



# 人工智能基础与进阶

## 大数据与人工智能

上海交通大学

# 目录 content



- 第一节      当代人工智能的基石：大数据**
- 第二节      大数据时代下的人工智能**
- 第三节      深度学习成功的启示**
- 第四节      大数据人工智能的伦理**

# 第一节 当代人工智能的基石：大数 据



# 当代人工智能的基石：大数据

信息交换、信息存储、信息处理

三个方面能力的大幅增长产生的  
了大数据

信息交换

通信和网络带宽的  
大幅增长

信息处理

整理、交换、分析数  
据的能力大幅增长

信息存储

计算机存储量的  
大幅增长



上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

# 当代人工智能的基石：大数据—信息交换

从1986年到2007年，信息数量增长了约**217倍**。

每个参与信息交换的节点都可以在**短时间内接收并存储大量数据**。

2018年微信每日发送信息450亿次，新浪微博日活跃用户2亿，微博视频/直播日均发布量为150万+；

2019年天猫双11最高时刻，秒钟订单额超10亿；

2019年京东“618”开场1小时下单金额50亿元；中国3万家综合性医院，每年新增数据量可达20Zbit。

# 当代人工智能的基石：大数据—信息存储



全球信息存储能力大约**每3年翻一番**

从1986年到2007年， 全球信息存储能力**增加了约120倍**

1986年时， 只能存储全世界所交换信息的大约1%

2007年这个数字已经增长到大约16%  
信息存储能力的增加为我们利用大数据提供了近乎无限的想象空间。

# 当代人工智能的基石：大数据—信息处理



谷歌、Facebook、亚马逊、百度、阿里等公司在数据量逐渐增大的同时，也相应建立了灵活、强大的**分布式数据处理集群**

数万台乃至数十万台计算机构成的**并行计算集群**每时每刻都在对累积的数据进行进一步加工和分析。

## 第二节

# 大数据时代下的人工智能



# 大数据时代下的人工智能

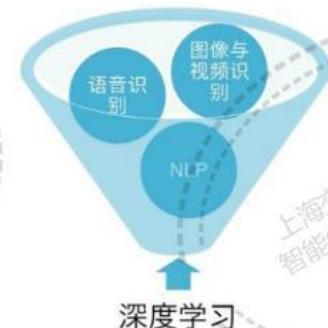
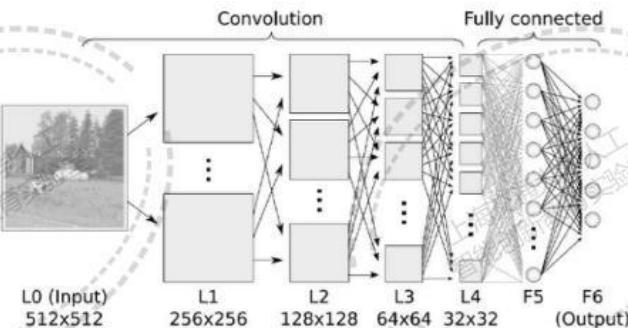
- 在有大数据之前，计算机并不擅长解决需要人类智能的问题
- 今天这些问题换个思路就可以解决了，其核心就是**变智能问题为数据问题**。
- 数据好比人类的新土壤，正是依托这片土壤，智能型的文明才得以滋生繁衍，土壤越广袤，其孕育的新文明才更有生机和活力
- 大量数据的使用，最大的意义在于它能让计算机完成一些过去只有人类才能做到的事情，这最终将带来一场**智能革命**。



# 大数据时代下的人工智能—大数据与深度学习

深度学习可以**从大数据中挖掘出以往难以想象的有价值的数据、知识或规律。**

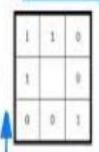
深度学习非常擅长从输入到输出的复杂非线性变换，实现对原始数据的表示学习或非线性建模。



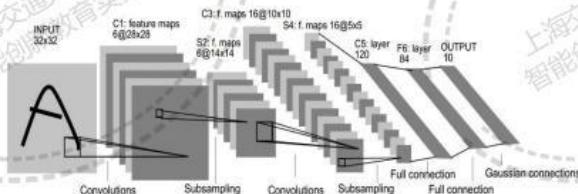
# 大数据时代下的人工智能—手工特征到自动特征



LBP



- 学者们还基于经验或专家知识设计了众多**人工特征**
- 在语音信号处理领域大量采用的LPCC、MFCC等倒谱系数，
- 在图像处理领域广泛使用SIFT、HOG、Gabor、LBP等特征。
- 以经验或专家知识驱动的模型**无法借力大数据**，不能从大数据中学习出蕴含于其中的丰富知识和规律。



# 大数据时代下的人工智能—神经网络复兴



Yoshua Bengio



Geoffrey Hinton



Yann LeCun

深度学习三巨头，2019年图灵奖得主

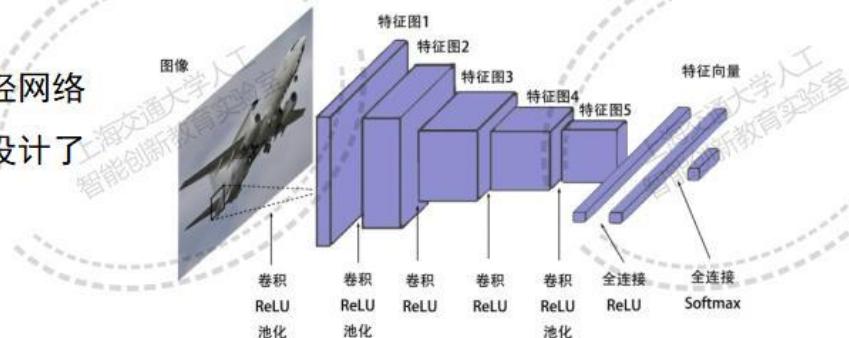
- 2006年，Hinton发文强调多隐层深度神经网络具有更优异的特征学习能力
- Bengio等人在国际会议NIPS2006也强调了分层训练深度网络的做法。
- 这些工作重新开启了**多层神经网络研究热潮**

# 大数据时代下的人工智能—神经网络复兴



- AlexNet就是在LeNet-5上的扩展与改进
- 夺得2012年ILSVRC比赛的冠军，并远超第一名

- 1989年LeCun等提出了卷积神经网络
- 1998年LeCun等在此基础之上设计了LeNet-5网络



# 大数据时代下的人工智能—应用



- 计算机可以记住全国所有通缉犯的长相。
- 大数据和深度学习一起，可以完成以前也许需要数万名人类警察才能完成的任务。

## 智能交通：大数据预测交通路况

Alpha Go不但从人类专业选手以往的数百万份棋谱中学习，还可以从自己和自己的对弈棋谱中学习。



# 大数据时代下的人工智能—应用

- 金融行业有大量客户的交易数据
- 更好地对客户进行风险防控，或针对特定客户进行精准营销
- 城市交通管理部门拥有大量交通监控数据
- 智能交通流量预测、智能交通疏导等



## 智能交通：大数据预测交通路况

将高速历史数据、实时数据与路网状况结合，基于阿里云大数据计算能力，预测未来1小时内的路况。结果显示，预测准确率稳定在91%以上。



常发拥堵分析  
节假日分析预测  
事件影响分析

### 第三节

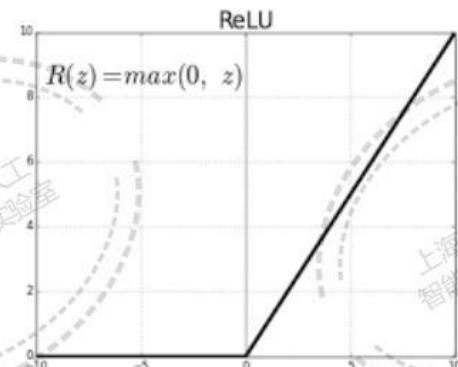
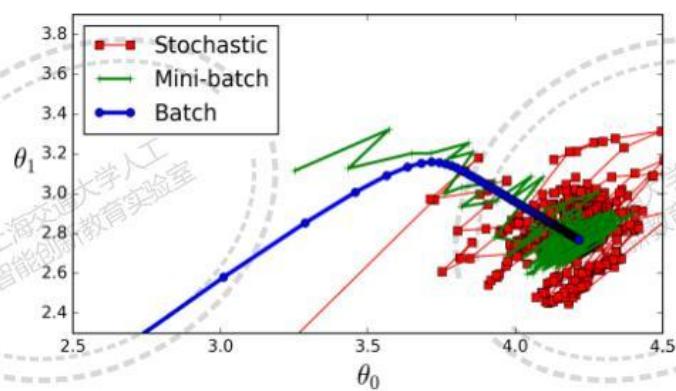
## 深度学习成功的启示



# 深度学习成功的启示—优化方法的变革

2006年Hinton等人开创了无监督、分层训练多层网络的先河

Mini-Batch SGD、ReLU激活函数、Batch Normalization、处理梯度消失问题的手段，对深度卷积网络不断加深、性能不断提升功不可没。



# 深度学习成功的启示—从手工特征到自动特征

- 人工特征严重影响了智能技术的有效性和通用性。
- 深度学习彻底颠覆了这种“人工特征”的思路，开启了数据驱动的表示学习的自动特征。
- 数据驱动的表示从而大大提高了人工智能算法的通用性。



# 深度学习成功的启示—从“分步治之”到“端到端”

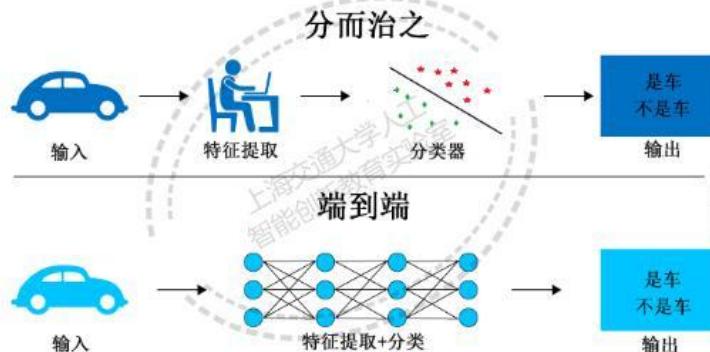
分治或分步的方法

分解问题所得到的子问题变得更简单，更容易解决

子问题的最优解未必是全局的最优

端到端

具有协同增效的优势



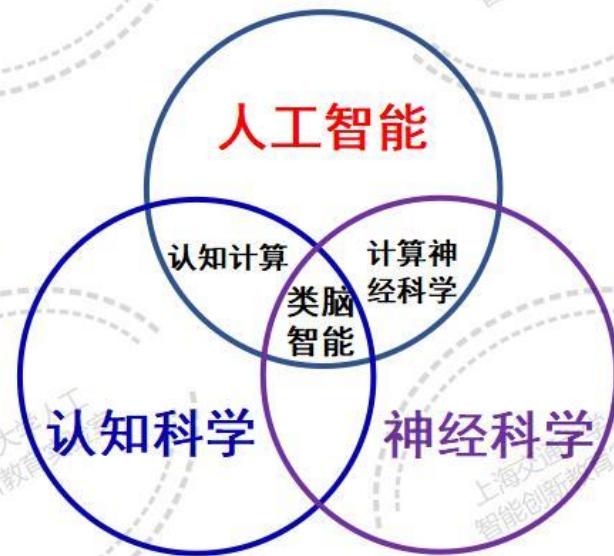
# 深度学习成功的启示—脑神经科学启发

上海交通大学人工  
智能创新教育实验室

认知机模型——模拟感受野逐渐变大、逐层提取由简及繁特征、语义逐级抽象的视觉神经通路。

脑神经科学的进步可以为深度模型发展提供更多的可能性。

近些年越来越多的神经科学研究表明，曾一度被认为功能极为特异化的神经细胞其实具有良好的可塑性。



上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

## 第四节 大数据人工智能的伦理



上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

# 大数据人工智能的伦理—安全与隐私

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

## 自动驾驶

交通运行的效率可以达到最优，交通事故率理论上可以趋近于零。

人为的交通事故→交通系统的**信息安全事故**。

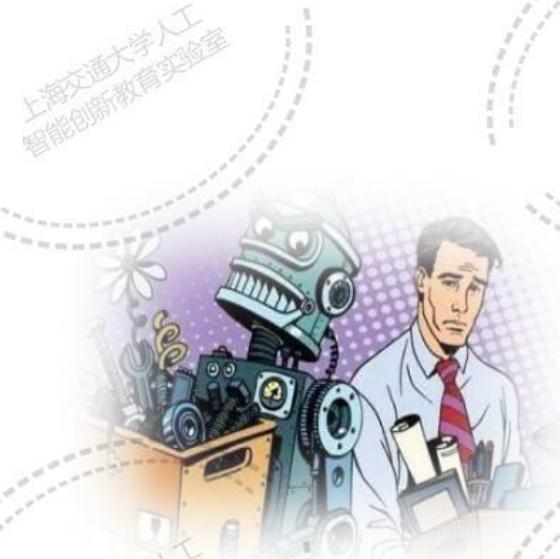
### 全事故。

- 黑客侵入智能汽车终端
- 破坏自动驾驶系统的信息采集和传输途径。

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室

上海交通大学人工  
智能创新教育实验室



# 大数据人工智能的伦理—安全与隐私



人工智能依赖于海量用户数据，这些都会导致用户**个人信息泄露**  
获得**生活便利**和保护**个人隐私**，目前已  
经成为一个**两难选择**

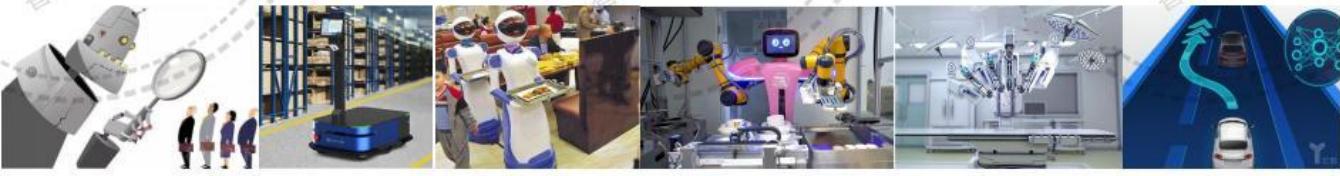
# 大数据人工智能的伦理—失业问题

在越来越进步的科技之下，许多以往借助于人力的劳动都被机器所取代。

涉及生活的方方面面，从餐厅服务、库房物品搬运、教育、医学诊断、新闻撰写到法律行业。

从短期看，我们也许很难避免某些行业、某些地区出现局部的失业现象，

从长远来看，这绝不是一种灾难性事件，而是人类的一次重新调整。



# 大数据人工智能的伦理—歧视

数据自身携带我们社会存在的**不公**；

算法可能会进一步**加大歧视**，造成恶性循环；

算法的**设计和学习过程**，也可能带来**偏见甚至歧视**；

不同出身的人所获取到的**信息差异**，会加剧而不是减少**阶级的固化**。

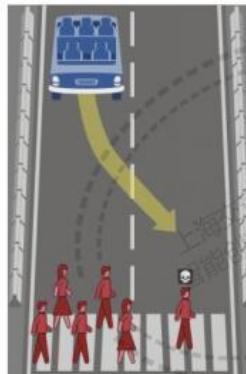
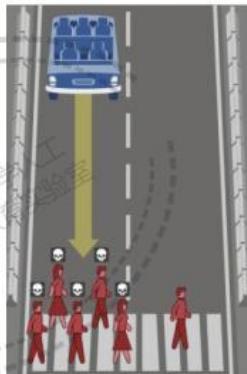


# 大数据人工智能的伦理—道德机器

机器逐步从被动工具向能动者转变

具有高度的甚至完全的**自主性**

人们需要对**智能机器**这一能动者提出类似的法律、伦理等道义要求，并且具有相应的外在约束和制裁机制。





谢谢聆听

THANKS FOR YOUR ATTENTION