他公共研究机构建立合作，统筹协调针对人工智能研究的计算能力。

2018 年，英国政府发布《产业战略：人工智能领域行动》，这是英国政府和产业界做出的首份发展人工智能的承诺，将采取切实行动推进人工智能发展，促进英国人工智能和数字驱动的经济蓬勃发展。英国政府在《人工智能领域行动》等多个人工智能方面的政策文件中，提出政府提高研发经费投入，优先支持关键领域的创新等措施。这些举措包括：未来 10 年，英国政府将研发经费（包括人工智能技术）占 GDP 的比例提高到 2.4%；2021 年研发投入将达 125 亿英镑；从“产业战略挑战基金”中拨款 9300 万英镑，用于机器人与 AI 技术研发等。英国政府也积极推出针对初创企业的激励政策。

近年来，英国政府不断加大政策、资金、人才、国际合作等方面的布局力度^[77]。在政策方面，据英国政府 2018 年推出的《工业战略：人工智能产业政策》报告显示，过去 3 年英国发布了包括人工智能产业在内的工业战略白皮书、人工智能产业政策等各项措施，并成立了人工智能发展委员会、数据伦理与创新中心、人工智能发展办公室及工业战略挑战基金等相关机构，以推动人工智能的发展。在资金方面，英国规划制定了金额超 9 亿英镑（约 78.7 亿元人民币）的一揽子人工智能产业扶持计划，还将投资谷歌、亚马逊、“人工智能元素”（Element AI）

⁷⁷ 英国更注重人工智能基础性研究，2019 年 03 月 12 日，经济参考报，<http://scitech.people.com.cn/n1/2019/0312/c1007-30971696.html>

以及“慧与科技”（HPE）等一系列跨国科技公司。2019年2月，英国政府宣布投资1300万英镑（约1.13亿元人民币）支持40个人工智能及数据分析项目，旨在提升生产力，改善英国的专业服务。在人才方面，自2017年起，英国计划将在4年内培育8000名计算机科学教师；未来7年，通过培训让5000名学生具备多样化的数字技术；支持新增450个与人工智能相关的博士点；加大包括人工智能人才在内的海外特殊人才引进力度，每年增加1000名至2000名人才引进。在国际合作方面，2018年7月，英国与法国签订五年协议，在人工智能等数字产业领域加强双方高端科研中心的合作；2019年1月，英国决定与日本在机器人、数据等领域加强深度合作。

德国：用 AI+工业 4.0 打造“人工智能德国造”品牌

德国政府早在2013年提出的“工业4.0”战略中，就已经涵盖了人工智能。2018年，德国联邦政府颁布了《高科技战略2025》，提出“推进人工智能应用，使德国成为人工智能领域世界领先的研究、开发和应用地点之一”，还明确提出建立人工智能竞争力中心、制定人工智能战略、组建数据伦理委员会、建立德法人工智能中心等。在《联邦政府人工智能战略》中制定三大战略目标，以及包括研究、技术转化、创业、人才、标准、制度框架和国际合作在内的12个行动领域，旨在打造“人工智能德国造”品牌。在资金投入方面，德国政府宣布将首先投入5亿欧元用于2019年及之后几年的人工智能发展，并将在2025年底累计投入30亿欧元。德国经济和能源部在2019年发布的《国家工业战略2030》（草案）中，也多次强调人工智能的重要性。

2020年1月15日，德国柏林工业大学宣布成立新的人工智能研究所，进一步开展人工智能科研和人才培养。德国联邦政府将在人工智能战略框架内对该研究所追加预算，预计到2022年时，研究所将获得3200万欧元财政支持。柏林市政府也将为研究所新增人工智能岗位^[78]。

⁷⁸ 德国：老牌工业强国用 AI 搭配工业 4.0，澎湃新闻，2020-07-06，<https://baijiahao.baidu.com/s?id=1671425654188905121&wfr=spider&for=pc>

日本：主张构建有效且安全应用的“AI-Ready 社会”

日本政府积极发布国家层面的人工智能战略、产业化路线图。2016 年成立了人工智能技术战略委员会，作为人工智能国家层面的综合管理机构，以制定人工智能研究和发展目标以及人工智能产业化路线图，负责推动总务省、文部省、经产省以及下属研究机构间的协作，进行人工智能技术研发。该委员会有 11 名成员，分别来自学术界、产业界和政府。2017 年，日本发布《人工智能技术战略》，确定了在人工智能技术和成果商业化方面，政府、产业界和学术界合作的行动目标。2018 年，日本发布《综合创新战略》提出要培养人工智能领域技术人才，确保在 2025 年之前每年培养和录用几十万名 IT 人才。此外，还发布了《集成创新战略》，将人工智能指定为重点发展领域之一，提出要加大其发展力度，同时强调要加强人工智能领域人才培养。2018 年 12 月 27 日，日本内阁府发布《以人类为中心的人工智能社会原则》推进人工智能发展，从宏观和伦理角度表明了日本政府的态度，主张在推进人工智能技术研发时，综合考虑其对人类、社会系统、产业构造、创新系统、政府等带来的影响，构建能够使人工智能有效且安全应用的“AI-Ready 社会”，于 2019 年 3 月正式公布。此原则是将人工智能(Artificial Intelligence, AI)视为未来的关键科技，但在研发应用上，须以联合国的可持续发展目标(Sustainable Development Goals, SDGs)为基础，以落实日本“超智能社会”(Society 5.0)为准则，其基本理念是 Dignity、Diversity & Inclusion 及 Sustainability，并且建构“尊重人类尊严”、“不同背景的大众皆能追求幸福”及“持续性”的社会。

韩国：提升领域竞争力发展成为“AI 强国”

韩国政府于 2019 年 12 月 17 日公布“人工智能(AI)国家战略”，以推动人工智能产业发展。该战略旨在推动韩国从“IT 强国”发展为“AI 强国”，计划在 2030 年将韩国在人工智能领域的竞争力提升至世界前列。并且，提出构建引领世界的人工智能生态系统、成为人工智能应用领先的国家、实现以人为本的人工智能技术。在人工智能生态系统构建和技术研发领域，韩国政府将争取至 2021 年全面开放公共数据，到 2024 年建立光州人工智能园区，到 2029 年为新

一代存算一体人工智能芯片研发投入约 1 万亿韩元。

其他国家

加拿大在 2017 年宣布了泛加拿大人工智能战略，承诺提供 1.25 亿加元的加拿大人工智能研究与开发。这一战略旨在增加加拿大的 AI 和毕业生人数。在埃德蒙顿、蒙特利尔和多伦多建立科学卓越中心。建立加拿大在 AI 经济、伦理、政策和法律研究方面的全球思想领导地位

法国于 2018 年 3 月发布 AI 战略，将投入 1.5 亿欧元把法国打造成 AI 研究、训练和行业的全球领导者。该计划由四个部分组成，一是宣布国家人工智能计划，将在法国各地建立一个由四五个研究机构组成的网络；二是将制定一项开放数据政策，推动人工智能在医疗等领域应用；三是政府将创建一个监管和金融框架，以支持国内“人工智能冠军企业”的发展；四是政府将制定道德规范。

印度在 2018 年 6 月发布《人工智能国家战略》，探求如何利用人工智能来促进经济增长和提升社会包容性，寻求一个适用于发展中国家的 AI 战略部署。该战略旨在提高印度人的工作技能，投资于能够最大限度地提高经济增长和社会影响的研究和部门，以及将印度制造的人工智能解决方案推广到其他发展中国家。

以色列于 2019 年 11 月发布了国家级人工智能计划，提出以色列要成为人工智能的世界五大国之一。并且政府以五年为一期，每年投资 10 至 20 亿新谢克尔（约 2.89 亿至 5.8 亿美元）开发人工智能技术，总共投资 100 亿新谢克尔（约 28.93 亿美元）于人工智能领域。

西班牙于 2019 年 3 月发布《西班牙人工智能研究、发展与创新战略》，认为最优先事项是建立一个有效的机制，以保障人工智能的研究、发展、创新，并评估人工智能对人类社会的影响。

9.1.2 中国 AI 发展支持政策

党和国家高度重视 AI 发展，从产业发展、教育等各个方面支持人工智能的发展。习近平总书记也曾多次强调用人工智能开辟社会治理新格局、人工智能为高质量发展赋能。早在 2015 年，《国务院关于积极推进“互联网+”行动的指导意见》就提出加快人工智能核心技术突破，促进人工智能在智能家居、智能终端、

智能汽车、机器人等领域推广应用的目标。近年来发布了一系列的支持人工智能发展政策，如图 9-1 所示。



图 9-1 中国人工智能发展重要支持政策

2016 年，人工智能写入国民经济和社会发展的“十三五”规划纲要。国务院先后发布了《“十三五”国家科技创新规划》和《“十三五”国家战略性新兴产业发展规划》。前者指出要重点发展大数据驱动的类人智能技术方法，在基于大数据分析的类人智能方向取得重要突破；后者强调要发展人工智能，培育人工智能产业生态，推动人工智能技术向各行业全面融合渗透。同年，国家发展改革委、科技部、工业和信息化部、中央网信办联合制定发布了《“互联网+”人工智能三年行动实施方案》，提出了总体思路和实施目标。

2017 年，人工智能首次被写入全国政府工作报告。国务院印发了《关于新一代人工智能发展规划的通知》确定新一代人工智能发展三步走战略目标，人工智能上升为国家战略层面。同年 10 月，人工智能被写入十九大报告，将推动互联网、大数据、人工智能和实体经济深度融合。工信部随后发布了《促进新一代人工智能产业发展三年行动计划(2018-2020 年)》，从推动产业发展角度出发，结合“中国制造 2025”，对《新一代人工智能发展规划》相关任务进行了细化和落实。

2018，政府工作报告中再次强调要加强新一代人工智能研发应用；在医疗、养老、教育等多领域推进“互联网+”；发展智能产业，拓展智能生活。教育部在《高等学校人工智能创新行动计划》中提出，到 2020 年，基本完成适应新一代人工智能发展的高校科技创新体系和学科体系的优化布局，高校在新一代人工

智能基础理论和关键技术研究等方面取得新突破，人才培养和科学研究的优势进一步提升，并推动人工智能技术广泛应用。工信部印发了《新一代人工智能产业创新重点任务揭榜工作方案》，提出通过在人工智能主要细分领域，选拔领头羊、先锋队，树立标杆企业，培育创新发展的主力军，加快我国人工智能产业与实体经济深度融合。

2019 年，政府工作报告中将人工智能升级为智能+，要推动传统产业改造提升，特别是要打造工业互联网平台，拓展“智能+”，为制造业转型升级赋能。中央全面深化改革委员会第七次会议审议通过了《关于促进人工智能和实体经济深度融合的指导意见》，提出要把握新一代人工智能的发展特点，结合不同行业、不同区域特点，探索创新成果应用转化的路径和方法，构建数据驱动、人机协同、跨界融合、共创分享的智能经济形态。同年，国家新一代人工智能治理专业委员会正式成立，并且发布了《新一代人工智能治理原则——发展负责任的人工智能》。随后，科技部发布《国家新一代人工智能创新发展实验区建设工作指引》，工业和信息化部《加快培育共享制造新模式新业态 促进制造业高质量发展的指导意见》与《“5G+工业互联网”512 工程推进方案》等具体鼓励措施。

进入 2020 年，国家大力推进并强调要加快 5G 网络、人工智能、数据中心等新型基础设施建设进度。人工智能技术被视为新一轮产业变革的核心驱动力量。此外，教育部、国家发展改革委、财政部联合发布了《关于“双一流”建设高校促进学科融合加快人工智能领域研究生培养的若干意见》，提出要构建基础理论人才与“人工智能+X”复合型人才并重的培养体系，探索深度融合的学科建设和人才培养新模式。7 月，国家标准化管理委员会、中央网信办、国家发展改革委、科技部、工业和信息化部联合印发《国家新一代人工智能标准体系建设指南》（国标委联〔2020〕35 号），以加强人工智能领域标准化顶层设计，推动人工智能产业技术研发和标准制定，促进产业健康可持续发展。

9.2 人工智能未来技术研究方向

人工智能经历几波浪潮之后，在过去十年中基本实现了感知能力，但却无法做到推理、可解释等认知能力，因此在下一波人工智能浪潮兴起时，将主要会去

实现具有推理、可解释性、认知的人工智能。2015 年，张钹院士提出第三代人工智能体系的雏形。2017 年，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三个方面，全面开展可解释性 AI 系统的研究。2018 年底，第三代人工智能的理论框架体系正式公开提出，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包括：（1）与脑科学融合，发展脑启发的人工智能理论；（2）探索数据与知识融合的人工智能理论与方法。虽然还没有明确第三代人工智能是什么，但是其趋势是清晰的。

Gartner 2020 年人工智能技术成熟度曲线图显示，如图 9-2 所示。2020 年人工智能技术成熟度曲线共有 30 项技术出现，其中有 17 项技术需要 2 到 5 年才能达到成熟期，有 8 项技术需要 5 到 10 年才能达到成熟期。出现的这些技术基本处于创新萌芽期、期望膨胀的顶峰期和泡沫低谷期，而“稳步爬升的光明期”和“实质生产的高峰期”出现的技术寥寥无几，仅有 Insight Engines（洞察引擎）和 GPU Accelerators（GPU 加速器）。

Hype Cycle for Artificial Intelligence, 2020

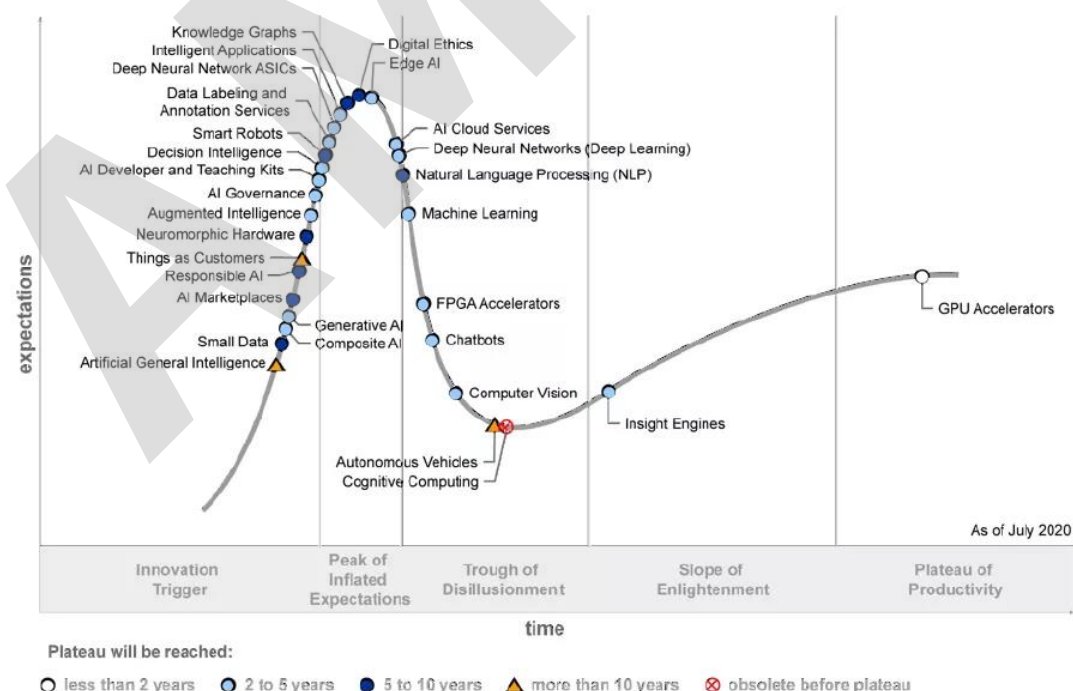


图 9-2 Gartner 2020 年人工智能技术成熟度曲线图

通过对 2020 年人工智能技术成熟度曲线分析,并结合人工智能的发展现状,本报告认为人工智能下一个十年重点发展的方向包括:强化学习(Reinforcement Learning)、神经形态硬件(Neuromorphic Hardware)、知识图谱(Knowledge Graphics)、智能机器人(Smart Robotics)、可解释性 AI(Explainable AI)、数字伦理(Digital Ethics)、自然语言处理(Natural Language Processing)等,这些技术目前多处于期望膨胀期,表明人们对其未来发展充满着很大期待,预计达到稳定期需要 5-10 年,是 AI 未来十年重点发展方向。

(1) **强化学习(Reinforcement Learning)**。强化学习用于描述和解决智能体在与环境的交互过程中通过学习策略以达成回报最大化或实现特定目标的问题。强化学习不受标注数据和先验知识所限制,而是通过接收环境对动作的奖励(反馈)获得学习信息并更新模型参数。由于智能体和环境的交互方式与人类和环境的交互方式类似,强化学习可以认为是一套通用的学习框架,可用来解决通用人工智能的问题。

(2) **神经形态硬件(Neuromorphic Hardware)**。神经形态硬件旨在用与传统硬件完全不同的方式处理信息,通过模仿人脑构造来大幅提高计算机的思维能力与反应能力。采用多进制信号来模拟生物神经元的功能,可将负责数据存储和数据处理的元件整合到同一个互联模块当中。从这一意义上说,这一系统与组成人脑的数十亿计的、相互连接的神经元颇为相仿。神经形态硬件能够大幅提升数据处理能力和机器学习能力,能耗和体积却要小得多,为人工智能的未来发展提供强大的算力支撑。

(3) **知识图谱(Knowledge Graphics)**。要实现真正的类人智能,机器还需要掌握大量的常识性知识,以人的思维模式和知识结构来进行语言理解、视觉场景解析和决策分析。知识图谱将互联网的信息表达成更接近人类认知世界的形式,提供了一种更好地组织、管理和理解互联网海量信息的能力,被认为是从感知智能通往认知智能的重要基石。从感知到认知的跨越过程中,构建大规模高质量知识图谱是一个重要环节,当人工智能可以通过更结构化的表示理解人类知识,并进行互联,才有可能让机器真正实现推理、联想等认知功能。清华大学唐杰教授在知识图谱的基础上提出的“认知图谱=知识图谱+认知推理+逻辑表达”,为

人工智能未来十年的发展提供了研究方向^[79]。

(4) **智能机器人(Intelligent Robot)**。智能机器人需要具备三个基本要素：感觉要素、思考要素和反应要素。感觉要素是利用传感器感受内部和外部信息，如视觉、听觉、触觉等；思考要素是根据感觉要素所得到的信息，思考出采用什么样的动作；反应要素是对外界做出反应性动作。智能机器人的关键技术包括多传感器信息融合、导航与定位、路径规划、智能控制等。由于社会发展的需求和机器人应用行业的扩大，机器人可以具备的智能水平并未达到极限，影响因素包括硬件设施的计算速度不够、传感器的种类不足，以及缺乏机器人的思考行为程序难以编制等。

(5) **可解释人工智能(Explainable AI)**。虽然深度学习算法在语音识别、计算机视觉、自然语言处理等领域取得令人印象深刻的性能，但是它们在透明度和可解释性方面仍存在局限性。深度学习的不可解释性已经成为计算机领域顶级会议（如 NIPS）火药味十足的讨论话题。一些方法尝试将黑盒的神经网络模型和符号推理结合了起来，通过引入逻辑规则增加可解释性。此外，符号化的知识图谱具有形象直观的特性，为弥补神经网络在解释性方面的缺陷提供了可能。利用知识图谱解释深度学习和高层次决策模型，是当前值得研究的科学问题，可以为可解释的 AI 提供全新视角的机遇。张钹院士指出当前人工智能的最大问题是不可解释和不可理解，并提倡建立具有可解释性的第三代人工智能理论体系^[80]。

(6) **数字伦理(Digital Ethics)**。作为新一轮科技革命和产业变革的重要驱动力，人工智能已上升为国家战略，人工智能将会在未来几十年对人类社会产生巨大的影响，带来不可逆转的改变。人工智能的发展面临诸多现实的伦理和法律问题，如网络安全、个人隐私、数据权益和公平公正等。为了让人工智能技术更好地服务于经济社会发展和人民美好生活，不仅要发挥好人工智能的“头雁”效应，也要加强人工智能相关法律、伦理、社会问题等方面的研究。数字伦理将是未来智能社会的发展基石，只有建立完善的人工智能伦理规范，处理好机器与人的新关系，我们才能更多地获得人工智能红利，让技术造福人类。

(7) **自然语言处理(Nature Language Processing)**。深度学习在自然语言处理取得了巨大突破，它能够高效学习多粒度语言单元间复杂语义关联。但是

^[79] CSDN. [EB/OL] <https://blog.csdn.net/tMb8Z9Vdm66wH68VX1/article/details/105462845>.

^[80] 张钹.人工智能进入后深度学习时代[J].智能科学与技术学报,2019,1(01):4-6.

仅仅依靠深度学习并不能完成对自然语言的深度理解。对自然语言的深度理解需从字面意义到言外之意的跃迁，这需要引入复杂知识的支持。丰富的语言知识能够提升模型的可解释性，可覆盖长尾低频语言单位的知识规则能够提升模型的可扩展性，而异质多样的知识与推理体系能够提升模型鲁棒性。因此有必要研究知识指导的自然语言处理技术，揭示自然语言处理和知识产生及表达的机理，建立知识获取与语言处理双向驱动的方法体系，实现真正的语言与知识智能理解。

透过历史可以发现，每一次的经济大发展都与科技的突破紧密相关。种种迹象表明科技正迎来新的爆发期，全球科技竞赛也将再次掀起高潮。中国想要在这轮科技革新中占得先机，就需要加强技术预判，找准方向，提早部署，特别是在一些基础性、突破性的领域精准布局。

9.3 人工智能发展面临的问题

9.3.1 安全与伦理挑战

随着人工智能的快速发展和应用，人们越来越重视随之而来的安全和伦理问题。AI 发展面临着诸多安全和伦理方面的挑战。安全挑战主要包括三个方面：一是人工智能可以替代体力劳动和脑力劳动，相应的岗位替代作用影响着人类就业安全；二是建立在大数据和深度学习基础上的人工智能技术，需要海量数据来学习训练算法，带来了数据盗用、信息泄露和个人侵害的风险。许多个人信息如果被非法利用，将会构成对隐私权的侵犯。三是人工智能具有强大的数据收集、分析以及信息生成能力，可以生成和仿造很多东西，甚至包括人类自身。随之而生的虚假信息、欺诈信息不仅会侵蚀社会的诚信体系，还会对国家的政治安全、经济安全和社会稳定带来负面影响。

人工智能发展面临的伦理挑战主要来自以下方面。一是**人们对智能化的过度依赖**。人工智能发展带来的简易、便捷的智能化工作和生活方式的同时，严重挤占了人们用于休息的自由时间、用于劳动的工作时间和用于个人全面发展的时间，由此催生了许多人的懒惰和对智能产品的过度依赖；同时，个性化新闻推荐或者自动生成的新闻，真假难辨的广告和宣传给人们封闭在“信息茧房”里。甚至逐

渐失去了独立自由决策的能力，成为数据和算法的奴隶。二是情感计算技术和类脑智能技术的创新融合发展，**可能扰乱人们对于身份和能动性的认知**。人类大脑与机器智能直接连接，会绕过大脑和身体正常的感觉运动功能；增强型神经技术的应用也可能改变人的体能和心智。这是对人类社会的道德社会规范和法律责任的挑战。三是**智能算法歧视将带来的偏见**。人工智能以大数据和深度学习为基础，数据、算法以及人为因素会导致计算结果的偏见和非中立性，比如性别歧视、种族歧视以及“有色眼镜”效应。数据和算法导致的歧视往往具有更大的隐蔽性，更难以发现和消除。例如，微软在 Twitter 上上线的聊天机器人 Tay 在与网民互动过程中，由于大量恶意数据的输入，成为集性别歧视、种族歧视等于一身的“流氓”，它不但辱骂用户，还发表了种族主义评论和煽动性的政治宣言。四是**人工智能对人类造成的威胁和伤害**。智能武器是可自动寻找、识别、跟踪和摧毁目标的现代高技术兵器，包括精确制导武器、智能反导系统、无人驾驶飞机、无人驾驶坦克、无人驾驶潜艇、无人操作火炮、智能地雷、智能鱼雷和自主多用途智能作战机器人等，它将成为未来战场主力军，信息处理和计算能力成为战争胜负的决定因素。人工智能武器是继火药和核武器之后战争领域的第三次革命。人工智能如果被赋予伤害、破坏或欺骗人类的自主能力，将是人类的灾难，后果难以想象。

面对人工智能带来的安全和伦理问题，受到越来越多各方关注和应对。2020 年，美国国防部下属的国防创新委员会推出了《人工智能伦理道德标准》，公布了人工智能五大伦理原则，即负责、公平、可追踪、可靠和可控。无论作战还是非作战人工智能系统，均须遵守上述原则，否则美国国防部将不予部署。牛津大学成立了人工智能伦理研究所（Institute for Ethics in AI），并委任了由 7 位哲学家组成的首个学术研究团队。中国人工智能学会伦理专业委员会也正着手进行中国人工智能伦理规范研究。

9.3.2 国家之间技术限制

科技是未来竞争的制高点。虽然科技无国界，但是科技公司有国界。当前世界各国对人工智能技术发展都不遗余力地投入和支持，同时，还使用不同方法来

保护自己的科技成果，封锁前沿技术和“卡脖子”技术外流路径，这将在一定程度上限制人工智能技术创新要素的自由流动。

从全球范围来看，中国和美国人工智能领域科研论文和专利产出数量最多的两个国家。但是近年来，中美两国在人工智能技术领域的贸易关系则存在摩擦。2018 年 11 月 19 日美国商务部工业安全署（BIS）出台了一份针对关键新兴技术和相关产品的出口管制框架，其中在人工智能领域包括神经网络和深度学习、进化和遗传计算、强化学习、计算机视觉、专家系统、语音和音频处理、自然语言处理、规划、AI 芯片组、AI 云技术、音频和视频操作技术共计 11 项技术。2019 年 10 月 7 日，美国 BIS 部门把 8 家计算机视觉领域的中国科技企业加入“实体清单”。清单中的实体须在有许可证的情况下才可购买美国技术与产品，但美政府有权拒绝许可申请。

在字节跳动 TikTok 公司出售在美业务的谈判过程中，2020 年 8 月 28 日，中国商务部、科技部调整发布了最新版的《中国禁止出口限制出口技术目录》（商务部 科技部公告 2020 年第 38 号）。在最新目录中，语音合成、人工智能交互界面、语音评测、基于数据分析的个性化信息推送服务、无人机、量子密码等技术均被列入“限制出口”名单^[81]。根据《中华人民共和国技术进出口管理条例》，凡是涉及向境外转移技术，无论是采用贸易还是投资或是其他方式，均要严格遵守《中华人民共和国技术进出口管理条例》的规定，其中限制类技术出口必须到省级商务主管部门申请技术出口许可，获得批准后方可对外进行实质性谈判，签订技术出口合同。

在大型跨国公司的收购过程中，相关国家政府批准出售是交易宣告成功的必要条件。对于字节跳动出售 TikTok 业务来说，有可能出现其中一个国家政府出面阻止交易的情况。

根据人工智能技术关键词获取中美两国的论文数据，生成中国和美国在不同领域的研究成果对比图，如图 9-3 所示。分析发现，在被限制出口的计算及服务技术中，中国在以人脸识别为代表的计算机视觉、语音识别与自然语言处理（特

⁸¹ 中国的 AI 算法，现在限制出口了，机器之心，2020 年 8 月 29 日，<https://mp.weixin.qq.com/s/8p6ibpFOHoEtIExtLZNaOw>

别是中文)上有着较美国领先的优势,主要体现在高水平论文发表量、专利申请量两方面。

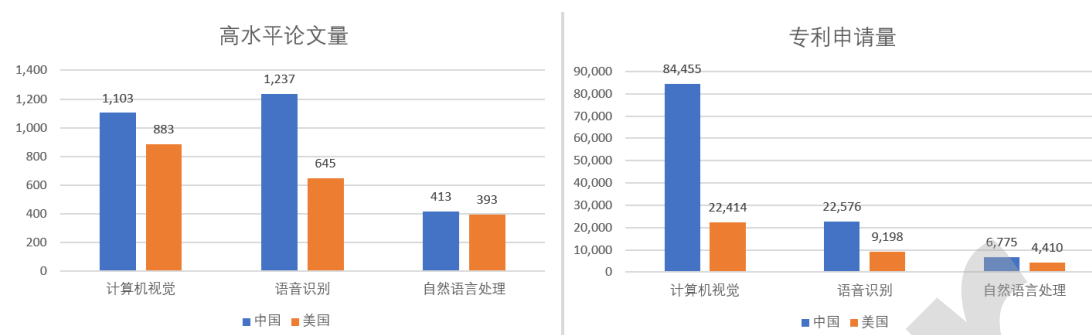


图 9-3 中国和美国 2011-2020 年在三个 AI 子领域的高水平论文量和专利申请量对比图

在语音识别、图像识别、自然语言处理技术上,中国国内市场提供了稳定庞大的用户与数据供应,以及政策支持为产业发展带来的所需资源和资本聚集,这些本土化优势,对于外国企业来说是不可复制的。

中国目前在计算机视觉领域的领先企业以 SenseTime, Face ++, YITU 和海康威视为代表,技术优势主要体现在人脸识别,在 2017 年中国在这一领域获得的专利数量大约是美国公司的 6 倍⁸²,其应用场景多为安全监控系统。相比而言,由于隐私政策,欧美的人脸识别技术难以发展实行,例如 2020 年 8 月 12 日英国法院裁决警察部门使用自动面部识别(AFR)违反了数据保护和平等法以及隐私权^[83]。

在语音识别领域,中国企业表现较优秀,特别是在中文识别和处理上。科大讯飞 iFlytek、依图科技 YITU、百度、腾讯、阿里巴巴等企业依靠中国庞大的中文用户,能获得远超美国能获得的中文语音数据库,这使得其语音识别 AI 有更好的语音识别学习条件。例如,腾讯可从其月活超 10 亿的微信用户那里获得中文语音数据。这一点是中国企业在中文语音识别技术上不可复制的优势^[84]。

⁸² Jeffrey Ding, interviewed by Frieda Klotz, Is China Taking the Lead in AI? April 30, 2020, <https://sloanreview.mit.edu/article/is-china-taking-the-lead-in-ai/>

⁸³ Police Facial Recognition Use Unlawful—U.K. Court Of Appeal Makes Landmark Ruling, Davey Winder, Aug 12, 2020, <https://www.forbes.com/sites/daveywinder/2020/08/12/police-facial-recognition-use-unlawful-uk-court-of-appeal-makes-landmark-ruling/#67922eaf75e0>

⁸⁴ The Future is Voice Activated: A Look into AI Voice Recognition in China, by Meiryum Ali, May 31, 2019, <https://lionbridge.ai/articles/the-future-is-voice-activated-a-look-into-ai-voice-recognition-in-china/>

在自然语言处理领域,百度的能力被认为超过微软和谷歌。受 Google 的 BERT 启发,百度的自然语言处理模型 ERNIE 最初是为理解汉语而开发的,但是它也能够更好地理解英语^[85]。Google 的模型在学习时会在每个序列中隐藏 15% 的单词,然后尝试根据上下文进行预测。基于类似的方法,百度团队通过让其 AI 模型预测文章中一串被隐藏的汉字,来学习文字组合的联系。不同于被微软和谷歌使用的英文,中文的特性要求 ERNIE 模型必须能够理解汉字组合后的出现的内在含义。结果显示,其在 GLUE 得分为第一名 90.1,超过微软和谷歌的模型得分。

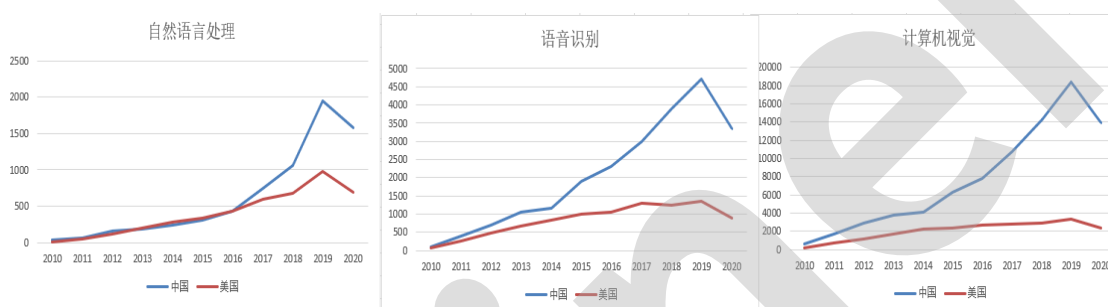


图 9-4 中国和美国 2011-2020 年在三个 AI 子领域专利公开趋势

我们研究发现,中国在人工智能和机器学习技术领域发展迅速,相关领域中国学者的论文发表量在 2008 年前后已经赶超美国。值得注意的是,中国杰出学者的国际合作对象国家不均衡,呈现出美国“一家独大”局面。以合发论文为产出指标看中国杰出学者开展国际合作的情况,中国杰出学者与美国合作紧密度最高,人数占比约 62.3%,其次是英国(14.7%)、德国(13.7%)、澳大利亚(9.5%)和新加坡(9.2%)。随着贸易战的不断升级蔓延,中美关系日益复杂,正常的科技与学术交流受阻,容易对我国的相关技术领域发展与人才培养造成不利影响。

在严峻的国际大环境下,未来人工智能技术自由交流发展将无疑受到影响。考虑到数据安全等多种因素,未来的基于数据分析的个性化信息推送服务技术,对外技术支持与技术服务出口都将受到限制。

⁸⁵ Baidu tops Microsoft and Google in teaching AI to understand human language, thanks to differences between Chinese and English, Tracy Qu, 30 Dec, 2019, <https://www.scmp.com/tech/big-tech/article/3043895/baidu-tops-microsoft-and-google-teaching-ai-understand-human-language>

附录 1 人工智能领域国际顶级期刊和会议

领域	期刊和会议	简称
经典 AI	AAAI Conference on Artificial Intelligence	AAAI
	International Joint Conference on Artificial Intelligence	IJCAI
机器学习	Annual Conference on Neural Information Processing Systems	NeurIPS
	International Conference on Machine Learning	ICML
	International Conference on Learning Representations	ICLR
计算机视觉	IEEE Conference on Computer Vision and Pattern Recognition	CVPR
	IEEE International Conference on Computer Vision	ICCV
	European Conference on Computer Vision	ECCV
自然语言处理	Annual Meeting of the Association for Computational Linguistics	ACL
	Conference on Empirical Methods in Natural Language Processing	EMNLP
	North American Chapter of the Association for Computational Linguistics	NAACL
机器人	IEEE International Conference on Robotics and Automation	ICRA
	IEEE RSJ International Conference on Intelligent Robots and Systems	IROS
知识工程	International Semantic Web Conference	ISWC
	International Conference on Principles of Knowledge Representation and Reasoning	KR
语音识别	IEEE International Conference on Acoustics, Speech and Signal Processing	ICASSP
数据挖掘	ACM SIGKDD International Conference on Knowledge Discovery and Data Mining	KDD
	ACM International Conference on Web Search and Data Mining	WSDM
信息检索与推荐	International ACM SIGIR Conference on Research and Development in Information Retrieval	SIGIR
	ACM Recommender Systems	RecSys
	International World Wide Web Conference	WWW
数据库	ACM SIGMOD International Conference on Management of Data	SIGMOD
	International Conference on Very Large Data Bases	VLDB
人机交互	ACM CHI Conference on Human Factors in Computing Systems	CHI
	ACM Conference on Computer-Supported Cooperative Work & Social Computing	CSCW
计算机图形	ACM SIGGRAPH Conference	SIGGRAPH
	ACM Transactions on Graphics	TOG
多媒体	ACM International Conference on Multimedia	MM
可视化	IEEE Transactions on Visualization & Computer Graphics	TVCG
	IEEE Visualization Conference	IEEE VIS
安全与隐私	ACM Conference on Computer and Communications Security	CCS
	IEEE Symposium on Security and Privacy	S&P
	USENIX Security Symposium	USS
计算机网络	ACM International Conference on Mobile Computing and Networking	MobiCom
	ACM SIGCOMM Conference	SIGCOMM
计算机系统	ACM Symposium on Operating Systems Principles	SOSP

	USENIX Symposium on Operating Systems Design and Implementation	OSDI
计算理论	ACM Symposium on Theory of Computing	STOC
	IEEE Annual Symposium on Foundations of Computer Science	FOCS
芯片技术	IEEE International Solid-State Circuits Conference	ISSCC
	Design Automation Conference	DAC
	ACM/SIGDA International Symposium on Field-Programmable Gate Arrays	FPGA
物联网	IEEE Internet of Things Journal	IoT-J
	IEEE Transactions on Wireless Communications	TWC

附录 2 全国 70 所高校人工智能学院、研究院、研究所名单

高校名称	学院/研究院 名称	成立时间
华中师范大学	人工智能教育学部	2020 年 5 月
北京师范大学	人工智能学院	2019 年 6 月
华东理工大学	人工智能与智能制造交叉研究院	2019 年 5 月
中国人民大学	高瓴人工智能学院	2019 年 4 月 22 日
北京大学	人工智能研究院	2019 年 4 月
厦门大学	人工智能研究院	2019 年 4 月
上海电力大学	上电-临港人工智能学院	2019 年 4 月
浙江理工大学	纺织工业人工智能研究院	2019 年 4 月
中国矿业大学（北京）	智慧矿山与机器人研究院	2019 年 4 月
西安交通大学	人工智能学院	2019 年 1 月 22 日
东南大学	人工智能学院&人工智能研究院	2019 年 1 月 4 日
华中科技大学	人工智能与自动化学院、人工智能研究院	2019 年 1 月
西北工业大学	虚拟现实与人工智能研究院	2019 年 1 月
西南大学	人工智能学院	2018 年 12 月 26 日
南京信息工程大学	人工智能海洋联合研究院	2018 年 12 月 1 日
山东大学	人工智能国际联合研究院	2018 年 12 月
中南大学	中南大学-深兰科技人工智能联合研究院	2018 年 12 月
大连理工大学	人工智能大连研究院	2018 年 10 月 22 日
福州大学	人工智能学院&人工智能研究院	2018 年 10 月 14 日
上海工程技术大学	人工智能产业研究院	2018 年 10 月
电子科技大学	人工智能研究院	2018 年 9 月 27 日
南京邮电大学	人工智能学院	2018 年 9 月 20 日
南通大学	南通大学智能信息技术研究中心	2018 年 9 月
四川大学	人工智能研究中心	2018 年 9 月
华东政法大学	人工智能与大数据指数研究院	2018 年 8 月
烟台大学	人工智能研究院	2018 年 7 月 31 日
北京科技大学	人工智能研究院	2018 年 7 月 21 日
北京邮电大学	人工智能研究院	2018 年 7 月 19 日
南京航空航天大学	人工智能学院&人工智能研究院	2018 年 7 月 2 日
南京理工大学	人工智能学院&人工智能研究院	2018 年 7 月 2 日
桂林电子科技大学	人工智能学院	2018 年 7 月
海南大学	大数据与人工智能研究院	2018 年 7 月
华东师范大学	人工智能研究所	2018 年 7 月
清华大学	人工智能研究院	2018 年 6 月 29 日
浙江理工大学	人工智能研究院	2018 年 6 月 25 日
重庆理工大学	重庆两江人工智能学院	2018 年 6 月 9 日
西南交通大学	人工智能研究院	2018 年 6 月 6 日

人工智能发展报告 2020

高校名称	学院/研究院 名称	成立时间
哈尔滨工程大学	人工智能研究院	2018 年 6 月
湖南工程学院	人工智能产业学院	2018 年 6 月
浙江大学	人工智能协同创新中心	2018 年 6 月
杭州电子科技大学	人工智能学院&人工智能研究院	2018 年 5 月 27 日
吉林大学	人工智能学院	2018 年 5 月 26 日
辽宁工程技术大学	人工智能学院	2018 年 5 月 17 日
南开大学	人工智能学院	2018 年 5 月 16 日
哈尔滨工业大学	人工智能研究院	2018 年 5 月 5 日
天津大学	人工智能学院	2018 年 5 月
天津师范大学	人工智能学院	2018 年 4 月 25 日
河北工业大学	人工智能与数据科学学院	2018 年 4 月 23 日
山东科技大学	人工智能学院	2018 年 4 月 21 日
清华大学&南京市	南京图灵人工智能研究院	2018 年 4 月 20 日
长春理工大学	人工智能学院&人工智能研究院	2018 年 4 月 16 日
南京大学	人工智能学院	2018 年 3 月 6 日
重庆邮电大学	人工智能学院	2018 年 2 月 7 日
上海交通大学	人工智能研究院	2018 年 1 月 18 日
湖北工业大学	人工智能学院	2018 年 1 月 13 日
湖南工业大学	人工智能学院	2018 年 1 月
武昌理工学院	人工智能学院	2018 年
北京交通大学	人工智能学院	2017 年 12 月 13 日
西南政法大学	人工智能法学院&人工智能法律研究院	2017 年 12 月 6 日
苏州大学	人工智能研究院	2017 年 11 月 19 日
西安电子科技大学	人工智能学院	2017 年 11 月 2 日
中国科学院大学	人工智能技术学院	2017 年 5 月 28 日
中山大学	智能工程学院	2017 年 5 月 11 日
同济大学	人工智能研究院	2017 年 5 月 7 日
中国科学技术大学	类脑智能技术及应用国家工程实验室	2017 年 5 月
北京协和医学院	医学人工智能研究中心	2017 年 4 月
北京联合大学	机器人学院	2016 年 5 月
湖南大学	机器人学院	2016 年 4 月 26 日
复旦大学	类脑智能科学与技术研究院	2015 年 3 月
山西农业大学	智能工程学院	2013 年 12 月
西安工业大学	人工智能与仿真研究所	1994 年

版权声明

AMiner 研究报告版权为 AMiner 团队独家所有，拥有唯一著作权。AMiner 咨询产品是 AMiner 团队的研究与统计成果，其性质是供用户内部参考的资料。

AMiner 研究报告提供给订阅用户使用，仅限于用户内部使用。未获得 AMiner 团队授权，任何人和单位不得以任何方式在任何媒体上（包括互联网）公开发布、复制，且不得以任何方式将研究报告的内容提供给其他单位或个人使用。如引用、刊发，需注明出处为“AMiner.org”，且不得对本报告进行有悖原意的删节与修改。

AMiner 研究报告是基于 AMiner 团队及其研究员认可的研究资料，所有资料源自 AMiner 后台程序对大数据的自动分析得到，本研究报告仅作为参考，AMiner 团队不保证所分析得到的准确性和完整性，也不承担任何投资者因使用本产品与服务而产生的任何责任。

顾问

李涓子 唐杰 刘德兵

编写

张淼 徐菁 叶静芸

李鹤飞 钱达行 赵薇

数据支持

赵慧军 李亚坤

封面设计

钱达行



报告编号 AITR-21-01

