

第四部分：交叉结构

- 1、Factorized Machine (FM)
- 2、深度交叉网络 (DCN)
- 3、LHUC网络
- 4、SENet & Bilinear Cross

1、Factorized Machine (FM)

线性模型

有d个特征，记作 $x = [x_1, \dots, x_d]$

线性模型： $p = b + \sum_{i=1}^d w_i x_i$

该模型有n+1个参数： $w = [w_1, \dots, w_d]$ 和 b

预测是特征的加权和（只加不乘，未涉及特征交叉）

二阶交叉特征

内容

有d个特征，记作 $x = [x_1, \dots, x_d]$

线性模型+二阶交叉特征： $p = b + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=i+1}^d u_{ij} x_i x_j$

（例如房屋单价和房屋大小做交叉，可以使房价估计得更准）

减少参数数量

模型有 $O(d^2)$ 个参数，如果模型很大，那么参数量将巨大，且容易出现过拟合

如何减少参数数量？

将权重 u_{ij} 存储在 $d \times d$ 矩阵中，然后对矩阵分解

V 有 d 行 k 列， V^T 有 k 行 d 列，其中 k 远远小于 d

将权重作下图的近似，即可得到FM

• 线性模型 + 二阶交叉特征：

$$p = b + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=i+1}^d u_{ij} x_i x_j.$$

$\approx \mathbf{v}_i^T \mathbf{v}_j$

矩阵 \mathbf{U}

矩阵 \mathbf{V}

矩阵 \mathbf{V}^T

Factorized Machine (FM)

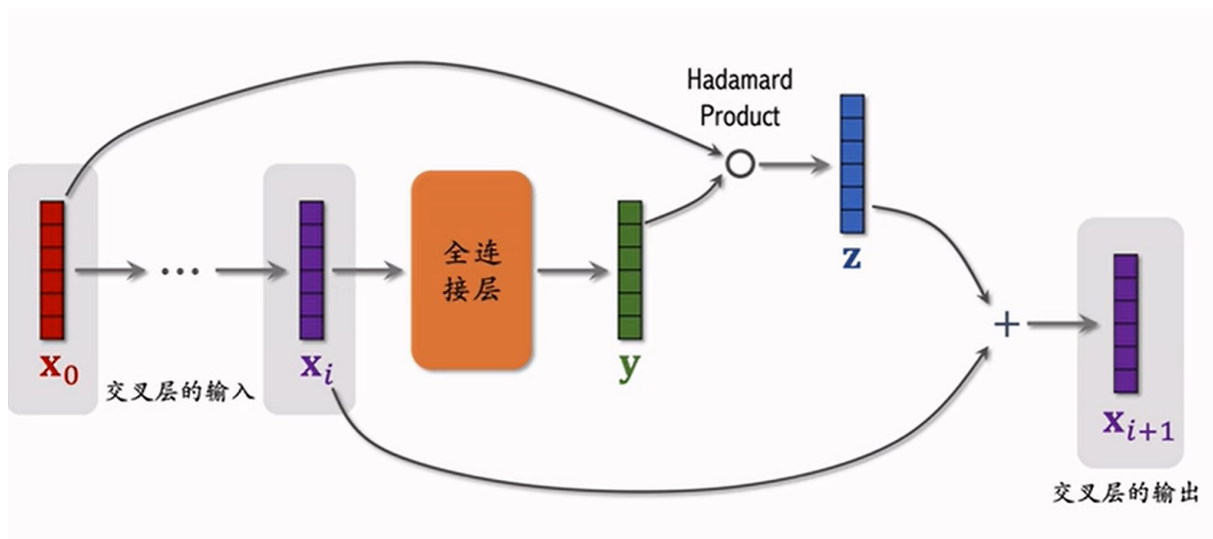
$$p = b + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=i+1}^d (v_i^T v_j) x_i x_j$$

FM模型有 $O(kd)$ 个参数，其中 k 远小于 d

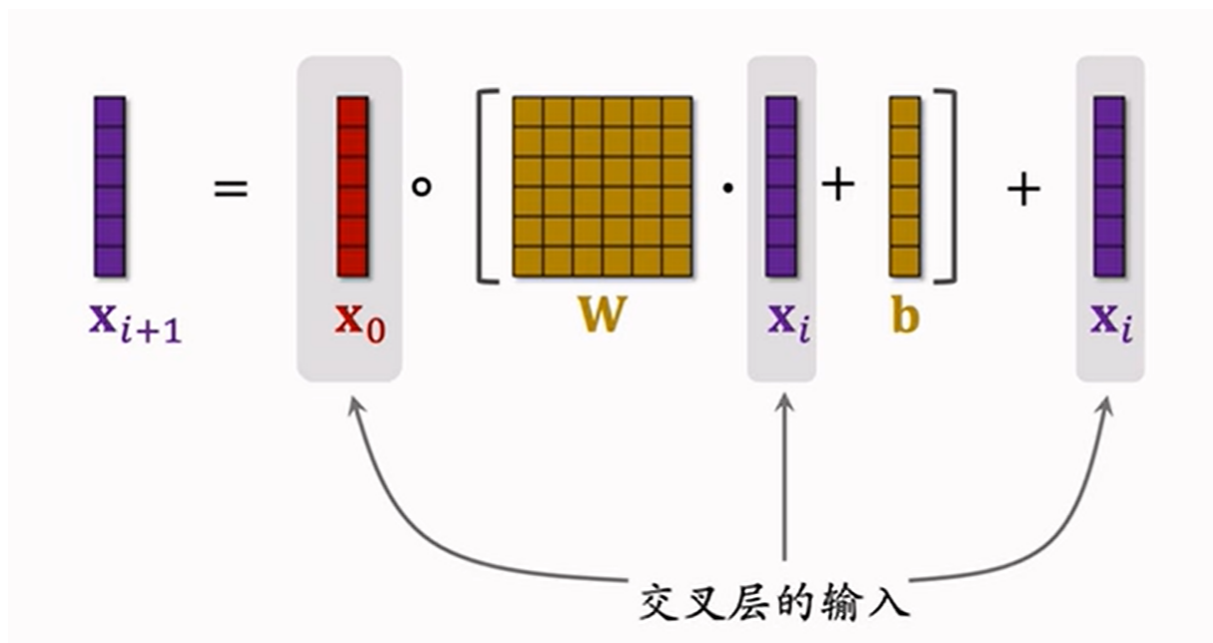
*FM现在在业界基本已经过时

2、深度交叉网络 (DCN)

交叉层 (cross layer)



可用下面公式表达：



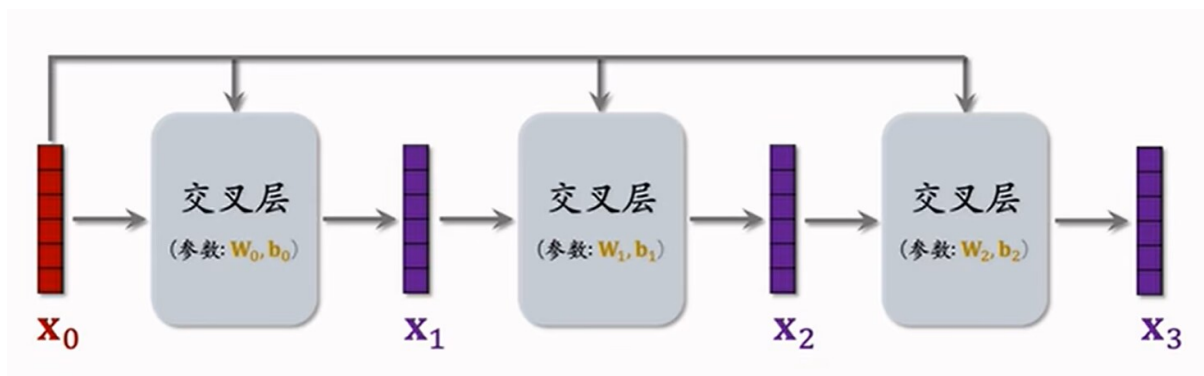
其中：

中括号里面 $W \cdot x_i + b$ 是全连接层中的操作

把输入与输出相加，相当于resnet连接，可以防止梯度消失

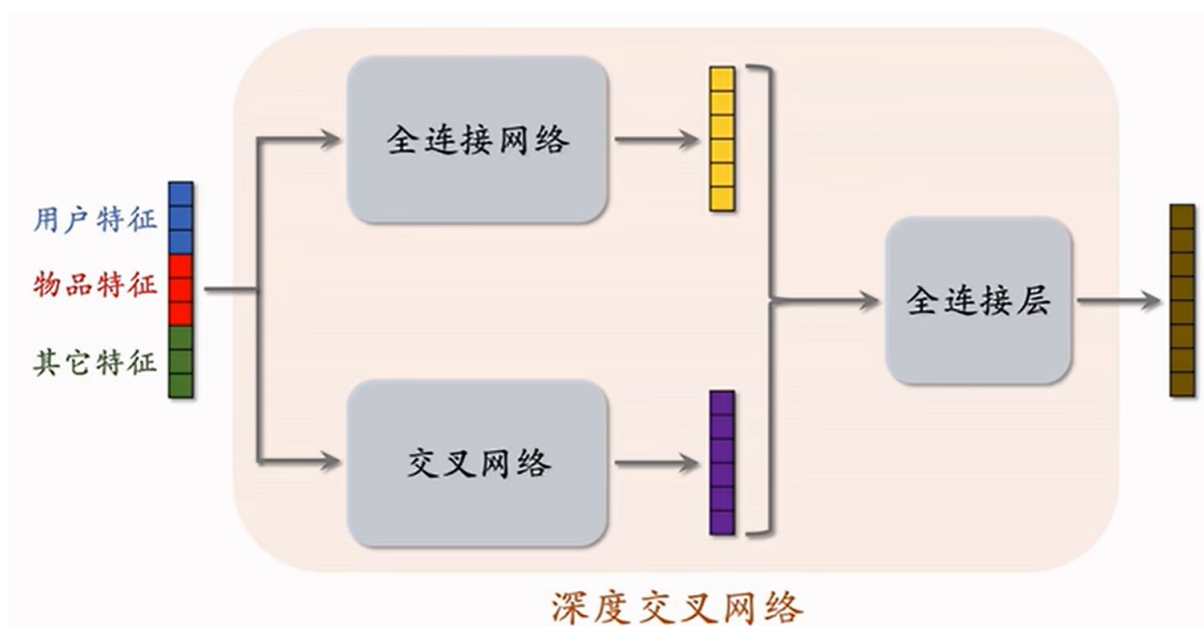
交叉网络 (cross network)

交叉网络的结构



深度交叉网络结构

DCN-V2并行式：

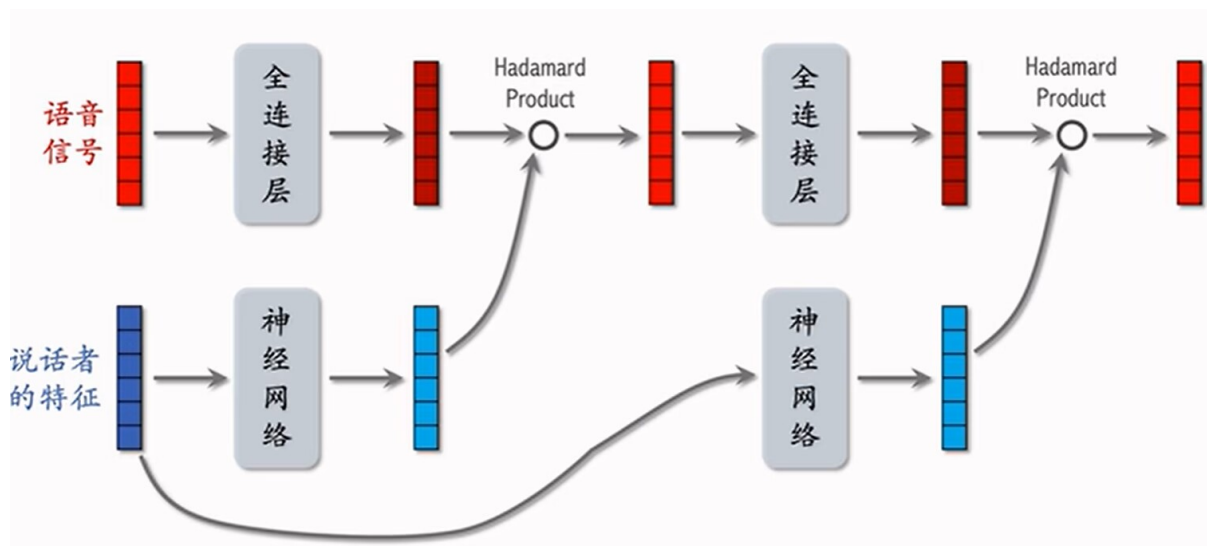


*之前用于召回/排序的模型，如双塔模型、多目标排序模型、MMoE等的神经网络部分都可以采用这种深度交叉网络

3、LHUC网络

语音识别中的LHUC

LHUC (learning hidden unit contributions) 起源于语音识别，后来快手将LHUC应用在推荐精排，称作PPNet



LHUC网络结构实现了语音信号与说话者特征的融合，说话者特征经过的神经网络通常是全连接层+sigmoid函数值x2

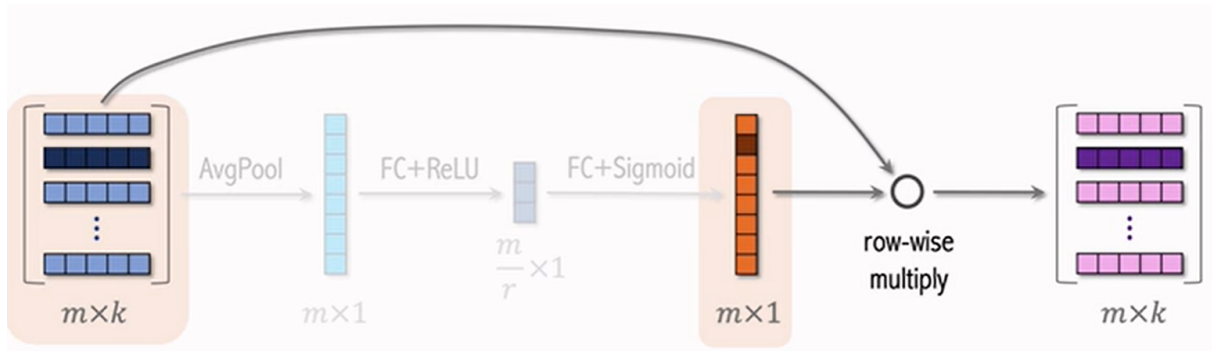
推荐系统中的LHUC

把语音信号、说话者的特征两个输入分别变成物品特征、用户特征，结构和语音识别的LHUC一致

4、SENet & Bilinear Cross

SENet

SENet的结构



先对特征进行embedding得到 $m \times k$ 的向量结构， m 行代表有 m 个特征（用户ID、物品ID、物品关键词等）， m 个向量的作用就是做特征的field-wise加权

（如果学出某个特点对任务的重要性不高，就对这个特点降权）

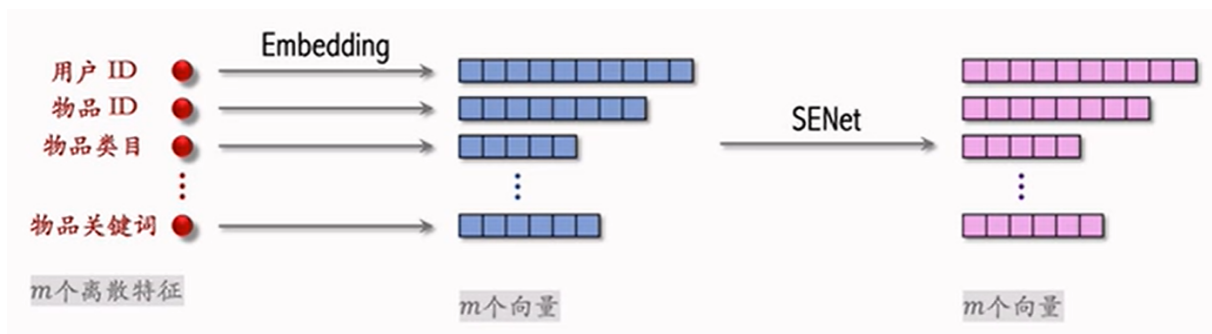
field-wise:

例如，用户ID embedding是64维向量，那么这64个元素算1个field，获得相同权重

如果有 m 个fields，那么权重向量就是 m 维

重要的field权重高，不重要的field权重低

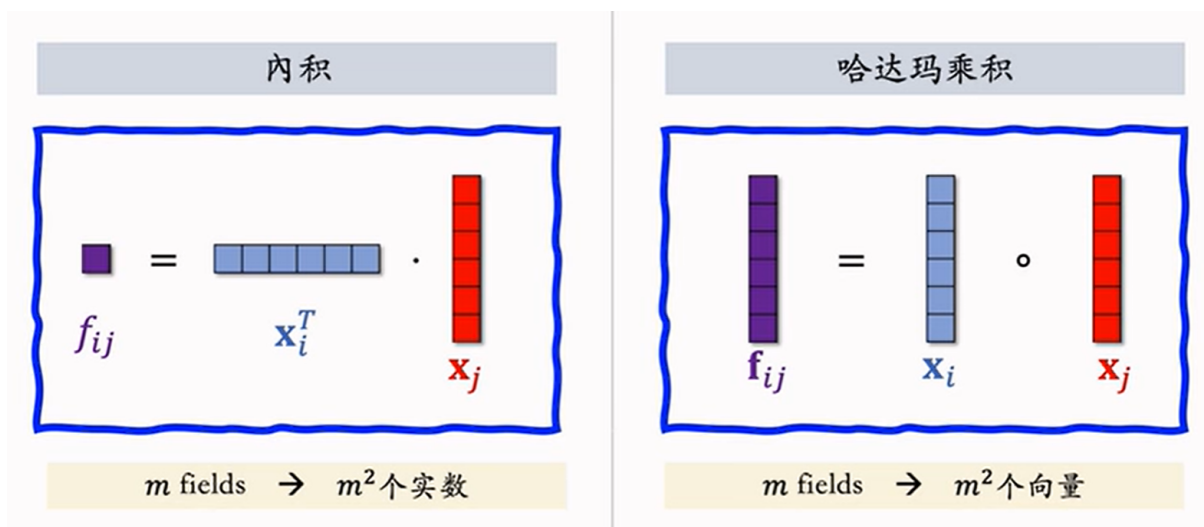
实际上，Embedding向量维度可以不同



Bilinear Cross

field间特征交叉

内积 & 哈达玛乘积

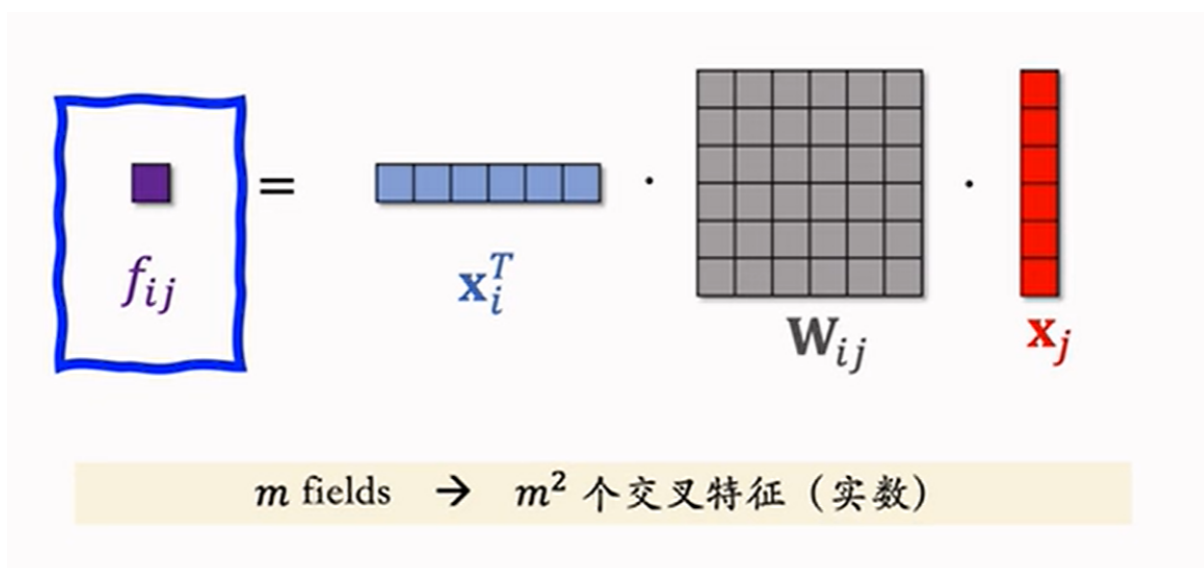


做这两种交叉必须要求 x_i 和 x_j 的维度一致

哈达玛乘积得到 m^2 个向量，相当夸张，通常需要人工选择一些特征进行交叉

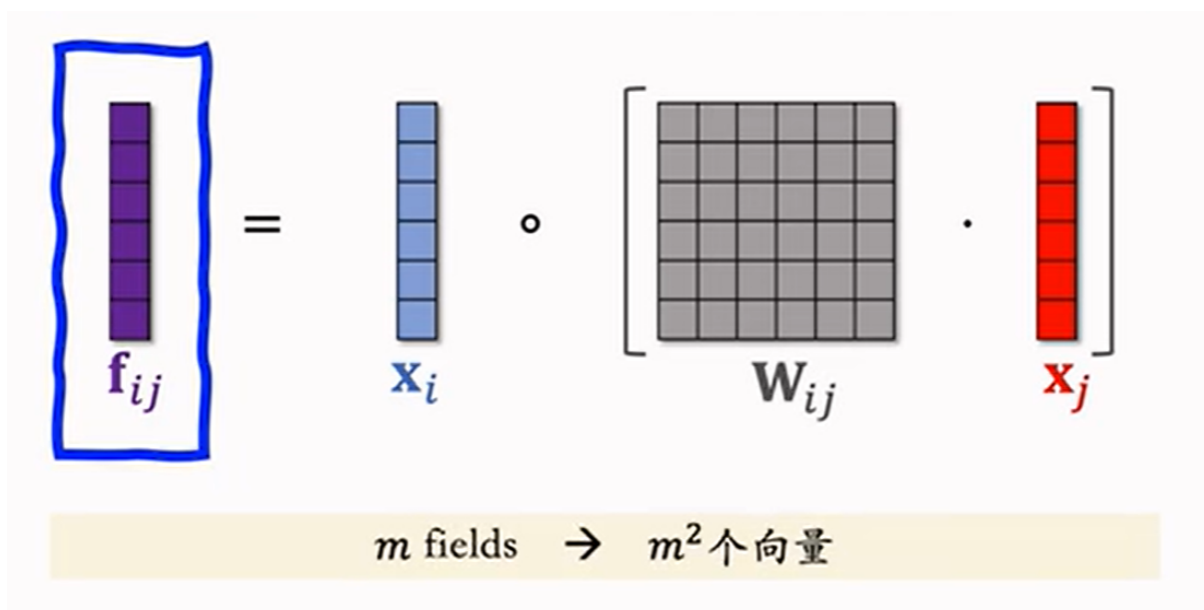
Bilinear Cross具体方法

1、内积型



如果有 m 个 fields，那么就有 $m^2/2$ 个参数矩阵，会导致参数数量极大，因此需要让重要的特征进行交叉，而不能所有特征都进行交叉

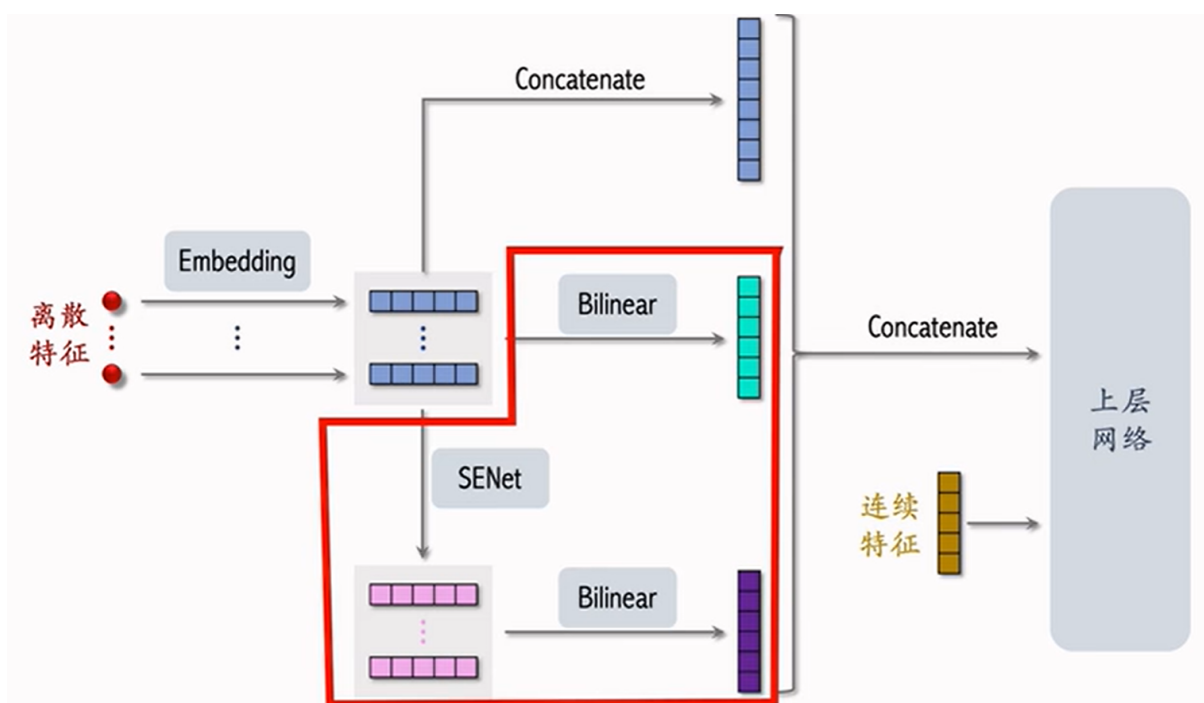
2、哈达玛乘积型



得到 m^2 个向量，容量太大了，而且很多都是无意义的特征，实际工程中需要指定一些特征做交叉

FiBiNet

这是将SENet & Bilinear Cross进行结合



相比起多目标排序模型来说，FiBiNet多了红色圈出来的部分，对特征的embedding一路进行SENet加权再bilinear交叉，另一路直接进行bilinear交叉，最后与embedding直接的串接再三路合并，和连续特征一起输入上层网络