让机器学习得更快

科大讯飞 鹿晓亮

科大讯飞 iFLYTEK

主要内容

- 深度学习在感知智能中获得巨大成功
- 面向感知及认知智能的深度学习平台
- 深度学习平台训练算法并行方式探讨
- 深度学习平台对讯飞超脑计划的支撑

计算智能

能存会算

感知智能

能听会说、能看会认

认知智能

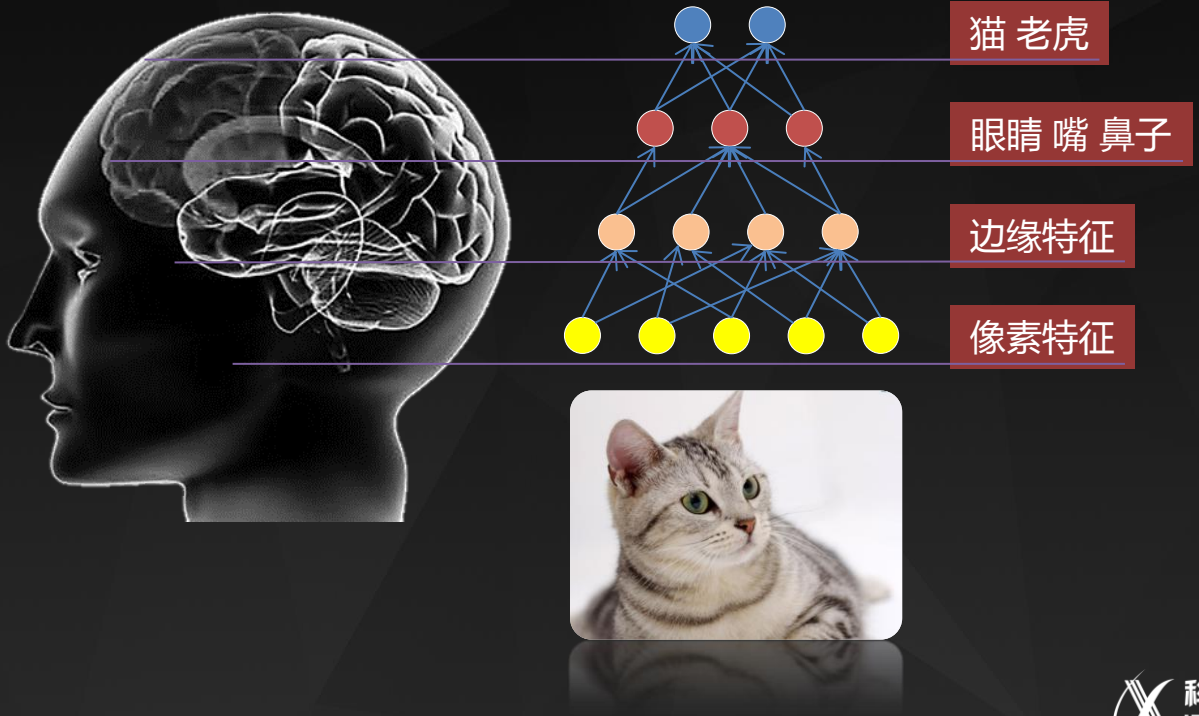
能理解会思考

科大讯飞 iFLYTEK

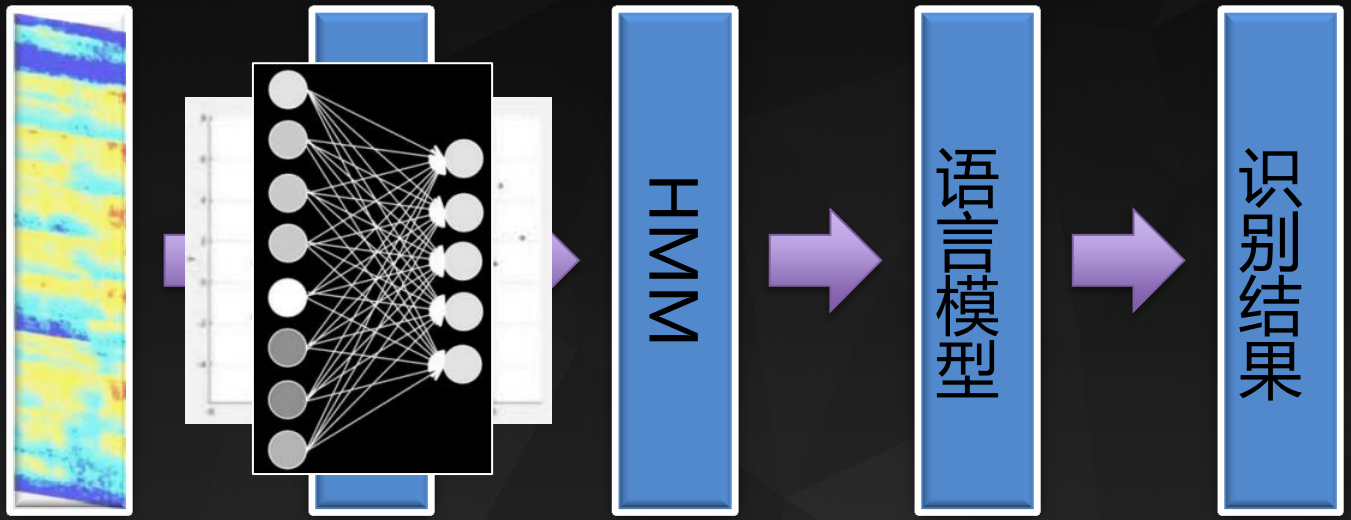
语音识别的血泪史

1920年代：RadioRex玩具狗
1950年代：Bell Lab Audry系统
6-70年代：DSP、DTW、Viterbi、HMM、DARPA
1980年代：特征提取、大规模语料、DARPA、NIST、Sphinx
1990年代：区分性训练、模型自适应、噪声鲁棒性、HTK
2000年后：更好的区分性训练技术等

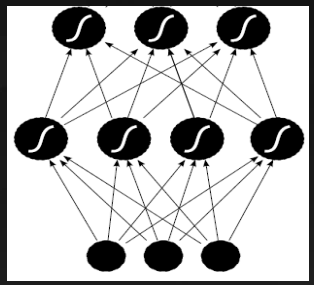
深度学习应用于语音识别

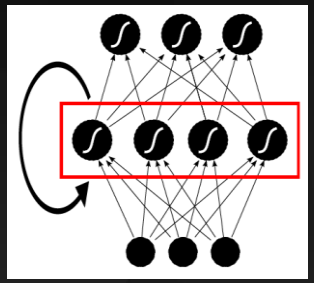


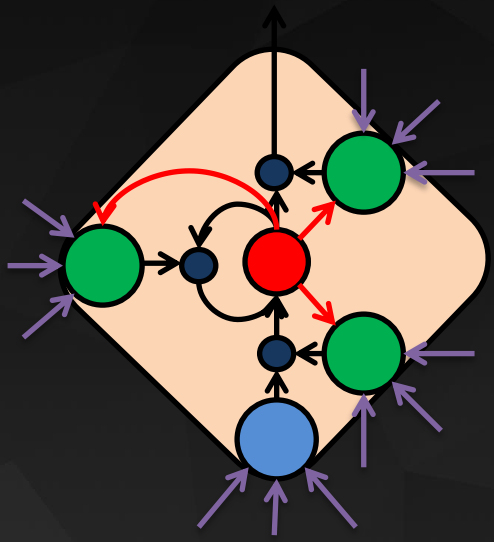
深度学习应用于语音识别



深度学习应用于语音识别







DNN

RNN

LSTM

大数据及云计算应用于语音识别

2010年10月28日，“语音云”在业界率先发布，为手机、汽车、智能家电等终端提供高质量语音合成、语音搜索、语音听写等智能语音交互服务能力

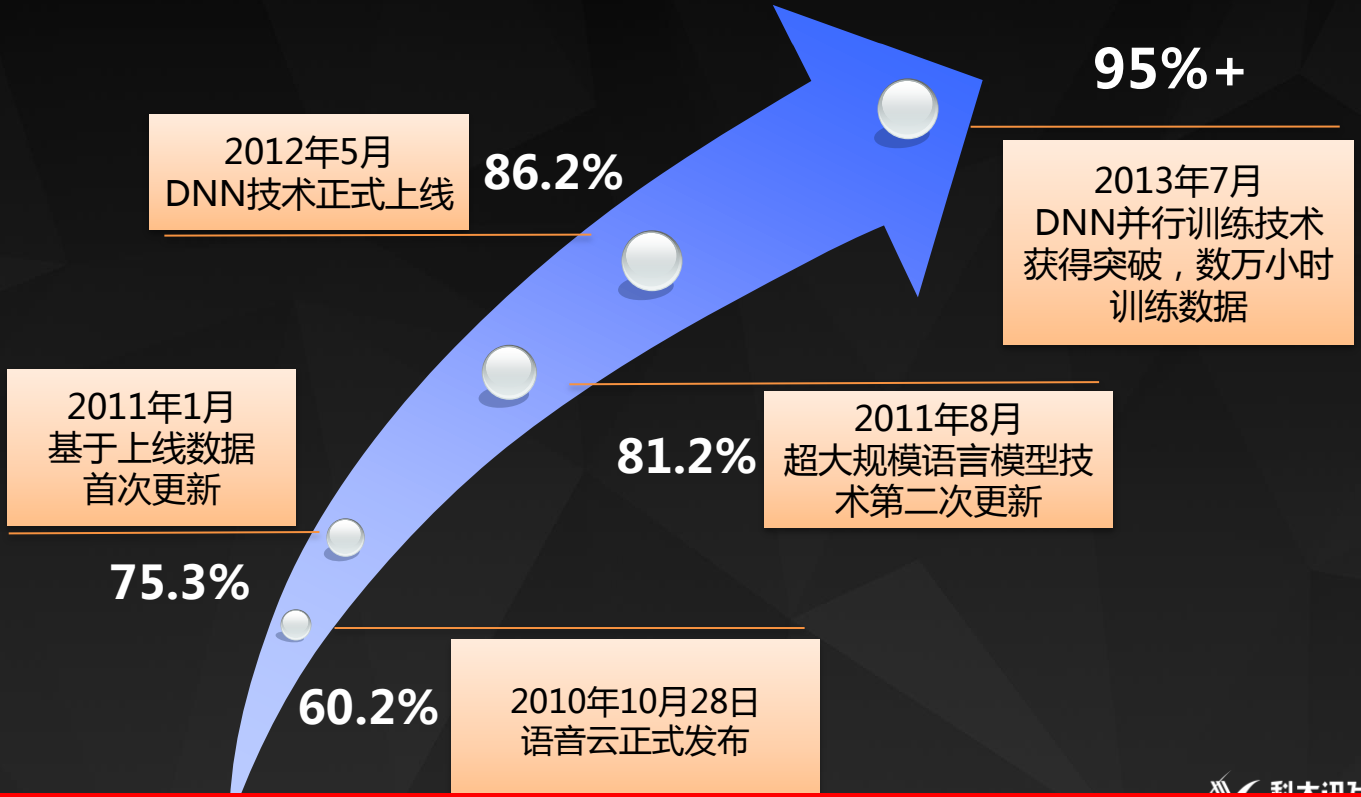
语音云启动仪式

周光召致辞

柳传志致辞

科大讯飞 iFLYTEK

深度学习和大数据的力量



未来几年将语音识别的句正确率提升到90%！

图像识别同样获得巨大成功

系统	方法	效果
DeepID3	DeepLearning	99.53%
Face++		99.50%
DeepID2+		99.47%
DeepID2		99.15%
DeepID		97.45%
DeepFace-ensemble		97.35%
FR+FCN		96.45%
GaussianFace	传统方法	98.52%
Betaface.com		98.08%
TL JointBayesian		96.33%
人眼		99.20%

主要内容

- 深度学习在感知智能中获得巨大成功
- 面向感知及认知智能的深度学习平台
- 深度学习平台训练算法并行方式探讨
- 深度学习平台对讯飞超脑计划的支撑

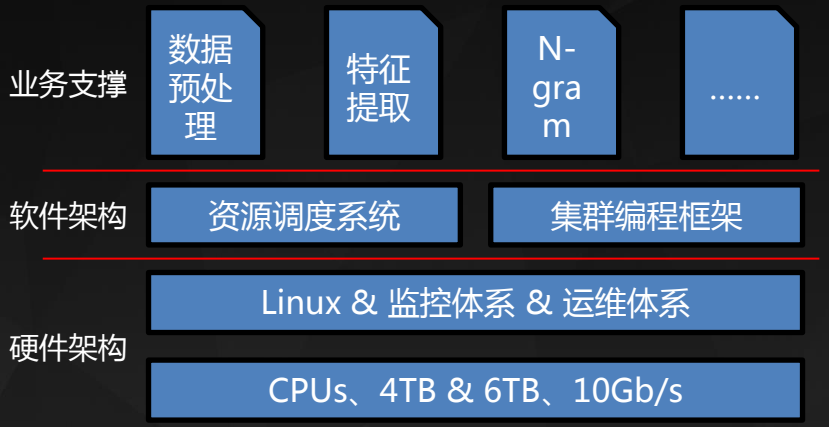
超算是人工智能的关键要素

- 深度学习技术的再度崛起，正在颠覆统计模式识别、机器学习和人工智能领域，相关专家成为“香饽饽”
- 大数据目前已经和深度学习融合，在语音识别及图像识别等感知人工智能方面发挥了巨大作用
- 超算平台是人工智能的基础，提供海量数据处理、存储以及高性能运算解决方案

科大讯飞 iFLYTEK

CPU集群

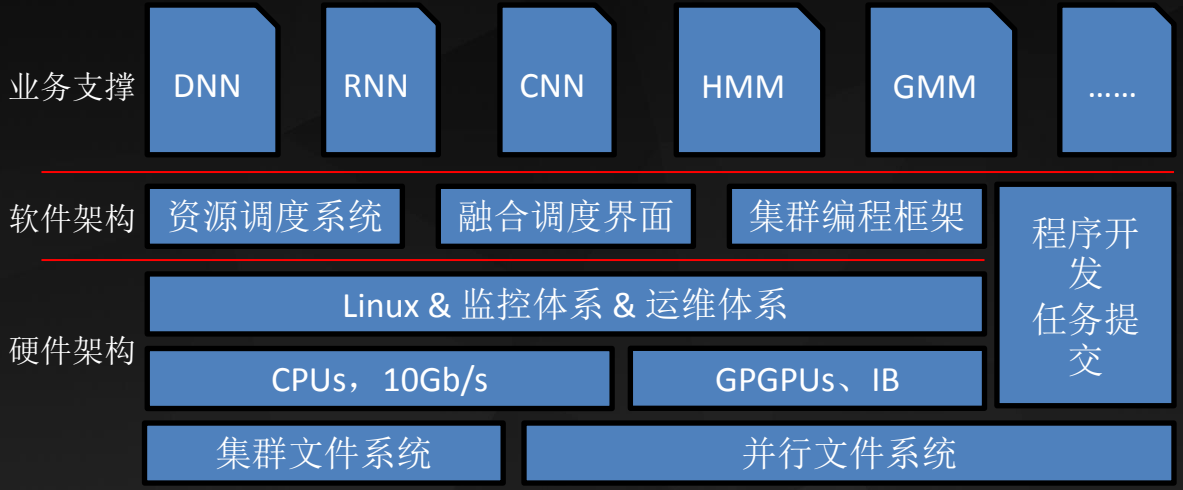
- 组成部分
  - 硬件组成
  - 软件调度
  - 支持业务
- 业务场景
  - 大规模数据预处理
  - 进行GMM-HMM等经典模型的训练



GPU集群

- 组成部分
  - 硬件组成
  - 软件调度
  - 支持业务
- 业务场景
  - 进行Deep Learning相关模型训练，如DNN、RNN、CNN等

深度学习平台



在硬件层面，全局设计网络方案、融合文件系统；在软件层面，重新设计并揉和调度界面、使HPC&BigData开发一体化；以提升程序开发效率和流程执行效率。

科大讯飞 iFLYTEK

主要内容

- 深度学习在感知智能中获得巨大成功
- 面向感知及认知智能的深度学习平台
- 深度学习平台训练算法并行方式探讨
- 深度学习平台对讯飞超脑计划的支撑

深度学习应用于语音识别

- **Acoustic model**

  DNN-HMM VS GMM-HMM

- **Computation of DNN in SR**

  model parameters : more than tens of millions

  speech corpus: more than ten thousand of hours

- **Acceleration**

  CPU – GPU – GPUs

深度学习应用于语音识别

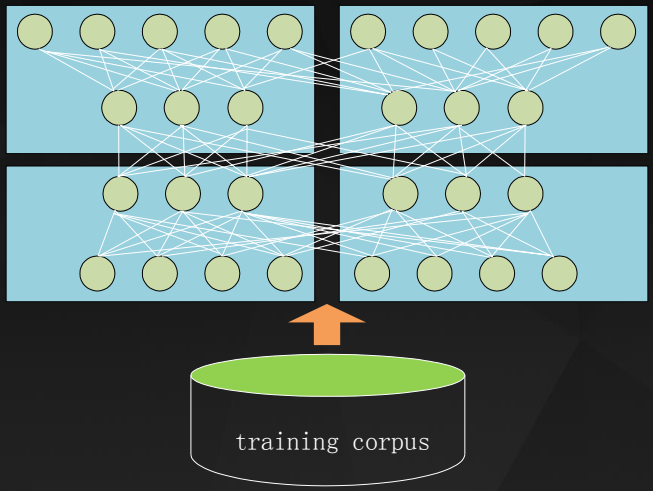


Fig. 2 Model parallelism

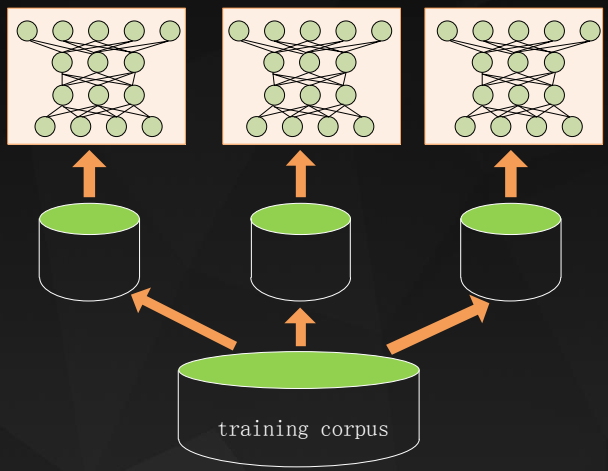


Fig. 3 Data parallelism

Tradeoff between Speed-up and Convergence

传统的异步SGD方案

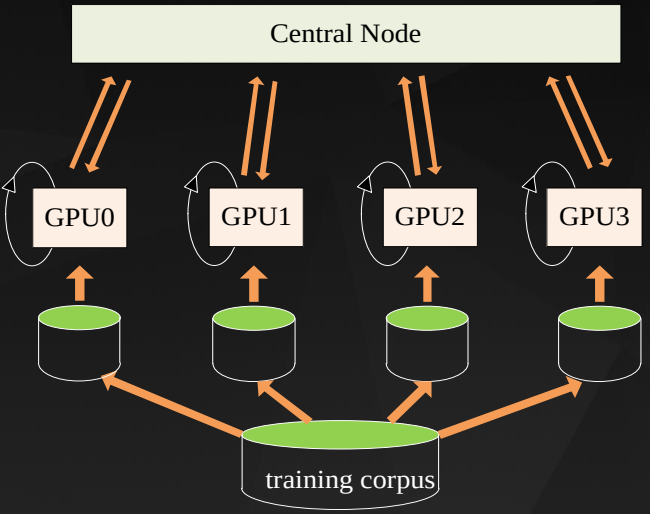


Fig. 4 ASGD applied to multi-GPU in a server [4][6]

- central node, high bandwidth requirement
- conflict between model latency and efficiency

环形并行学习策略

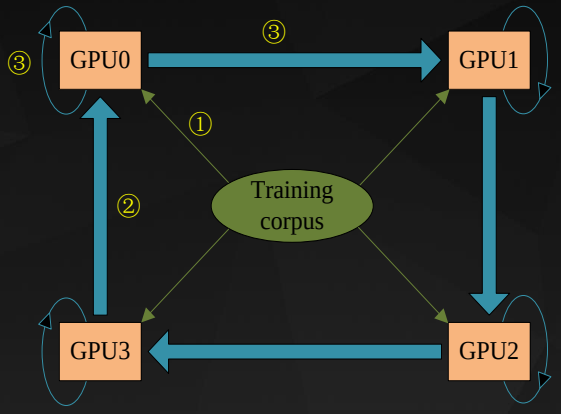


Fig. 5 Ring structure parallel strategy for multiple GPUs

① get mini-batch from training corpus

② receive the model from the previous node, and merge the local gradient to generate a new model

③ send the new model to the next node and train the next mini-batch simultaneously

环形并行学习策略

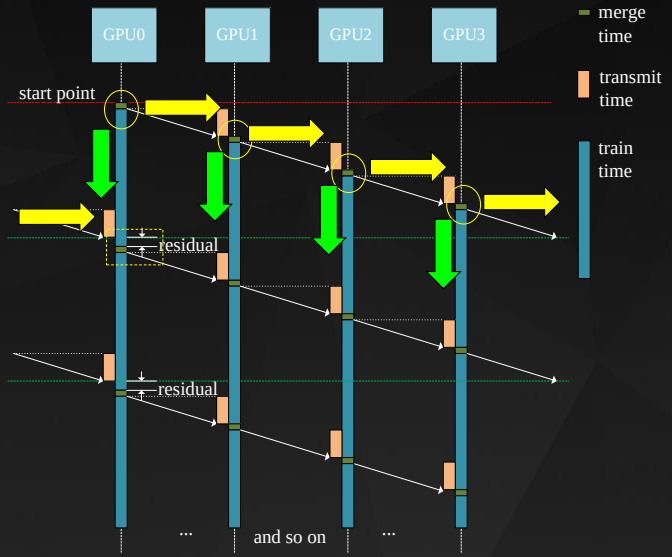


Fig. 6 Timing analysis of the RSPS

- asynchronous mode
- no central node, one transmission per mini-batch for each node, low bandwidth requirement
- easy to hide transmission

环形并行学习策略

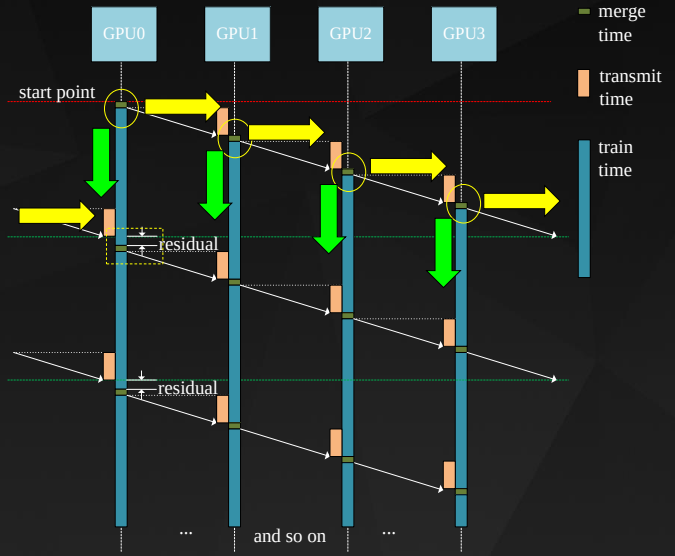


Fig. 6 Timing analysis of the RSPS

overlap of transmission and computation

$$T_{residual} = T_{calc} - [nT_{transmit} + (n-1)T_{merge}] \geq 0$$

$$n(T_{transmit} + T_{merge}) \leq T_{calc} + T_{merge}$$

$$n \leq \frac{T_{calc} + T_{merge}}{T_{transmit} + T_{merge}}$$

科大讯飞 iFLYTEK

环形并行学习策略

$$T_{wait} = \max\{-T_{residual}, 0\} = \max\{nT_{transmit} + (n-1)T_{merge} - T_{calc}, 0\}$$

$$Speedup = \frac{T_{\text{single}}}{T_{multipe}} = \frac{n(T_{calc} + T_{merge})}{T_{calc} + T_{merge} + T_{wait}}$$

$$Speedup = \begin{cases} n & if \quad n \le \dfrac{T_{calc} + T_{merge}}{T_{transmit} + T_{merge}} \\ \dfrac{T_{calc} + T_{merge}}{T_{transmit} + T_{merge}} & else \end{cases}$$

科大讯飞 iFLYTEK

环形并行学习策略

$$Speedup_{\max} = \frac{T_{calc} + T_{merge}}{T_{transmit} + T_{merge}}$$

- T_{calc} (larger mini-batch, eg. rectified linear units)
- $T_{transmit}$ (compress transmission data, eg. quantize the gradient)
- T_{merge} (overlap merging, eg. pipelining, hierarchical merging)

实验结论

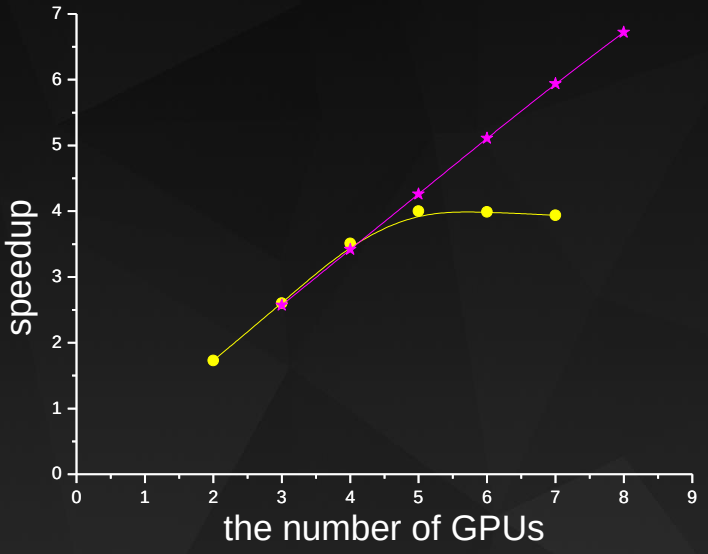
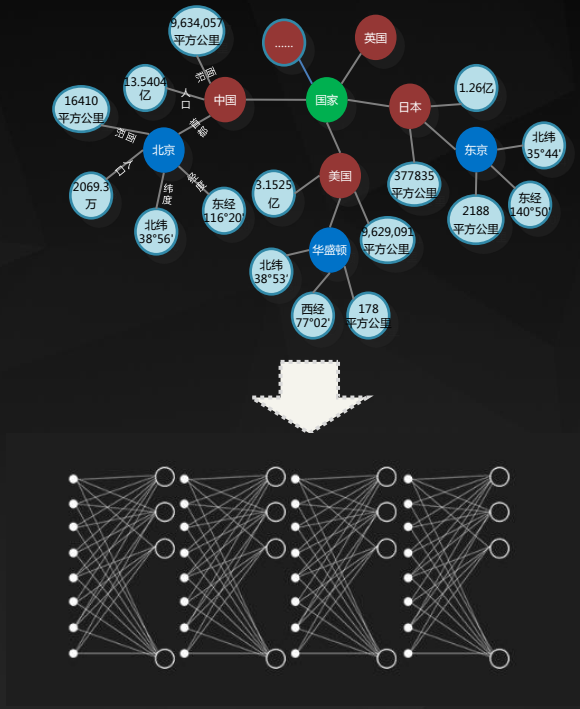


Fig. 7 Relationship between the speedup and the number of GPUs

主要内容

- 深度学习在感知智能中获得巨大成功
- 面向感知及认知智能的深度学习平台
- 深度学习平台训练算法并行方式探讨
- 深度学习平台对讯飞超脑计划的支撑

讯飞超脑计划

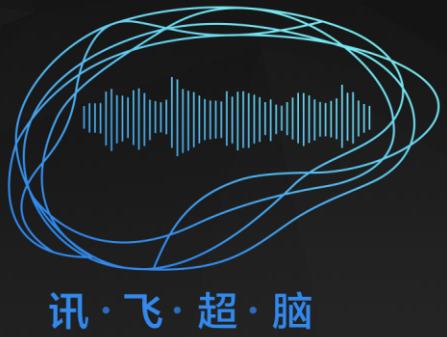


- **讯飞超脑的三大研究方向：**
  - 更加贴近人脑认知机理的人工神经网络设计，更好的支撑认知智能的实现
  - 实现与人脑神经元复杂度可比的超大人工神经网络（相当于目前感知智能网络规模的1000倍）
  - 实现基于连续语义空间分布式表示的知识推理及自学习智能引擎

科大讯飞 iFLYTEK

讯飞超脑预期成果

实现世界上第一个中文认知智能计算引擎！



- 通过模拟人脑的知识表示达到联想和推理
- 通过自动学习获取新的知识实现不断进化
- 通过自然交互（语音、文字）更加拟人化

科大讯飞 iFLYTEK

超算平台对讯飞超脑的支持

数千倍训练数据及数千倍模型参数的巨大挑战！

更大规模的超算平台集群建设

更优的深度学习并行化算法及集群调度算法

深度定制的人工神经网络专属芯片

科大讯飞 iFLYTEK

THANK YOU!

科大讯飞 iFLYTEK