百度语音技术产品介绍

百度多媒体组
贾磊

2013.2.24

通用语音识别服务简介

个人账号分享信息

个人用户账号信息

声学训练数据

语言训练数据

声学模型

输出分布 输出分布 输出分布

DNN

语言模型

语言模型自适应

解码空间组织和构建

网络接口协议

网络接口协议

网络接口协议

任务调度模块

解码器

识别结果返回

无线网络

端点检测

语音压缩

手机本地解析

返回识别结果/命令

语音服务内容

……

互联网下的语音识别的技术特色

1. 网络化的识别构架

- 客户端采集声音，压缩后上传
- 服务器对压缩语音进行识别，然后识别结果下发
- 网络通讯质量还对语音识别服务有很大的制约

2. 海量语言模型训练语料和语音层信息的快速更新

- 庞大的训练语料群，服务器集中识别方式允许采用上百亿文法的语言模型
- 语音识别系统的语言层信息的快速更新
- 各种网络服务为语言层信息更新提供了良好的语料基础

3. 海量的来自各种平台的声音特征

- 海量的用户群保证了语音识别系统的声音的多样性和语音通道的多样性
- 机器学习算法可以自动的挖掘有利于提高系统识别精度的声音特征

4. 庞大的计算资源和服务平台

- 需要大规模的计算机群进行海量声音、文本语料的相关模型训练
- 需要数量众多的计算机进行线上服务，资源耗费严重

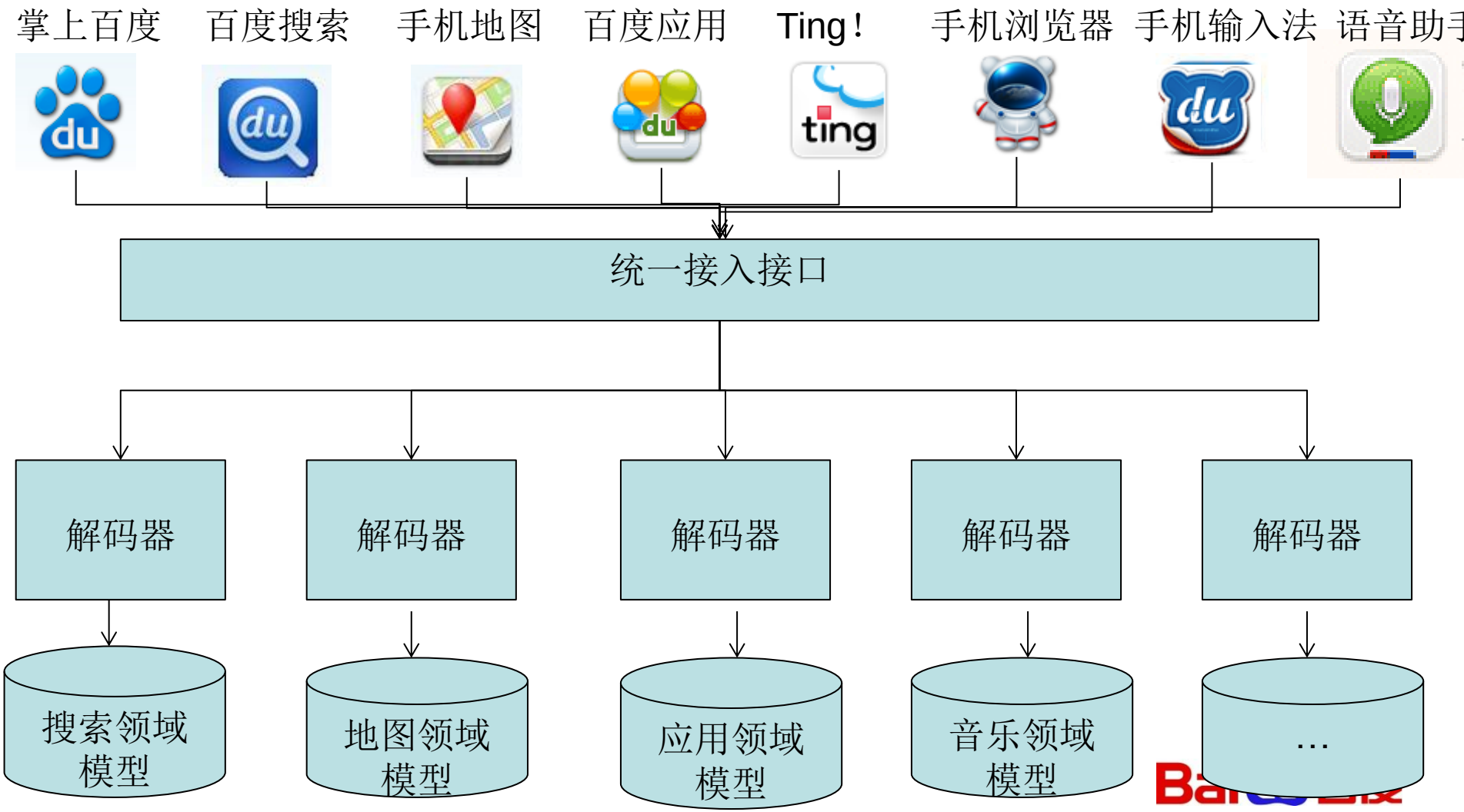
Baidu 百度

整合百度搜索资源的语音搜索

1. 网页搜索
2. 音乐搜索
3. 百度知道
4. APP搜索
5. 百度百科
6. 百度框计算

Baidu 百度

一套网络架构支持多样化产品

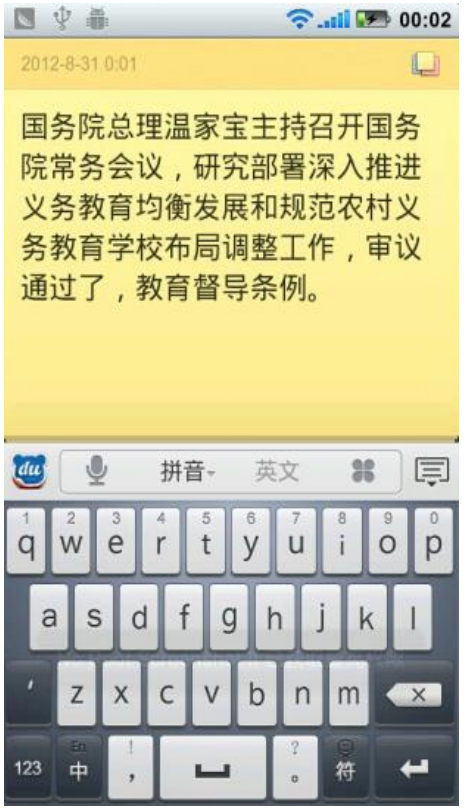


适合多种文体的高精度语音输入

短信输入

微博输入

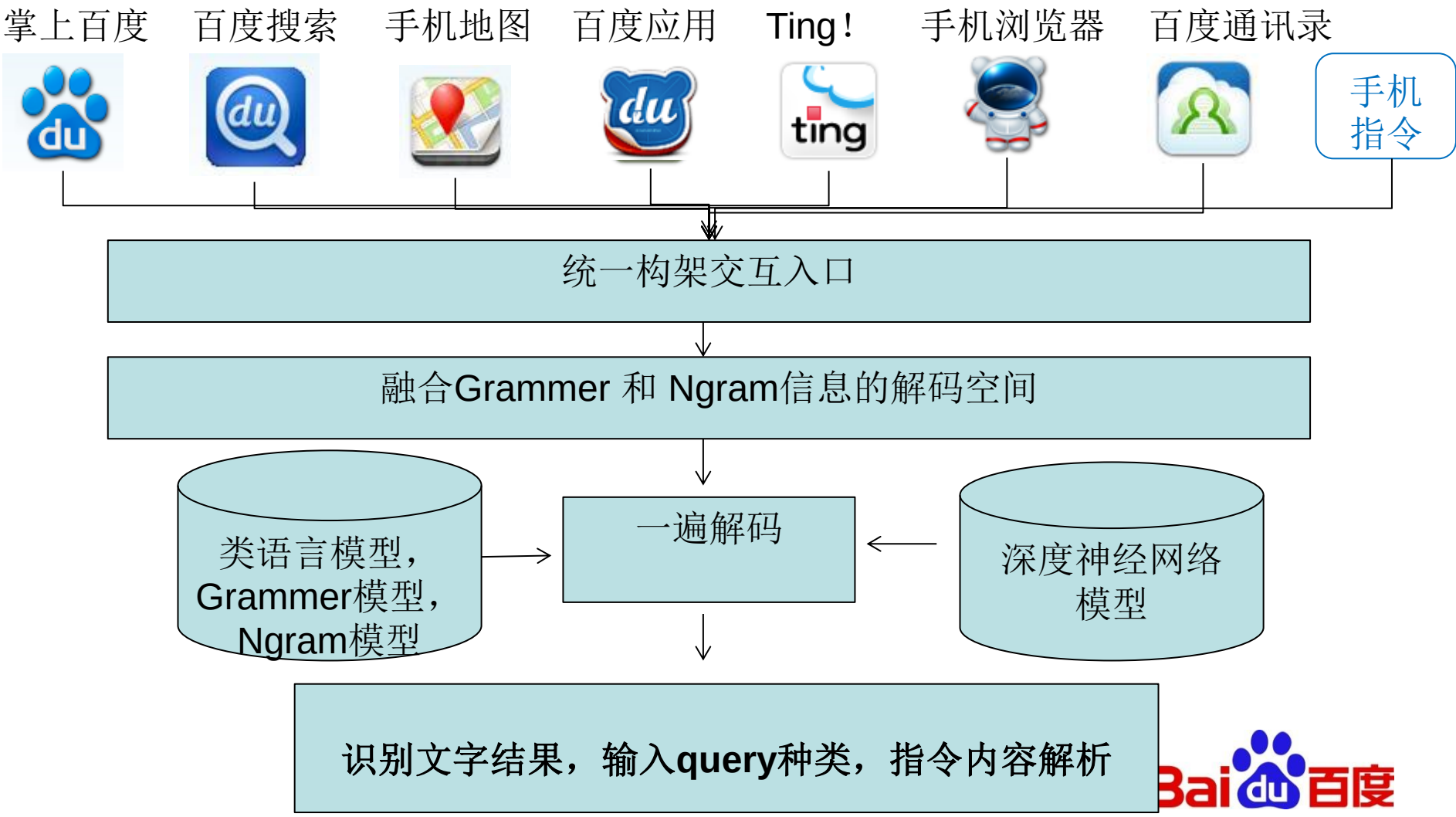
古诗输入



新闻输入

Baidu 百度

一套解码器支持多种应用（统一入口技术）



百度语音项目-手机语音助手

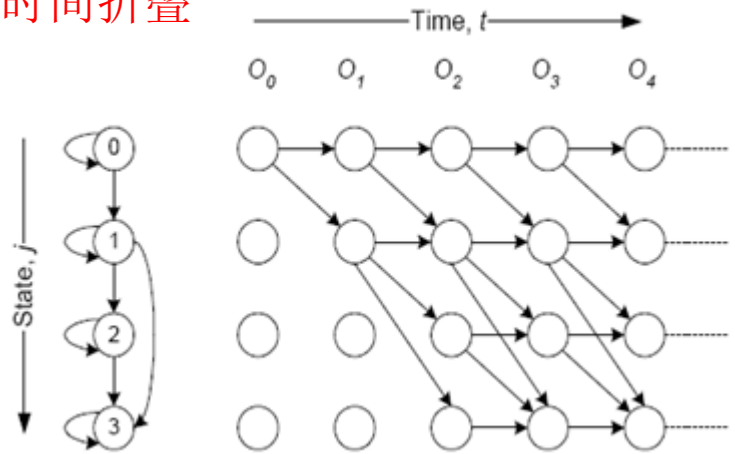
1. 高精度的手机语音指令识别
2. 高精度的通用搜索和垂直领域语音搜索
3. 语音自动问答和对话理解

内容 丰富优质内容知识库

声学建模简单介绍

1. One word： 我 是 中国 人
2. Prouncation: W o Sh i Zh ong G uo R en
3. Context dependent Demi-syllable modeling:
   Zh ： i-**Zh**-ong
   Ong : Zh-**ong**-G
   G ： ong-**G**-uo
   Uo: G-**uo**-R

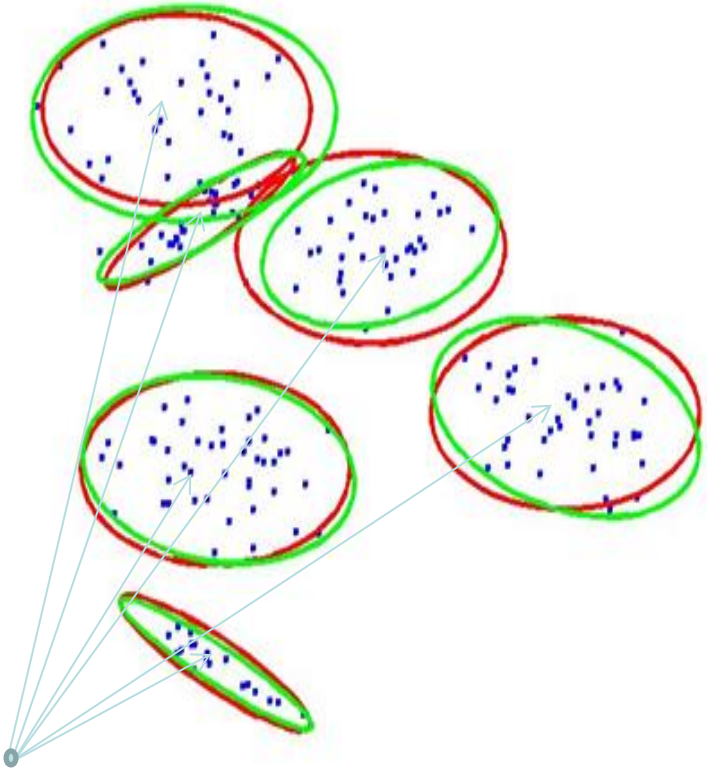
动态时间折叠



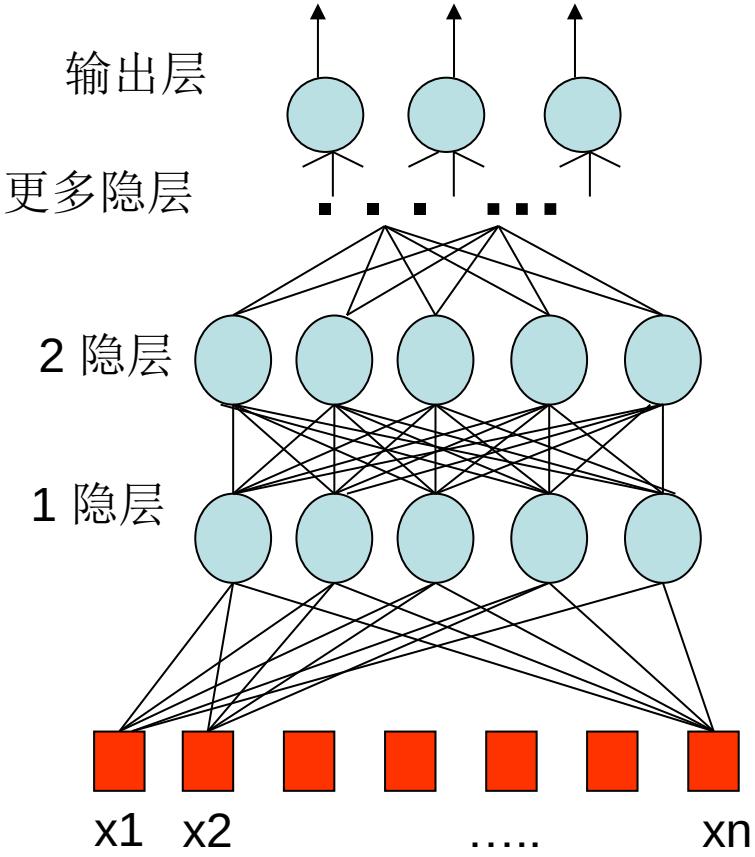
混合高斯模型

$$p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i \, g(\mathbf{x}|\mu_i, \Sigma_i)$$

混合高斯模型和ＤＮＮ模型



混合高斯模型



深度神经网络模型

$$P(O|S)=\sum_{i=1}^{M}\varpi_i N(O|\mu_i,\Sigma_i)$$

$$P(O|S)=P(S|O)/P(S)$$

$$P(S|O)=\exp(O_S^{L-1})\Big/\sum_j \exp(O_j^{L-1})$$

语音识别中的深度神经网络

1. 在1000小时数据的训练上，相对于mpe和fmpe的区分度系统而言，实现了相对误识别率的降低超过20%。

2. 训练后的DNN网络最终是稀疏的，经过优化后，可以适应CPU的线上服务要求。最后的概率计算打分在12核12线的条下满足时时解码要求。

3. 有希望克服了SGD缓慢训练的问题，使用异步混乱梯度法或者是基于二阶优化信息的优化算法，有希望实现了DNN的并行海量数据训练，解决了DNN训练时间过长的难题。

4. **DNN在百度已经取代了GMM！**

Baidu 百度

语音、图像中使用的深度神经网络

1. 多层结构，输入特征500维，每层2000 – 2500个节点，输出层10000维，全连接。
2. 海量训练样本，10亿样本训练级别 。如果考虑抗噪，会到百亿 训练样本级别。
3. 传统训练方法： stochastic gradient decent (SGD)

   要点：混乱随机、mini-bache、pre-training
4. **语音实践使用的难题：** 多核并行计算时候的在线解码计算量
5. **语音实践使用的难题：**目前的SGD训练，工业界中支持训练样本数目为10亿量级，谷歌使用4个GPU，20亿训练样本，训练半年出一个模型。

Baidu 百度

基于二阶统计信息的DNN训练

Deep learning: 一个拥有千万个未知参数的数学优化问题。

$$P^L_{\mathbf{s}|\mathbf{v}}(s|v^L) = \frac{e^{z^L_s(v^L)}}{\sum_{s'} e^{z^L_{s'}(v^L)}} = \mathrm{softmax}_s(z^L(v^L))$$

Hassian-Free Deep Learning:

(1) 神经网络的输出损失函数（交互熵，最小二乘，**softmax)** 是凸函数

(2) 采用高斯牛顿法近似整个神经网络的损失函数

$$H = \frac{\partial}{\partial \vec{w}}(J_{\mathcal{L}\circ M}J_{\mathcal{N}}) = J'_{\mathcal{N}}H_{\mathcal{L}\circ M}J_{\mathcal{N}} + \sum_{i=1}^{0}(J_{\mathcal{L}\circ M})_i H_{\mathcal{N}_i},$$

高斯牛顿法的二阶矩阵G

(3) 由于G是正定的，因此构建下面的二阶辅助目标函数

$$\text{Let } q_\theta(\mathrm{d}) = \nabla\mathcal{L}(\theta)^T\mathrm{d} + \tfrac{1}{2}\mathrm{d}^T(\mathbf{G}(\theta) + \lambda\mathbf{I})\mathrm{d}$$

(4) 共轭梯度法 优化二阶辅助目标函数

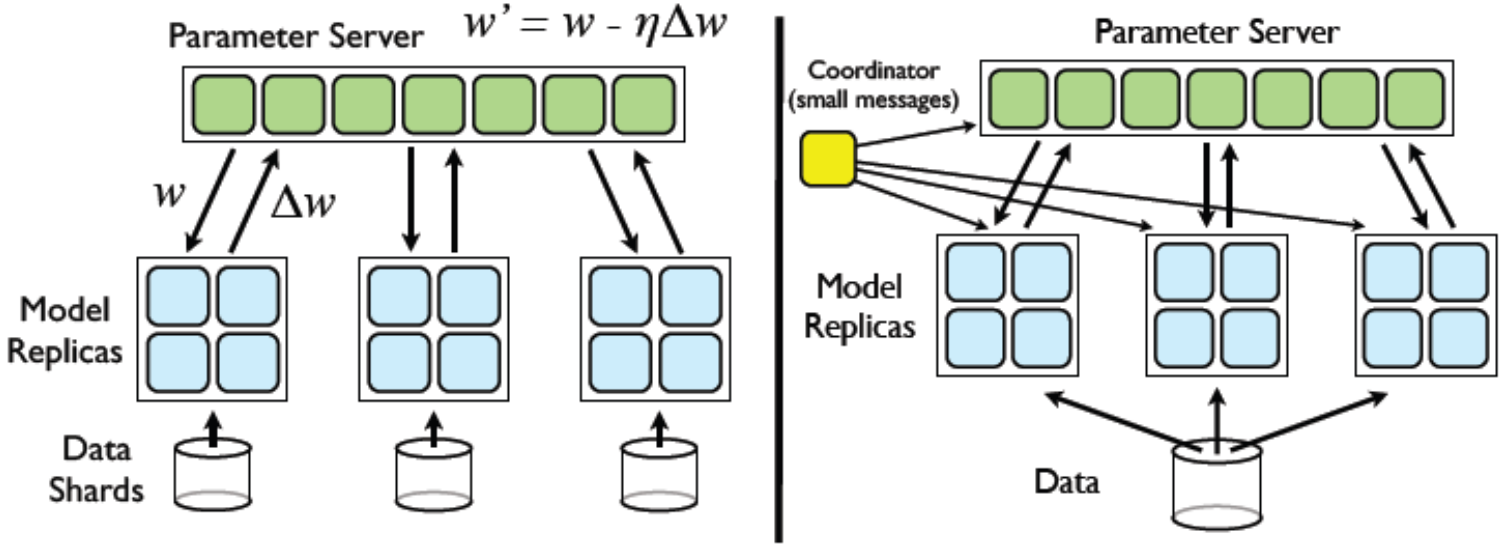
(5) 核心**Trick1:** **Gd**

核心**Trick2:** **Mini-bache** 高斯牛顿估计

核心**Trick3:** **Back-tracing**

Baidu 百度

异步SGD训练和LBFGS



Down-pure SGD (on-line method)

1. Robust to computer failure
2. Possible sub-set model parameter sharing
3. Introduce more stochasticity
4. Asynchronous model update

LBFGS Bache mode

1. Much less bandwidth requirement
2. Bache mode learning
3. LBFG not suitable for large-scale network training ???

Baidu 百度

DNN未来技术展望

各种应用

各种应用： 语音识别， 图像识别。。。

技术核心

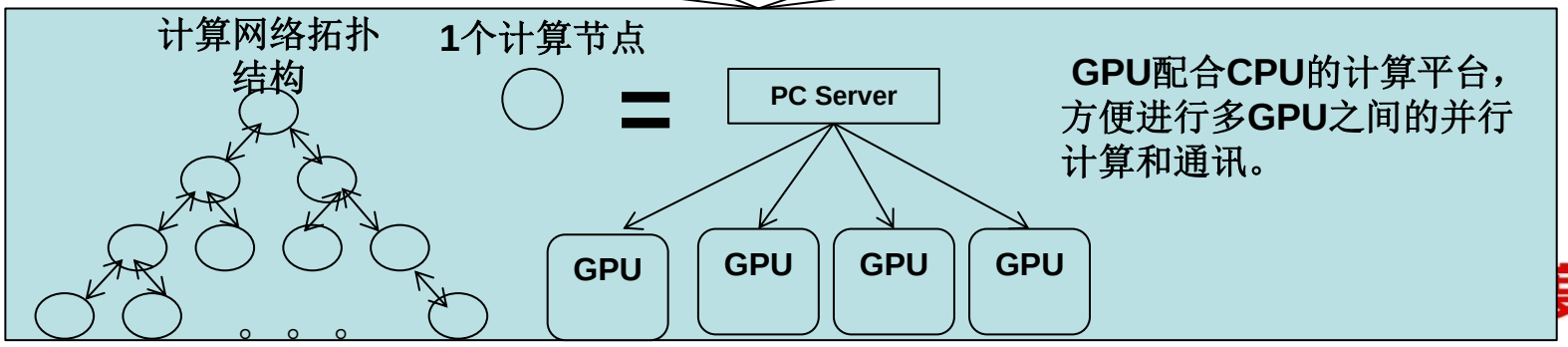
海量数据的获取和生成

DNN训练技术
1. SGD
2. 利用近似二阶梯度信息

输入数据的特征抽取优化(CNN…)

区分度DNN:优化目标和实际问题的结合(MPE, BMMIE)

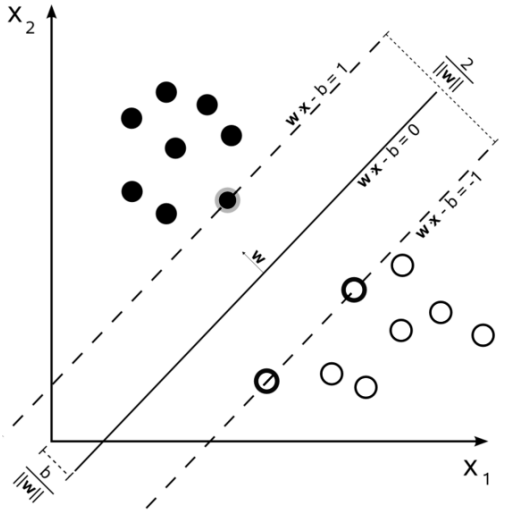
计算平台(Deep Brain)



深度神经网络解决非稠密问题

1. **Deep Neural Network**作为一种强力的模式识别分类工具，广泛成功的应用于语音、图像等多媒体领域。而在文本分类领域，其应用并未获得压倒性的优势。

2. **SVM**是高维小样本分类的经典
   - 最大分类边界
   - 原问题和对偶问题的差异
   - 核函数
   - 凸优化



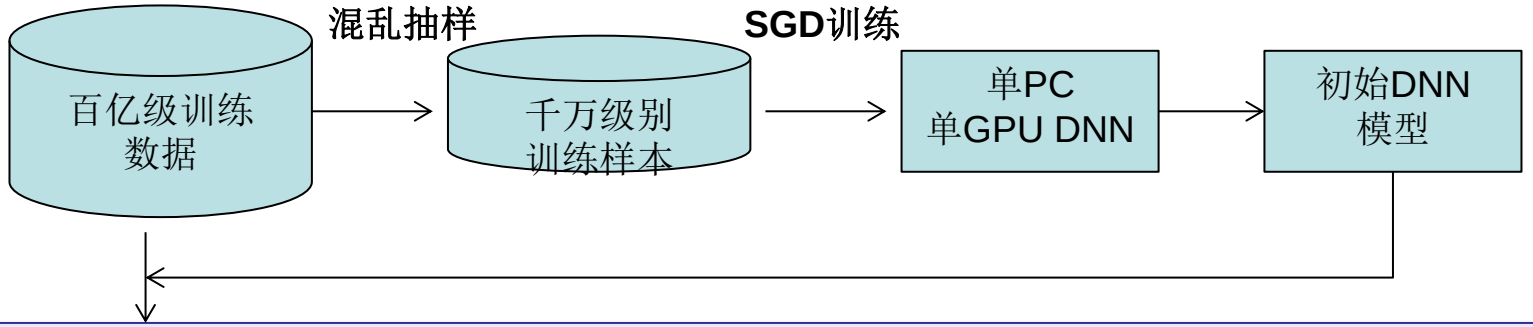
3. **Deep Neural Network： why deep?**

4. **Deep learning for non-sparse problem**
   (learn from SVM…)

Baidu 百度

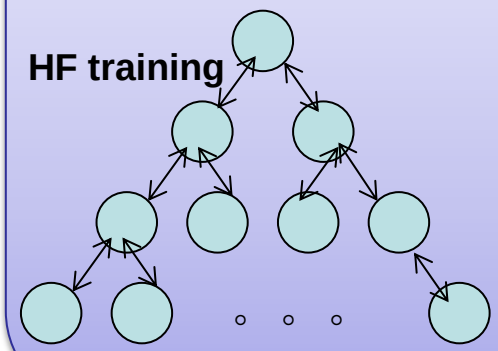
Deep Brain

1. 技术的平台建设：
   (1) 希望彻底解决DNN训练的时间过长的技术瓶颈
   (2) 希望彻底解决DNN训练的网络结构和权重共同学习问题
2. 平台构想

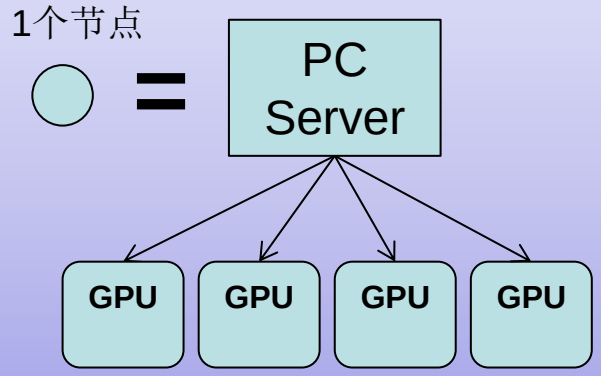


Deep Brain

计算网络拓扑结构



节点结构



1. 进行网络参数和网络结构学习。
2. 可用于语音、图像等多媒体任务中的神经网络训练
3. 可用于任何机器学习场合。

招聘信息（多名）

1. 语音合成
2. 嵌入式语音识别
3. 大规模机器学习
4. 语音识别抗噪技术
5. 解码技术
6. 语言模型技术
7. 语法规则自动学习技术

Baidu 百度

结束...

谢谢大家

Baidu 百度