

AiCon
全球人工智能与机器学习技术大会

FFM及DeepFFM模型 在推荐系统的探索

张俊林

新浪微博

主办方 **Geekbang** **InfoQ**
极客邦科技

TABLE OF

CONTENTS 大纲

- 大规模推荐系统
 - 线性排序模型：从LR到FFM模型
 - FFM模型改进版：双线性FFM (Bilinear-FFM) 模型
 - 深度排序模型：从Wide&Deep到xDeepFM模型
 - DeepFFM模型

推荐系统的应用



音乐推荐



电影推荐



商品推荐

微博推荐 / 排序任务应用场景



关系流Feed排序

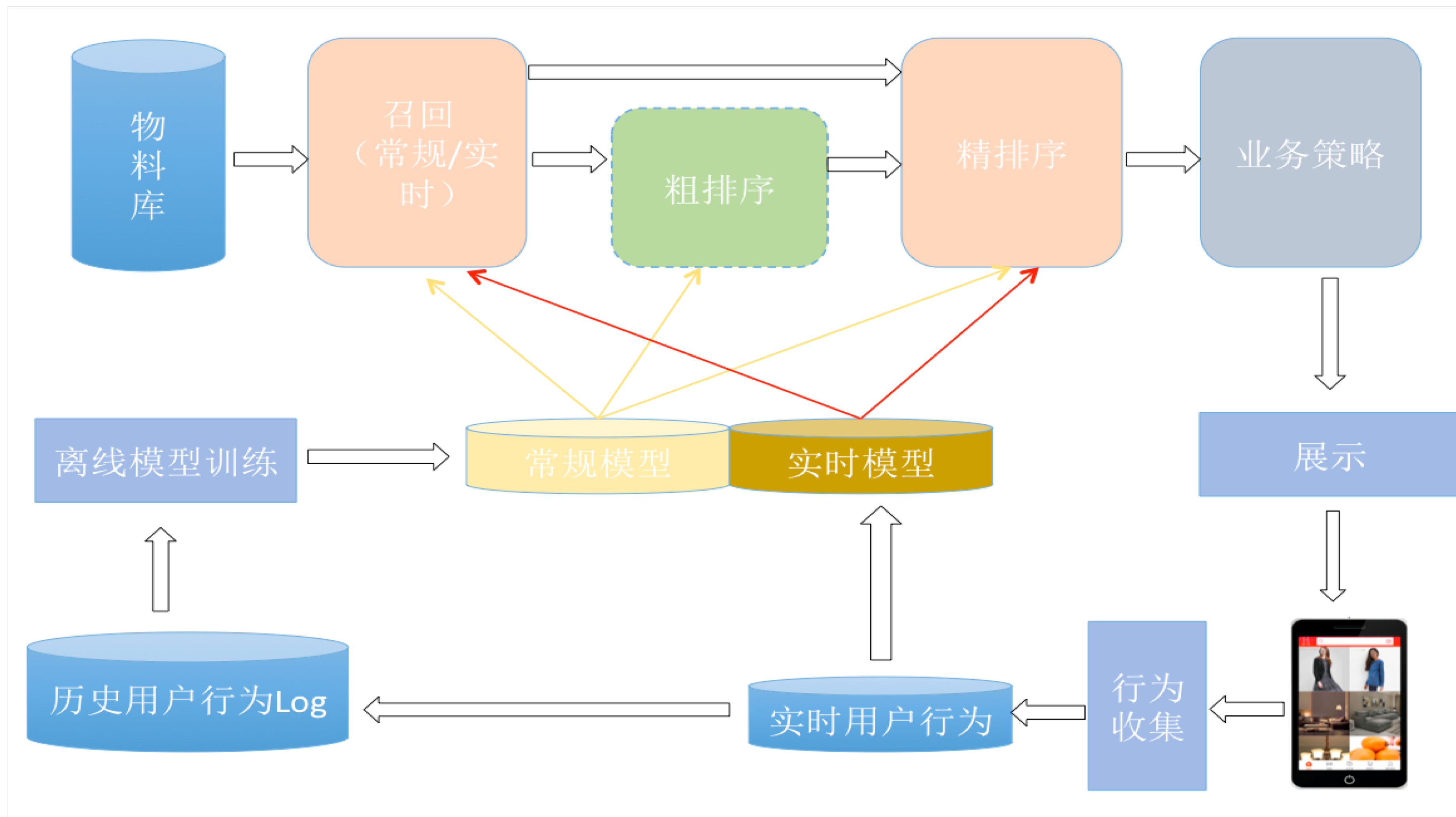


热门流个性化排序



正文页推荐

大规模推荐系统架构：以微博为例



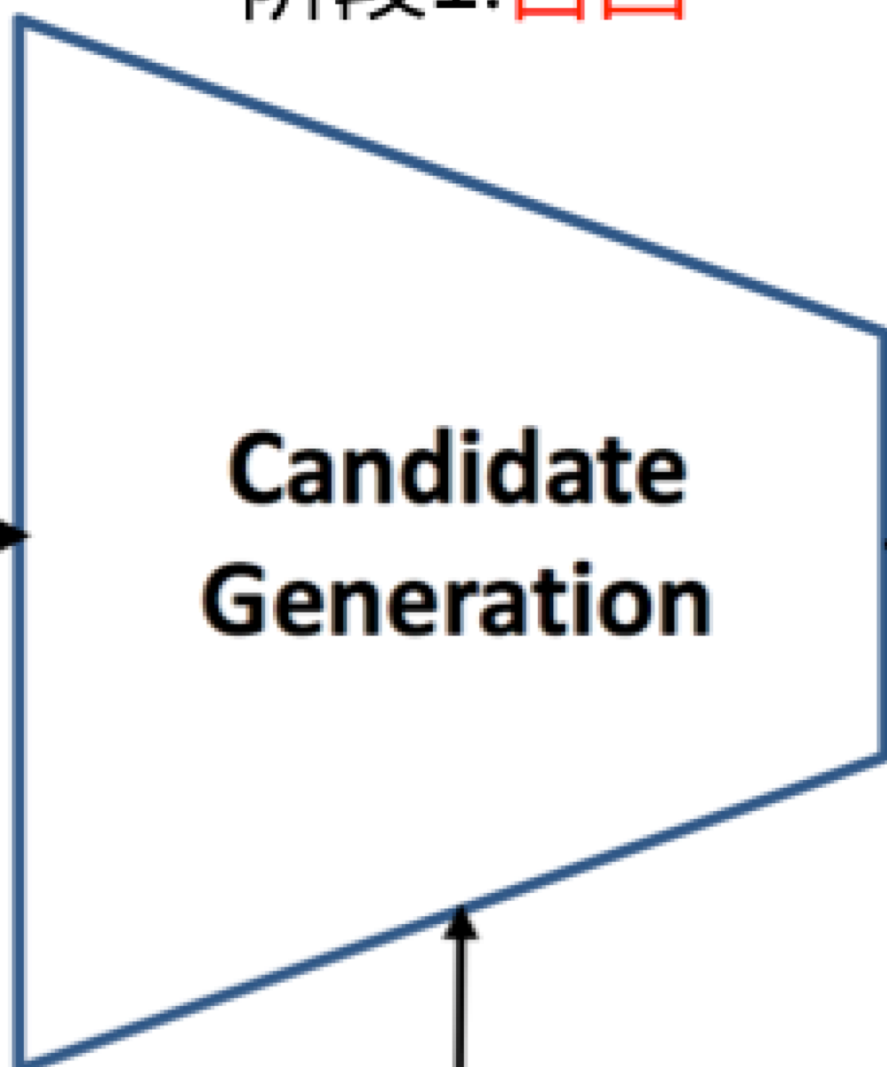
线上推荐系统的两个阶段

处理数据量大；
速度要够快；
模型不能太复杂；
使用较少特征；



Millions

阶段1:召回



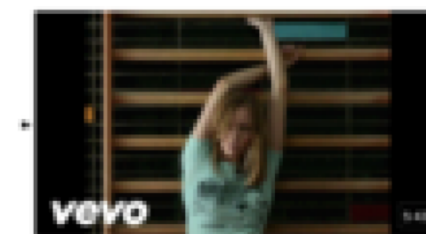
Hundreds

阶段2:排序



Tens

处理数据量少；
模型要够准；
可以上复杂模型；
可以使用较多特征；



⋮

今天我们主题在这里！

模型：

基于表示学习

基于特征组合

特征源：

User History and Contexts

All other Side Info

排序模型：工业界算法的演进路线

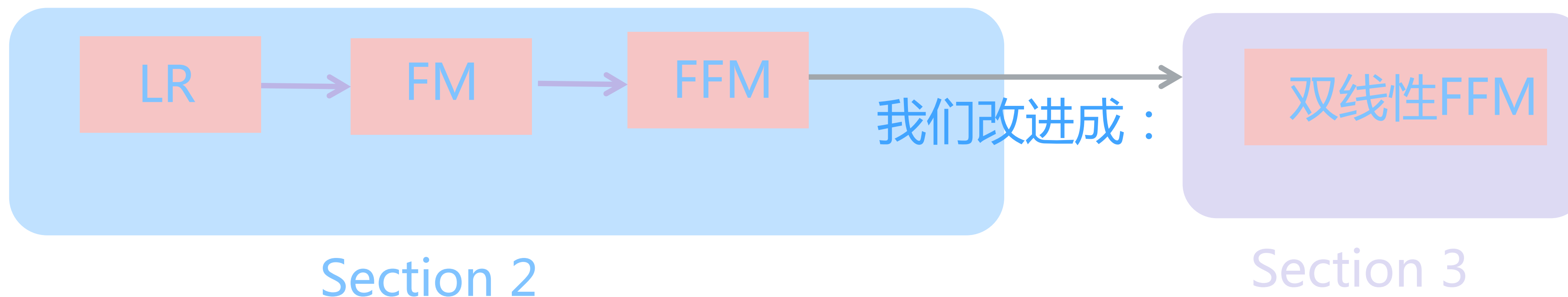


“最关键的是特征及特征组合能力！”

特征组合 特征1：年龄="80s"，特征2：性别="男"，特征3：三级标签="AngelaBaby"，组合特征="80s_男_AngelaBaby"

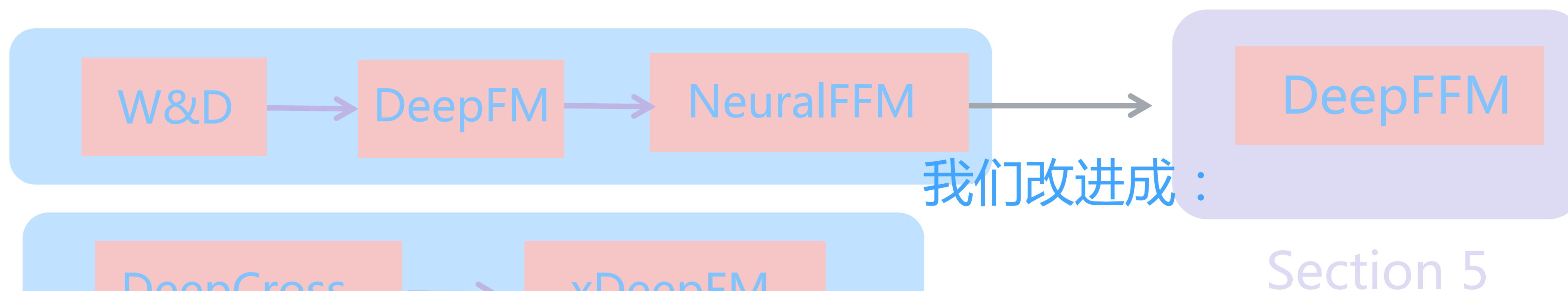
本次分享技术路线图

线性排序模型：



Deep排序模型：

演化分支1:



演化分支2:

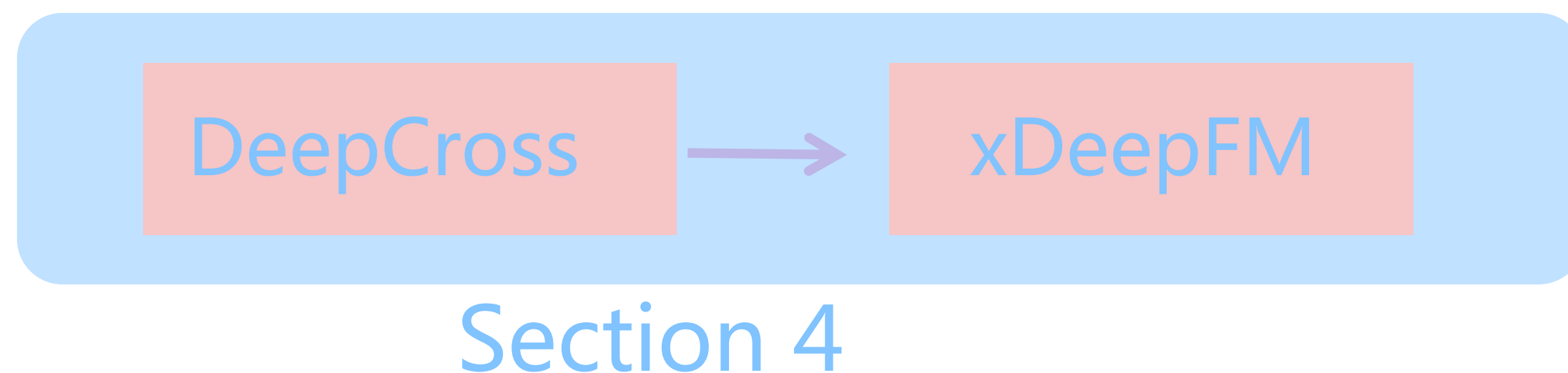


TABLE OF

CONTENTS 大纲

- 大规模推荐系统
- 线性排序模型：从LR到FFM模型
- FFM模型改进版：双线性FFM (Bilinear-FFM) 模型
- 深度排序模型：从Wide&Deep到xDeepFM模型
- DeepFFM模型

线性模型：思路及问题

$$\text{Linear: } \hat{y}(x) := w_0 + \sum_{i=1}^n w_i x_i$$

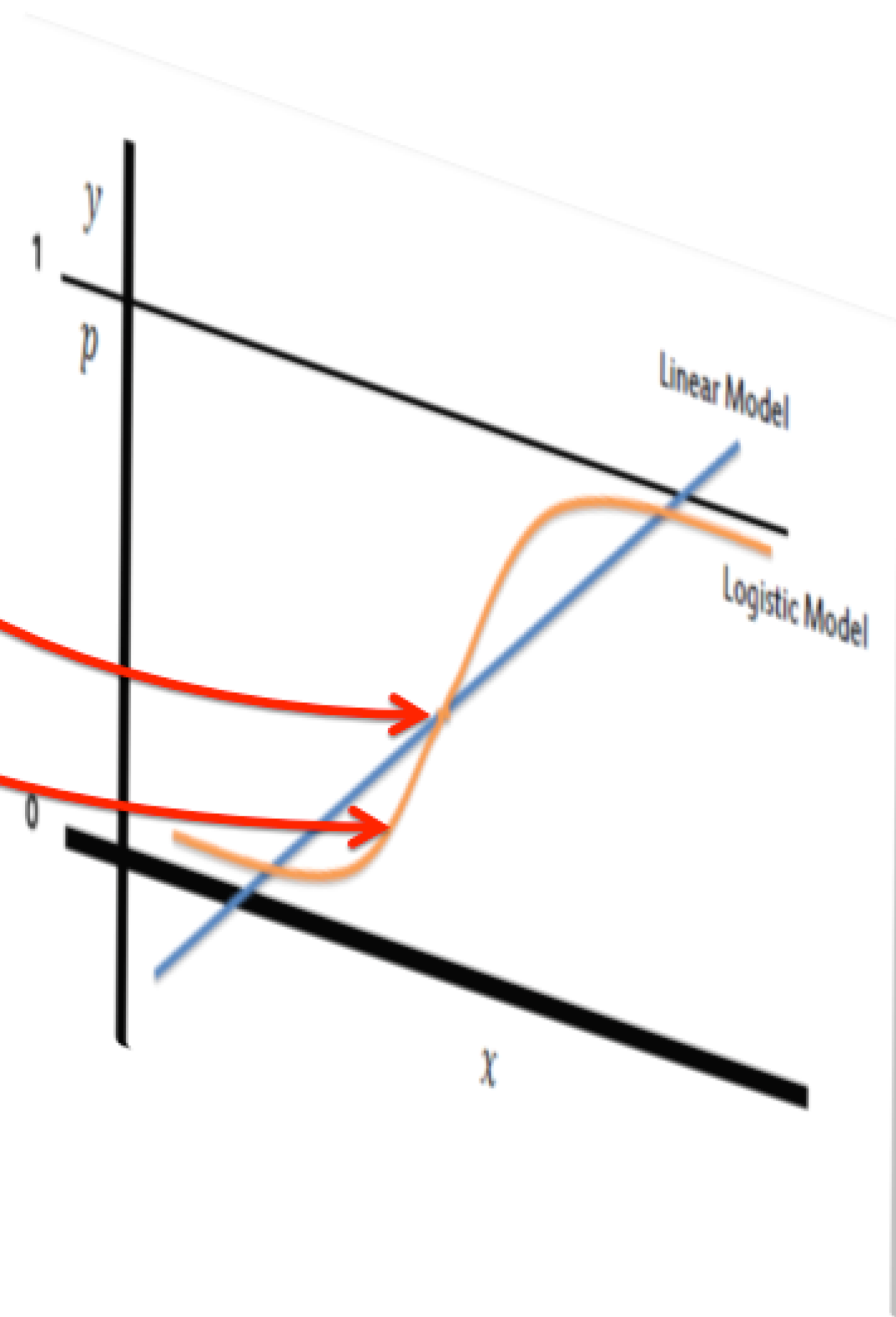
$$\text{LR: } \hat{y}(x) = \frac{1}{1 + w_0 \exp(-w^T x)}$$

优势：

简单；可解释；易扩展

缺点：

难以捕获特征组合



线性模型改进：加入特征组合

改进版本：

$$\hat{y}(x) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n w_{i,j} x_i x_j$$

两两特征组合

优势：

直接将两两组合特征引入模型

缺点：

组合特征泛化能力弱

$$w_{i,j} = 0 \quad \text{if 在训练数据中 } x_i x_j = 0$$

这其实就是SVM

FM模型

$$\text{FM: } \hat{y}(x) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j$$

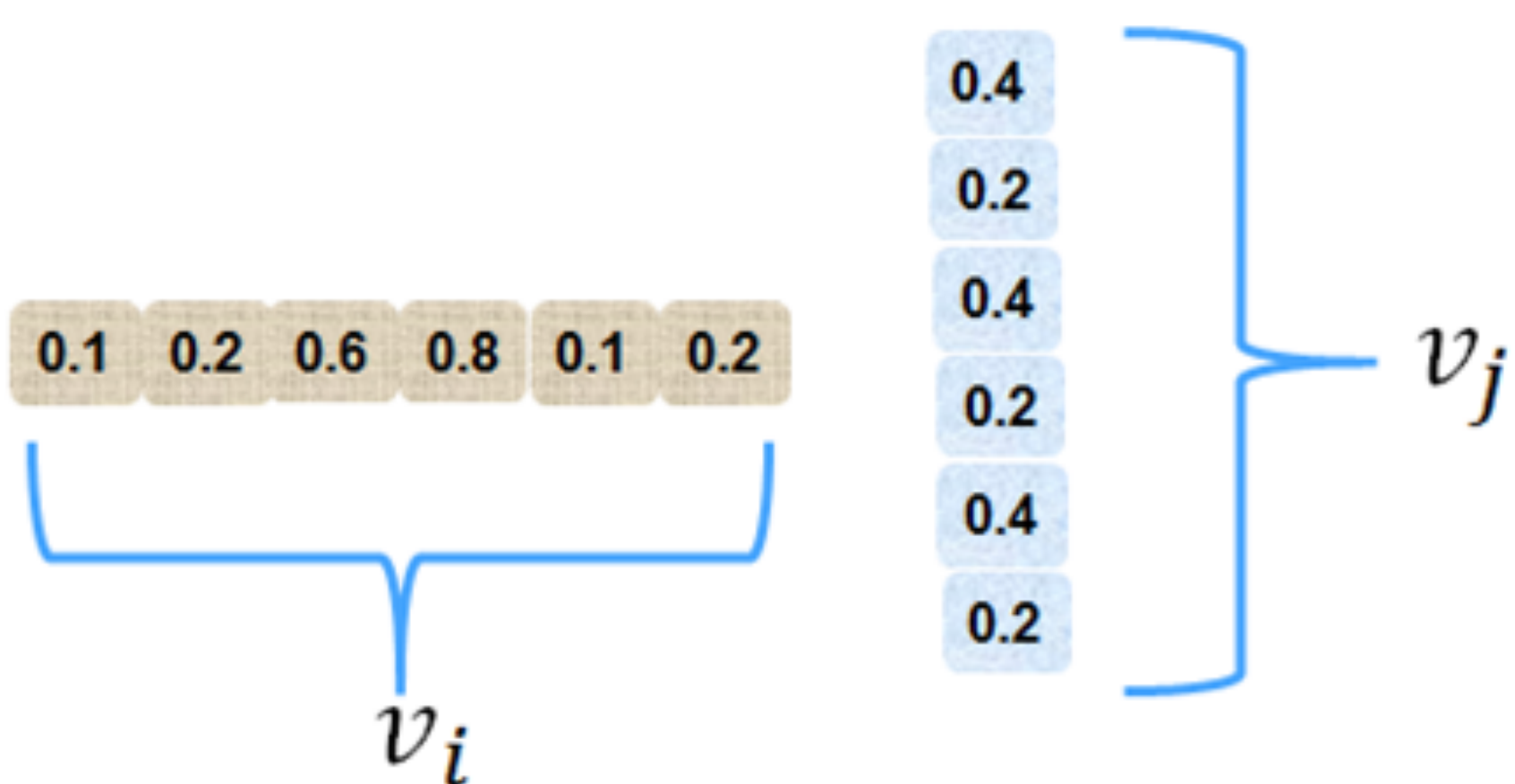
LR模型

Dense化两两特征

$$\langle v_i, v_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f} =$$

v_i v_j

FM模型

$$\langle v_i, v_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f} =$$


v_1	0.3	0.2	0.6	0.8	0.1	0.2
v_2	0.1	0.8	0.6	0.8	0.4	0.6
v_3	0.4	0.2	0.7	0.2	0.1	0.2
v_4	0.1	0.2	0.6	0.8	0.5	0.2
v_{n-1}	0.3	0.2	0.6	0.8	0.1	0.2
v_n	0.5	0.8	0.9	0.8	0.4	0.6



$$w_{i,j} = \langle v_i, v_j \rangle \neq 0$$

even if 训练数据中 $x_i x_j = 0$

only if 在训练数据中存在k使得 $x_i x_k \neq 0$



FM模型泛化能力强

FFM模型：基本思想

FM: $\phi_{FM}(V, x) = \langle V_{ESPN}, V_{Nike} \rangle + \langle V_{ESPN}, V_{Male} \rangle + \langle V_{Nike}, V_{Male} \rangle$

V_{ESPN} 0.1 0.2 0.6 0.8 0.1 0.2

一个特征一个向量

Clicked	Publisher (P)	Advertisor (A)	Gender (G)
Yes	ESPN	Nike	Male

FM是FFM的特例；
FFM更细致地刻画特征；

V_{ESPN}

0.2 0.6 0.1 0.3 0.7 0.9
0.3 0.2 0.8 0.2 0.1 0.5

一个特征(F-1)个向量

FFM: $\phi_{FFM}(V, x) = \langle V_{ESPN,A}, V_{Nike,P} \rangle + \langle V_{ESPN,G}, V_{Male,P} \rangle + \langle V_{Nike,G}, V_{Male,A} \rangle$

FFM模型

FM: $\langle v_i, v_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f} =$

0.1 0.2 0.6 0.8 0.1 0.2



参数扩大F倍 (F=特征fields个数)

FFM: $\langle v_{ij}, v_{ji} \rangle :=$

v_{ij}



v_{ji}

v_j

FFM模型

参数扩大F倍 (F=特征fields个数)

效果好于FM模型

实用化困境：参数量太大，太耗内存

如何改造？：效果接近于FFM；消耗内存远小于FFM模型

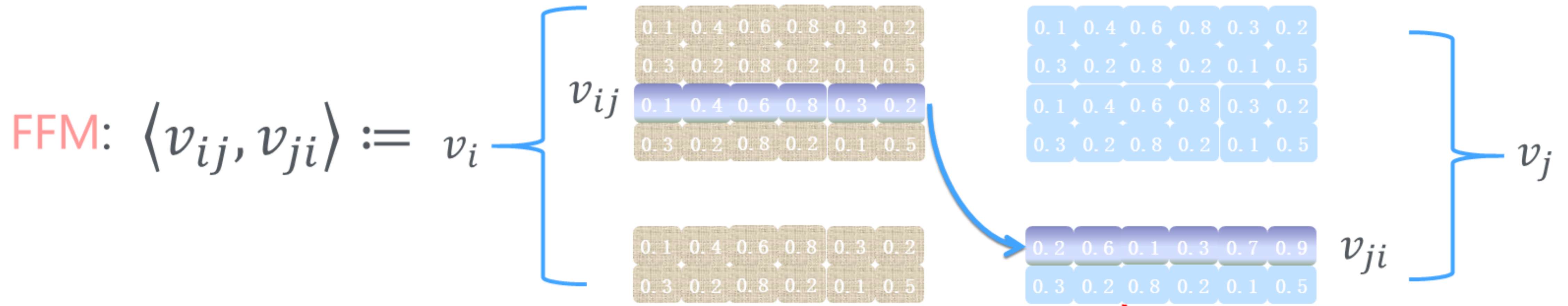
我们（张俊林&黄通文）提出的改进模型：双线性FFM

TABLE OF

CONTENTS 大纲

- 大规模推荐系统
- 线性排序模型：从LR到FFM模型
- **FFM模型改进版：双线性FFM（Bilinear-FFM）模型**
- 深度排序模型：从Wide&Deep到xDeepFM模型
- DeepFFM模型

双线性FFM模型：基本思路



Bilinear-FFM:

$$v_i W v_j :=$$



双线性FFM模型：3个W

Bilinear-FFM:

$$v_i W v_j :=$$



类型1: 所有特征共享W
新增参数量 = $k*k$

k : 特征embedding size

F : 特征Fields个数

类型2: 一个Field一个 W_i
新增参数量 = $F*k*k$

类型3: 一个Field组合一个 W_{ij}
新增参数量 = $F*F*k*k$

Option: 可以加入Layer Norm

双线性FFM模型：效果对比

Criteo和Avazu是两个工业级的CTR数据

Model Name	Criteo		Avazu	
	AUC	Logloss	AUC	Logloss
LR	0.7808	0.4681	0.7633	0.3891
FM	0.7923	0.4584	0.7745	0.3832
FFM	0.8001	0.4525	0.7795	0.3810
Bi-FFM-ALL	0.7935	0.4573	0.7754	0.3830
Bi-FFM-EACH	0.7963	0.4550	0.7738	0.3838
Bi-FFM-INTER	0.7995	0.4525	0.7781	0.3820
Bi-FFM-ALL-LN	0.8004	0.4518	0.7763	0.3837
Bi-FFM-EACH-LN	0.8007	0.4511	0.7745	0.3843
Bi-FFM-INTER-LN	0.8035	0.4484	0.7765	0.3829

一些结论：

1. FM显著好于LR；
2. FFM好于FM；
3. 双线性FFM可以在大量减少参数情况下效果接近于FFM；
4. 通常情况下，LN有助于提升效果

微博机器学习团队正在大规模工程化：双线性FFM

双线性FFM模型：新增参数量对比

以Criteo数据集合为例子（4500万数据）：

特征数量：230万

特征Field数量：39

假设特征embedding大小：10

FFM参数量：

$$230万 * 39 * 10 = 8.97亿$$

相差38倍！

双线性FFM参数量：

$$FM : 230万 * 10 = 2300万$$

$$类型1: 新增 $10 * 10 = 100$$$

$$类型2 : 新增 $39 * 10 * 10 = 3900$$$

$$类型3 : 新增 $39 * 39 * 10 * 10 = 15万$$$

双线性FFM模型：总结

1.性能优于或者接近于FFM模型

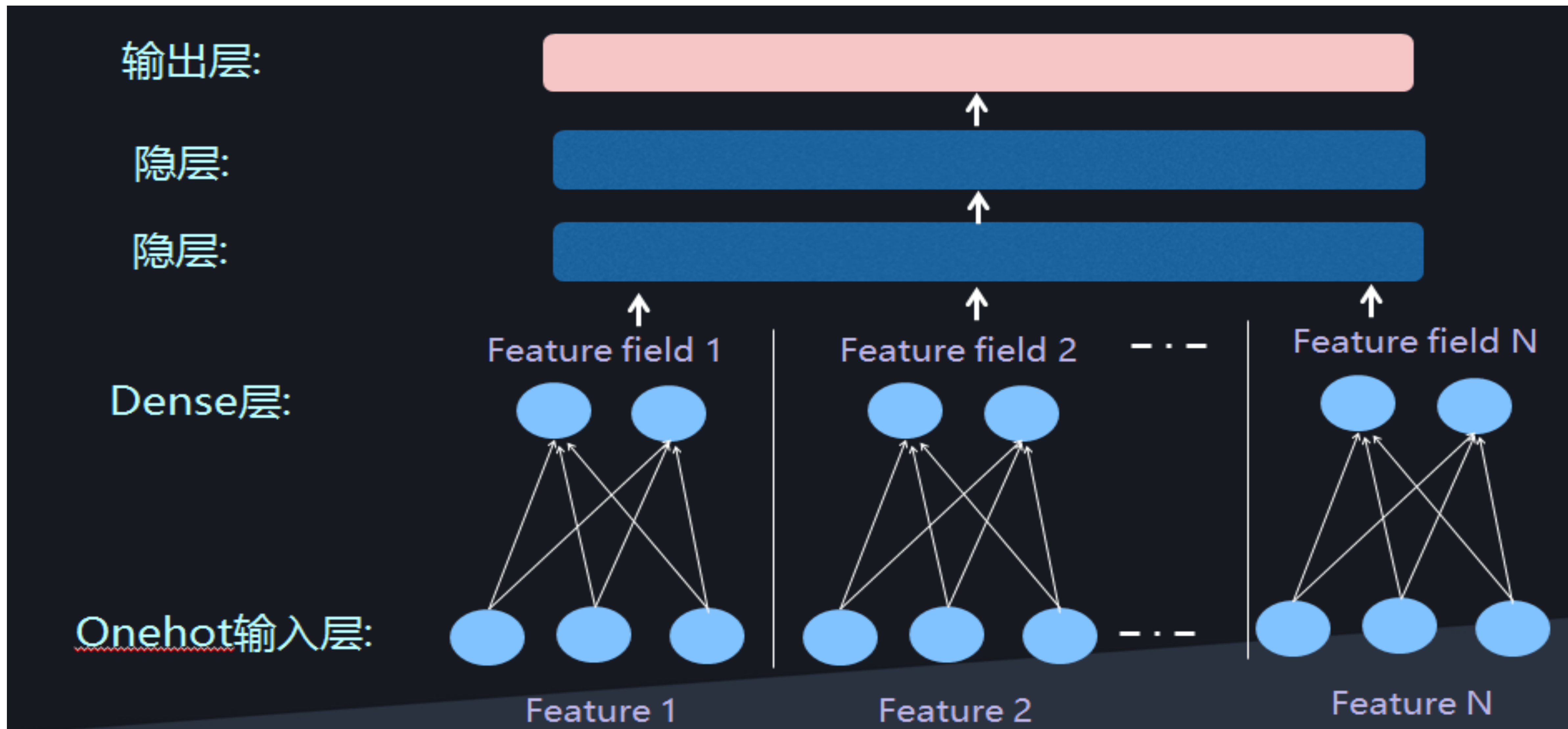
2.参数量是FFM模型的2.6%

TABLE OF

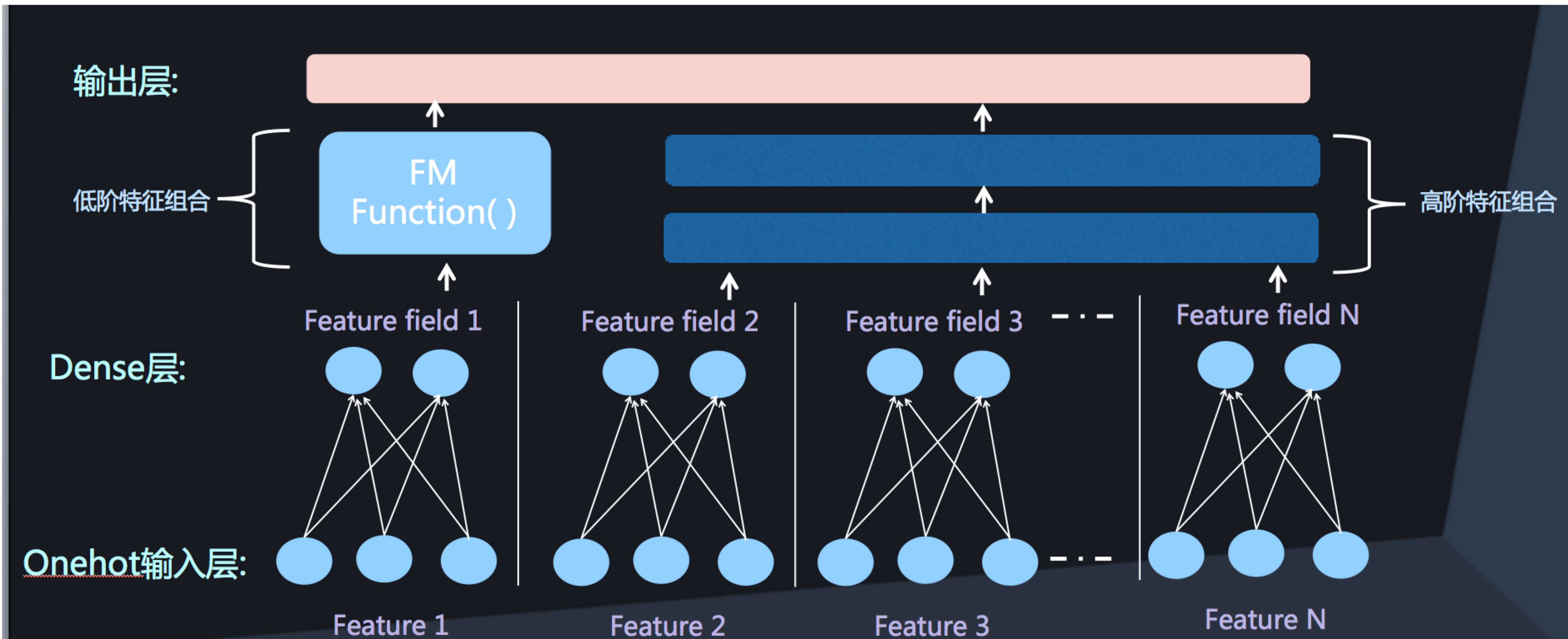
CONTENTS 大纲

- 大规模推荐系统
- 线性排序模型：从LR到FFM模型
- FFM模型改进版：双线性FFM (Bilinear-FFM) 模型
- 深度排序模型：从Wide&Deep到xDeepFM模型
- DeepFFM模型

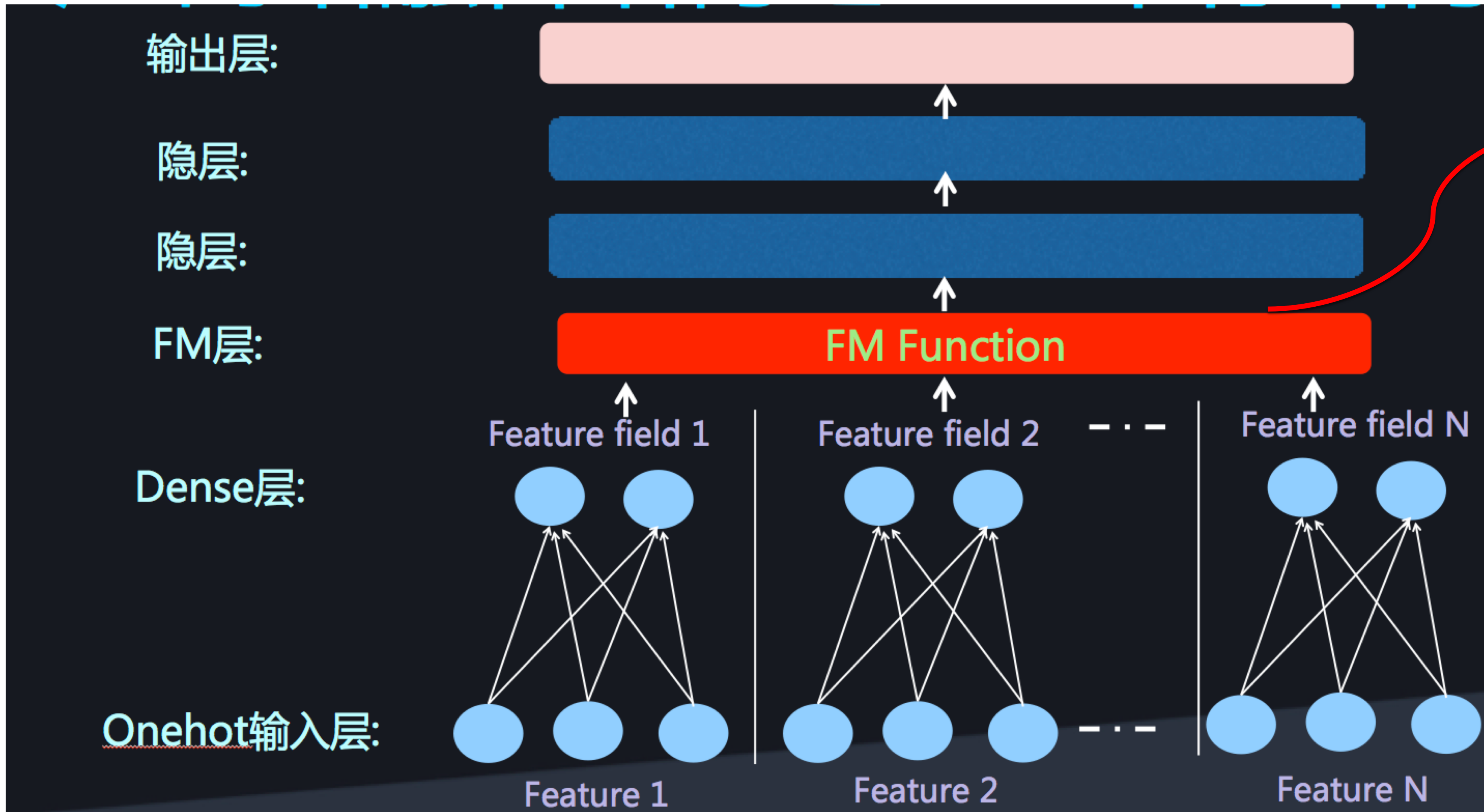
深度排序模型公共组件：DNN部分



深度排序模型（并行结构）：DNN+特征组合



深度排序模型（串行结构）：DNN+特征组合



PNN/NFM/AFM
等深度模型

深度排序模型的两条演进路线

如何更有效地捕获特征组合是核心，沿着以下两个演进路线发展：

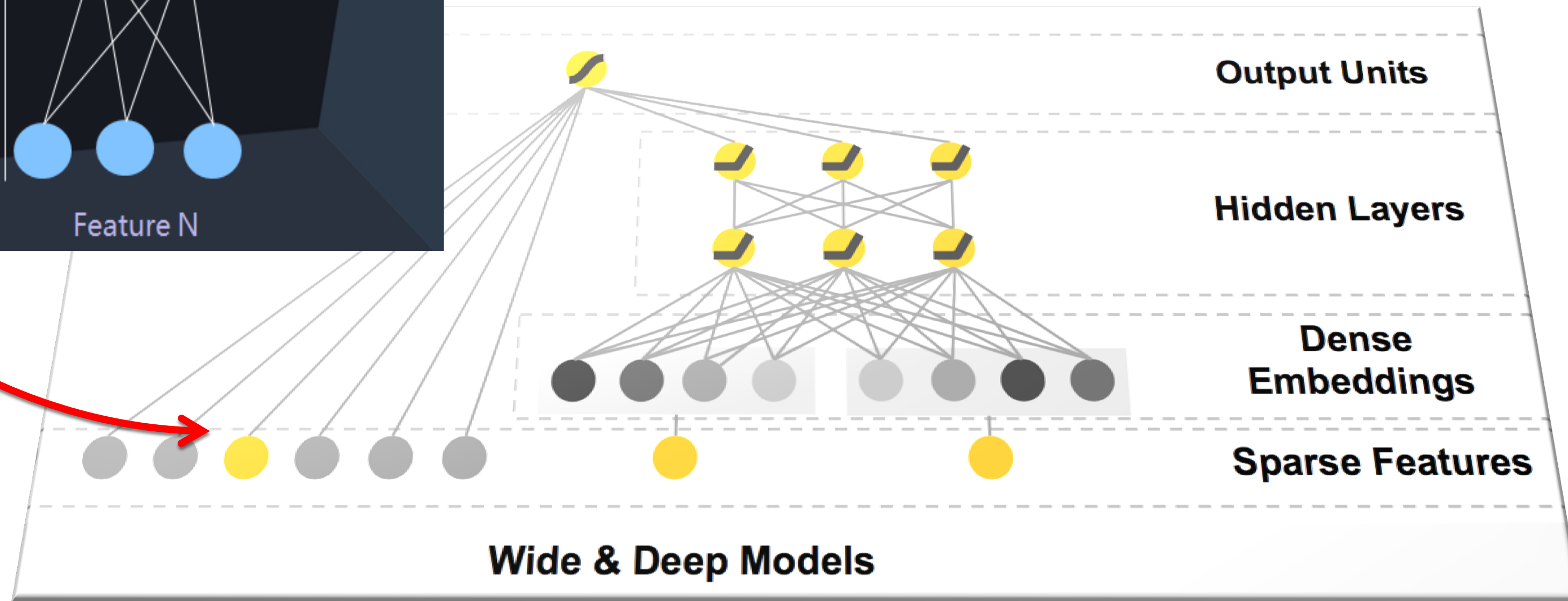
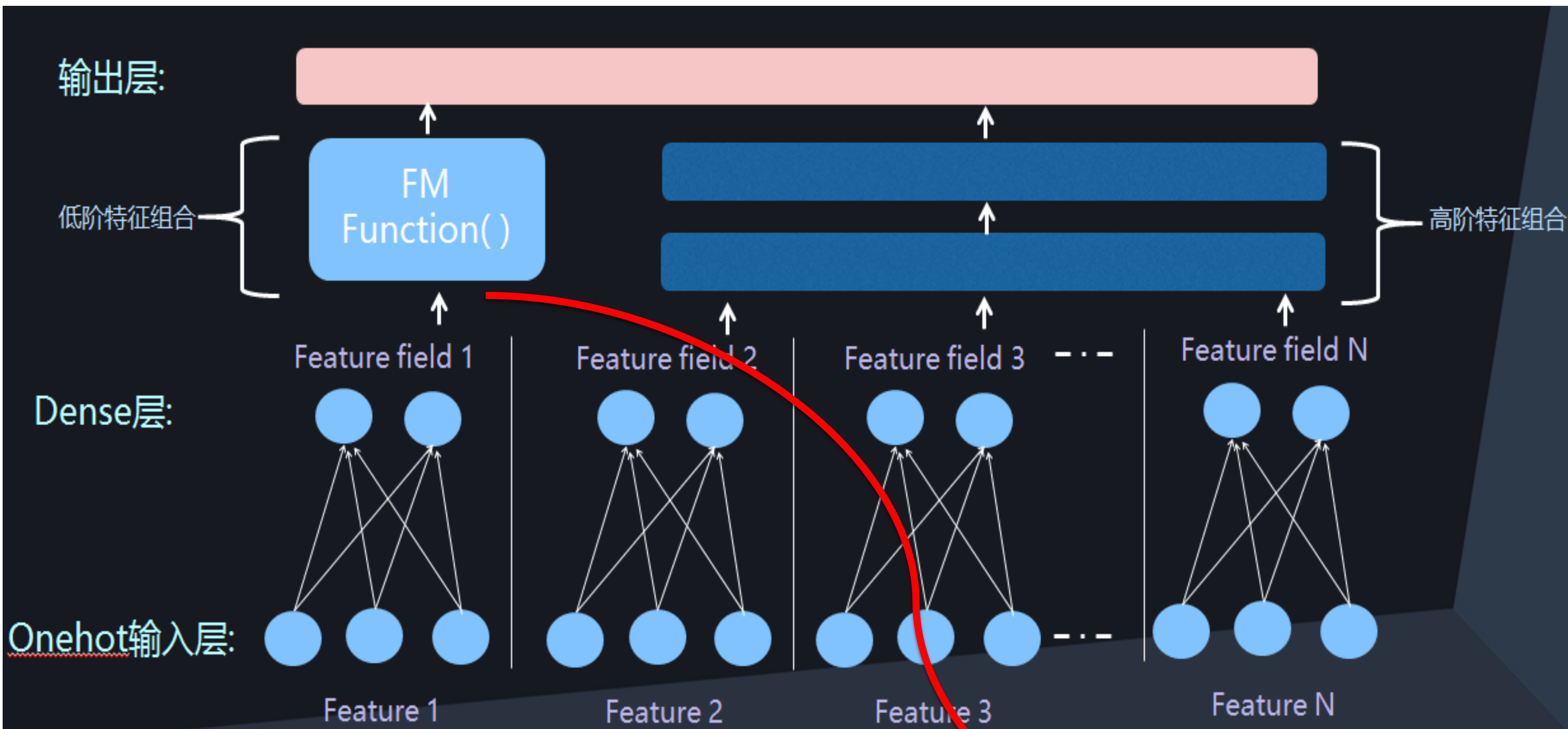
演进路线一：新型的FM function

- (1) 更有效地捕获二阶特征
- (2) Wide & Deep \rightarrow DeepFM \rightarrow NeuralFFM \rightarrow DeepFFM

演进路线二：显示建模2阶 / 3阶 / 4阶....K阶特征组合

- (1) 捕获高阶特征组合，获得额外受益；
- (2) 目前研究表面，2阶，3阶，4阶有正向受益
- (3) 再高阶组合没什么用
- (4) DeepCross \rightarrow xDeepFM

新型FM Function代表：Wide & Deep



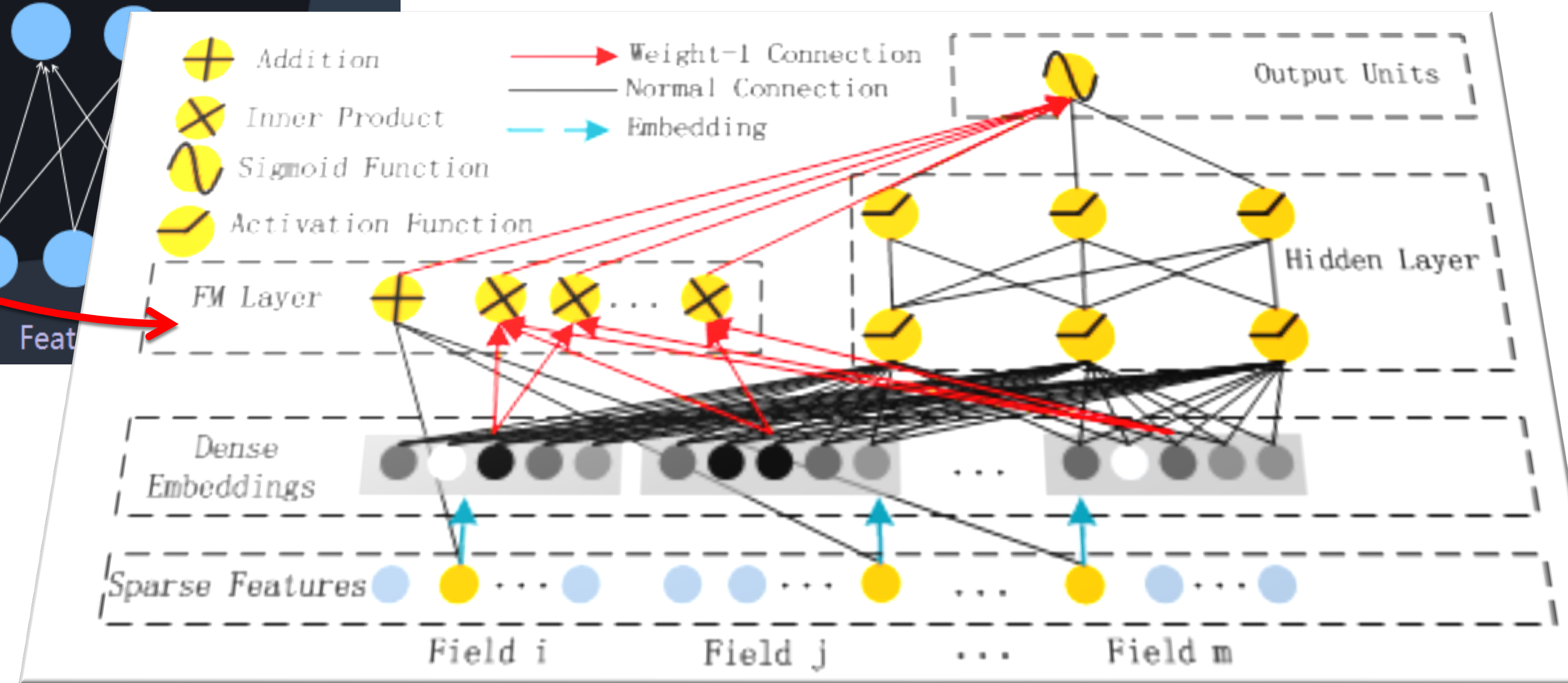
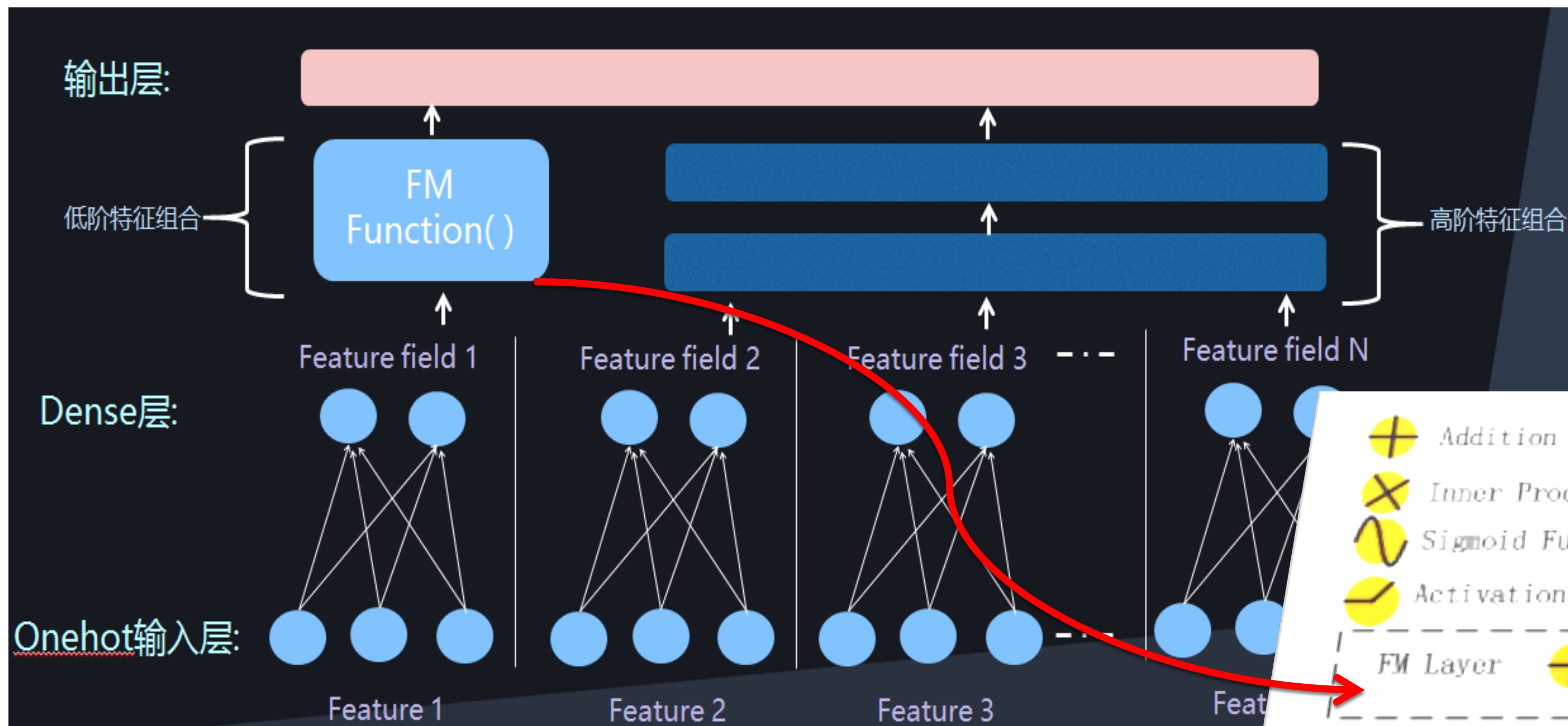
Wide网络LR模型，无FM组合

两个不同的输入结构

Wide部分手工特征工程+Cross 特征

Wide & Deep是相对原始的深度模型

新型FM Function代表：DeepFM

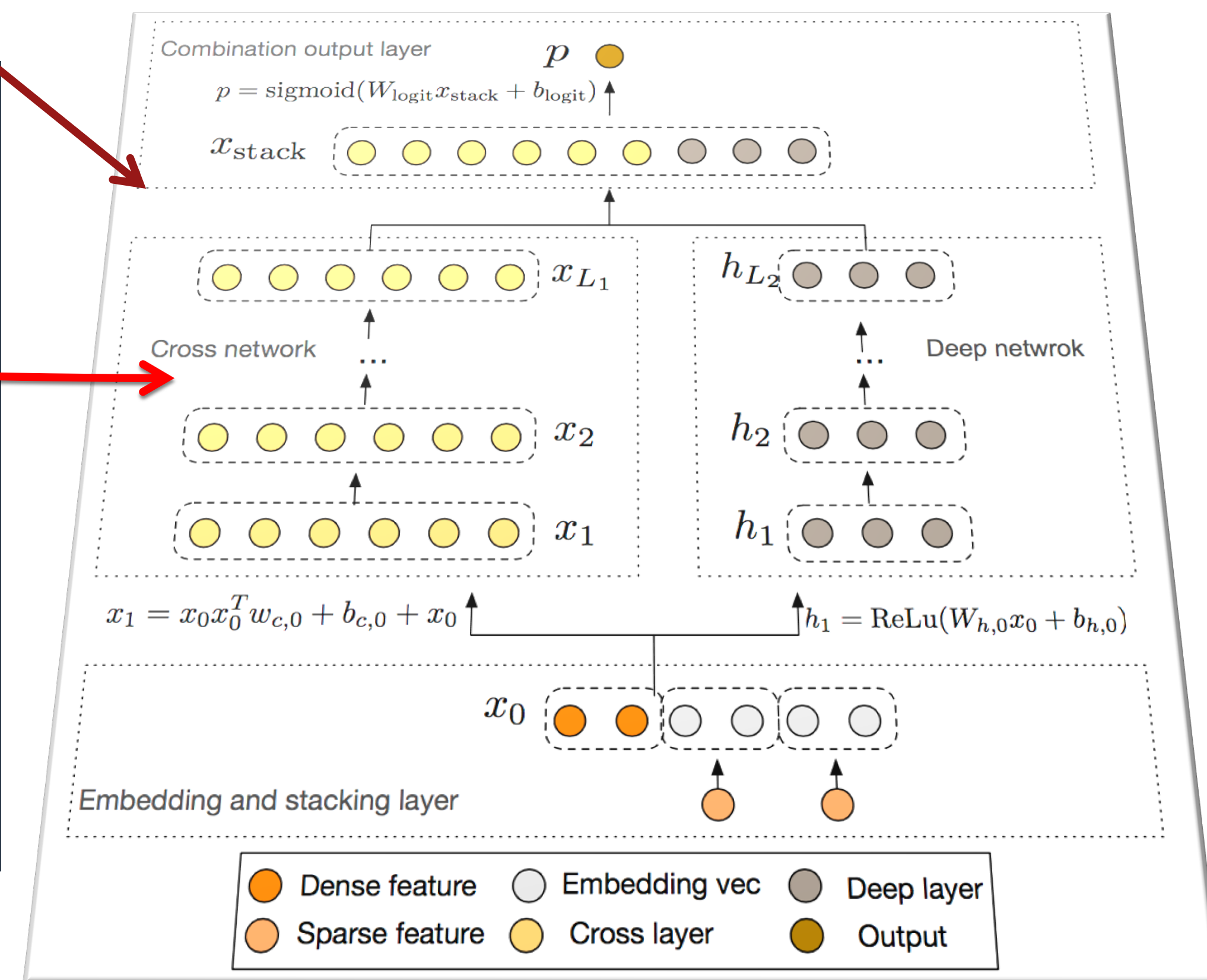
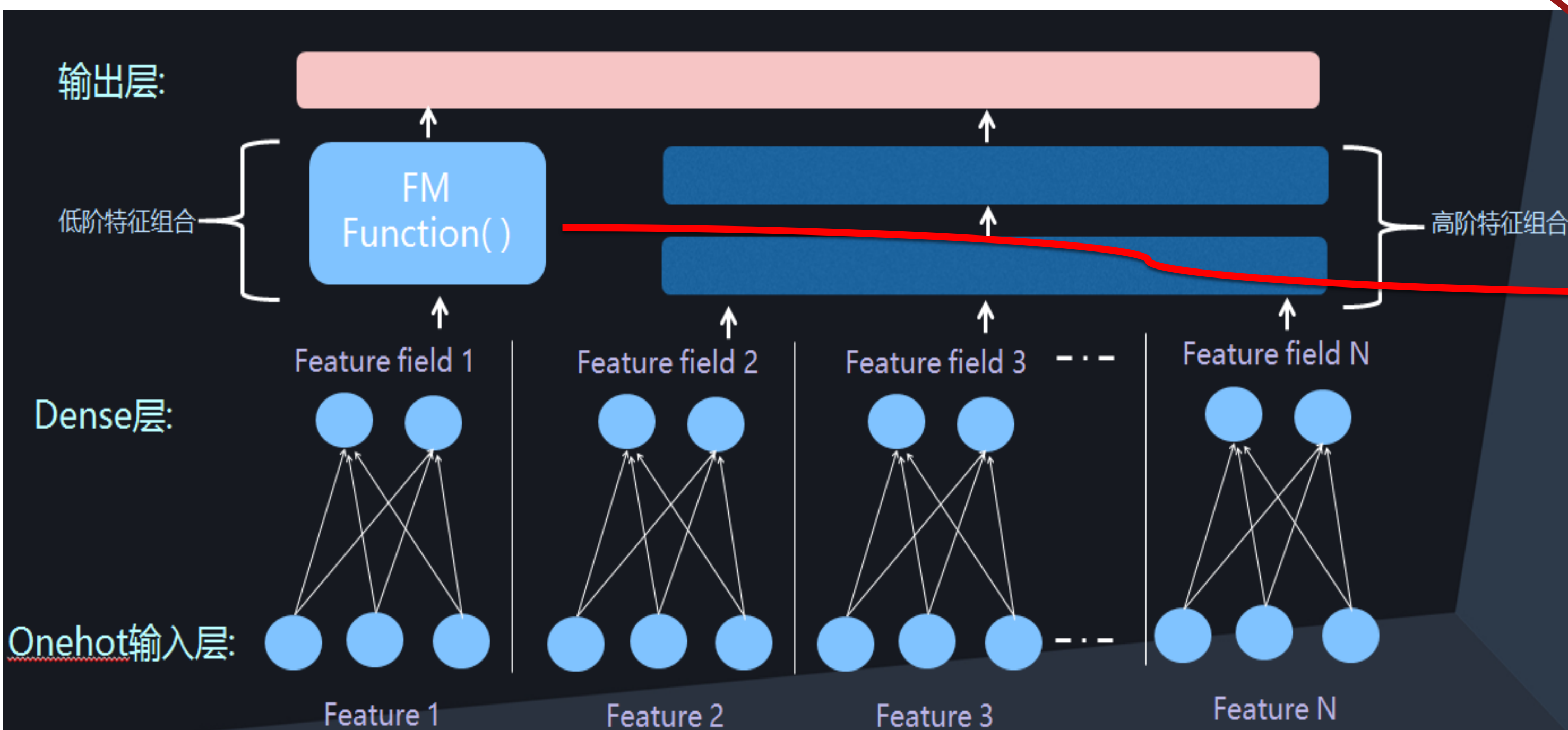


DeepFM:
可以看作把Wide & Deep的LR换成FM;
特征组合部分+DNN部分共享输入层;

目前State of the Art 深度模型

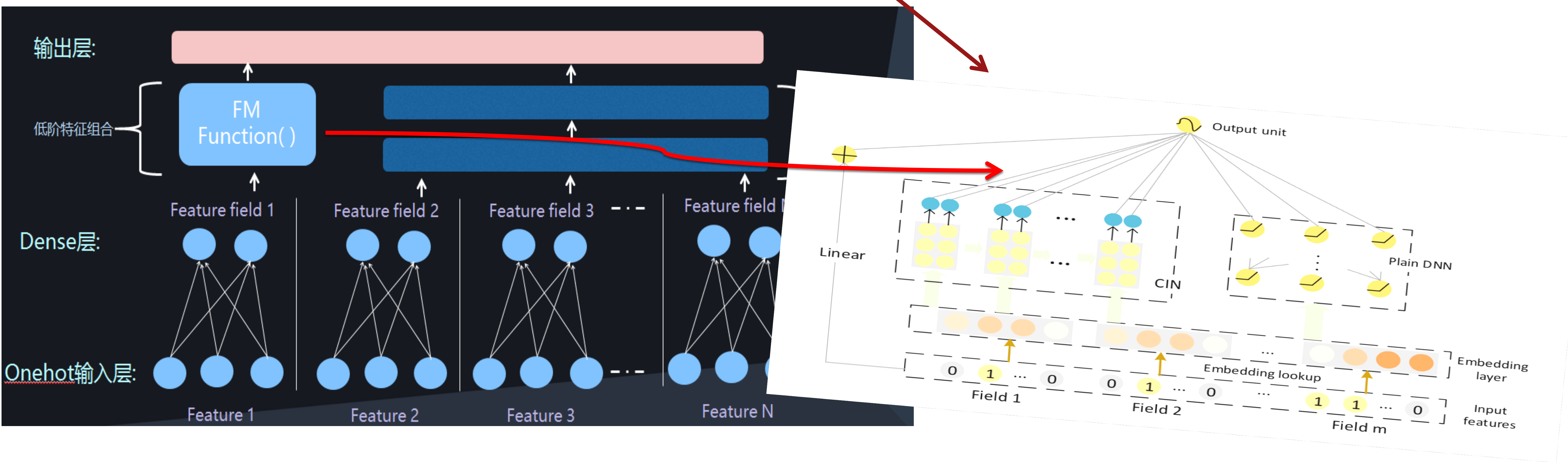
显示高阶组合：Deep & Cross

FM Function=Cross Network=ResNet



显示高阶组合：xDeepFM

FM Function=多阶组合=类CNN的结构



关于深度CTR模型的一些个人看法

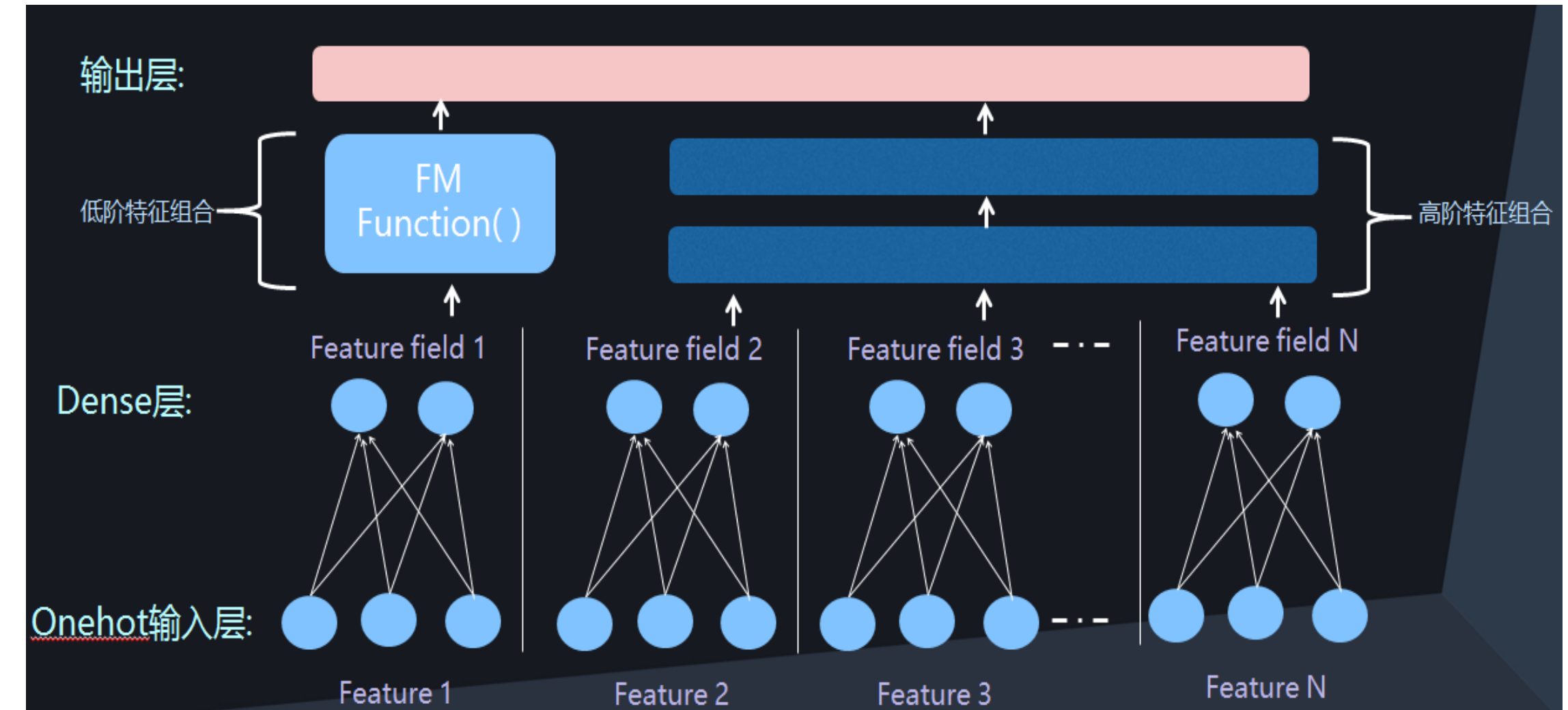
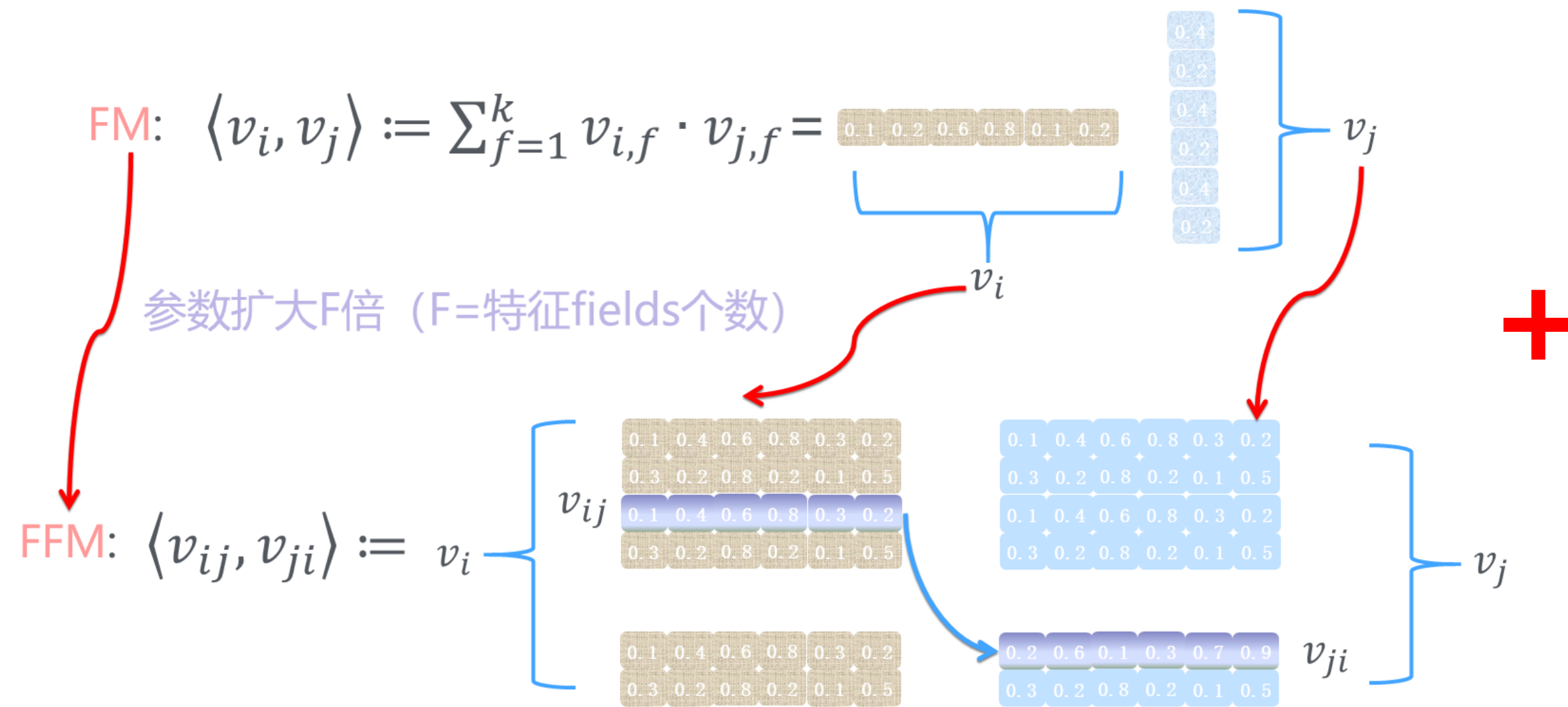
- 如下观点
 - 模型结构趋同：串行结构vs.并行结构；
 - 深度模型中的输入问题基本解决：[onehot](#)→[field embedding](#)；
 - **关键所在**:在于2-order特征组合网络结构如何设计([其实不同方案区别不大](#))；
 - 多特征组合分层表示是个趋势：2-order/3-order/4-order有效，[再高阶用处不大](#)
 - MLP这种加性捕获特征组合能力不强，乘性结构捕获组合特征比较合适；
 - 多模态融合是个趋势，也是DL绝对可以发挥作用的地方，但是需要较强的工程能力；

TABLE OF

CONTENTS 大纲

- 大规模推荐系统
- 线性排序模型：从LR到FFM模型
- FFM模型改进版：双线性FFM（Bilinear-FFM）模型
- 深度排序模型：从Wide&Deep到xDeepFM模型
- **DeepFFM模型**

FFM模型：如何改造成神经网络版本？



?

DeepFFM模型

NeuralFFM:

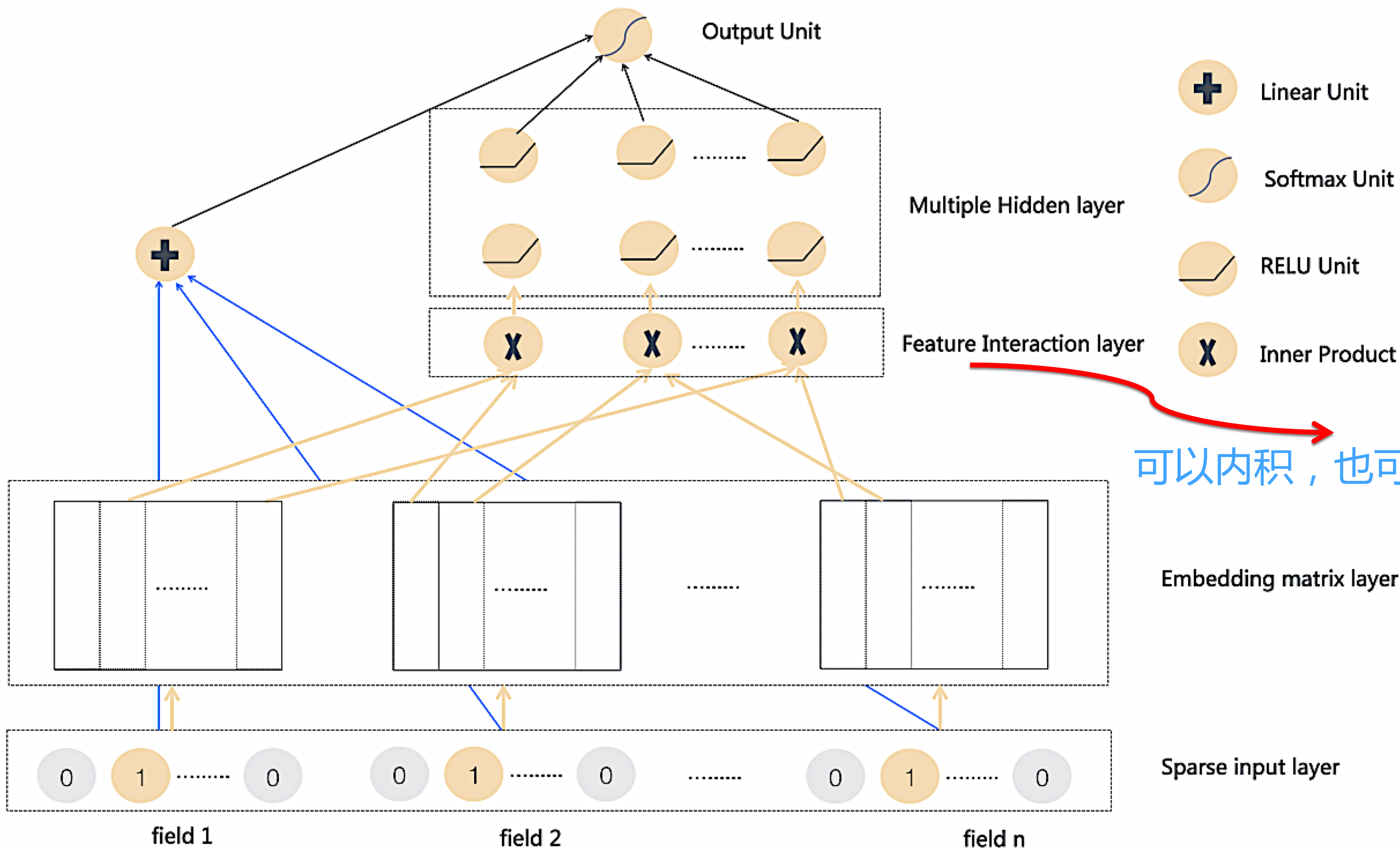
南京大学杨毅等提出；

在2017年参加腾讯社交竞赛中获得很好的
比赛名次（单模型第三名）；

DeepFFM:

微博（张俊林/黄通文，2018年10月）做了小改进；
引入Attention机制，称为DeepFFM模型；

FFM的神经网络版本：NeuralFFM



可以内积，也可以哈达马积

内积与哈达马积

向量内积：

$$\begin{bmatrix} 0.1 & 0.2 & 0.4 & 0.2 \end{bmatrix} * \begin{bmatrix} 0.3 & 0.2 & 0.1 & 0.5 \end{bmatrix} = 0.21$$

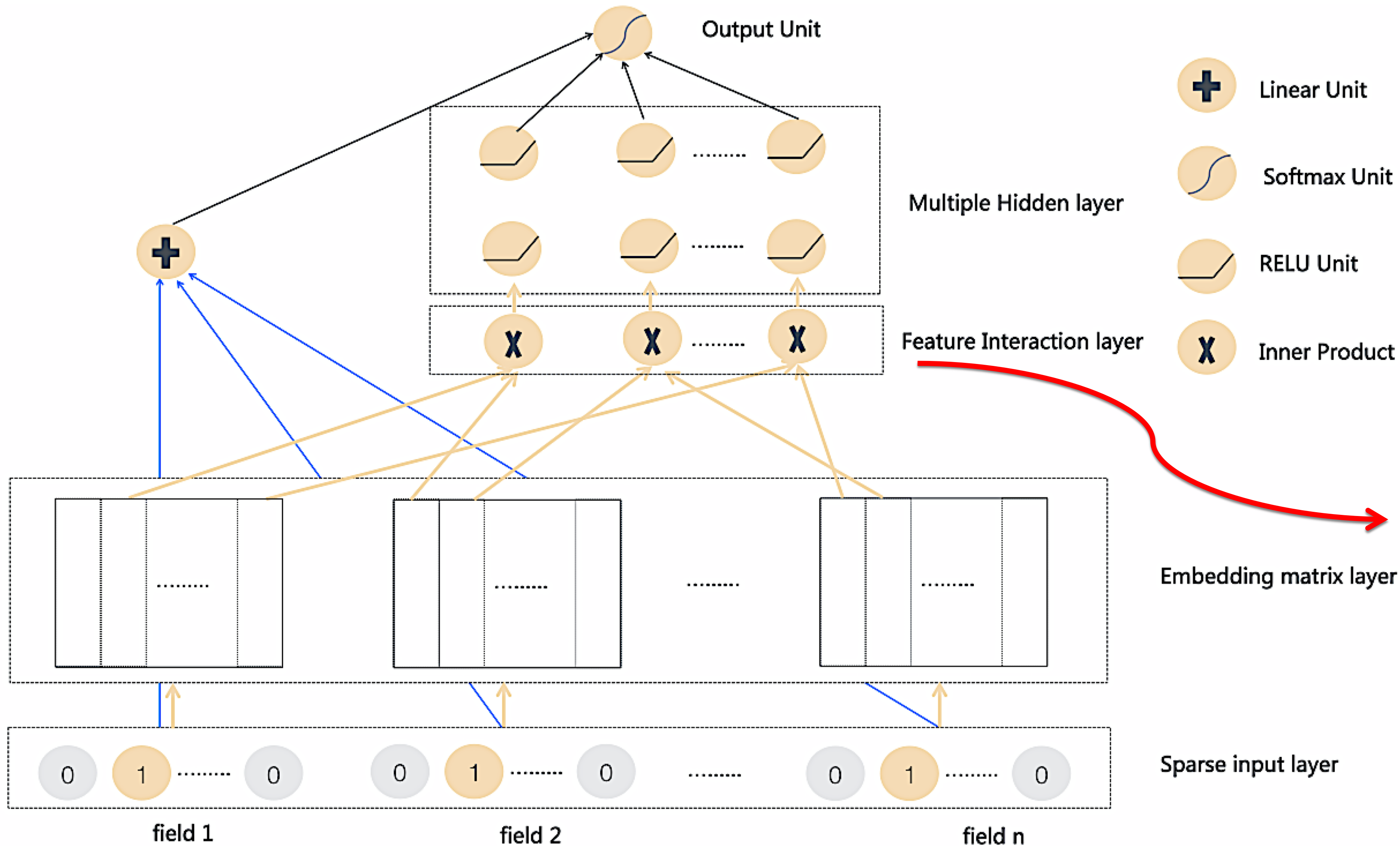
数值

哈达马积：

$$\begin{bmatrix} 0.1 & 2 & 0.4 & 2 \end{bmatrix} \times \begin{bmatrix} 3 & 0.2 & 1 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.3 & 0.4 & 0.4 & 1 \end{bmatrix}$$

向量

我们提出的改进版本：DeepFFM



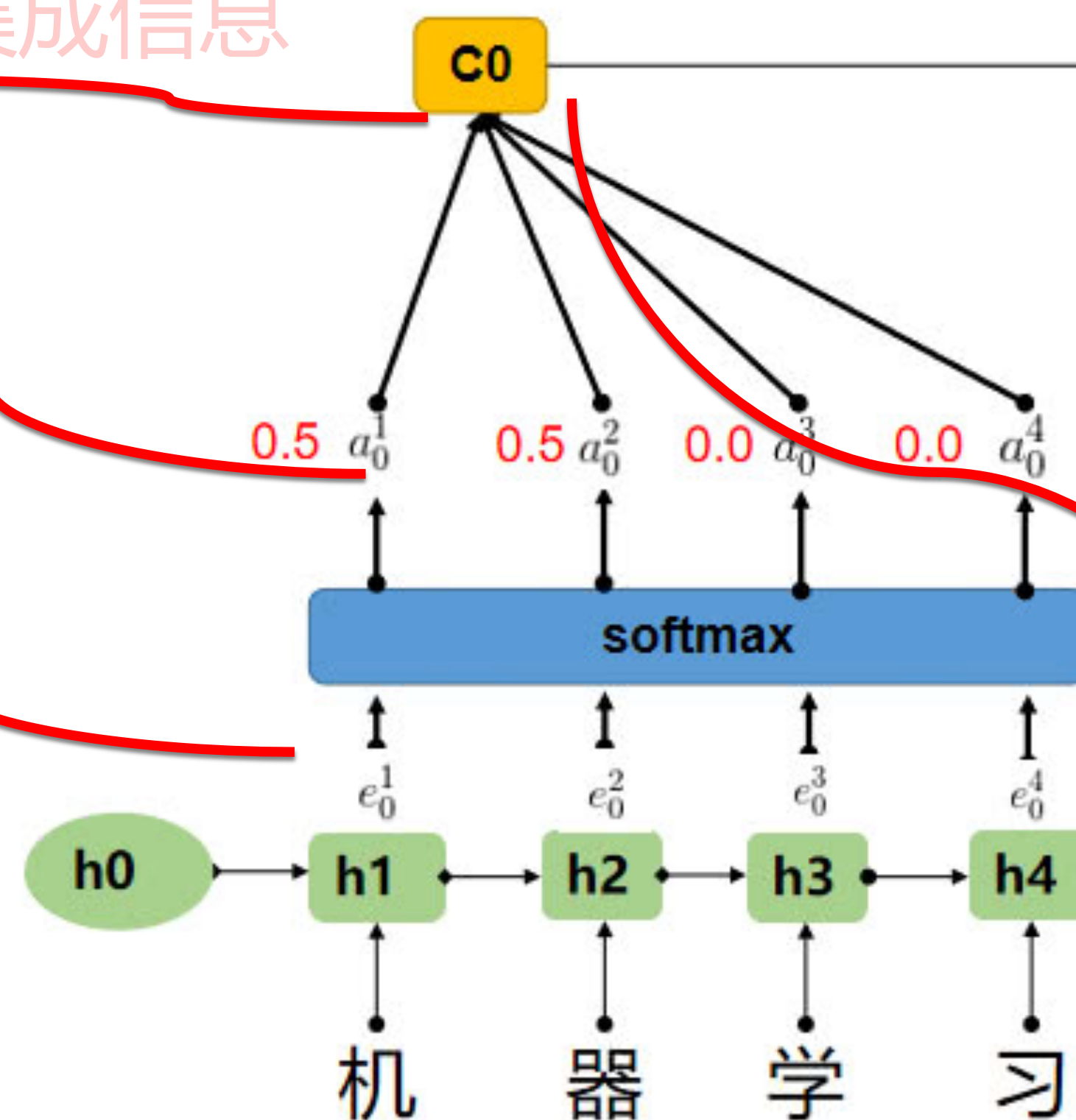
对NeuralFFM的小改进：
1.在特征交互层加入Attention
2.Attention代表对不同特征组合重要性判断；
3.特征交互层可以是内积，也可以是哈达马积

Attention：自动学习特征权重

3. 权重* 特征Value: 加权集成信息

2. 权重归一化

1. 学习特征权重



$$C_0 = \sum a_0^i h^i$$
$$= 0.5h^1 + 0.5h^2$$

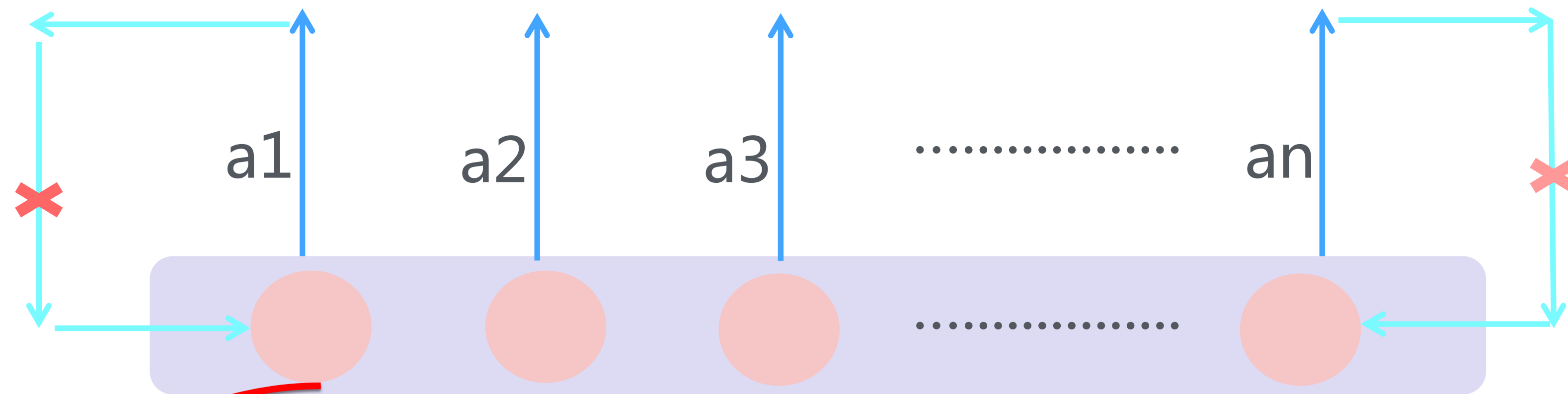
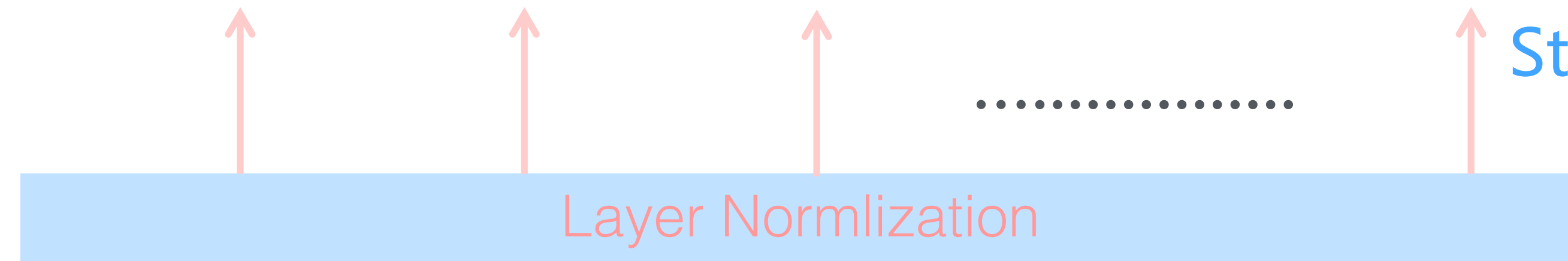
以Encoder-Decoder机器翻译为例子

DeepFFM-I : 特征交互层加入Attention

Step1:学习权重 a_1 — a_n ;

Step2:将权重乘以原始激活值;

Step3:Layer Norm规范化;

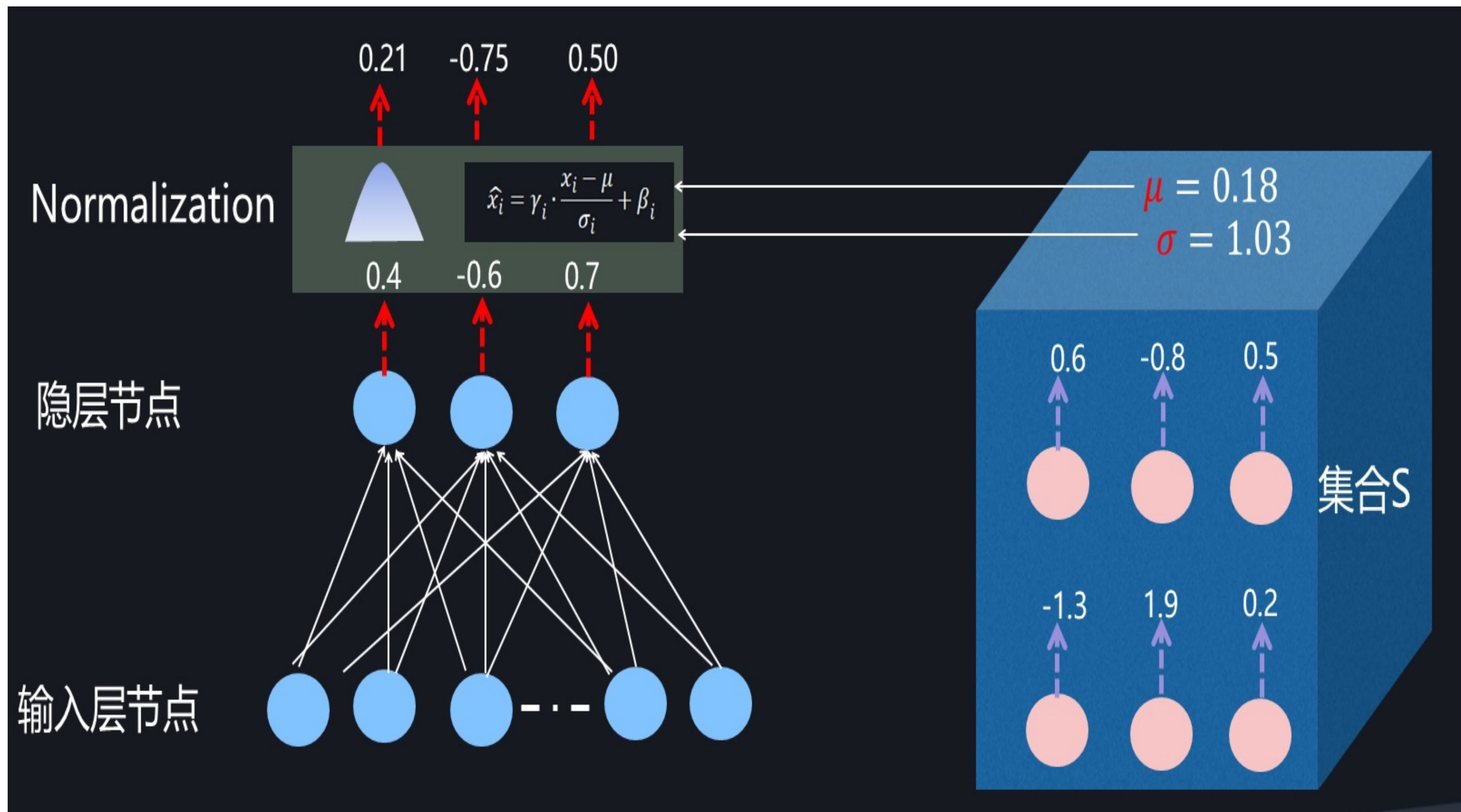


内积版Attention

特征交互层 :

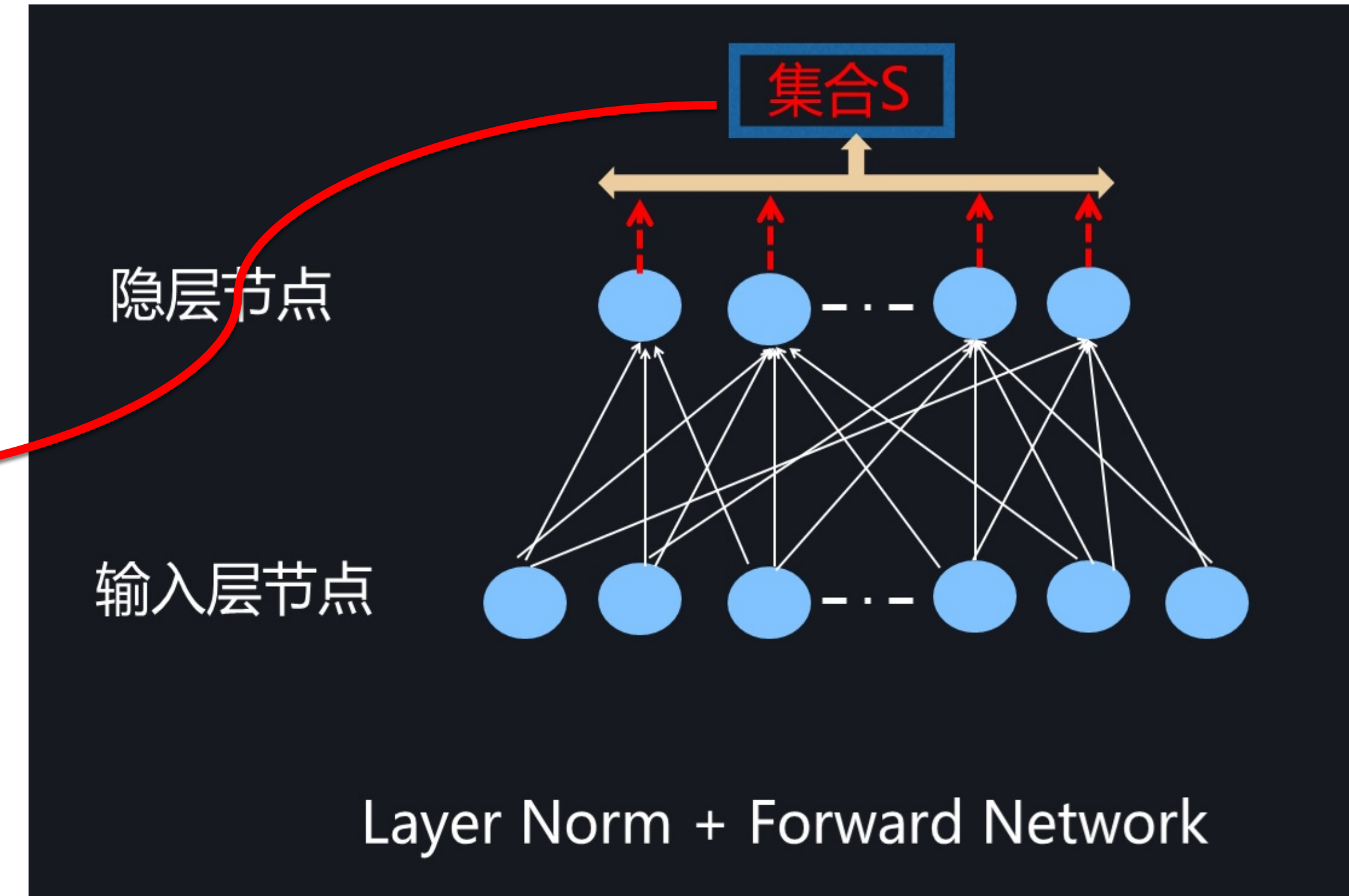
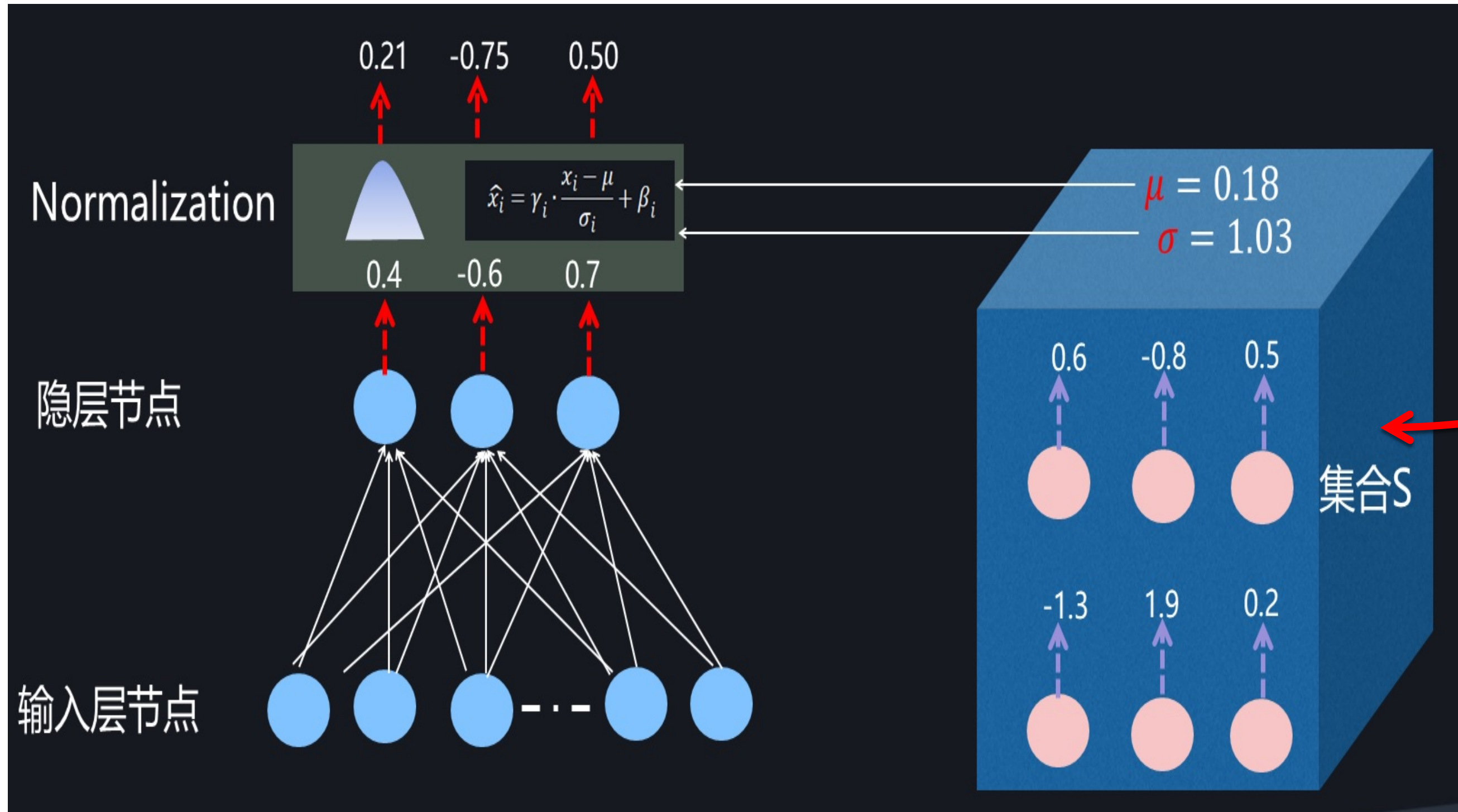
一个特征组合

DL中的Normalization方法



- BatchNorm
- LayerNorm
- GroupNorm
- InstanceNorm

DL中的Normalization方法:LayerNorm



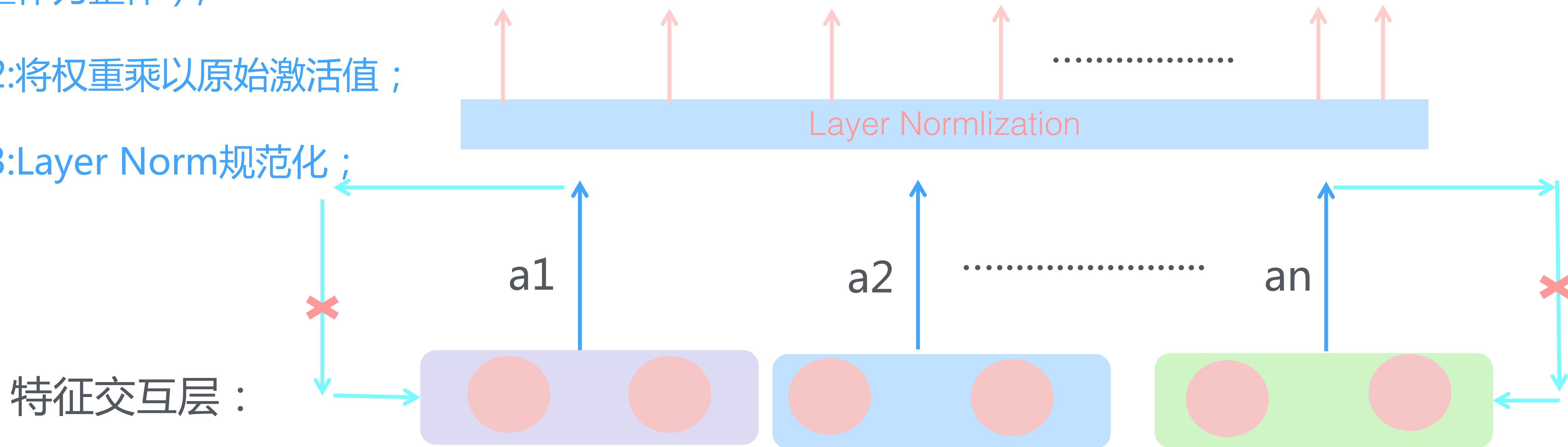
LayerNorm

DeepFFM-II : 特征交互层加入Attention

Step1:学习权重 a_1 — a_n (哈达马积向量作为整体);

Step2:将权重乘以原始激活值;

Step3:Layer Norm规范化;



特征交互层:

一个特征组合

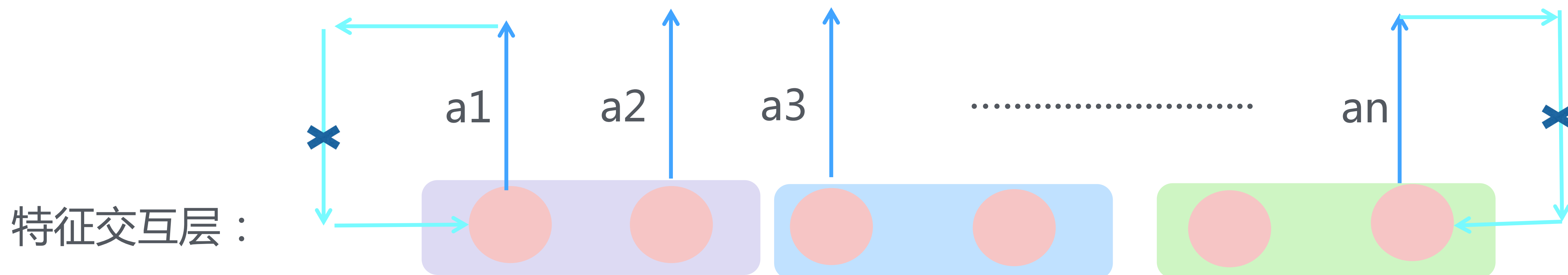
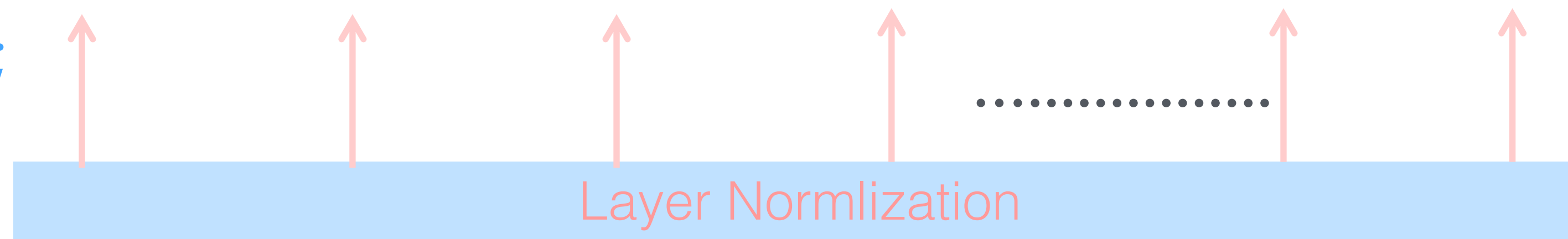
哈达马积版Attention : 一个组合一个权重

DeepFFM-III : 特征交互层加入Attention

Step1:学习权重 a_1 — a_n (哈达马积向量每位各自权重);

Step2:将权重乘以原始激活值;

Step3:Layer Norm规范化;



一个特征组合

哈达马积版Attention : 一个位一个权重

DeepFFM模型：效果对比

Criteo和Avazu是两个工业级的CTR数据

Model Name	Criteo		Avazu	
	AUC	Logloss	AUC	Logloss
DeepFM	0.8085	0.4445	0.7786	0.3810
DeepCross	0.7977	0.4617	0.7680	0.3940
xDeepFM	0.8091	0.4461	0.7808	0.3819
NeuralFFM-I	0.8087	0.4434	0.7839	0.3783
NeuralFFM-H	0.8088	0.4434	0.7835	0.3782
DeepFFM-I	0.8085	0.4440	0.7860	0.3767
DeepFFM-II	0.8091	0.4427	0.7862	0.3764
DeepFFM-III	0.8090	0.4443	0.7861	0.3770

一些结论：

1. 基准方法里，xDeepFM效果很好
2. NeuralFFM是有效的；
3. DeepFFM是有效的，效果最好；
4. DeepFFM-I/II/III效果差不多；
5. 相对而言，DeepFFM-II效果最好

由于神经网络版本FFM效率问题，如果实用化仍需进一步改进！

总结

- 在线推荐系统两阶段：召回+排序；
- 本分享主要集中在排序阶段
 - 线性排序模型：
 - 传统线性排序模型：LR->FM->FFM
 - 我们提出了FFM模型的改进模型：双线性FFM模型（减少内存）
 - Deep排序模型：LR->FM->FFM
 - 排序模型演进路线1：W&D->DeepFM->NeuralFFM
 - 排序模型演进路线2：Deep&Cross->xDeepFM
 - 我们提出了NeuralFFM模型的改进模型：DeepFFM模型（效果好）

THANKS

AiCon
全球人工智能与机器学习技术大会



QCon

全球软件开发大会

北京·2019

更多技术干货分享，北京站精彩继续

提前参与，还能享受更多优惠

识别二维码
查看了解更多

2019.qconbeijing.com



人工智能基础课

“通俗易懂的人工智能入门课”

王天一
工学博士，
副教授



扫一扫, 回复「AI」
领取 AI 学习图谱

极客邦企业培训与咨询

「帮助企业和技术人成长」

10 余年
经验技术专家

200+
国内外一线技术
专家团队

800+
企业研发团队的
选择

10000+
学员参与学习
交流

助力企业提升技术竞争壁垒，让技术驱动业务发展



扫码了解更多官方咨询

TGO 鲲鹏會

汇聚全球科技领导者的高端社群

🏠 全球9大城市

👤 700+ 高端科技领导者

使命
Mission

为社会输送更多
优秀的科技领导者

愿景
Vision

构建全球顶尖的有技术背景的
优秀人才成长平台



扫码了解更多内容