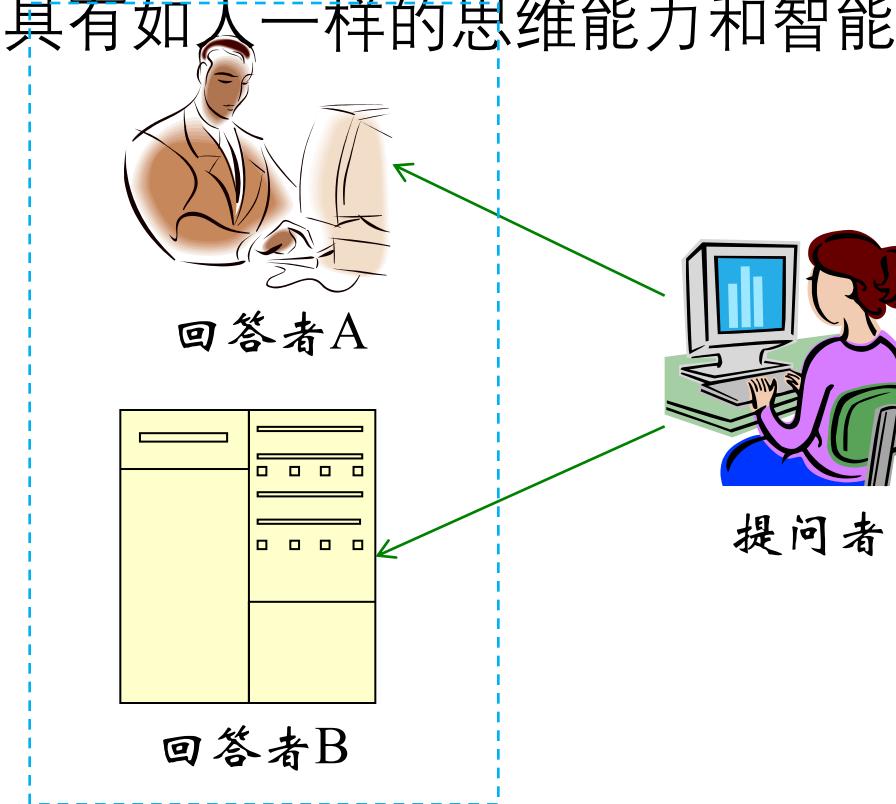


# 图灵测试

## □ 怎样判断机器具有智能—图灵测试

1950年Alan Turing的文章“.”  
(*Mind*, Vol. 59, No. 236) 提出图灵测试，检验一台机器或Computing Machinery and Intelligence电脑是否具有如人一样的思维能力和智能

电脑和人分别封闭在不同的房间，测试者不知道哪个房间是人，哪个房间是电脑，他向双方提出测试问题，电脑和人给出各自的答案，如果一系列的测试问题之后，测试者分不出哪些是电脑的答案，哪些是人的答案，则电脑通过测试，确实具有与人一样的智能。



# 计算智能

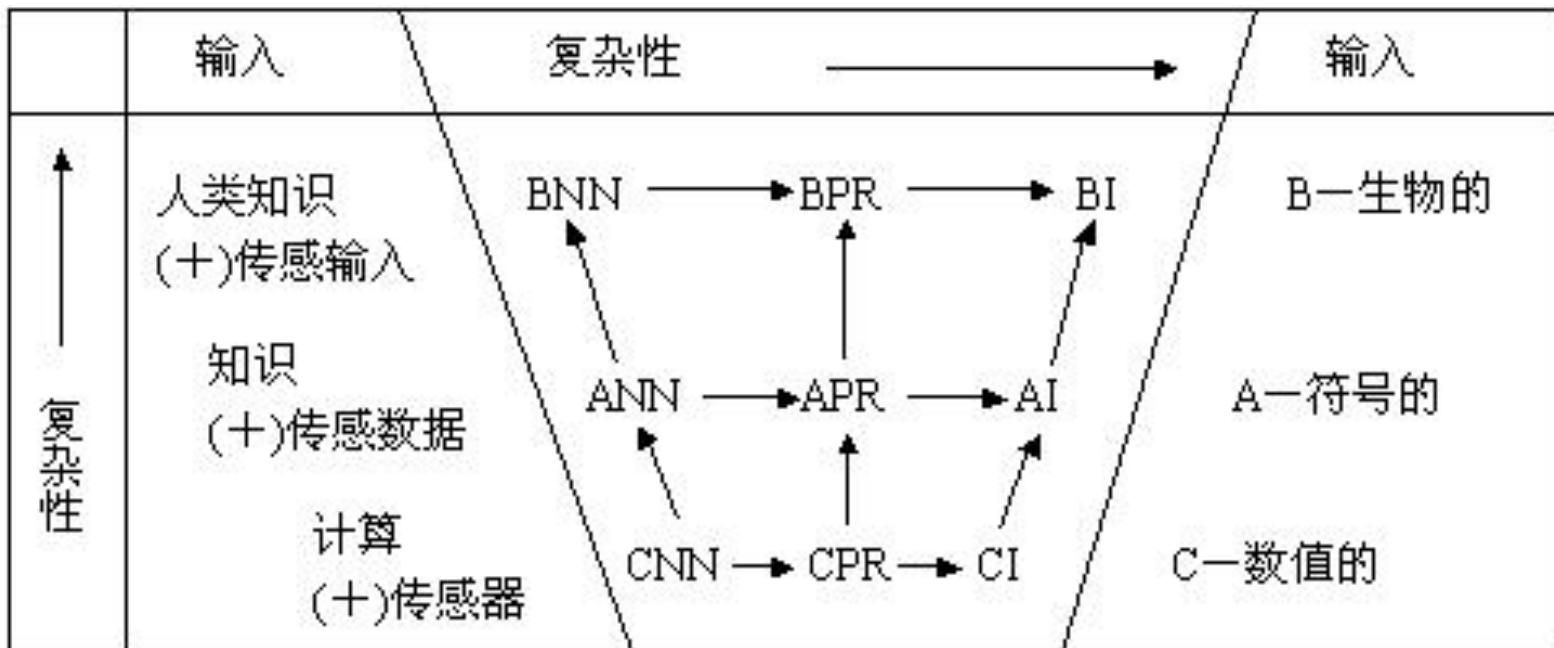


- 计算智能主要借鉴仿生学和拟物的思想，基于人们对生物体智能机理和某些自然规律的认识，采用数值计算的方法去模拟和实现人类的智能、生物智能、其它社会和自然规律。
- 从学科范畴看，计算智能是在神经网络（Neural Networks, NN）、进化计算（Evolutionary Computation, EC）及 模糊系统（Fuzzy System, FS） 这3个主要领域发展相对成熟的基础上形成的一个统一的学科概念。

# 计算智能与人工智能的关系

- 目前，对计算智能与人工智能的关系有2种不同观点，一种观点认为计算智能是人工智能的一个子集，另一种观点认为计算智能和人工智能是不同的范畴。
- 第一种观点的代表人物是贝慈德克。

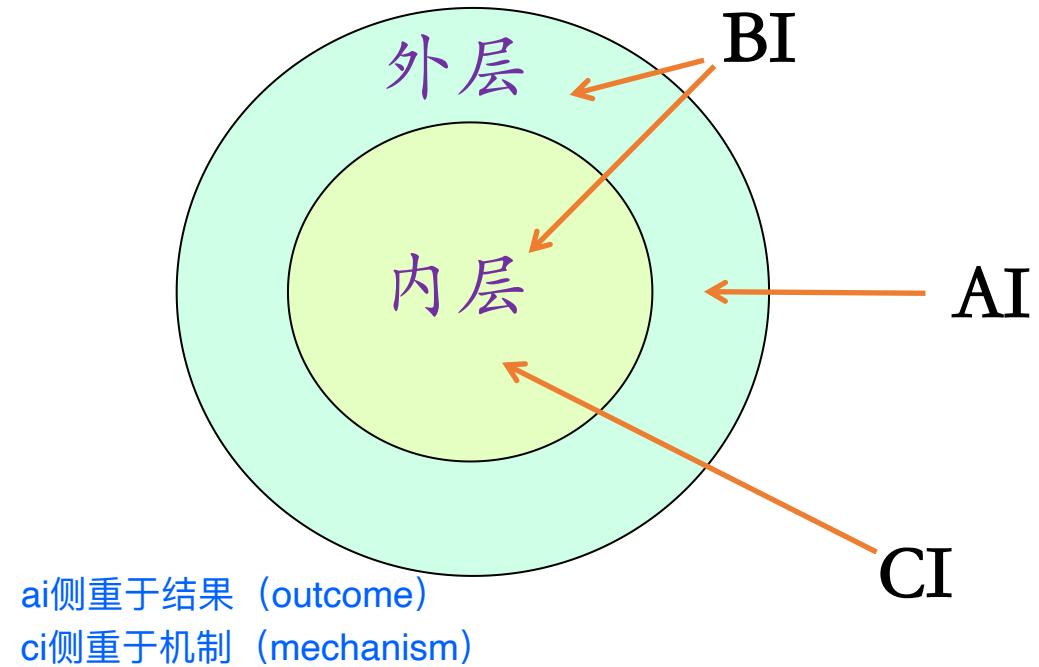
CI < AI < BI



A—Artificial, 表示人工的（非生物的），即人造的  
B—Biological, 表示物理的+化学的 = 生物的  
C—Computational, 表示数学+计算机 = 计算的

# 计算智能与人工智能的关系

□ 第二种观点是大多数学者所持有的观点，其代表人物是艾伯哈特（R.C.Eberhart）：虽然人工智能与计算智能之间有重合，但计算智能是一个全新的学科领域，无论是生物智能还是机器智能，计算智能都是其**最核心**的部分，而人工智能则是**外层**。



- 事实上，CI和传统的AI只是智能的两个不同层次，各自都有自身的优势和局限性，相互之间只应该互补，而不能取代。
- 大量实践证明，只有把AI和CI很好地结合起来，才能更好地模拟人类智能，才是智能科学技术发展的正确方向。

# 进化计算（演化计算）

## □ 旅行商问题（Travelling Salesman Problem）

设有 $n$ 个城市，任意两个城市之间的距离如矩阵 $D=(d_{ij})_{n \times n}$  ( $i,j=1,2,\dots,n$ ) 所示，其中  $d_{ij}$  表示从城市  $I$  到城市  $j$  的距离。

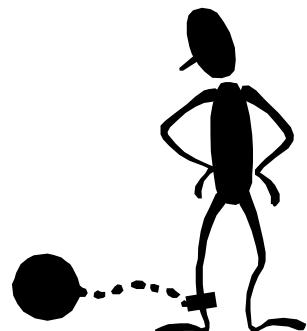
寻找这样一种周游方案：周游路线从某个城市开始，经过每个城市一次且仅一次，最后回到出发城市，使得周游路线的总长度最短。

$$f = \min \sum_{i=1}^n d_{\pi(i)} d_{\pi(i+1)}$$

$\pi(1)$ 为周游序列中第 $i$ 个城市的编号，且 $\pi(n+1)=\pi(1)$

$(n-1)!/2$

如果使用枚举法，复杂度将随着城市数目的增加急剧增加。

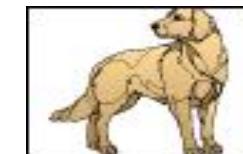
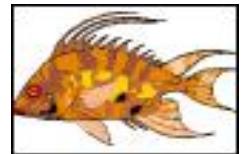


# 进化计算（演化计算）— 思考

低级、简单  $\xrightarrow{\text{进化}}$  高级、复杂

优化问题求解，能否借助这种方法呢？

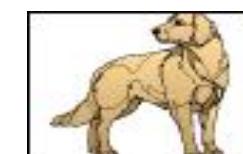
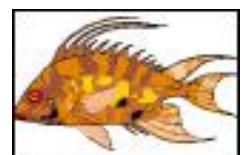
进化计算的概念：(Evolutionary Computation, EC) 是在达尔文 (Darwin) 的进化论和孟德尔 (Mendel) 的遗传变异理论的基础上产生的一种在基因和种群层次上模拟自然界生物进化过程与机制，进行问题求解的自组织、自适应的随机搜索技术。它以达尔文进化论的“物竞天择、适者生存”作为算法的进化规则，并结合孟德尔的遗传变异理论，将生物进化过程中的繁殖、变异、竞争和选择引入到算法中，是一种对人类智能的进化模拟方法。



# 进化计算（演化计算）— 思考

## □ 进化计算的发展史

- 20世纪60-70年代：**萌芽阶段**  
体系建立、计算量大、未受重视
- 20世纪80年代：**发展阶段**  
体系完善、计算机的发展、人工智能瓶颈
- 20世纪80年代以后：**成熟阶段**  
众多研究机构和学者的高度重视、研究热点

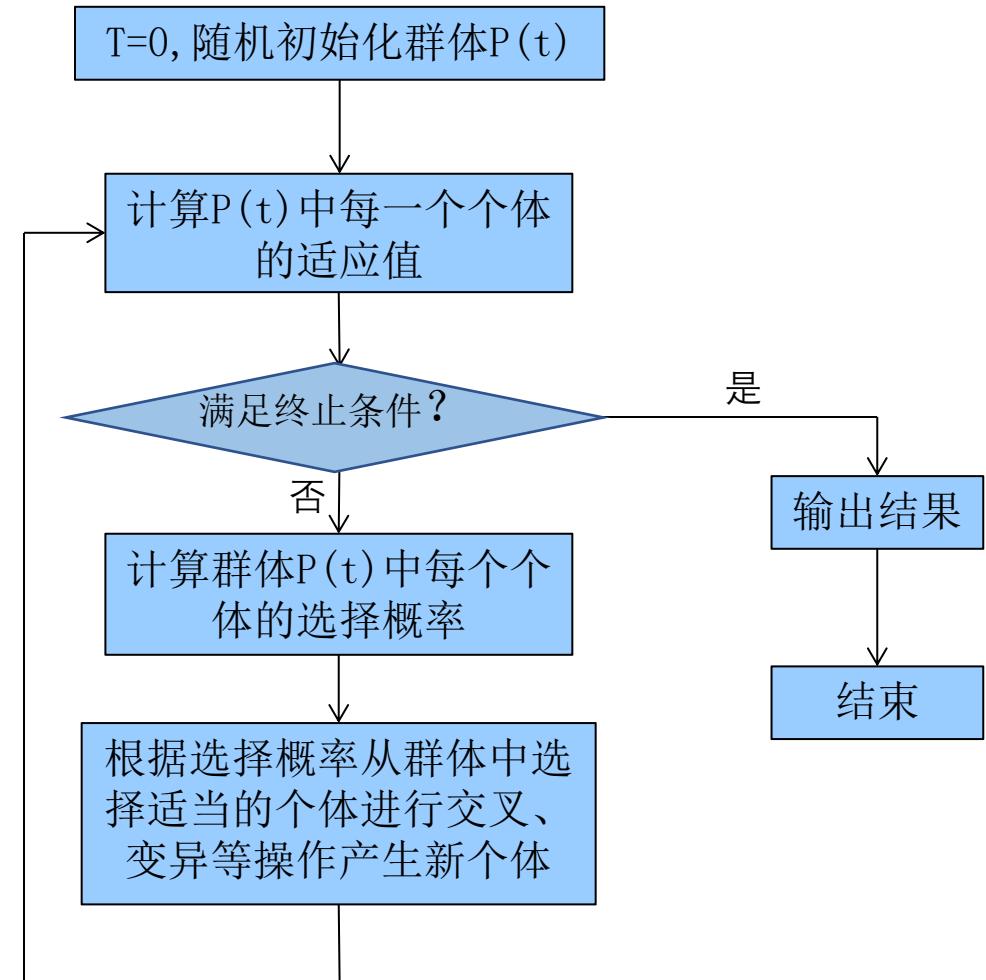


- 进化计算的主要分支：**遗传算法(Genetic Algorithms)**、**遗传规划 (Genetic Programming)**、**进化策略(Evolution Strategies)**三大分支。其中，遗传算法是进化计算中最初形成的一种具有普遍影响的模拟进化优化算法。



# 进化计算（演化计算）— 思考

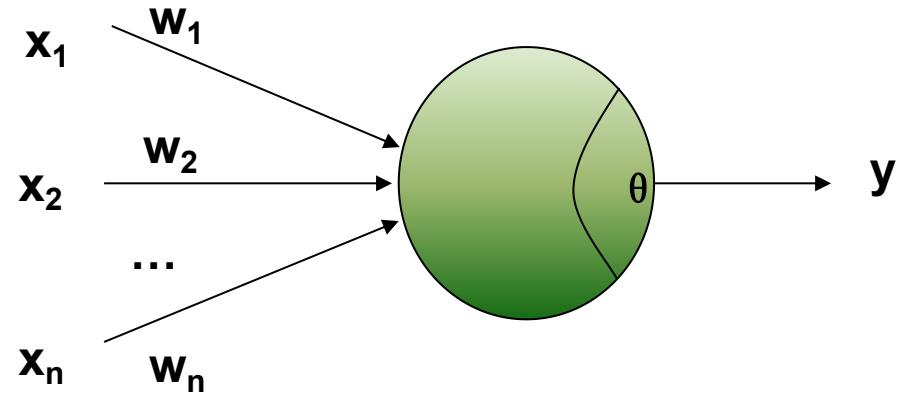
- 遗传算法（GA）模拟达尔文的遗传选择和自然淘汰的生物进化过程
- 1975年首先由Holland提出用于自然和人工系统的自适应行为研究(Adaptation in Natural and Artificial Systems)
- 组成：个体与群体、适应值函数、遗传操作、终止条件
- 特点：隐含并行性、过程性、非确定性、群体性、内在学习、统计性、稳健性、整体优化



# 神经网络（神经计算）

## □ 神经元具有的特性：

- 1) 多个输入
- 2) 一个输出
- 3) 输出只有抑制、兴奋两种状态



## □ 神经元工作过程：

累加各部分的加权输入，然后比较累加和是否超过阈值，如果超过，输出为兴奋状态，否则，输出为抑制状态。

□ 可见，输入权值和阈值对于输出状态影响巨大。

□ 输入权值和阈值是神经元学习的基础，权值和阈值调整的过程就是学习的过程。

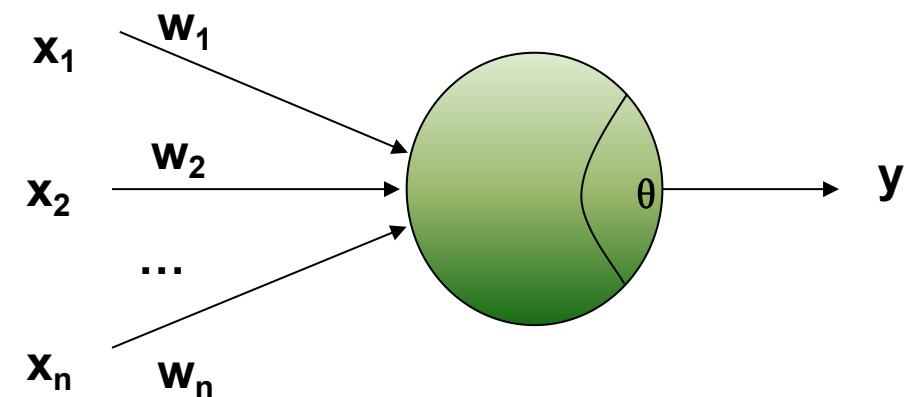
# 神经网络（神经计算）

## □ 神经元的学习：学习识别模式A和模式B

假定遇到模式A输出为0，模式B输出为1

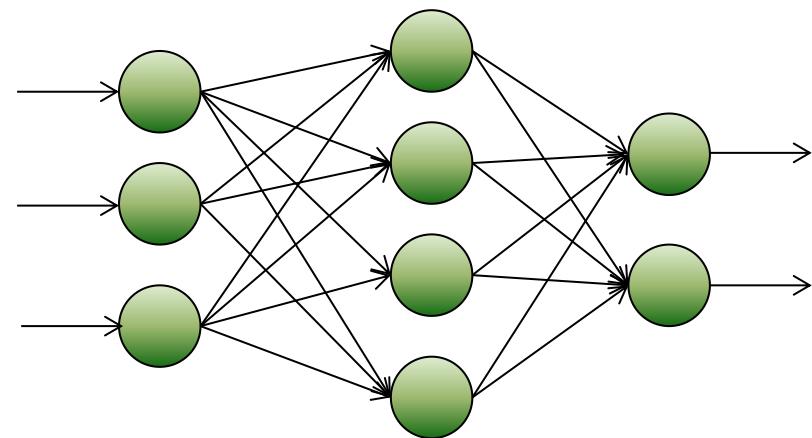
- 1) 随机化输入权值和阈值
- 2) 如果输入的是模式A，输出为1，调整权值，使之输出为0
- 3) 如果输入的是模式B，输出为0，调整权值，使之输出为1
- 4) 重复步骤2, 3, 直到训练完成

可以确定的是：如果此时输入的是模式A中的一个新样本，即使没有学过，也会有很大的概率判为模式A。



# 神经网络（神经计算）

## □ 神经网络



- 1) 可以充分逼近任意复杂的非线性关系
- 2) 采用并行分布处理方法，使得快速进行大量运算成为可能
- 3) 可学习和自适应不知道或不确定的系统

## □ 神经网络的特点：

- 1) 信息是通过神经元上的兴奋模式分布在网络上；
- 2) 信息处理是通过神经元之间同时相互作用的动态过程来完成的。

这就是人工智能研究的“结构（联接）主义”。

# 模糊系统（模糊计算）

## □ 从“秃头悖论”说起

有位先生的头发长的很好，即他的头“**不秃**”。

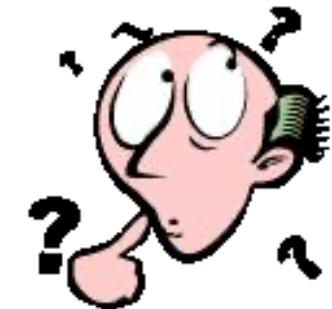
我们可以肯定：若将他头发拔去一根，他的头仍不秃；

拔去两根，他的头也不会秃；

再假定：拔去k根头发他还“**不秃**”；

则依常理，再拔1根（共拔去K+1根）头发，他也不会是“**秃**”。

则依数学归纳法，可得结论：如此拔下去，头发拔光了，他也“**不秃**”！



# 模糊理论的产生

在模糊集合中，给定范围内元素对它的隶属关系不一定只有“是”或“否”两种情况，而是用介于0和1之间的实数来表示隶属成程度，还存在中间过渡状态

## □ 精确性的数学与现实生活中的模糊的矛盾

- 在日常生活中，经常遇到许多模糊事物，没有分明的数量界限，要使用一些模糊的词句来形容、描述。
- 需要寻找一种描述和加工模糊信息的数学工具，这就推动数学家深入研究模糊理论。

## □ 1965年，扎德 (Zadeh) 提出了模糊集合论 (Fuzzy Set) 。

1. 提出用“模糊集合”作为表现模糊事物的数学模型。
2. 提出“隶属函数”概念



# 模糊理论—应用

- 1974年，印度裔英国学者马德尼(E.H.Mamdani)首先将模糊理论用于锅炉和蒸汽机的控制
- 1983 Takagi T. and Sugino M. 提出具有线性规则后项的模糊控制器(**TS模糊控制器**)
- 80年代以后，采用模糊控制技术的家电产品大量上市，模糊技术在图像识别、自动控制、市场预测、人工智能等领域普遍应用

## Fuzzy Related Journals:

- *IEEE Trans. On Fuzzy Systems*
- *Fuzzy Sets and Systems*
- *Journal of Intelligent and Fuzzy Systems*
- *Fuzzy Optimization and Decision Making*



# AI三要素



# 大数据

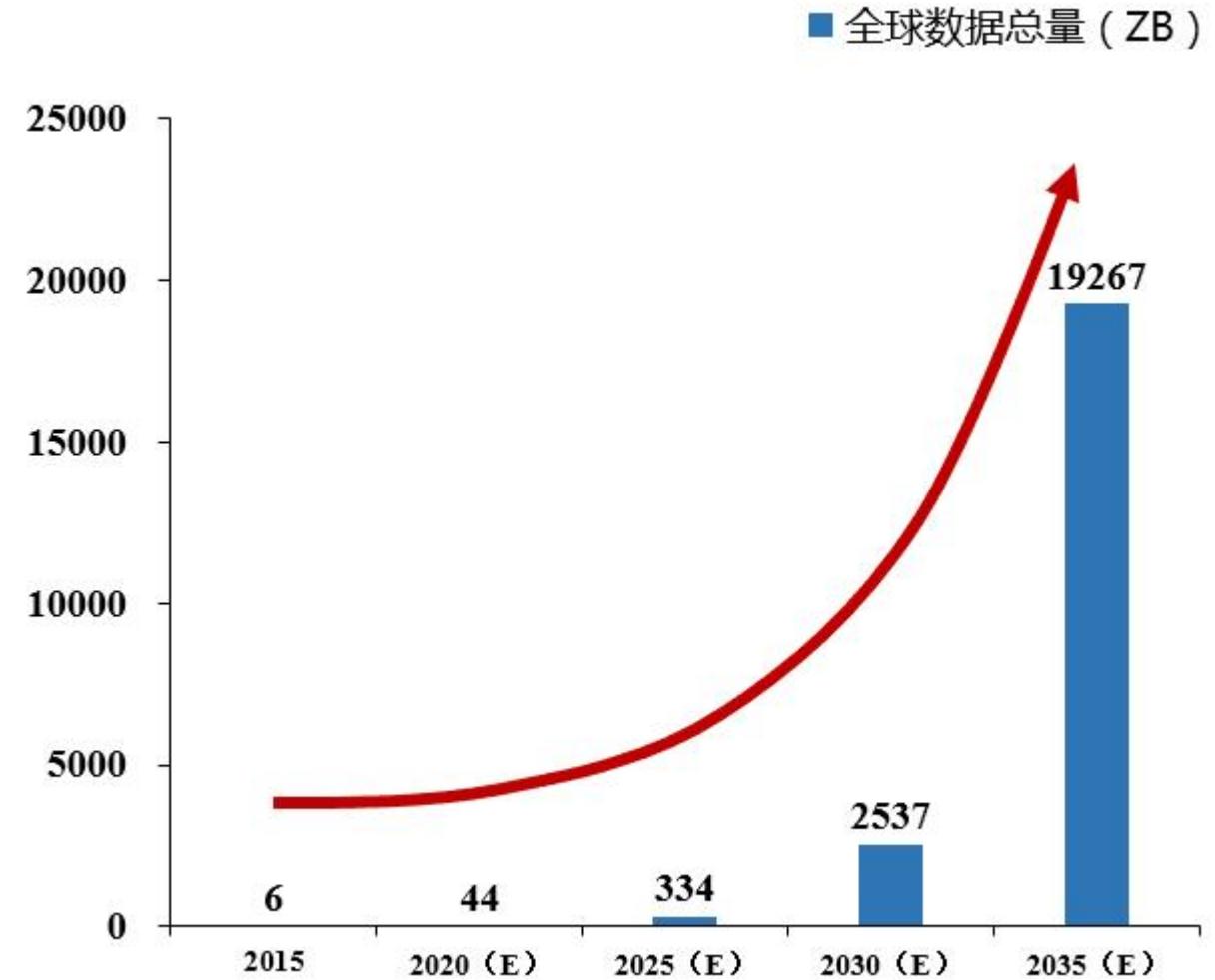
## 何为大? — 数据度量

- **1Byte=8Bit**
- **1KB=1.024 Bytes**
- **1MB=1,024KB=1,048,576 Bytes**
- **1GB=1,024MB=1048,576KB=1073,741824 Bytes**
- **1TB=1,024GB=1,048,576MB=1099,511,627.776 Bytes**
- **1PB=1024TB=1048576GB=1125,899,906,842624Bytes**
- **1EB=1,024PB=1,048,576TB=1,152921,504,606846976 Bytes**
- **1ZB=1,024EB =1,180,591,620,717,411,303,424 Bytes**
- **1YB=1,024ZB=1.208,925,819,614,629,174,706,176 Bytes**

# 大数据

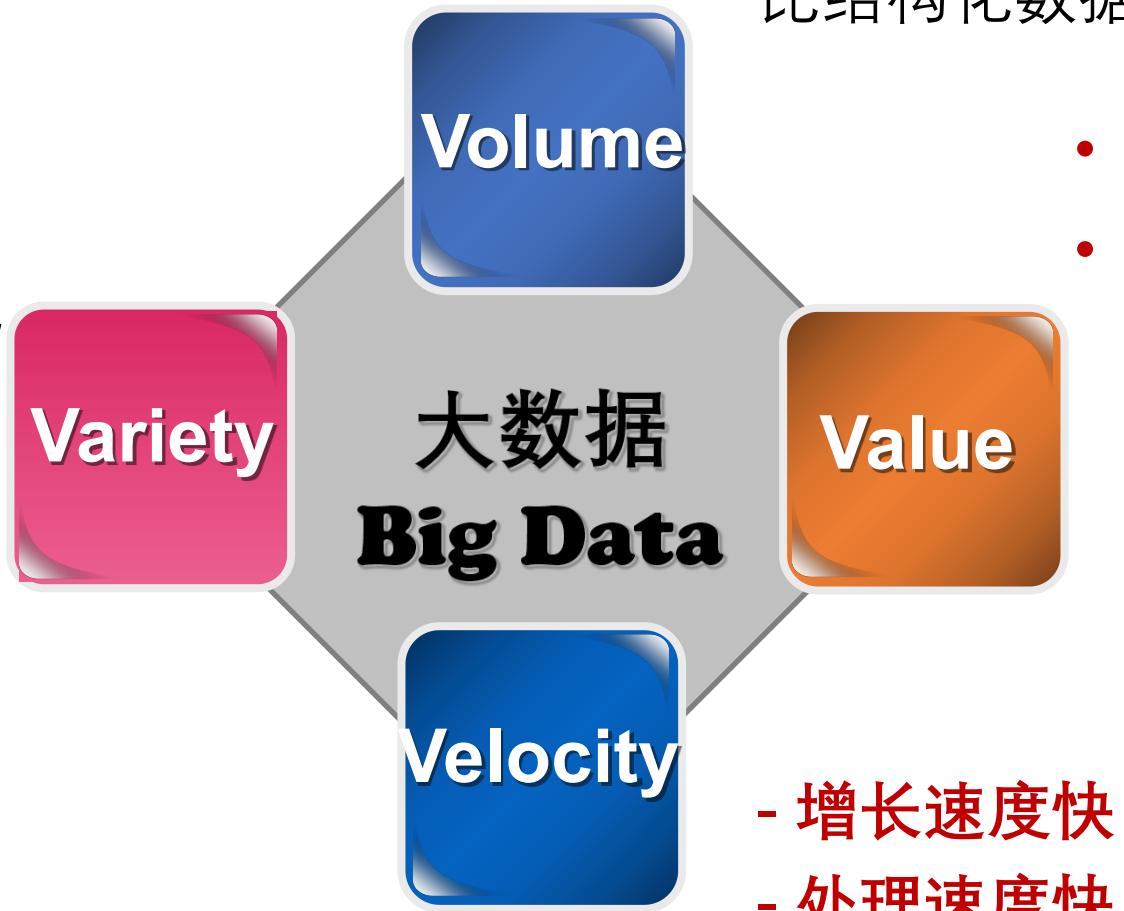
麦肯锡研究院：大数据是指其大小超过了常规数据库工具获取、储存、管理和分析能力的数据集

杰姆·格雷 (Jim Gray) 提出著名的“**新摩尔定律**”，即人类有史以来的数据总量，每过18个月就会翻一番。



# 大数据的4V特征

- **数据来源多：**物联网数据、互联网数据、行业/企业内部数据
- **数据类型多：**结构化数据少，80%以上为图片、音频、视频、模型、文档等非结构化和半结构化数据
- **关联性强：**数据之间频繁交互



- 非结构化数据的超大规模和增长
- 占总数据量的**80-90%**
- 比结构化数据**增长快**10倍到50倍
  - 大量的不相关信息
  - 大数据有巨大的潜在价值，但同其呈几何指数爆发式增长相比，单元模块数据的**价值密度较低**，给开发海量数据增加了难度和成本。
- 增长速度快
- 处理速度快

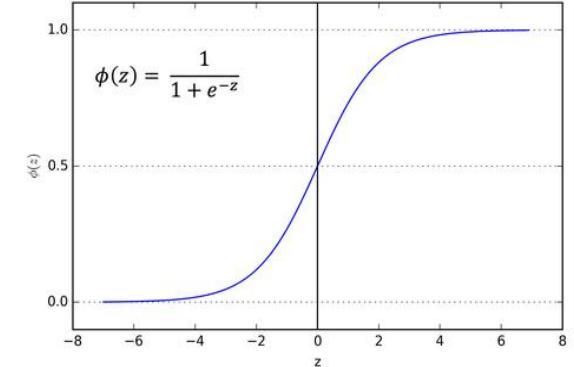
# 算法

## 1. 起步发展期：1943年—20世纪60年代

- 1943年，美国神经科学家麦卡洛克（Warren McCulloch）和逻辑学家皮茨（Water Pitts）提出**神经元**的数学模型，这是现代人工智能学科的奠基石之一。
- 1950年，艾伦·麦席森·图灵（Alan Mathison Turing）提出“**图灵测试**”，让机器产生智能这一想法开始进入人们的视野。
- 1950年，克劳德·香农（Claude Shannon）提出**计算机博弈**。
- 1956年，达特茅斯会议正式使用了人工智能（artificial intelligence, AI）。这是人类历史上**第一次人工智能**研讨，标志着人工智能的诞生。
- 1957年，弗兰克·罗森布拉特（Frank Rosenblatt）在一台IBM-704计算机上模拟实现了一种“**感知机**”（Perceptron）的神经网络模型。

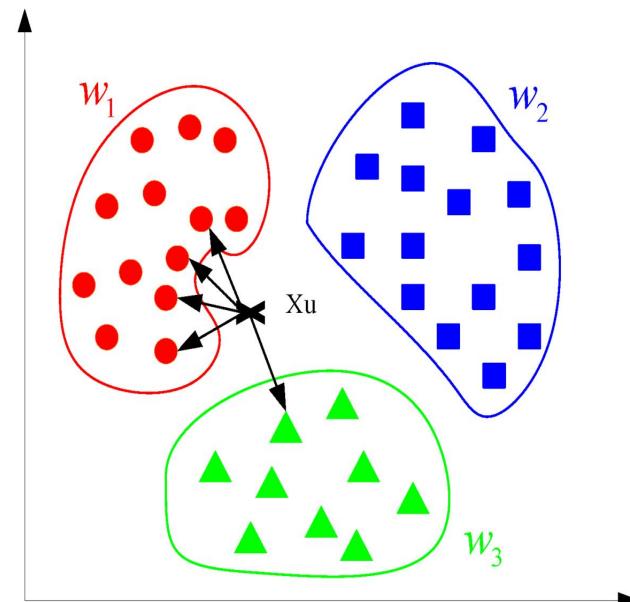
# 算法

- 1958年，David Cox提出了**Logistic Regression**。
- 1959年，Arthur Samuel **机器学习**概念：Field of study that gives computers the ability to learn without being explicitly programmed.  
(研究如何让计算机不需要显式的程序也可以具备学习的能力)。
- 1961年，Leonard Merrick Uhr 和 Charles M Vossler发表了题为A Pattern Recognition Program That Generates, Evaluates and Adjusts its Own Operators 的**模式识别**论文。
- 1965年，吉德 (I. J. Good) 发表了一篇对人工智能未来可能对人类构成威胁的文章——“**AI威胁论**”的先驱。他认为机器的超级智能和无法避免的智能爆炸最终将超出人类可控范畴。



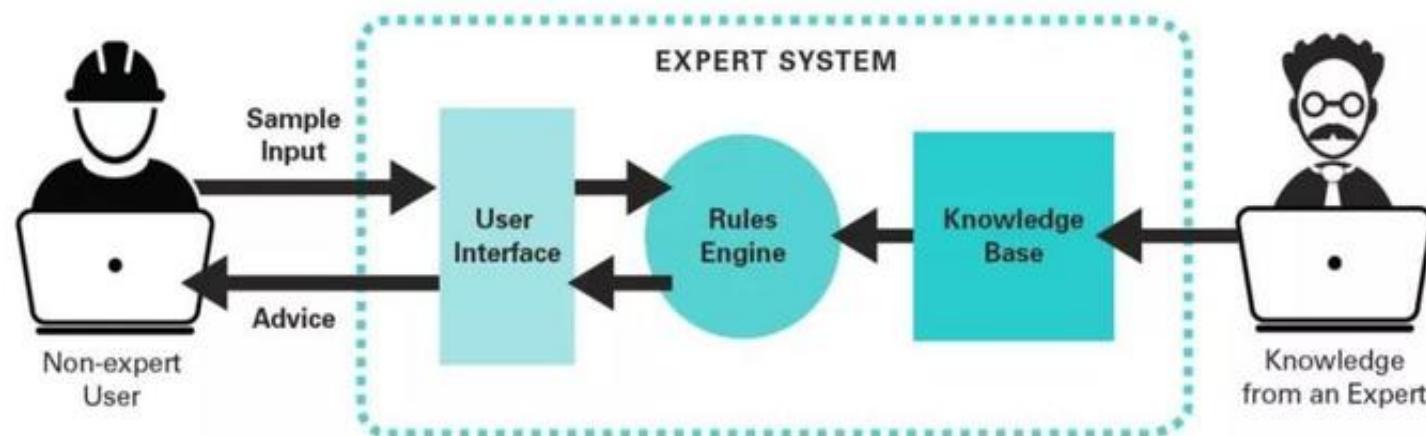
# 算法

- 1966 年， 麻省理工学院科学家 Joseph Weizenbaum 在 ACM 上发表了题为《ELIZA-a computer program for the study of natural language communication between man and machine》 文章描述了 ELIZA 的程序如何使人与计算机在一定程度上进行**自然语言对话**成为可能。
- 1967年， Thomas 等人提出**K最近邻算法** (The k nearest neighbor algorithm, **KNN**) 。



# 算法

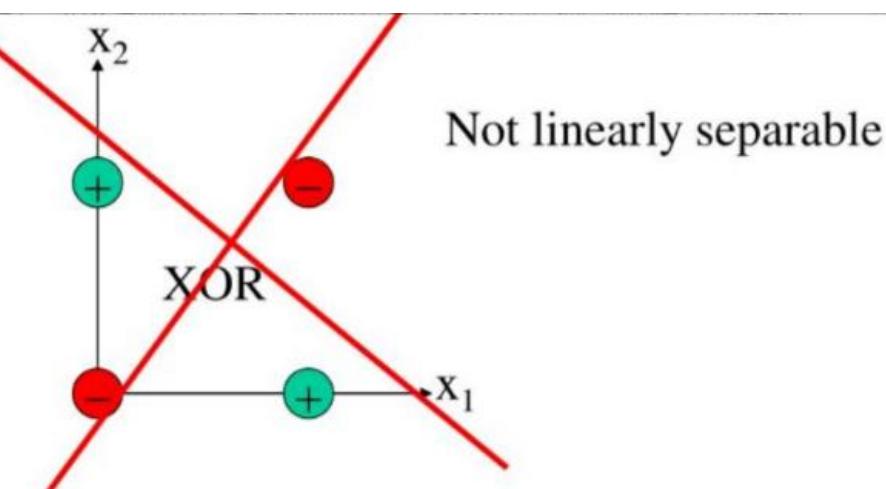
- 1968年，爱德华·费根鲍姆（Edward Feigenbaum）提出首个**专家系统**DENDRAL，并对**知识库**给出了初步的定义，这也孕育了后来的第二次人工智能浪潮。



The user inputs information about: the equipment,  
the component  
and the oil being tested ... The Expert System does the rest.

# 算法

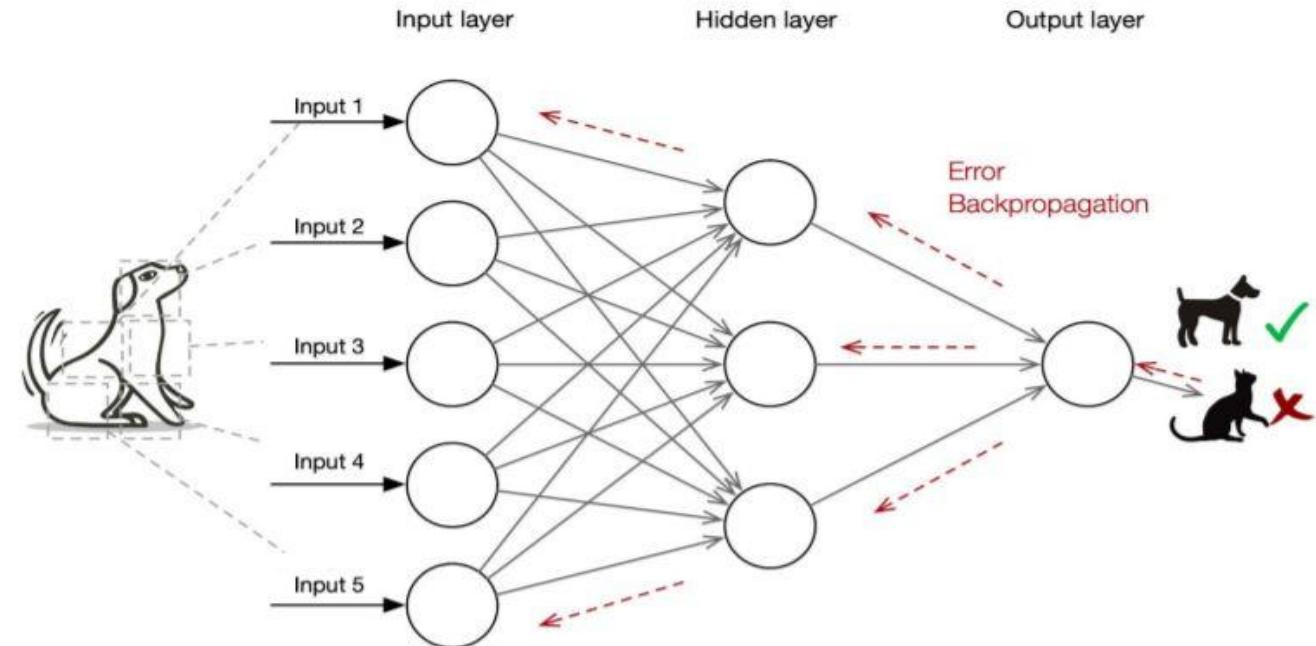
- 1969年，马文·明斯基（Marvin Minsky）的著作《感知器》提出对 XOR **线性不可分**的问题：单层感知器无法划分XOR原数据，解决这个问题需要引入更**高维非线性网络**（**MLP**, 至少需要两层），但多层网络并无有效的训练算法。这给神经网络研究以沉重的打击，神经网络的研究走向长达10年的低潮时期。



# 算法

## 2. 反思发展期：20世纪70年代

- 1974年，哈佛大学沃伯斯(Paul Werbos)博士论文里，首次提出了通过误差的**反向传播(BP)**来训练人工神经网络，但在该时期未引起重视。
- 1975年，马文·明斯基(Marvin Minsky)在论文《知识表示的框架》(A Framework for Representing Knowledge)中提出用于人工智能中的**知识表示学习框架**理论。



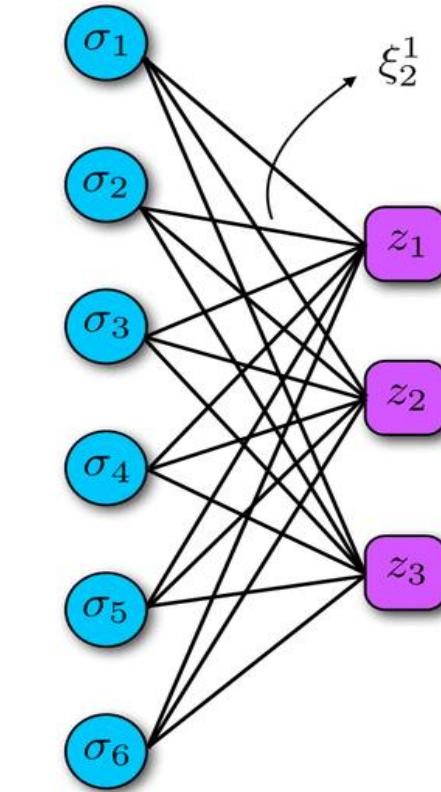
# 算法

## 3. 应用发展期：20世纪80年代

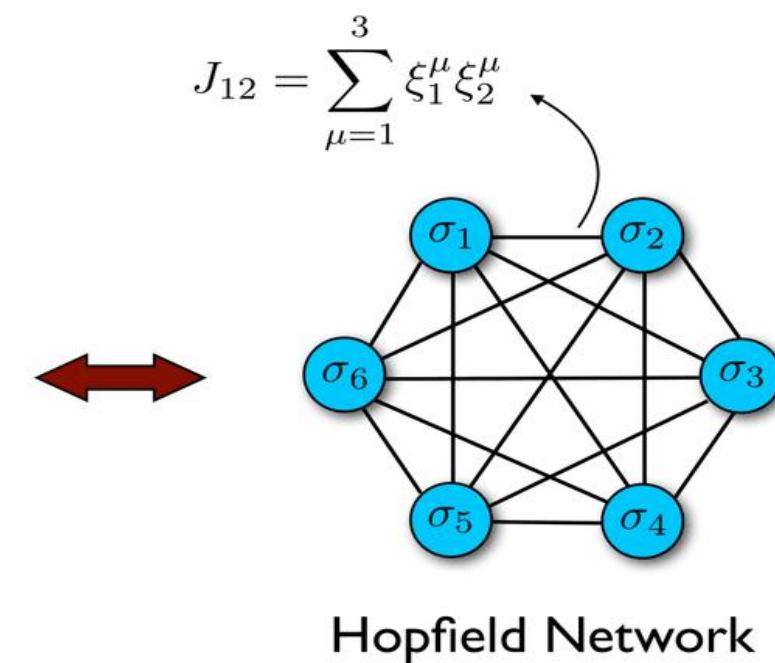
- 1980年，在美国卡内基梅隆大学(CMU)召开了**第一届机器学习国际研讨会**，标志着机器学习研究已在全世界兴起。
- 1980年，德鲁·麦狄蒙（Drew McDermott）和乔恩·多伊尔（Jon Doyle）提出非单调逻辑，以及后期的**机器人系统**。
- 1981年，保罗（R.P.Paul）出版**第一本机器人学**课本，“Robot Manipulator: Mathematics, Programmings and Control”，标志着**机器人学科**走向成熟。

# 算法

- 1982年，约翰·霍普菲尔德（John Hopfield）发明了**霍普菲尔德网络**，这是最早的RNN的雏形。霍普菲尔德神经网络模型是一种**单层反馈神经网络**。
- 1983年，Terrence Sejnowski, Hinton等人发明了**玻尔兹曼机**（Boltzmann Machines），本质是一种无监督模型，用于对输入数据进行重构以提取数据特征做预测分析。



Boltzmann Machine

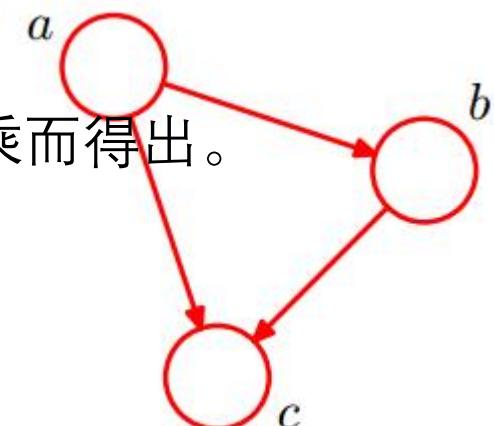


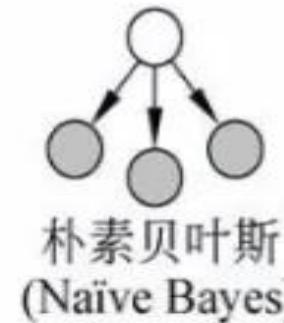
# 算法

- 1985年，朱迪亚·珀尔提出**贝叶斯网络(Bayesian network)**，他以倡导人工智能的概率方法和发展贝叶斯网络而闻名，还因发展了一种基于结构模型的因果和反事实推理理论而受到赞誉。

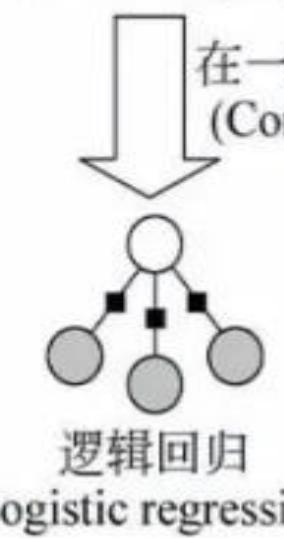
- 贝叶斯网络拓扑结构是一个**有向无环图(DAG)**，通过把某个研究系统中涉及的随机变量，根据是否条件独立绘制在一个有向图中，以描述随机变量之间的条件依赖，用圈表示**随机变量**(random variables)，用箭头表示**条件依赖**(conditional dependencies)就形成了贝叶斯网络。

- 对于任意的随机变量，其联合概率可由各自的局部条件概率分布相乘而得出。如图中b依赖于a(即：  $a \rightarrow b$ )， c依赖于a和b， a独立无依赖，根据贝叶斯定理有  $P(a,b,c) = P(a) * P(b|a) * P(c|a,b)$

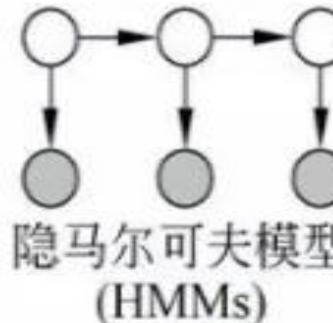




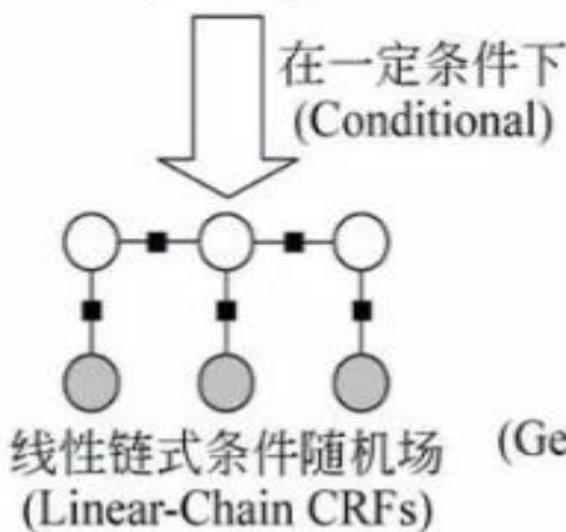
序列  
(Sequence)



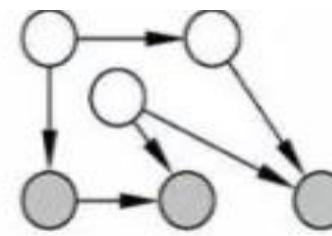
在一定条件下  
(Conditional)



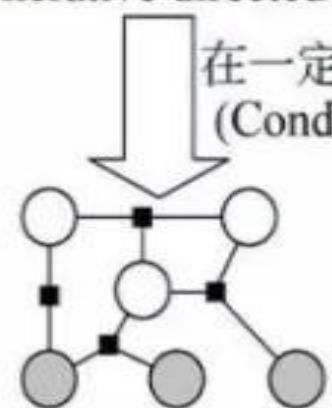
一般图  
(General graphs)



一般图  
(General graphs)



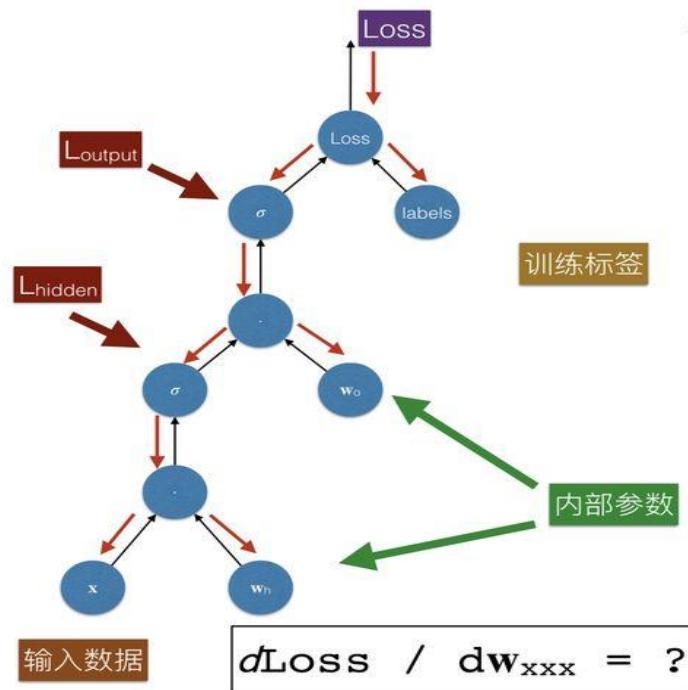
生成式有向图模型  
(Generative directed models)



在一定条件下  
(Conditional)

- 1986年，辛顿(Geoffrey Hinton)等人先后提出了**多层感知器(MLP)**与**反向传播(BP)****训练相结合**的理念（该方法在当时计算力上还是有很多挑战，基本上都是和链式求导的梯度算法相关的），这也解决了单层感知器不能做非线性分类的问题，开启了神经网络新一轮的高潮。

## 反向传播



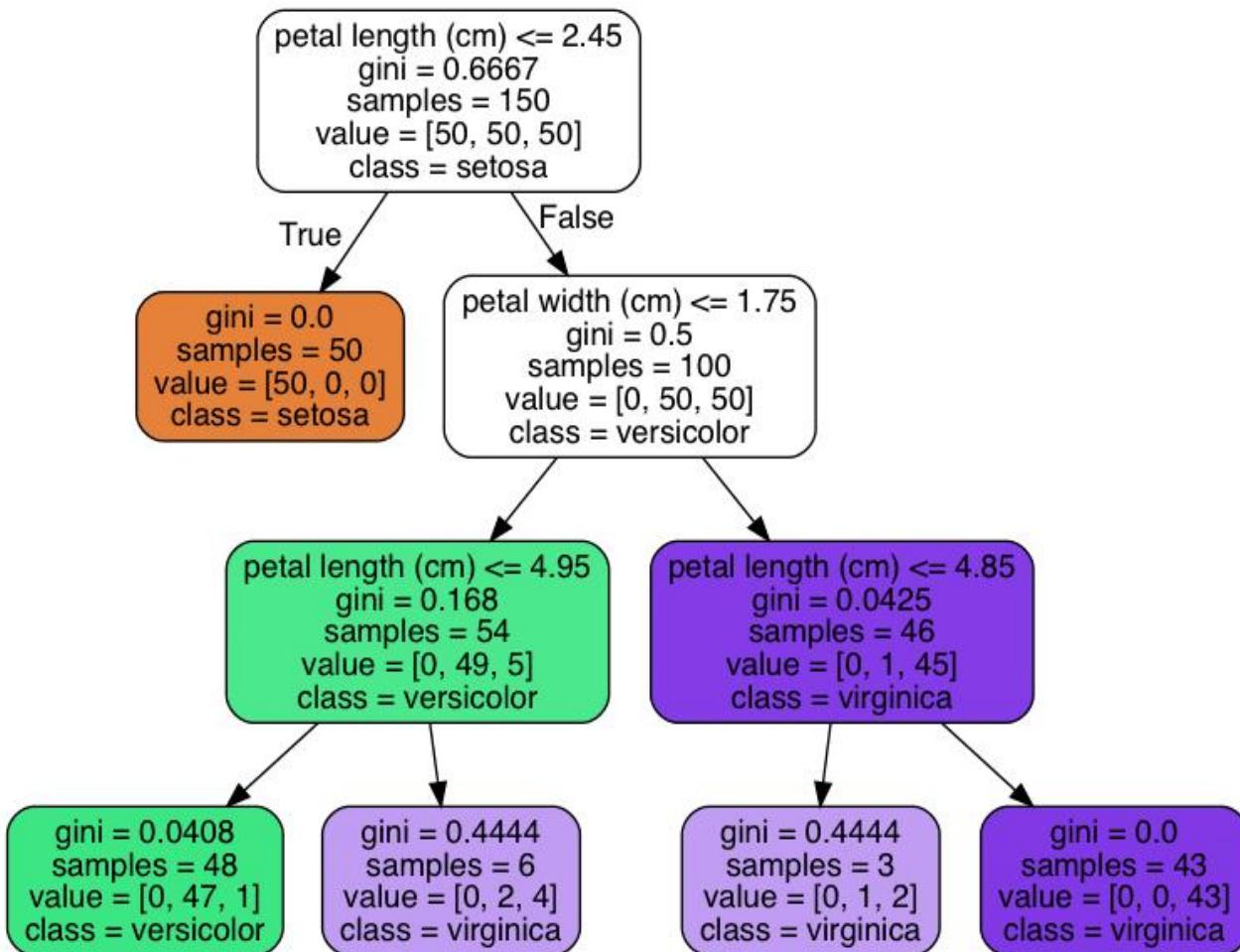
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

$$d\text{Loss} / d\mathbf{w}_o = \\ (\text{dLoss}/\text{dL}_{\text{Output}}) \cdot (\text{dL}_{\text{Output}}/\text{d}\mathbf{w}_o)$$

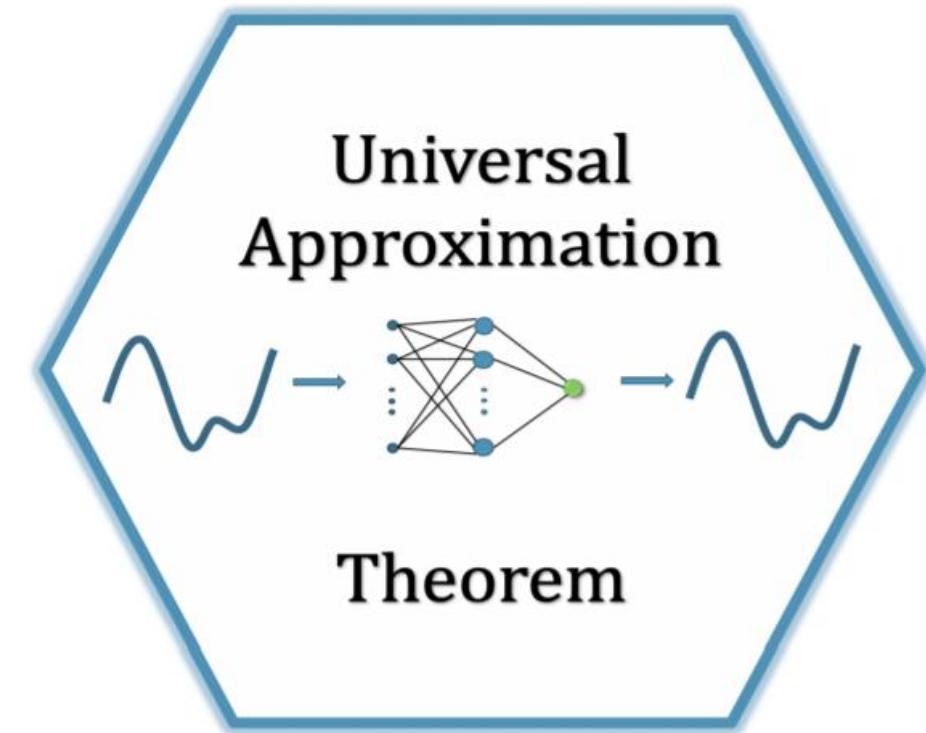
$$d\text{Loss} / d\mathbf{w}_h = \\ (\text{dLoss}/\text{dL}_{\text{Output}}) \cdot (\text{dL}_{\text{Output}}/\text{dL}_{\text{hidden}}) \cdot \\ (\text{dL}_{\text{hidden}}/\text{d}\mathbf{w}_h)$$

- 1986年，昆兰（Ross Quinlan）提出ID3决策树算法。

- 决策树模型可视为多个规则(if, then)的组合，与神经网络黑盒模型截然不同的是它拥有良好的模型解释性。

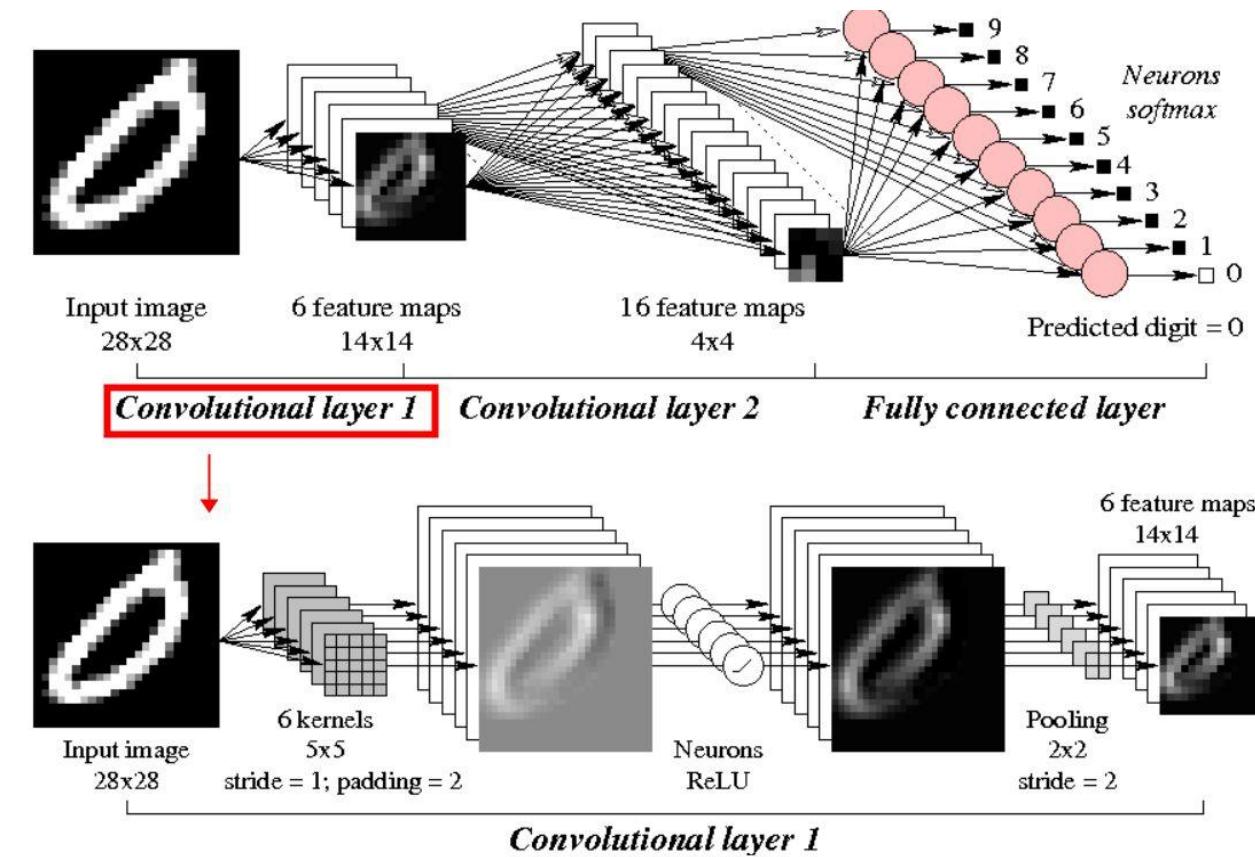
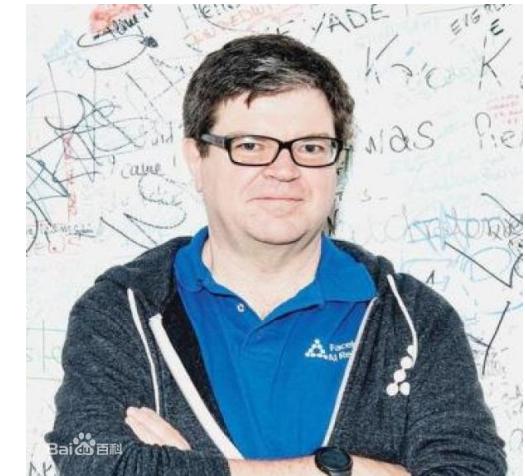


- 1989年，George Cybenko证明了“**万能近似定理**”（universal approximation theorem）。简单来说，多层前馈网络可以近似任意函数，其表达力和图灵机等价。这就从根本上消除了Minsky对神经网络表达力的质疑。
- “万能近似定理”可视为神经网络的基本理论：一个前馈神经网络如果具有线性层和至少一层具有“挤压”性质的激活函数（如 sigmoid 等），给定网络足够数量的隐藏单元，它可以以任意精度近似任何从一个有限维空间到另一个有限维空间的 borel 可测函数。



- 1989年， LeCun (CNN之父) 结合反向传播算法与权值共享的卷积神经层发明了**卷积神经网絡 (Convolutional Neural Network, CNN)**，并首次将卷积神经网络**成功应用**到美国邮局的手写字符识别系统中。

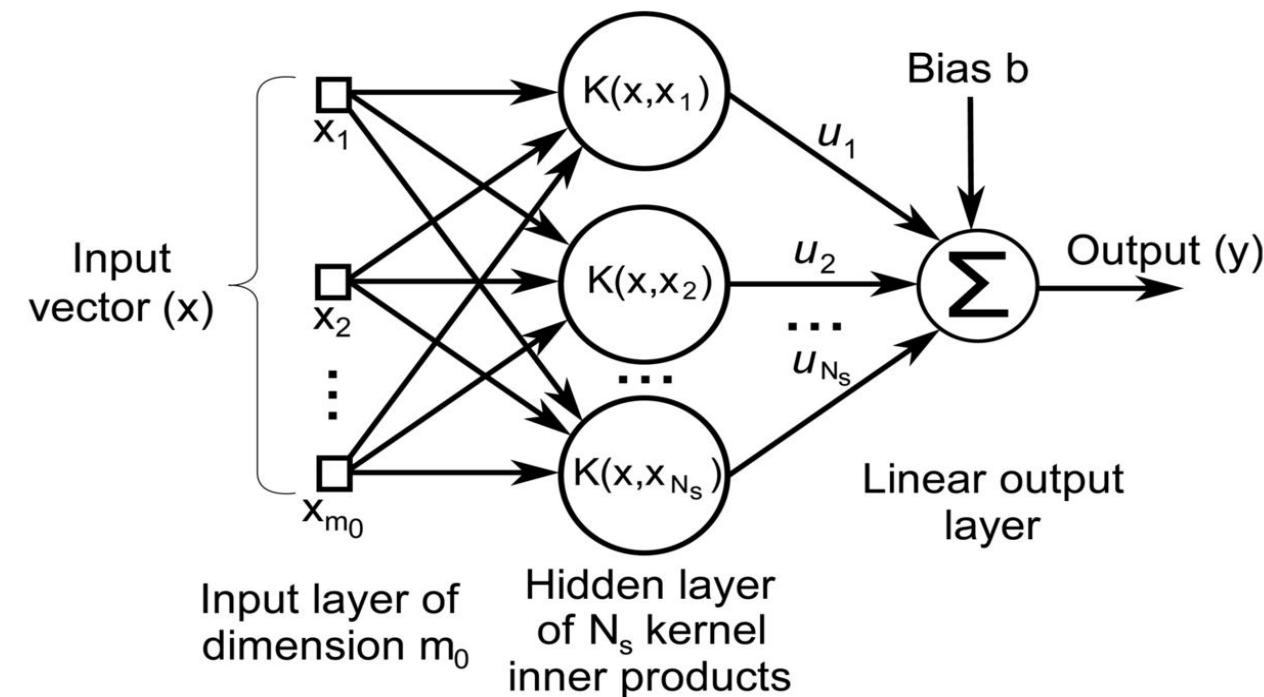
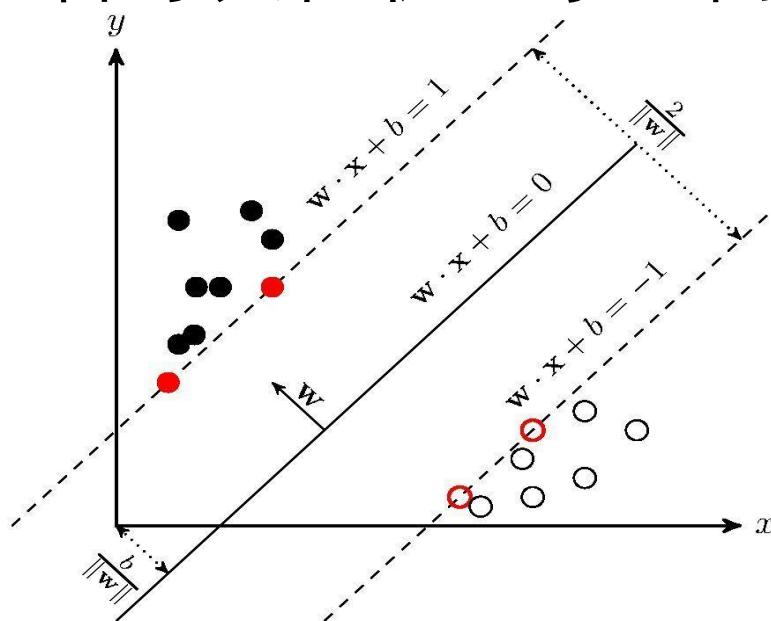
杨立昆 (Yann LeCun) ，于1960年出生于法国巴黎附近，担任Facebook首席人工智能科学家和纽约大学教授，2018年图灵奖 (Turing Award) 得主。



# 算法

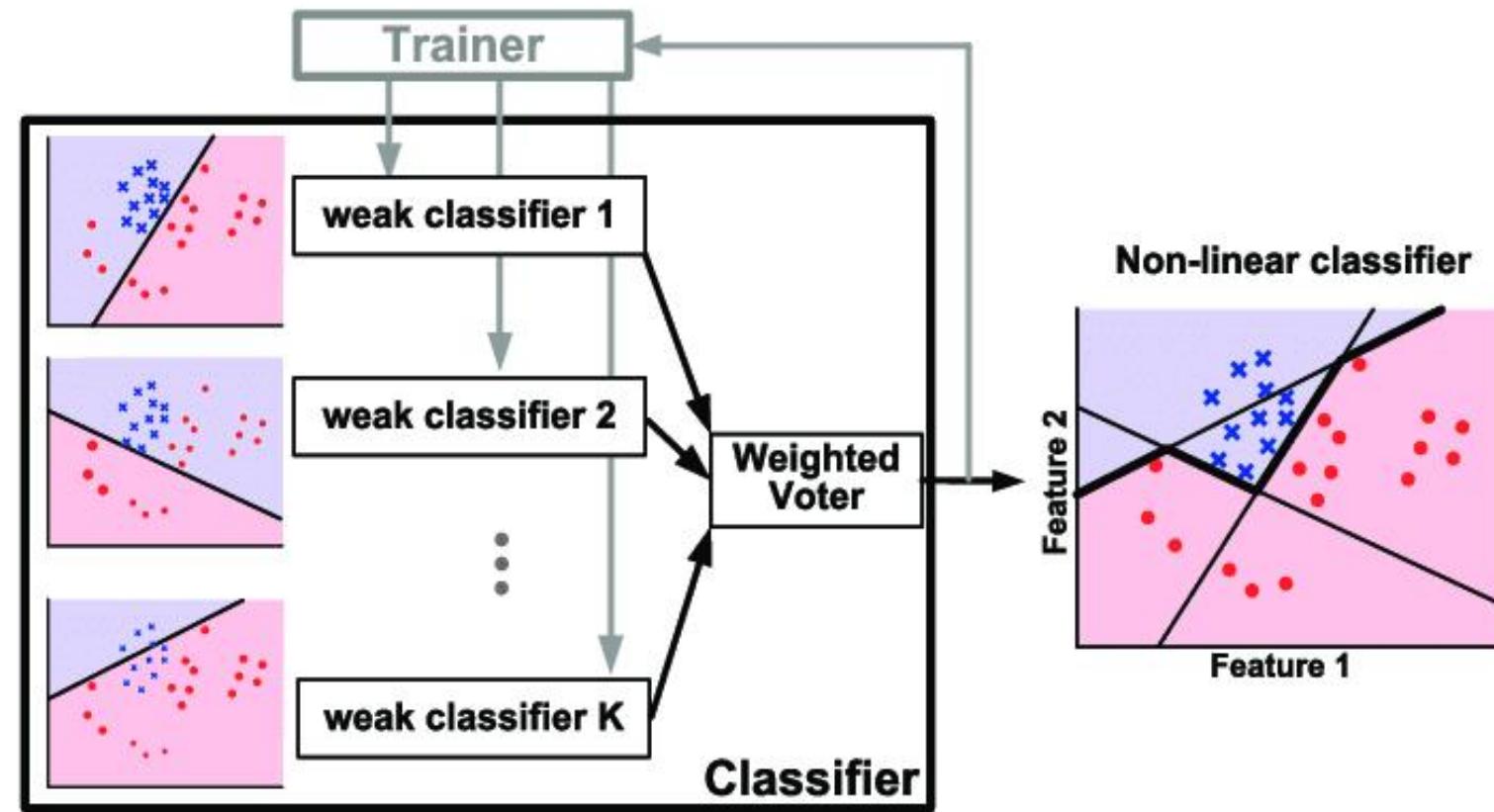
## 4. 平稳发展期：20世纪90年代—2010年

- 1995年，Cortes和Vapnik提出联结主义经典的**支持向量机(Support Vector Machine)**，它在解决小样本、非线性及高维模式识别中表现出许多特有的优势，并能够推广应用到函数拟合等其他机器学习问题中。

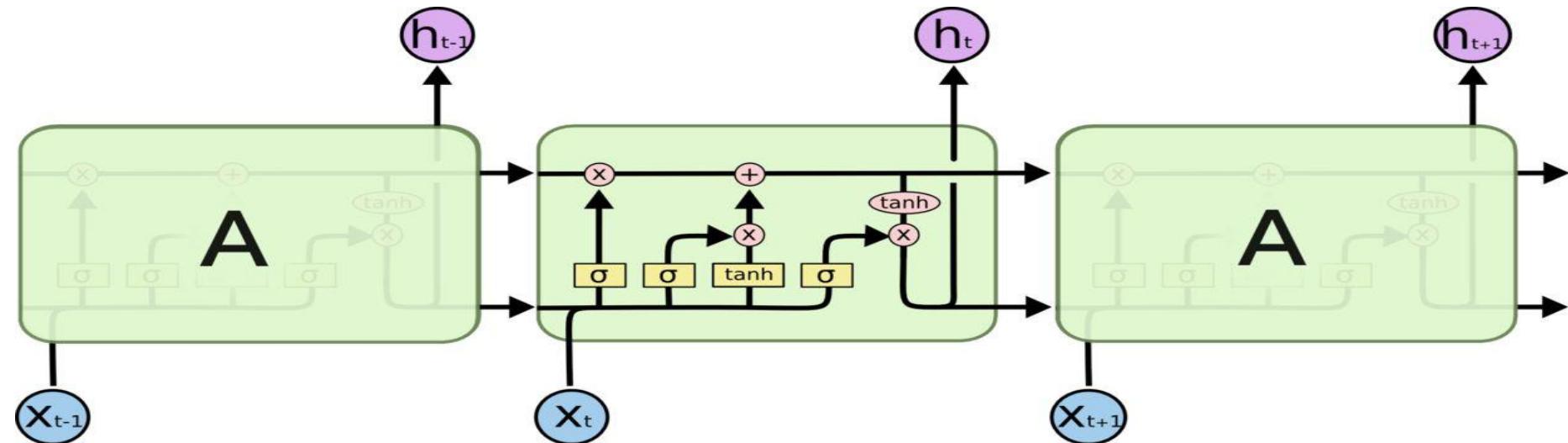


- 1995年，Freund和schapire提出了**AdaBoost (Adaptive Boosting)算法**。AdaBoost采用的是Boosting集成学习方法——串行组合弱学习器以达到更好的泛化性能。

- Adaboost迭代算法基本思想主要是通过调节的每一轮各训练样本的权重(错误分类的样本权重更高)，串行训练出不同分类器。最终以各分类器的准确率作为其组合的权重，一起加权组合成强分类器。

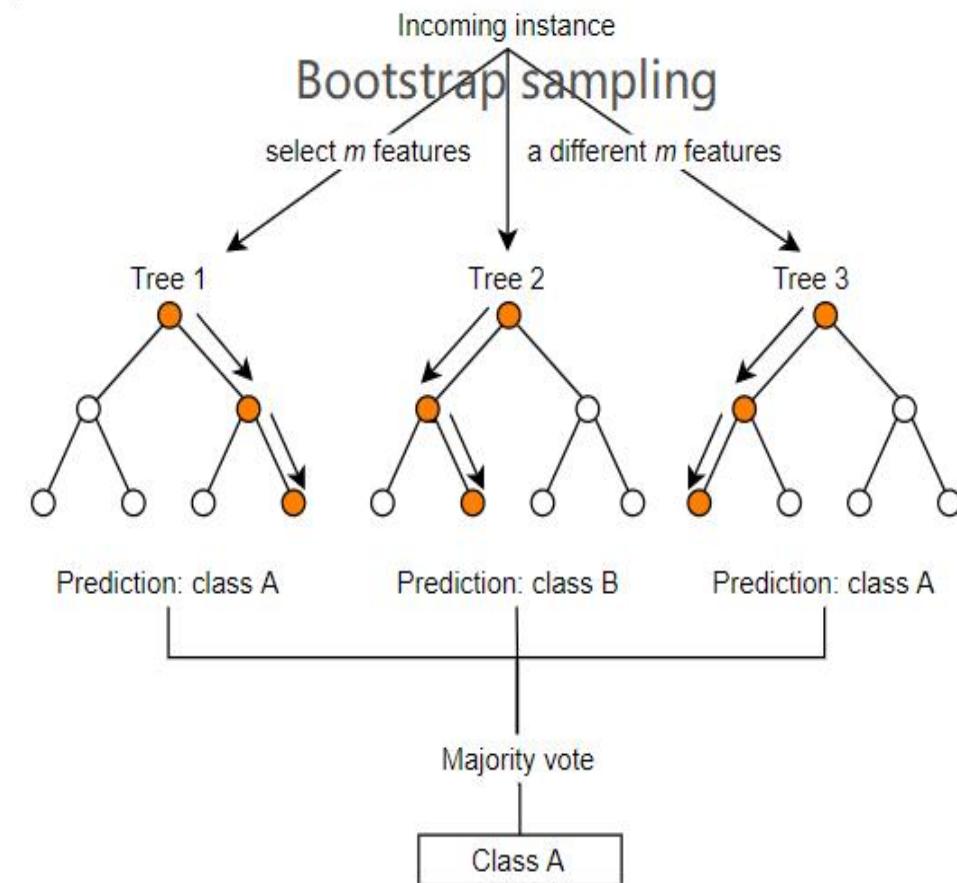


- 1997年，国际商业机器公司（IBM）**深蓝超级计算机战胜了国际象棋世界冠军**卡斯帕罗夫。深蓝是基于暴力穷举实现国际象棋领域的智能，通过生成所有可能的走法，然后执行尽可能深的搜索，并不断对局面进行评估，尝试找出最佳走法。
- 1997年，Sepp Hochreiter 和 Jürgen Schmidhuber提出了**长短期记忆神经网络(LSTM)**。

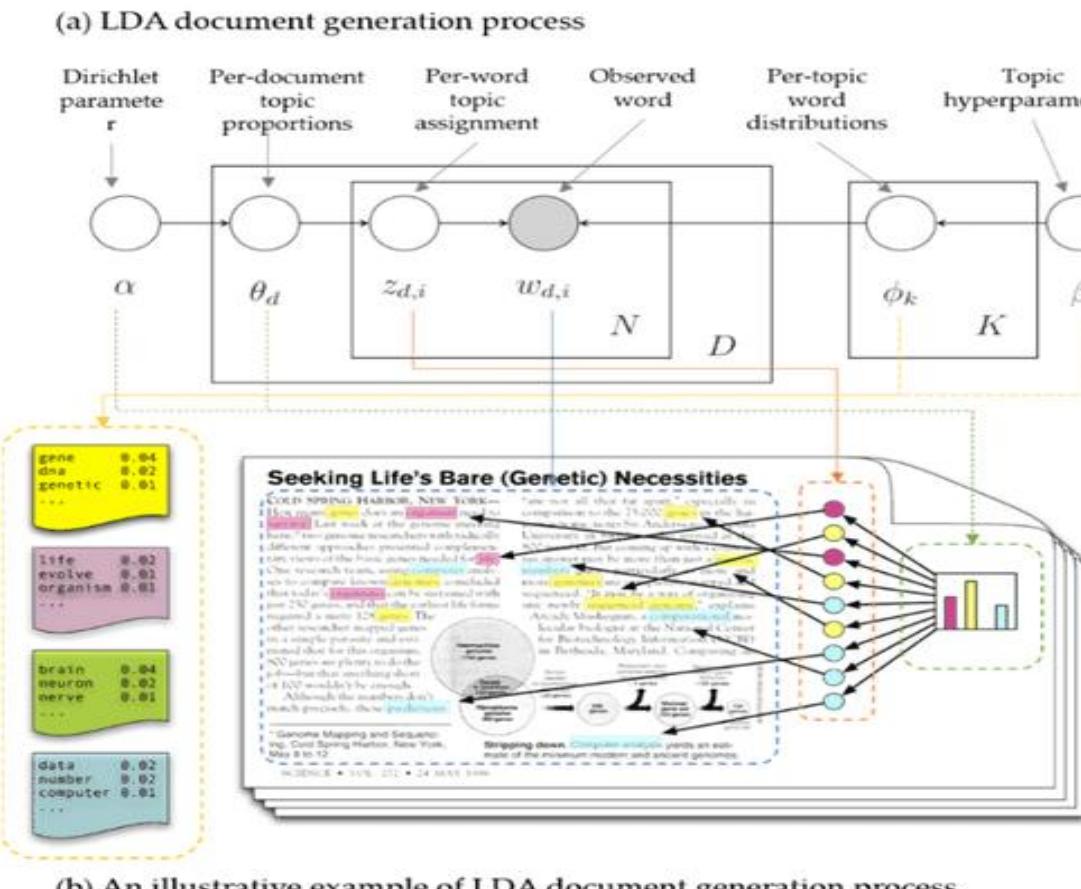


- 1998年，万维网联盟的蒂姆·伯纳斯·李(Tim Berners-Lee)提出**语义网(Semantic Web)**的概念。其核心思想是：通过给万维网上的文档(如HTML)添加能够被计算机所理解的语义(**Meta data**)，从而使整个互联网成为一个基于语义链接的通用信息交换媒介。
- 2001年，John Lafferty首次提出**条件随机场模型 (Conditional random field, CRF)**。CRF是基于贝叶斯理论框架的判别式概率图模型，给定条件随机场 $P(Y | X)$ 和输入序列 $x$ ，求条件概率最大的输出序列 $y^*$ 。在许多自然语言处理任务中，如分词、命名实体识别等表现尤为出色。

- 2001年，布雷曼博士提出**随机森林（Random Forest）**，将多个有差异的弱学习器(决策树)Bagging并行组合，通过建立多个拟合较好且有差异的模型去组合决策，以优化泛化性能的一种**集成学习方法**。多样差异性可减少对某些特征噪声的依赖，降低方差（过拟合），组合决策可消除些学习器间的偏差。
- 基本思想：对于每一弱学习器(决策树)抽样构造其训练集，并随机抽取其可用特征子集，即以训练样本及特征空间的多样性训练出N个不同的弱学习器，最终结合N个弱学习器的预测（类别或者回归预测数值），取最多数类别或平均值作为最终结果。



- 2003年，David Blei, Andrew Ng和Michael I. Jordan于2003年提出**LDA (Latent Dirichlet Allocation)**。LDA是一种无监督方法，用来推测文档的主题分布，将文档集中每篇文档的主题以概率分布的形式给出，可以根据主题分布进行主题聚类或文本分类。



(c) Two outputs of LDA

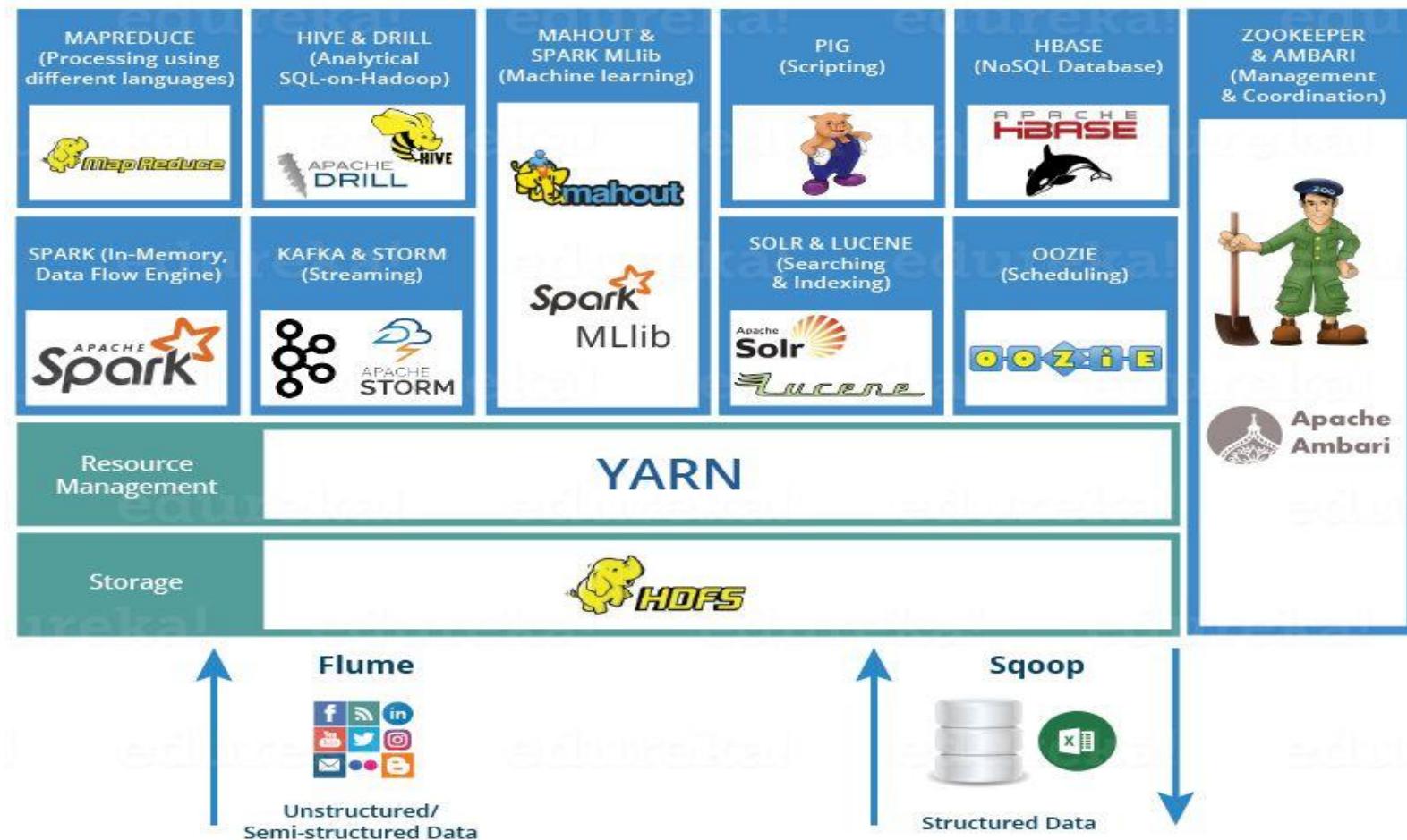
(c-1) Per-document topic proportions ( $\theta_d$ )

|       | Topic 1 | Topic 2 | Topic 3 | ... | Topic K |
|-------|---------|---------|---------|-----|---------|
| Doc 1 | 0.20    | 0.50    | 0.10    | ... | 0.10    |
| Doc 2 | 0.50    | 0.02    | 0.01    | ... | 0.40    |
| Doc 3 | 0.05    | 0.12    | 0.48    | ... | 0.15    |
| ...   | ...     | ...     | ...     | ... | ...     |
| Doc N | 0.14    | 0.25    | 0.33    | ... | 0.14    |

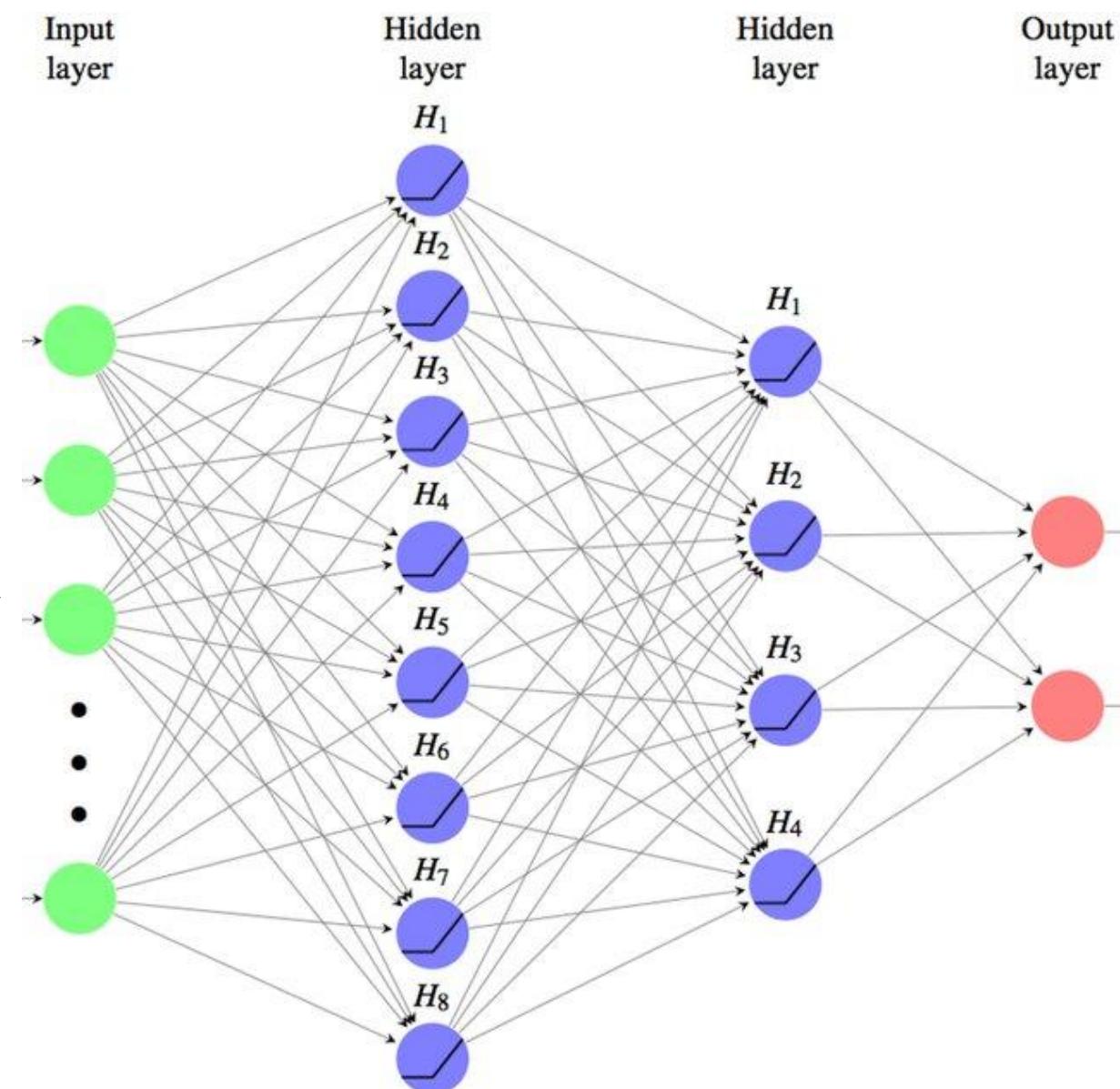
(c-2) Per-topic word distributions ( $\phi_k$ )

|        | Topic 1 | Topic 2 | Topic 3 | ... | Topic K |
|--------|---------|---------|---------|-----|---------|
| word 1 | 0.01    | 0.05    | 0.05    | ... | 0.10    |
| word 2 | 0.02    | 0.02    | 0.01    | ... | 0.03    |
| word 3 | 0.05    | 0.12    | 0.08    | ... | 0.02    |
| ...    | ...     | ...     | ...     | ... | ...     |
| word N | 0.04    | 0.01    | 0.03    | ... | 0.07    |

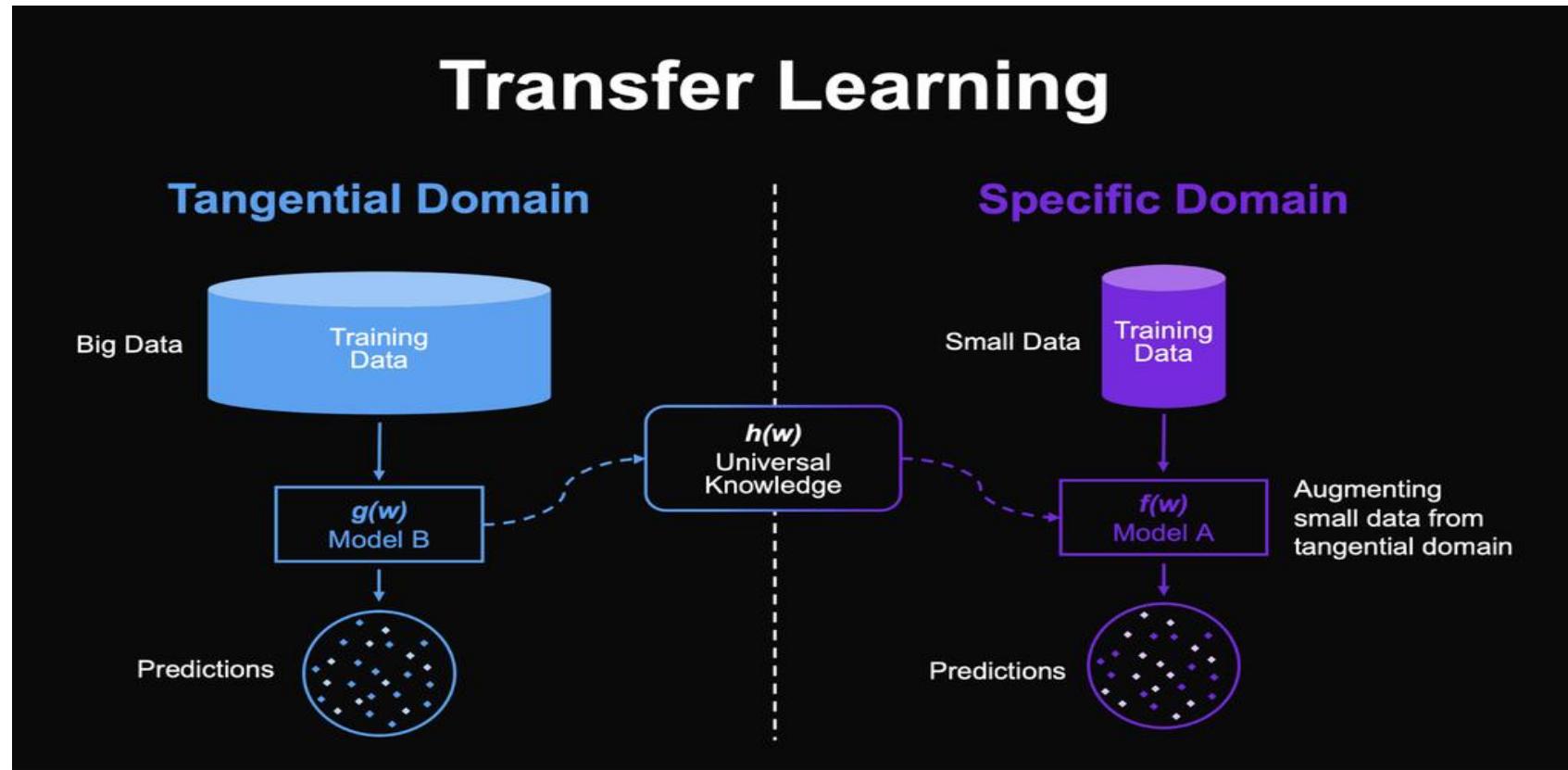
- 2003年，Google公布了3篇大数据奠良性论文，为大数据存储及分布式处理的核心问题提供了思路：**非结构化文件分布式存储（GFS）**、**分布式计算（MapReduce）** 及**结构化数据存储（BigTable）**，并奠定了现代大数据技术的理论基础。



- 2005 年，波士顿动力公司推出一款**动力平衡四足机器狗**，有较强的通用性，可适应较复杂的地形。
- 2006年，杰弗里·辛顿以及他的学生鲁斯兰·萨拉赫丁诺夫正式提出了**深度学习**的概念（**Deeping Learning**），开启了深度学习在学术界和工业界的浪潮。2006年也被称为**深度学习元年**，杰弗里·辛顿也因此被称为**深度学习之父**。



- 2010年, Sinno Jialin Pan和Qiang Yang发表文章《A Survey on Transfer Learning》——**迁移学习**

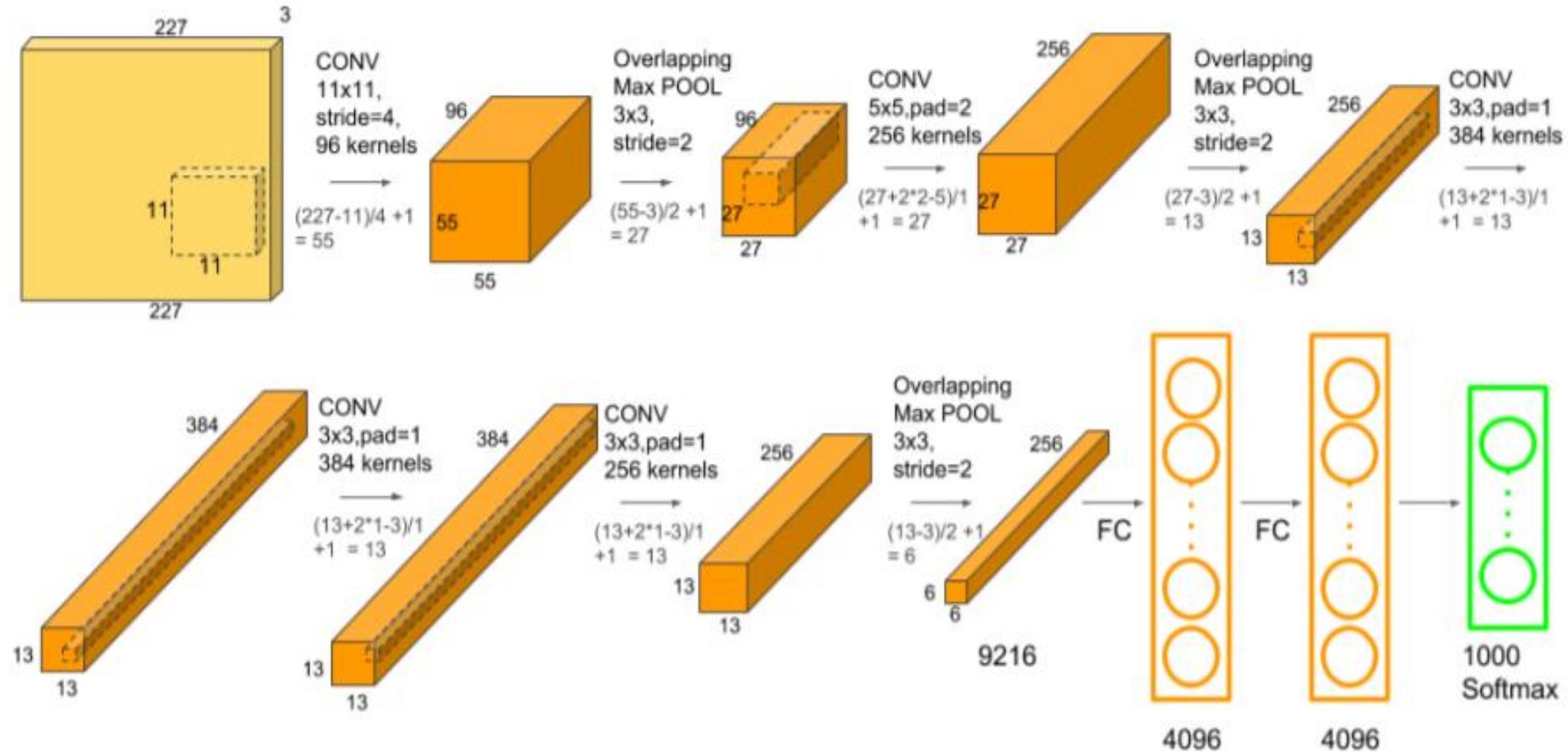


# 算法

## 5. 蓬勃发展期：2011年至今

- 2011年，IBM Watson问答机器人参与Jeopardy回答测验比赛最终赢得了冠军。Watson是一个集自然语言处理、知识表示、自动推理及机器学习等技术实现的**电脑问答（Q&A）系统**。
- 2012年，Hinton和他的学生Alex Krizhevsky设计的**AlexNet神经网络模型**在ImageNet竞赛大获全胜，这是史上第一次有模型在ImageNet 数据集表现如此出色，并引爆了神经网络的研究热情。

- AlexNet是一个经典的CNN模型，在数据、算法及算力层面均有较大改进，创新地应用了Data Augmentation、ReLU、Dropout和LRN等方法，并使用GPU加速网络训练。

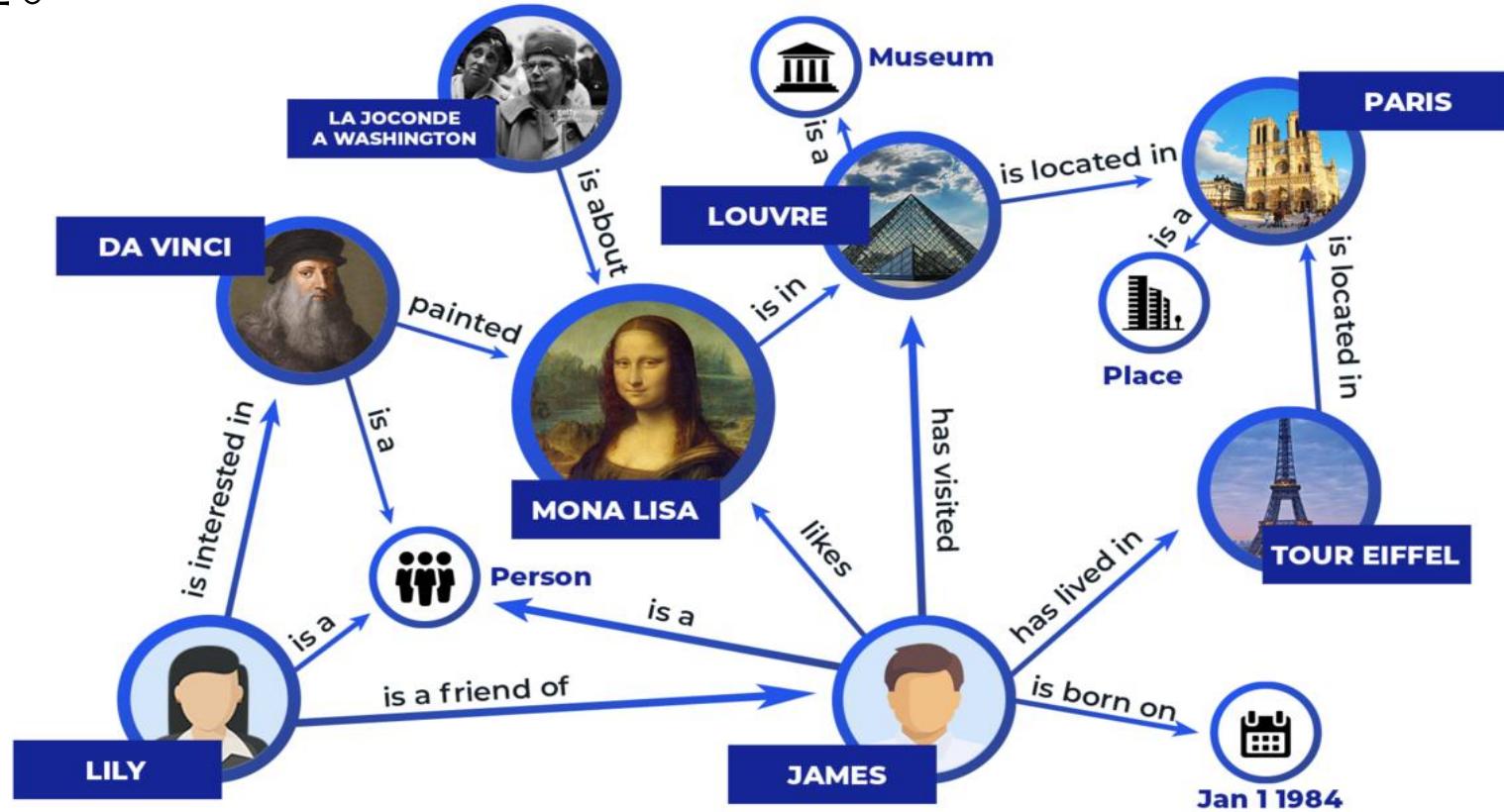


```
ImageNet_Root/  
├─ train/          # 训练集 (1,281,167张)  
|  ├─ n01440764/    # 类别目录 (WordNet ID命名)  
|  |  └─ n01440764_18.JPG  
|  |  ...  
|  |  ...          # 1000个类别  
├─ val/           # 验证集 (50,000张)  
|  └─ ILSVRC2012_val_00000001.JPG  
|  ...  
└─ test/          # 测试集 (100,000张, 无标签)  
└─ meta/  
  └─ ILSVRC2012_devkit_t12.tar.gz  # 开发工具包  
└─ data/  
  ├─ meta.mat      # MATLAB 格式的类别元数据  
  └─ labels.txt    # 标签映射文件
```

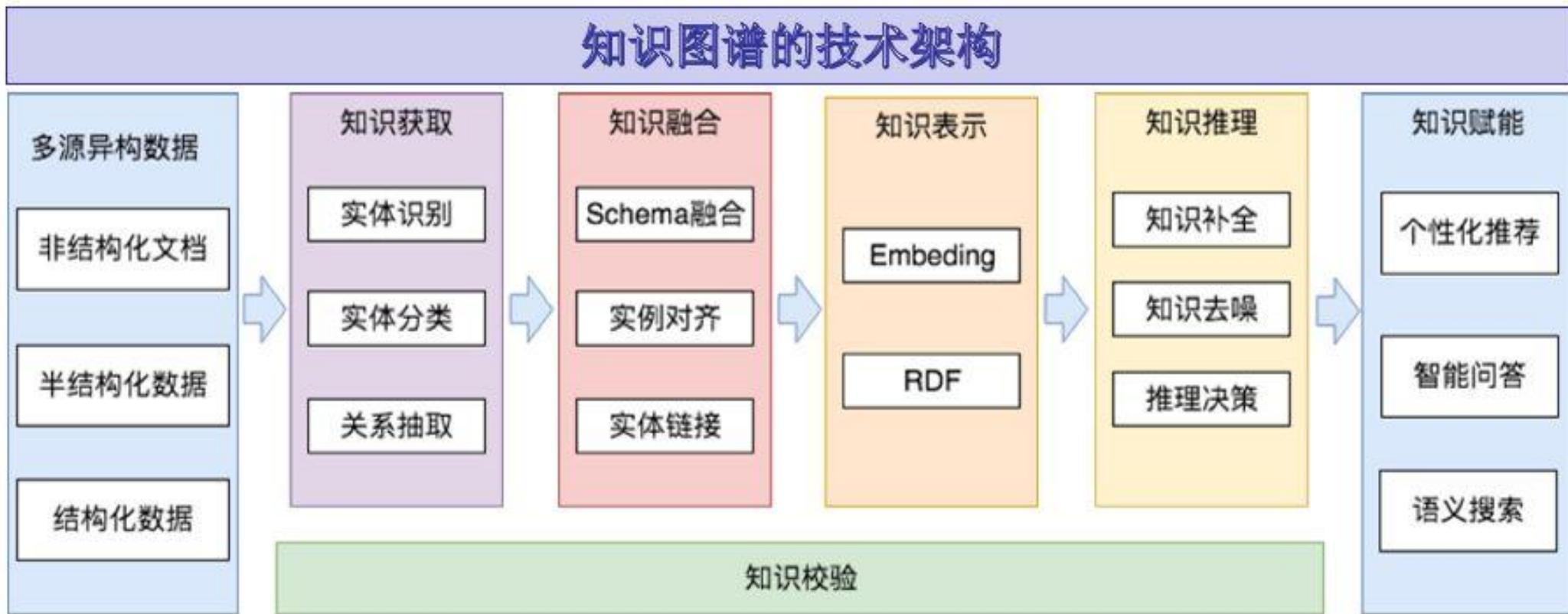
从2010年起，每年ImageNet 的项目组织都会举办一场基于ImageNet 的大规模视觉识别竞赛（ImageNet Large Scale VisualRecognition Challenge , ILSVRC ）。ILSVRC使用ImageNet的一个子集，总共有大约120万个训练图像，50,000个验证图像，以及150,000个测试图像；1000类别标记。

CIFAR-10是一个更接近普适物体的彩色图像数据集。CIFAR-10 是由Hinton 的学生Alex Krizhevsky 和Ilya Sutskever 整理的一个用于识别普适物体的小型数据集。一共包含10 个类别的RGB 彩色图片：飞机（airplane）、汽车（automobile）、鸟类（bird）、猫（cat）、鹿（deer）、狗（dog）、蛙类（frog）、马（horse）、船（ship）和卡车（truck）。

- 2012年，谷歌正式发布谷歌**知识图谱(Google Knowledge Graph)**，它是Google的一个从多种信息来源汇集的知识库，通过Knowledge Graph在普通的字串搜索上叠一层相互之间的关系，协助使用者更快找到所需资料的同时，也可使以知识为基础的搜索更近一步，以提高Google搜索的质量。

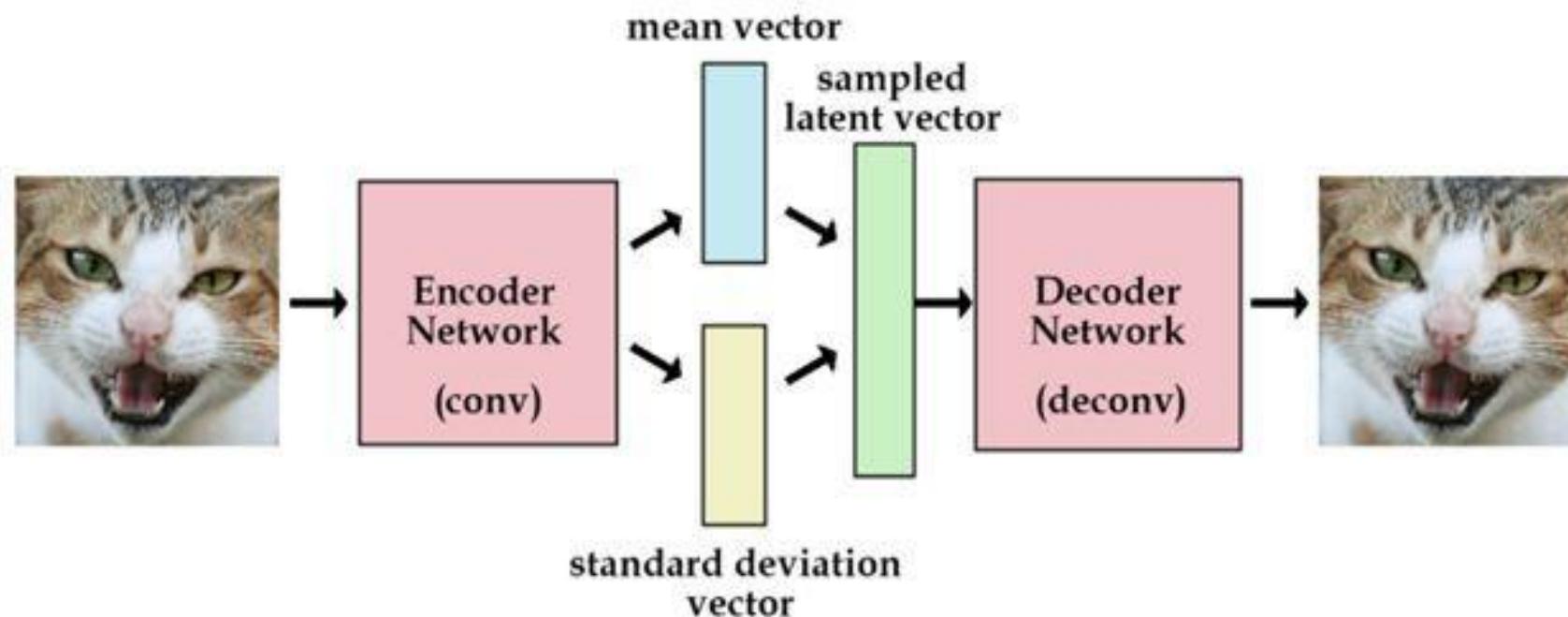


- 知识图谱是结构化的语义知识库，是符号主义思想的代表方法，用于以符号形式描述物理世界中的概念及其相互关系。其通用的组成单位是RDF三元组(实体-关系-实体)，实体间通过关系相互联结，构成网状的知识结构。

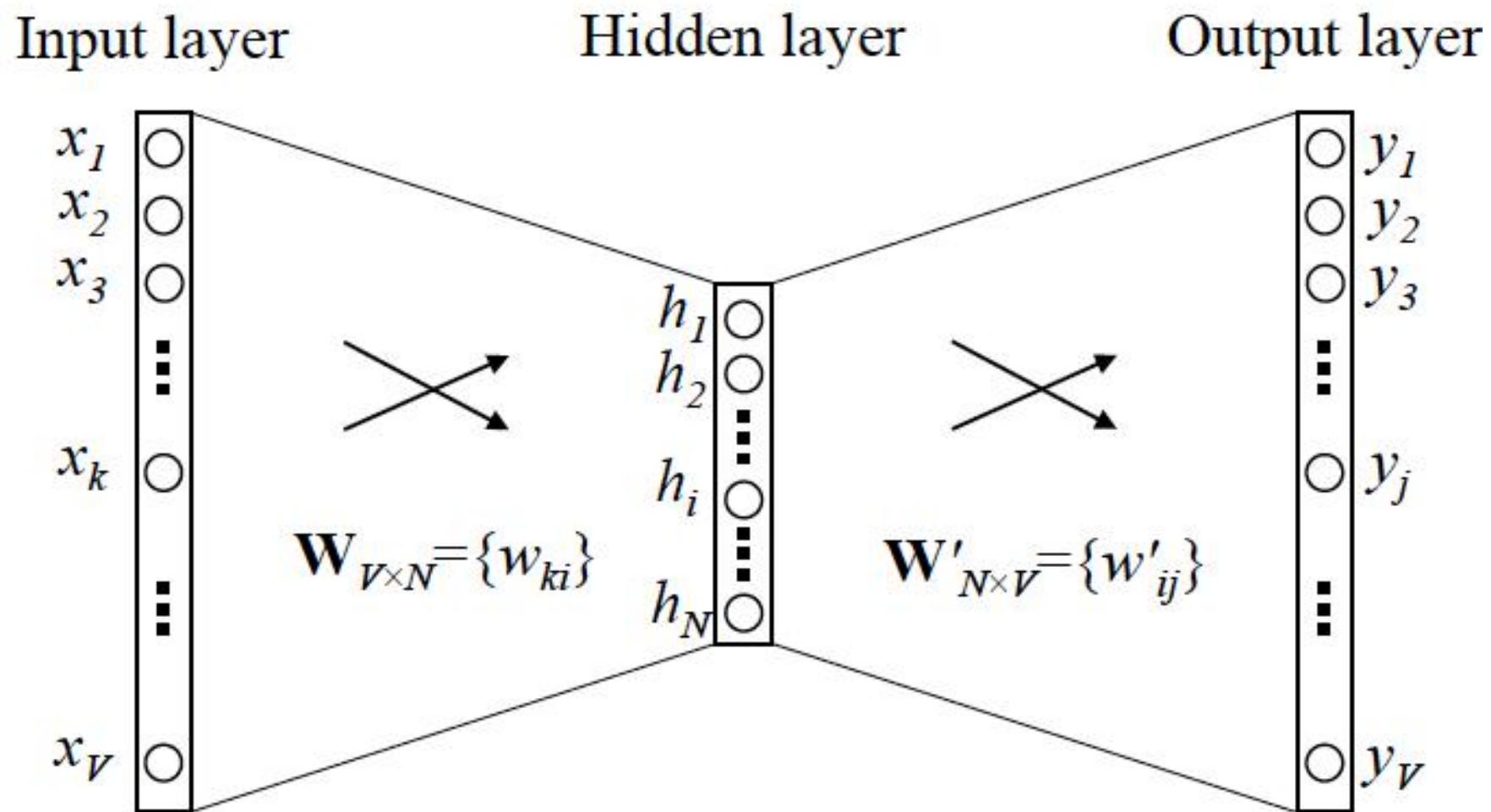


- 2013年，Durk Kingma和Max Welling在ICLR上以文章《Auto-Encoding Variational Bayes》提出**变分自编码器（Variational Auto-Encoder, VAE）**。

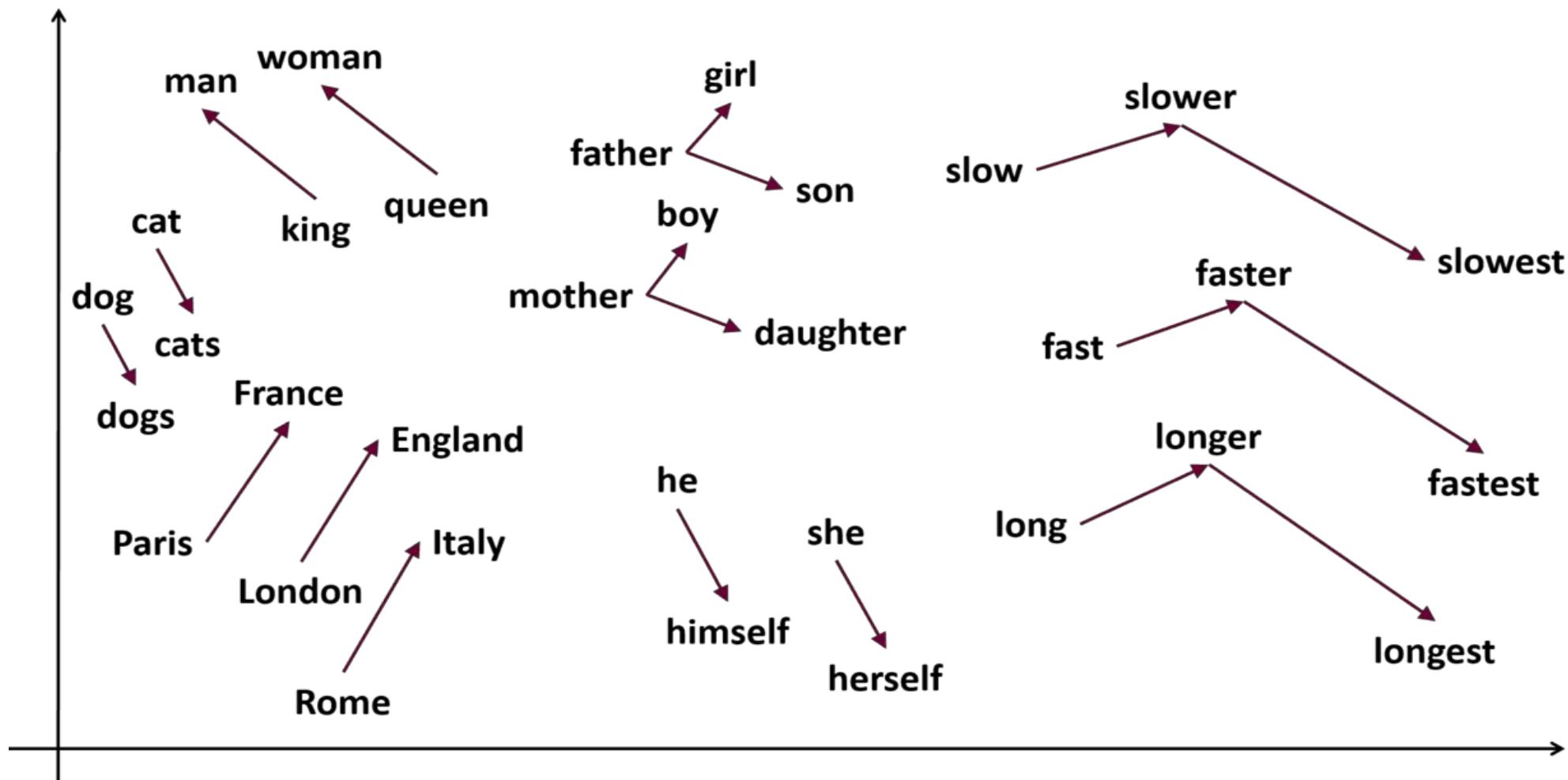
- VAE基本思路是将真实样本通过编码器网络变换成为一个理想的数据分布，然后把数据分布再传递给解码器网络，构造出生成样本，模型训练学习的过程是使生成样本与真实样本足够接近。



- 2013年，Google的Tomas Mikolov 在《Efficient Estimation of Word Representation in Vector Space》提出经典的 **Word2Vec模型**用来学习单词分布式表示，因其简单高效引起了工业界和学术界极大的关注。

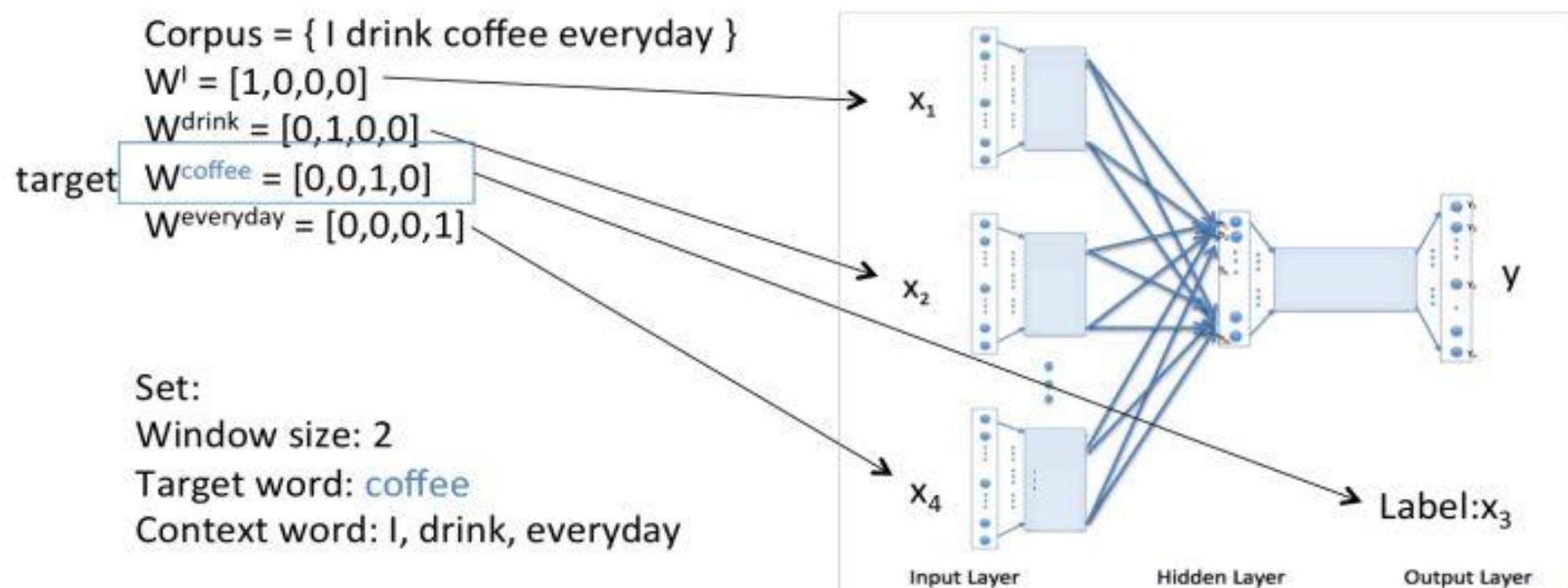


- Word2Vec基本的思想是学习每个单词与邻近词的关系，从而将单词表示成低维稠密向量。通过这样的分布式表示可以学到单词的语义信息，直观来看，语义相似的单词距离相近。

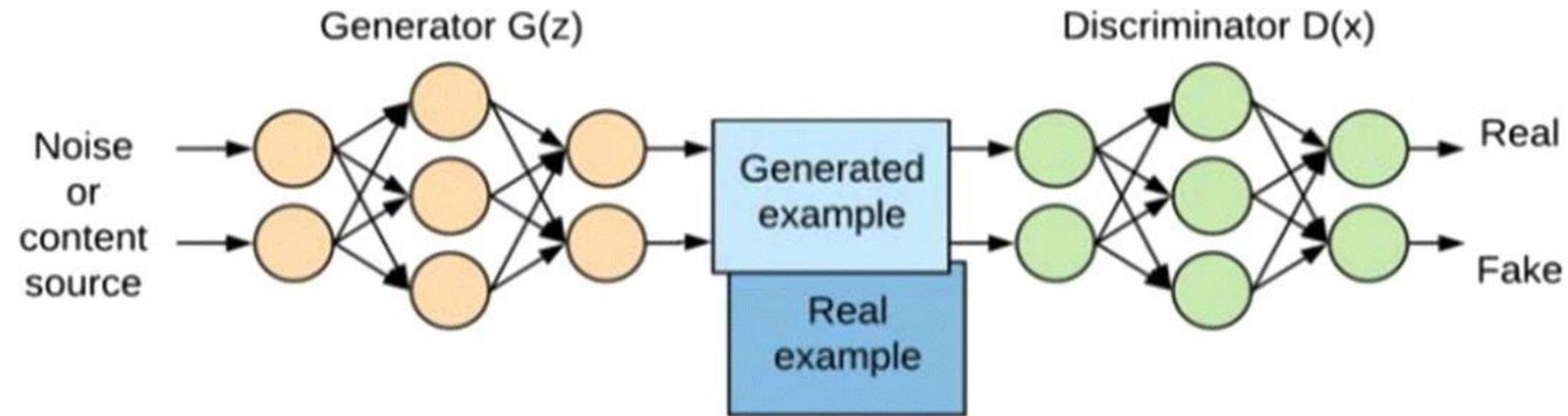


- Word2Vec网络结构是一个浅层神经网络（输入层-线性全连接隐藏层->输出层），按训练学习方式可分为**CBOW模型**(以一个词语作为输入，来预测它的邻近词)或**Skip-gram模型**(以一个词语的邻近词作为输入，来预测这个词语)。

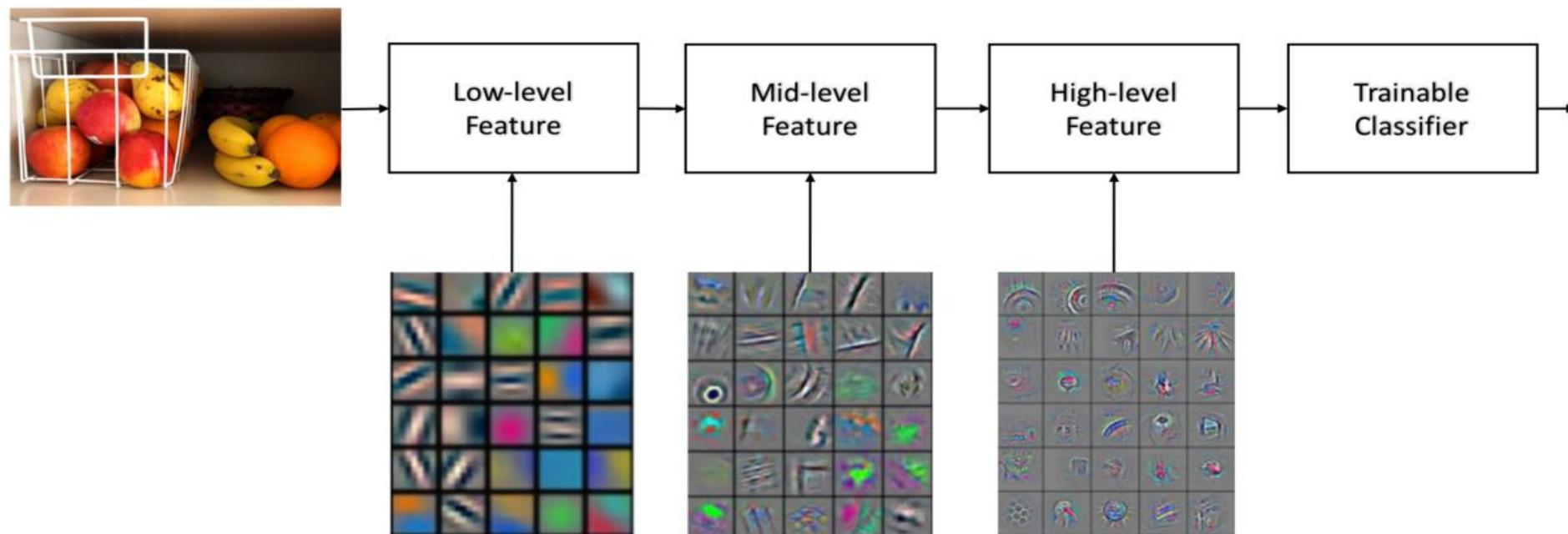
## An example of CBOW Model



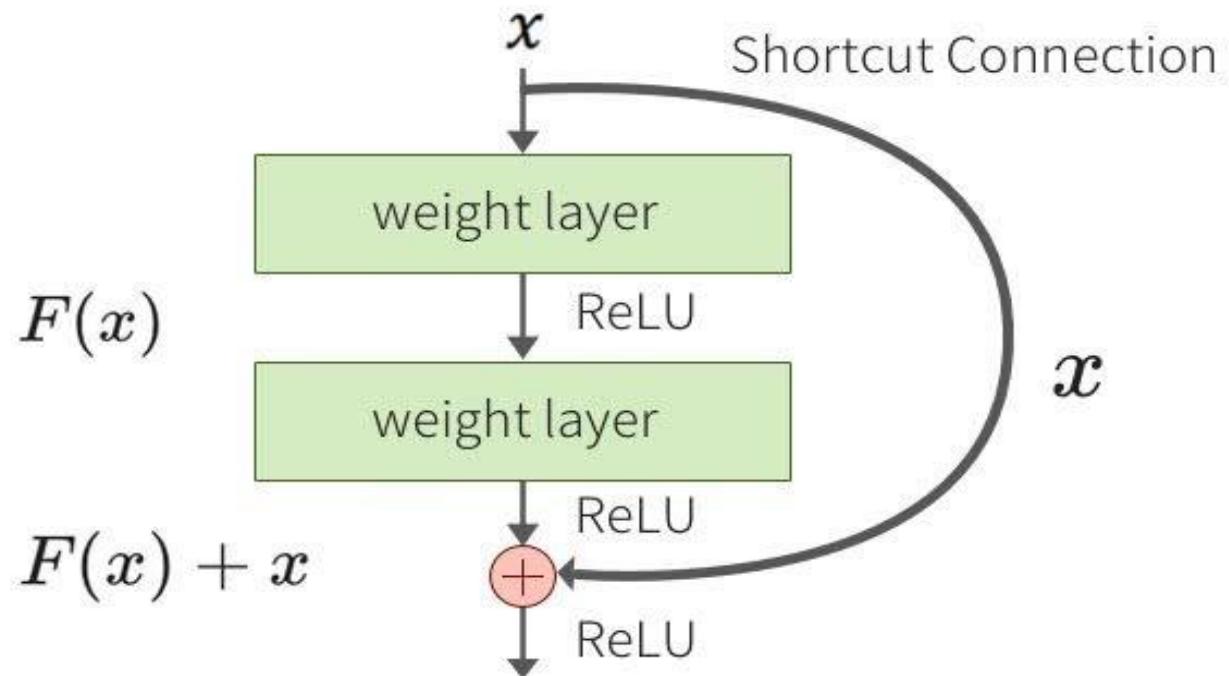
- 2014年，聊天程序“尤金·古斯特曼”（Eugene Goostman）在英国皇家学会举行的“2014图灵测试”大会上，首次“通过”了图灵测试。
- 2014年，Goodfellow及Bengio等人提出**生成对抗网络（Generative Adversarial Network, GAN）**，被誉为近年来最酷炫的神经网络。



- 2015年，为纪念人工智能概念提出60周年，深度学习三巨头**LeCun**、**Bengio**和**Hinton**(他们于2018年共同获得了图灵奖)推出了深度学习的联合综述《Deep learning》。
  - 《Deep learning》文中指出深度学习就是一种特征学习方法，把原始数据通过一些简单的但是非线性的模型转变成为更高层次及抽象的表达，能够强化输入数据的区分能力。通过足够多的转换的组合，非常复杂的函数也可以被学习。

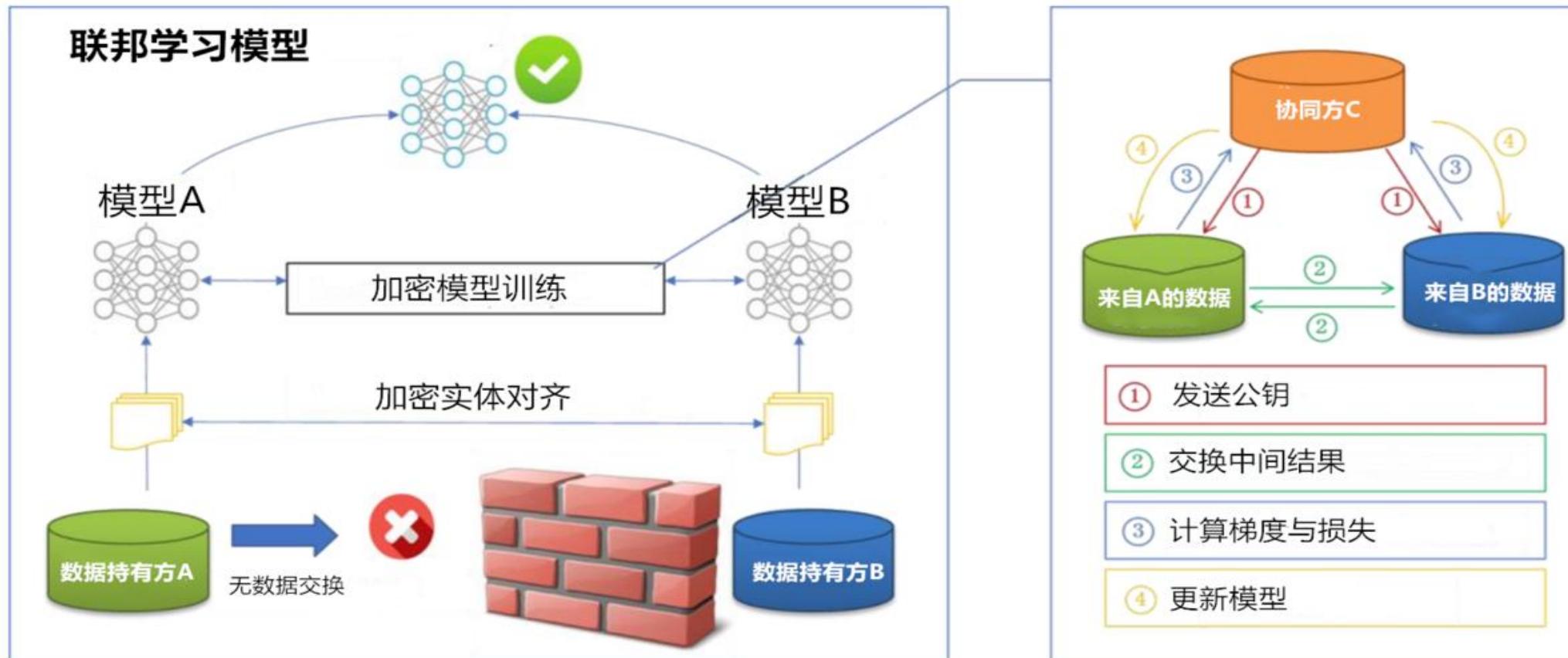


- 2015年，Microsoft Research的Kaiming He等人提出的**残差网络（ResNet）**在ImageNet大规模视觉识别竞赛中获得了图像分类和物体识别的优胜。
- 残差网络的主要贡献是发现了网络不恒等变换导致的“退化现象（Degradation）”，并针对退化现象引入了“**快捷连接（Shortcut connection）**”，缓解了在深度神经网络中增加深度带来的梯度消失问题。



- 2015年，谷歌开源**TensorFlow**框架。它是一个基于数据流编程 (dataflow programming) 的符号数学系统，被广泛应用于各类机器学习 (machine learning) 算法的编程实现，其前身是谷歌的神经网络算法库DistBelief。
- 2015年，马斯克等人共同创建**OpenAI**。它是一个非营利的研究组织，使命是确保通用人工智能 (即一种高度自主且在大多数具有经济价值的工作上超越人类的系统) 将为全人类带来福祉。其发布热门产品的如：**OpenAI Gym**, **GPT**等。
- 2016年，谷歌提出**联邦学习**方法，它在多个持有本地数据样本的分散式边缘设备或服务器上训练算法，而不交换其数据样本。

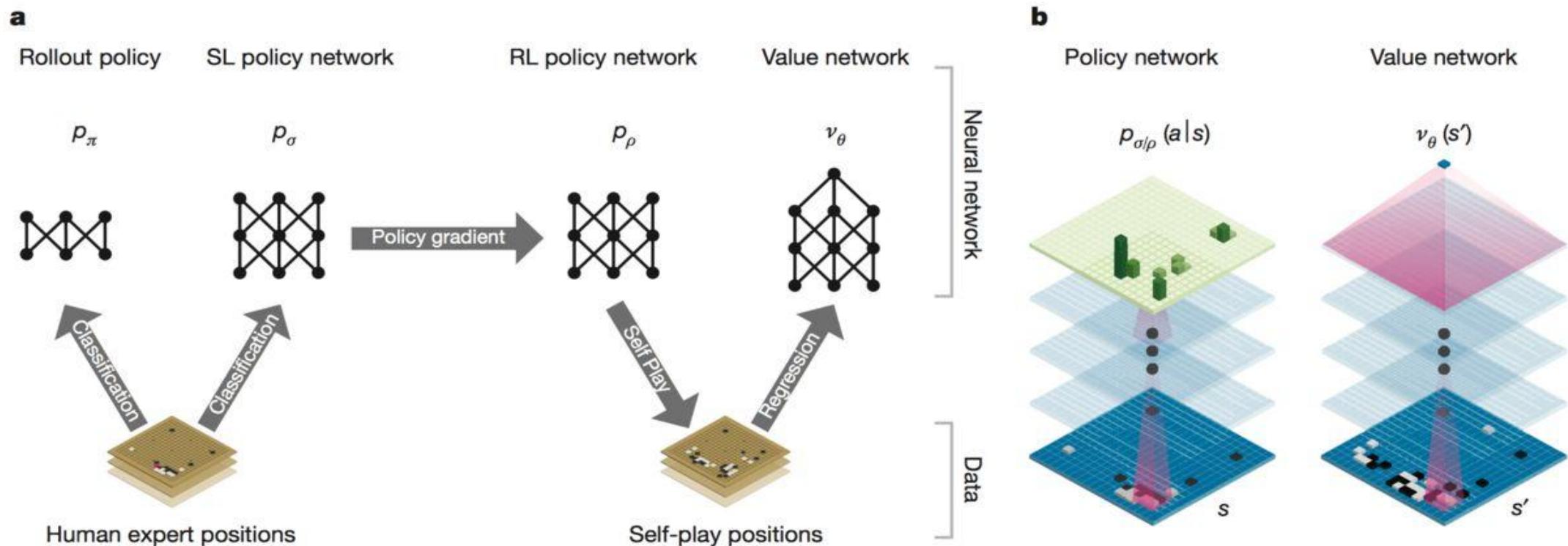
- 联邦学习保护隐私方面最重要的三大技术分别是：差分隐私（Differential Privacy）、同态加密（Homomorphic Encryption）和隐私保护集合交集（Private Set Intersection），能够使多个参与者在不共享数据的情况下建立一个共同的、强大的机器学习模型，从而解决数据隐私、数据安全、数据访问权限和异构数据的访问等关键问题。



- 2016年，AlphaGo（一款围棋人工智能程序，其主要工作原理是“深度学习”）与围棋世界冠军、职业九段棋手李世石进行围棋人机大战，以**4:1**的总比分获胜。



- AlphaGo是一款围棋人工智能程序，其主要工作原理是“深度学习”，由以下四个主要部分组成：策略网络（Policy Network）给定当前局面，预测并采样下一步的走棋；快速走子（Fast rollout）目标和策略网络一样，但在适当牺牲走棋质量的条件下，速度要比策略网络快1000倍；价值网络（Value Network）估算当前局面的胜率；蒙特卡洛树搜索（Monte Carlo Tree Search）树搜索估算每一种走法的胜率。



- 在2017年更新的AlphaGo Zero，在此前的版本的基础上，结合了强化学习进行了自我训练。它在下棋和游戏前完全不知道游戏规则，完全是通过自己的试验和摸索，洞悉棋局和游戏的规则，形成自己的决策。

随着自我博弈的增加，神经网络逐渐调整，提升下法胜率。更为厉害的是，随着训练的深入，AlphaGo Zero还独立发现了游戏规则，并走出了新策略，为围棋这项古老游戏带来了新的见解。

## AlphaGo Zero

### Reinforcement learning in AlphaGo Zero

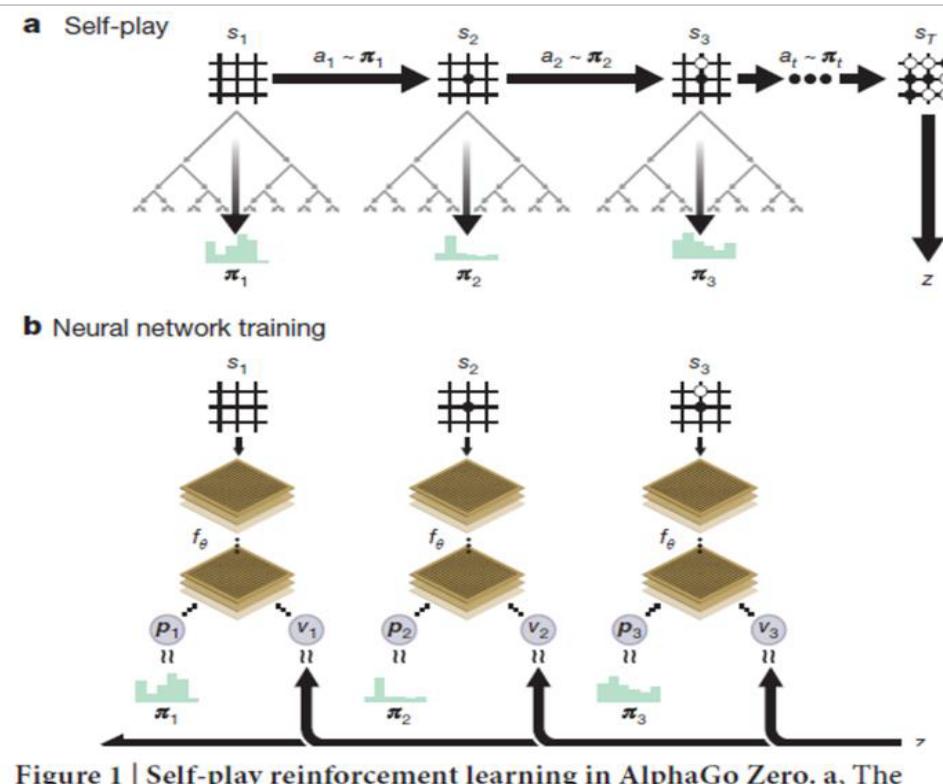
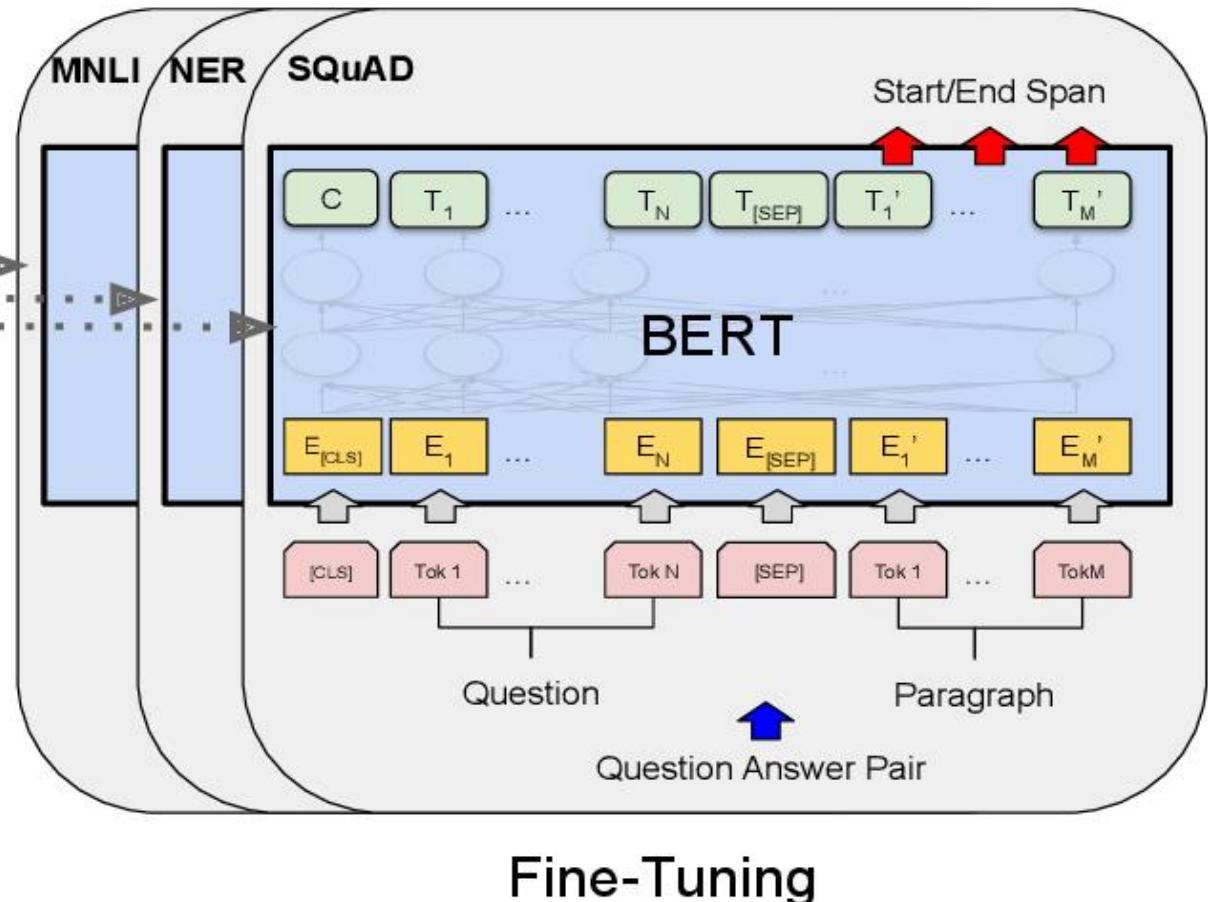
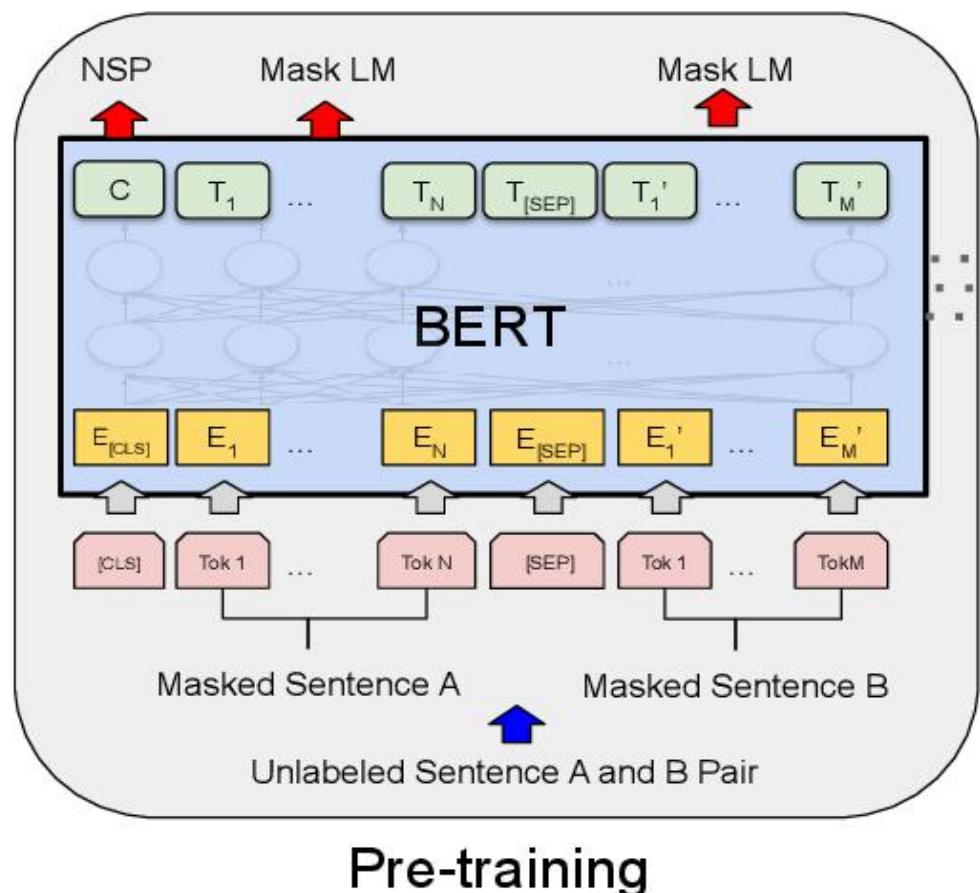


Figure 1 | Self-play reinforcement learning in AlphaGo Zero. a, The

- 2017年，中国香港的汉森机器人技术公司（Hanson Robotics）开发的**类人机器人索菲亚**，是历史上首个获得公民身份的一台机器人。索菲亚看起来就像人类女性，拥有橡胶皮肤，能够表现出超过62种自然的面部表情。其“大脑”中的算法能够理解语言、识别面部，并与人进行互动。



- 2018年，Google提出论文《Pre-training of Deep Bidirectional Transformers for Language Understanding》并发布**Bert(Bidirectional Encoder Representation from Transformers)模型**，成功在 11 项 NLP 任务中取得 state of the art 的结果。

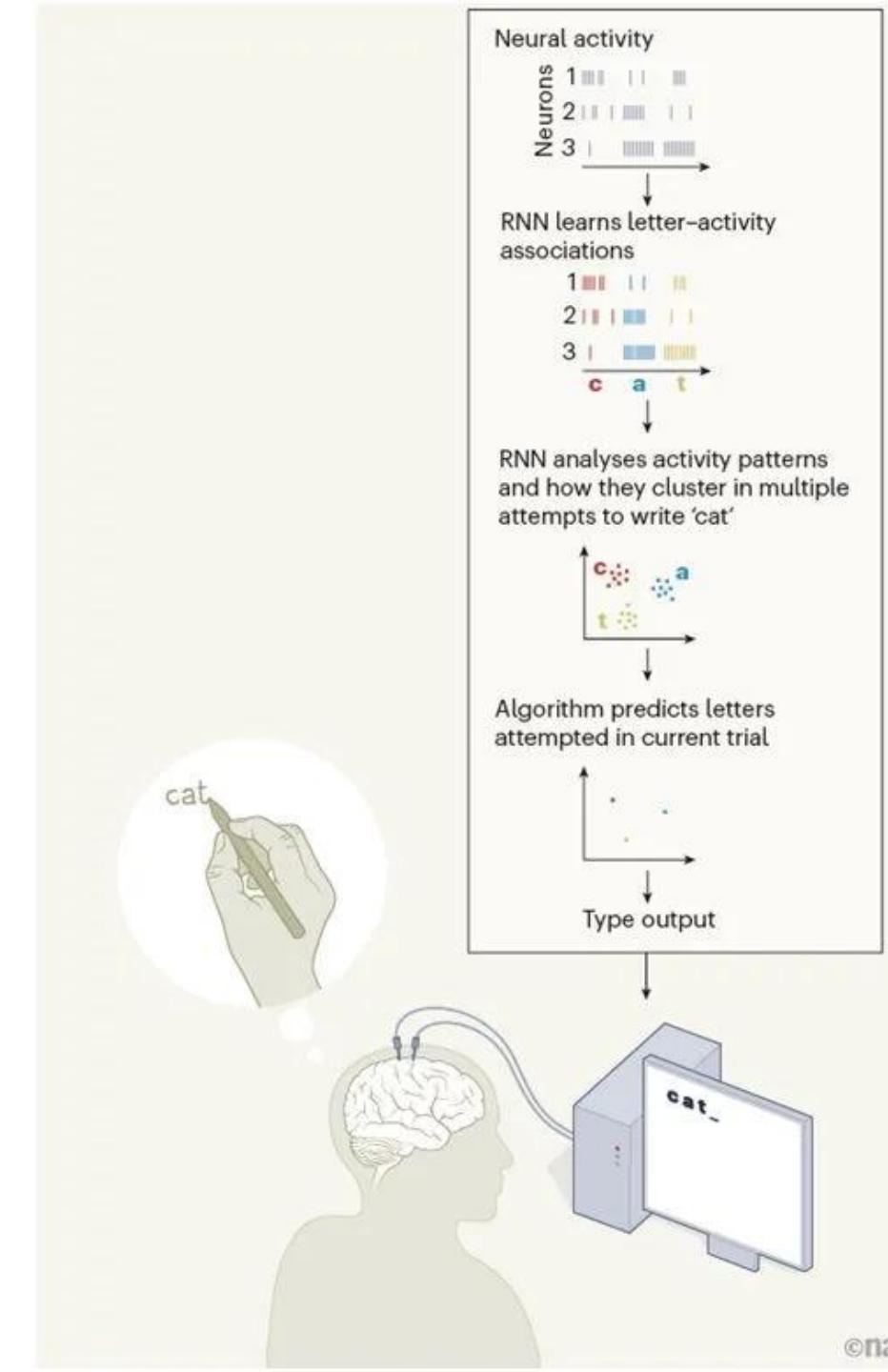


- 2019年，IBM宣布推出Q System One，它是世界上第一个专为科学和商业用途设计的**集成通用近似量子计算系统**。
- 2020年，Google与Facebook分别提出**SimCLR**与**MoCo**两个无监督学习算法，均能够在无标注数据上学习图像数据表征。两个算法背后的框架都是对比学习（contrastive learning），对比学习的核心训练信号是图片的“可区分性”。
- 2020年，OpenAI开发的文字生成(text generation)人工智能**GPT-3**，它具有1,750亿个参数的自然语言深度学习模型，比以前的版本GPT-2高100倍，该模型经过了将近0.5万亿个单词的预训练，可以在多个NLP任务（答题、翻译、写文章）基准上达到最先进的性能。

- 2020年，马斯克的脑机接口（brain–computer interface, BCI）公司**Neuralink**举行现场直播，展示了植入Neuralink设备的实验猪的脑部活动。
- 2020年，谷歌旗下DeepMind的AlphaFold2人工智能系统有力地解决了蛋白质结构预测的里程碑式问题。它在国际蛋白质结构预测竞赛（CASP）上击败了其余的参会选手，精确预测了蛋白质的三维结构，准确性可与冷冻电子显微镜（cryo-EM）、核磁共振或X射线晶体学等实验技术相媲美。
- 2020年，中国科学技术大学潘建伟等人成功构建76个光子的**量子计算原型机“九章”**，求解数学算法“高斯玻色取样”只需200秒，而目前世界最快的超级计算机要用6亿年。

- 2021年，OpenAI提出两个连接文本与图像的神经网络：**DALL-E** 和 **CLIP**。DALL-E 可以基于文本直接生成图像，CLIP 则能够完成图像与文本类别的匹配。
- 2021年，德国Eleuther人工智能公司于今年3月下旬推出开源的文本AI模型**GPT-Neo**。对比GPT-3的差异在于它是**开源免费**的。

- 2021年，美国斯坦福大学的研究人员开发出一种用于打字的脑机接口(brain-computer interface, BCI)，这套系统可以从运动皮层的神经活动中解码瘫痪患者想象中的手写动作，并利用递归神经网络(RNN)解码方法将这些手写动作实时转换为文本。相关研究结果发表在2021年5月13日的**Nature**期刊上，论文标题为“**High-performance brain-to-text communication via handwriting**”。



# Transformer架构的奠基（2017年）

- 2017年，Google提出了基于自注意力机制的神经网络结构——Transformer架构。这一架构彻底重塑了自然语言处理领域，为后续的预训练大模型奠定了坚实的基础。Transformer遵循编码器-解码器的整体架构，通过自注意力机制捕捉输入序列中不同位置之间的全局依赖关系，实现了对序列数据的高效处理。

Vaswani等人在**2017**年通过其开创性论文“**Attention is All You Need**”引入了Transformer架构。

# 预训练大模型的崛起（2018年-2020年）

- 2018年，BERT和GPT等大模型崭露头角，显著提升了上下文理解和文本生成能力。这些模型通过大规模的数据预训练，学习到了丰富的语言知识和模式，为后续的自然语言处理任务提供了强有力的支持。
- 2020年，OpenAI推出了拥有1750亿参数的GPT-3模型，展示了卓越的“少样本”和“零样本”学习能力。GPT-3的出现标志着大模型在参数规模和性能上取得了巨大的突破。

# 幻觉问题的挑战与应对（2021年-2022年）

- 随着大模型的广泛应用，幻觉问题逐渐凸显。即生成内容与事实不符，甚至出现“一本正经地胡说八道”的现象。为了应对这一挑战，OpenAI在2022年开发了“对话式”的ChatGPT模型，采用了监督微调和基于人类反馈的强化学习技术，显著提高了生成内容的准确性和可信度。

# 多模态与推理能力的突破（2023年-2024年）

- 2023年，GPT-4等多模态模型的发布，整合了文本、图像和音频处理能力，使大模型能够以更接近人类的“听”、“说”、“看”能力进行交互。这一突破极大地扩展了大模型的应用场景和范围。
- 2024年，随着技术的不断进步，大模型在复杂问题解决和深度推理能力上取得了显著进展。OpenAI-o1和DeepSeek-R1等推理模型的推出，标志着人工智能在模拟人类思维模式上迈出了重要一步。

# DeepSeek-R1的开创性贡献（2025年）

- 2025年初，中国推出了具有开创性且高性价比的大型语言模型 DeepSeek-R1。该模型采用大规模专家混合（MoE）架构，参数规模高达6710亿，是许多流行开源大模型的10倍。DeepSeek-R1在数学计算、代码生成、自然语言推理等关键领域表现出色，性能已比肩OpenAI的GPT-01正式版。此外，DeepSeek-R1还以其“超成本效益”和“开源”设计挑战了AI领域的传统规范，推动了先进大模型的普及，并促进了各行业的创新。

# DeepSeek-R1的开创性贡献（2025年）

- 英伟达、微软、亚马逊等科技巨头纷纷接入DeepSeek-R1模型，进一步丰富了其AI服务生态，加速了AI技术的落地应用。这一事件不仅彰显了DeepSeek-R1模型的强大实力，也预示着AI领域将迎来新一轮的变革与发展。
- 综上所述，从Transformer架构的奠基到DeepSeek-R1模型的开创性贡献，大模型的发展历程充满了挑战与突破。未来，随着技术的不断进步和应用的深入拓展，我们有理由期待AI将为我们的生活和社会带来更多的惊喜和改变。
- DeepSeek-V3 (2024-12)、DeepSeek-R1-Zero 和 DeepSeek-R1 (2025-01)

# GPT-4V：视觉遇见语言

- 2023年，OpenAI推出了GPT-4V，将GPT-4的语言能力与先进的计算机视觉相结合。它可以解释图像、生成标题、回答视觉问题，并推断视觉中的上下文关系。其跨模态注意力机制允许文本和图像数据的无缝集成，使其在医疗保健（如分析医学图像）和教育（如互动学习工具）等领域具有价值。

**Prompt:**  
Describir la imagen en 20 idiomas diferentes



**GPT-4V:**

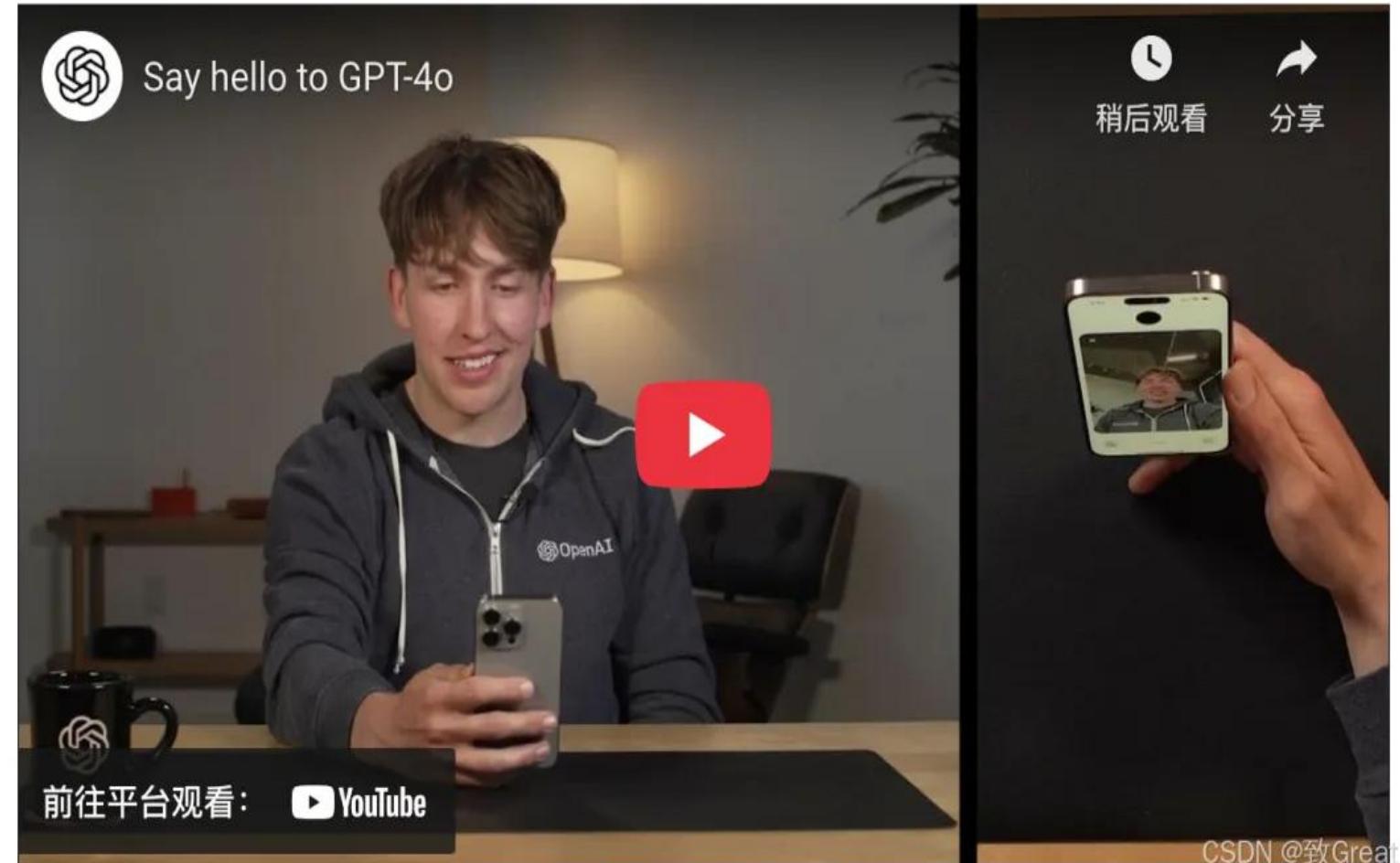
1. Inglés: This is a beautiful view of Machu Picchu in Peru.
2. Español: Esta es una hermosa vista de Machu Picchu en Perú.
3. Francés: C'est une belle vue de Machu Picchu au Pérou.

**GPT-4V output continued**

4. Alemán: Dies ist eine schöne Aussicht auf Machu Picchu in Peru.
5. Italiano: Questa è una bella vista di Machu Picchu in Perù.
6. Portugués: Esta é uma bela vista de Machu Picchu no Peru.
7. Russo: Это красивый вид на Мачу-Пикчу в Перу.
8. Chino: 这是秘鲁马丘比丘的美丽景色。
9. Japonés: これはペルーのマチュピチュの美しい景色です。
10. Coreano: 이것은 페루의 마추 피추의 아름다운 경치입니다.
11. Hindi: यह पेरू के माचू पिच्चु का सुंदर हस्य है।
12. Árabe: هذا منظر جميل لماشو بيتشو في بيرو.
13. Turco: Bu, Peru'daki Machu Picchu'nun güzel bir manzarasıdır.
14. Griego: Αυτή είναι μια όμορφη θέα του Μάτσου Πίτσου στο Περού.
15. Hebreo: זהו נוף יפה של מאצ'ו פיצ'ו בפרו.
16. Sueco: Det här är en vacker vy av Machu Picchu i Peru.
17. Noruego: Dette er en vakker utsikt over Machu Picchu i Peru.
18. Danés: Dette er en smuk udsigt over Machu Picchu i Peru.
19. Holandés: Dit is een prachtig uitzicht op Machu Picchu in Peru.
20. Polaco: To piękny widok na Machu Picchu w Peru.

# GPT-4o：全模态前沿

- 到2024年初，GPT-4o通过整合音频和视频输入进一步推进了多模态。它在一个统一的表示空间中运行，可以转录音频、描述视频或将文本合成音频。实时交互和增强的创造力——如生成多媒体内容——使其成为娱乐和设计等行业的多功能工具。



# 算力 (Computer Power)

- **基础算力（基于CPU芯片）、智能算力（基于GPU和NPU芯片）、超算算力**
  - 智能算力的增长规模远超通用算力，比重将超过70%
  - 工信部数据显示，截至2021年底，我国在用数据中心机架总规模达520万标准机架，在用数据中心服务器规模1900万台，算力总规模超过140 EFLOPS。全国在用超大型和大型数据中心超过450个，智算中心超过20个。

# 算力

- **云计算：**21世纪初，CPU、内存、硬盘、GPU（显卡）这些资源被整合调用，以软件形式呈现，成为一个可以无限扩张的虚拟算力池。
  - 亚马逊、阿里巴巴、腾讯
- **算力网络与东数西算**
  - 2022年2月，“东数西算”工程正式全面启动
  - 2022年5月，西南地区最大的人工智能计算中心——成都智算中心正式上线
  - 2022年8月，阿里云宣布正式启动张北超级智算中心
  - 2022年8月，内蒙古枢纽节点建成了中国最大的自动驾驶智算中心——“扶摇”
  - 2022年9月，“百度智能云-昆仑芯（盐城）智算中心”正式揭牌上线

# AI赋能各行各业

## 1. AI+传媒=最先受益的应用领域

- **AI+文字:** ChatGPT (OPENAI) 、文心一言 (百度) 、通义千问 (阿里巴巴) 、盘古 (华为)
- **AI+图片:** SAM分图大模型 (META脸书) 、萤火虫Firefly (Adobe推出的生成式AI) 、AI绘图工具DALL-E (OPENAI) 、“AI简笔画”(美图公司)
- **AI+视频:** 万兴播爆 (万兴科技)

## 2. AI+计算机=“程序员”: ChatGPT (OPENAI)

# 传感器与智能传感器

- **传感器定义**（国家标准GB/T7665-2005）：能感受被测量并按照一定规律转换成可用输出信号的器件或装置，通常由敏感元件和转换元件组成。
- **智能传感器**：集感知、信息处理与通信于一体；能提供以数字量方式传播具有一定知识级别的信息，是传感器集成化与微处理器相结合的产物。可定义为基于人工智能理论，利用微处理器实现智能处理功能的传感器。



火灾传感器



温湿度智能传感器



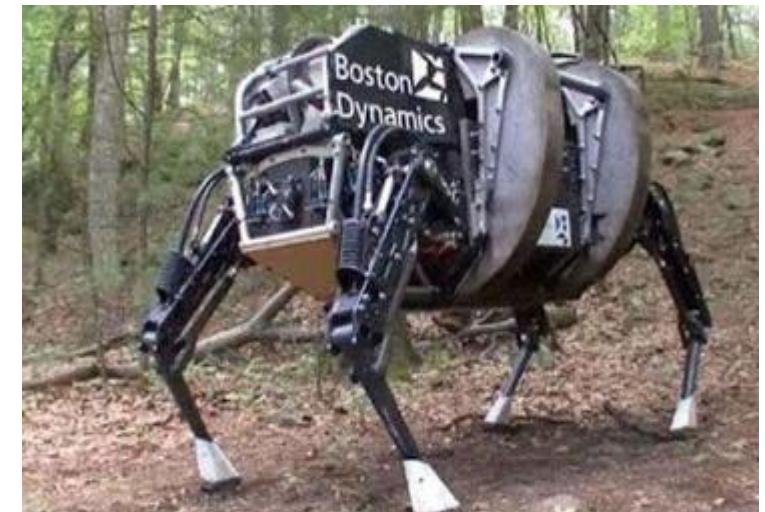
# 传感器与智能传感器

- 智能传感器的主要功能：

1. 自补偿功能
2. 自计算和处理功能
3. 自学习与自适应功能
4. 自诊断功能
5. 其它：用于数据交换通信接口功能、数字和模拟输出功能、使用备用电源的断电保护功能等。



Google无人驾驶汽车



“大狗”机器人（Big Dog）

内部安装的各种传感器不仅可让它根据环境的变化调整行进姿态，还能够保障操作人员实时地跟踪“大狗”的位置并监测其系统状况。

# 关键环节

## 预测、感知、决策、控制

- **预测：**将历史信息和当前感知数据作为输入，通过模型计算预估未来状态的过程
- **感知：**利用各种传感器感知周围环境并产生相关反馈的过程
- **决策：**基于感知数据和先验知识实现路径规划、避障决策、异常处理决策等
- **控制：**利用计算机、传感器和执行器等设备，实现实行进过程中的运动控制、姿态控制、目标追踪、避碰等

# 环境感知的传感系统

主要由机器视觉识别系统、雷达系统、超声波传感器和红外线传感器所组成。

## 1. 机器视觉识别系统

- 利用CCD等成像元件全方位拍摄车外环境，根据搜集到的信息得到反映真实场景的图像**数据**，综合运用各种道路检测和目标识别等**算法**，提取出车道线、道路边界以及车辆的方位信息，判断汽车是否有驶出车行道的危险，当情况十分危险时，会通过报警系统给驾驶员发出提示和警报
- 同时，图像测控系统还可以根据视觉导航的输出，对车辆的执行机构发出指令，从而自主决定车辆当前的前进方向和控制车辆自身的运动状态。

## 2. 雷达系统

- 利用电磁波探测目标的距离、速度、方位等。
- 不要复杂的设计与繁复的计算；
- 不受光线、天气等因素的干扰；
- 雷达系统是靠电磁波反射原理工作的，这会导致相近的不同雷达电磁波之间的相互干扰而影响工作效能；
- 提供远距离的车辆和障碍物信息，用在汽车ACC系统、防碰撞系统以及驾驶支援系统中。



车载线扫描激光雷达检测前方障碍物



车载三维激光雷达环境感知

### 3. 超声波传感器

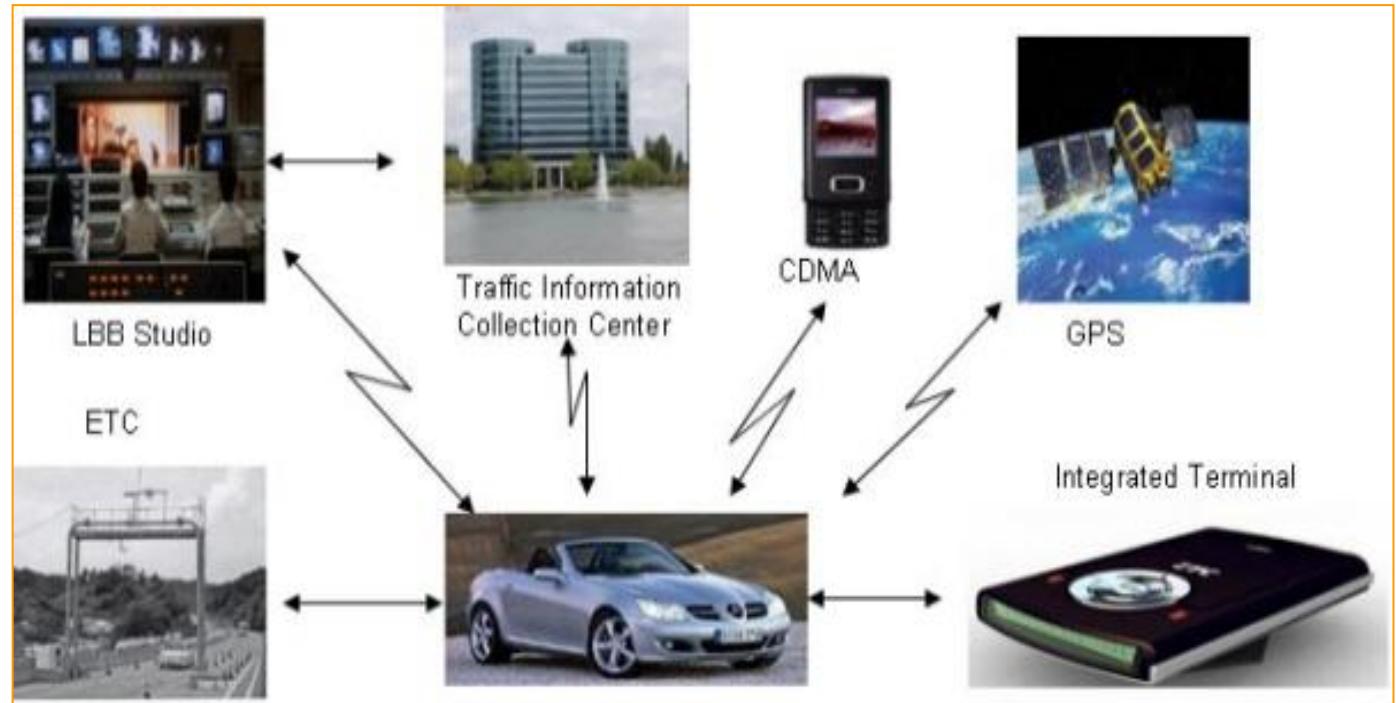
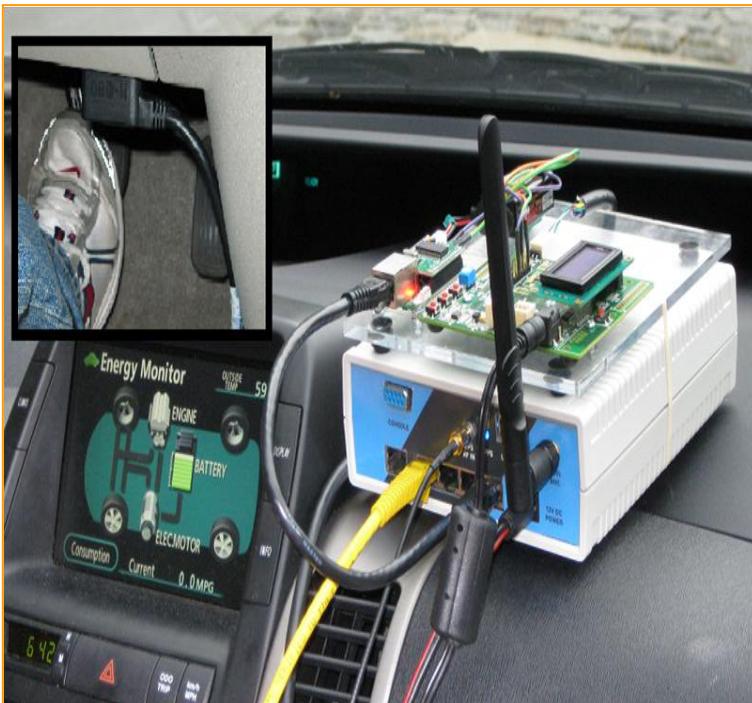
利用超声波为检测方法的传感器。一般检测距离大约1m到5m，但检测不出来详细的位置信息，使用超声波探测得来的数据处理简单、快速。主要用于近距离障碍物的检测，比如在倒车防撞系统中。

### 4. 红外线传感器

利用红外线来进行测量工作，不受黑暗、风、沙、雨、雪、雾的阻挡，环境适应性好，且功耗低。红外线传感器可以增强机器视觉识别的可靠性，使黑夜如同白昼，因此常被用于智能汽车中的夜视系统中。

## 5. 信息交换系统

基于无线、网络等近、远程通讯技术让车辆之间信息交换，从而获取车辆行驶周边环境信息。



## 6. 多传感器信息融合

为了克服传感器的数据可靠性低、有效探测范围小等局限性，保证在任何时刻都能获取完全可靠的环境信息，通常使用多个传感器进行数据采集，利用**传感器信息融合技术**对检测到的数据进行分析、综合、平衡，根据各个传感器互补特性进行容错处理，扩大系统的时频覆盖范围，增加信息维数，避免单个传感器的工作盲区，从而得到所需要的环境信息。



# 多模态融合（Multimodal Fusion）

- 结合来自**不同模态（如视觉、听觉、文本等）的数据**，以提升信息处理和理解能力的技术方法。**多模态数据**通常具有**不同的物理性质和信息特征**，通过融合这些多模态信息，可以获得更全面和准确的理解。这种融合过程可以发生在：
  - 数据层融合**：直接对不同模态的数据进行融合。
  - 特征层融合**：提取不同模态的数据特征后进行融合。
  - 决策层融合**：对不同模态的处理结果进行融合。

- **数据层融合的基本方法和步骤：**
  - **数据预处理：**对不同模态的数据进行预处理，包括数据清洗、归一化和对齐等。
  - **数据拼接：**将不同模态的数据在输入层直接拼接或组合在一起，形成统一的输入表示。
  - **模型训练：**使用融合后的数据训练一个单一模型，进行下游任务的执行。

- 特征层融合的基本方法和步骤：
  - 特征提取：使用专门的模型从不同模态的数据中提取高维特征表示。
  - 特征对齐：对提取的特征进行对齐，以确保在融合时能够正确结合。
  - 特征融合：将对齐后的特征进行融合，形成综合特征表示。
  - 模型训练：使用融合后的特征训练一个下游任务模型。

- **决策层融合的基本方法和步骤：**

- 独立特征提取：分别从每个模态中提取特征。
- 独立模型训练：对每个模态分别训练独立的模型。
- 决策结果融合：将每个模态的独立决策结果进行融合，形成最终的决策。

# 数据集

- **KITTI数据集：**德国卡尔斯鲁厄理工学院（KIT）和丰田美国技术研究院(TTIC)联合创办，是目前国际上最大的自动驾驶场景下的计算机视觉算法评测数据集，用于评测立体图像(stereo)，光流(optical flow)，视觉测距(visual odometry)，3D物体检测(object detection)和3D跟踪(tracking)等计算机视觉技术在车载环境下的性能。
- KITTI包含市区、乡村和高速公路等场景采集的真实图像数据，每张图像中最多达15辆车和30个行人，还有各种程度的遮挡与截断，整个数据集由389对立体图像和光流图，39.2 km视觉测距序列以及超过200k的3D标注物体图像组成，以10Hz的频率采样及同步。

# 智能无人系统

**军事：地区冲突、反恐战争、**

**战场利器：**智能无人系统

- **陆地无人系统：**侦察无人车、运输无人车、作战无人车、破障无人车、排爆无人车、无人车编队与指挥系统
- **空中无人系统：**侦察无人机、作战无人机、后勤运输无人机及无人机编队
- **海洋无人系统：**侦察无人艇、作战无人艇、后勤运输无人艇、巡逻搜救无人艇、侦察无人潜航器、作战无人潜航器以及岸基支持系统

# 智能无人系统

- 复杂环境下**自主感知与理解**技术
  - 多模态传感融合自主感知、复杂场景目标识别与理解
- 行为决策与轨迹规划技术
  - 多源异构环境下的行为决策、动/静环境下的轨迹规划、复杂场景下的轨迹跟踪
- 自主**导航定位**技术
  - 基于惯性/卫星深度信息融合的导航定位、基于惯性/天文信息融合的导航定位、基于视觉跟踪的导航、地球物理辅助导航技术
- 多场景**自主学习与智能控制**技术
  - 任务的自主技能学习技术、自主作业交互控制技术、类人智能控制的无人系统运动控制技术
- 无人集群**协同控制**技术
  - 智能无人系统集群局部规则控制技术、智能无人系统集群软控制技术、智能无人系统集群领航控制技术、智能无人系统人工势场控制技术
- 自然**人机交互**技术
  - 人机交互技术、增强现实与混合现实技术、脑机接口技术