大数据分析的
深度神经网络方法

章毅
四川大学计算机学院
2016.03.25 重 庆

提纲

- 大数据简介
- 大脑新皮层的神经网络结构
- 大脑新皮层神经网络的记忆
- 大数据分析GPU深度神经网络计算平台

大数据历史

1997 “大数据”一词诞生

2000 大数据商业应用诞生

2004 深度神经网络在语音大数据、图像大数据领域接连取得巨大突破

2006 大数据分布式存储与并行计算技术诞生

2011 美国奥巴马政府发布了大数据计划，将大数据战略上升为美国国家意志

2012 G. E. Hinton教授在《Science》发表文章，深度神经网络兴起

2013 Google Brain计划，激起大规模神经网络在大数据分析中的应用

2014 全世界兴起大数据热潮

2015 国务院发布大数据发展行动纲要，大数据正式上升为我国国家意志

我国大数据产业兴起

基于深度神经网络的大数据分析方法在学术界和工业界均取得巨大成功

工信部白皮书指出：大数据分析是大数据研究的重要环节，其中大数据分析的神经网络方法具有良好前景

大数据概念

大数据的概念

体量浩大
Volume

多源异构
Variety

大数据的
基本特点

生成快速
Velocity

价值稀疏
Value

大数据的目标

实现大数据转换为价值

问题：怎样实现大数据的目标？

大数据关键技术

大数据关键技术

展示平台

- 大数据知识展示
- 大数据产品

分析平台

- 大数据分析方法
- 高性能计算平台

数据平台

- 大数据采集，标记
- 大数据存储，管理

大数据分析是大数据转换为价值的最重要的环节，否则，大数据仅仅是一堆数据而已。

大数据分析是大数据转化为价值的桥梁

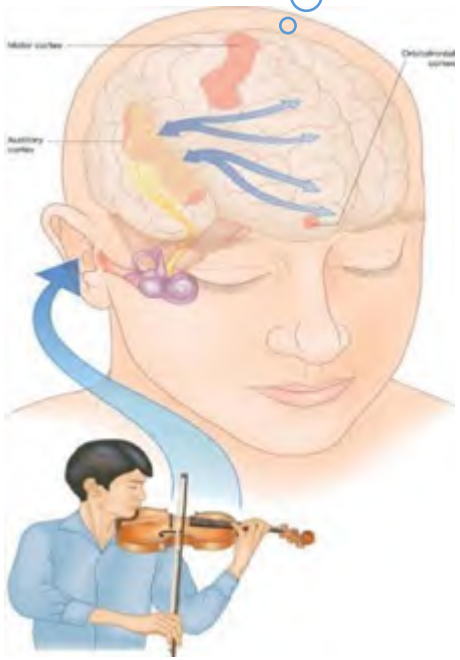
大数据分析

价值

问题：怎样设计大数据分析方法？

大数据关键技术

进入大脑的信息被编码为某种数据，进而由大脑神经网络处理



人类大脑是天然的大数据处理器！

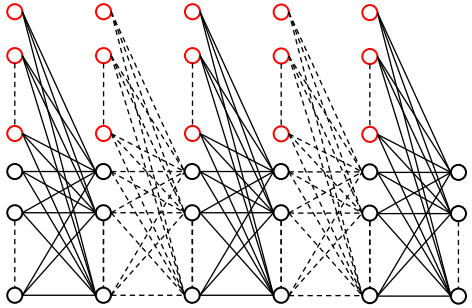
- 每秒信息传递和交换1000亿次，PB级数据
- 同步处理声音、温度、气味、图像等数据
- 50亿本书的存储容量
- 每秒人眼数据量140.34GB
- 在识别、判断、预测等智能行为方面展现出十分强大的能力
- 优秀的大数据处理器

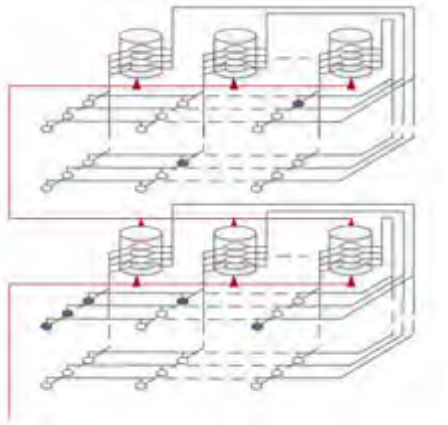
大数据关键技术

大数据 + 深度神经网络 = 巨大成功

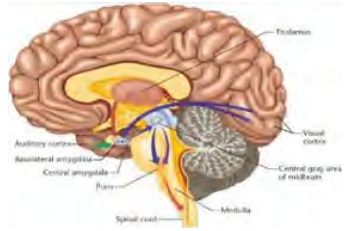
大数据分析
最成功的方法

深度神经网络方法





神经网络是一种模拟
大脑神经网络的计算方法

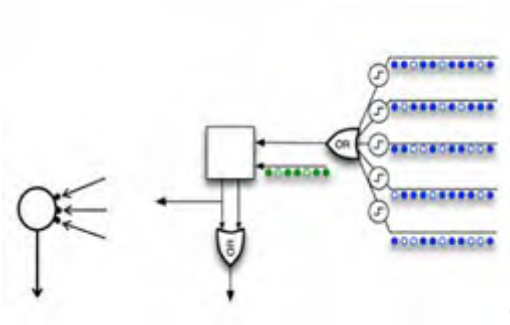


大数据分析

价值

大数据分析是大数据转化为价值的桥梁

深度神经网络历史



Jeff Hawkins

2004年
HTM
神经网络；
智能理论
新框架

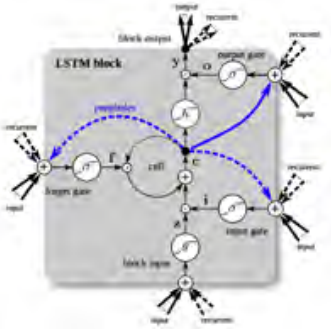
1960s.
第一代神
经网络
感知机

1985年
第二代神
经网络
BP

2006年
深度学习；
前馈深度
神经网络

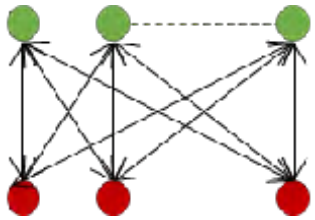
Geoffrey Hinton

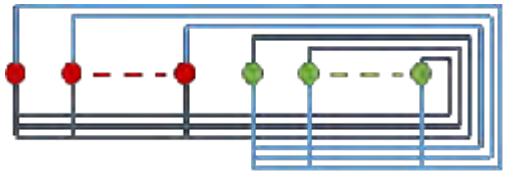
2009年
Recurrent
神经网络
异军突起



Jürgen Schmidhuber

至今
深度神经
网络在众
多领域均
取得重大
成功





IBM

Google

facebook

Microsoft

深度神经网络方法

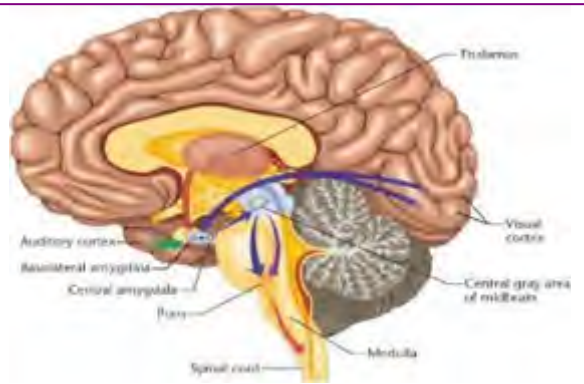
深度神经网络方法

神经网络结构⊕学习算法

大脑新皮层
神经网络结构

大脑新皮层
神经网络的记忆机制

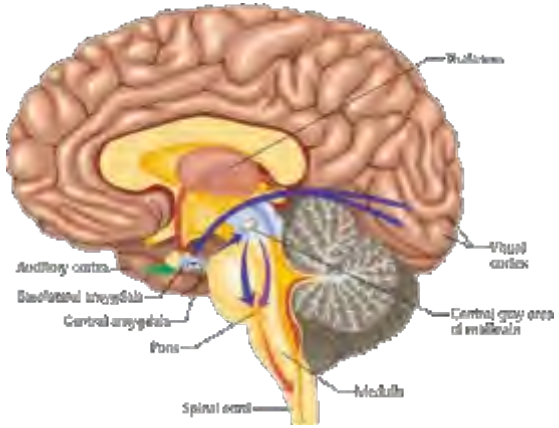
模拟大脑神经计算的方法

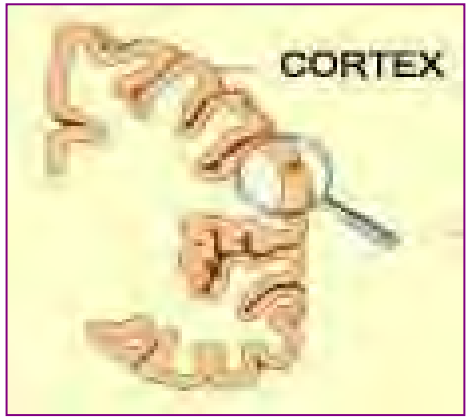


提纲

- 大数据简介
- 大脑新皮层的神经网络结构
- 大脑新皮层神经网络的记忆
- 大数据分析GPU深度神经网络计算平台

大脑新皮层的神经网络结构

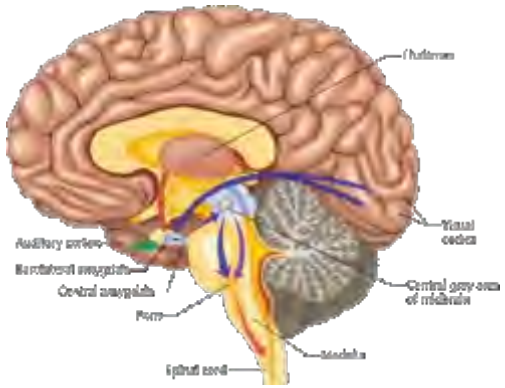


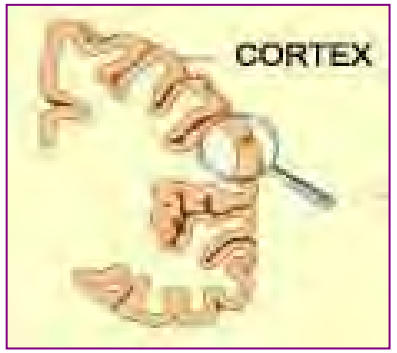


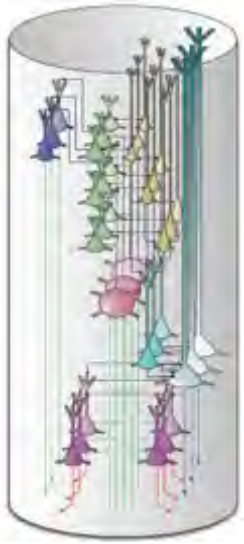
- 人脑的新皮层
  - 一张餐巾大小（约1000cm^2)
  - 六张扑克牌厚（约2mm ）
  - 每平方毫米的面积包含了约10万个神经元
  - 300亿个神经元
  - 100 万亿突触连接

- 神经科学认为：智能是由大脑神经网络的活动产生的。
- 几乎所有我们所认识到的与智能有关的内容，如感知、语言、想象力、数学、艺术、规划等等，都发生于大脑新皮层。
- 新皮层的神经网络对智能的产生起至关重要的作用。

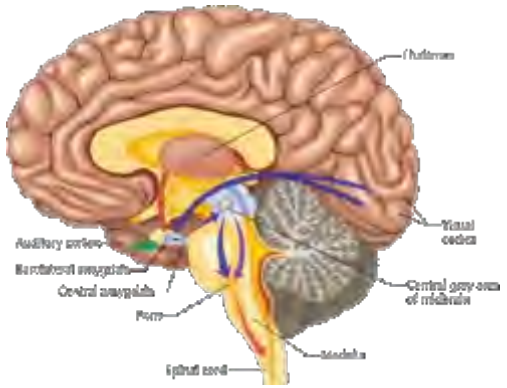
大脑新皮层的神经网络结构



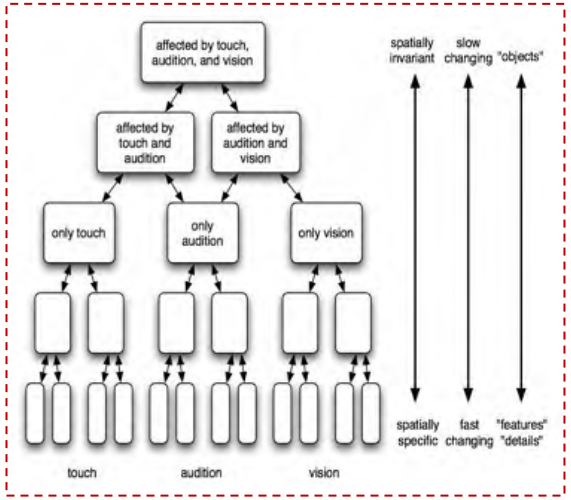


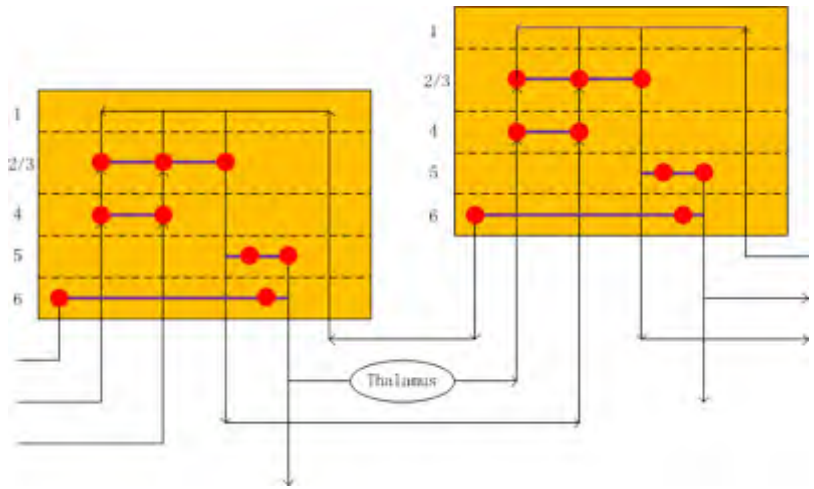


大脑新皮层的神经网络结构



按照功能区
构成层级结构

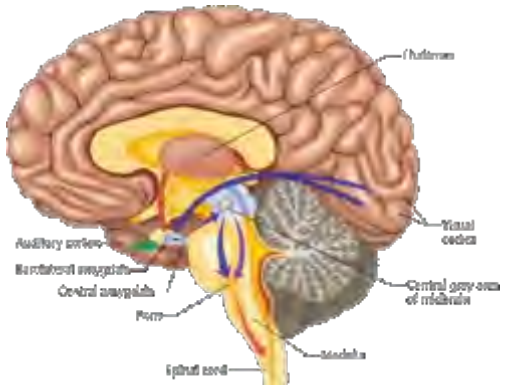


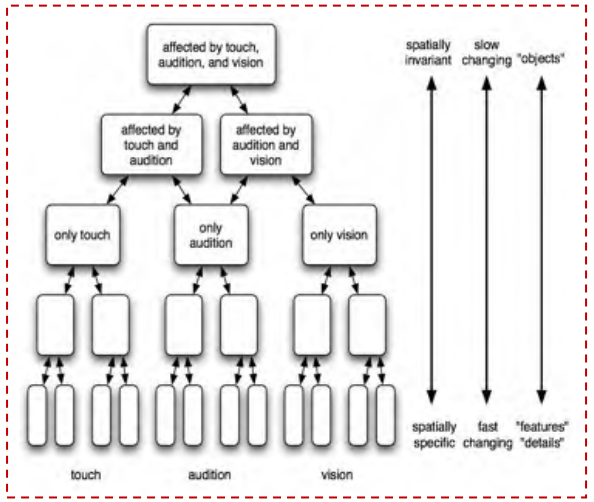


不同层级区域间依靠神经柱内复杂的连接进行通讯

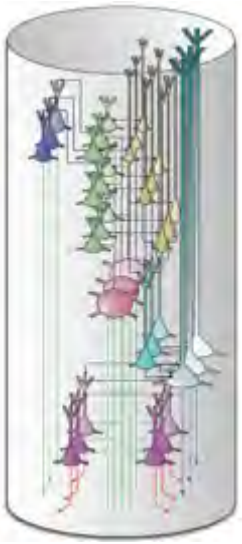
基因决定大脑皮层的整个结构，包括
各个区域之间的相互连接的具体细节

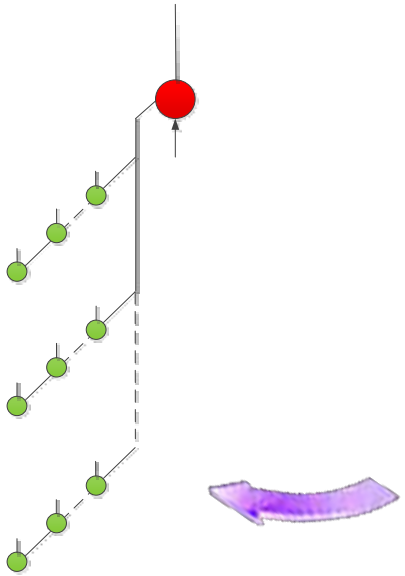
大脑新皮层的神经网络结构





新皮层连接的拓扑结构

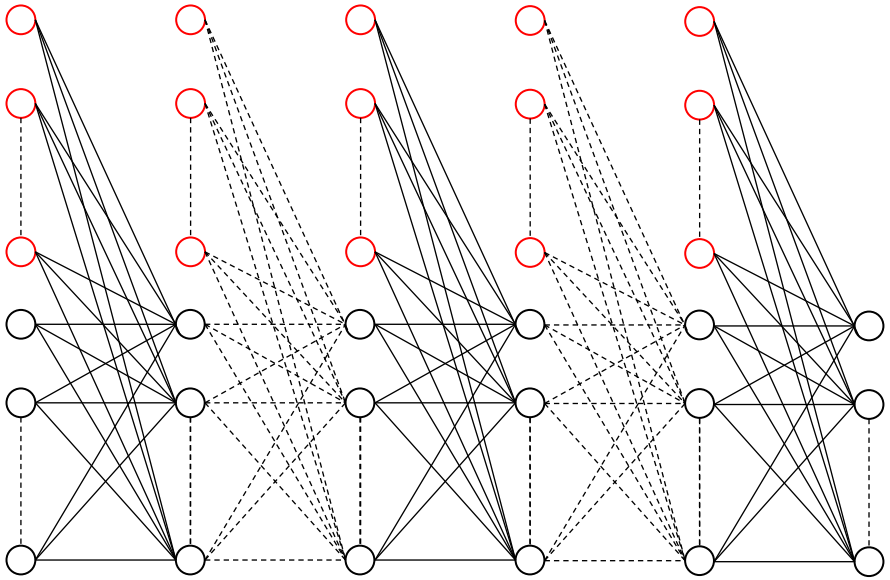


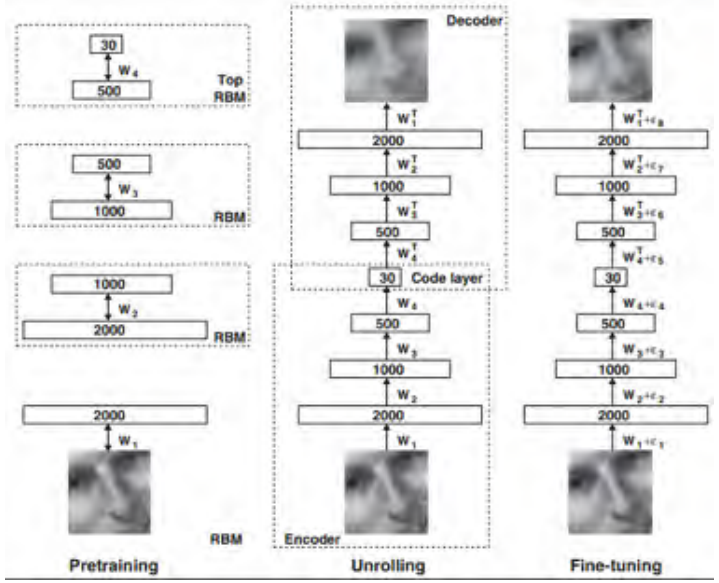


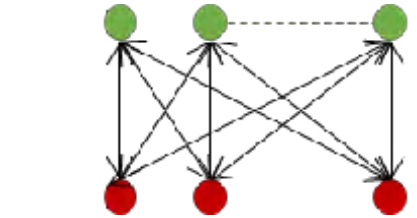
当前使用的简单神经网络结构

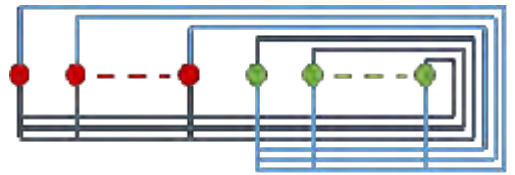
前馈深度神经网络

----无限逼近任何非线性函数









RBM神经网络

Geoffrey Hinton

2006年，G. E. Hinton教授在《Science》发表文章，基于前馈深度神经网络的大数据分析方法兴起

当前使用的简单神经网络结构

大数据 ⊕ 前馈深度神经网络

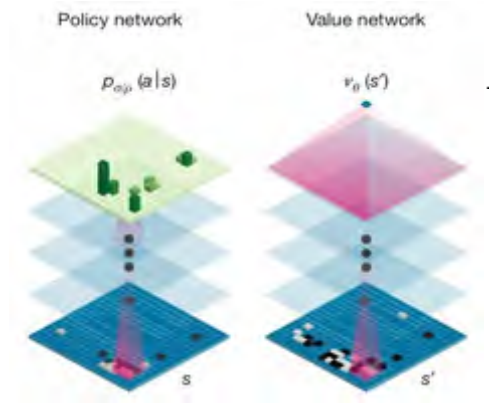
应用领域

- 语音识别：2012，RBM深度模型降低错误率30%，是10年来最大突破
- 图像识别：大规模对象识别(ImageNet)正确率(95.06%)超过人类(94.9%)

应用产品

4 AlpahGo Google : 1 李世石 世界冠军

2016年03月 AlphaGo



2012年10月
微软使用深度神经网络于语音识别，发布同声传译产品

2014年5月
香港中文大学DeepID2模型人脸识别率超越人类水平

IMAGENET

2015年2月
微软亚洲研究院大规模对象识别正确率超过人类

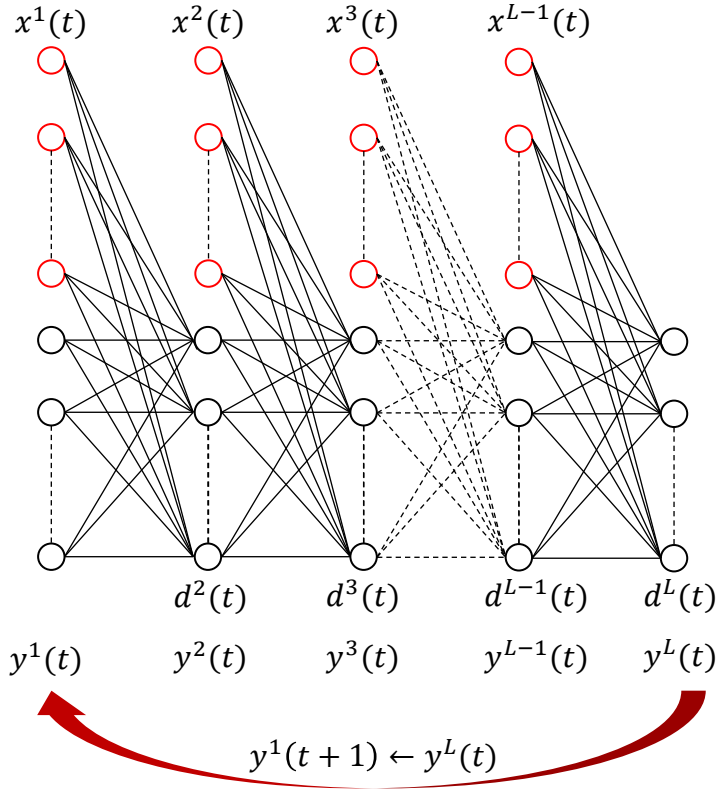
enlitic

2016年01月
美国企业Enlitic开发的基于深度学习的癌症检测系统，肺癌检出率超过放射技师

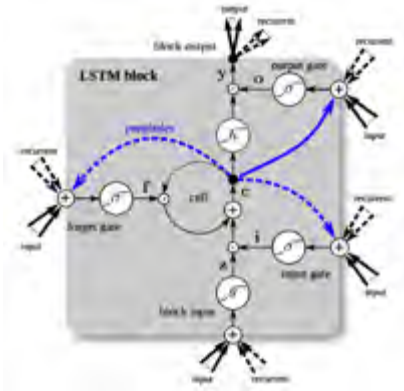
当前使用的简单神经网络结构

Recurrent 神经网络

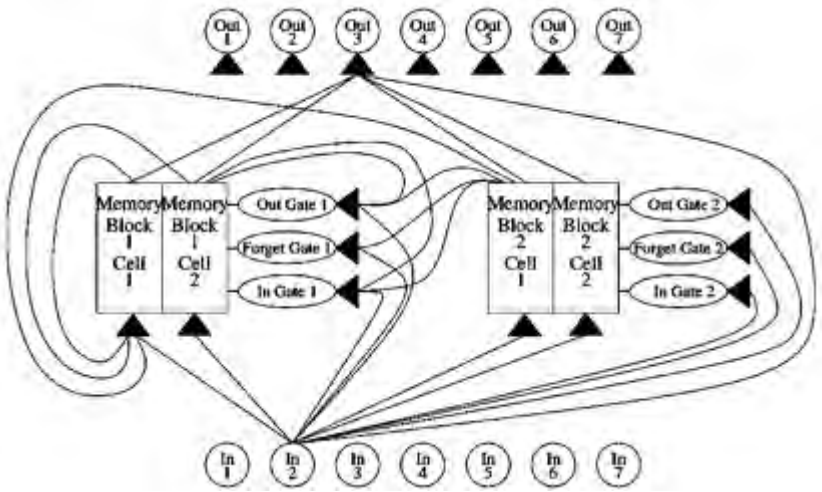
----无限逼近非线性动力学系统



Continuously Recurrent Computing



Jürgen Schmidhuber



LSTM神经网络

以LSTM神经网络为代表的，Recurrent深度神经网络迅速兴起，并广泛应用于时序大数据分析

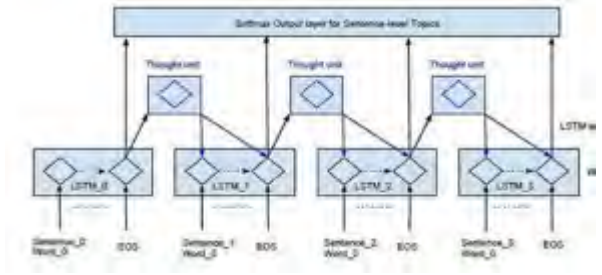
当前使用的简单神经网络结构

大数据 ⊕ Recurrent深度神经网络

应用领域

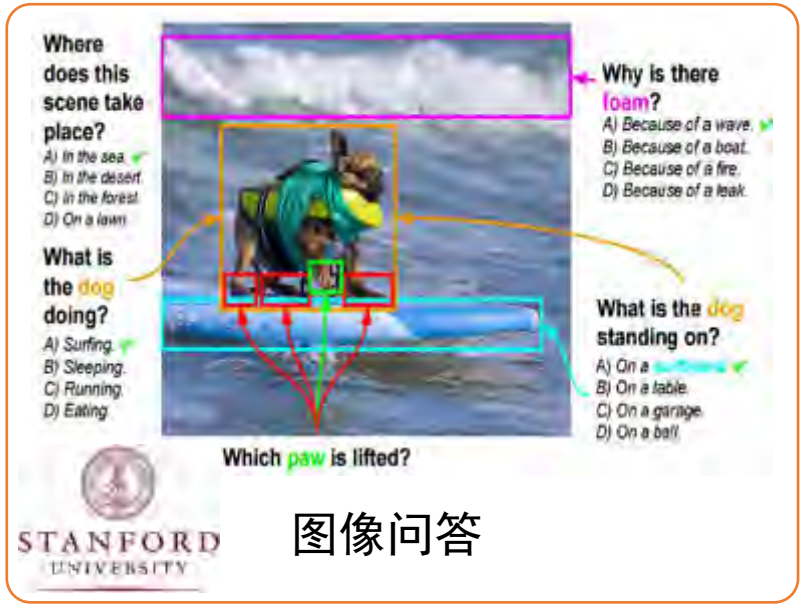
- **自然语言处理**：颠覆传统自然语言处理模式，突破自然语言处理前沿难关
- **视觉内容理解**：将视觉对象和自然语言相结合，打造可用的视觉内容理解产品
- **语音识别**：语音识别率大幅上升，入选MIT科技评论2016年十大突破技术

应用产品



在NLP的关键任务（接续语句预测）上，能做到20%的提升。这是问答系统得以发展的重要基础。

Google

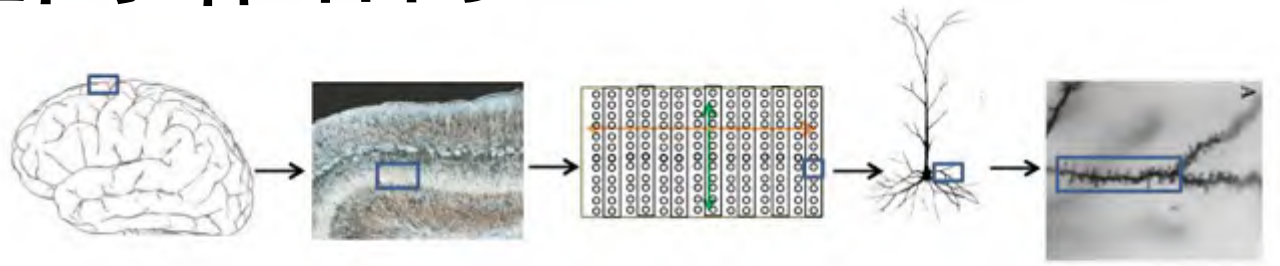


图像问答

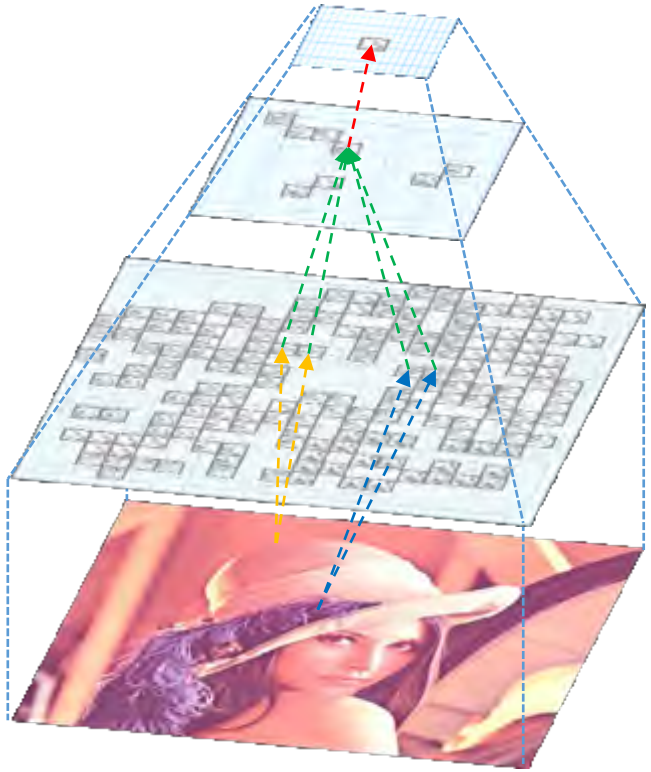
语音识别

当前使用的简单神经网络结构

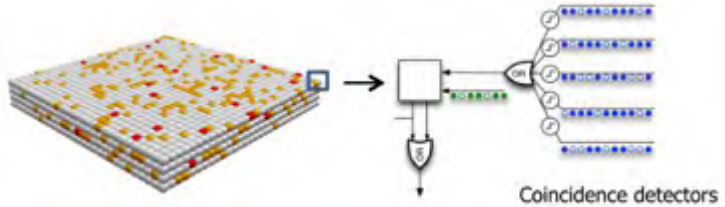
HTM神经网络



----模拟大脑新皮层活动的网络



HTM神经网络



Jeff Hawkins

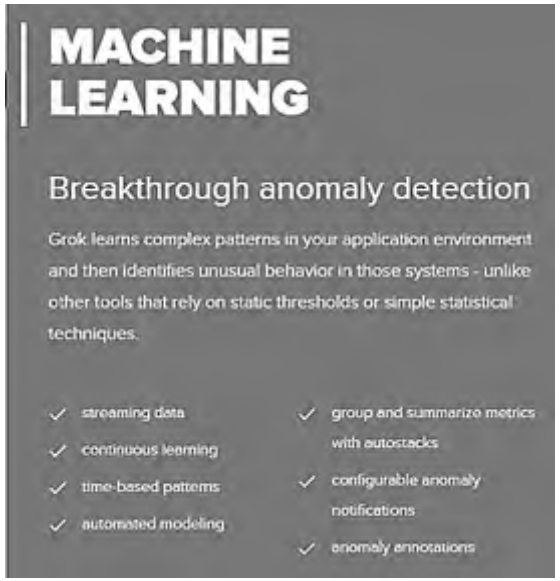
HTM神经网络以稀疏分布式表达为基础，实现“基于记忆的预测”的智能理论新框架，并将其应用于多种大数据分析

当前使用的简单神经网络结构

Numenta ⇨ GROK

大数据 ⊕ HTM神经网络

http://grokstream.com/



产品：Grok2.0, 序列异常检测

应用实例：

异常增涨预测

异常代码提交检测

提纲

- 大数据简介
- 大脑新皮层的神经网络结构
- 大脑新皮层神经网络的记忆
- 大数据分析GPU深度神经网络计算平台

深度神经网络方法

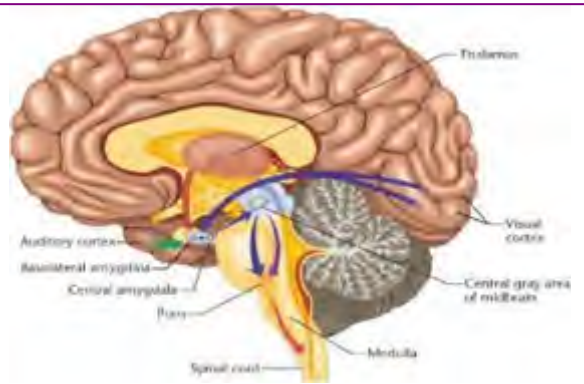
深度神经网络方法

神经网络结构⊕学习算法

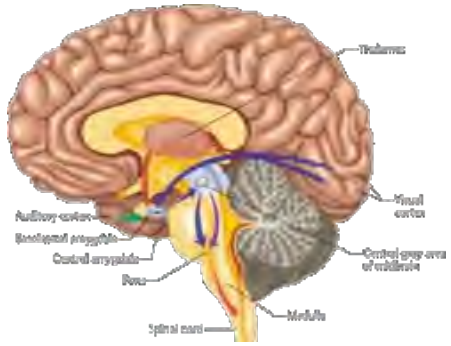
大脑新皮层
神经网络结构

大脑新皮层
神经网络的记忆机制

模拟大脑神经计算的方法

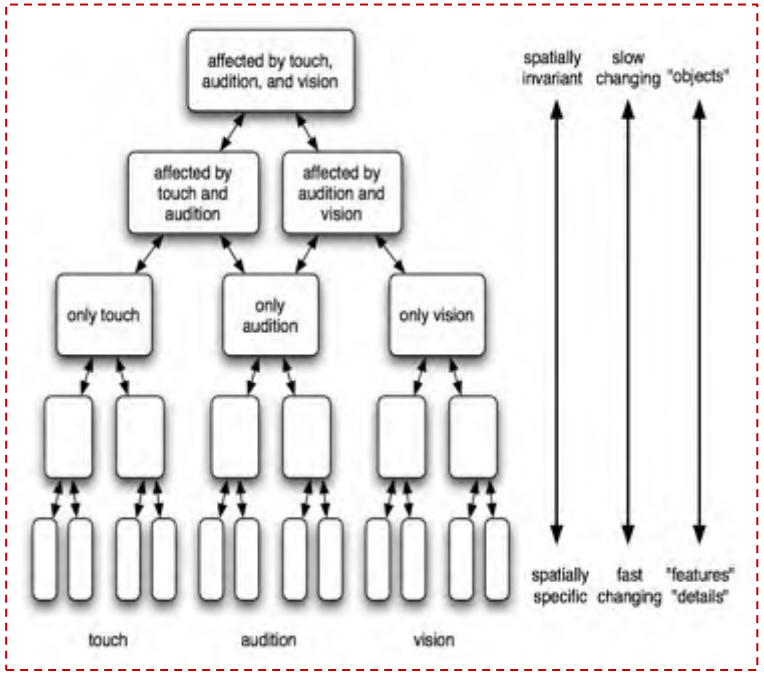


大脑新皮层神经网络的记忆



- 整个大脑皮层是一个记忆系统
- 大脑皮层以恒定形式存储模式
- 大脑皮层按照层级结构存储模式
- 大脑皮层可以存储模式序列
- 大脑皮层以自联想方式回忆模式
- 记忆通过学习来完成

新皮层连接的拓扑结构

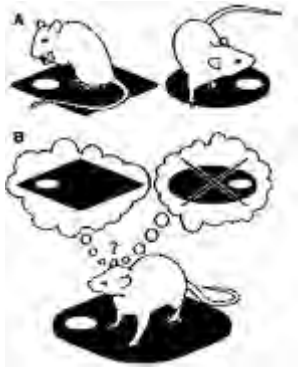


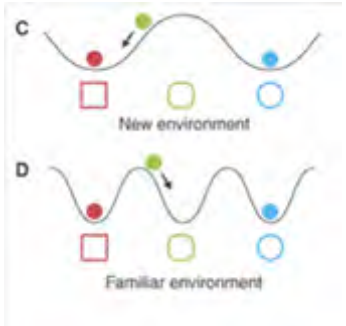
问题：记忆的学习机制是什么？

大脑新皮层神经网络的记忆

神经科学：记忆以动力学系统

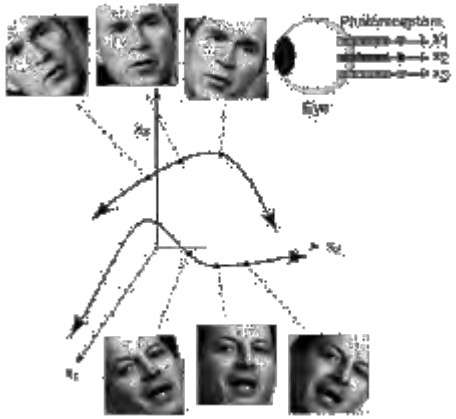
吸引子的方式存在



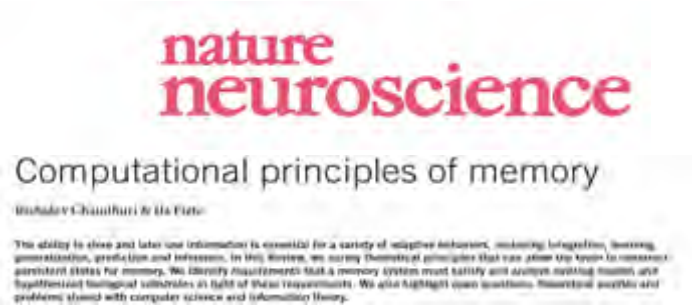


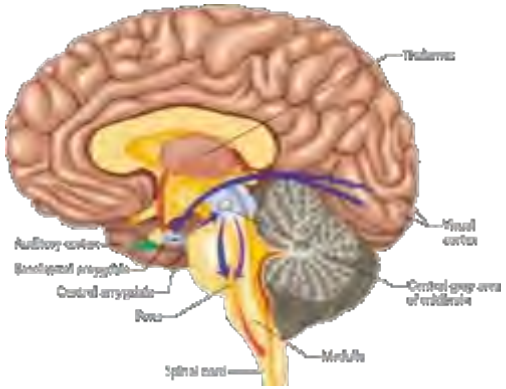
Daniel J. Amit

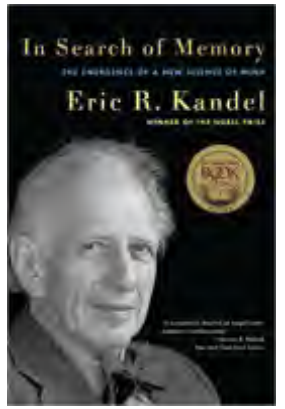
现代神经网络理论奠基人



Seung, H. Sebastian, and Daniel D. Lee. "The manifold ways of perception." *Science* 290.5500 (2000): 2268-2269.







Eric R. Kandel

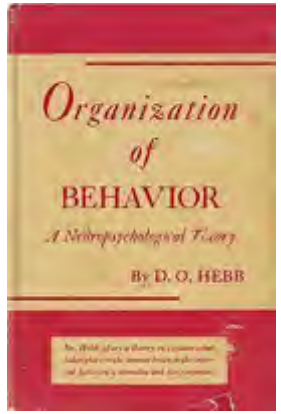
Columbia University 教授

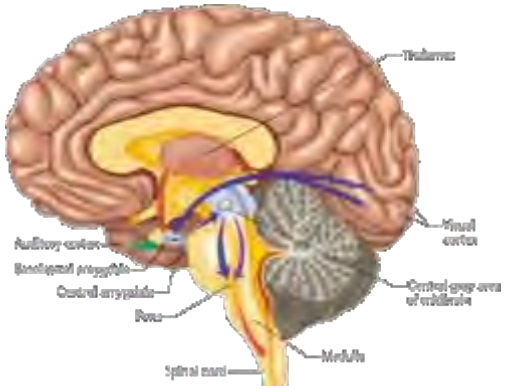
2000年诺贝尔生理学医学奖获得者

大脑新皮层神经网络的记忆

行为可以由神经元的
活动解释
-- D. O. Hebb, 1949

D. O. Hebb
英国皇家学会会士





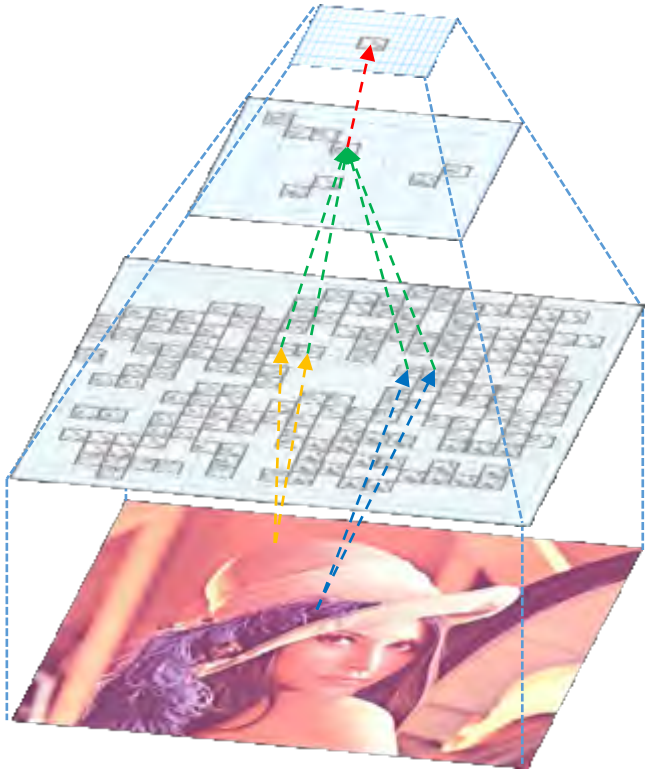
神经元之间的突触连接会随着突触两端神经元的激活而增强。

----Hebb学习规则

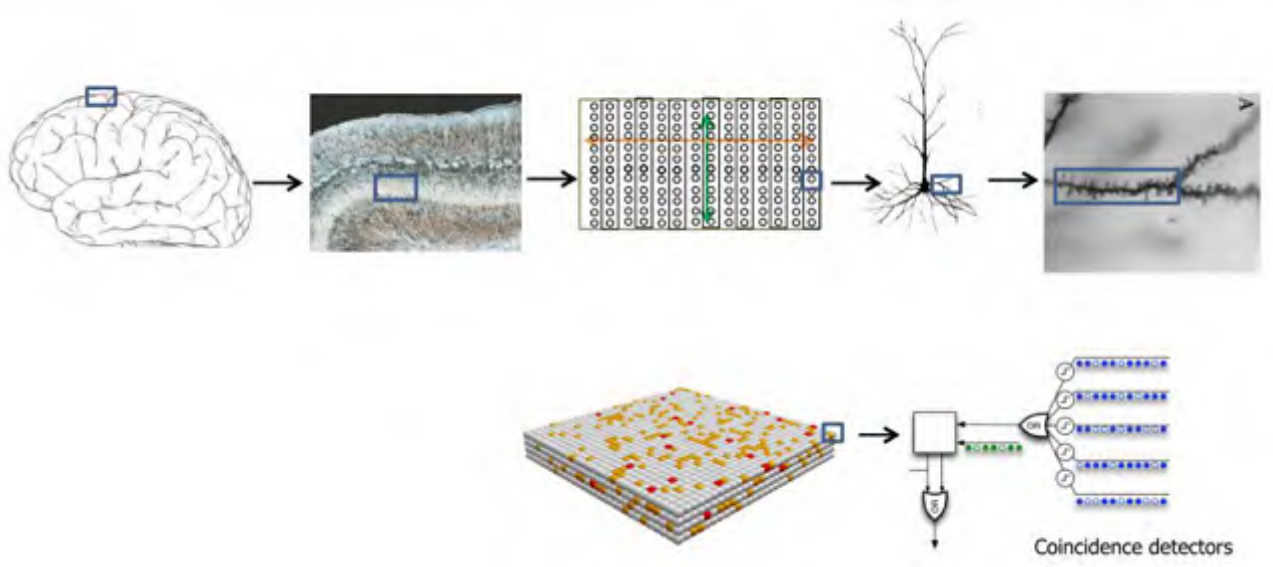
大脑新皮层神经网络的记忆

神经元之间的突触连接会随着突触两端神经元的激活而增强。
----Hebb学习规则

Jeff Hawkins



HTM神经网络常用学习算法：Hebb Learning

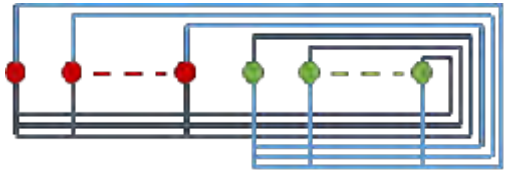


大脑新皮层神经网络的记忆

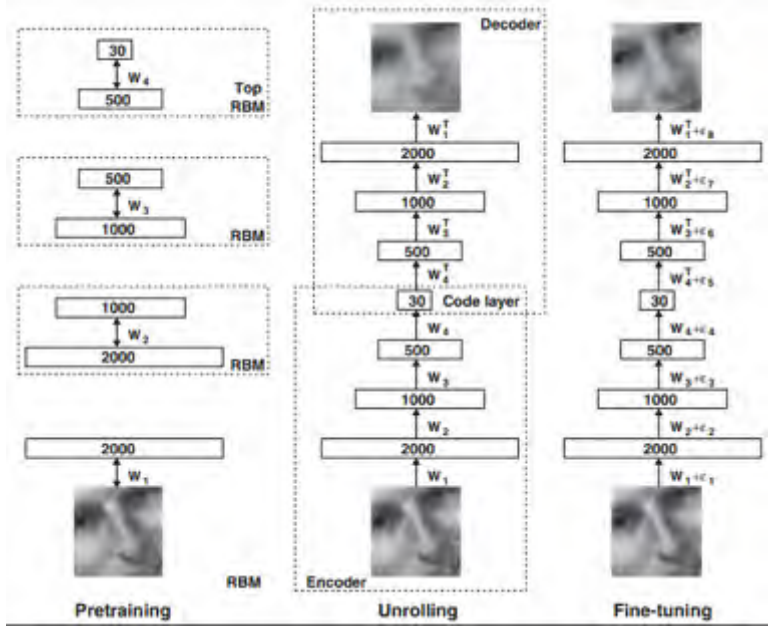
从数学的优化理论直接设计学习算法

前馈深度神经网络常用学习算法：
Back propagation，Autoencoder，RBM

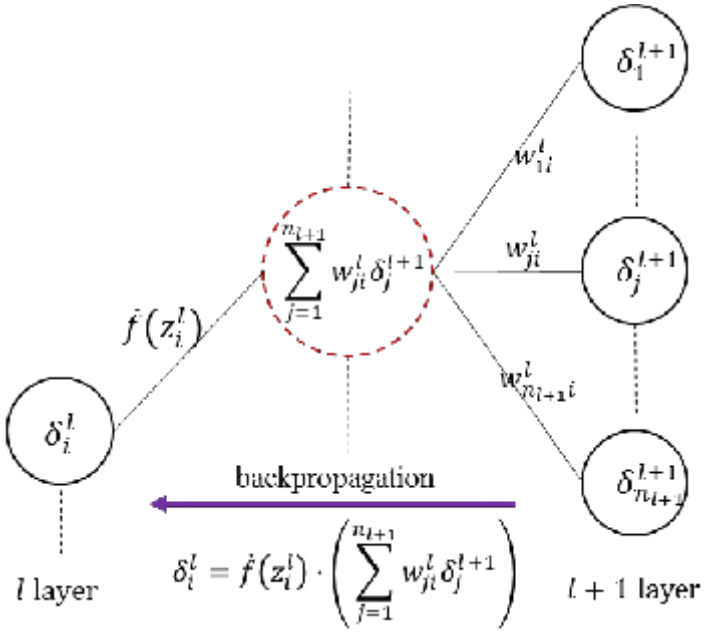
Geoffrey Hinton



RBM



AutoEncoder

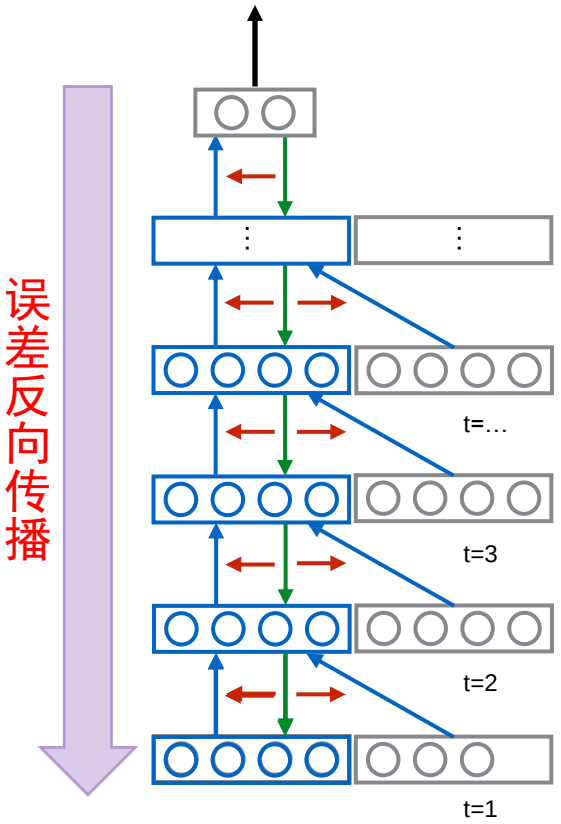


Back propagation algorithm

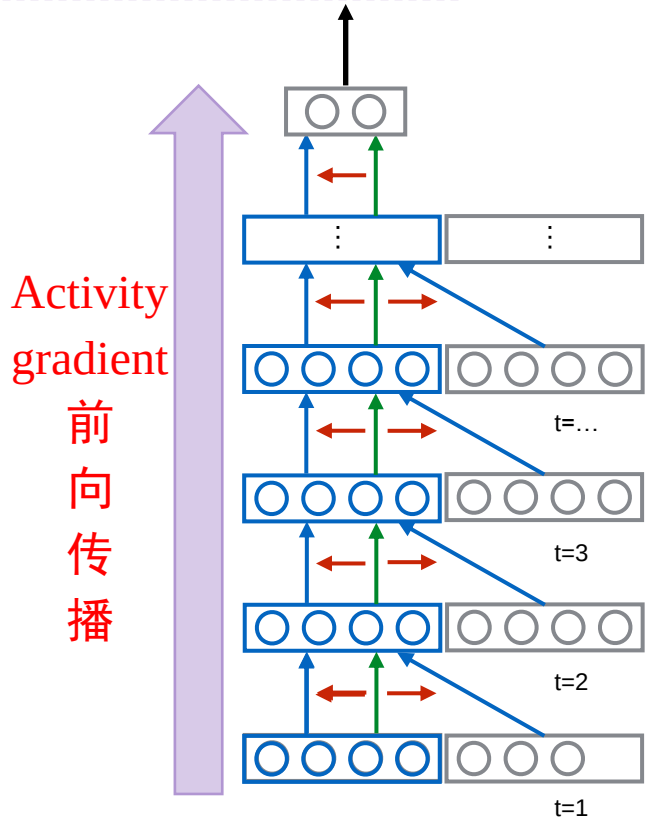
大脑新皮层神经网络的记忆

从数学的优化理论直接设计学习算法

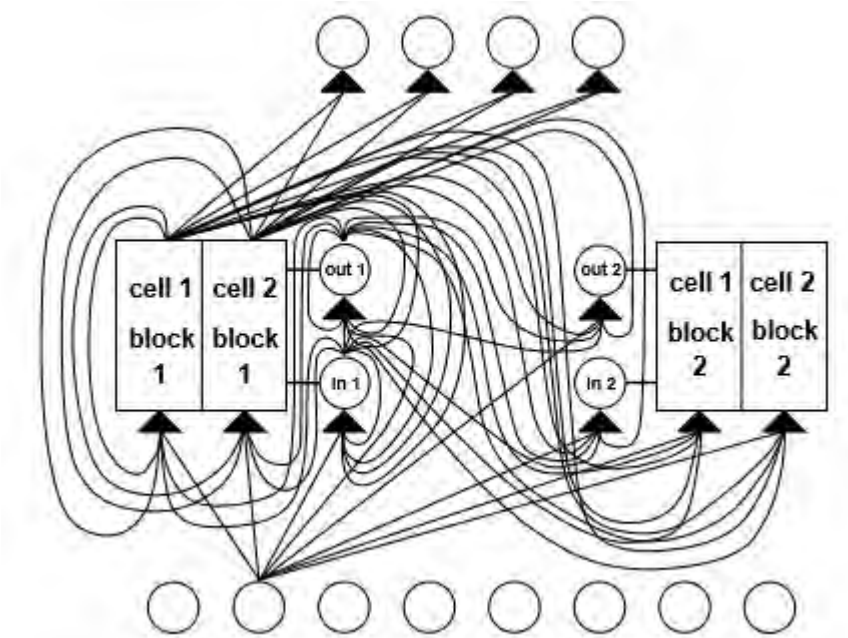
Recurrent深度神经网络常用学习算法：
BPTT，RTRL，LSTM



Backpropagation Through Time



Recurrent Real Time Learning

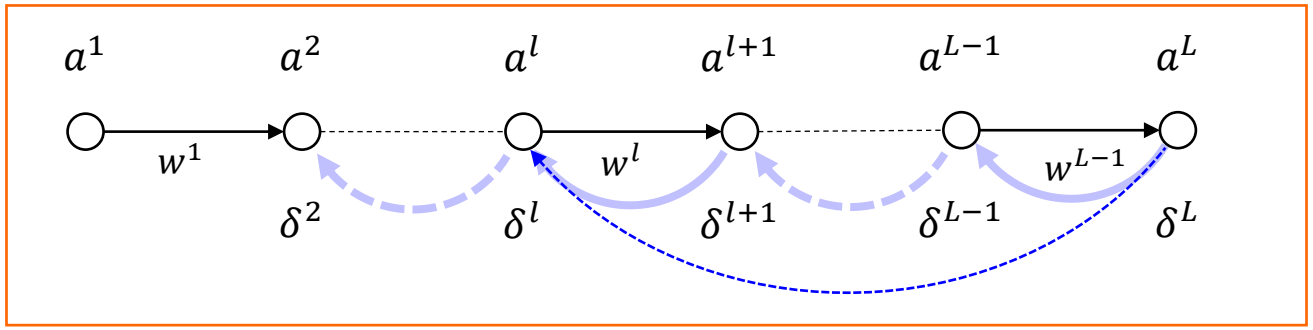


Long short-term memory

大脑新皮层神经网络的记忆

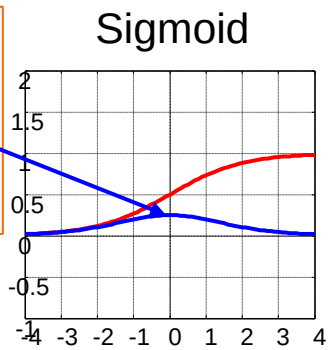
现有学习算法的问题：

- 梯度消失问题
- 学习冲突问题

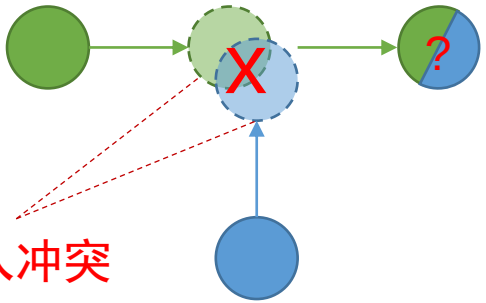


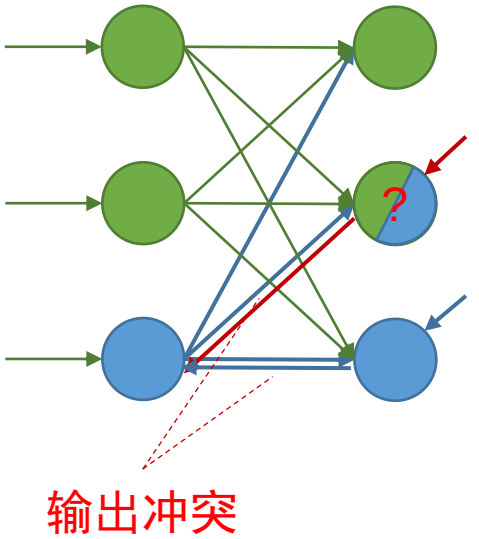
$$\left|\frac{\partial \delta^l}{\partial \delta^L}\right| = \prod_{m=L-1}^{l} \left| w \cdot \dot{f}(z^m) \right|, \text{ where } \dot{f}(z^m) \leq 0.25$$

which indicates δ^l will descent exponentially.



梯度消失问题



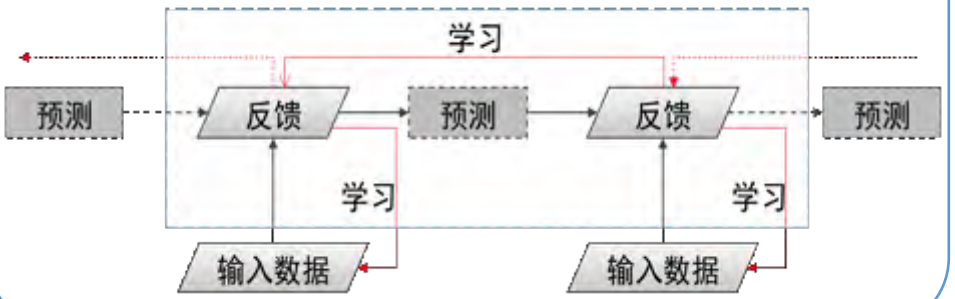


深度神经网络方法

深度神经网络方法

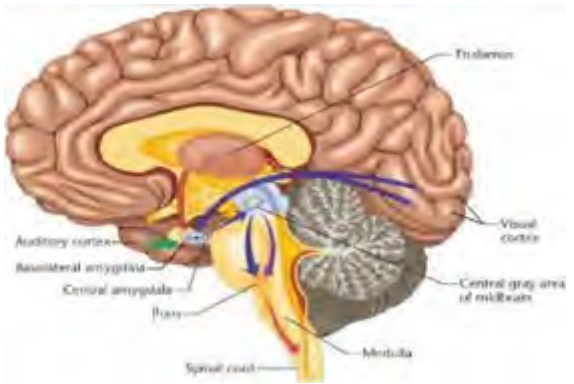
神经网络结构⊕学习算法

认知计算⊕数学方法



大脑新皮层
神经网络结构

大脑新皮层
神经网络的记忆机制



提纲

- 大数据简介
- 大脑新皮层的神经网络结构
- 大数据分析的深度神经网络方法
- 大数据分析GPU深度神经网络计算平台

大数据关键技术

大数据分析是大数据转换为价值的最重要的环节，否则，大数据仅仅是一堆数据而已。

大数据关键技术

展示平台

- 大数据知识展示
- 大数据产品

分析平台

- 深度神经网络
- 高性能计算平台

数据平台

- 大数据采集，标记
- 大数据存储，管理

大数据分析是大数据转化为价值的桥梁

大数据分析

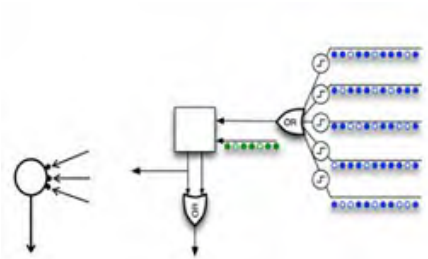
价值

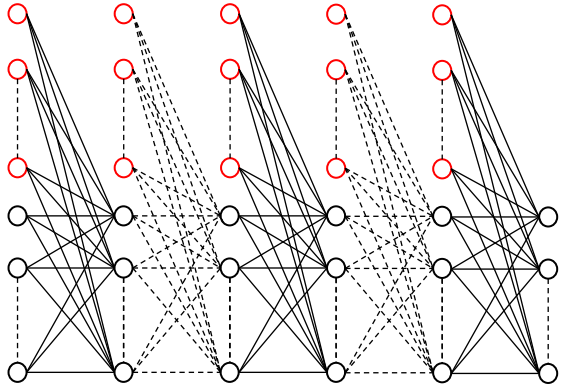
问题：如何构建大数据分析平台？

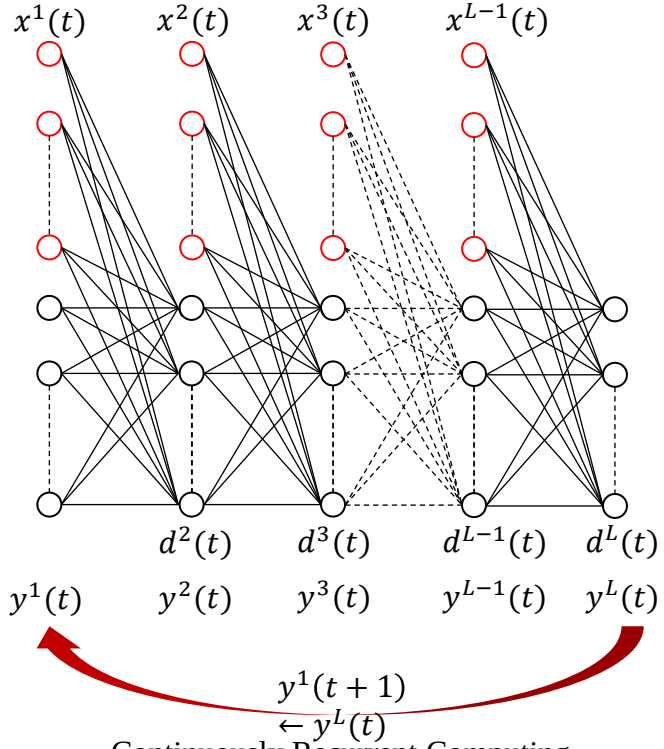
大数据分析平台

分析方法

深度神经网络方法





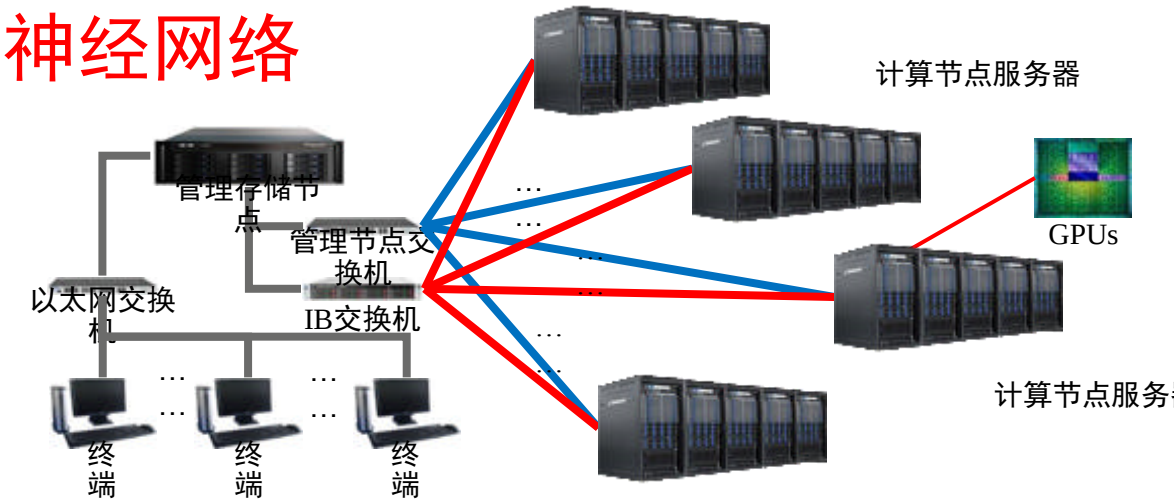


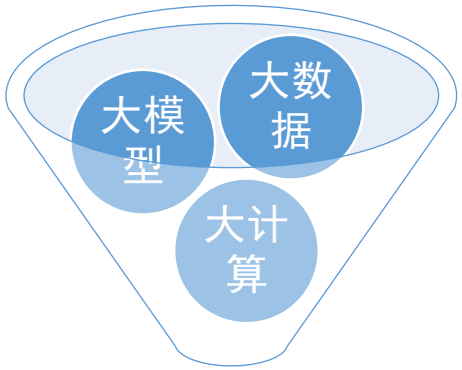
Continuously Recurrent Computing

\+

计算平台

GPU深度神经网络计算平台



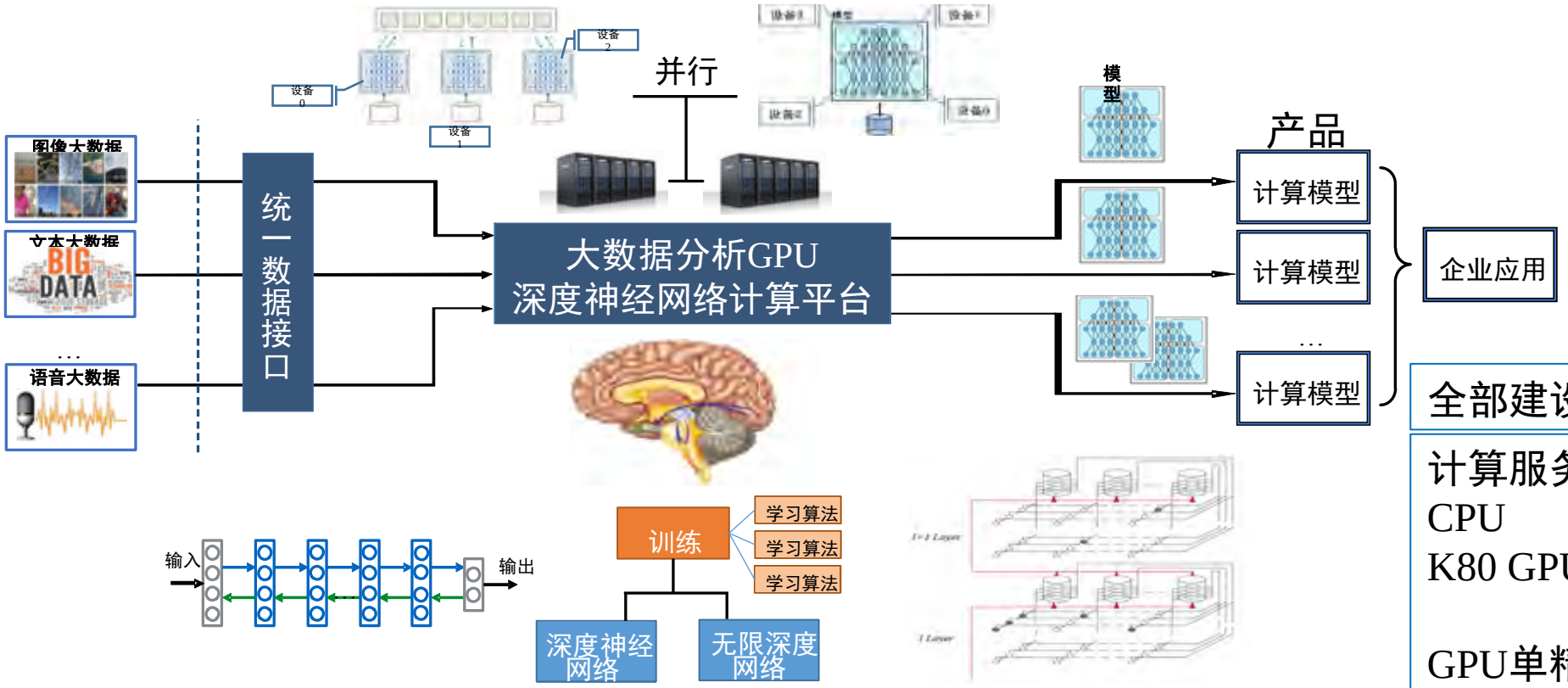


大规模计算平台

大数据分析GPU深度神经网络计算平台

四川大学

大数据分析GPU深度神经网络计算平台



全部建设完成后总体规模

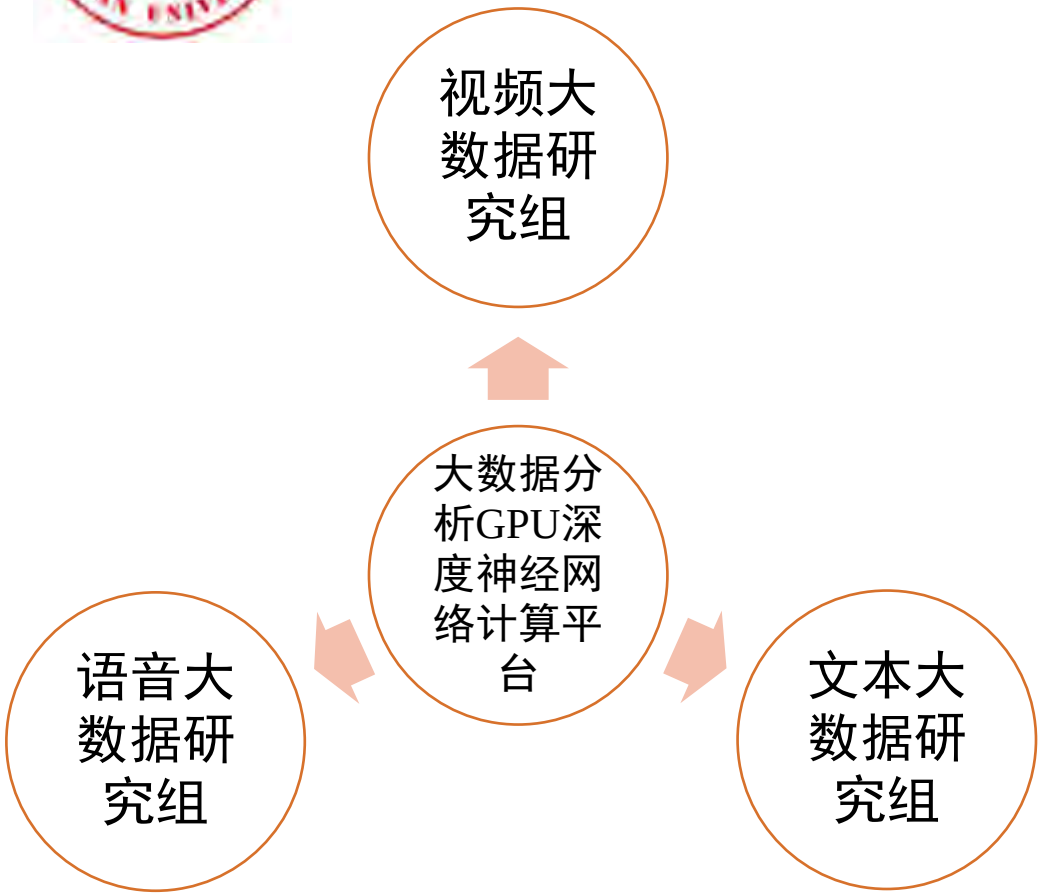
计算服务器	15台
CPU	120块
K80 GPU 计算卡	60块

GPU单精度浮点计算能力 500 Tflops

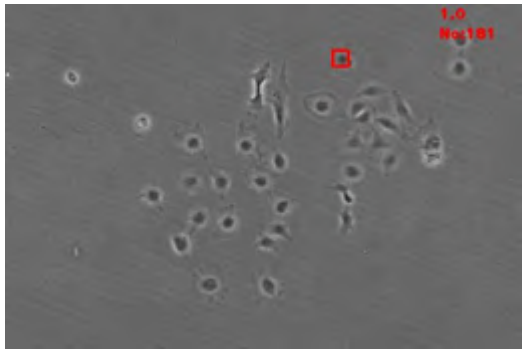
集群总计算能力达到 700 Tflops

大数据分析GPU深度神经网络计算平台

四川大学大数据研究团队



图像内容理解



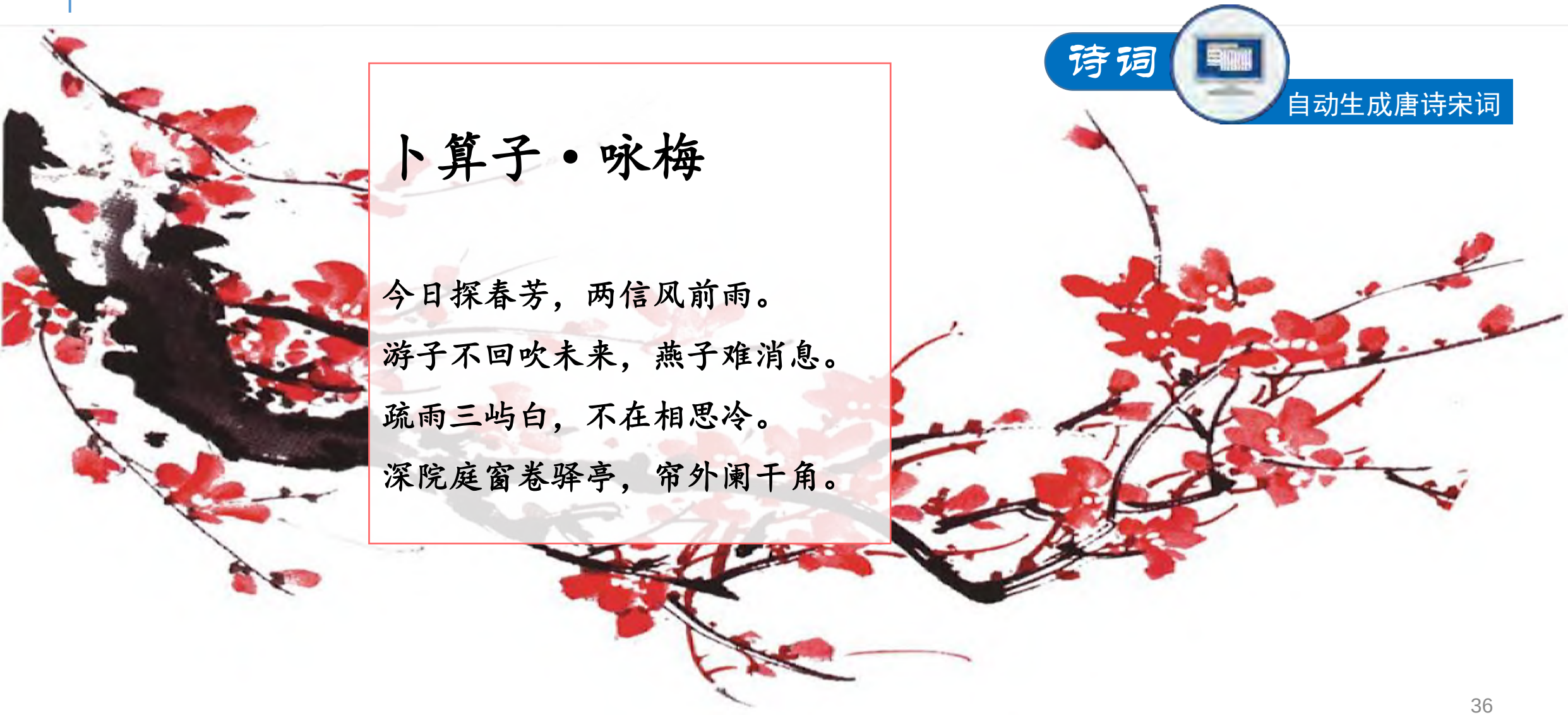
干细胞追踪系统

四川话识别

诗词精灵

在研产品

文本大数据研究产品-诗词精灵



诗词

自动生成唐诗宋词

卜算子·咏梅

今日探春芳，两信风前雨。
游子不回吹未来，燕子难消息。
疏雨三屿白，不在相思冷。
深院庭窗卷驿亭，帘外阑干角。

文本大数据研究产品-诗词精灵

诗词 自动生成唐诗宋词

虞美人·咏梅

朱阑缺后苍梅影。
间识番陶倒。
西风自倚半阑干。
舞地穿花深夜、旋笙箫。

人间紫燕怜飞去。
未重关风切。
城南楼外柳风丝。
遥想苍苔院宇、照江楼。

临江仙·梨花雨

压架清风清晓月，
未曾飞雨丁宁。
人间半醉赋诗词。
倩谁还说与，
未负大瓶花。

底事屈郎能几许，
持觞劝有光丝。
宝钗斜插半云莲。
温存风急暮，
断送暮春烟。

谢谢！