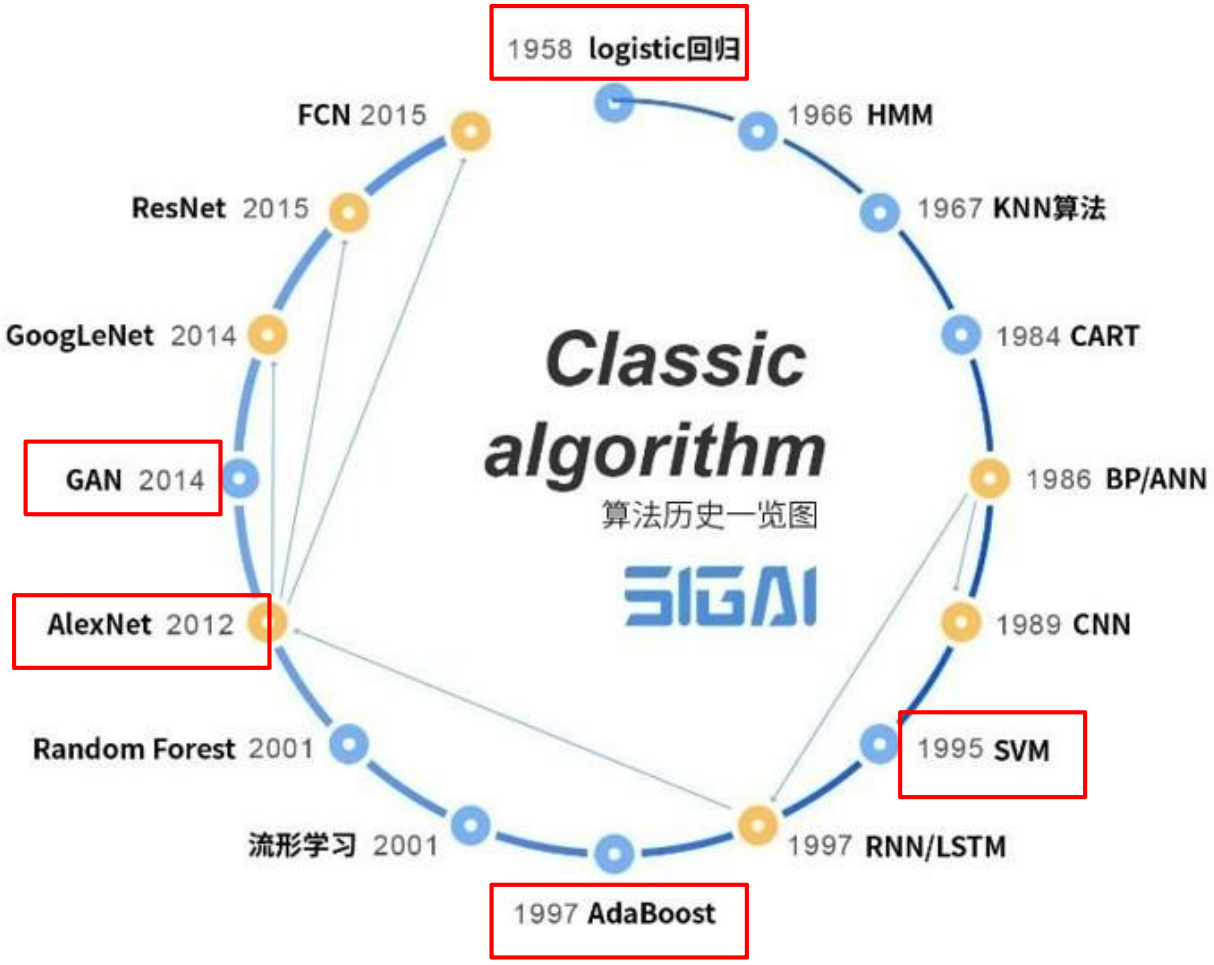
人工智能原理与技术

一、绪论

2. 深度学习概述

机器学习模型

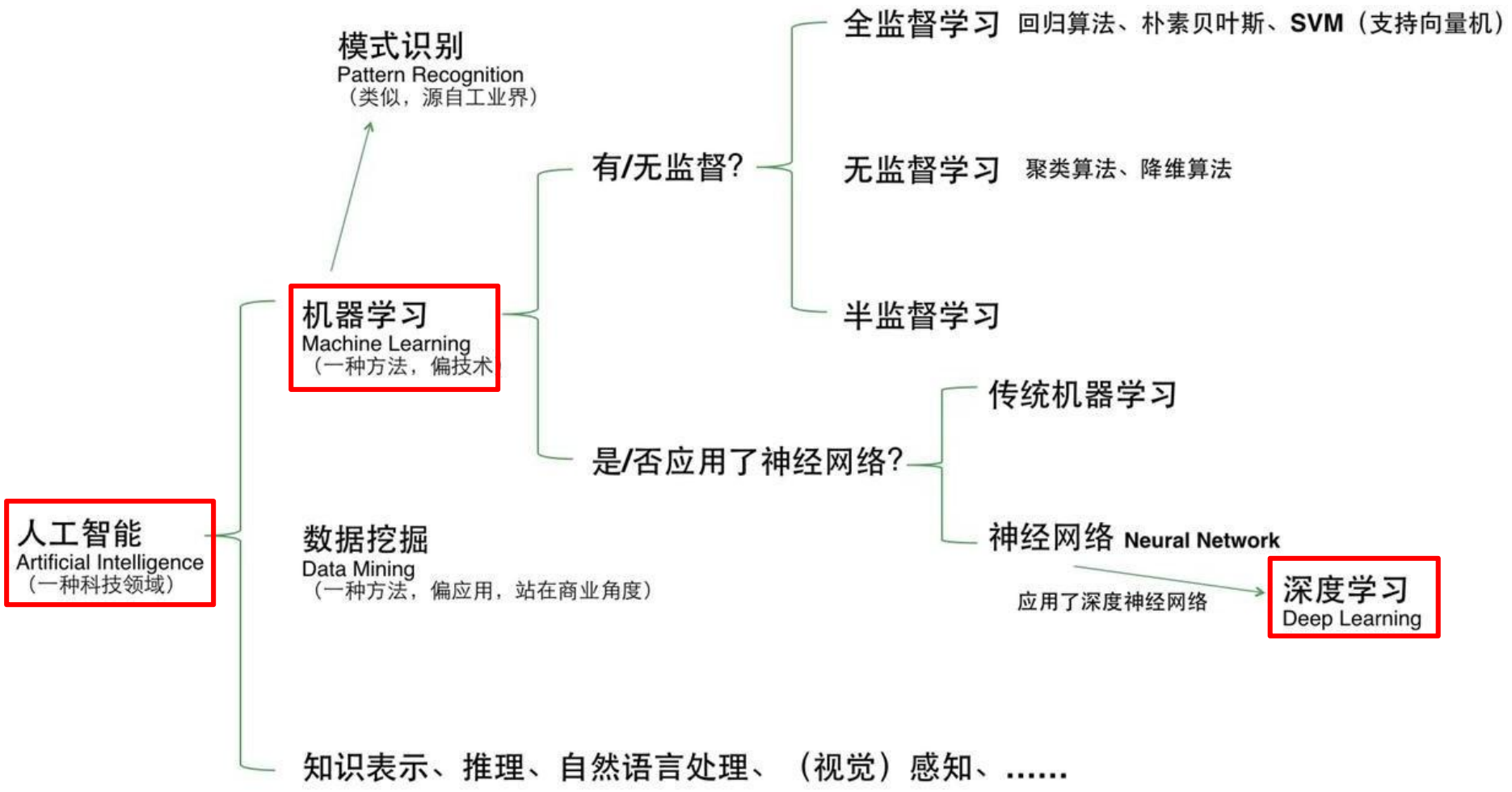


从2012年以一来，主流算法已经被深度学习占领了

深度学习无所不在

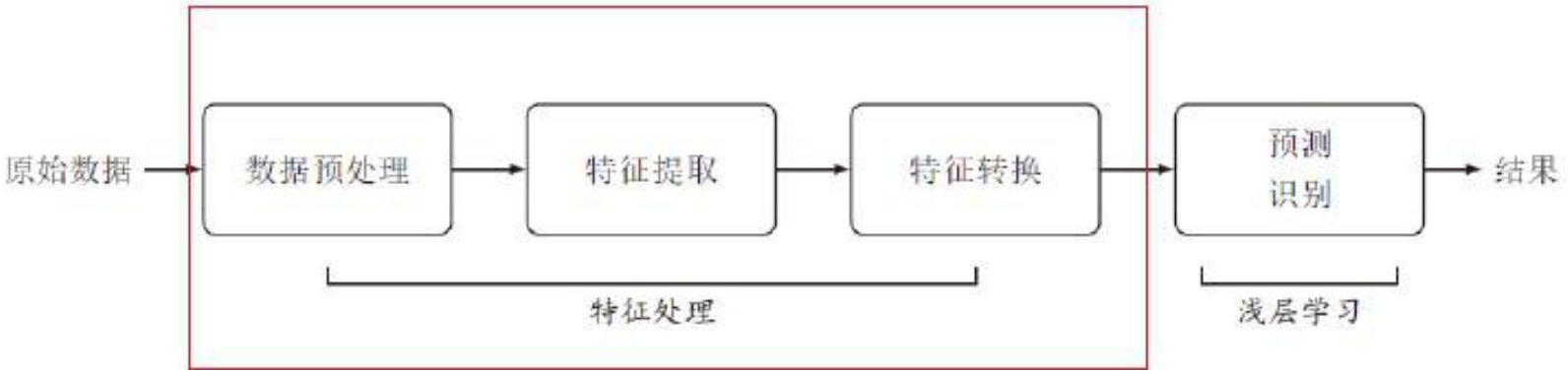

DEEP LEARNING EVERYWHERE
INTERNET & CLOUD
Image Classification
Speech Recognition
Language Translation
Language Processing
Sentiment Analysis
Recommendation
MEDICINE & BIOLOGY
Cancer Cell Detection
Diabetic Grading
Drug Discovery
MEDIA & ENTERTAINMENT
Video Captioning
Video Search
Real Time Translation
SECURITY & DEFENSE
Face Detection
Video Surveillance
Satellite Imagery
AUTONOMOUS MACHINES
Pedestrian Detection
Lane Tracking
Recognize Traffic Sign

人工智能>机器学习>深度学习



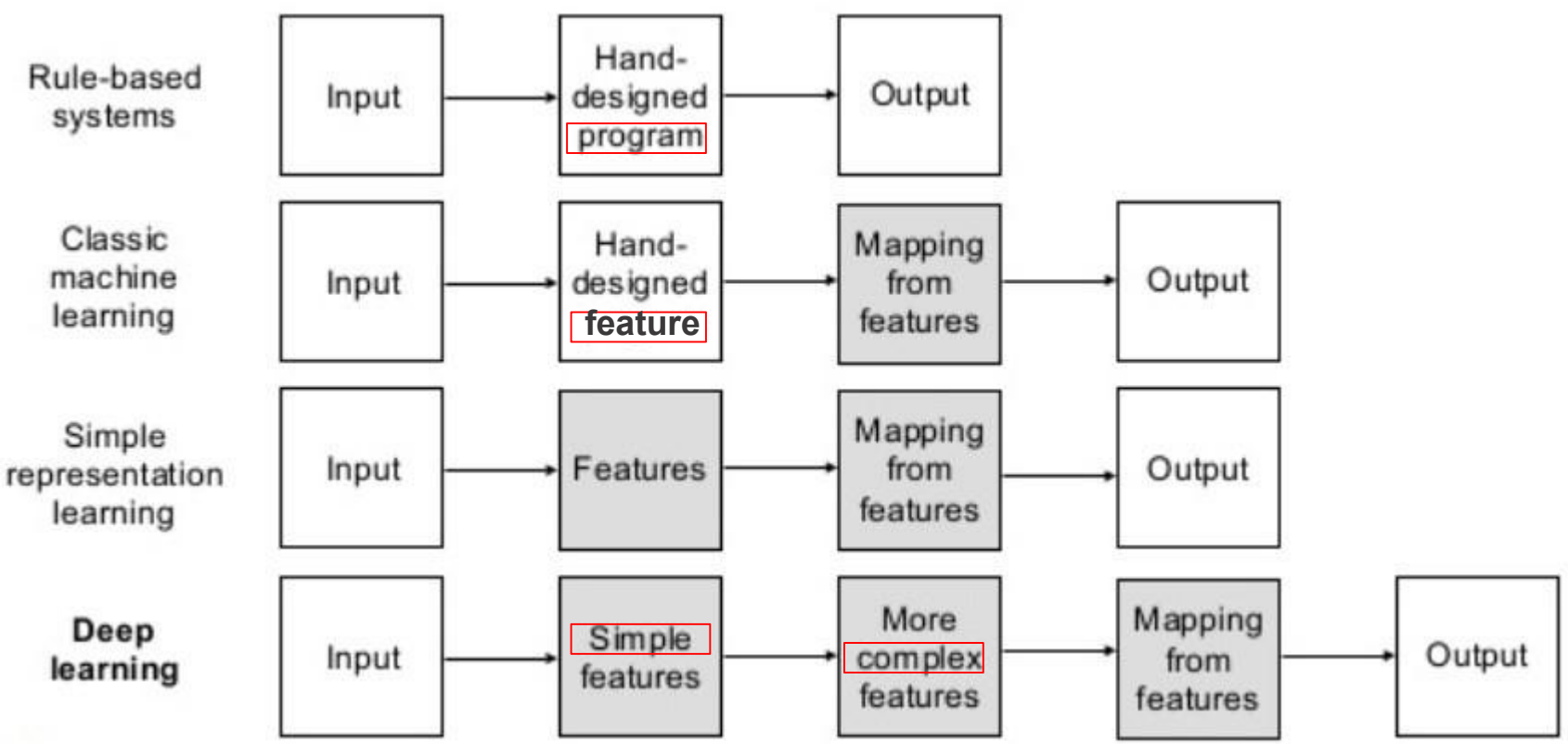
传统机器学习：人工设计特征

- 在实际应用中，特征往往比分类器更重要
  - 预处理：经过数据的预处理，如去除噪声等。比如在文本分类中，去除停用词等。
  - 特征提取：从原始数据中提取一些有效的特征。比如在图像分类中，提取边缘、尺度不变特征变换特征等。
  - 特征转换：对特征进行一定的加工，比如降维和升维。降维包括
    - 特征抽取（Feature Extraction）：PCA、LDA
    - 特征选择（Feature Selection）：互信息、TF-IDF



举例子：眼睛的检测，人脸的检测

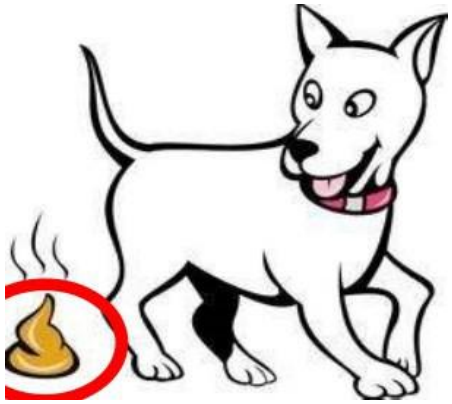
传统机器学习 VS 深度学习



思考：为什么需要自动学习特征？

传统机器学习 VS 深度学习

从巡逻机器人的狗便便检测需求说起...



本页来自山世光老师PPT

传统机器学习 VS 深度学习

前深度学习时代，我们这么做...

- 步骤1：花几天时间收集并标注几百张便便图
- 步骤2：花几个月观察便便图，并绞尽脑汁选择或设计一些特征
  - 形状，颜色，纹理；SIFT，HOG，Gabor，LBP，Haar...
- 步骤3：用某种分类器训练和测试，结果不好回到步骤2

\+ 专家知识驱动的特征设计 \+ 专家选择的 的分类器

思考： 这么一个很小的任务到接近实用化，用多长时间？

本页来自山世光老师PPT

传统机器学习 VS 深度学习

前深度学习时代，我们这么做...

- 步骤1：花几天时间收集并标注几百张便便图
- 步骤2：花几个月观察便便图，并绞尽脑汁[illegible]设计一些特征
  - 形状，颜色，纹理；SIF[illegible]..
- 步骤3：用某种[illegible]

需要多久？1年甚至更久
行人检测用了10年
人脸检测用了20年...

[illegible]驱动的特征设计

+

专家选择的 的分类器

本页来自山世光老师PPT

传统机器学习 VS 深度学习

深度学习时代，我们这么做...

- 步骤1：**花几个星期时间收集并标注(框出狗便便位置)数万张便便图**
- **步骤2：花1个星期，挑几个深度模型，选几组模型超参数**
- 步骤3：交给机器绞尽脑汁优化学模型中的数千万/数亿权重参数

\+

专家选择 深度模型

\+

机器优化 深度模型

本页来自山世光老师PPT

传统机器学习 VS 深度学习

深度学习时代，我们这么做...

- 步骤1：花几个星期时间收集并标注(框出狗屎位置)数万张便便图
- 步骤2：花1个星期，挑几个深[illegible]招参数
- 步骤3：交给机器[illegible]亿权重参数

需要多久？2个月

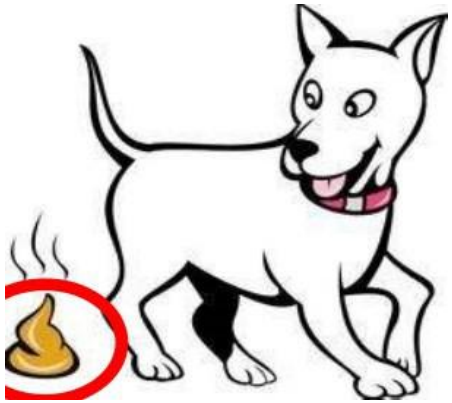
专家选择深度模型

机器优化 深度模型

本页来自山世光老师PPT

传统机器学习 VS 深度学习

后深度学习时代，我们怎么做？



本页来自山世光老师PPT

传统机器学习 VS 深度学习

后深度学习时代，我们期待这么做...

- 步骤1：花几分钟时间收集并标注几张便便图
- 步骤2：交给机器绞尽脑汁完成任务

+

机器选择和 优化模型

本页来自山世光老师PPT

传统机器学习 VS 深度学习

10月19日，谷歌旗下人工智能团队DeepMind在今天对外发布了一款zhi全新的AlphaGo程序

这款软件名为AlphaGo Zero，与之前击败了李世石的 AlphaGo 进行对弈，胜率高达100%

脱离人类干预

从下棋说起

Game Playing Machine: from Chess to Go

1769年匈牙利作家兼发明家Wolfgang von Kempelen建造了机器人TheTurk，能够跟国际象棋高手对弈，但最终谜底揭开，机器人之所以会下棋是因为箱子里藏着一个人。

《2001太空漫游》（2001: A Space Odyssey）是1968年上映的，由斯坦利·库布里克执导，根据科幻小说家亚瑟·克拉克小说改编的科幻电影。片中一个情节是机器人HAL与人Frank下国际象棋，人类在机器面前甘拜下风

HAL→IBM

TIME

1997 年在 IBM 的深蓝战胜国际象棋世界冠军 Kasparov后，时代杂志提出了一项新的挑战：让计算机与人类下围棋吧，它获胜的机会很小。“**计算机要在围棋上战胜人类，还要再过一百年，甚至更长的时间**”时代杂志的文章中这些写道。

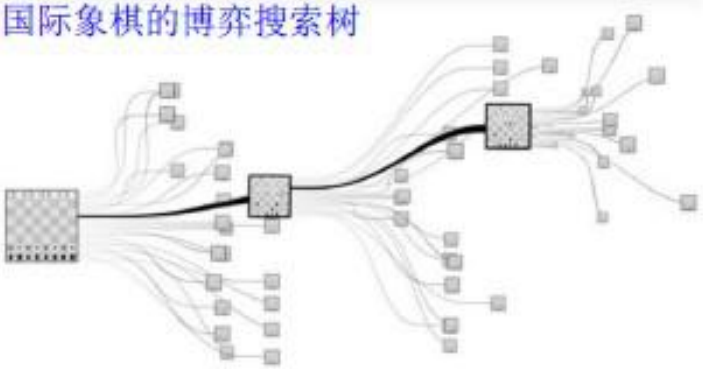
20 Years, NOT 100 Years

Deep Blue依靠运算能力搜索最佳走棋，而AlphaGo更像人类，通过自学来提高棋力。

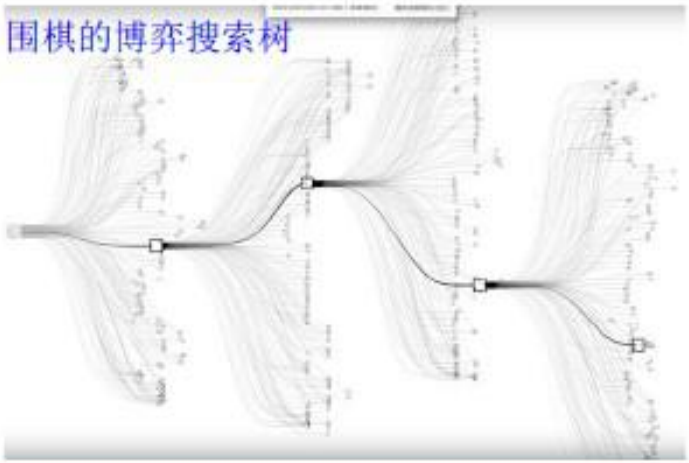
AlphaGo Shock：AlphaGo冲击波在韩国，投资30亿美元的5年计划，研究人工智能。

Deeper Blue *vs.* AlphaGo

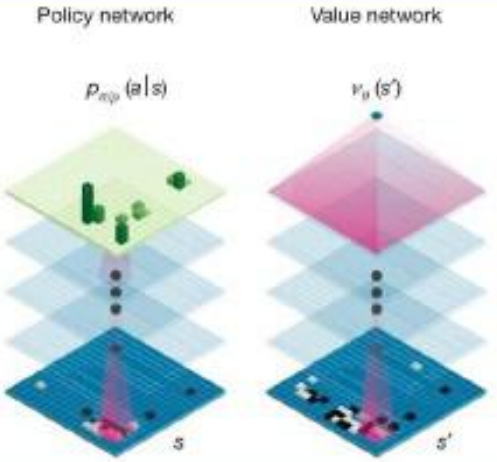
国际象棋的博弈搜索树



围棋的博弈搜索树



两个网络：
走子网络
评估网络

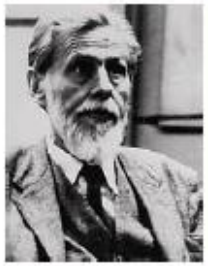


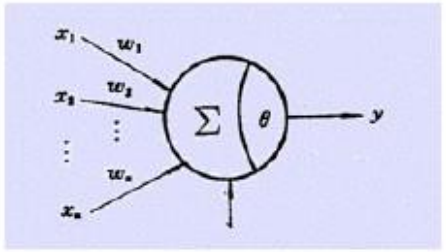
Hinton 在访谈中说：我们会得到非常好的证据，最终学术界会改变观点. 也许那些和你争论的科学家永远不会回心转意，但是年轻一代会，这就是在深度学习领域正在发生的事. 传统的人工智能领域的老家伙们还是不信，但是年轻一代的研究生们都看到了事情在朝什么方向发展.

一切的开始：感知器

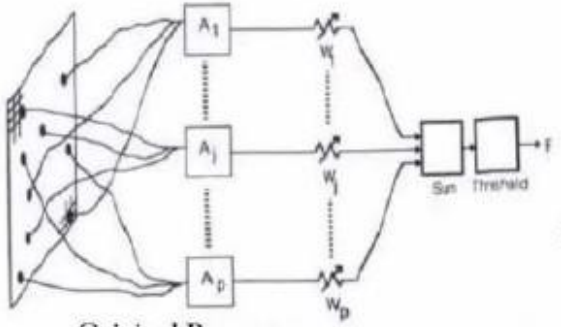
Rosenblatt & Perceptron

- 计算模型：1943年最初由Warren McCulloch 和Walter Pitts 提出
- 感知器 (Perceptron)：康奈尔大学 Frank Rosenblatt 1957年提出
- Perceptron是第一个具有自组织自学习能力的数学模型
- Rosenblatt 乐观预测：感知器最终可以 “学习, 做决定, 翻译语言”
- 感知器技术六十年代一度走红，美国海军曾出资支持，期望它 “以后可以自己走, 说话, 看, 读, 自我复制, 甚至拥有自我意识”





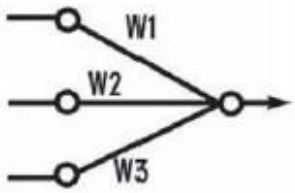
Perceptron (1957)



Original Perceptron

(From Perceptrons by M. L Minsky and S. Papert, 1969, Cambridge, MA: MIT Press. Copyright 1969 by MIT Press.

Frank Rosenblatt
(1928-1971)

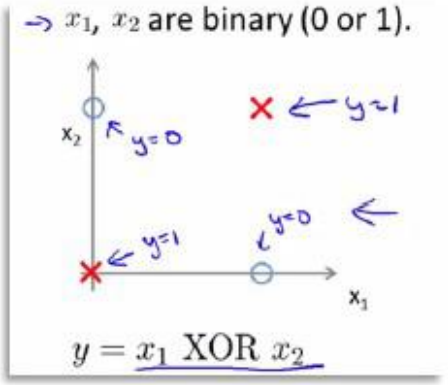


Simplified model:

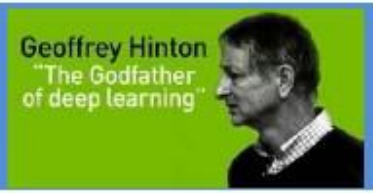
23

Rosenblatt *vs.* Minsky

- Rosenblatt 和 Minsky是间隔一级的高中校友。但是六十年代，两个人在感知器的问题上展开了长时间的激辩：R认为感应器将无所不能，M则认为它应用有限
- 1969 年, Marvin Minsky 和 Seymour Papert 出版了新书: "感知器: 计算几何简介". 书中论证了感知器模型的两个关键问题:
  - 第一, 单层的神经网络无法解决不可线性划分的问题, 典型例子如异或门
  - 第二, 更致命的问题是, 当时的电脑完全没有能力完成神经网络模型所需要的超大的计算量
- 此后的十几年, 以神经网络为基础的人工智能研究进入低潮, 相关项目长期无法得到政府经费支持, 这段时间被称为业界的核冬天
- Rosenblatt 自己则没有见证日后神经网络研究的复兴。1971年他43岁生日时, 不幸在海上开船时因为事故而丧生



Geoffrey Hinton & NNs



- 1970年, 当神经网络研究的第一个寒冬降临时, 在英国的爱丁堡大学, 一位23岁的年轻人 Geoffrey Hinton, 刚刚获得心理学的学士学位.
- Hinton 六十年代还是中学生时就对脑科学着迷。当时一个同学给他介绍关于大脑记忆的理论是: 大脑对于事物和概念的记忆, 不是存储在某个单一的地点, 而是像全息照片一样, 分布式地存在于一个巨大的神经元的网络里.
- 分布式表征 (Distributed Representation)和传统的局部表征 (Localized Rep.) 相比
  - 存储效率高很多：线性增加的神经元数目, 可以表达指数级增加的大量不同概念
  - 鲁棒性好：即使局部出现硬件故障, 信息的表达不会受到根本性的破坏
- 这个理念让 Hinton 顿悟, 使他40多年来一直在神经网络研究的领域里坚持
  - 本科毕业后, Hinton 选择继续在爱丁堡大学读研, 把人工智能作为自己的博士研究方向
  - 1978 年, Hinton在爱丁堡获得博士学位后, 来到美国继续他的研究工作

Rumelhart &BP Algorithm

- 神经网络被 Minsky 诟病的问题：巨大的计算量; XOR问题
- 传统的感知器用所谓 "梯度下降"的算法纠错时, 耗费的计算量和神经元数目的平方成正比. 当神经元数目增多, 庞大的计算量是当时的硬件无法胜任的
- 1986年7月, Hinton 和 *David Rumelhart* 合作在Nature杂志上发表论文：Learning Representations by Back-propagating Errors, 第一次系统简洁地阐述BP算法及其应用
  - 反向传播算法把纠错的运算量下降到只和神经元数目本身成正比
  - BP算法通过在神经网络里增加一个所谓隐层 (hidden layer), 解决了XOR难题
  - 使用了BP算法的神经网络在做如形状识别之类的简单工作时, 效率比感知器大大提高，八十年代末计算机的运行速度, 也比二十年前高了几个数量级
- 神经网络及其应用的研究开始复苏！

Yann Lecun & CNN

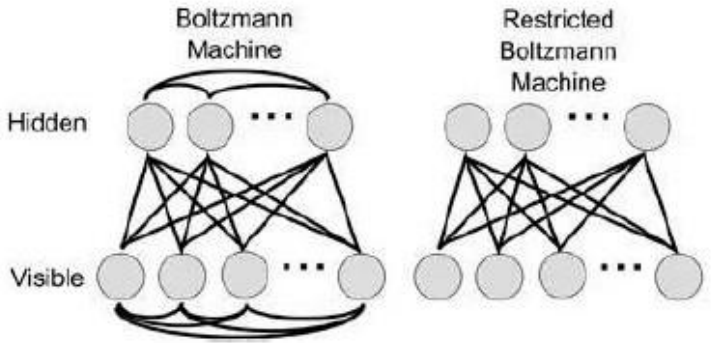
- Yann Lecun于1960年出生于巴黎
- 1987年在法国获得博士学位后，他曾追随 Hinton 教授到多伦多大学做了一年博士后的工作，随后搬到新泽西州的 Bell Lab 继续研究工作
- 在 Bell Lab，Lecun1989年发表了论文，“反向传播算法在手写邮政编码上的应用”，他用美国邮政系统提供的近万个手写数字的样本来训练神经网络系统，训练好的系统在独立的测试样本中，错误率只有5%
- Lecun进一步运用一种叫做“卷积神经网络” (Convolutional Neural Networks) 的技术，开发出商业软件，用于读取银行支票上的手写数字，这个支票识别系统在九十年代末占据了美国接近20%的市场
- 此时就在Bell Lab，Yann Lecun临近办公室的一个同事Vladmir Vapnik的工作，又把神经网络的研究带入第二个寒冬！

Hinton & Deep Learning

- 2003年, Geoffrey Hinton, 还在多伦多大学, 在神经网络的领域苦苦坚守
- 2003年在温哥华大都会酒店, 以Hinton 为首的十五名来自各地的不同专业的科学家, 和加拿大先进研究院 (Canadian Institute of Advanced Research, CIFAR) 的基金管理负责人 Melvin Silverman 交谈
  - Silverman 问大家, 为什么 CIFAR 要支持他们的研究项目
  - 计算神经科学研究者, Sebastian Sung (现为普林斯顿大学教授) 回答道： "喔, 因为我们有点古怪. 如果CIFAR 要跳出自己的舒适区, 寻找一个高风险, 极具探索性的团体, 就应当资助我们了!"
  - 最终 CIFAR 同意从2004年开始资助这个团体十年, 总额一千万加元. CIFAR 成为当时世界上唯一支持神经网络研究的机构
- Hinton 拿到资金支持不久, 做的第一件事, 就是把"神经网络" 改名换姓为 "深度学习"
- 此后，Hinton 的同事不时会听到他突然在办公室大叫： "我知道人脑是如何工作的了!".

DBN & RBM

- 2006年Hinton 和合作者发表论文：A Fast Algorithm for Deep Belief Nets
- 算法上借用了统计力学中"玻尔兹曼分布"概念 ：一个微粒在某个状态的几率, 和那个状态的能量的指数成反比, 和它的温度的倒数之指数成反比。使用所谓的"限制玻尔兹曼机" (RBM)来学习
  - RBM 相当于一个两层网络, 同一层神经元之间不可连接 (所以叫 "限制"), 可以对神经网络实现 "unsupervised training" 。深度置信网络DBN就是几层 RBM 叠加在一起
  - RBM 可以从输入数据进行预先训练, 自己发现重要的特征, 对神经网络连接的权重进行有效的初始化, 被称作：特征提取器 (Feature Extractor)或自动编码器 (Autoencoder)
  - Hinton 指出：深度学习的突破除了计算蛮力的大幅度提高以外, 聪明有效地对网络链接权重的初始化也是一个重要原因
  - 经过六万个MNIST 数据库的图像训练后, 对于一万个测试图像的识别错误率最低降到了只有 1.25%



Andrew Y. Ng & GPU

- 2007年之前，用GPU编程缺乏一个简单的软件接口，编程繁琐，Debug困难
  2007年 Nvidia 推出 CUDA 的GPU 软件接口后才真正改善
- 2009年6月，斯坦福大学的 Rajat Raina 和吴恩达合作发表论文：Large-scale Deep Unsupervised Learning using Graphic Processors (ICML09)；论文采用DBNs模型和稀疏编码(Sparse Coding)，模型参数达到一亿（与Hinton模型参数的对比见下表）
- 论文结果显示：使用GPU运行速度和用传统双核CPU相比，最快时要快近70倍，在一个四层，一亿个参数的深信度网络上使用GPU把程序运行时间从几周降到一天

Published source	Application	Params
Hinton et al., 2006	Digit images	1.6mn
Hinton & Salakhutdinov	Face images	3.8mn
Salakhutdinov & Hinton	Sem. hashing	2.6mn
Ranzato & Szummer	Text	3mn
Our model		100mn

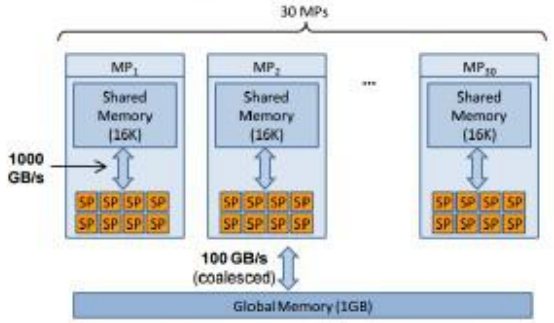
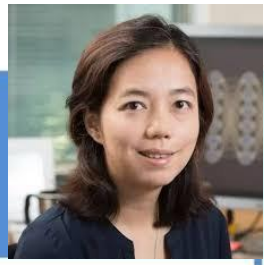


Figure 1. Simplified schematic for the Nvidia GeForce GTX 280 graphics card, with 240 total cores (30 multiprocessors with 8 stream processors each).

Jen-Hsun Huang & GPU

- 黄仁勋, 1963年出生于台湾
- 1993年从斯坦福大学硕士毕业后不久创立了 Nvidia
- Nvidia 起家时做的是图像处理的芯片, 主要面对电脑游戏市场. 1999 年Nvidia推销自己的 Geforce 256 芯片时, 发明了 GPU (Graphics Processing Unit)这个名词
- GPU 的主要任务, 是要在最短时间内显示上百万、千万甚至更多的像素. 这在电脑游戏中是最核心的需求. 这个计算工作的核心特点, 是要同时并行处理海量的数据
- 传统的 CPU 芯片架构, 关注点不在并行处理, 一次只能同时做一两个加减法运算. 而 GPU 在最底层的算术逻辑单元 (ALU, Arithmetic Logic Unit), 是基于所谓的 Single Instruction Multiple Data (单指令多数据流)的架构, 擅长对于大批量数据并行处理
- 一个 GPU, 往往包含几百个 ALU, 并行计算能力极高. 所以尽管 GPU 内核的时钟速度往往比 CPU的还要慢, 但对大规模并行处理的计算工作, 速度比 CPU 快许多
- 神经网络的计算工作, 本质上就是大量的矩阵计算的操作, 因此特别适合于使用 GPU

Big Data: ImageNet



- 2009年, 一群在普林斯顿大学计算机系的华人学者 (第一作者为 Jia Deng)发表了论文：ImageNet: A large scale hierarchical image database, 宣布建立了第一个超大型图像数据库供计算机视觉研究者使用
- 数据库建立之初,包含了320万个图像. 它的目的, 是要把英文里的8万个名词,每个词收集5百到1千个高清图片,存放到数据库里. 最终达到5千万以上的图像
- 2010 年, 以 ImageNet 为基础的大型图像识别竞赛, ImageNet Large Scale Visual Recognition Challenge 2010 (ILSVRC2010) 第一次举办

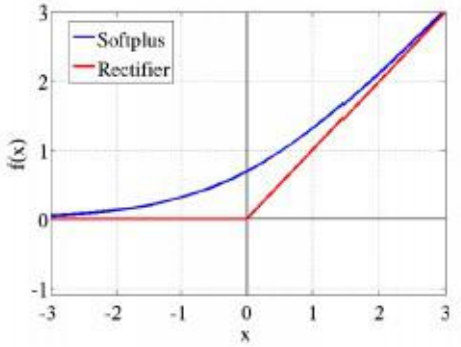
- 竞赛最初的规则：以数据库内120万个图像为训练样本. 这些图像从属于1千多个不同的类别, 都被手工标记。
经过训练的程序, 再用于5万个测试图像评估分类准确率

Image Classification : ILSVRC竞赛

- Top Five Category：计算机会对图像的分类，答出最有可能的头五个类别，如果正确答案都不在里面即为错误
- 2010年冠军：NEC 和伊利诺伊大学香槟分校的联合团队，用支持向量机 (SVM) 的技术，识别分类的错误率为 28%
- 2011年冠军：用 Fisher Vector 的计算方法 (类似SVM)，将错误率降到了 25.7%
- 2012年冠军：Hinton 和两个研究生 Alex Krizhevsky, Illya Sutskever ，利用CNN+Dropout 算法 + RELU激励函数，用了 两个 Nvidia 的 GTX 580 CPU (内存 3GB，计算速度 1.6 TFLOPS)，花了接近六天时间，错误率只有 15.3%
- 2012年10月13日，当竞赛结果公布后，学术界沸腾了 这是神经网络二十多年来，第一次在图像识别领域，毫无疑义的，大幅度挫败了别的技术
- 这也许是人工智能技术突破的一个转折点

Yoshua Bengio & RELU

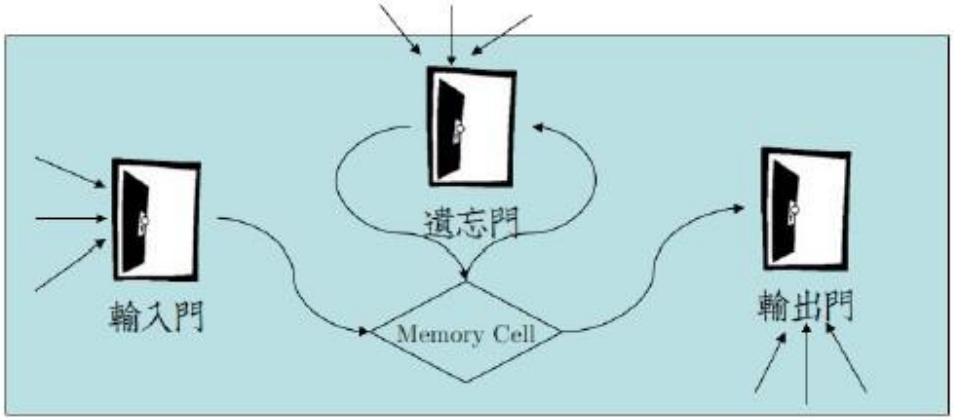
- 2011 年, 加拿大的蒙特利尔大学学者 Xavier Glorot 和 Yoshua Bengio 发表论文：Deep Sparse Rectifier Neural Networks
- 论文的算法中使用一种称为“修正线性单元” (REctified Linear Unit, RELU) 的激励函数. 对于特定的输入, 统计上有一半神经元是没有反应, 保持沉默
- 和使用别的激励函数的模型相比, RELU 不仅识别错误率普遍更低, 而且其有效性, 对于神经网络是否进行“预先训练”过并不敏感



- 传统的激励函数, 计算时要用指数或者三角函数, 计算量要比简单的RELU至少高两个数量级
- RELU的导数是常数, 非零即一, 不存在传统激励函数在反向传播计算中的"梯度消失问题"
- 由于统计上约一半的神经元在计算过程中输出为零, 使用 RELU 的模型计算效率更高, 而且自然而然的形成了所谓 “稀疏表征” (sparse representation), 用少量的神经元可以高效, 灵活, 稳健地表达抽象复杂的概念

Schmidhuber & LSTM

- 1997年瑞士 Lugano 大学的 Schmidhuber 和他的学生 Sepp Hochreiter 合作, 提出了长短期记忆 (LSTM, Long Short-Term Memory) 的计算模型
- LSTM：背后要解决的问题, 是如何将有效信息, 在多层循环神经网络传递之后, 仍能输送到需要的地方去
- LSTM 模块, 是通过内在参数的设定 (如图, input gate, output gate, forget gate), 决定某个输入信息在很久以后是否还值得记住, 何时取出使用, 何时废弃不用



Generative Adversarial Networks (GANs)

Ian Goodfellow, OpenAI Research Scientist

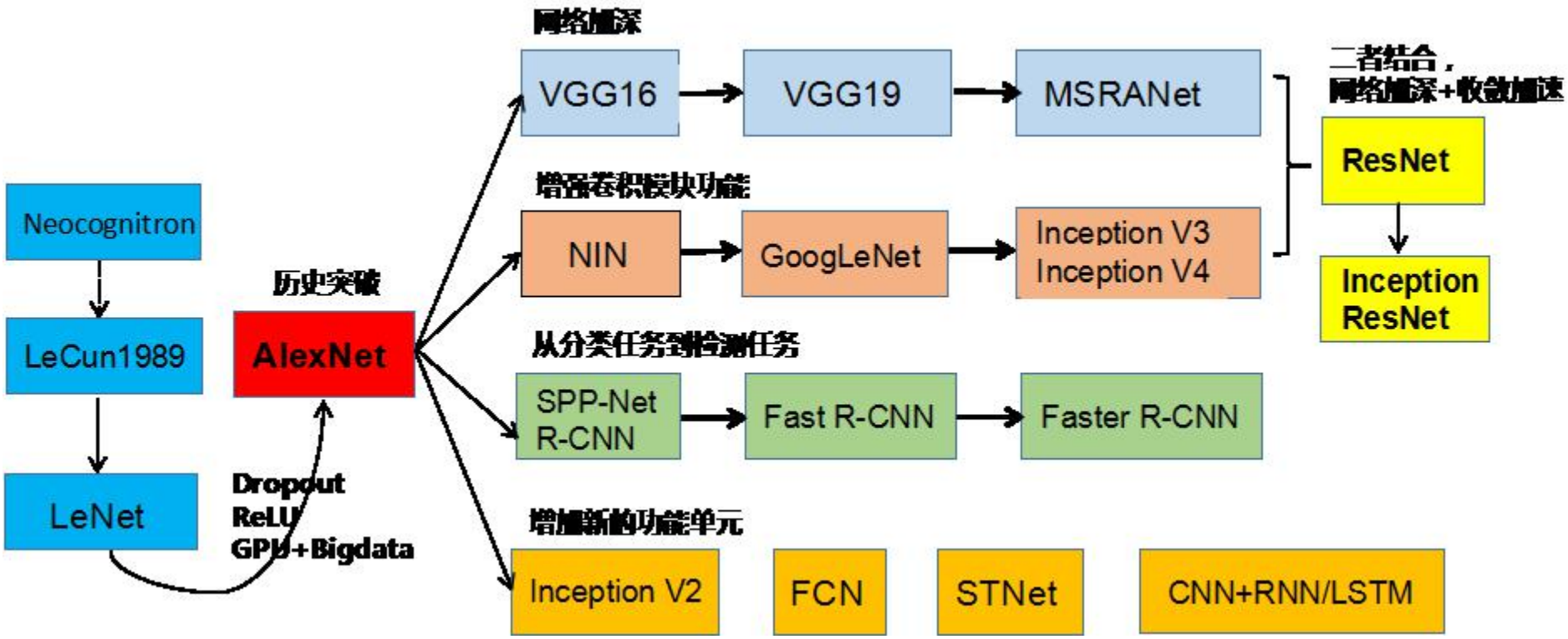
NIPS 2016 tutorial

Barcelona, 2016-12-4

OpenAI

休息，休息一会儿 … …

神经网络结构的发展

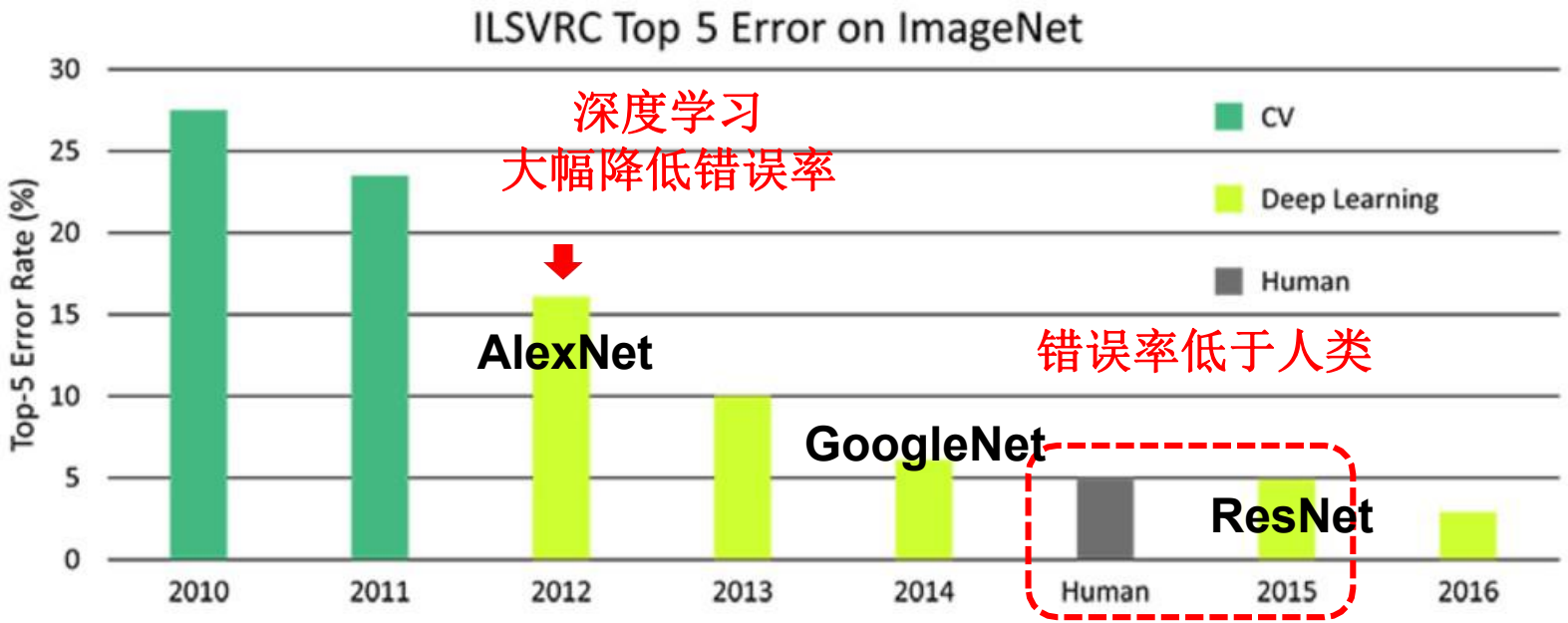


ImageNet的进击

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

2014 Teams	%error
GoogLeNet	6.6
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	12.1



source: https://www.dsiac.org/resources/journals/dsiac/winter-2017-volume-4-number-1/real-time-situ-intelligent-video-analytics

深度学习：学霸？游戏王？

10分钟，2000万

一年，300

深度学习的“琴棋书画”

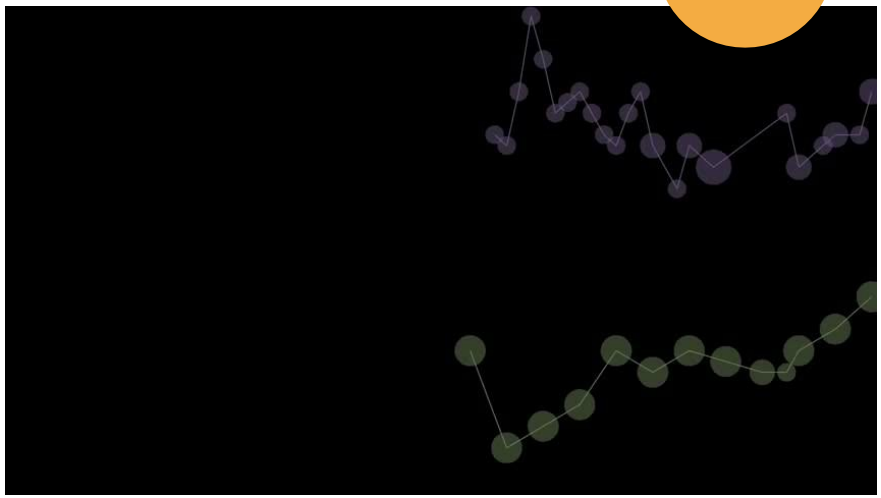
https://webapps.msxiaobing.com/V3/Portal?task=poem
http://poem.bosonnlp.com/

下棋

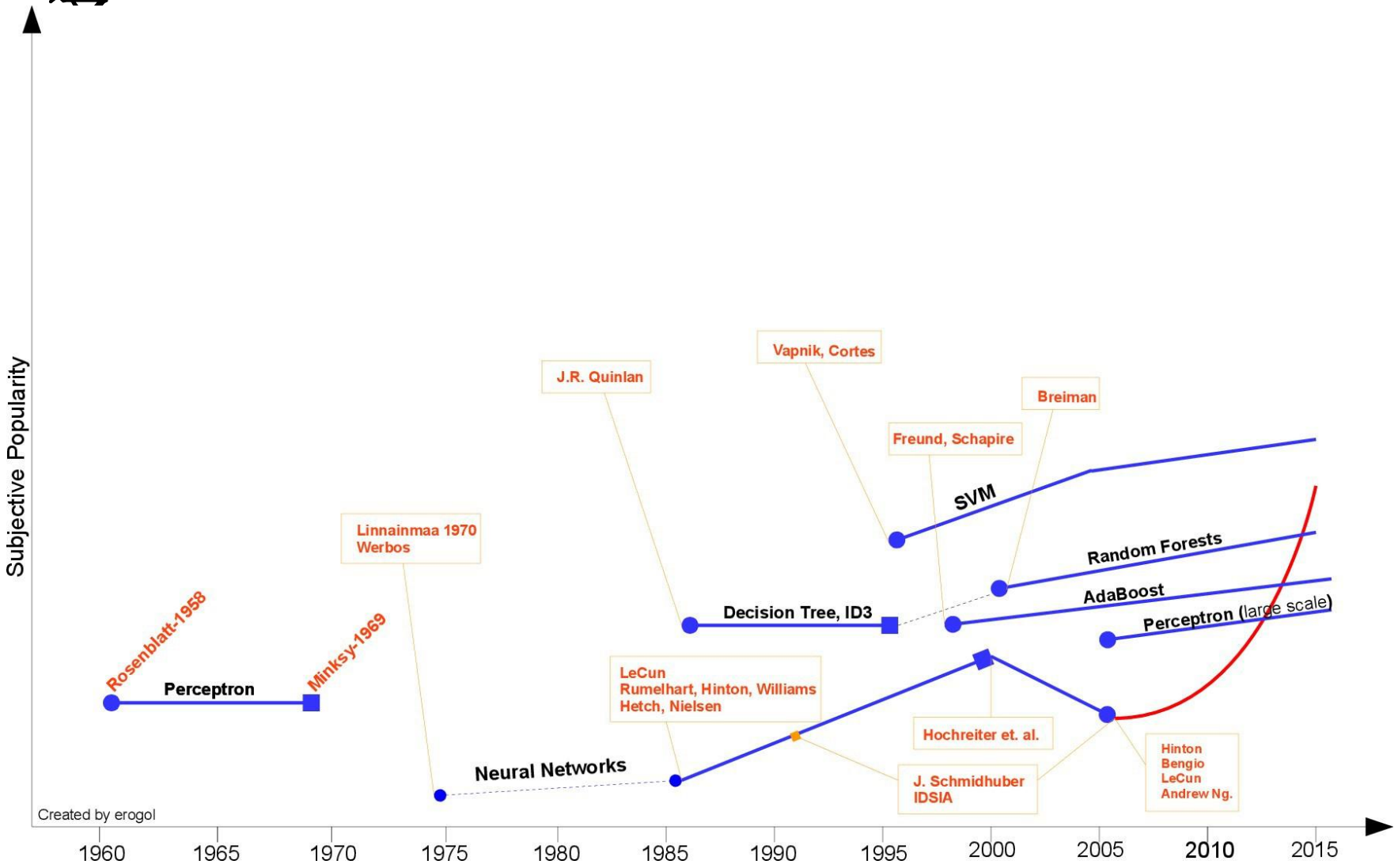
写诗

作曲

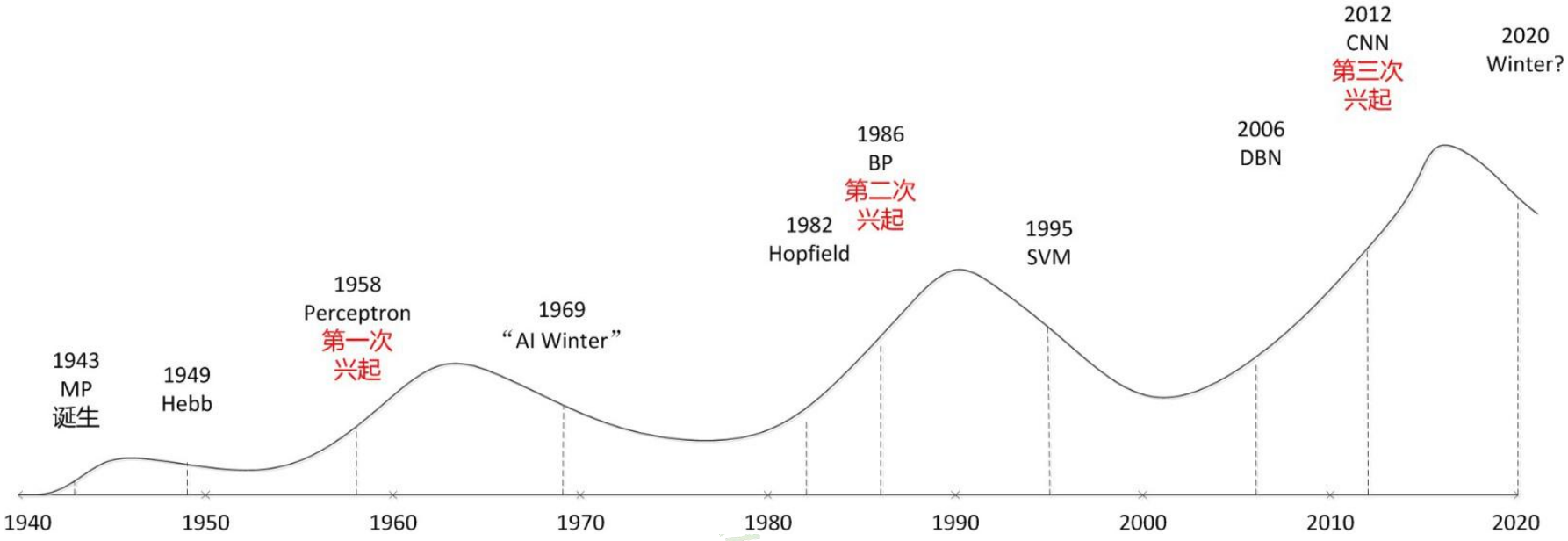
艺术画



流行机器学习模型的演变

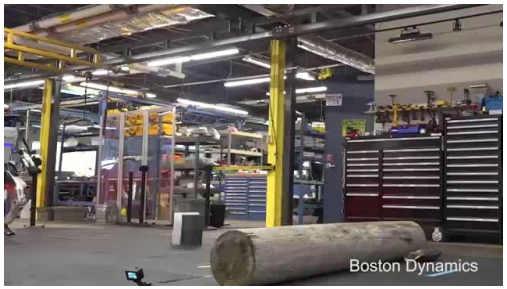


神经网络的三起两落


1943
MP
诞生
1949
Hebb
1958
Perceptron
第一次
兴起
1969
“AI Winter”
1982
Hopfield
1986
BP
第二次
兴起
1995
SVM
2006
DBN
2012
CNN
第三次
兴起
2020
Winter?
1940
1950
1960
1970
1980
1990
2000
2010
2020

神经网络的三起两落

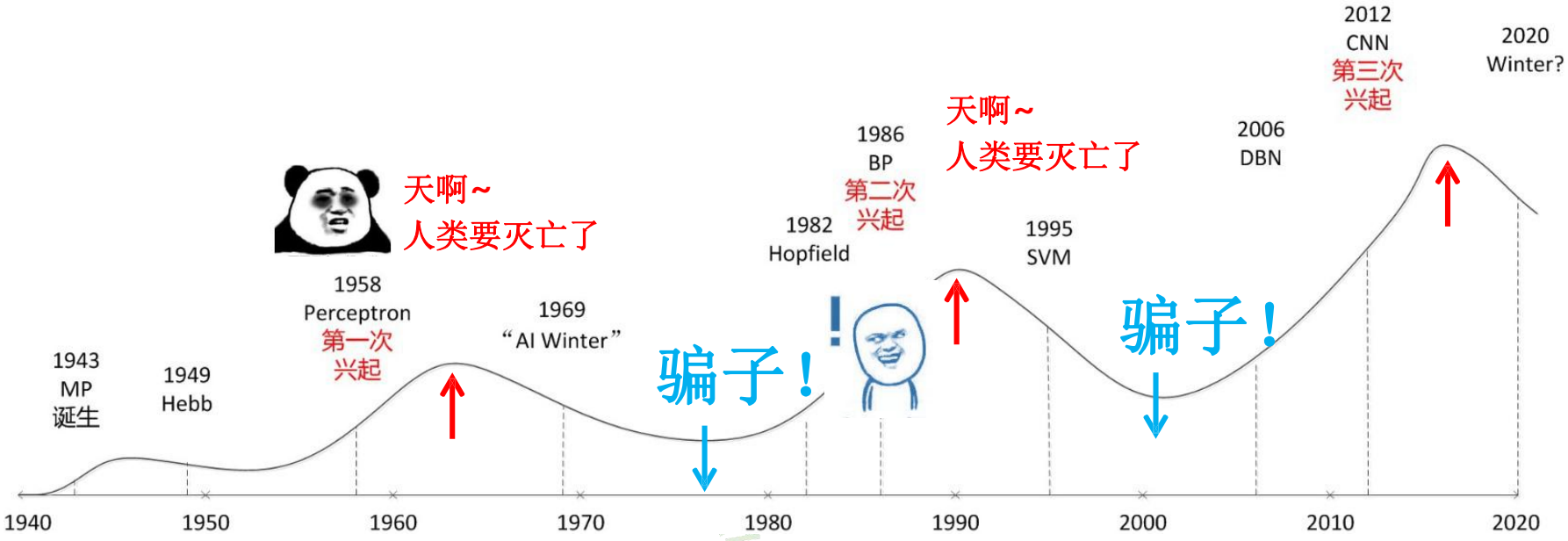
+



Atlas

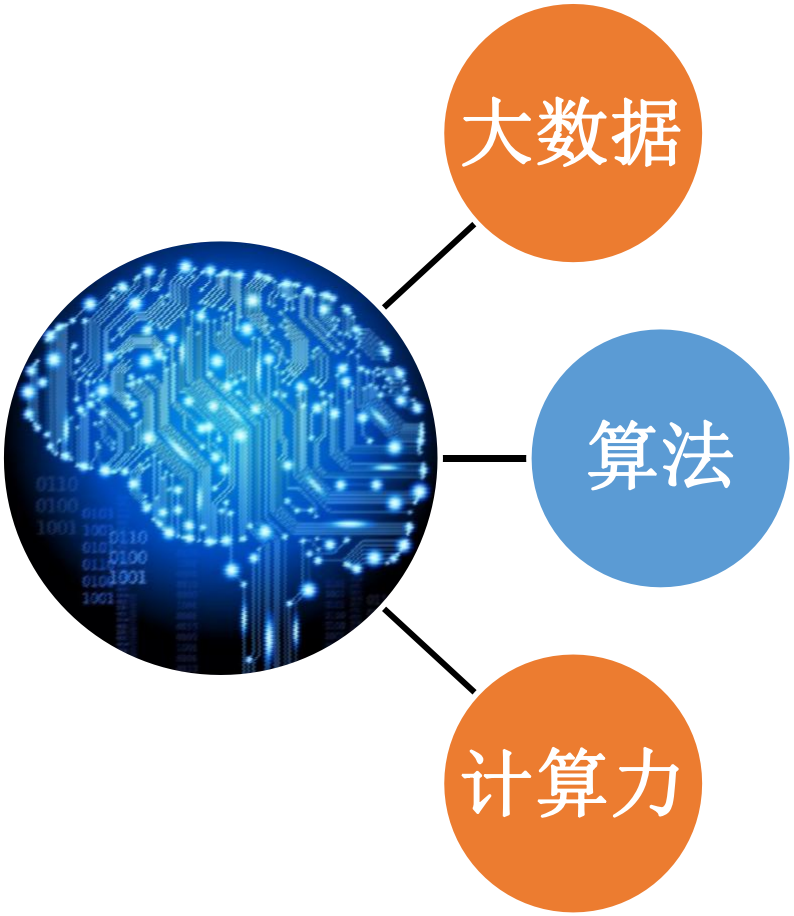
=

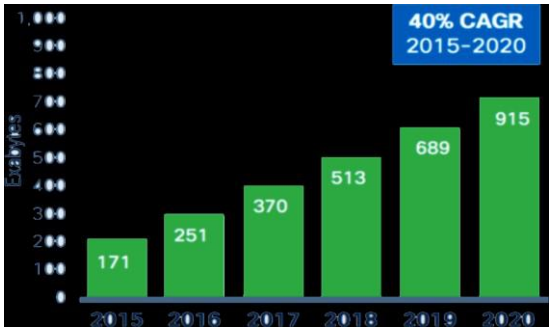
天啊~
人类要灭亡了



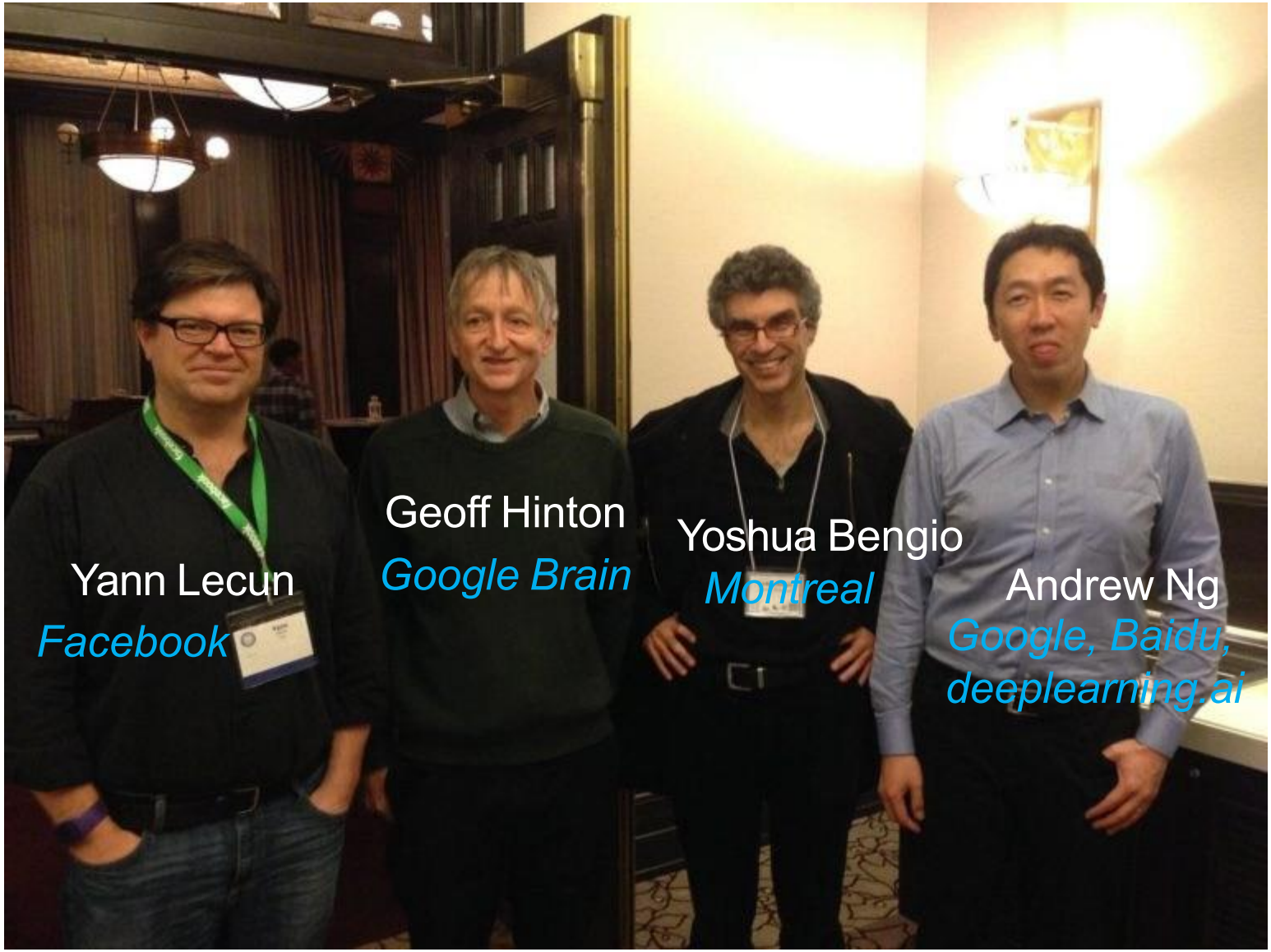
清华无人自行车

深度学习: 三个助推剂

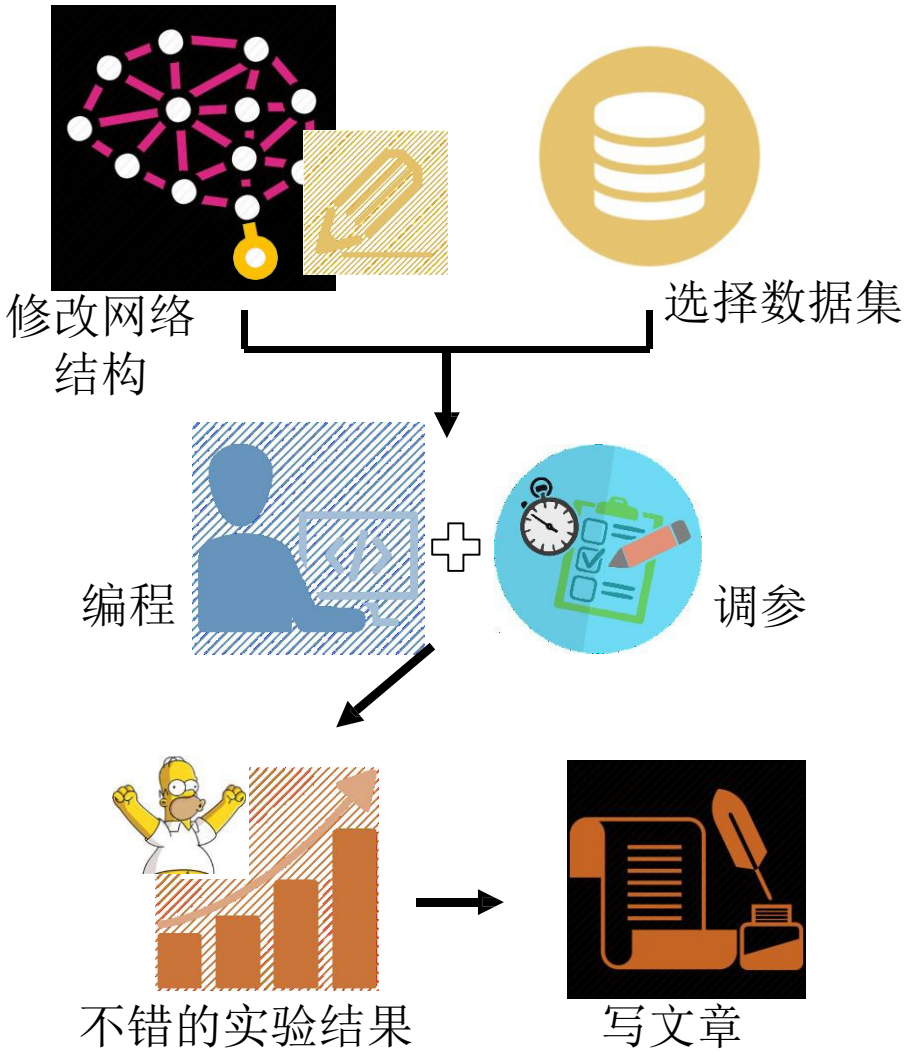


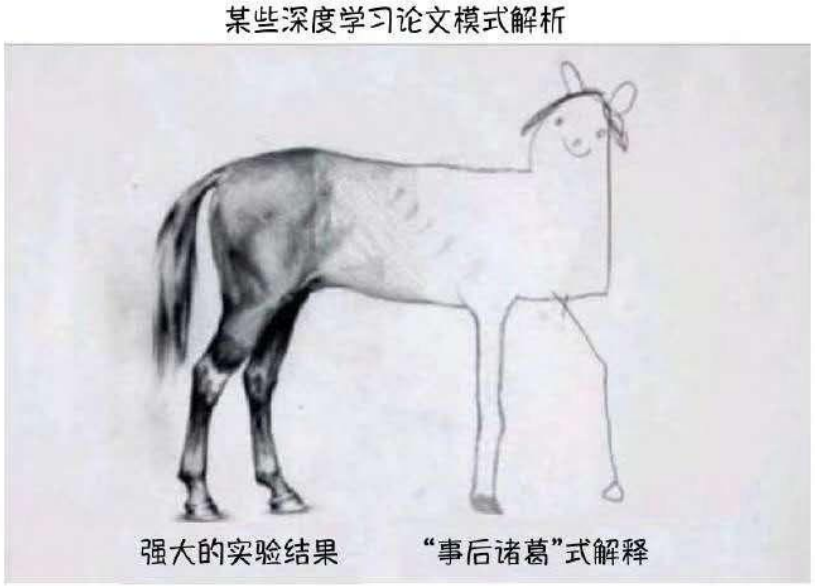


全球数据中心数据量在未来几年年均增速**40%**
— Cisco Global Cloud Index

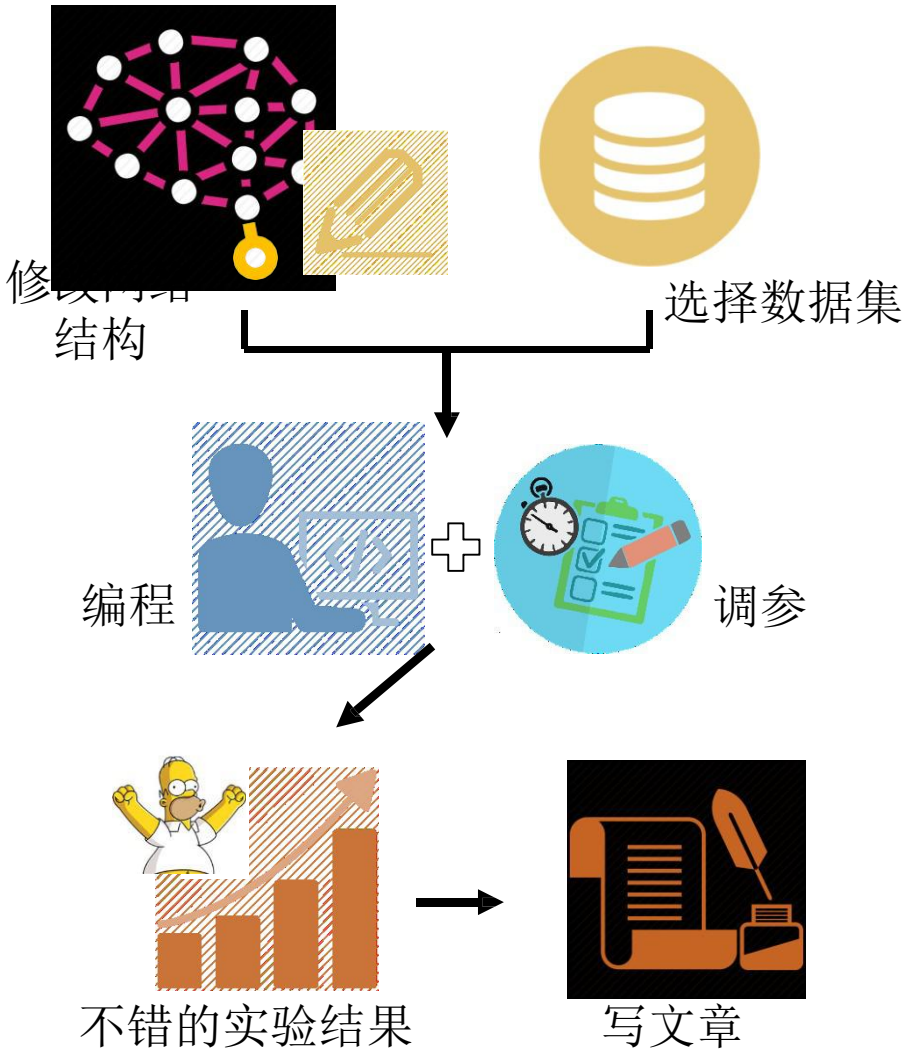

Yann Lecun
Facebook
Geoff Hinton
Google Brain
Yoshua Bengio
Montreal
Andrew Ng
Google, Baidu,
deeplearning.ai

深度学习: 工程技术 or 科学研究?





深度学习: 工程技术 or 科学研究?



EDITORIAL

National Science Review
0: 1, 2017
doi: 10.1093/nsr/nwx071
Advance access publication 24 June 2017

Science and technology, not SciTech

Guanrong Chen

The recurrent phrase 'science and technology' ranks one ahead of the other, but the two words have been treated the same by many and, in practice, their priorities have often been swapped.

Science, more precisely natural science, refers to a system or notion of acquiring knowledge through experimentation, simulation and analysis to understand and explain natural phenomena. Within the context of this discussion, it could also include mathematical science. Technology, on the other hand, refers to the collection of techniques, methods, skills and processes that are applicable to the generation of products or services beneficial to human society.

Between the two, science provides a foundation for technology to develop. Conversely, advancement in technology continuously generates new motivation and poses new questions to science. The late great scientist Qian Xuesen (Hsue-Shen Tsien, 1911–2009) believed that there is an important component, which he named engineering science, connecting the two together.

Scientific advances have mostly been driven by human curiosity to understand the basic principles governing the natural world, rather than the desire to meet human needs. Many incidences of discovery emerged unexpectedly, beyond human prediction or planning, and they might not be recognized as such within a short time. To name a couple of examples, mathematical number theory has a 3000-year-old history but it was considered particularly useful only when it was successfully applied to modern cryptography. The esoteric theory of general relativity of Albert Einstein had been placed in Heaven but recently stepped down to Earth with the GPS application. The structure of the DNA double helix was discovered due to the curiosity of James Watson and Francis Crick about genetic inheritance, which has lately revolutionized both life sciences and biotechnology.

It thus has become clear that, in promoting science and technology, one should not take the same approach and, in particular, one should not simply borrow the ideas from technology development to pave the way for science to evolve. Methodologies and policies from technology management should not be simply applied to managing science. However, it is not uncommon today that many administrative decision makers in academia rely on their 'technological thinking' to target everything including science, believing that centralized planning, big money and fast-track promotions alike would be able to spur science to develop and excel. Furthermore, prevailing views and policies measure the values of scientific research based solely on whether it is useful in providing services to the society or whether it is able to deliver marketable products in the foreseeable future. In so doing, some long-term fundamental scientific research would be ruled out because it could be labeled 'useless' from a technological point of view, especially at its initial stage.

In responding to such science and technology governing, Helmut Schwarz, President of the Alexander von Humboldt Foundation, recently points out that 'most breakthroughs in research are not and could not be planned. Rather, they appear, like Puck, in entirely unexpected corners. Because it is the passion of individuals that sparks major discoveries or inventions, choosing outstanding people and providing intellectual freedom and generous funding are key to the success of academic institutions' (On the usefulness of useless knowledge. *Nature Reviews* 2017; doi: 10.1038/S41570-016-0001).

Notably, in the common Chinese wording of SciTech (科技), this compound abbreviation of 'science and technology' is usually understood and presented as one single subject, leading to the widespread misconception of science and technology as synonym, to be viewed and managed in the same way. This is a problem throughout the long history of China. Cumulated observations and evidence suggest that this view of SciTech may be one of the reasons that modern science did not emerge in China. In fact, most Chinese ancient advances were developed towards technology for their practical values but did not evolve into building fundamental scientific knowledge and theories. For example, the discovery of gunpowder did not lead to modern theoretical chemistry, the creation of the compass did not lead to modern electromagnetics theory or theoretical physics and the ancient Chinese remainder theorem did not lead to modern number theory in mathematics.

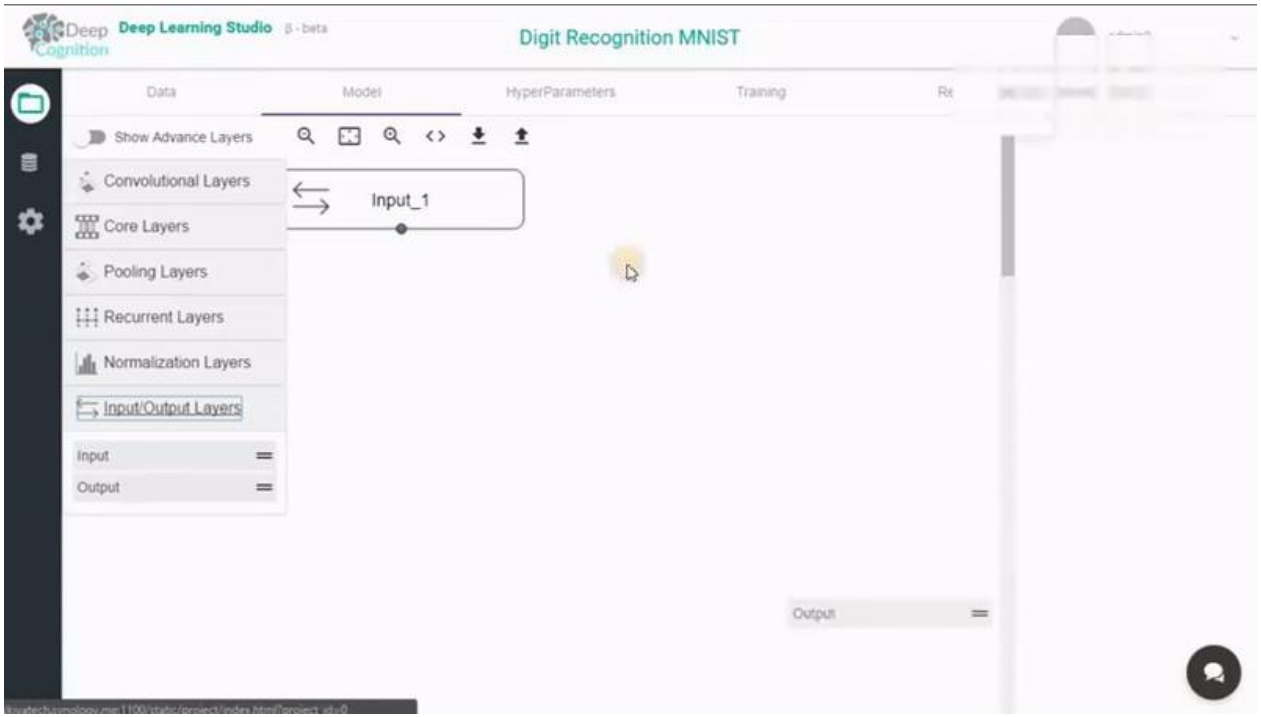
That technological innovations were not accompanied by the establishment of modern science has long been a big puzzle that remains for Chinese scientists and technologists to be fully unraveled which, if well resolved, might quickly lead Chinese modern science to the forefront.

Guanrong Chen
Chair Professor & Director, Centre for Chaos and Complex Networks,
City University of Hong Kong
Editorial Board Member of NSR
E-mail: eegchen@cityu.edu.hk

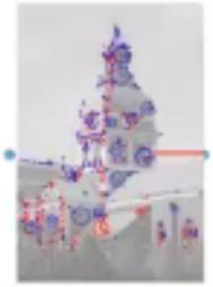
工程低门槛 & 科研高门槛

- Keras框架的作者François Chollet称：“深度学习研究已经进入了瓶颈期。将深度学习应用于解决现实生活问题的应用正在迎来一个大爆发。”

Deep Cognition



深度学习的应用发展： 视觉 + 语言



特征提取

物体检测

语义标注

实体标注

描述生成

What is the mustache made of?

多媒体问答

多媒体叙事

Description:
A falcon is eating during sunset.
The falcon is standing on earth.

Poem:
Like a falcon by the night
Hunting as the black knight
Waiting to take over the fight
With all of its mind and might

语言：为你写诗

深度学习的应用发展： 视觉 + 语言

随着深度学习的应用拓展，从探索深度学习“能”做什么，发展到思考深度学习“不能”做什么？

深度学习的“不能”

Deep Learning: A Critical Appraisal

Gary Marcus

(Submitted on 2 Jan 2018)

Although deep learning has historical roots going back decades, neither the term "deep learning" nor the approach was popular just over five years ago, when the field was reignited by papers such as Krizhevsky, Sutskever and Hinton's now classic (2012) deep network model of Imagenet. What has the field discovered in the five subsequent years? Against a background of considerable progress in areas such as speech recognition, image recognition, and game playing, and considerable enthusiasm in the popular press, I present ten concerns for deep learning, and suggest that deep learning must be supplemented by other techniques if we are to reach artificial general intelligence.

Comments: 1 figure
Subjects: **Artificial Intelligence (cs.AI)**; Learning (cs.LG); Machine Learning (stat.ML
MSC classes: 97R40
ACM classes: I.2.0; I.2.6
Cite as: **arXiv:1801.00631 [cs.AI]**
(or **arXiv:1801.00631v1 [cs.AI]** for this version)

Deep learning thus far

3.1. is data hungry

3.2. is shallow & has limited capacity for transfer

3.3. has no natural way to deal with hierarchical structure

3.4. has struggled with open-ended inference

3.5. is not sufficiently transparent

3.6. has not been well integrated with prior knowledge

3.7. cannot inherently distinguish causation from correlation

3.8. presumes a largely stable world

3.9. its answer often cannot be fully trusted

3.10. is difficult to engineer with

Gary Marcus. 2018. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631* (2018).

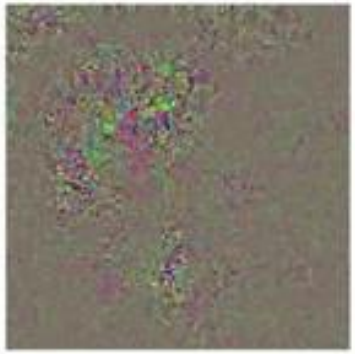
深度学习的“不能”(1)

- 算法输出不稳定，容易被“攻击”

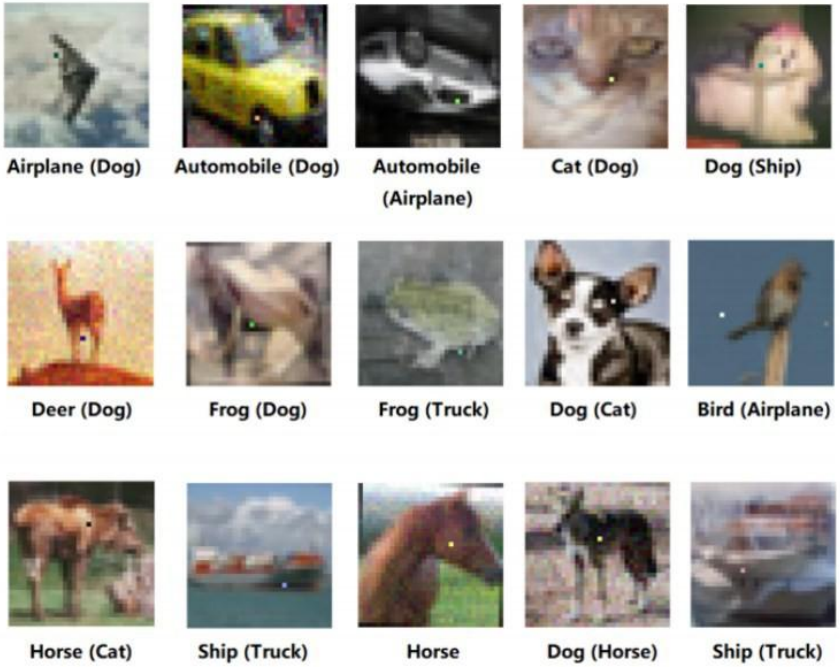
“African elephant”

“koala”

difference



10x difference



one pixel attack

深度学习的“不能”(1)

- 算法输出不稳定，容易被“攻击”

SQuAD
The Stanford Question Answering Dataset

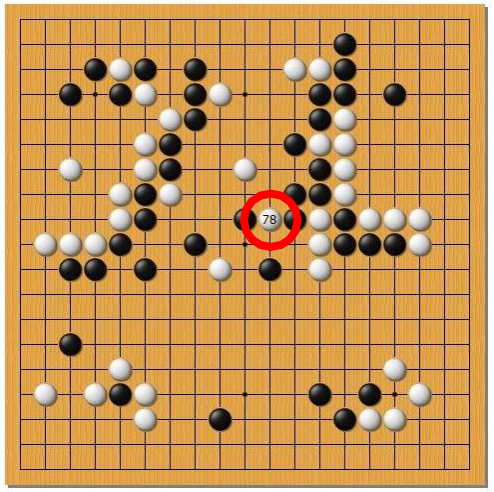
Article: Super Bowl 50
Paragraph: *“Peyton Manning became th back ever to lead two different teams to Bowls. He is also the oldest quarterba in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver’s Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.”*
Question: *“What is the name of the quarterback who was 38 in Super Bowl XXXIII?”*
Original Prediction: John Elway
Prediction under adversary: Jeff Dean

	Match Single	Match Ens.	BiDAF Single	BiDAF Ens.
Original	71.4	75.4	75.5	80.0
ADDSENT	27.3	29.4	34.3	34.2
ADDONESENT	39.0	41.8	45.7	46.9
ADDANY	7.6	11.7	4.8	2.7
ADDCOMMON	38.9	51.0	41.7	52.6

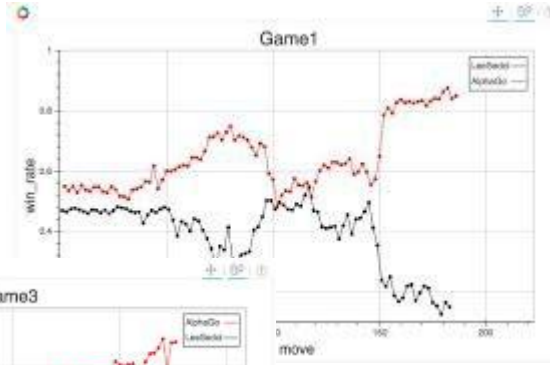
75%
36%
7%

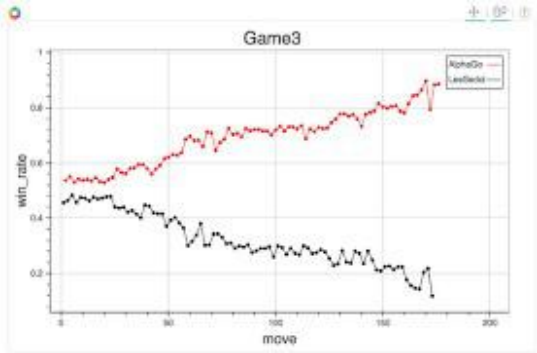
深度学习的“不能”(2)

- **模型复杂度高，难以纠错和调试**



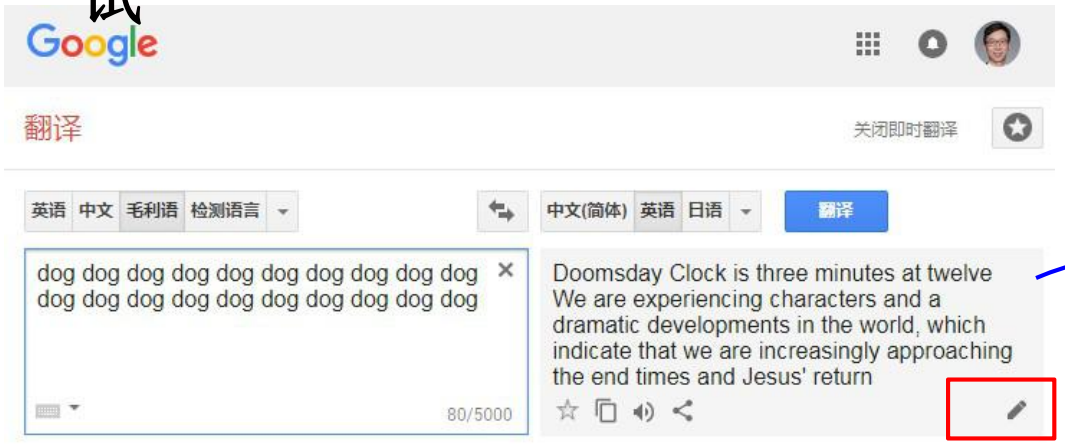
实际的我们





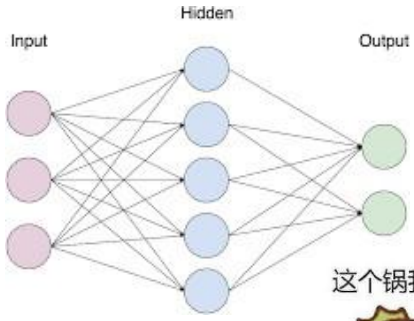
深度学习的“不能”(2)

- 模型复杂度高，难以纠错和调试



“世界末日时钟还差3分钟到12点。我们正在经历世界上的人物和戏剧性的发展，这表明我们越来越接近末日和耶稣的回归。”

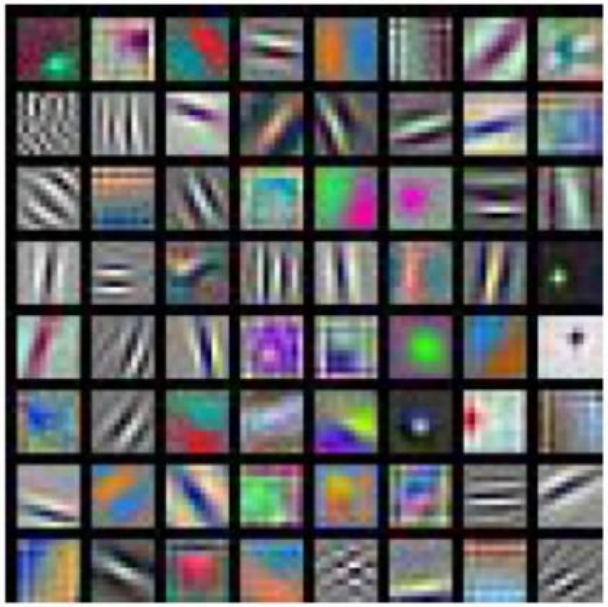
恶作剧？



Deep Dream

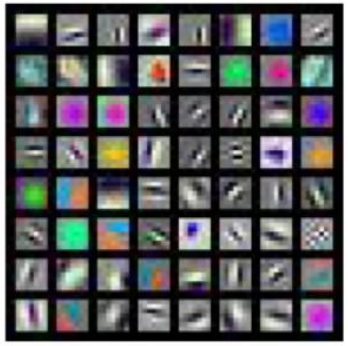
深度学习的“不能”(3)

- **模型层级复合程度高，参数不透明**

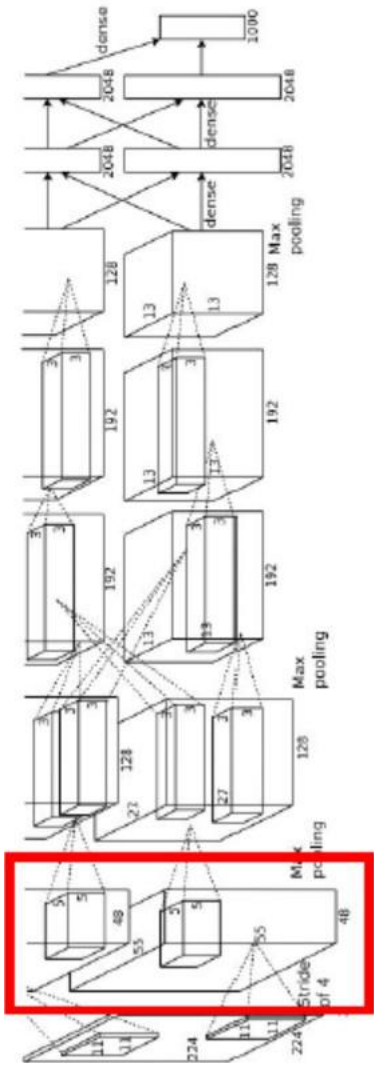


AlexNet:
64 x 3 x 11 x 11

ResNet-101:
64 x 3 x 7 x 7



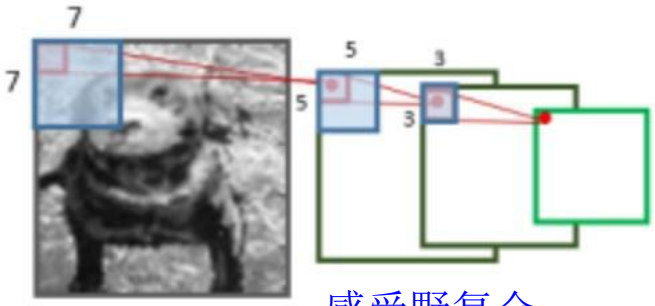
DenseNet-121:
64 x 3 x 7 x 7



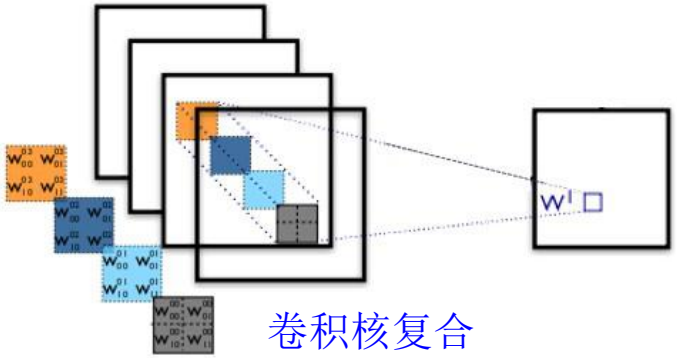
深度学习的“不能”(3)

- **模型层级复合程度高，参数不透明**

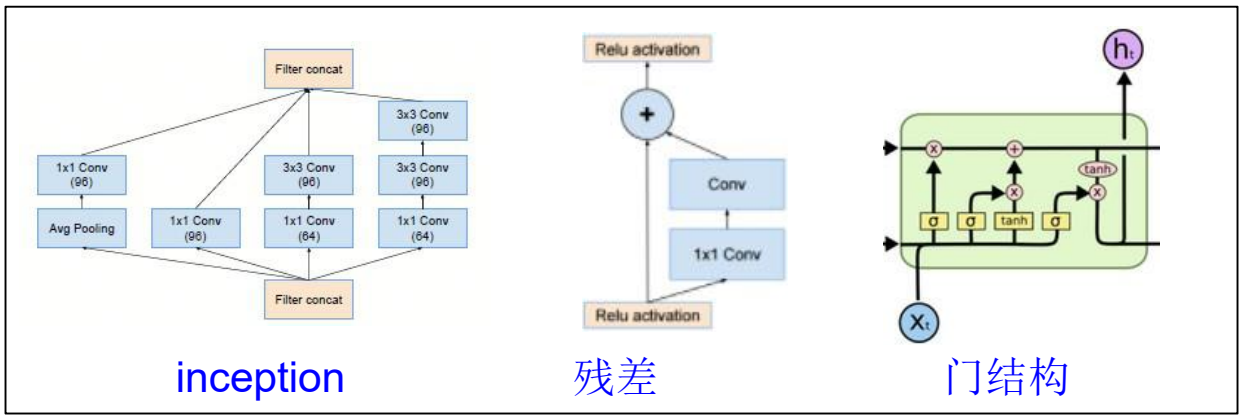
第一层：16*3*3*3 第二层：N*16*3*3



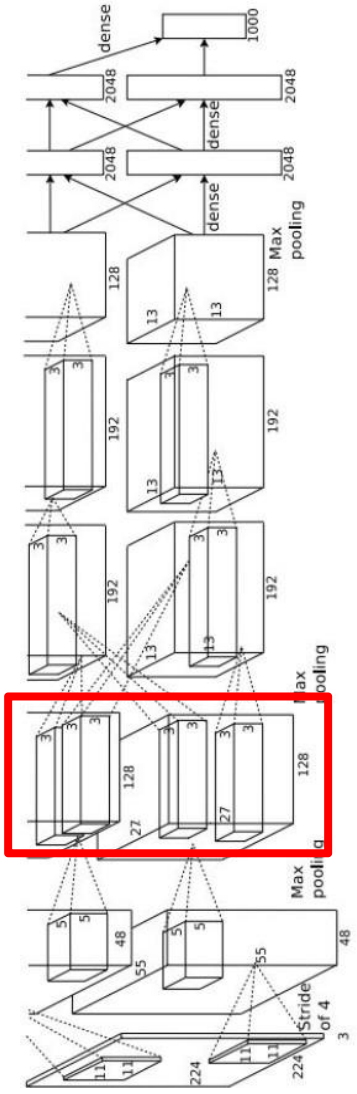
感受野复合



卷积核复合



inception 残差 门结构



深度学习的“不能”(4)

- 端到端训练方式对数据依赖性强，模型增量性

$$\text{Test loss} - \text{training loss} \leq \sqrt{\frac{N}{m}}$$

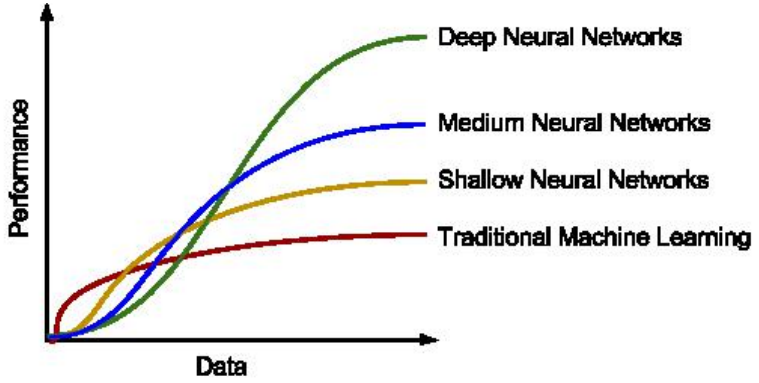
泛化误差　　训练误差

m = #训练样本

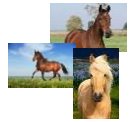
N = effective capacity (模型有效容量)

↑ 上界

#参数 / VC维 / Rademacher复杂度



当样本数据量小的时候，深度学习无法体现强大拟合能力


马


草地


农舍


山

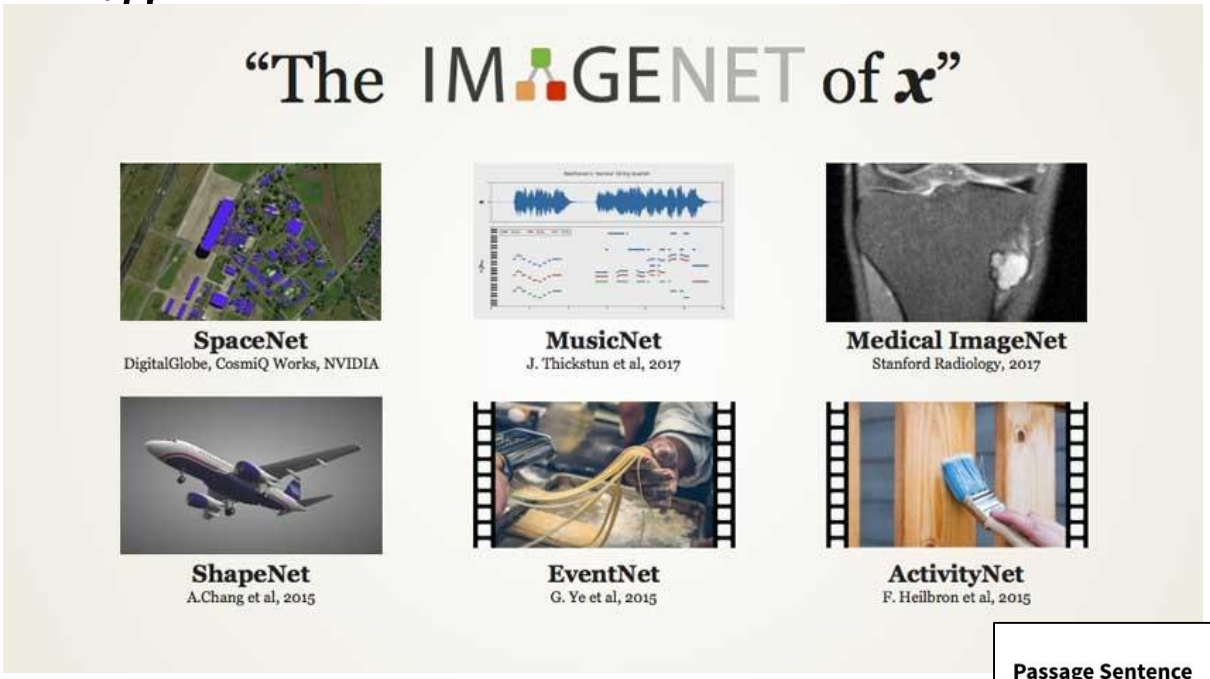
语义标注

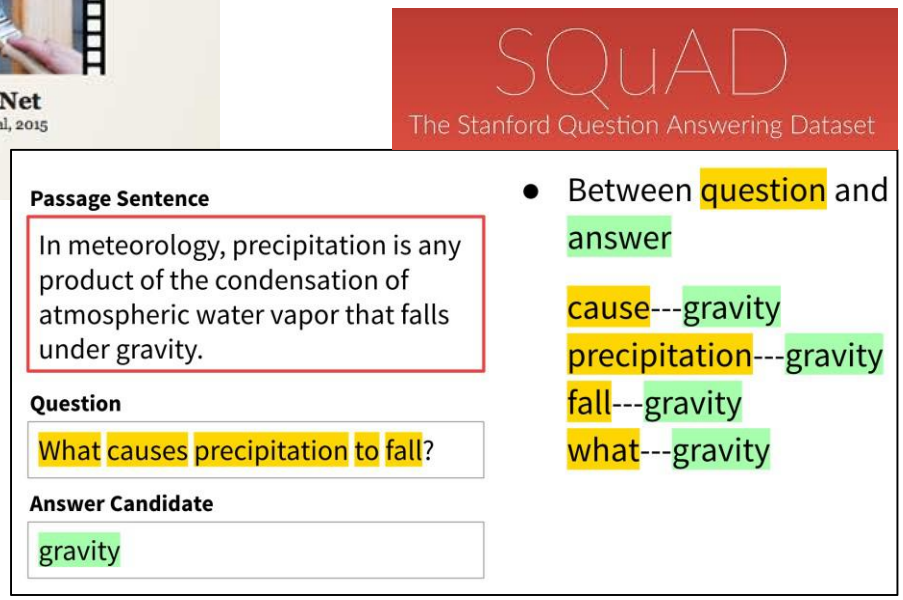
关系检测

图像描述 ?

深度学习的“不能”(5)

- 专注直观感知类问题，对开放性推理问题无能为力



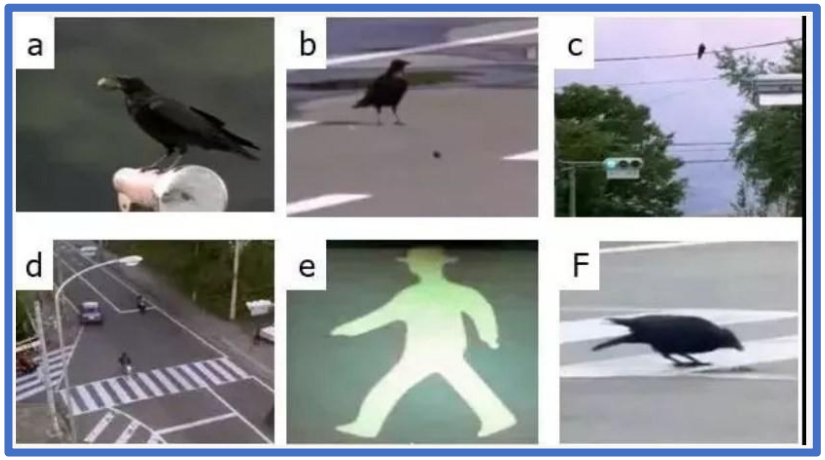


深度学习的“不能”(5)

- 专注直观感知类问题，对开放性推理问题无能为力

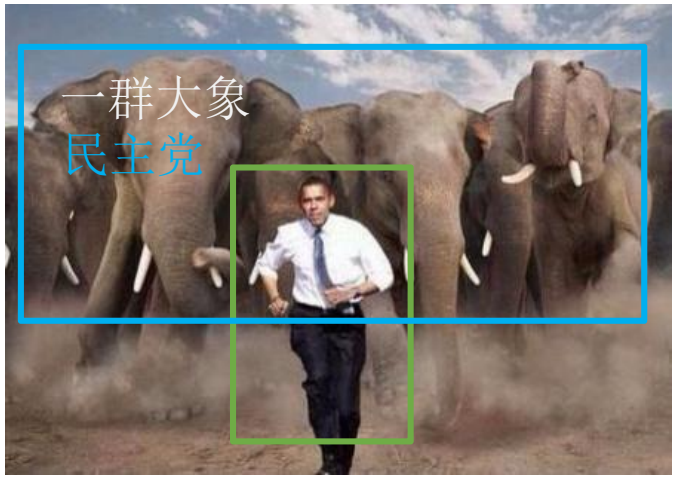
“鹦鹉”智能

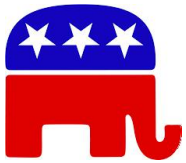
“乌鸦”智能



深度学习的“不能”(5)

- 专注直观感知类问题，对开放性推理问题无能为





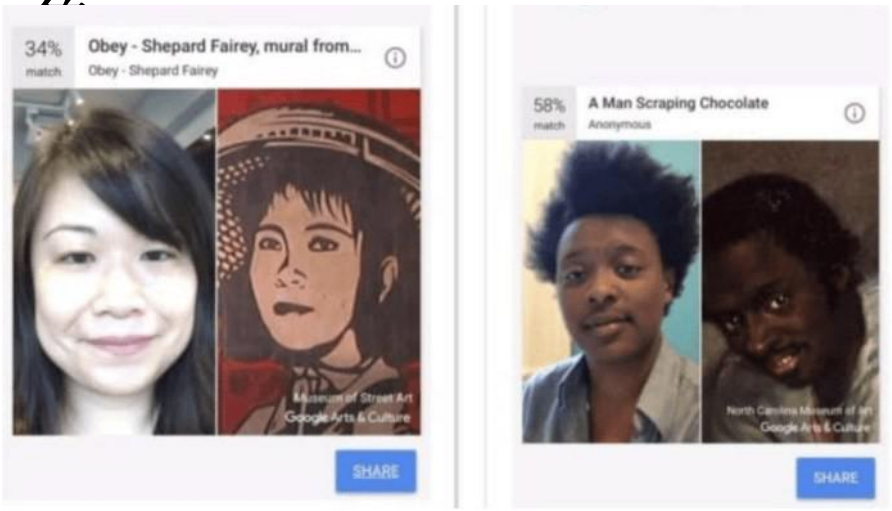
大象

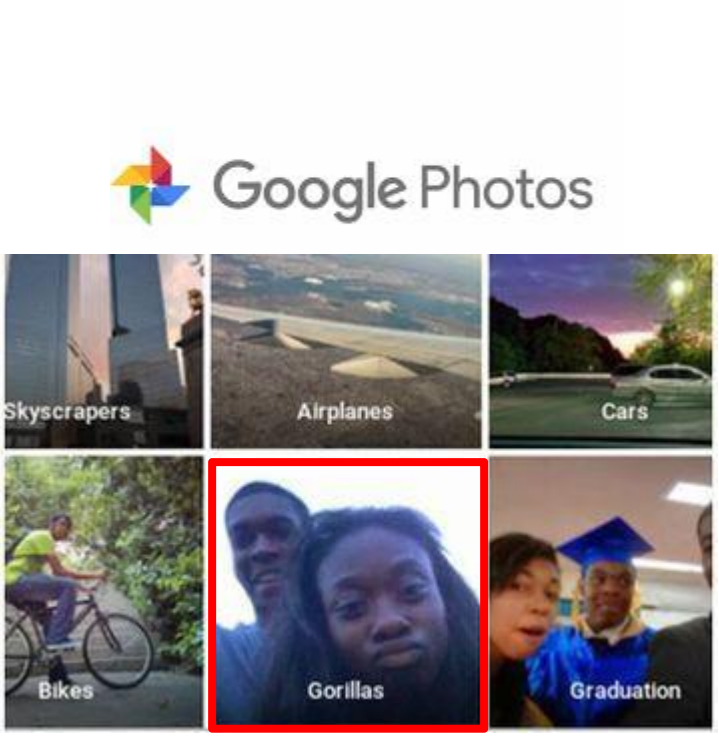
|

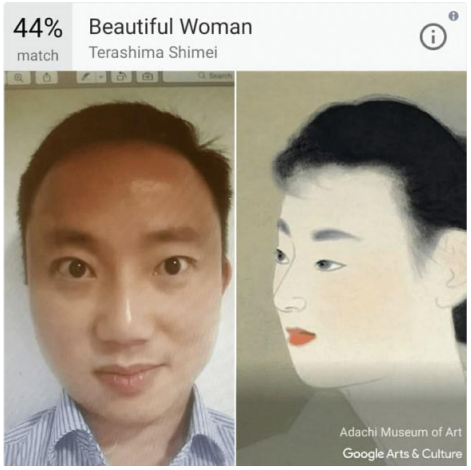
民主党

深度学习的“不能”(6)

- 人类知识无法有效引入进行监督，机器偏见难以避







深度学习的“不能”(6)

- 人类知识无法有效引入进行监督，机器偏见难以避免

Man:Woman as King:Queen

Father:Doctor as Mother: Nurse

Man:Computer_Programmer as Woman: Homemaker

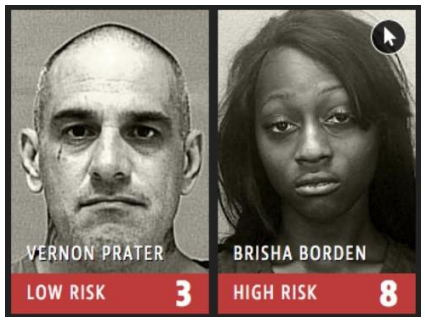
Gender stereotype *she-he* analogies

sewing-carpentry	registered nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	lovely-brilliant

深度学习的“不能”(6)

- 人类知识无法有效引入进行监督，机器偏见难以避免
  - 微软开发了**Tay**聊天机器人，模仿年轻网民的语言模式
  - 试用**24**小时后，被引入歧途，成为偏激的种族主义者，甚至发出

- 美国法院用以评估犯罪风险的算法**COMPAS**，被证明对黑人造成了系 统性歧视。



算法依赖于大数据，但数据不是中立的：从真实社会中抽取，必然带有社会固有的不平等、排斥性和歧视。

深度学习的“不能”(6)

- 人类知识无法有效引入进行监督，机器偏见难以避免

阿西莫夫机器人三定律

- 机器人不得伤害人类，或坐视人类受到伤害；
- 除非违背第一法则，否则机器人必须服从人类命令；
- 除非违背第一或第二法则，否则机器人必须保护自己

深度学习的“不能”与解释性

深度学习的“不能”

- 稳定性低
- 可调试性差
- 参数不透明
- 机器偏见
- 增量性差
- 推理能力差

深度学习的“不能”与解释性

深度学习的“不能”

- 稳定性低
- 可调试性差
- 参数不透明
- 机器偏见
- 增量性差
- 推理能力差

解释性的三个层次

找得到

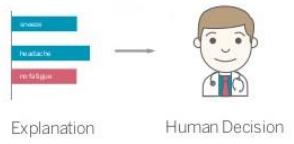
“对症下药”

知道哪些特征对输出有重要影响，出了问题准确快速纠错

看得懂

不再“对牛弹琴”

双向：算法能被人的知识体系理解 +利用和结合人类知识



留得下

“站在巨人的肩膀上”

知识得到有效存储、积累和复用

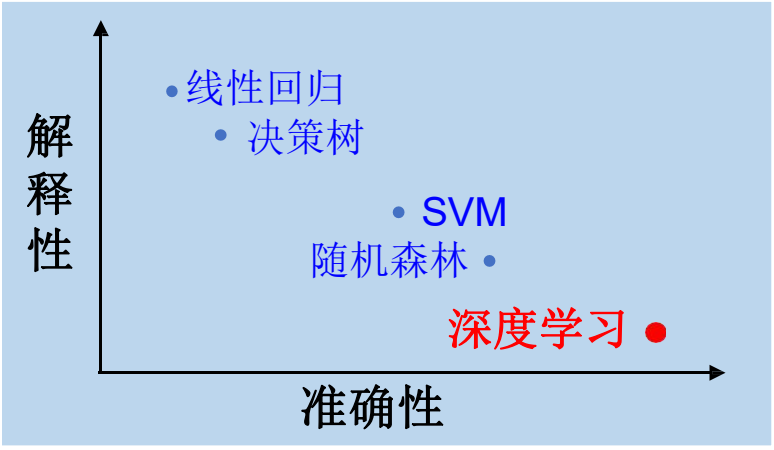
→ 越学越聪明

解释性 vs 泛化性

2017年4月，美国国防部先进研究项目局DARPA资助了可解释性人工智能项目Explainable AI, 提高AI决策的解释能力。

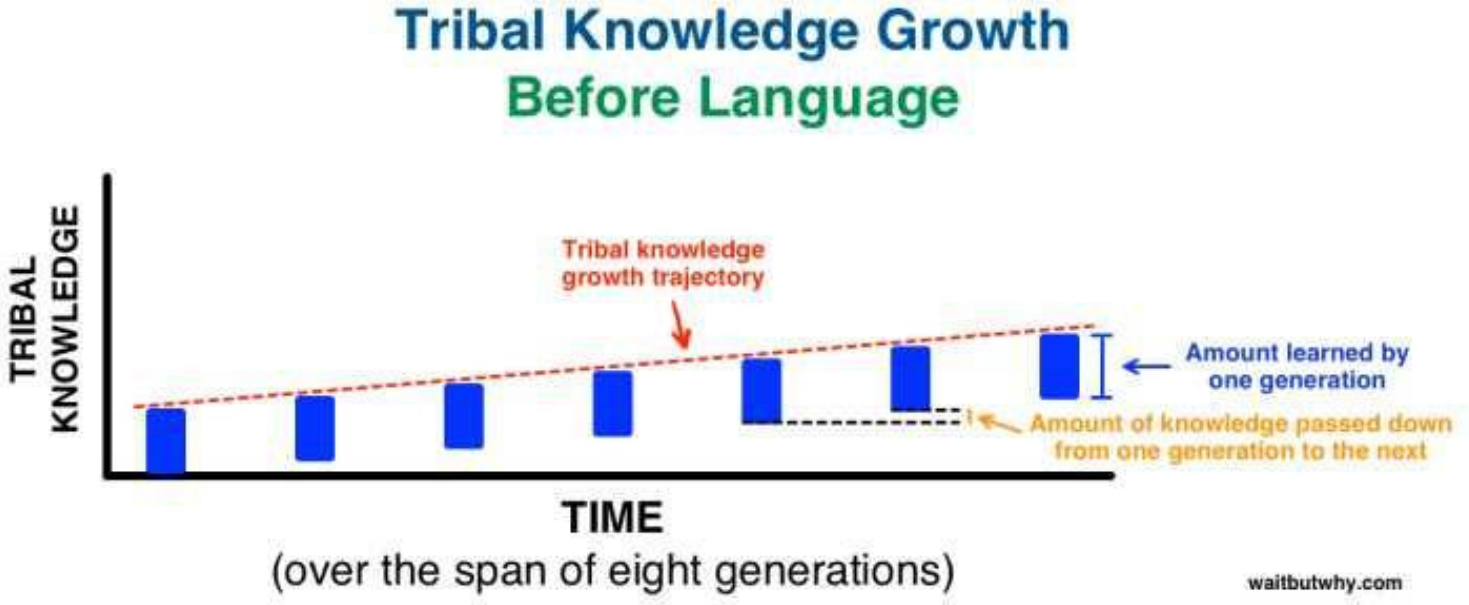
2017年7月，国务院在《新一代人工智能发展规划》中提出“实现具备高可解释性、强泛化能力的人工智能”

《一般数据保护条例》规定：自2018 年5 月25 日起要求所有算法解释其输出原理，保证数据隐私和算法公平性



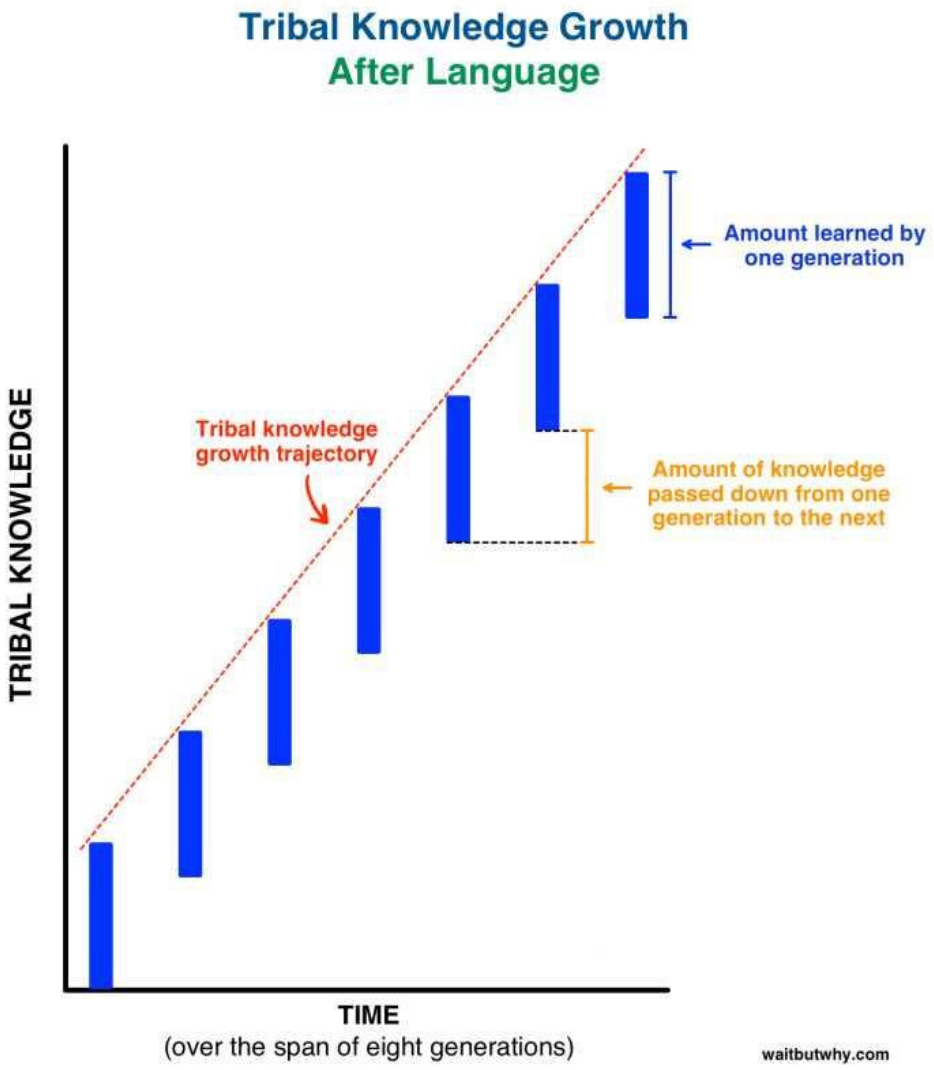
知识的表达和积累

- 语言出现以前，每一代人类都需要自身重新习得很多知识
- 知识无法积累：从上一代到下一代一增长缓慢



知识的表达和积累

- 语言出现后，知识传递有了载体
- 同时间尺度下增长迅速



连接主义 vs 符号主义：从对立到合作

1956

达特茅斯会议

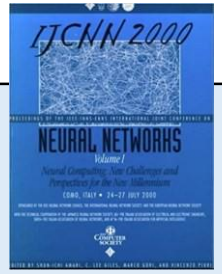
- 人工智能被定义为“对智能行为的符号化建模”
- 符号主义自上而下，由控制者主导
- 连接主义自下而上，崇尚分布式、平等的知识表示和计算

1991

明斯基

- “并不存在一种最佳的知识表示方法…同时利用多种知识表示的优势”
  承认了连接主义和符号主义必须结合

2000

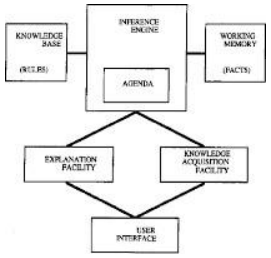


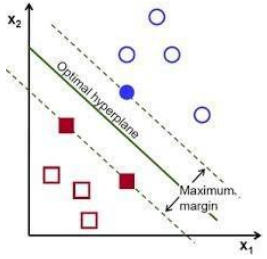
IJCNN Panel

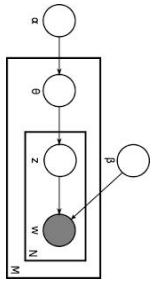
- 从控制论中找到依据支持神经网络规则抽取的合理性
- “strong connectionism dies, weak connectionism stands”

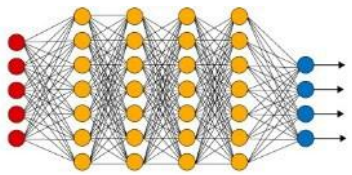
连接主义 + 符号主义

从专家系统、统计机器学习到概率图模型再到深度学习，模型准确率不断提高，解释性却没有提高





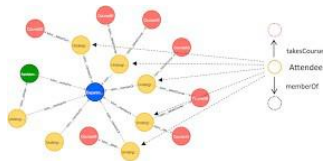




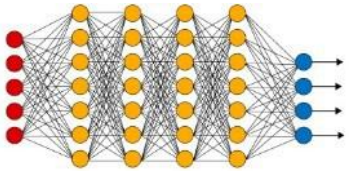
符号主义以逻辑为核心，相比以统计为核心的连接主义方法通常有较好的解释性

以知识图谱为代表的符号主义和以深度学习为代表的连接主义，逐渐走向协同发展的方向

模拟心智 → 符号主义 → 逻辑结构 → 知识图谱



模拟结构 → 联结主义 → 神经网络 → 深度学习



- 深度学习 + 图谱
- 数　　据 + 知识

THANKS