

大规模深度学习算法 Deep Belief Network及其应用

刘凯 (花名：拜阳)

共享业务事业部-数据应用

2015-06-27



自我介绍

- 2006-2011年毕业于中科院自动化所，模式识别与智能系统
- 2011-2014年在IBM中国研究院任研究员，大数据挖掘与管理
- 2014年加入阿里，从事分布式机器学习算法研发，以及在阿里的各种业务场景推广应用

Outline

- 为什么要用深度学习
- 分布式DBN架构和原理
- 典型应用
 - 淘宝广告评论检测
 - 手淘天天特价个性化推荐
- 总结与未来工作

Outline

- 为什么要用深度学习
- 分布式DBN架构和原理
- 典型应用
 - 淘宝广告评论检测
 - 手淘天天特价个性化推荐
- 总结与未来工作

机器学习在互联网界的典型应用场景

- 为什么要机器学习
- 为什么需要更好的机器学习模型
- 为什么要分布式学习

个性化推荐

天天抢好货 吐血推荐

买疯了@买一送一... ¥69.00 ¥668.00	夏棉麻9分裤小脚亚... ¥37.00 ¥468.00	超显瘦胖瘦MM都... ¥55.00 ¥499.00
-------------------------------	--------------------------------	-------------------------------

广告CTR prediction

广告库 $D = \{ad_1, ad_2, \dots, ad_i\}$

广告请求 $q^{(i)}$

CTR 预估系统

广告的预估CTR

$Pctr_ad_1^{(i)}$

$Pctr_ad_2^{(i)}$

...

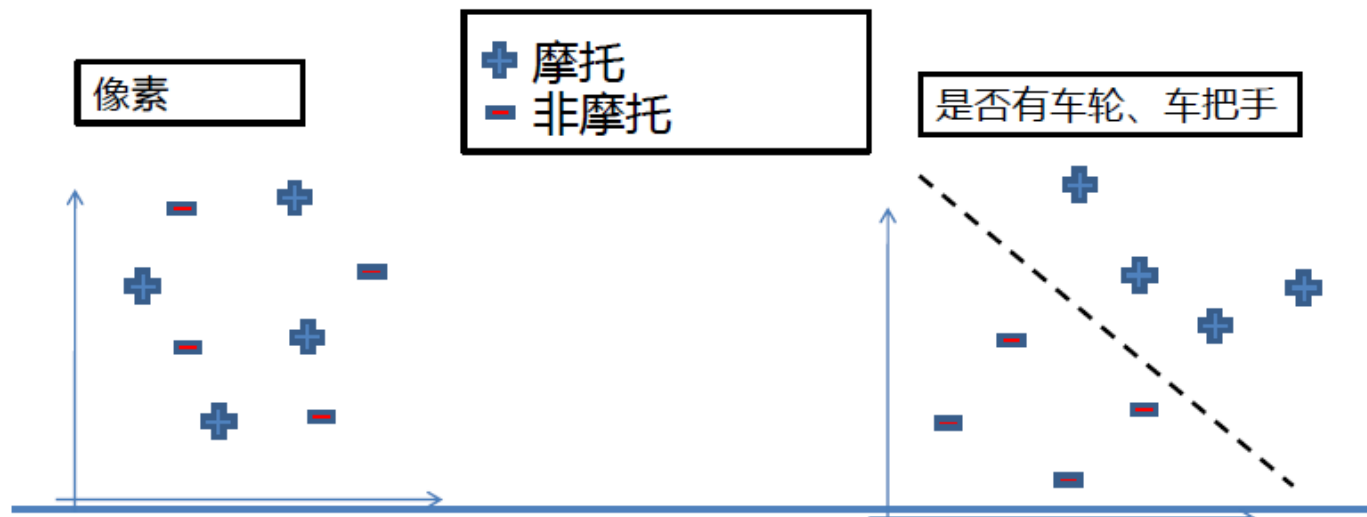
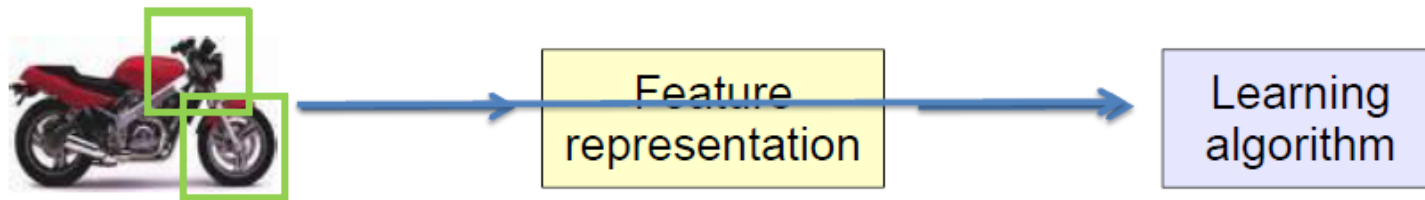
$Pctr_ad_n^{(i)}$

- ◆ 从海量用户行为数据中学习用户的行为偏好，为每个用户提供个性化的服务
- ◆ 需要更精确的分布式机器学习模型高效学习含有丰富用户行为的大数据

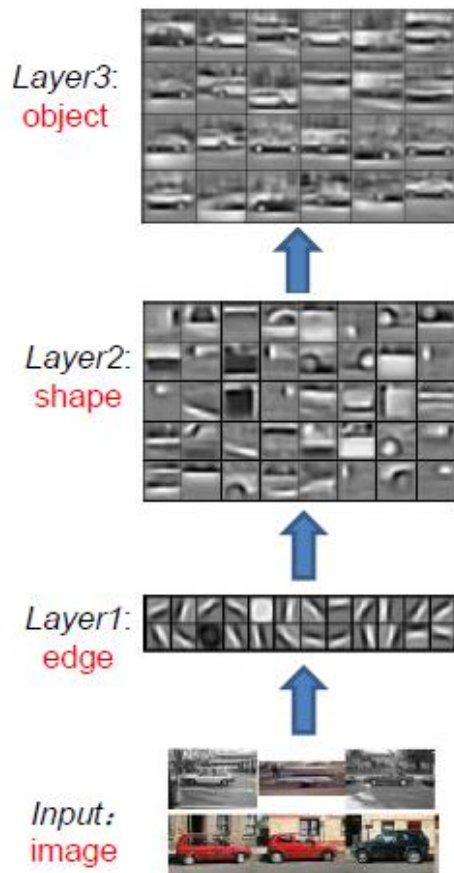
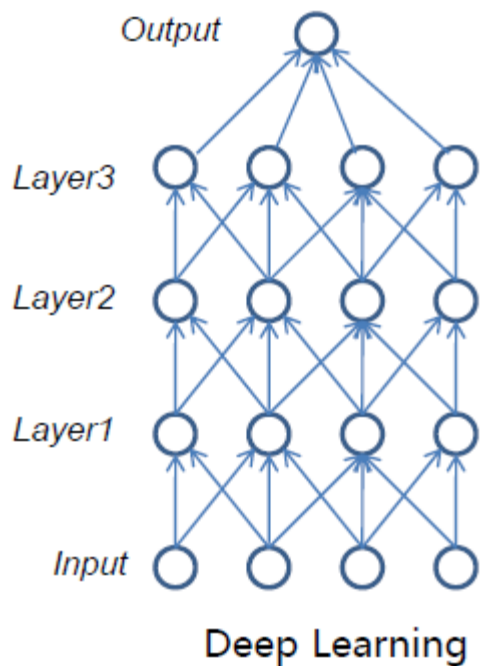
深度学习模型具有更强的特征表示能力

- 线性模型：如果输入特征线性不可分，则精度较差(如逻辑回归)

□ 特征表示

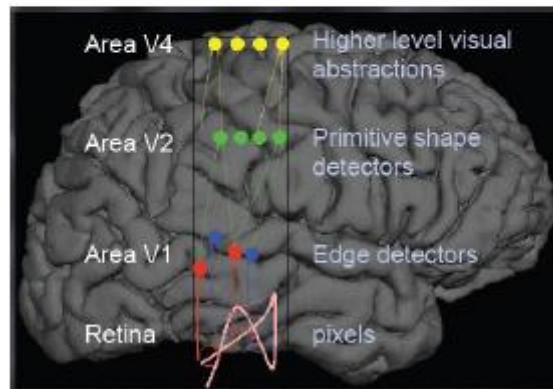


深度学习：结构决定一切



Deep Learning
输入、中间结果

Lee, Grosse, Ranganath, Ng.
ICML 09.

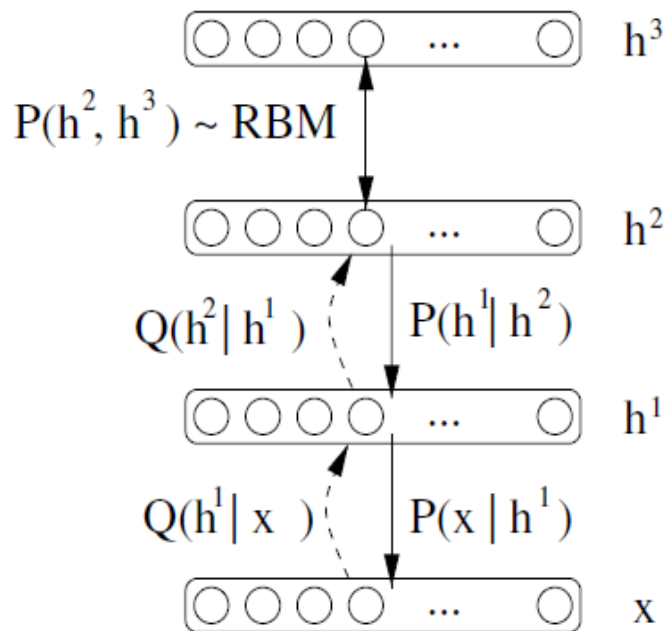


人类视觉

Deep Belief Network(DBN)的优势

一般的DNN中，随着网络深度增加，在迭代中容易掉入到某个局部极小点而导致训练失败

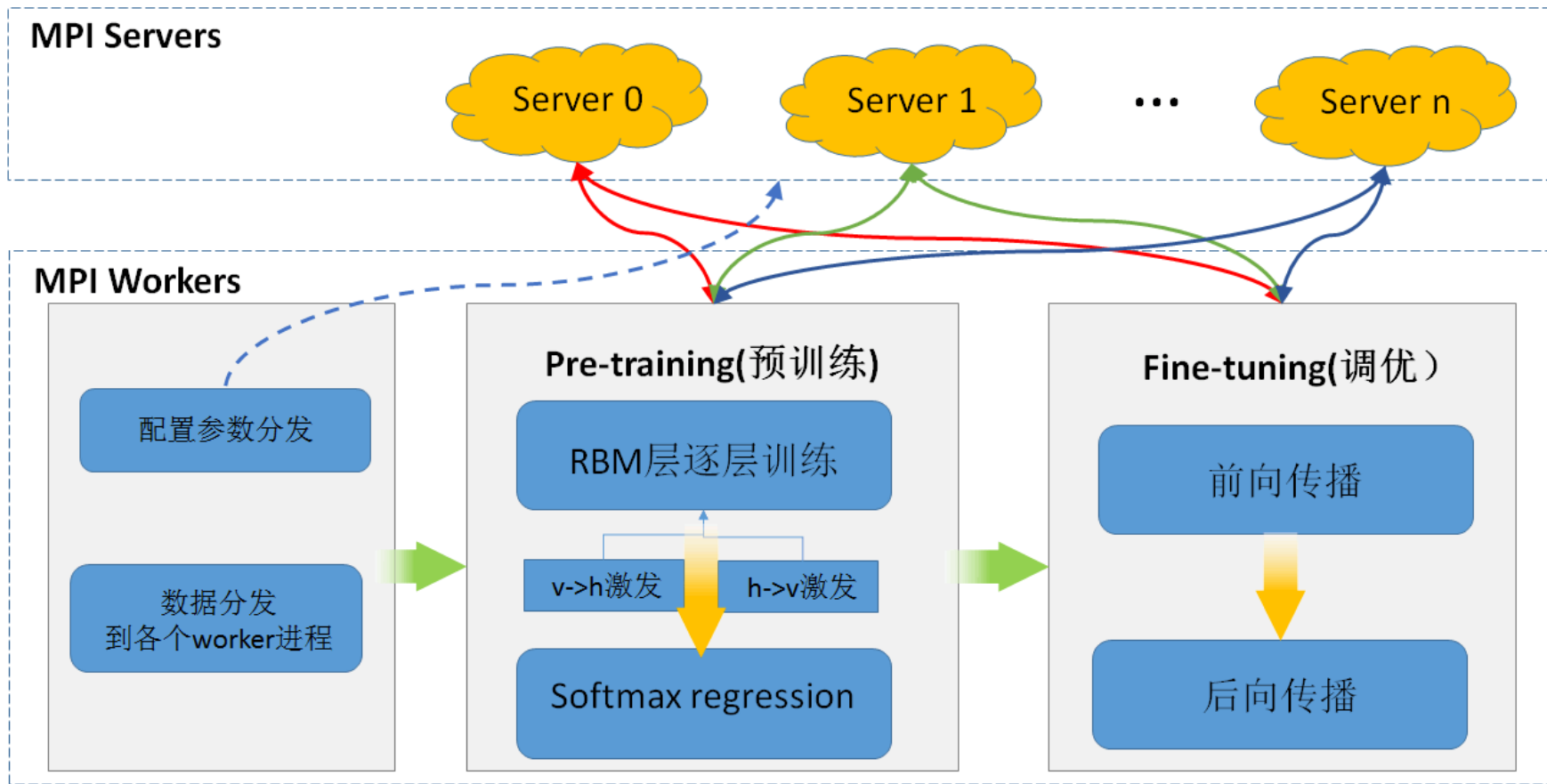
- DBN: 多个RBM逐层堆积，逐层预训练所组成的多层神经网络
- DBN的优势：在BP有监督训练之前，对网络参数进行无监督的预训练
- 预训练可将网络参数初始化到靠近某个较优的局部极小点甚至全局最优点附近



Outline

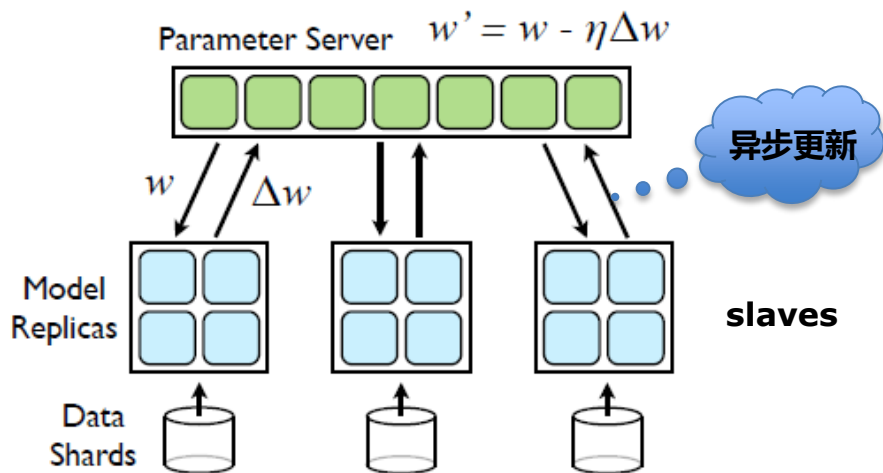
- 为什么要用深度学习
- 分布式DBN架构和原理
- 典型应用
 - 淘宝广告评论检测
 - 手淘天天特价个性化推荐
- 总结与未来工作

分布式DBN架构



RBM: 限制玻尔兹曼机 v->h: 从可见层到隐藏层 h->v: 从隐藏层到可见层

迭代、参数更新模式

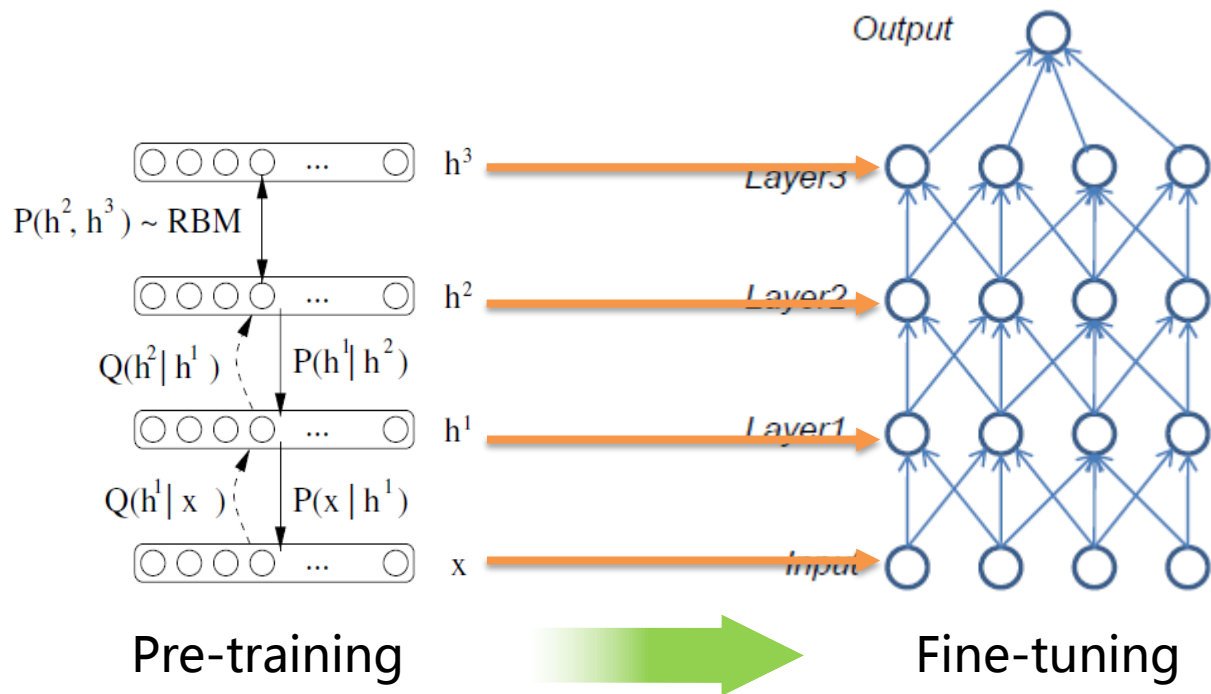


- 采用异步随机梯度下降技术
- 每个worker之间不需要统一步调

- Master为存储网络参数的server进程
 - 每个server作为一个parameter server，管理着一层网络的参数
 - 它与各个worker进程通信以完成参数梯度的更新和网络最新参数的分发
- Slave为模型训练的工作进程，它们是主要的算法执行者
 - 对数据进行切分，在各个slave的训练都是基于这一部分数据进行
 - 采用mini-batch方法迭代，加快收敛速度

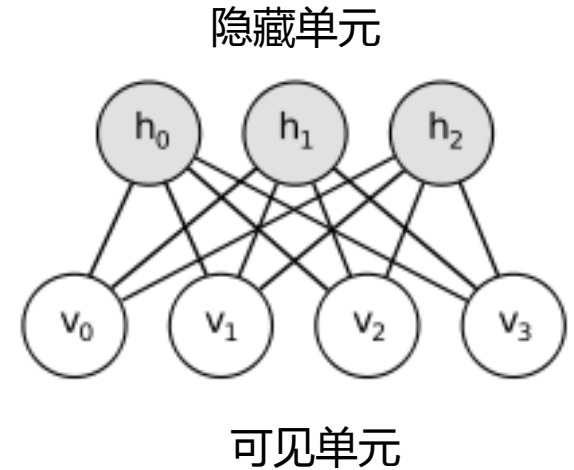
核心模块

- pre-training : 通过逐层训练对网络参数进行初始化, 使得模型比较靠近某个较优局部最优点甚至全局最优区域
- fine-tuning : 基于pre-training模块得到的初始化网络, 进行自底向上的前向传播、自顶向下的后向传播的多轮有监督训练



Restricted Boltzmann Machines (RBM)

- Restrict the connectivity to make learning easier
 - Only one layer of hidden units
 - No connections between hidden units or visible units
- In a RBM, the hidden units are conditionally independent given the visible states



RBM代价函数的梯度计算：Gibbs采样

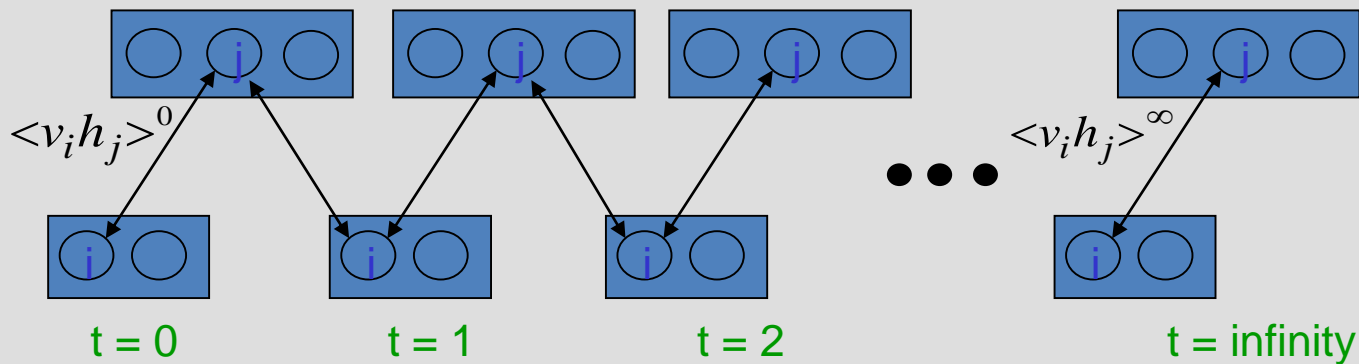
- Log-likelihood代价函数的梯度计算：

$$\frac{\partial(-\log p(v))}{\partial \theta} = \frac{\partial(-\sum_h \log p(v))}{\partial \theta} = E\left[\frac{\partial(-energy(v, h))}{\partial \theta} \mid v\right] - E\left[\frac{\partial(-energy(v, h))}{\partial \theta}\right]$$

- 直接计算梯度计算复杂度很大

- 将原来在 (v, h) 上求期望转变为使用Gibbs采样的操作

$$\frac{\partial \log p(v)}{\partial w_{ij}} = \langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^\infty$$

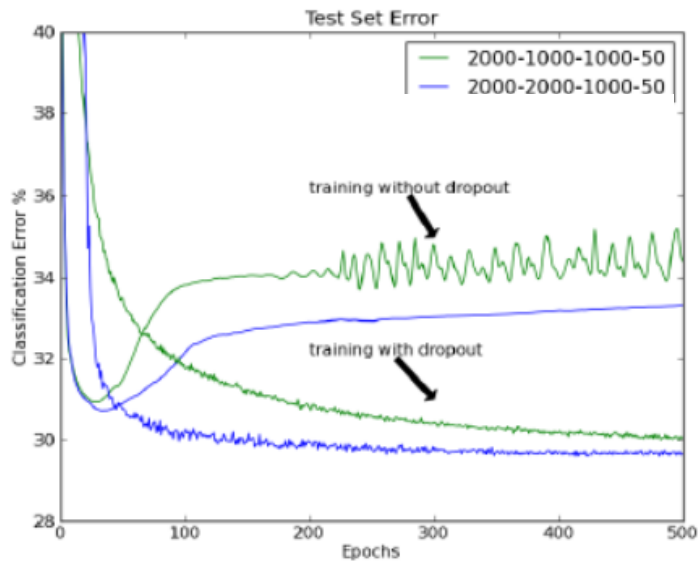
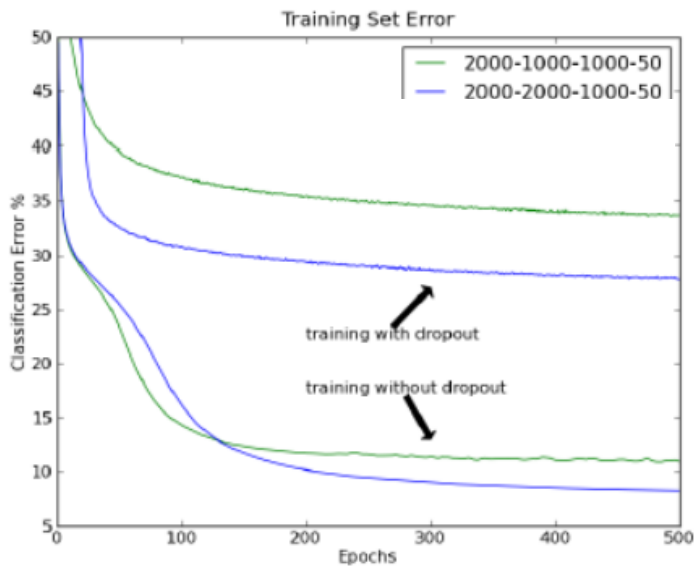


Supervised Fine-tuning

- 使用有标签的训练数据继续训练模型
- 采用一般的梯度后向传播方法：
 - forward propagation
 - backward propagation
 - weight updating

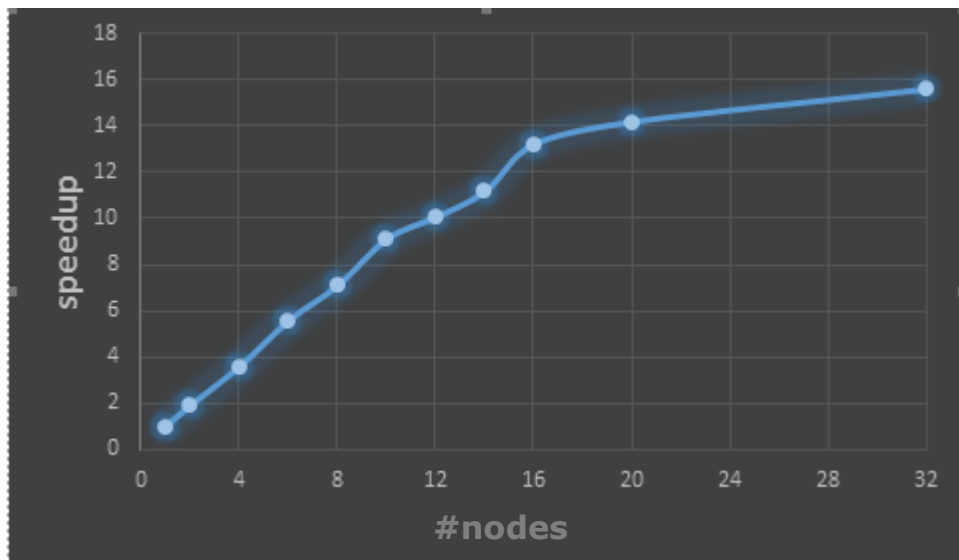
Dropout策略避免overfitting

- 若在进行模型训练时训练数据集并不够大，从而可能训练过拟合而导致在测试数据集上表现较差
- Dropout策略：在模型训练时对于每一个训练样本随机让某些隐藏层节点的权重不工作，不工作的那些节点可以暂时认为不是网络结构的一部分
- 它的权重需保留下来（只是暂时不更新），下个样本输入时它可能又得工作



并行化的可扩展性

- speedup基本上随着节点的增加呈现线性提速
- 处理样本的数量受限于各个计算节点的内存大小
- 每个worker节点的空间复杂度为 $O(N*(F+M)/p)$ ，其中N为sample个数，F为特征维数，M为多层神经网络的神经元总数，p为进程数

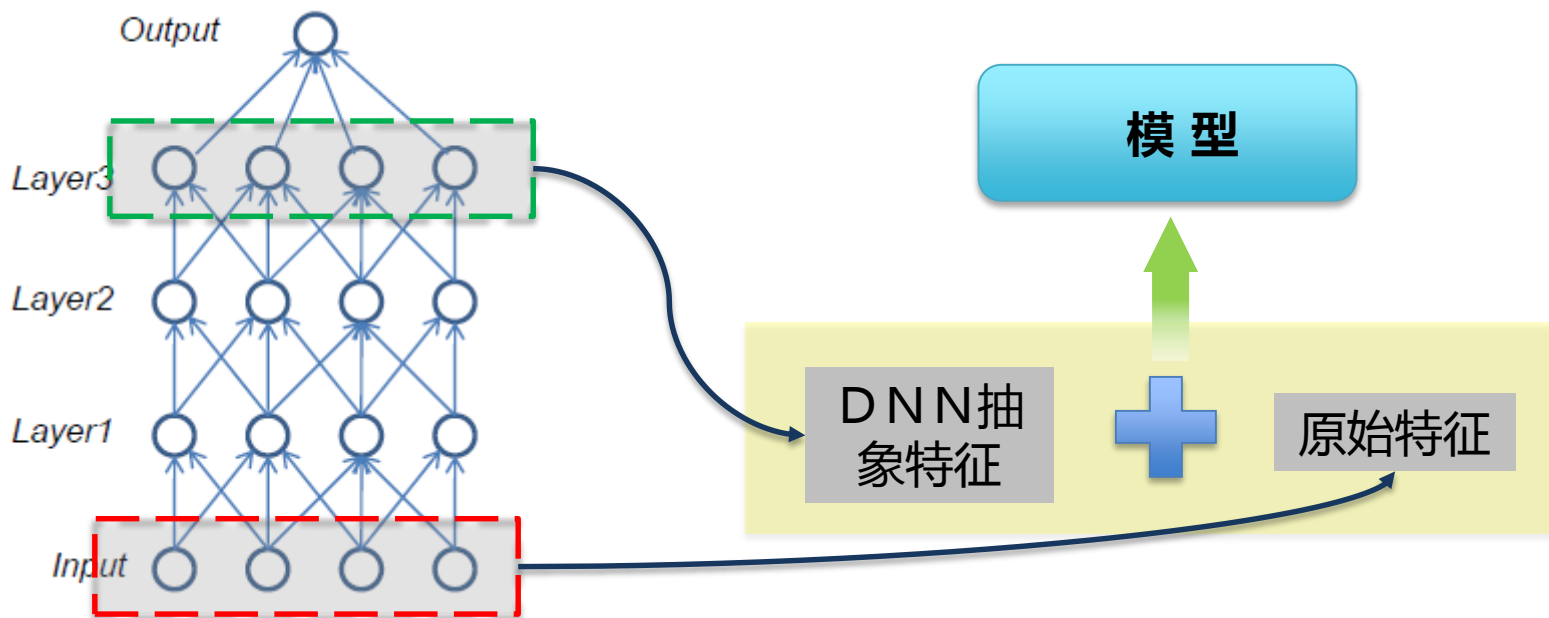


不同节点数的加速比

- 训练数据、验证数据完全相同
- 总量计算复杂度相同
- Mini_batch_size=10
- 每10次迭代worker与server数据交互一次

使用DNN抽象特征

- 充分利用深度模型对于特征的抽象表征能力
- 模型融合：使用最后一层隐藏层输出的高维抽象特征，与原始特征连接成一个新样本，训练其他模型



Outline

- 为什么要用深度学习
- 分布式DBN架构和原理
- 典型应用
 - 淘宝广告评论检测
 - 手淘天天特价个性化推荐
- 总结与未来工作

淘宝广告评论检测

- 淘宝详情页评论内容有不少是各种形式的广告类评论
- 每天新增评论千万级，广告十万级，人工删除基本不可能，亟需借助机器学习自动化地屏蔽、删除广告评论
- 对模型要求高：既要准确召回广告评论，又要尽量降低假阳性的比例

淘宝广告评论检测

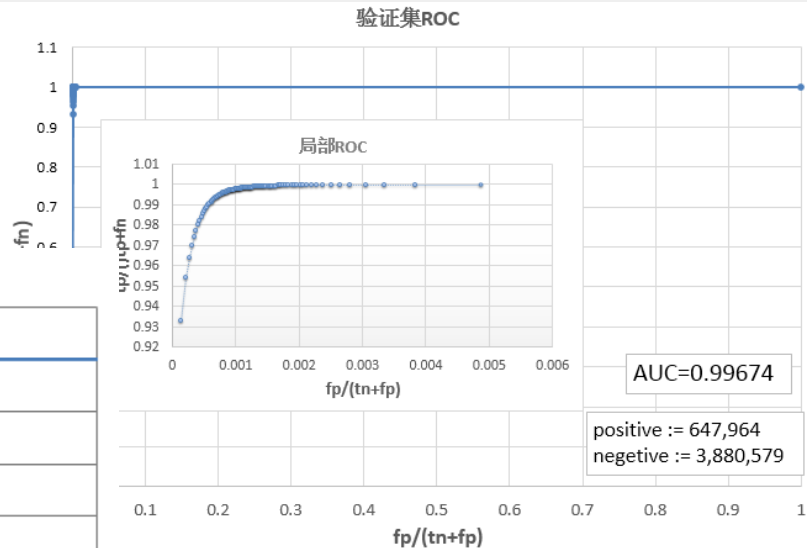
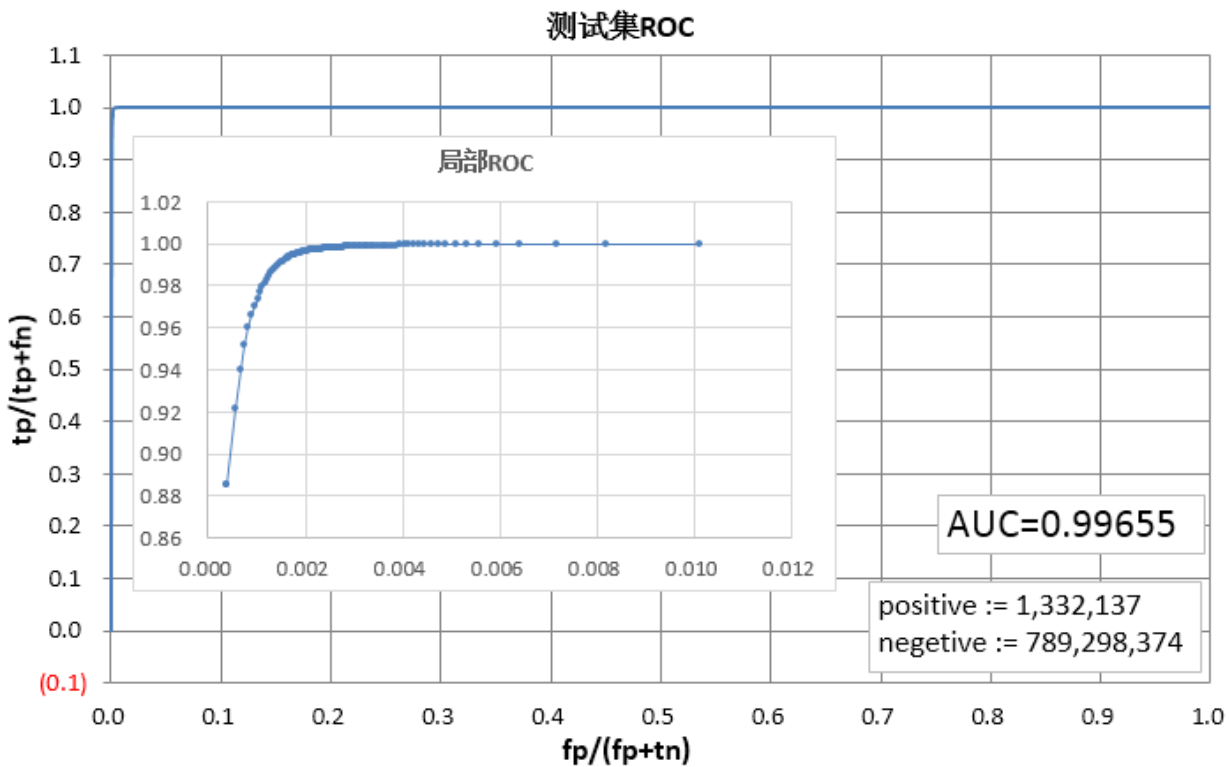
- 67维特征，包含用户购买相关特征、评论行为与语义特征、用户属性特征等
- 训练样本数据约为1800万条，正负样本比例为1:6
- 测试样本数据约为8亿条，正负样本比例为1:593

整体 accuracy	0.99763
正样本recall	0.99765
正样本precision	0.41537
负样本recall	0.99763
负样本precision	0.999996

阈值为默认
0.5时的指标

- 输出的阳性数据中，保留了绝大部分的正样本，正负样本的比例调整为1:1.5
- 如果对oversample负样本数量，同时复制正样本几次，正样本precision可进一步提升

淘宝广告评论检测——ROC&AUC



(0.1)

淘宝广告评论检测——模型融合

- 从训练好的DBN模型最后一层隐藏层输出200个高维抽象特征，与原始特征连接，一起作为GBDT模型的训练样本
- 从不同模型预测precision上看，DBN高维特征可以生成有效模型

Features	Precision	Recall threshold
基础特征	98.31%	0.985
基础特征+DBN特征	98.65%	0.996
基础特征+最大熵特征	98.14%	0.992

手淘天天特价个性化推荐

- 移动端屏幕尺寸更小，用户花在手淘上的时间更加碎片化，因此更需要精准地每个用户推荐最有可能喜好的商品
- 尝试使用DBN为每个用户匹配的选品候选集中的商品排序



给某女士的推荐结果

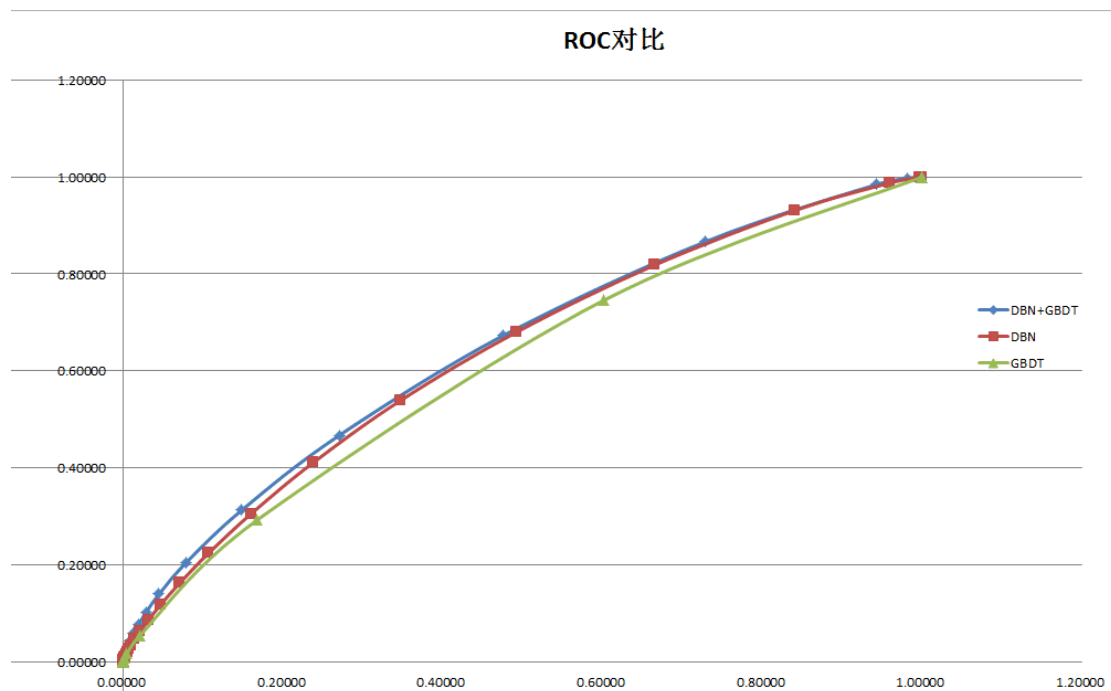


给本人的推荐结果

手淘天天特价个性化推荐——ROC

- 训练数据集为5400w，正负样本比例约为1:18
- 使用卖家属性特征、商品统计特征、卖家属性和统计特征、买家到类目的统计特征、买家对品牌和卖家的偏好特征等62维特征
- ctr为20%，从浏览到购买的cvr为0.45%

算法	特征	auc
gdbt	原始特征	0.6140
dbn	原始特征	0.6320
gdbt	dbn高维特征 +原始特征	0.6431



Outline

- 为什么要用深度学习
- 分布式DBN架构和原理
- 典型应用
 - 淘宝广告评论检测
 - 手淘天天特价个性化推荐
- 总结与未来工作

并行化DBN

- **采用ParaServer-Worker的并行架构**，将网络的参数存储与模型训练分离，各个worker分别异步地与server通信，在保证训练收敛前提下显著提高模型训练的效率
- **采用pre-training+fine-tuning的方式进行并行化训练**，pre-training先对网络参数进行初始化，再使用fine-tuning进行模型调优；
- **dbn可特殊化为一般的dnn**：忽略pre-training（将该阶段的迭代步数设为0），直接进行基于BP的fine-tuning
- **支持多类别(>2)分类**
- **充分利用深度模型对于特征的抽象能力**：可使用最后一层隐藏层输出的高维抽象特征，与原始特征连接成一个新样本，训练其他模型

未来工作

- 算法优化：由于数据的有效性会随着时间衰减，未来我们将增加算法的增量训练模块，支持对增量数据的快速学习，使得模型能很好地表达最近数据的趋势
- 算法应用推广：在个性化推荐多个场景以及其他业务中推广应用

A scenic landscape featuring a paved road with a dashed yellow center line that recedes into the distance. The road is flanked by a vast, flat, arid plain. In the background, there are rolling, low mountains under a bright blue sky filled with large, white, fluffy clouds. The overall atmosphere is one of openness and vastness.

THANK YOU

邮箱 : baiyang.lk@alibaba-inc.com
kailiusci@163.com