

# 高校算法大赛

到底对不对

2017-07-06

freeze()

for

<=Date()

9

+

8

if

Your

var

time



# 目录

CATALOG

01



团队成员

02



赛题详情

03



解题思路

04



团队故事

05



总结展望





## 赛题详情

比赛  
场景

以移动App广告为研究对象，给定广告、用户和上下文信息

预测  
目标

预测App广告点击后被激活的概率

评价  
指标

对数损失函数



## 解题思路

- 业务分析
- 特征工程
- 模型改进
- 模型融合



## 业务分析

点击

是否下载?

下载

是否激活?

激活

主观因素

是否习惯/活跃下载  
偏好下载哪类APP  
已经下载过哪些APP

用户与广告的交互

客观因素

iOS非WiFi下载限制  
更换设备后再下载  
广告内生转化率

交叉, 时间相关, 转化特征

激活统计策略

广告主较长上报时间  
连续点击统计策略

清洗数据, 连续点击特征





## 特征工程

基础特征

交互特征

转化特征

冷启动特征

交叉特征

时间相关特征

- APP平台
- APP一级类别
- APP类别
- 广告站点ID
- 广告位类别
- 联网方式
- 服务商
- .....



## 特征工程

基础特征

交互特征

转化特征

冷启动特征

交叉特征

时间相关特征

- 十分钟内是否连续点击
- 前四次点击的转化序列
- 用户在install表中下载此类别APP次数
- 用户当天第几次点击当前APP
- 用户是否下载过当前APP
- 用户距上次下载此APP的时间
- .....



## 特征工程

基础特征

交互特征

转化特征

冷启动特征

交叉特征

时间相关特征

- APP ID 转化率
- APP类别转化率
- 广告主转化率
- 推广计划转化率

- 广告位转化率
- 广告位类别转化率
- 广告转化率
- .....





## 特征工程

基础特征

交互特征

转化特征

冷启动特征

交叉特征

时间相关特征

- 用户是否第一天出现
- 广告主是否第一天出现
- APP是否第一天出现
- 广告位是否第一天出现
- 广告是否第一天出现
- .....



## 特征工程

基础特征

交互特征

转化特征

冷启动特征

交叉特征

时间相关特征

- APP-用户婚否
- APP-用户教育程度
- APP-用户性别

- APP-联网方式
- APP-运营商
- .....



## 特征工程

基础特征

交互特征

转化特征

冷启动特征

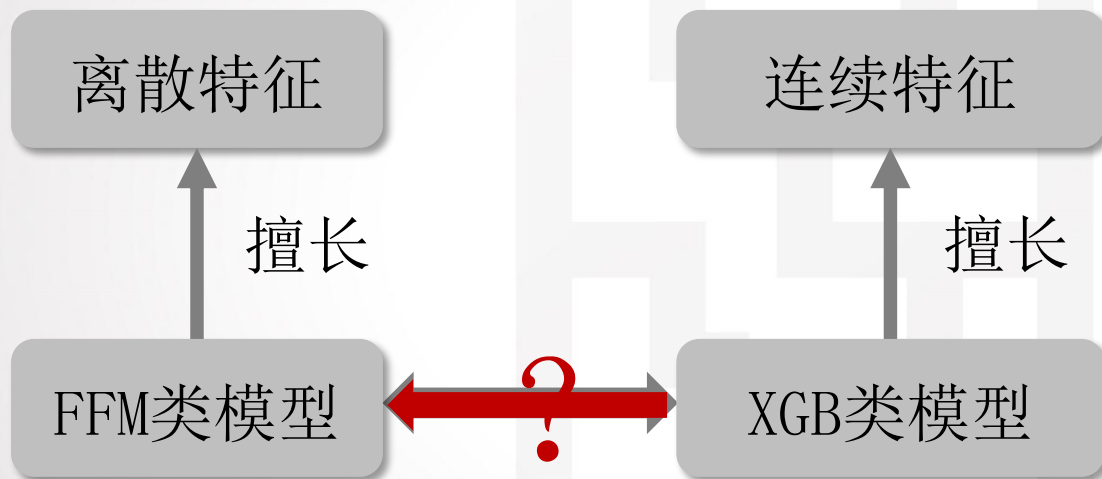
交叉特征

时间相关特征

- 用户当天点击次数
- 距离用户上一次点击时间
- 距离用户下一次点击时间
- 用户是否更换运营商
- 用户是否更换APP平台
- .....



## 特征工程



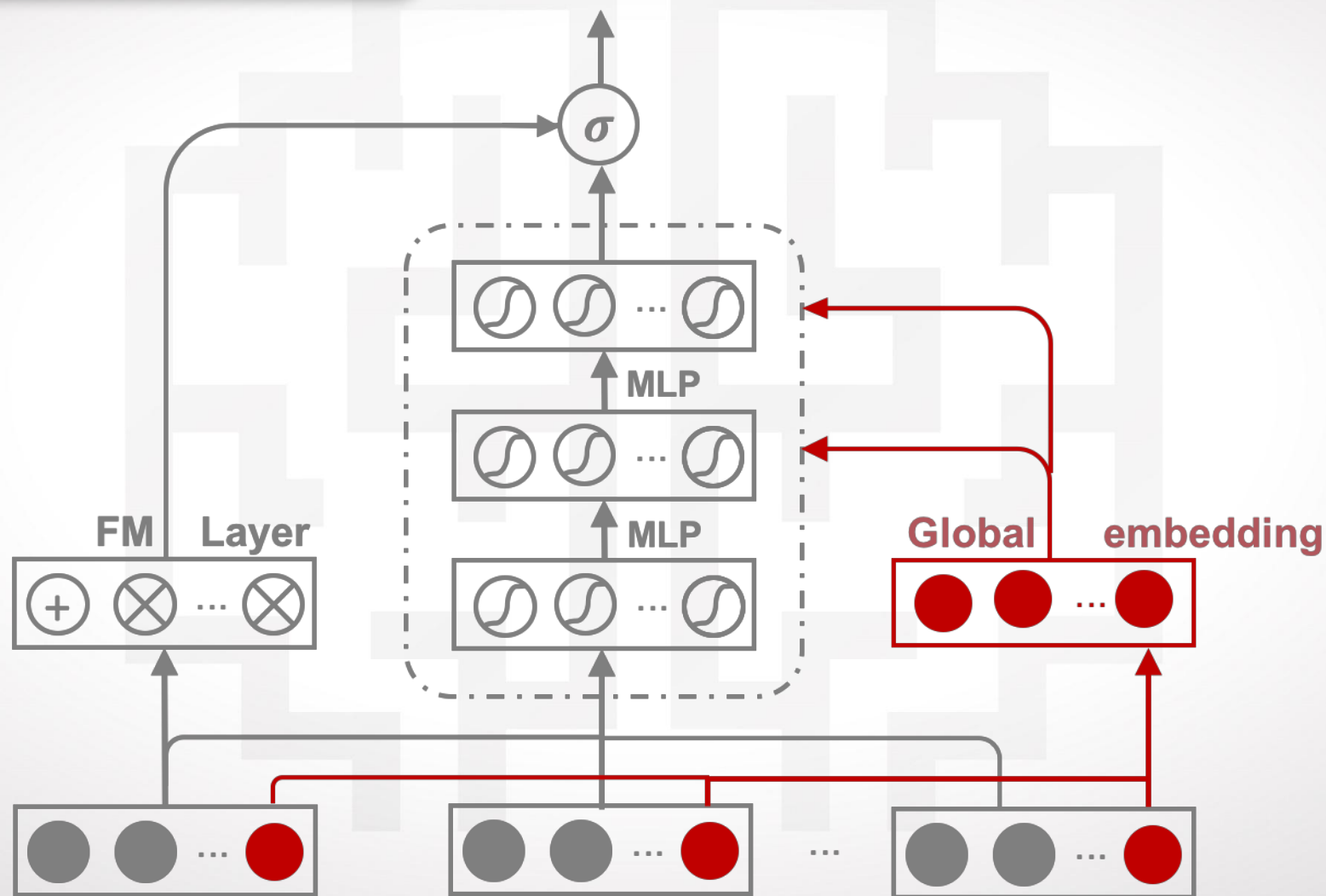
叶子节点类别

FFM+XGB叶子节点特征 = Top10



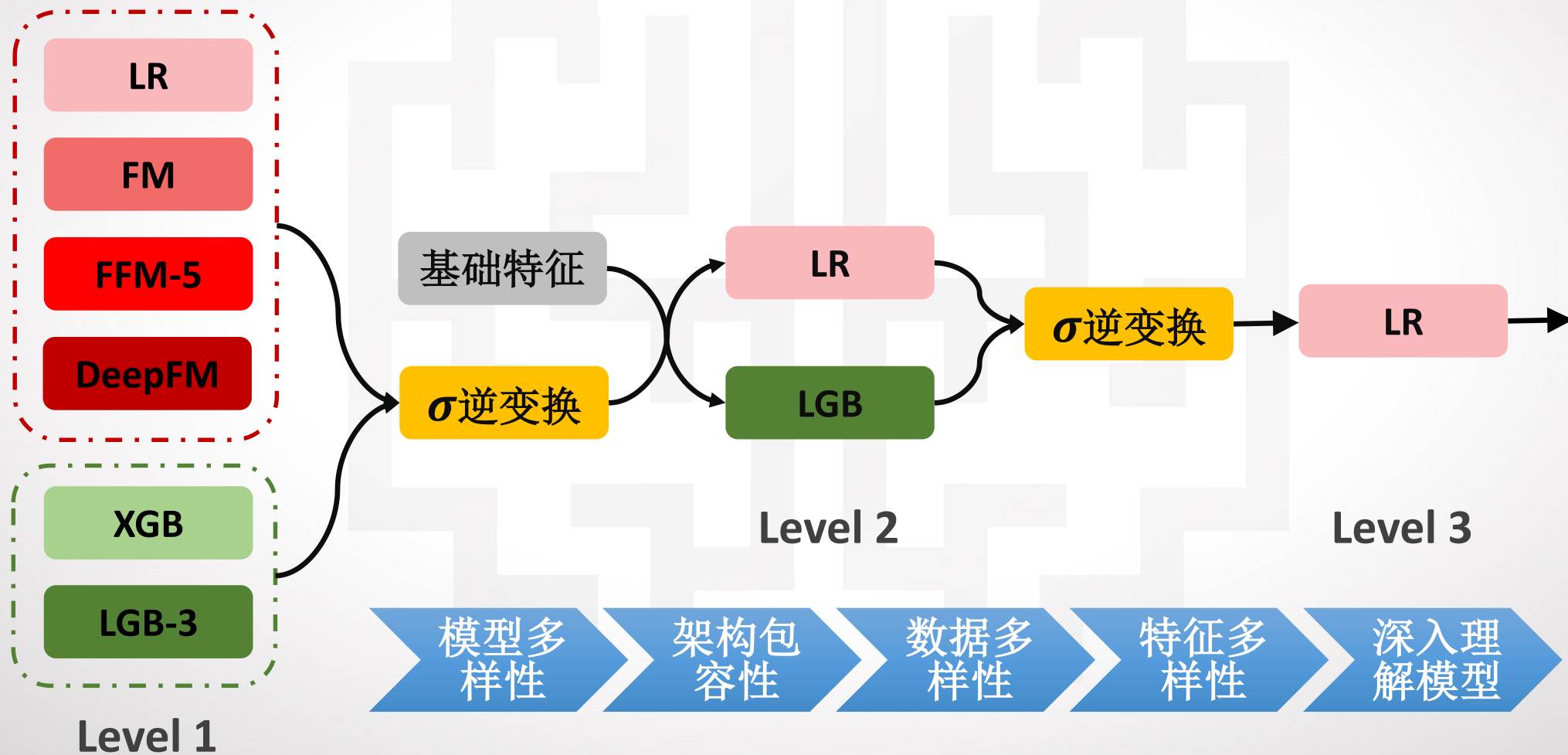


## 模型改进-DeepFM





## 模型融合





## 困难挑战

### 模型效果疲软

原因：模型架构存在缺陷、模型单一

- 改进DeepFM模型使更加适应场景
- 尝试不同性质的模型，增加多样性

02  
ITEM

### 树模型线上线下一致

原因：信息泄露，过拟合严重

- 删除垃圾特征
- 对转化率特征进行平滑修正
- 调整模型参数，克服过拟合

04  
ITEM

01  
ITEM

### 特征表现力差

原因：缺乏对业务的理解和思考

- 梳理业务流程，环环必究
- 挖掘强特征

03  
ITEM

### 迭代速度太慢

原因：实现缺乏并行化，代码耦合性强

- 重构代码
- 实现特征“热插拔”





## 历史成绩

Top1

### 初赛中期

+连续点击特征  
+点击次序特征

Top30

初赛前期  
基础特征

### 初赛后期

+DeepFM改进

Top1

### 决赛前期

重构代码  
特征“热插拔”

Top20

Top2

### 决赛中期

+冷启动特征  
+LGB叶子特征

### 决赛后期

12个模型  
3层融合

Top4







## 团队故事



### Team Story

#### 翻身当广告主

注册腾讯广告推广账户  
真正化身广告主  
深入业务流程走一遭



#### 高效圆桌会议

高效开会解决问题  
Trello团队进程管理  
随时微信QQ探讨



#### 站在巨人肩膀上

研究DeepFM算法论文  
研究Kaggle、有道、京  
东等算法比赛经验分享





## 总结展望

深入探索业务逻辑



人工特征+模型特征

改进现有模型



多模型有效融合

腾讯社交广告好不好

就看算法到底对不对



THANKS

freeze()

for

<=Date()

2017-07-06

Your

var

time

if