



中国科学技术大学

University of Science and Technology of China

大数据智能

i创会®
www.ichuanghui.org

吴枫

类脑智能技术及应用国家工程实验室主任

中科大信息学院执行院长

□ 大数据的成因

□ 大数据时代

□ 智能时代

□ 合肥市发展人工智能优势





一张图表的人类历史



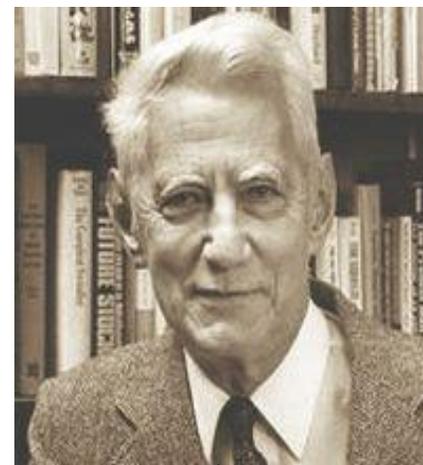
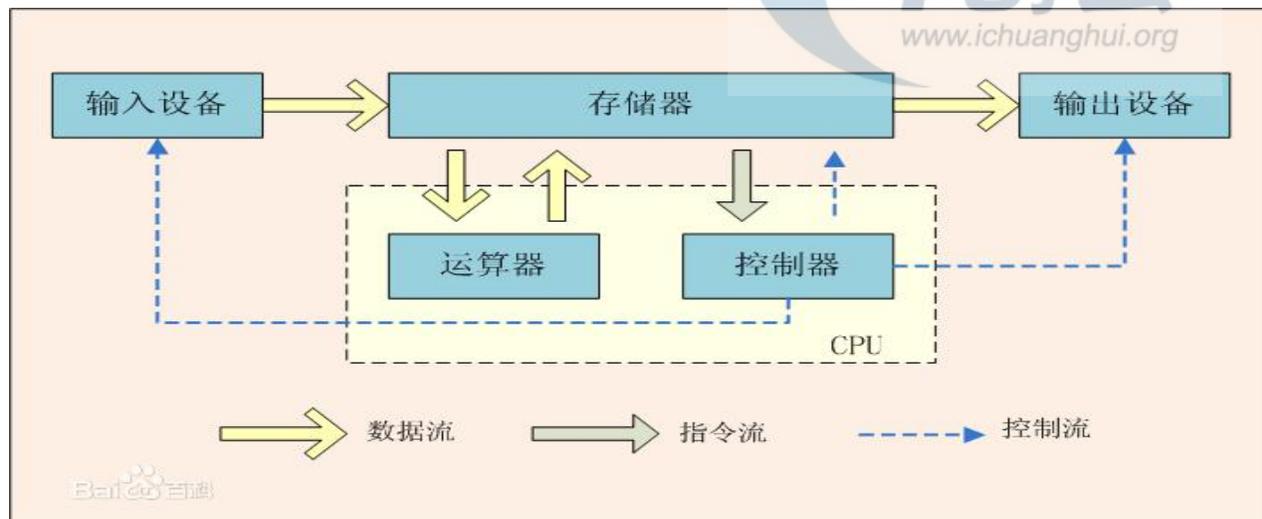
上世纪70-80年代

数千年以来，人类发展曲线一直呈现出平缓向上的趋势。社会的发展与进步及其缓慢，以至于几乎无法辨清。牲畜和农业、战争和帝国、哲学和宗教都没有发挥更大的影响力。但正好是在2000多年以前的时候，剧变发生了，人口和社会发展使得人类历史的发展曲线几乎弯曲了90度。

■ 香农信息论（1948）

- 熵 - 信息度量
- 通信信道的容量
- 无线通信发展

■ 冯诺依曼体系架构（1946）



香农



冯·诺依曼

互联网的发展



中国科学技术大学
University of Science and Technology of China

中国

投资与资本市场

73亿美元

1004亿美元

3月18日三家公司总市值近4500亿美元

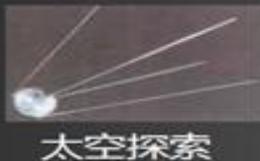
7月18日Google市值4780亿美元

图灵/香农

二战、技术应用



人类登上月球
互联网元年



太空探索



中国发出
第一封电子邮件

TCP/IP
协议



NSFNET
WWW



1939 1945 1957 1969 1983 1987 1991 1995 2000 2005 2010 2015 2045

互联网
用户规模

全球
中国

1600万
1万

4亿
2000万

10亿
1亿

35亿
7.5亿

技术进步
及融合

信息革命的
基础设施

计算机（终端/计算）

通信（联结/交互）

信息（数字化）

大型机—小型机—PC—Mobil—智能手机—传感器—云计算

电报—电话—有线—无线—宽带—数字移动—WIFI—量子通信

符号—文字—音频—视频—多媒体—大数据

手机终端

- 2016年全球智能手机用户达到23亿，占全球总人口的31%，
- 2020年预计将达到36亿
- 无线流量超过有线流量

	2013	2014	2015	2016	2017	2018
1. China*	436.1	519.7	574.2	624.7	672.1	704.1
2. US**	143.9	165.3	184.2	198.5	211.5	220.0
3. India	76.0	123.3	167.9	204.1	243.8	279.2
4. Japan	40.5	50.8	57.4	61.2	63.9	65.5
5. Russia	35.8	49.0	58.2	65.1	71.9	76.4



大数据的倍增器

- 搜索、导航、新闻、打车、...

□ 大数据的成因

□ 大数据时代

□ 智能时代

□ 对合肥市建议



□ Google

- 创建于1998年9月，使命是整合全球信息，使人人皆可访问并从中受益
- 通过对由超过50000万个变量和20亿个词汇组成的方程进行计算，PageRank 能够对网页的重要性做出客观的评价
- 2004年4月1日，Gmail邮箱提供每个用户10GB以上的存储空间
- 2005年6月20日，Google Maps提供全球地图服务，包括局部详细的卫星照片
- 2005年1月25日，Google公司推出Google Video服务，Google公司以16.5亿美元宣布收购YouTube
- 2007年11月5日，Google公司在全球公布了Android手机操作系统

□ Google 流感预测

- 谷歌在2008年推出的流感趋势系统监测全美的网络搜索
- 人们输入的搜索关键词代表了他们的即时需要，反映出用户情况
- 设计人员编入“一揽子”流感关键词，包括温度计、流感症状、肌肉疼痛、胸闷等
- 只要用户输入这些关键词，系统就会展开跟踪分析，创建地区流感图表和流感地图
- 利用这些搜索来提前9个星期预测可能与流感相关的就医量

□ 这项研究结果令人深入了解了大规模数据集分析给予技术公司的巨大力量

□ 阿里巴巴

- 创建于1999年9月，阿里巴巴零售平台2015财年交易额突破3万亿元，全球最大的零售交易平台
- 11月11日销售额从2009年的5200万到2017年1682亿，相比较2015年感恩节网购销售额仅为19.3亿美元，黑色星期五为33.4亿美元

截至2014年6月，我国网络购物用户规模达到3.32亿

截至2014年6月，我国使用网上支付的用户规模达到2.92亿



i创会®
www.ichuanghui.org



- “啤酒与尿布”
- 网上购物对实体店带来巨大冲击

□ 腾讯

- 创建于1998年11月，中国最大的社交和通信服务提供商
- 1999年2月，开通QQ，2004年4月注册用户数突破3亿大关，2014年4月11日，最高同时在线账户数突破2亿
- 2011年初推出微信，2015年微信月活跃账户数达6.97亿户，微信正在改变人们的沟通方式和生活方式

□ 社交网络对社会的影响力巨大

- 颜色革命 - 论坛、博客、维基百科、视频分享网站等作为教室，把Twitter、Facebook、YouTube作为活动的指挥中心
- 土耳其政变 - 埃尔多安打了一通 Facetime 视频通话，通过 CNN 土耳其台发表演说，鼓励土耳其人民上街反对政变；土耳其民众利用特殊的网络连接软件，把街头发生的一切通过 Twitter 上的图片、视频和 Periscope 直播拨给了全世界大部分国家

□ 滴滴

- 2012年6月6日创建，利用移动互联网和大数据技术，创新交通出行服务模式
- 滴滴打车利用移动互联网特点，颠覆了路边拦车概念，改变传统出租司机等客方式
- 2014年8月，滴滴专车上线，主打中高端商务用车服务市场，车辆和司机可以来自于个人
- 2015年滴滴出行全平台订单总量达到14.3亿，相当于美国2015年所有出租车订单量的近两倍

□ 共享经济的开始

- 《网络预约出租汽车经营服务管理暂行办法》
- 满足条件的私家车可按一定程序转为网约车，从事专车运营

□ Airbnb (AirBed and Breakfast, 空中食宿)

- 2008年8月创建, Airbnb是一个旅行房屋租赁社区, 用户可通过网络或手机应用程序发布、搜索度假房屋租赁信息并完成在线预定程序。
- Airbnb用户遍布190个国家近34000个城市, 发布的房屋租赁信息达到5万条

□ 共享经济

- 有空闲的资源就可以出租, 就可以提高闲置资源利用率从而获得最大收益

Airbnb已经变成了一种商业模式

这些归类在 Airbnb for X 的公司, 意思是出租空间



Hipcamp
找漂亮露营地
的服务



Boatbound
私人船只
出租服务



Airpnp
分享
自家厕所



Breather
找个地方静静



SPOT Park
停车位分享



SupperShare
家中厨房
出租服务



Misterbnb
同性恋酒店
/住宿分享



Root
家中柜子
和车库出租



3D Hubs
3D打印机
出租

大数据时代的思考



中国科学技术大学
University of Science and Technology of China

□ 大公司拥有大数据

大数据种类	拥有者
搜索数据	百度、谷歌
购物数据	阿里巴巴、亚马逊、京东
社交数据	腾讯、Facebook
地图	高德、百度、谷歌
交通出行	滴滴、Uber
...	...

□ 挑战

- 整合
- 创造
- 管理

□ 大数据的成因

□ 大数据时代

□ 智能时代

□ 合肥市在人工智能的优势



□ 人类下一个技术社会形态

智能化



信息化



工业化



- 主体：智能技术
- 优势：发现和利用知识
- 资源：大数据、深度神经网络

智能机器网络化群体协作
机器智能与人类智能合作

- 主体：信息技术
- 优势：生产和获取信息
- 弊端：信息处理和计算欠智能

机器无法有效地辅助和替代人类

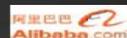
发展预测



中国科学技术大学
University of Science and Technology of China

中国

投资与资本市场



3月18日三家公司总市值近**4500亿美元**

73亿美元

1004亿美元

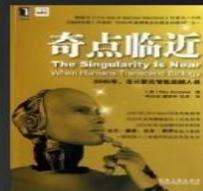
7月18日Google市值**4780亿美元**



人类登上月球
互联网元年



中国发出
第一封电子邮件



图灵/香农

二战\技术应用



太空探索

TCP/IP
协议



1939 1945 1957 1969 1983 1987 1991 1995 2000 2005 2010 2015 2045

互联网
用户规模

全球
中国

1600万

4亿

10亿

35亿

1万

2000万

1亿

7.5亿

技术进步
及融合

信息革命的
基础设施

计算机 (终端/计算)

通信 (联结/交互)

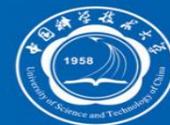
信息 (数字化)

大型机—小型机—PC—Mobil—智能手机—传感器—云计算

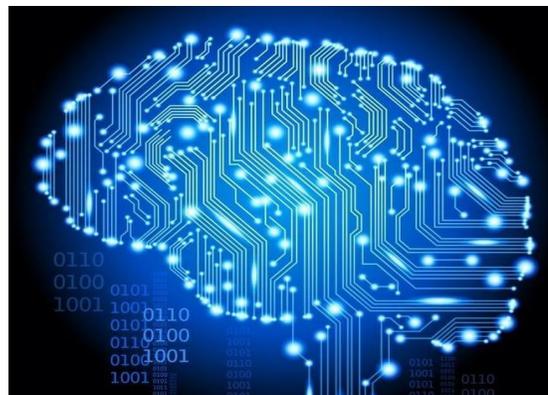
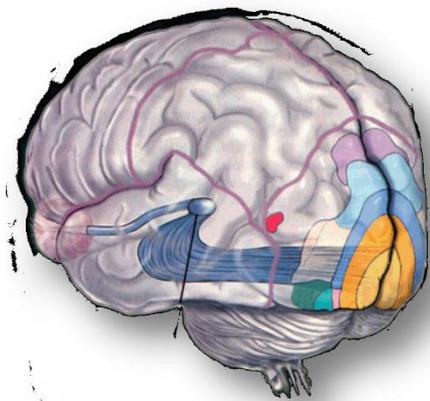
电报—电话—有线—无线—宽带—数字移动—WIFI—量子通信

符号—文字—音频—视频—多媒体—大数据

两大技术途径



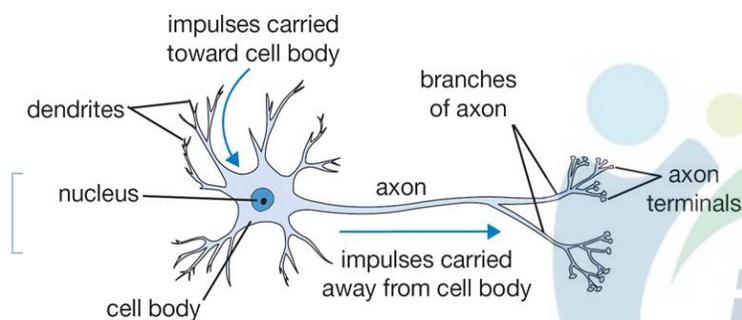
中国科学技术大学
University of Science and Technology of China



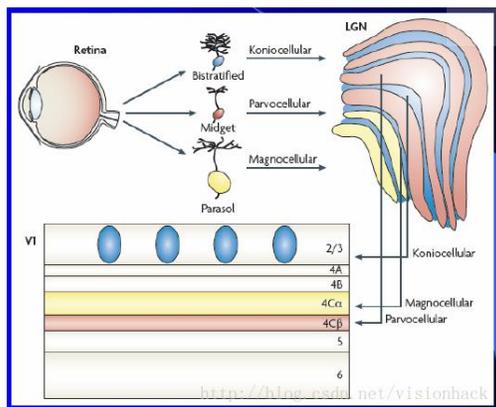
类脑计算是通向“强人工智能”之路，在脑科学基础研究重大投入，并与信息科学形成紧密融合，是主攻方向

大数据人工智能可以充分利用机器学习和大数据方向的最新进展，能够在近期取得阶段性突破，面临严峻挑战

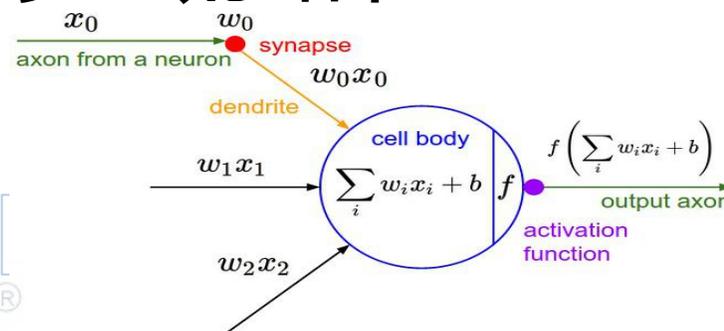
深度神经网络与大数据的结合成为当前解决实际人工智能问题的主流路径



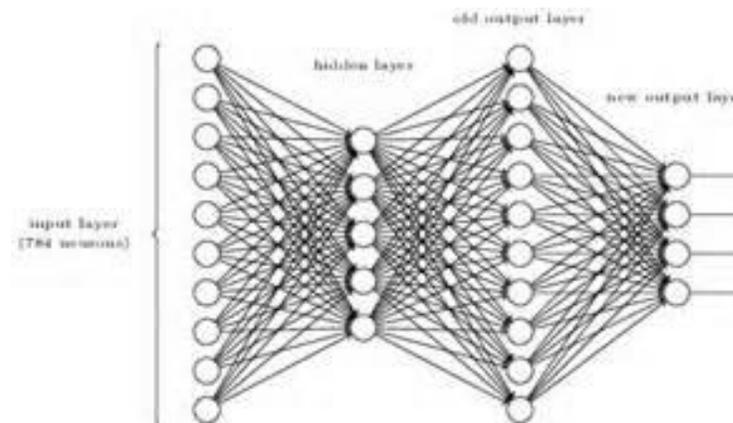
大脑神经元细胞



大脑皮层V1区分层结构



人工神经元



人工神经网络

i创会
www.ichuanghui.org

图像识别



IMAGENET

- 百万级的训练样本
- 百万级的参数
- 模型越来越复杂
- 计算复杂度越来越高
- 存储开销大

Revolution of Depth

152层

3.57

22层

6.7

19层

7.3

11.7

8层

16.4

8层

25.8

浅层

28.2

IISVRC'15

IISVRC'14

IISVRC'14

IISVRC'13

IISVRC'12

IISVRC'11

IISVRC'10

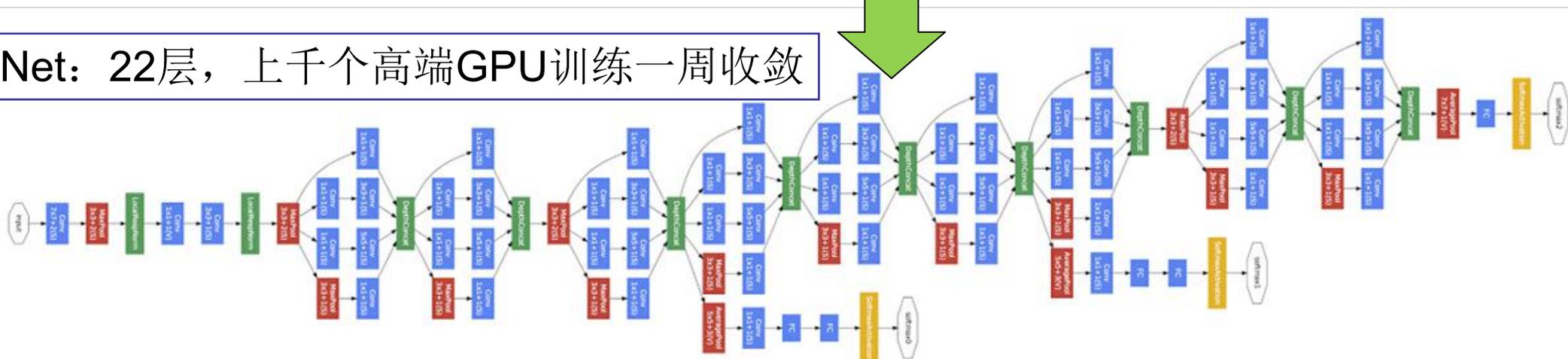
ResNet

GoogleNet

VGG

AlexNet

GoogleNet: 22层, 上千个高端GPU训练一周收敛



传统语音模式识别



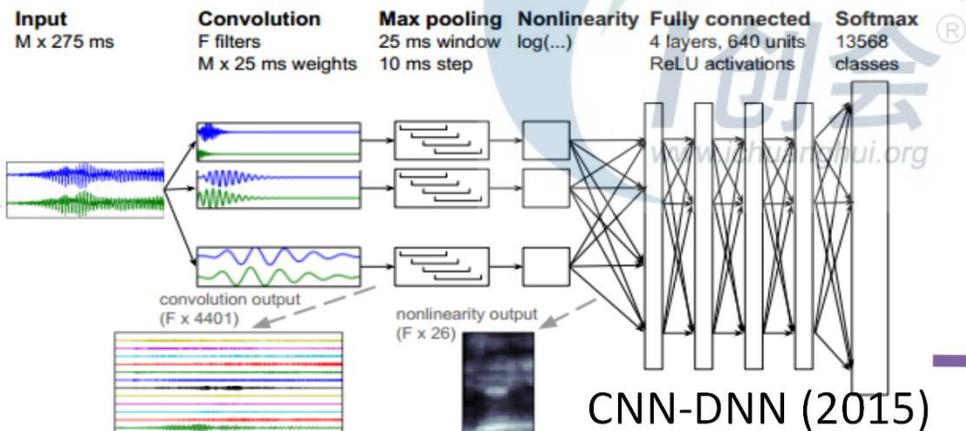
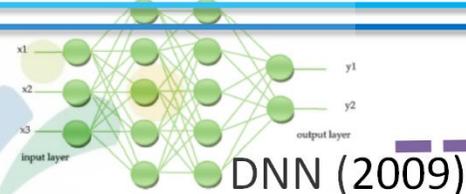
识别文本

频谱表征

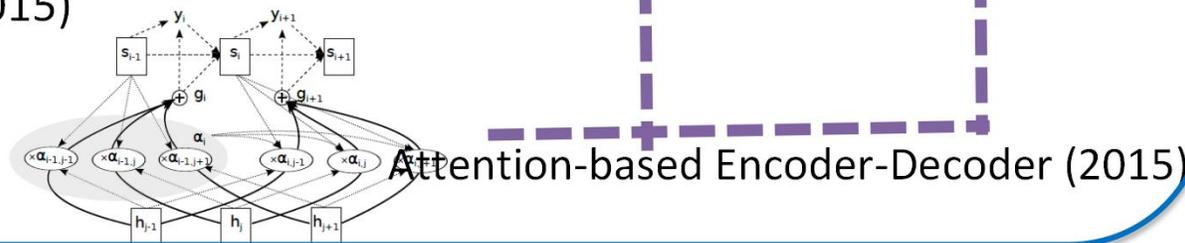
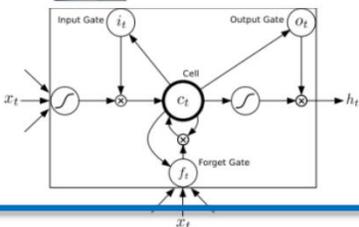
特征提取

声学模型

语言模型



LSTM-CTC (2013)



- 微软（亚洲）互联网工程院^[1] 在2014年5月29日发布一款人工智能伴侣虚拟机器人，并取名“微软小冰”
 - “微软小冰”集合了中国近7亿网民多年来积累的、全部公开的文献记录
 - 凭借微软在大数据、自然语义分析、机器学习和深度神经网络方面的技术积累
 - 精炼为几千万条真实而有趣的语料库，通过理解对话的语境与语义，实现了超越简单人机问答的自然交互



□ 计算机是如何利用大数据和运算能力

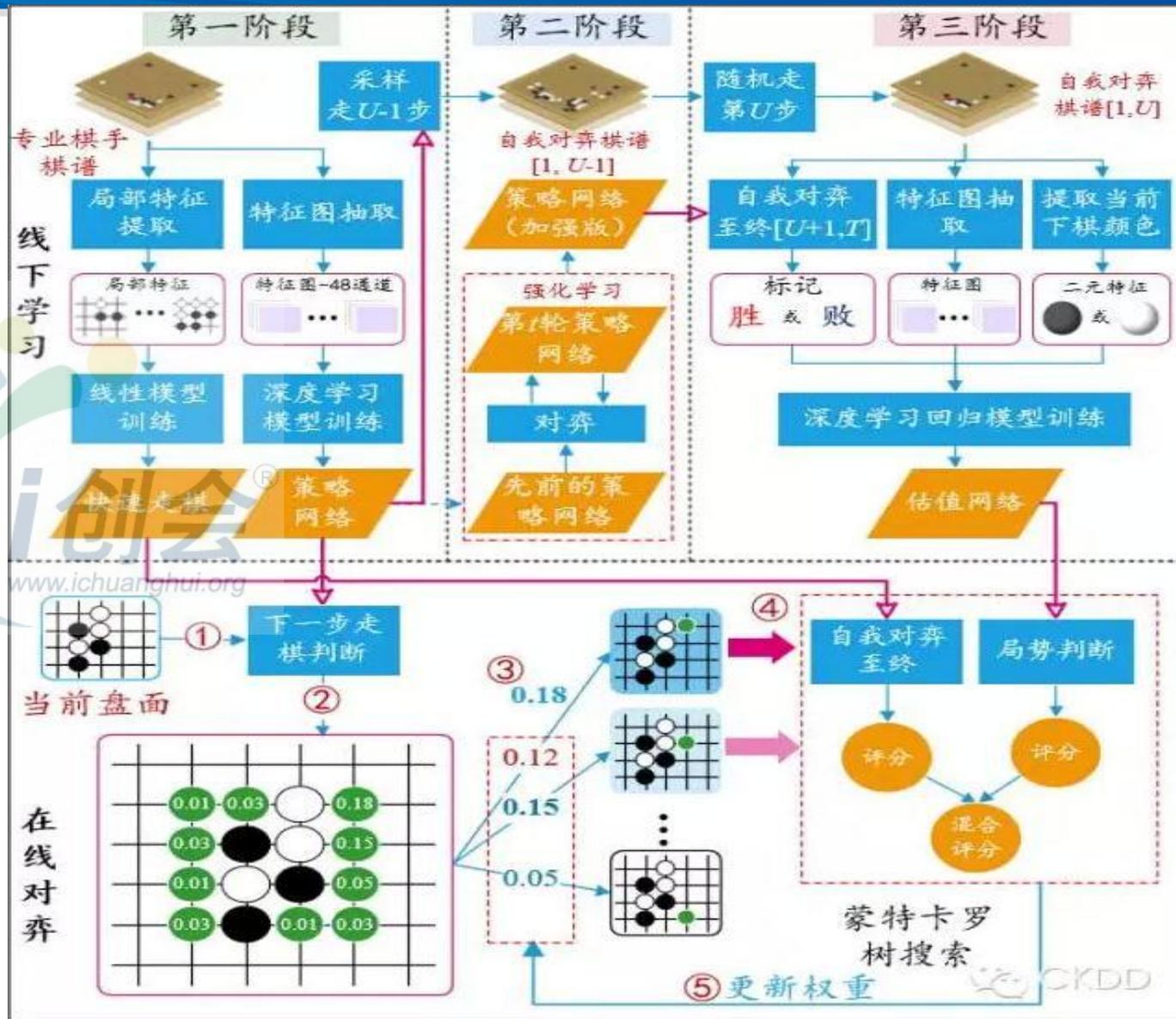
- 学习了人类围棋的规则，但是思考方式不完全和人类相同
- 学习了16万盘人类对弈，生成了3000万盘进行自我学习
- 深度神经网络的统计方法与规则搜索（计算能力）相结合



围棋



DCNN在围棋程序中的成功应用有效说明了深度神经网络可以很好的应用在机器辅助决策系统中



□ 大数据的成因

□ 大数据时代

□ 智能时代

□ 对合肥市建议



□ 任正非

- 华为技术有限公司主要创始人
- 华为技术有限公司总裁



□ 2016年科技大会发言

- 从技术的角度来看，未来二、三十年人类社会将演变成为一个智能社会，其深度和广度我们还想象不到
- 越是前途不确定，越需要创造，这也给千百万家企业公司提供了千载难逢的机会

智能社会的演变



中国科学技术大学
University of Science and Technology of China

□ 高校

- 中科大信息学院，计算机学院，国家示范性微电子学院，网络空间安全学院，大数据学院，软件学院
- 组建人工智能与大数据学部

□ 企业

- 讯飞、商汤、科大智能等人工智能领跑企业

□ 平台

- 类脑智能国家工程实验室
- 语音语言国家工程实验室

□ 类似硅谷和中关村，产学研一体的人工智能产业园





中国科学技术大学
University of Science and Technology of China

