



**QCon** 全球软件开发大会  
INTERNATIONAL SOFTWARE  
DEVELOPMENT CONFERENCE

BEIJING 2017

# 搜狗深度学习技术在广告推荐领域的应用

舒鹏



促进软件开发领域知识与创新的传播



关注InfoQ官方信息  
及时获取QCon软件开发者  
大会演讲视频信息



扫码，获取限时优惠



全球架构师峰会 2017 [深圳站]

2017年7月7-8日 深圳·华侨城洲际酒店

咨询热线：010-89880682



全球软件开发大会 [上海站]

2017年10月19-21日

咨询热线：010-64738142

# 目录

## CONTENTS

01

搜索广告背景知识

02

深度学习在搜狗搜索广告的一些应用

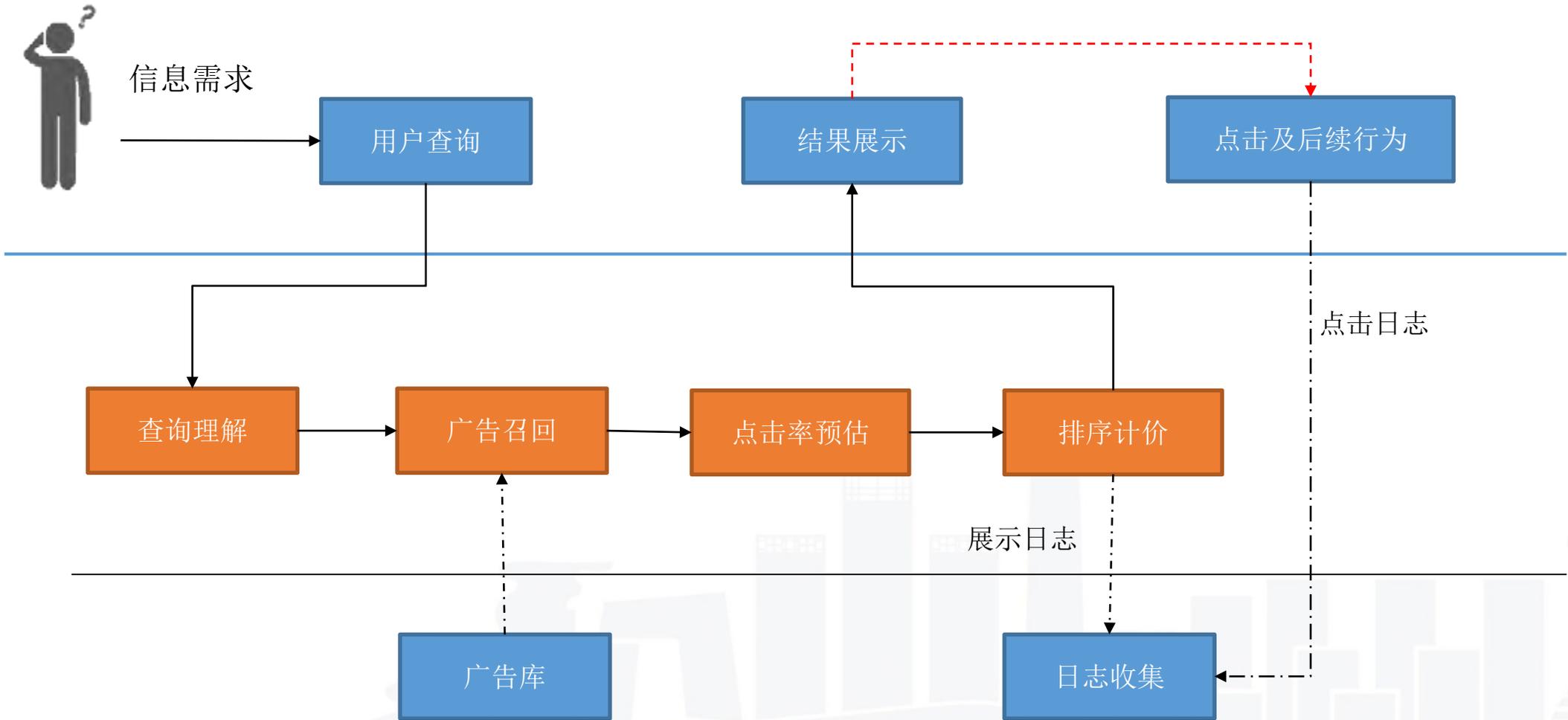
03

基于多模型融合的CTR预估

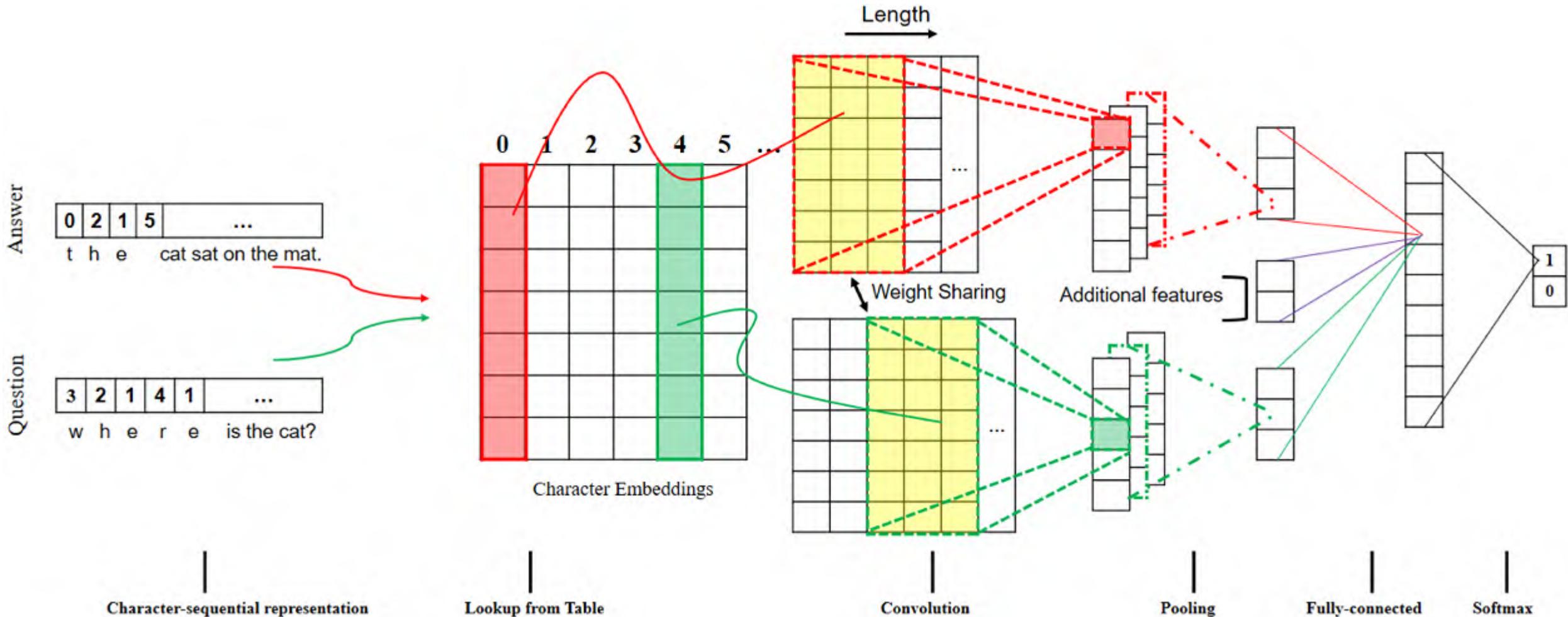
04

若干思考

# 搜索广告背景知识



### 文本相关性计算

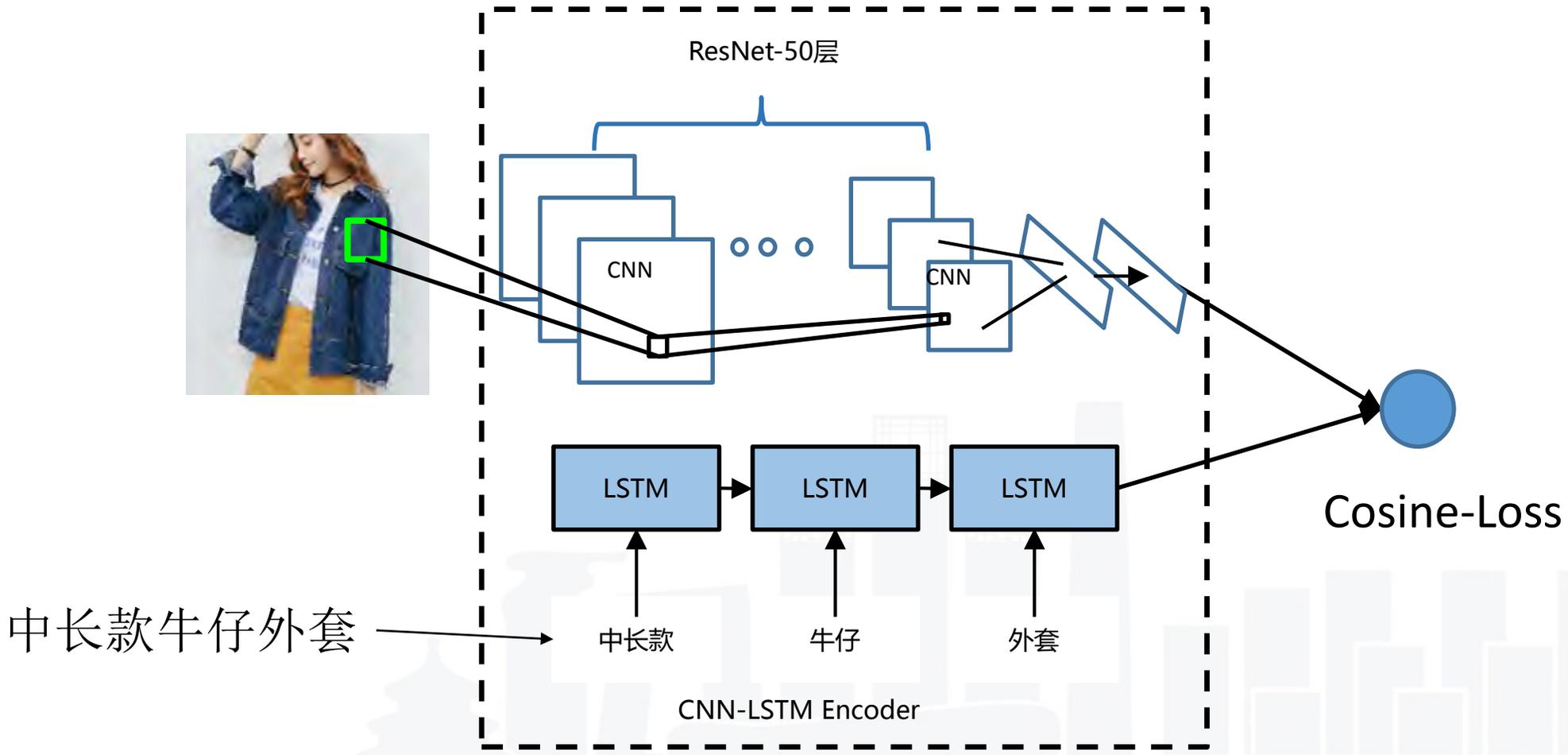


无需分词：基于字符粒度表达的问答系统设计

L.X Meng, Y.Li, M.Y Liu, P Shu. Skipping Word: A Character-Sequential Representation based Framework for Question Answering. *CIKM2016*, pages 1869-1872, 2016. Sogou Inc

# 深度学习在搜狗搜索广告的一些应用

## 广告物料推荐



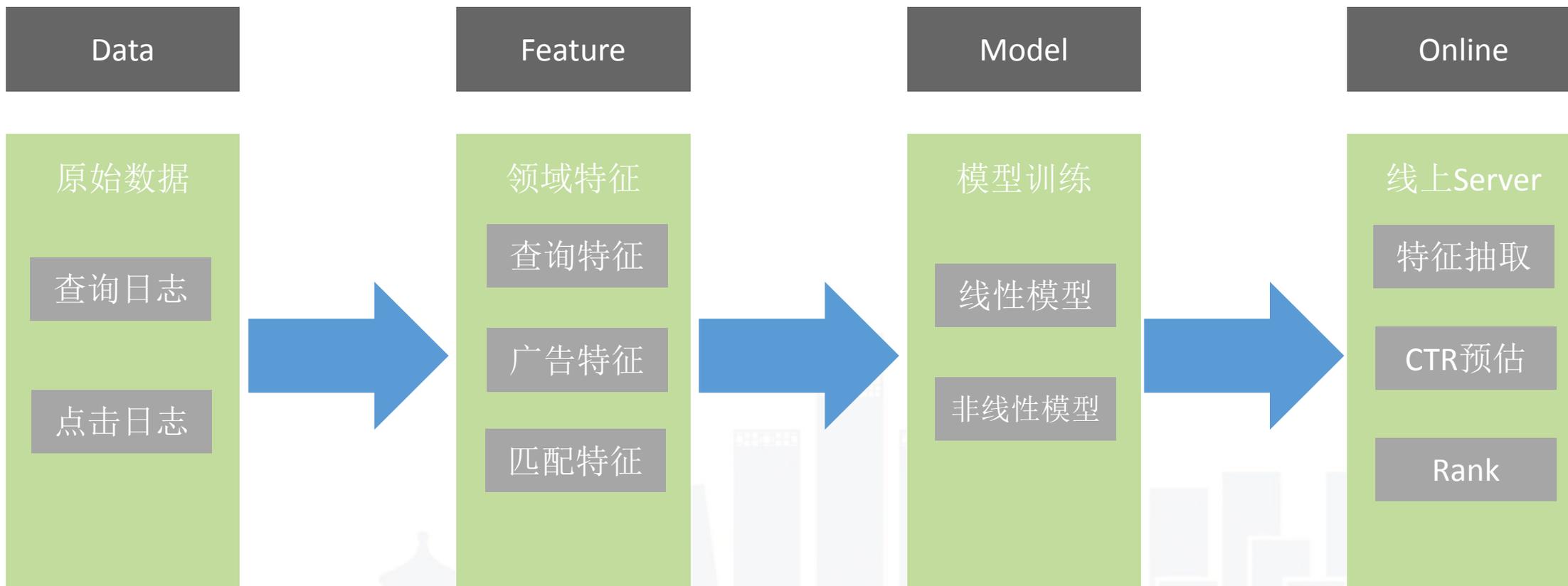
# 深度学习的搜狗搜索广告的一些应用

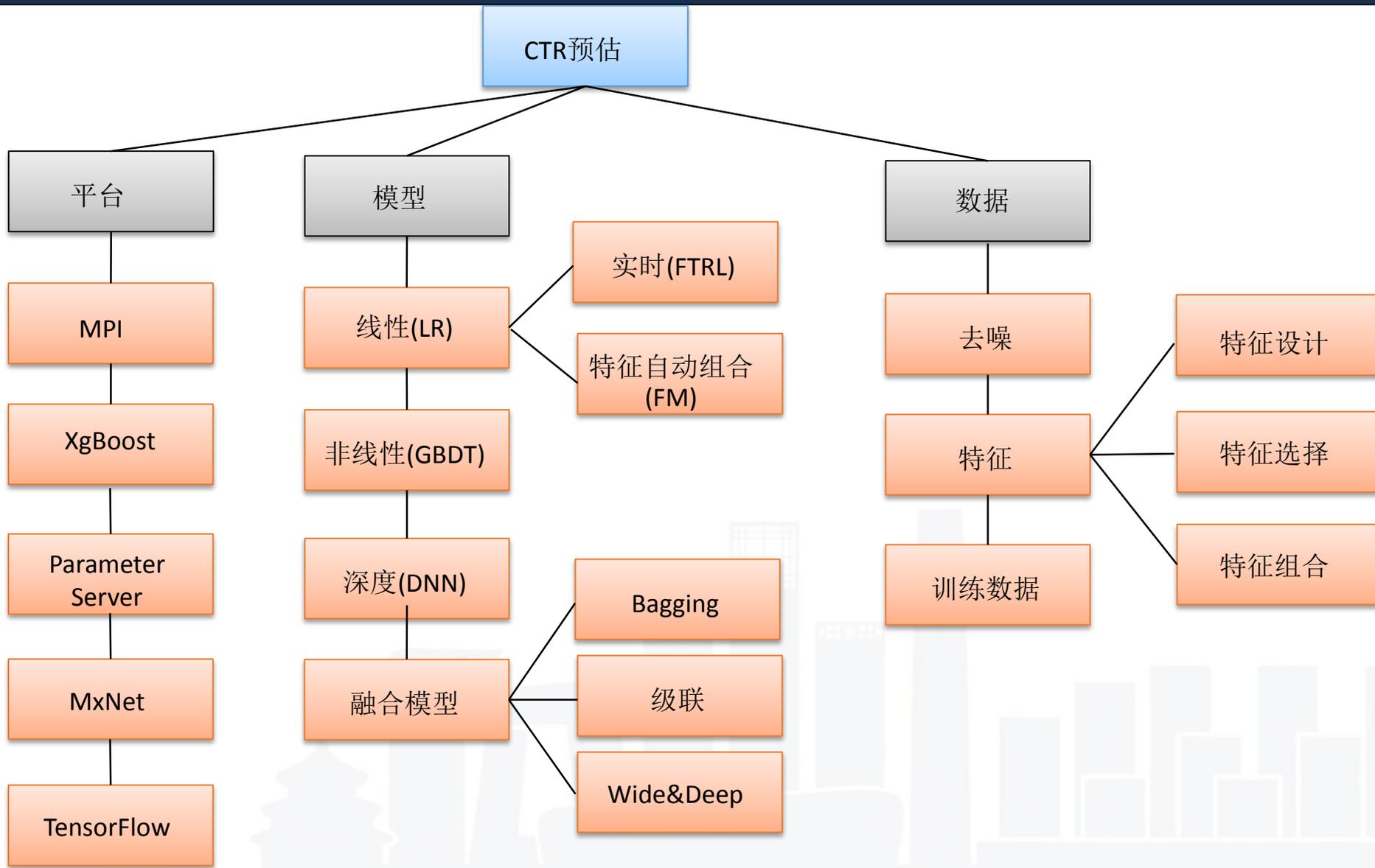
方向	用途	相关技术
图像理解	图片物料推荐	CNN
文本相关性	广告召回、创意生成	Word2Vec、CSR、LSTM
CTR预估	广告排序、特征挖掘	DNN、MxNet、TensorFlow



03

## 基于多模型融合的CTR预估





## 离散特征

---

容易设计；刻画细致；特征稀疏；

特征量巨大；模型复杂度受限



## 连续特征

---

需要仔细设计；定长；特征稠密

特征量相对较小，可以使用多种模型训练

- 简单、处理特征量大、稳定性好
- 需借助交叉特征
- Logistic Regression

线性

## 模型类别

非线性

- 能够学习特征间非线性关系
- 模型复杂、计算耗时
- DNN、GBDT

## CTR bagging

- 将多个模型的输出CTR加权平均
- 实现方法简单，模型之间不产生耦合
- 可调参数有限，改进空间相对较小

## 模型融合

- 任一模型的输出作为另一模型的特征输入
- 实现方法复杂，模型之间有依赖关系
- 实验方案较多，改进空间较大

## 目标

- 可支持多个不同模型的加载和计算
- 可支持模型之间的交叉和CTR的bagging
- 可通过配置项随时调整模型融合方案
- 避免不必要的重复操作，减少时间复杂度

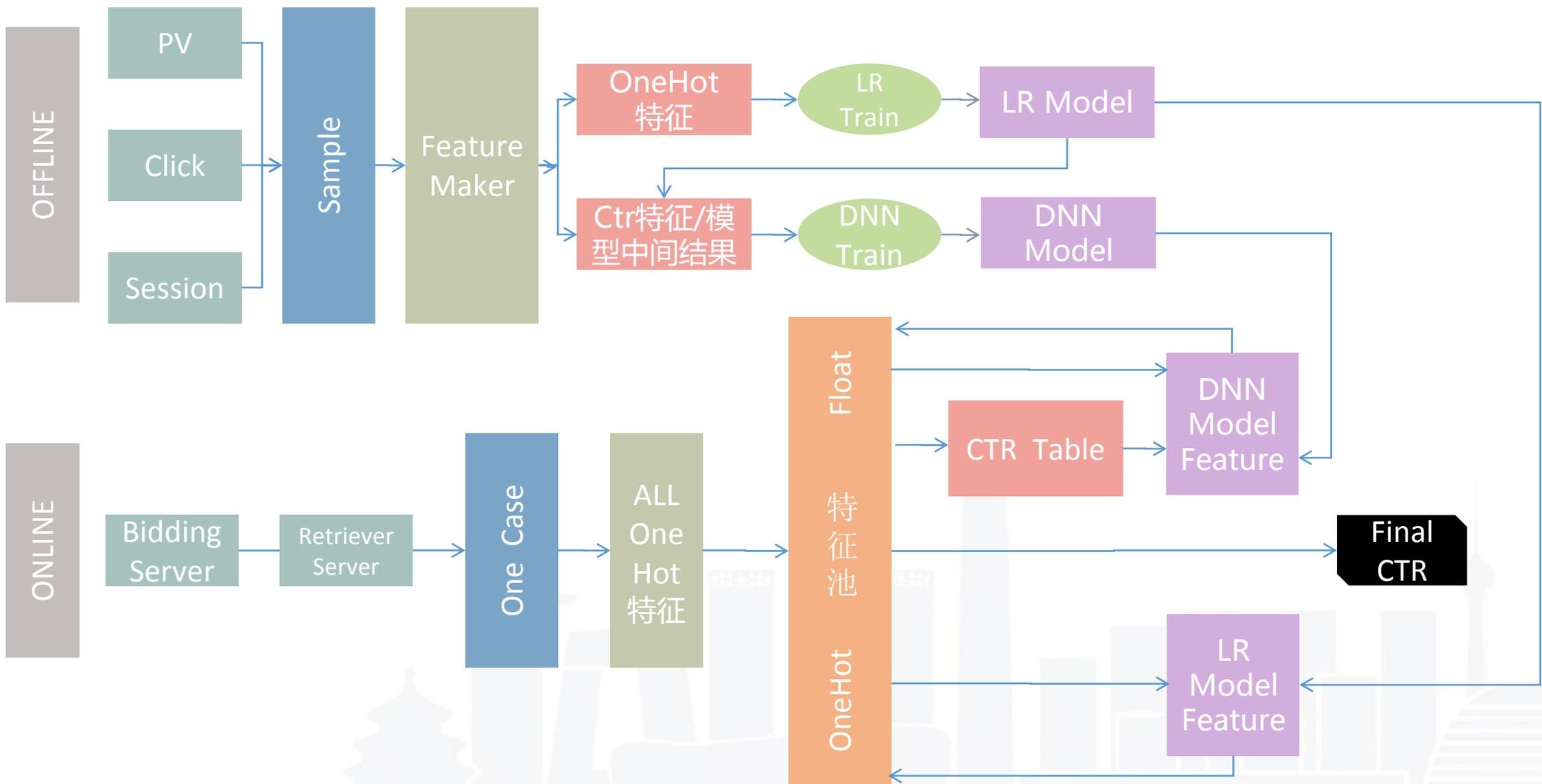
## 解决方案（引入ModelFeature的概念）

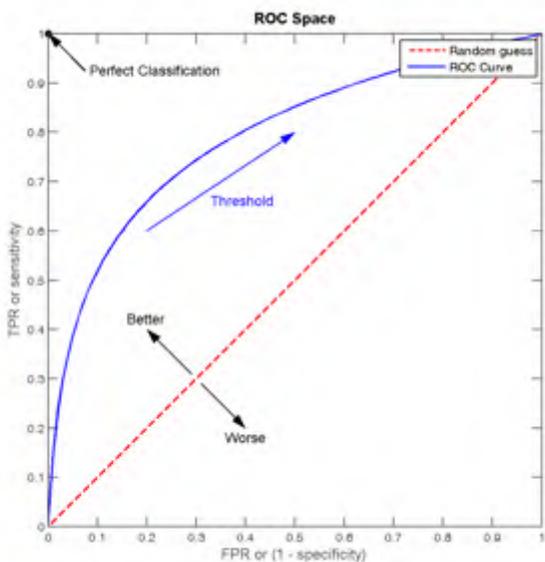
- 模型本身也看做一个抽象特征
- 模型特征依赖于其它特征，通过计算得到新的特征
- 模型特征输出可作为CTR，也可作为特征为其它模型使用
- 限定ModelFeature的计算顺序，即可实现bagging/模型交叉等功能

## 关键点

- 数据一致性
- 流程稳定

# 模型融合





AUC



上线收益

Survival Bias  
特征覆盖率

是否一致？

- 加大数据量，提升模型稳定性
- 加大数据量，提升模型收益

诉求

## 并行化训练

方案

- MxNet支持多机多卡, 使用成本低
- 构建多机多卡GPU集群, 优化训练效率, 提高加速比

## 现状

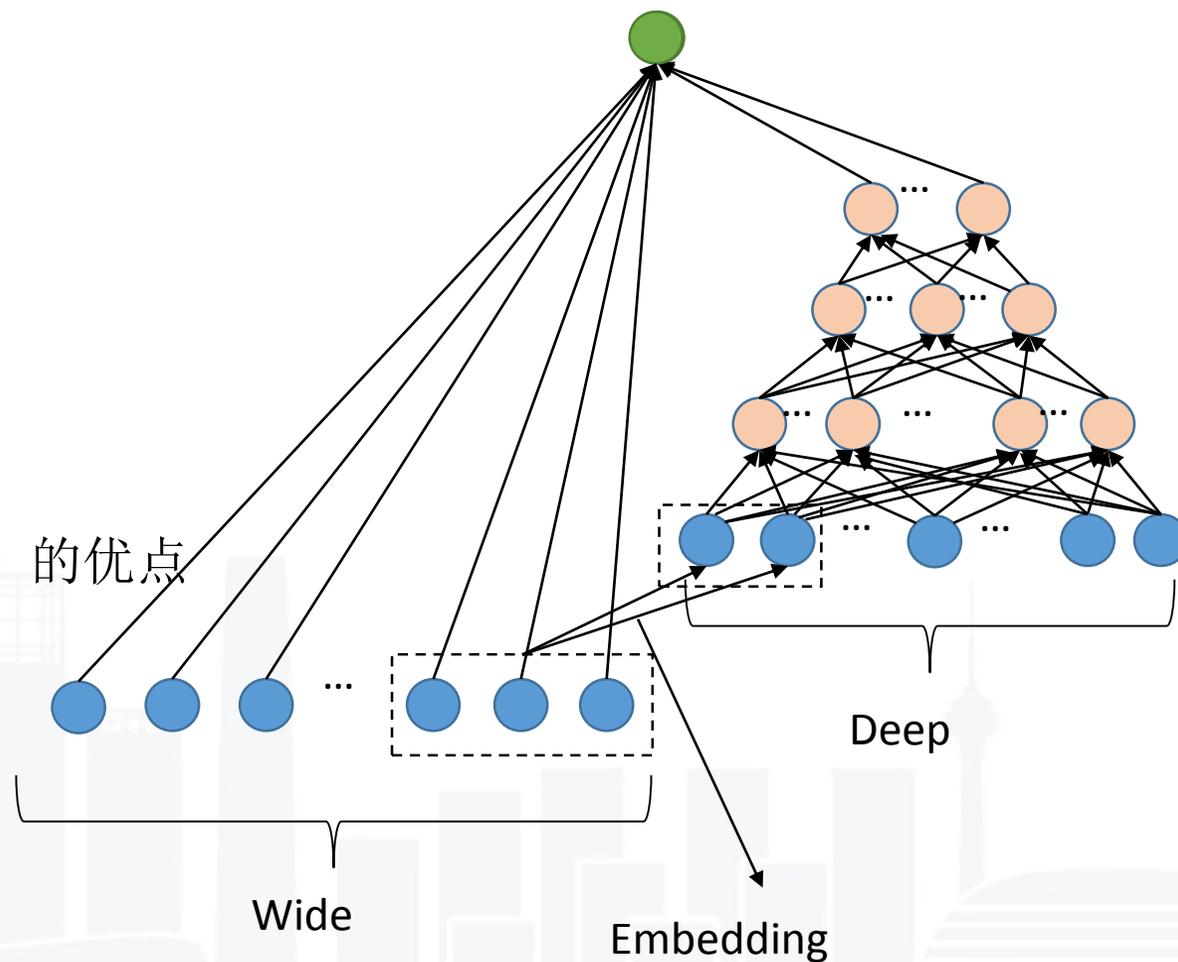
- 已经实现LR+DNN融合模型的上线，收益较好
- 受限于线上计算资源，模型复杂度有限
- 线下训练流程有依赖，繁琐易出错

## 计划

- 线上服务拆分，独立出深度学习计算模块，采用低功耗GPU加速
- 支撑更宽、更深、更复杂的网络结构
- 采用Wide & Deep，线下训练流程解耦

## 背景和优势

- Google于16年6月份发表相应论文
- 用于应用商店中推荐APP的排序
- 基于TensorFlow平台，可兼具业界流行模型（LR、DNN）的优点
- 一次训练给出两个模型，流程简洁稳定，效果更佳



04

## 若干思考

1

## DL的强项

输入不规整  
结果确定  
容易获取的海量训练数据

2

## CTR预估

特征有明确含义  
场景相关，以用户为导向  
很难界定“Ground Truth”  
训练样本“有限”

3

## 方向

特定业务场景  
模型融合  
提升效率，降低成本



关注QCon微信公众号，  
获得更多干货！

# Thanks!



INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

主办方 **Geekbang** > **InfoQ**  
极客邦科技