

CTR预估模型演变

王老师 / 2021-07-17



OUTLINE

1

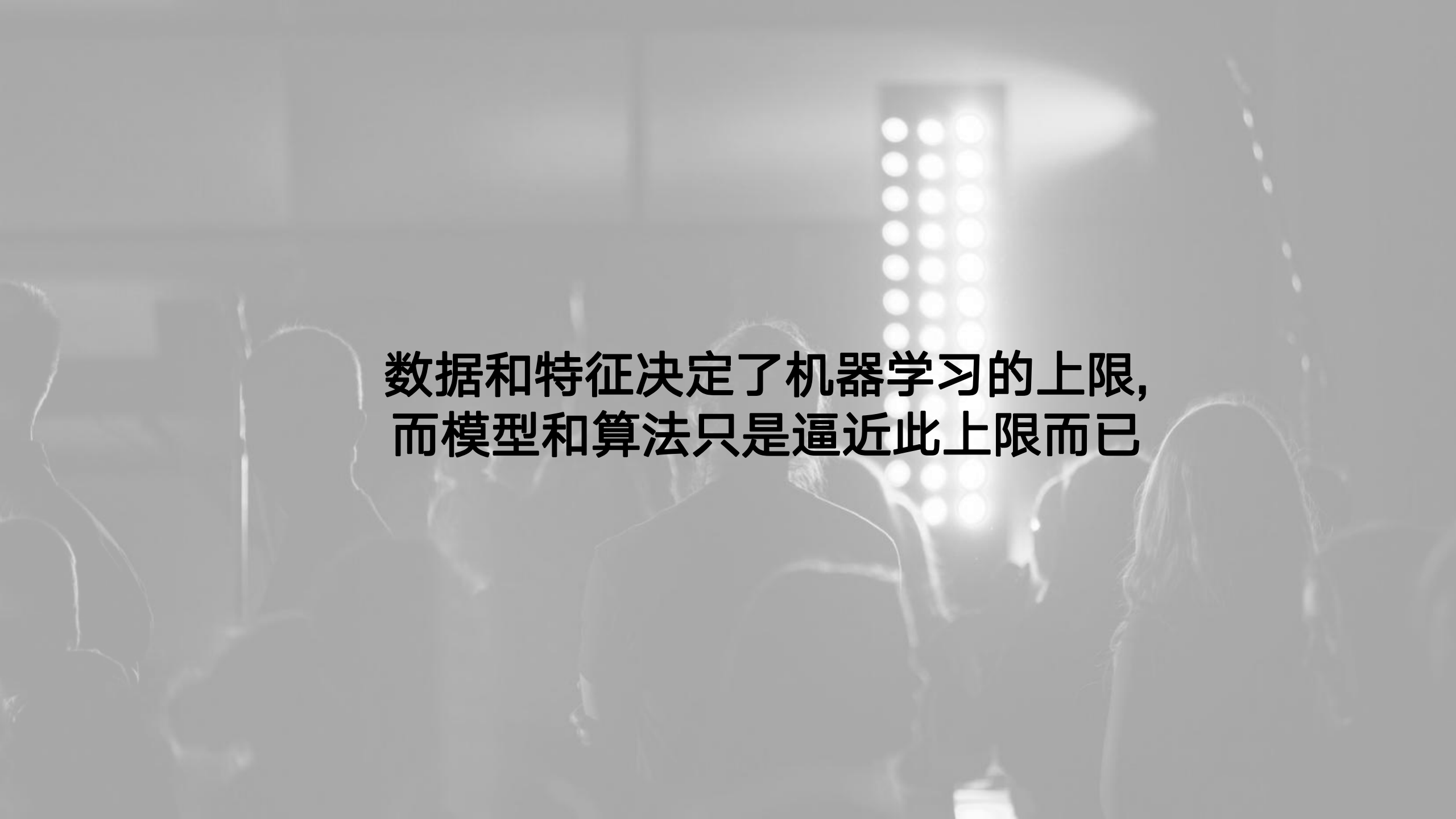
发展初期：人工特征+线性模型

2

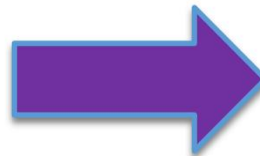
加速发展期：自动特征交叉+线性模型

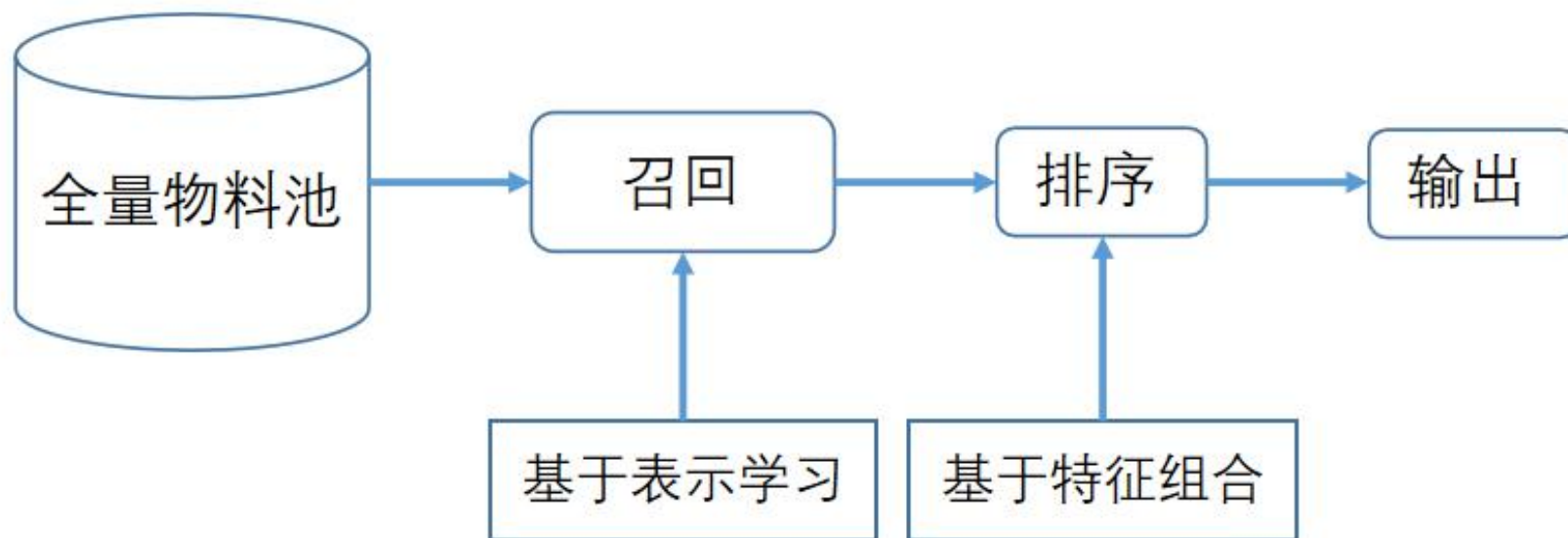
3

深度发展期：深度模型阶段



**数据和特征决定了机器学习的上限，
而模型和算法只是逼近此上限而已**



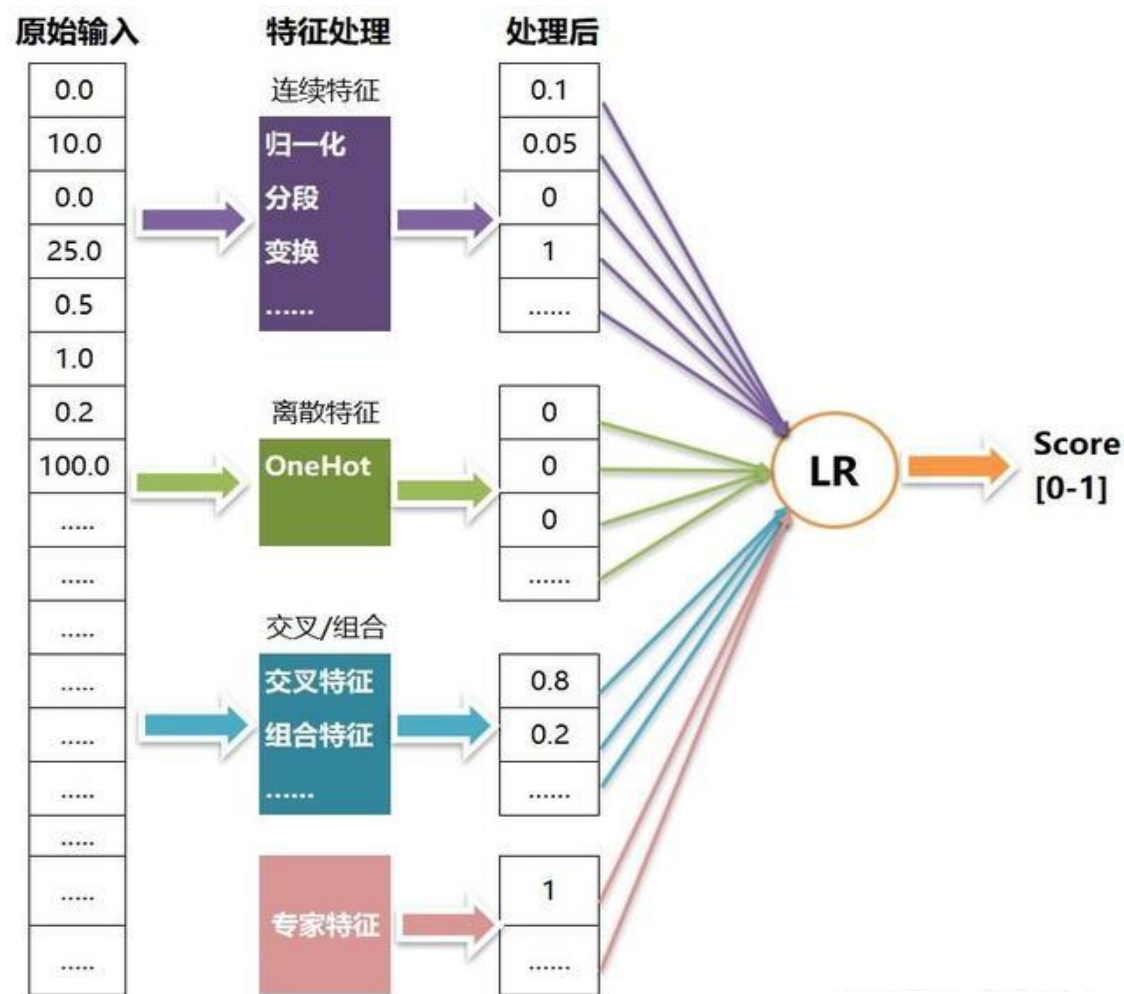


排序环节是推荐系统最关键，也是最具有技术含量的部分，目前大多数推荐技术其实都聚焦在这块。排序环节主要包含两部分：**特征工程**和**模型**。

CTR预估模型-演变发展

发展初期（2010年前）：人工特征+ 线性模型阶段

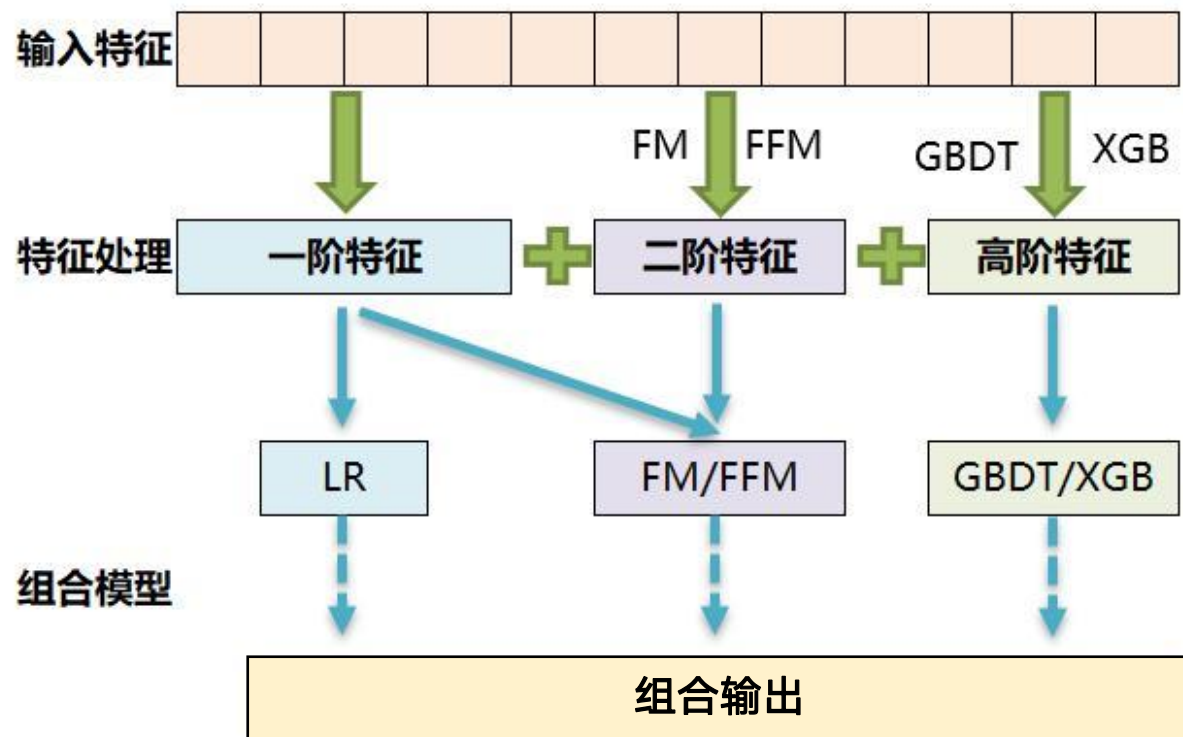
- 1) 原始输入特征量级在百、千、万之间；
- 2) 经过处理后的特征量级在万、十万、百万、千万级别都有；
- 3) 模型虽然简单，但是参数空间小，性能高，效果好；
- 4) 效果提升需要依赖人工特征建设，基于对业务的理解，通过大量的人工经验和数据分析挖掘有效的特征组合；



CTR预估模型-演变发展

加速发展期（2010年-2015年）：自动特征交叉 + 线性模型阶段（FM/FFM、GBDT+LR、XGBoost）

- 1) 有监督自动进行二阶、高阶特征交叉，这样就可以记忆各种有效的特征组合。
- 2) 通过参数可以控制特征交叉的参数空间，比如控制FM隐向量的长度，树模型的棵数和深度。
- 3) 联合低阶、二阶、高阶进行联合训练学习，主要目的是为了在同一个空间中强化记忆每一个特征或者特征组合对预测结果的正向或者负向的影响权重。



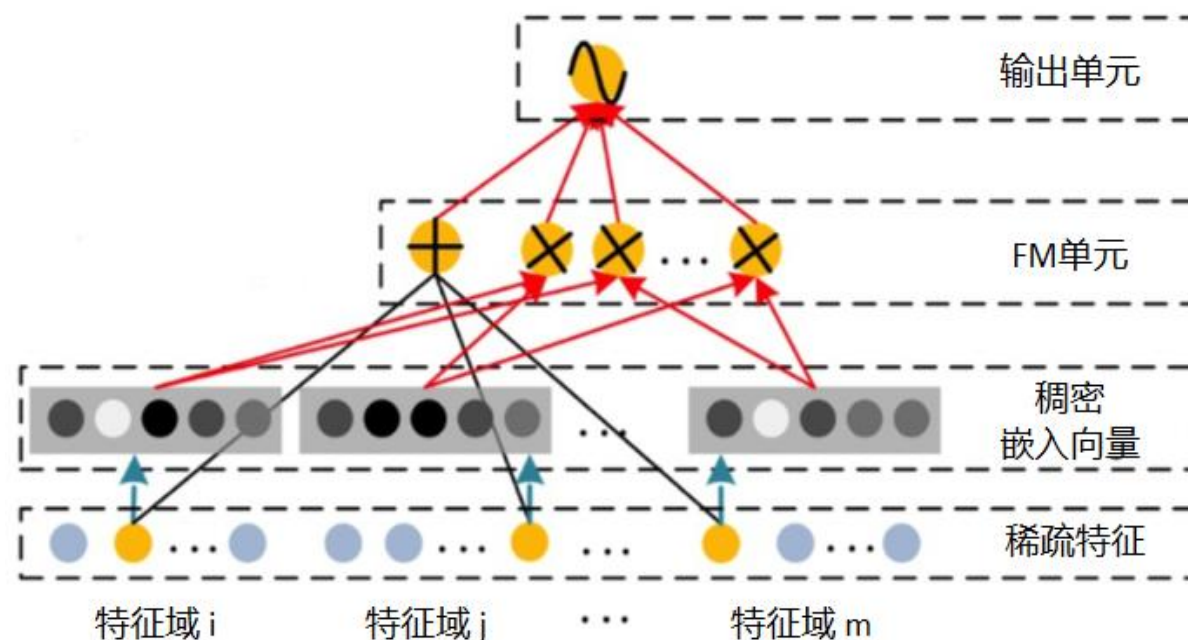
CTR预估模型-演变发展

FM: FM模型在LR模型的基础上增加了自动二阶特征交叉的部分

FFM: 在FM的基础上引入了“场 (Field)” 的概念而形成的新模型

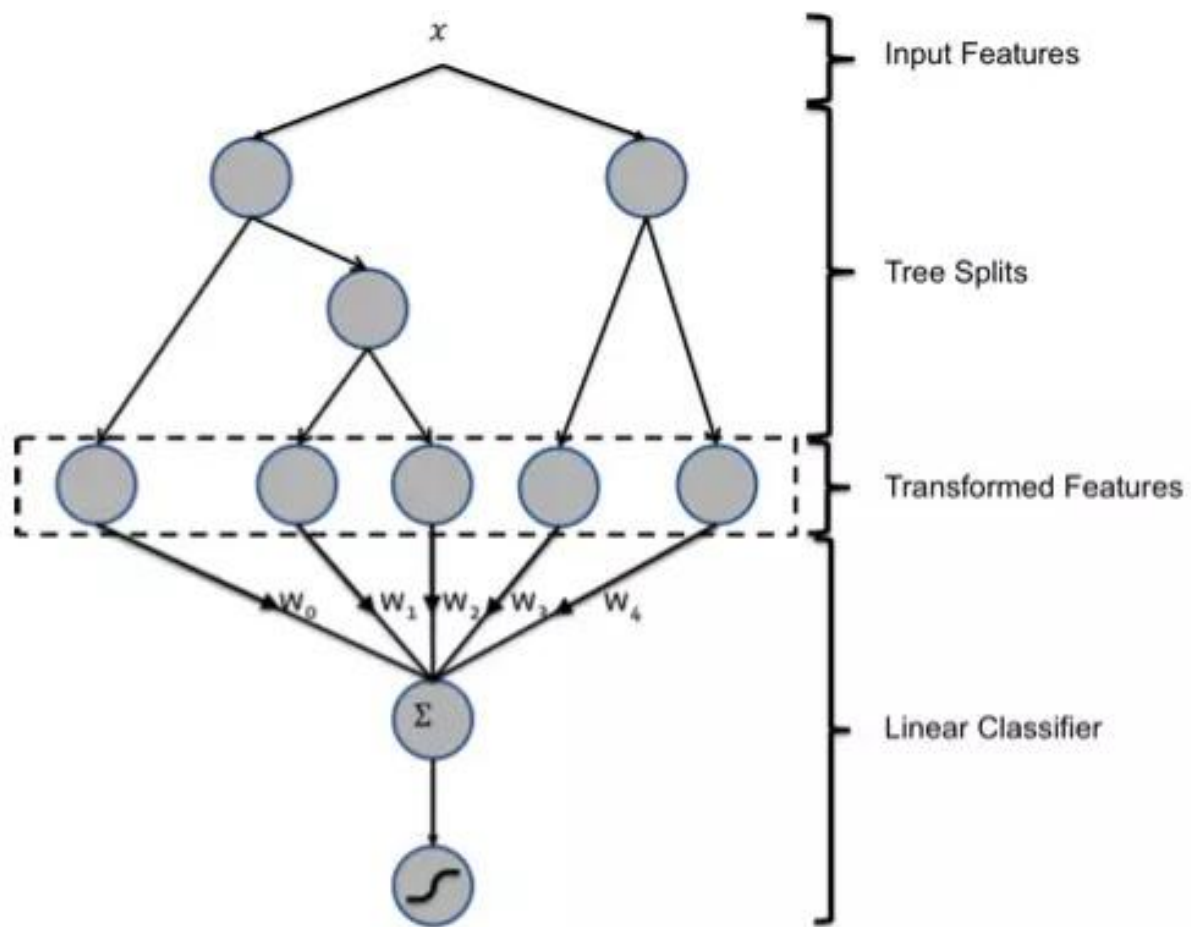
$$\phi(w, x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle V_i, V_j \rangle x_i x_j$$

$$\phi(w, x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle V_{i, f_2}, V_{j, f_1} \rangle x_i x_j$$



CTR预估模型-演变发展

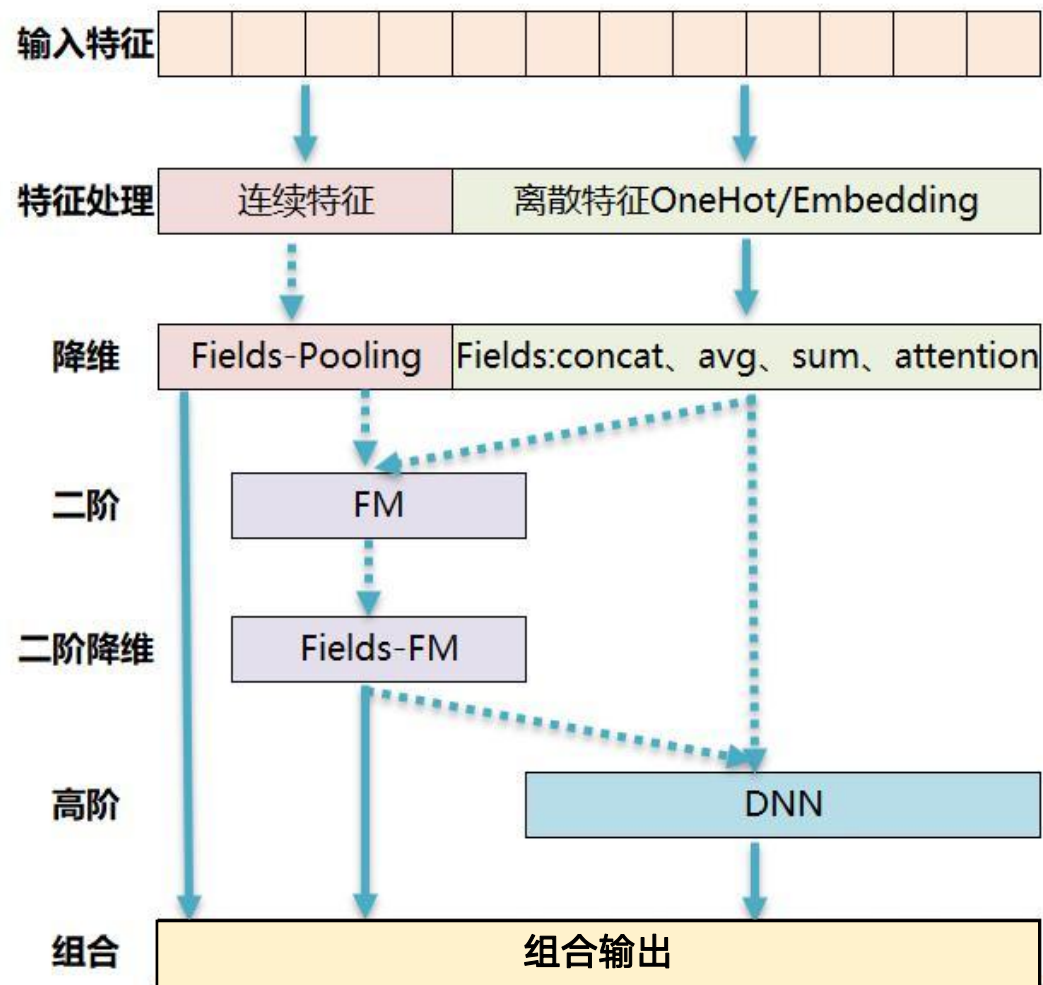
GBDT+LR: 用GBDT进行特征组合, 自上而下每条路径代表一种高阶特征交叉, 然后叶子节点编号作为LR的输入。



CTR预估模型-演变发展

深度发展期（2016年-至今）：深度模型阶段，开始将特征映射至多维空间中，然后再通过多层网络去学习特征之间的相关性（FNN、PNN、Wide & Deep、NFM、AFM、DeepFM、xDeepFM）

- 1) 离散特征处理向Embedding空间发展，采用低维空间更加合理的表达特征，这里即对特征空间进行了压缩，也对离散特征进行了合理的表示；
- 2) 在各个阶段通过降维减少参数空间量级；
- 3) 挖掘上下文与目标之间的关系，比如设计XXX序列特征进行挖掘与目标之间相关性的规律；
- 4) 采用DNN进行高阶特征挖掘；
- 5) 融合低阶、二阶、高阶进行联合学习，低阶和二阶主要强化记忆能力，高阶主要强化泛化能力；



CTR预估模型-演变发展

FNN: 使用FM得隐向量作为embedding初始值

PNN: 在embedding层和MLP之间加入product层

CTR

Fully Connected

Hidden Layer (l2)

Fully Connected

Hidden Layer (l1)

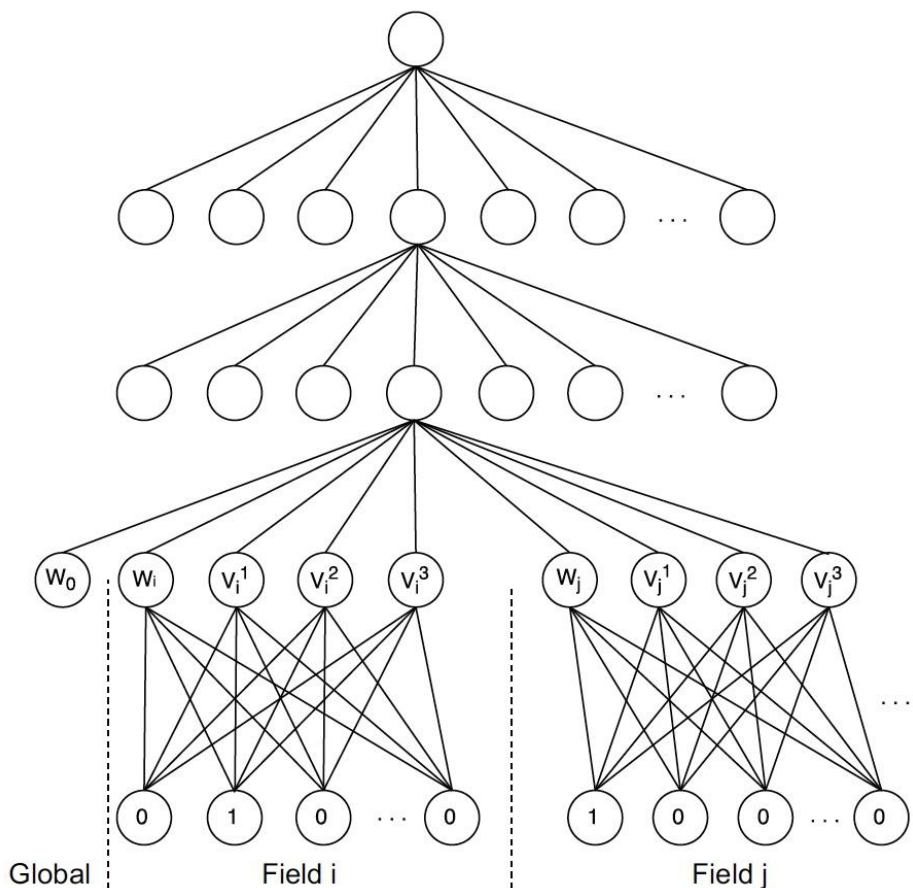
Fully Connected

Dense Real Layer (z)

Initialised by FM's
Weights and Vectors.

Fully Connected within
each field

Sparse Binary
Features (x)



CTR

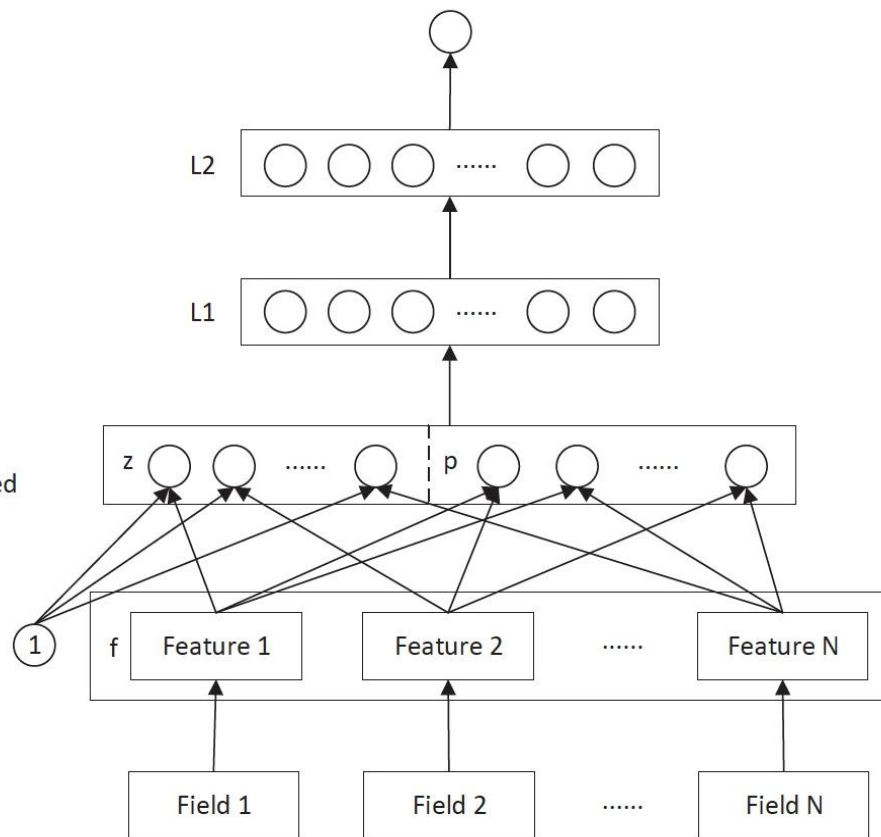
Hidden Layer 2
Fully Connected

Hidden Layer 1
Fully Connected

Product Layer
Pair-wisely Connected

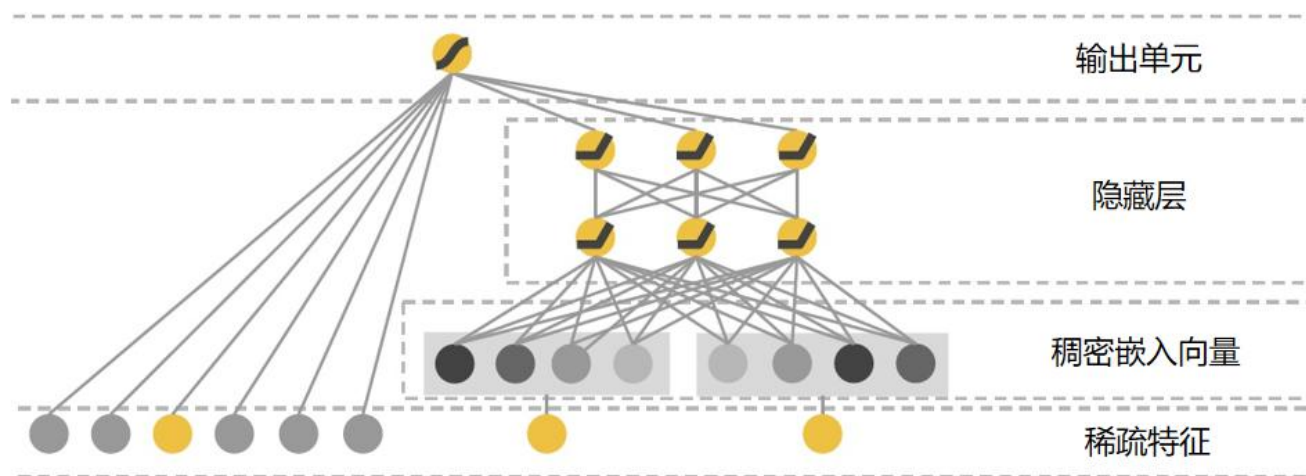
Embedding Layer
Field-wisely Connected

Input

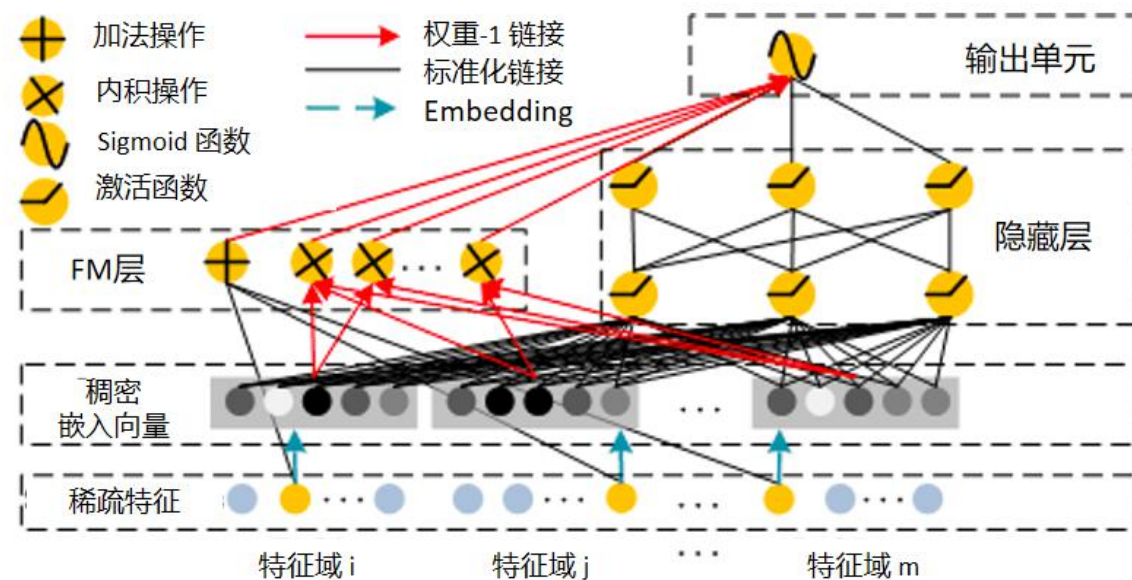


CTR预估模型-演变发展

记忆与泛化结合：Wide&Deep



$$Y_{wide\&deep} = \text{sigmoid}(w_{wide}^T \cdot [x, \phi(x)] + w_{deep}^T \cdot a^{(l)} + b)$$

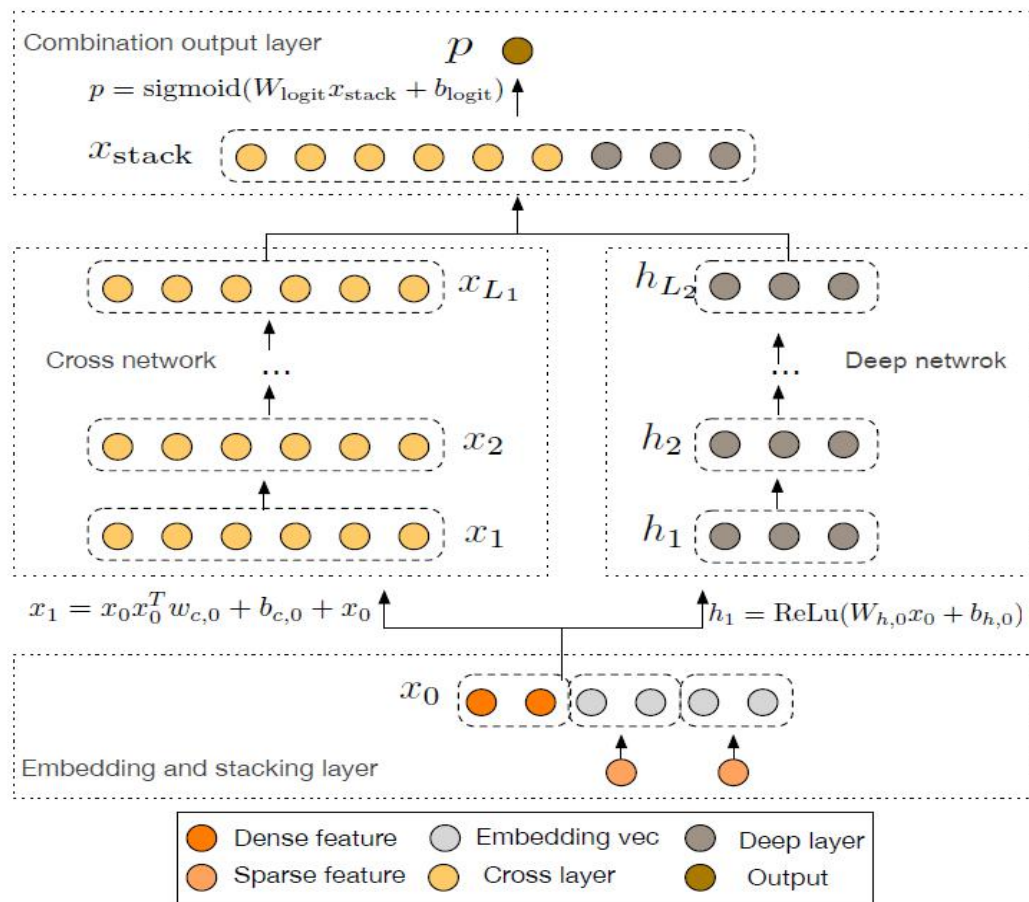


$$Y_{FM} = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle V_i, V_j \rangle x_i x_j$$

$$Y_{DeepFM} = \text{sigmoid}(Y_{FM} + Y_{DNN})$$

CTR预估模型-演变发展

深度交叉网络模型：Deep & Cross



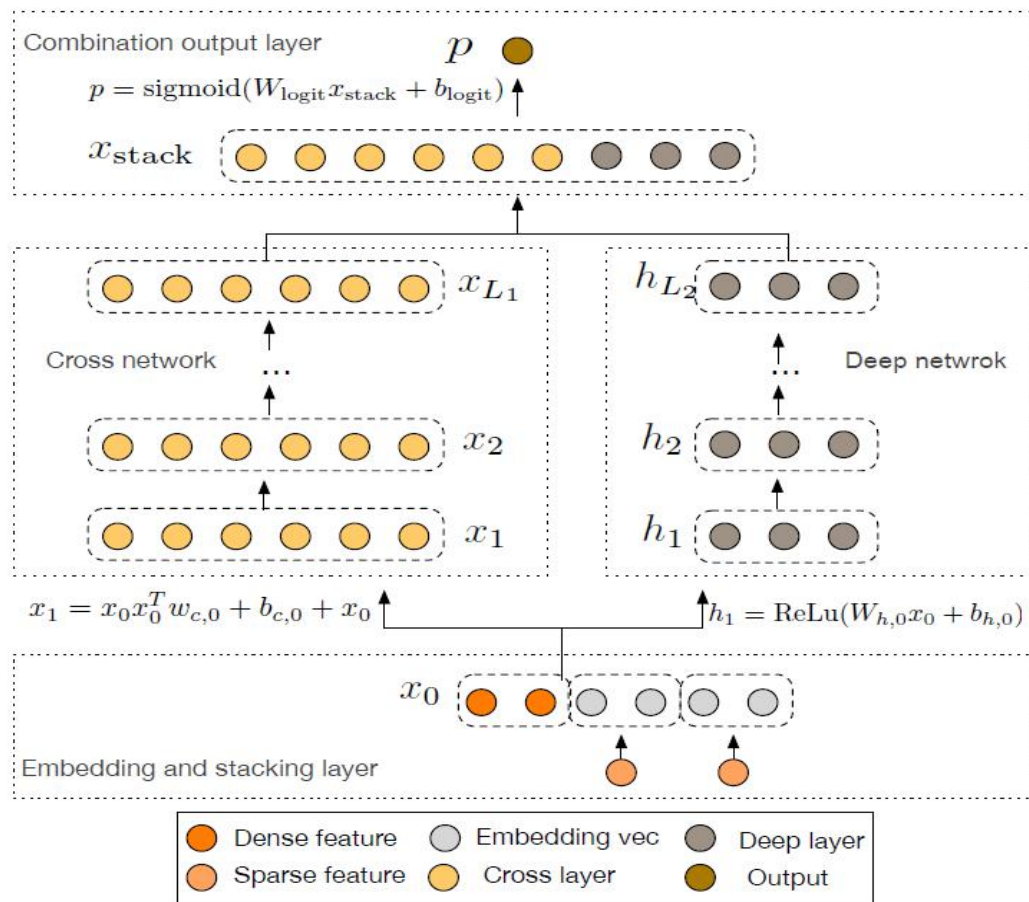
优势

- 较强的cross feature刻画能力
 - 自动构造有限高阶的交叉特征，无需人工特征工程
 - 性能高效，计算代价小
- 较好的可迁移能力
 - Cross net是任务无关的

Figure 1: The Deep & Cross Network

CTR预估模型-演变发展

深度交叉网络模型：Deep & Cross



优势

- 较强的cross feature刻画能力
 - 自动构造有限高阶的交叉特征，无需人工特征工程
 - 性能高效，计算代价小
- 较好的可迁移能力
 - Cross net是任务无关的

Figure 1: The Deep & Cross Network

CTR预估模型-演变发展

深度交叉网络模型：Deep & Cross

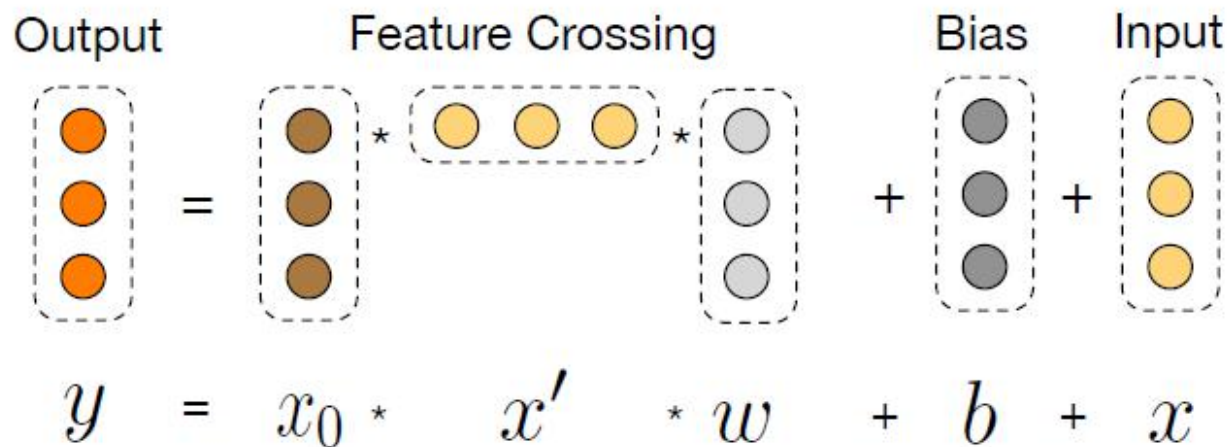


Figure 2: Visualization of a cross layer.

$$x_{l+1} = x_0 x_l^T w_l + b_l + x_l = f(x_l, w_l, b_l) + x_l,$$

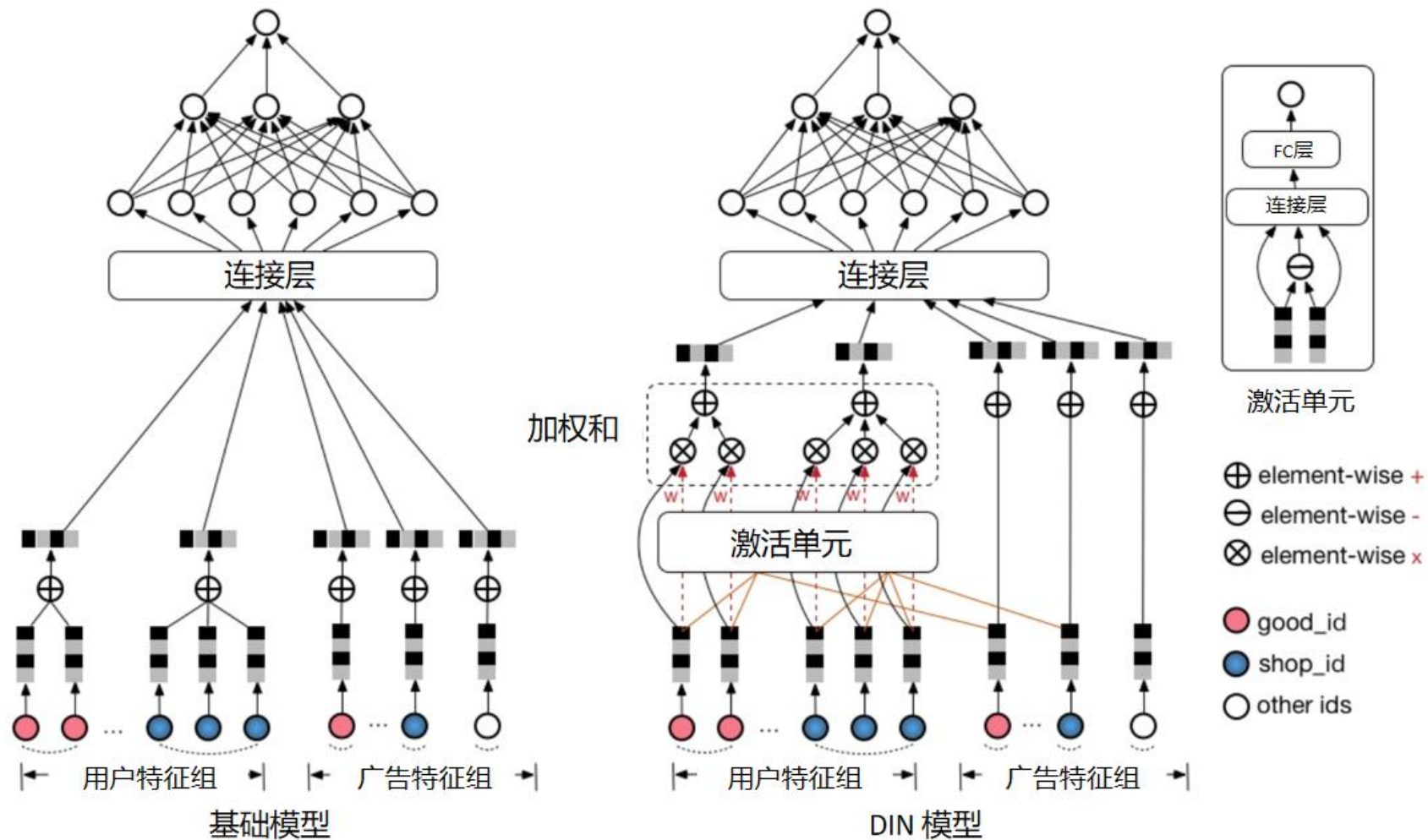


Cross Network

- 每层的神经元个数都相同，输入输出维度都是相同的
- 受残差网络（Residual Network）结构启发，每层函数拟合的其实是残差，可以缓解梯度消失，使网络深度更深
- 网络深度（L）决定了特征叉乘的阶数（L+1）
- 自动叉乘，Cross输出了原始特征从一阶到L+1阶的所有叉乘特征组合，而模型参数量仅仅随输入维度（d）成线性增长：2*d*L
- **参数共享**：不同叉乘项对应的权重不同，但并非每个叉乘组合对应独立的权重（指数数量级），通过参数共享，Cross有效降低了参数量。此外，参数共享还使得模型有更强的泛化性和鲁棒性

CTR预估模型-演变发展

兴趣网络模型：DIN和DIEN

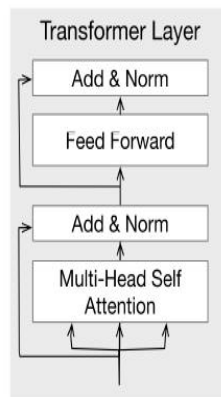
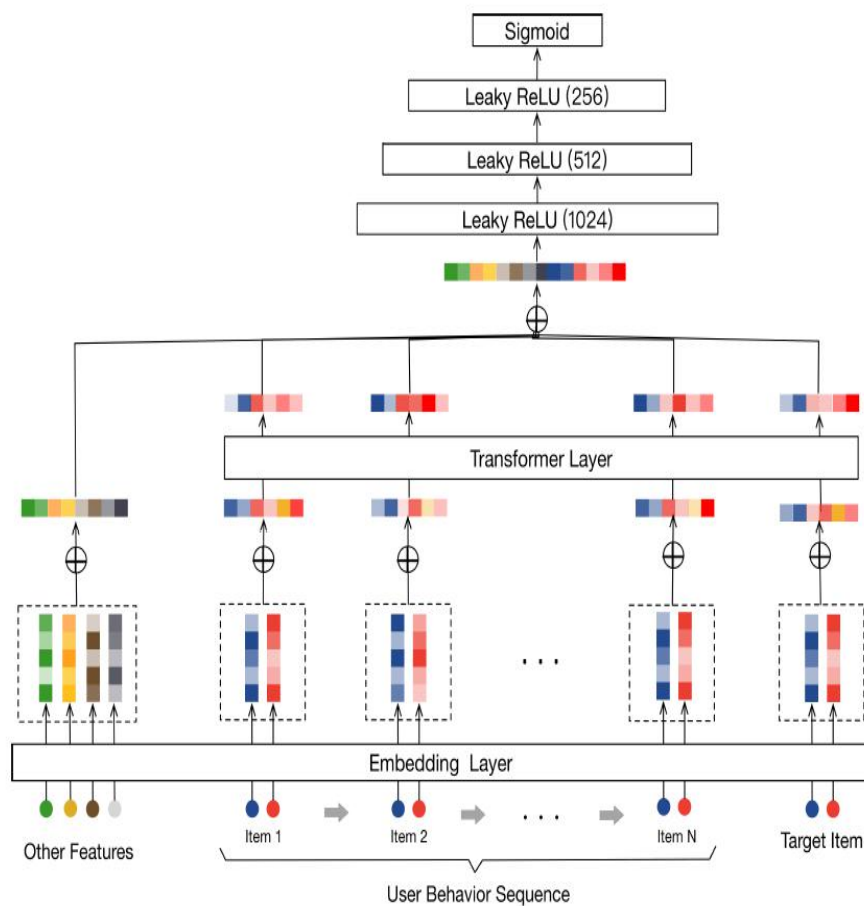


$$V_u = \sum_{i=1}^N w_i * V_i = \sum_{i=1}^N G(V_i, V_a) * V_i$$

首先是把u和v以及u v的
element wise差值向量合并起来
作为输入，然后喂给全连接层，
最后得出权重

CTR预估模型-演变发展

Behavior Sequence Transformer Model (BST)



捕捉用户行为序列的
序列信息

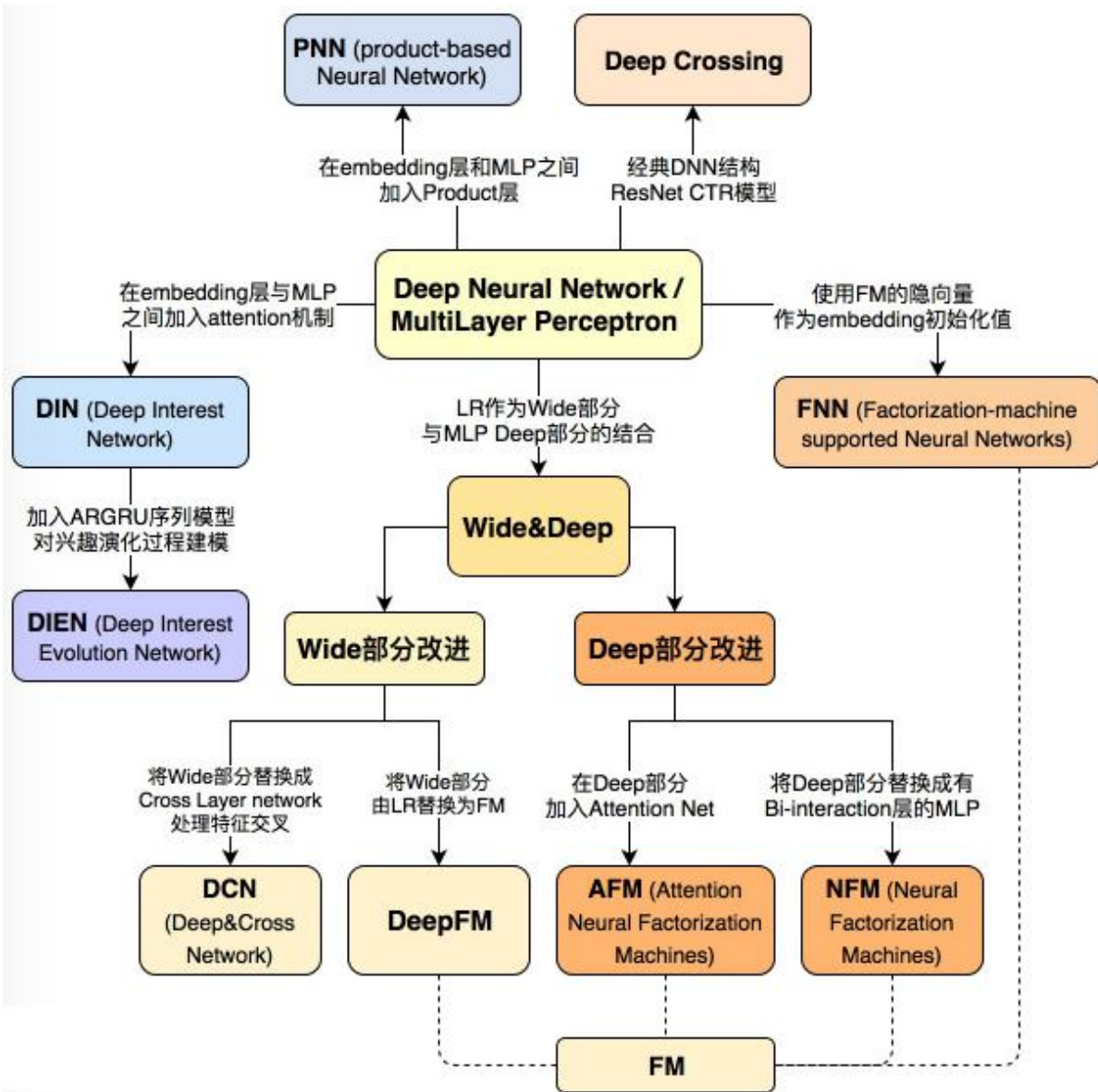
- \oplus Concatenate
- User Profile Features
- Item Features
- Context Features
- Cross Features
- Positional Feature
- Sequence Item Features

优势

较强的用户兴趣建模能力

Self-Attention能有效捕获目标Item和用户行为之间的关系，还能对用户行为序列中的序进行有效建模

总结思考



- 改变神经网络的复杂结构
- 改变特征的交叉方式
- 模型组合
- FM算法的深度模型演进
- 注意力机制和推荐模型的结合
- 序列模型与推荐模型的结合
- 强化学习与推荐算法模型的结合
- 多任务模型



<http://www.julyedu.com/employment/rs10>