



DNN在凤巢CTR预估中应用

2014/02

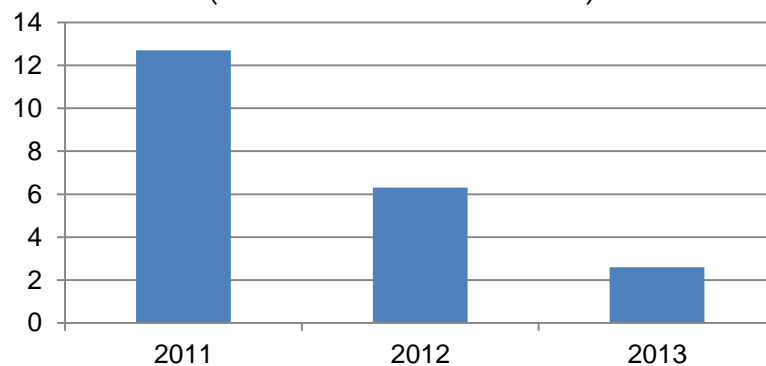
陈雨强@ECOM

离散LR模型遇到瓶颈

- 凤巢CTR预估模型为离散值LR

- 规模大，刻画细致
- 靠特征提升模型
- 提升逐年渐难

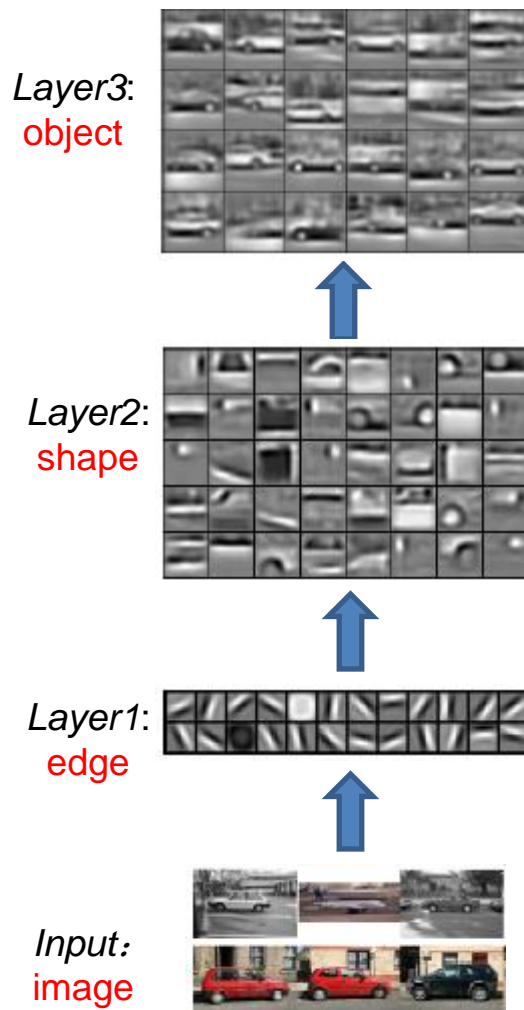
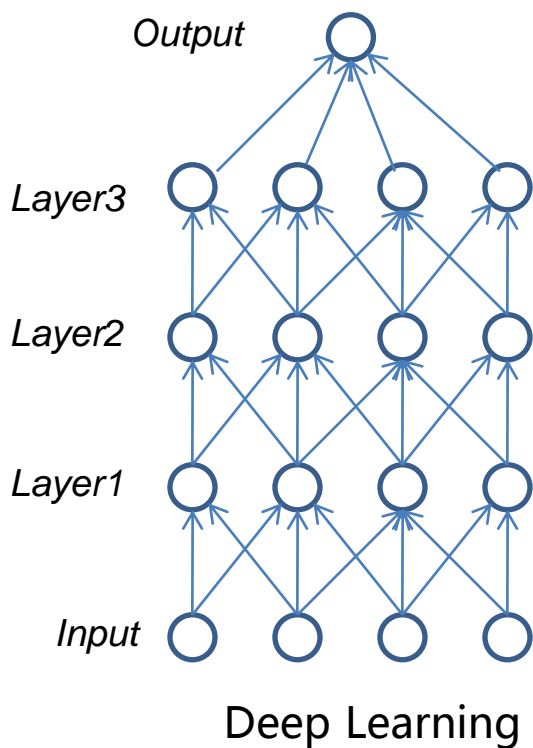
近年特征工程带来的收益
(归一化到CPM1 %)



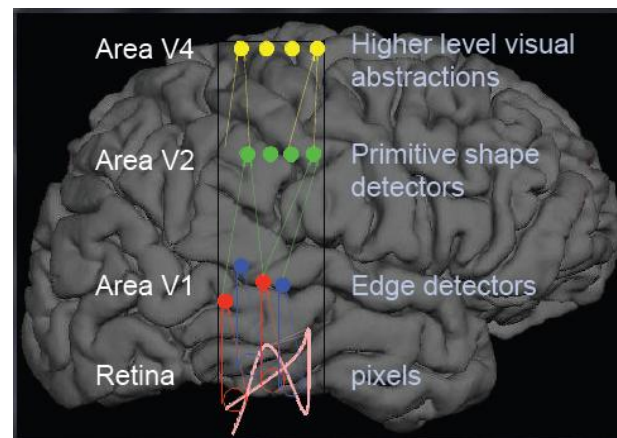
- 业界普遍有两条路

- 大规模特征+线性模型 (Google, Baidu)
- 中小规模特征+非线性模型 (Yahoo, Bing, FB, Taobao)

模型：深度学习——结构决定一切



Lee, Grosse, Ranganath, Ng.
ICML 09.



人类视觉

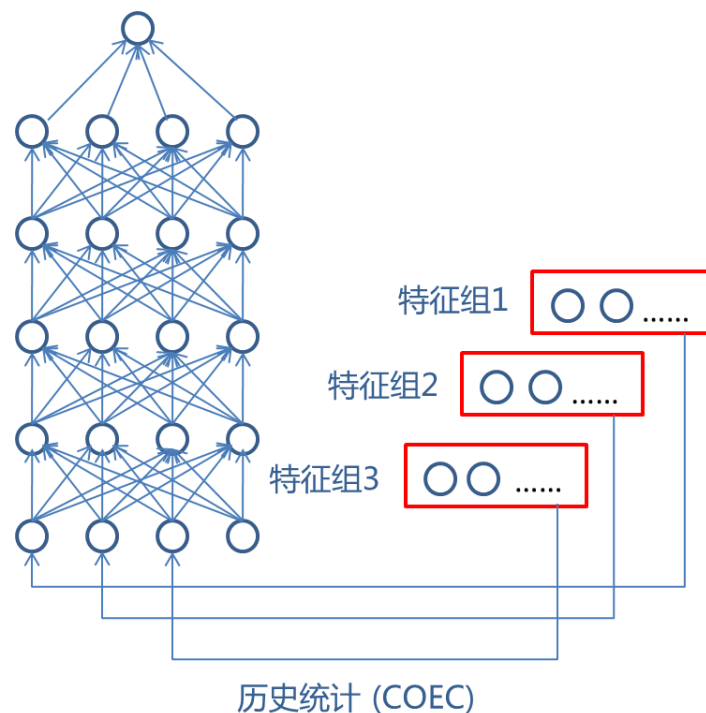
Deep Learning
输入、中间结果

DNN for CTR难点：特征

- "大²数据" → "大数据"
 - 大规模ins × 大规模特征 → 大规模ins × 小规模特征
 - 学术界鲜有合适算法
 - 也并未采用Google的技术方案

- 使用历史统计特征

- 训练数据用单一位置



模型替换之痛

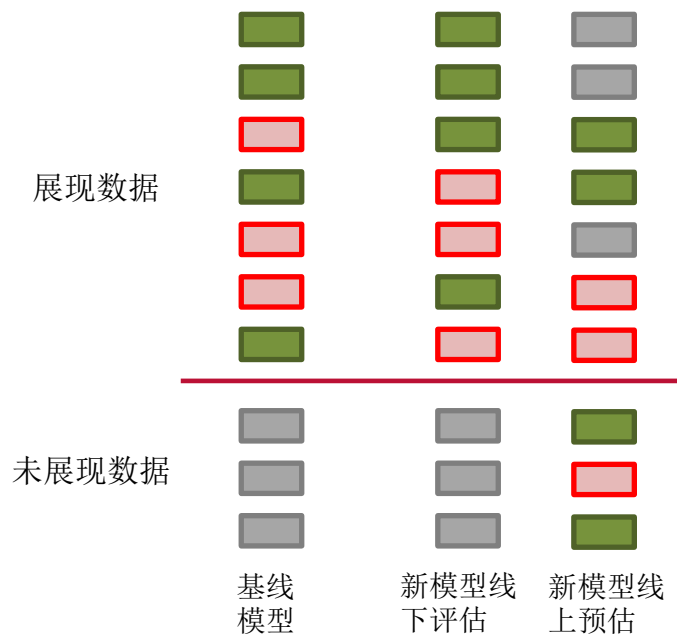
- 实验出现了线上线下一致
 - 修改训练时间窗
 - 线下AUC持平
 - 线上CTR上涨0.5%
 - 修改模型复杂度
 - 线下AUC显著上涨1个百分点
 - 线上CTR下跌1%
 - 在之前优化LR特征的过程中从未出现过

数据分布：一致？不一致？

- 一个看似一致的不一致

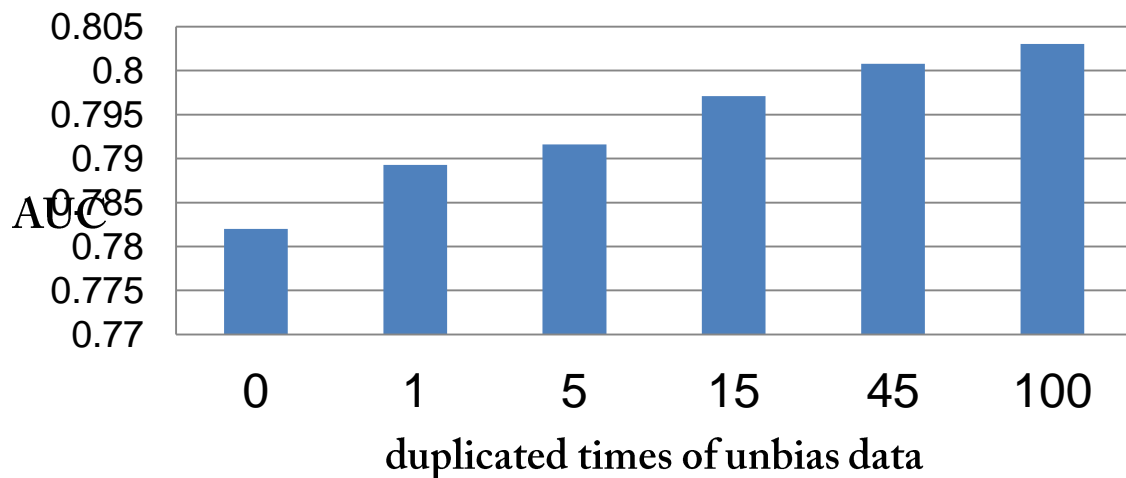
- 线下使用线上展现点击的日志
- 只能评估展现出来的广告的质量
- 我们不知道为未展现广告上算法的好坏，但未展示广告数远大于展示广告数，是模型效果的关键
- 为什么之前LR添加特征没有这个问题

- 变化小
- 参与迭代



解决不一致性：无偏数据

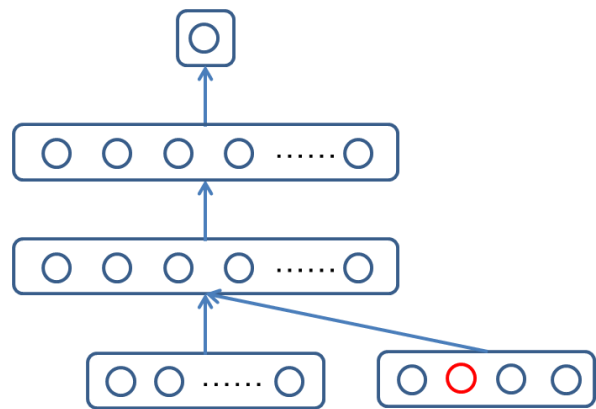
- DNN模型无法直接直接上线参与迭代：可能的损失太大
- 折中：小流量生成无偏数据
 - 不让任何策略作用在广告排序阶段
 - 圈流量随机展示广告
- 使用数据Over Sampling大幅提升无偏效果



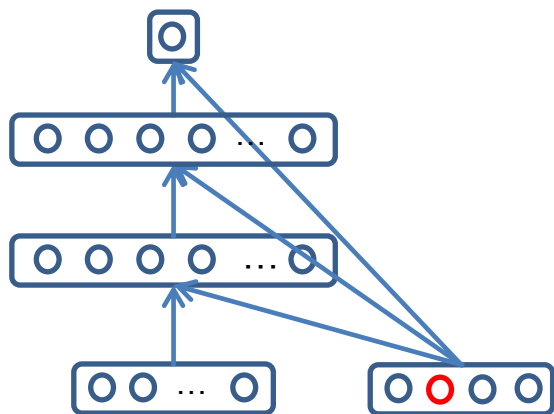
系统优化 – 数据、特征、模型

数据优化

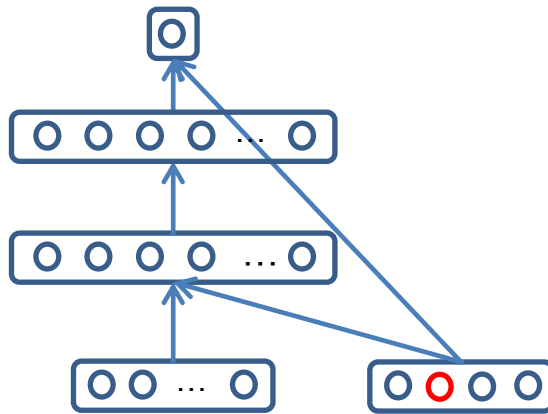
- Position Bias
 - 点击行为差别很大
 - 无法使用所有位置数据
- 类比LR解决
 - 加入位置特征
 - 在深层网络中不可行



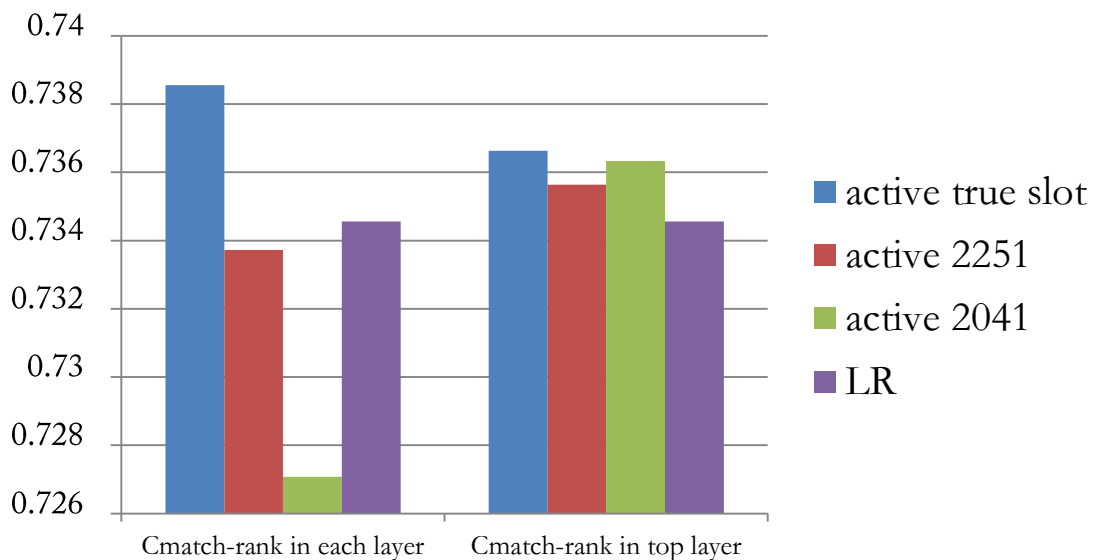
数据优化方法：Positon Bias修正



每一层加入Position bias



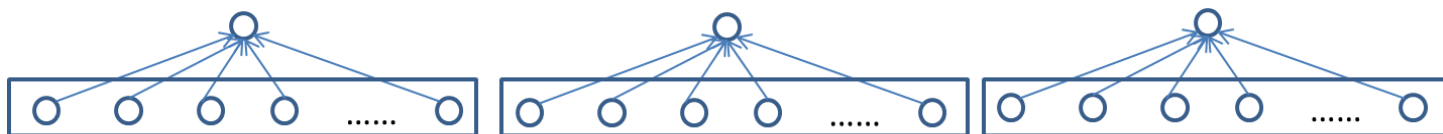
顶层加入Position bias



使用所有位置数据进行训练DNN整体AUC **+1.2**个百分点

特征：使用LR特征

- 历史统计可以被看成单特征LR

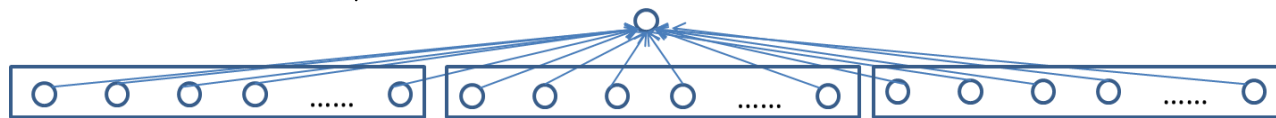


- COEC可以看做单特征+Pos bias LR

- Feature correlation对效果有显著提高

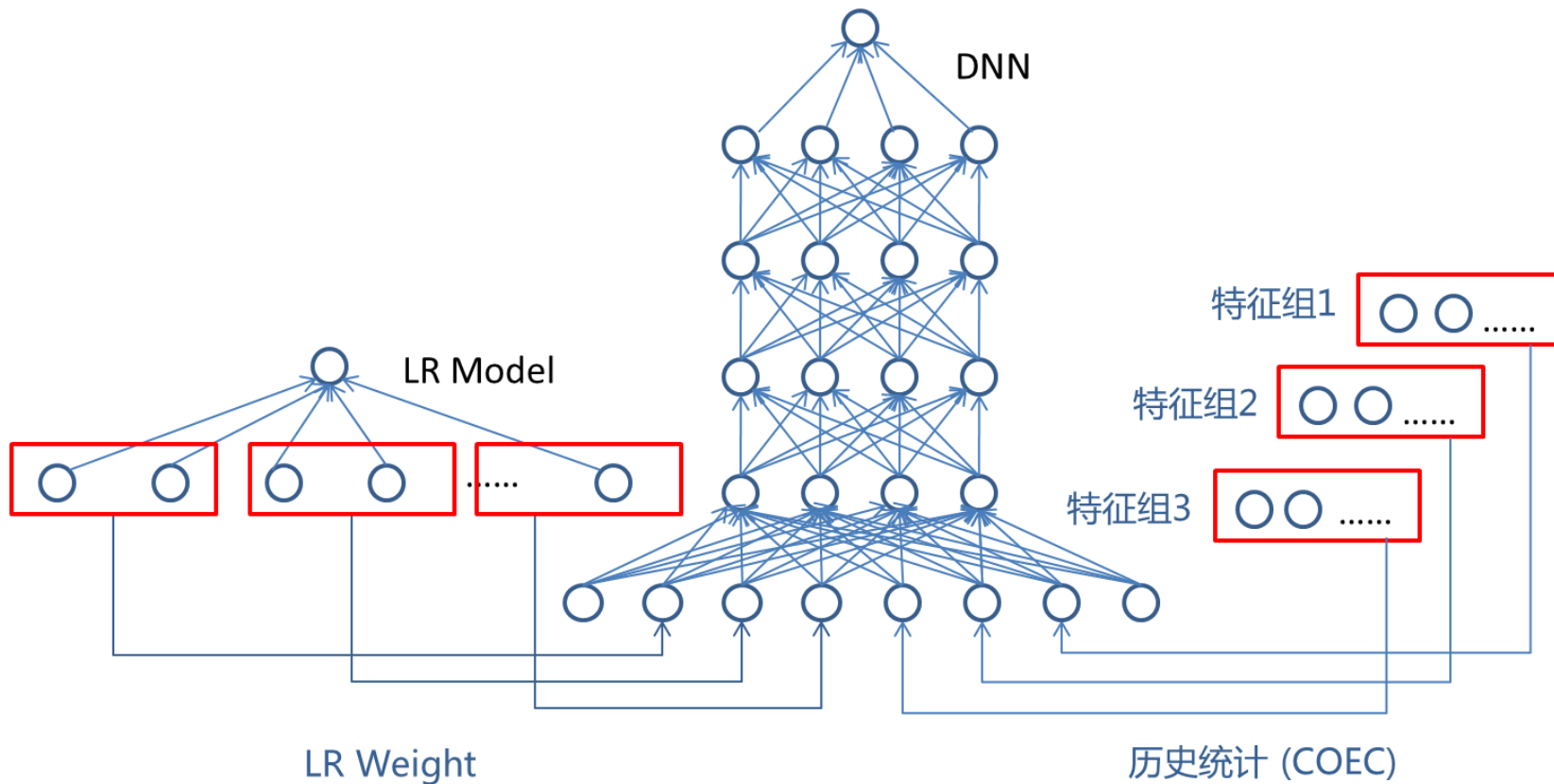
- 更进一步，考虑全特征correlation

- LR权重特征



使用LR Weight作为特征的DNN

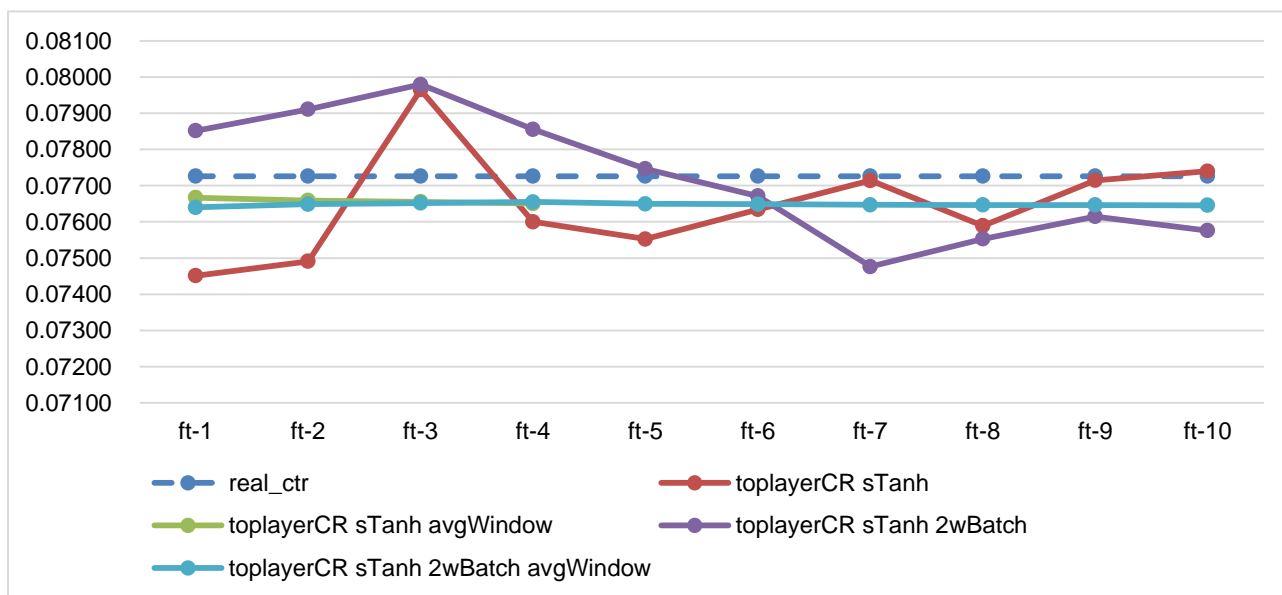
$q = \text{DNN Model Prediction}$



AUC +1个百分点

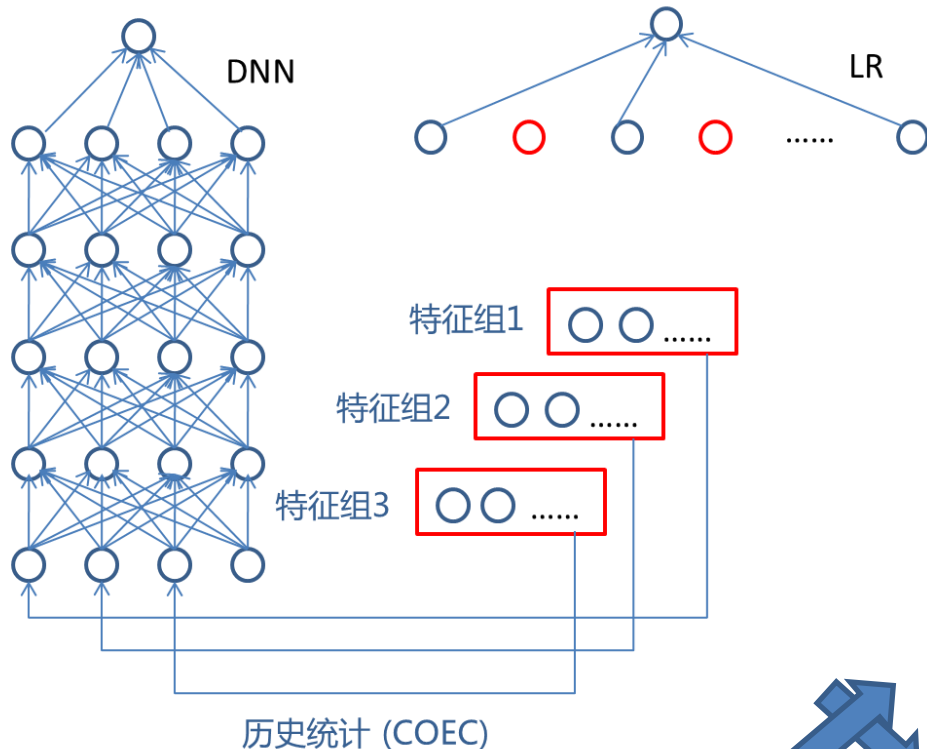
模型

- Stanh, RELU激活
 - 防止梯度消失
 - 脱离预训练，加速学习(大Batch_size)
- Average windows保持预估稳定性



怎么用DNN 小心谨慎，逐步切换

$$q = \lambda * \text{DNN Model Prediction} + (1 - \lambda) * \text{LR Model Prediction}$$



ctr, ctr * bid推左门限

$$Ads_1 = (q_1, bid_1)$$

$$Ads_2 = (q_4, bid_4)$$

$$Ads_5 = (q_5, bid_5)$$

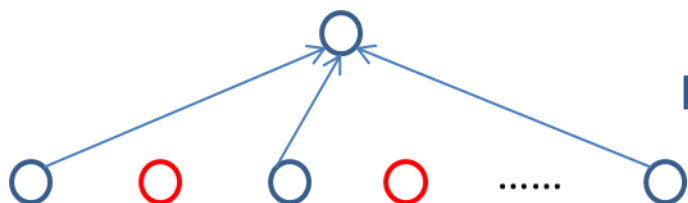
$$Ads_2 = (q_2, bid_2)$$

$$Ads_3 = (q_3, bid_3)$$

DNN作用于凤巢CTR预估三期

深度学习模型首次服务凤巢代

LR Model Prediction



ctr * bid排序

$$CPM_4 = q_4 * bid_4$$

$$CPM_1 = q_1 * bid_1$$

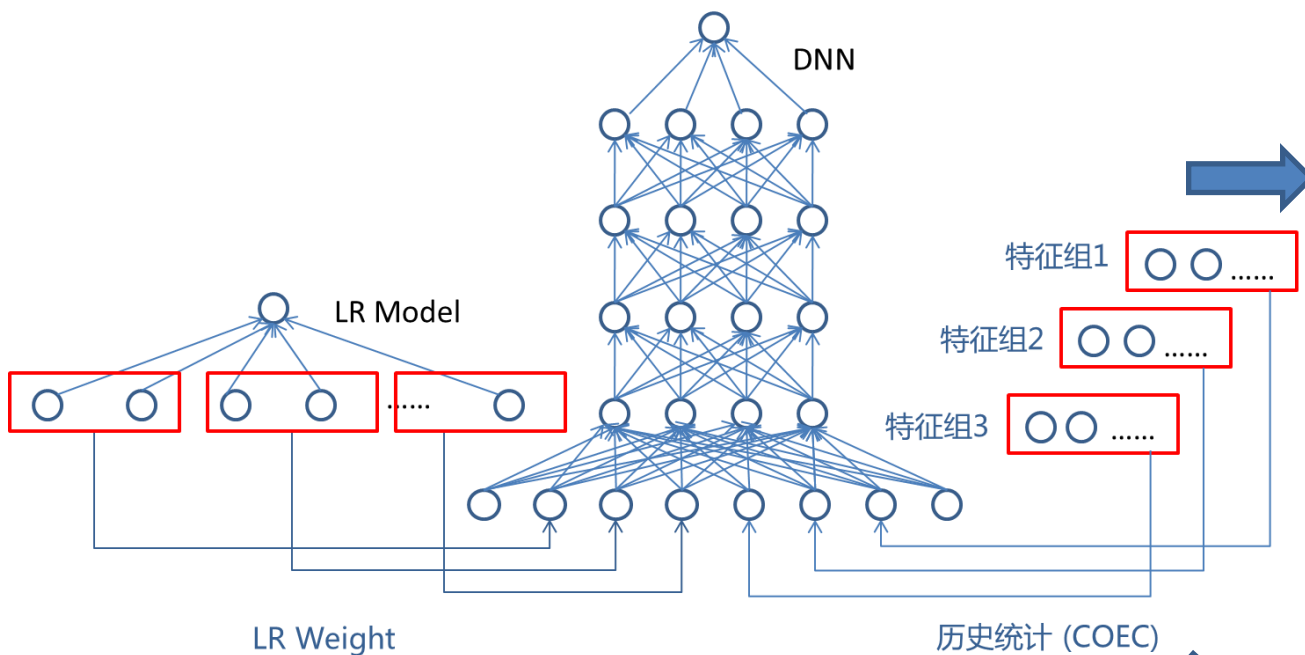
$$CPM_5 = q_5 * bid_5$$

GSP计费

$$price_1 = \frac{q_1 * bid_1}{q_4}$$

$$price_1 = \frac{q_5 * bid_5}{q_1}$$

$q = \text{DNN Model Prediction}$



ctr, ctr * bid推左门限

$$Ads_1 = (q_1, bid_1)$$

$$Ads_2 = (q_4, bid_4)$$

$$Ads_5 = (q_5, bid_5)$$

$$Ads_2 = (q_2, bid_2)$$

$$Ads_3 = (q_3, bid_3)$$

DNN作用于凤巢CTR预估三期

新一代凤巢CTR预估系统的里程碑

ctr * bid排序

$$CPM_4 = q_4 * bid_4$$

$$CPM_1 = q_1 * bid_1$$

$$CPM_5 = q_5 * bid_5$$

GSP计费

$$price_1 = \frac{q_1 * bid_1}{q_4}$$

$$price_1 = \frac{q_5 * bid_5}{q_1}$$

DNN背后 - 持续的工程优化

工程优化

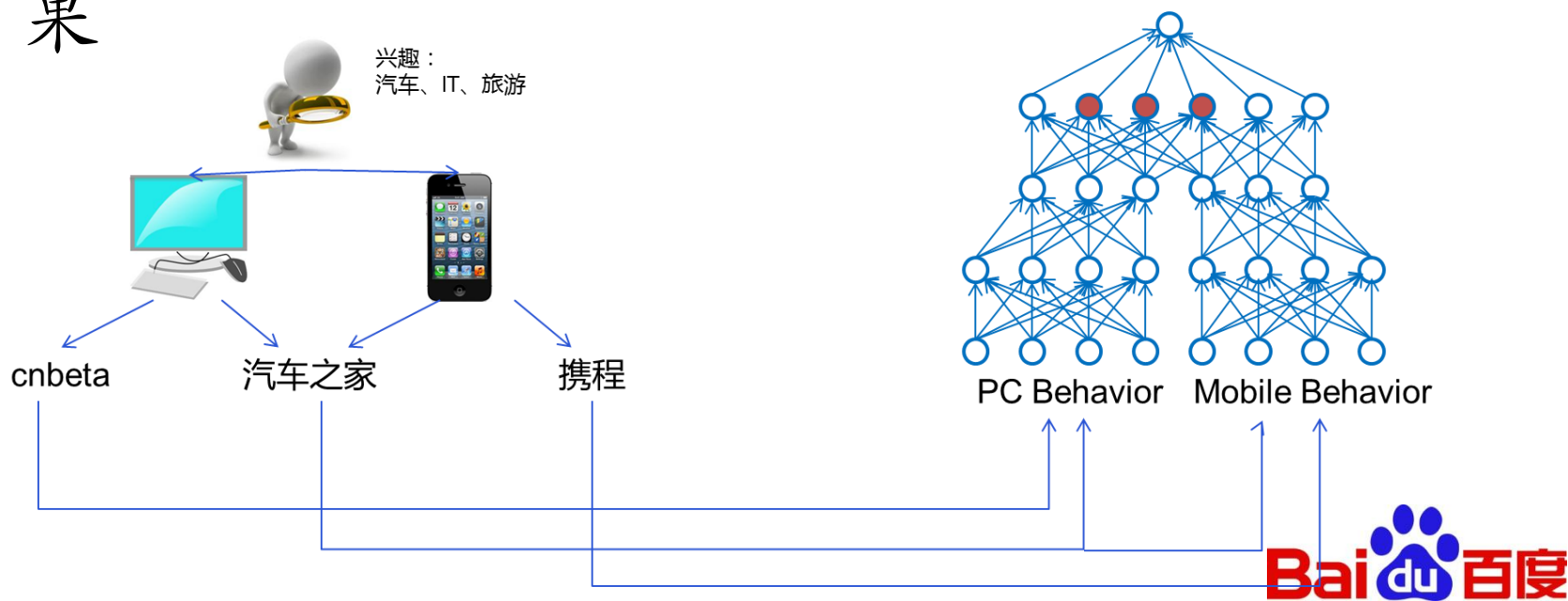
	DNN一期	DNN二，三期
模型大小	<10w 参数	50w 参数
数据规模	20亿	100亿
训练数据存取	本地存放	基于HDFS
线下训练速度		4.5x (异步训练、下载，多卡并行)
线上预估速度	数值/指令集优化	exp()函数优化
历史统计数据流	滑窗统计数据流	数据延时3h->1h
K-V查询	凤巢大规模使用分布式K-V第一步	支持增量(5h->1h)，节省一半内存，更高效率 (qps 3000->10000)

项目效果

- 2012/2 – 2013/5 **深度学习在搜索广告中的应用一期**
 - DNN首次全流量服务于凤巢CTR预估
 - 全流量左侧整体CTR2 **+3.67%**, PVR*ASN **-4.10%**
- 2013/6 – 2013/11 **深度学习在搜索广告中的应用二期**
 - 实现了深度学习模型首次加入凤巢迭代
 - 全流量左侧整体CPM1 **+0.65%**, CTR2 **+3.28%**, PVR*ASN **-2.82%**
- 2013/11 – 2014/1 **深度学习在搜索广告中的应用三期**
 - 完全替换LR, 成为新一代CTR预估模型
 - 全流量左侧整体CPM1持平, CTR2 **+6.19%**, PVR*ASN **-6%**

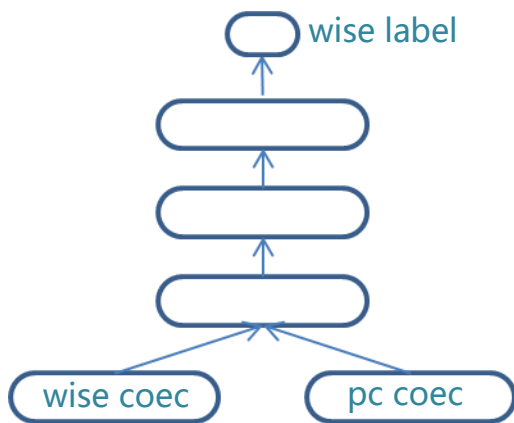
无线搜索广告上的应用

- 在无线上迁移了现有PC搜索广告的结构
 - 全流量基线AUC **+1.5**个百分点
- 用数据更丰富的PC提升无线搜索广告的效果

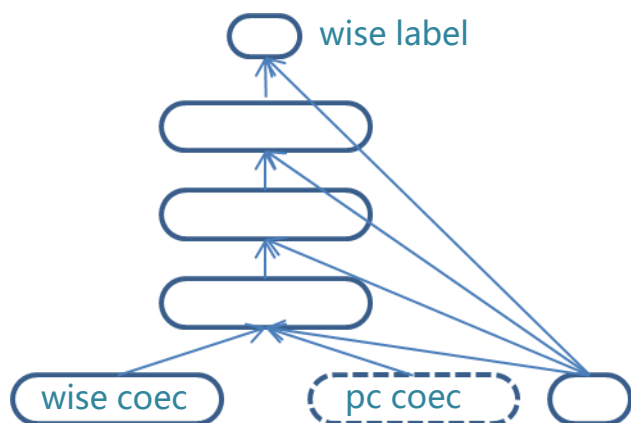


A Transfer Learning Road Map

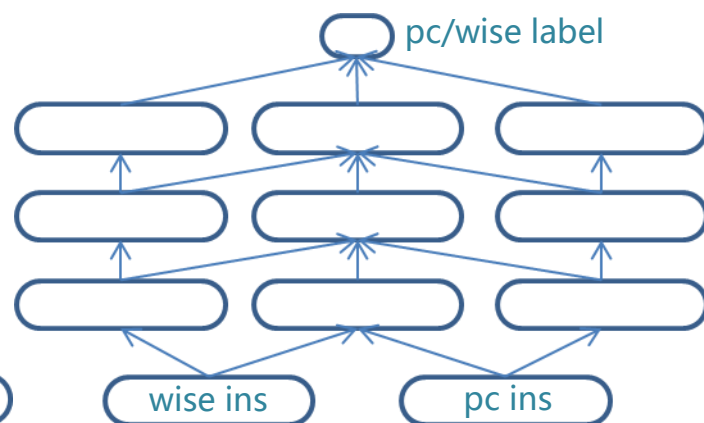
- 由于问题不同, 直接混合训练一般会失败
- 由易到难的迁移知识
 - 迁移特征
 - 迁移PC搜索广告数据 Label
 - 迁移PS自然搜索结果 Label



新增特征



特征分布变化
Label分布不变



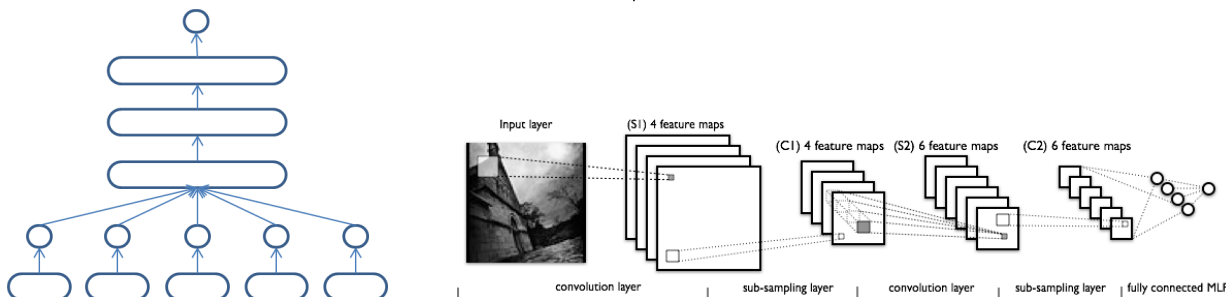
特征分布变化
Label变化

Future Work

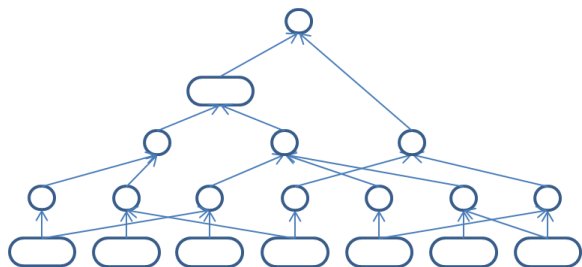
- 新一代CTR Model应该怎样
 - 联合优化而非两阶段的模型
 - 更泛化、更多考虑语义的模型
 - 大规模稀疏的模型
- 工程上
 - 流式实时的模型
 - 支持线上预估使用大模型

大规模稀疏的模型

- 越复杂的模型越难以学习
- 人工结构先验可以降低学习的难度



- 我们可以加入更多结构先验



Dict, Wiki, Relation, Clustering

谢谢！

Q & A