

# xDeepFM [2018]

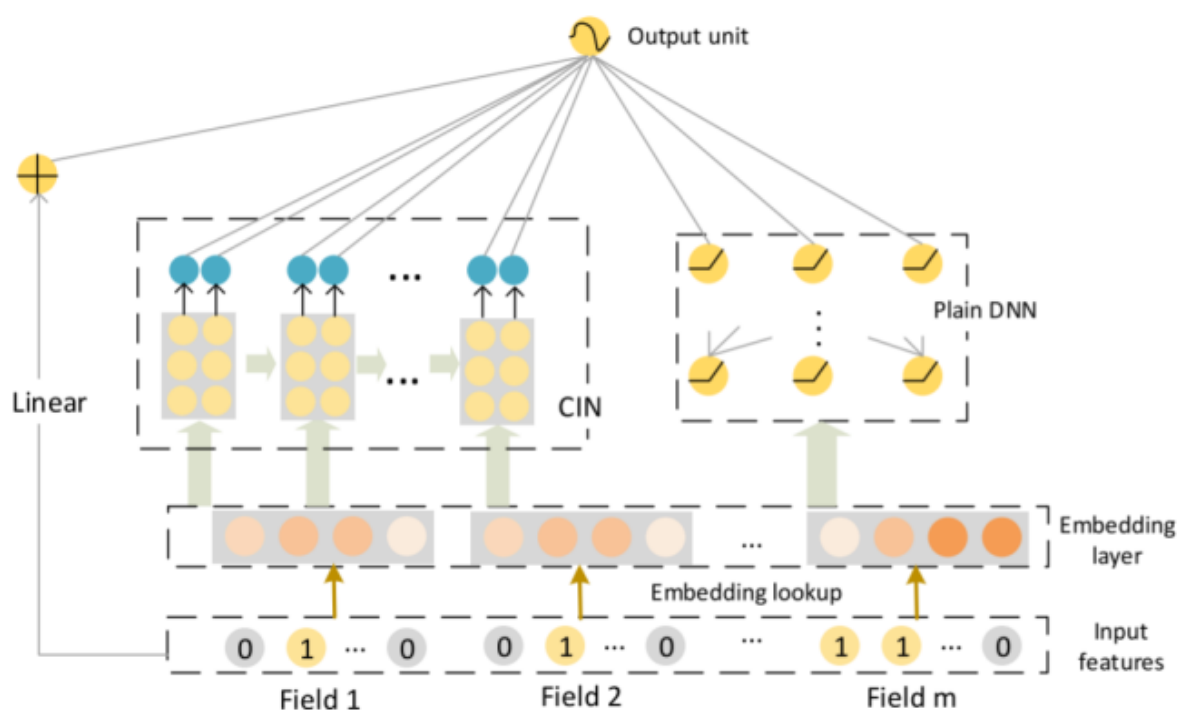
xDeepFM的卖点还是在于显式构造高阶交叉特征，只是和DCN不一样，xDeepFM的特征交叉是**vector-wise**而不是bit-wise的。也就是说，xDeepFM**区分了不同field的embedding**，实现的是"真正的"、像FM一样的特征交叉，而不是像DCN那样"虚假的"特征交叉。

文章上来先diss了一下DCN，指出DCN的本质只是每层给 $x_0$ 乘上不同的系数而已，表达能力有限。而且，DCN的特征交叉是bit-wise的。

## 0x01. 模型结构

为了解决DCN的问题，xDeepFM中提出了compressed interaction network(CIN)模块，来替代DCN中的cross network。

### 1. 整体模型结构



该模型主要分为三个部分：

- linear: 捕捉**线性**特征，对应Logistic Regression
- DNN: **隐式地、bit-wise**地学习高阶交叉特征
- ★ CIN: **显式地、vector-wise**地学习高阶交叉特征

### 2. CIN模块

其中，CIN模块是xDeepFM模型的核心。CIN模块的目的是完成**显式的、vector-wise**的特征交叉，同时把复杂度控制在**多项式时间**(不能随着阶数上升而指数增长)。

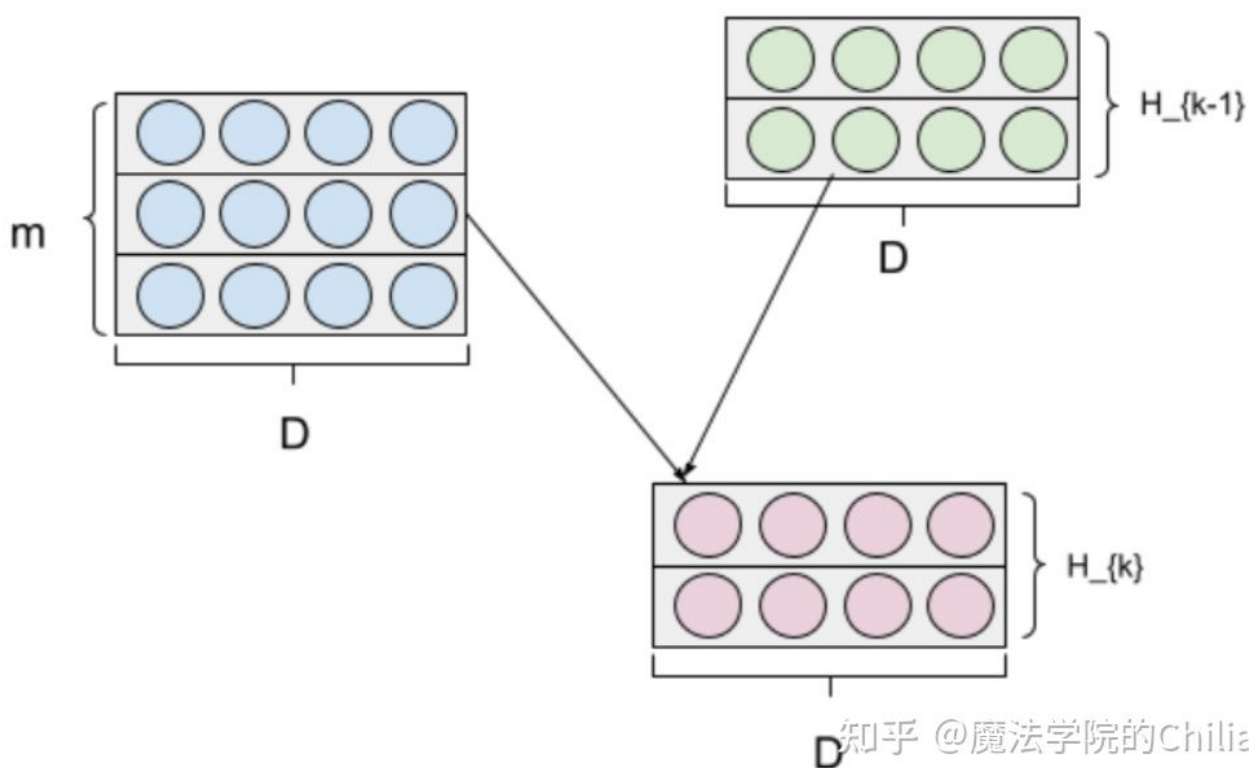
## 2.1 交互、压缩

对于输入特征，变为embedding之后可以全部拼接起来组成一个矩阵  $X^0 \in \mathbb{R}^{m \times D}$ ，其中m是field的个数,D是embedding size。第 k-1 层的矩阵记为  $X^{k-1}$ ，但它的第一维是  $H_{k-1}$ 。得到第  $k$  层的操作可以表示为：

$$X^k[h, :] = \sum_{i=1}^{H_{k-1}} \sum_{j=1}^m W_{i,j}^{k,h} (X^{k-1}[i, :] \odot X^0[j, :])$$

$X^0[j, :]$  就表示取出这个矩阵的第  $j$  行。也就是说，新的feature map中的**每一行**，都是先让上一个feature map的**每一行**，和每一个输入embedding的**每一行**做element-wise乘（哈达玛积），再用一套独有的W做变换后加起来融合的。

原论文中的图比较难以理解，于是我做了一个新的图：



最左边蓝色的图是  $X_0$ ，右边绿色的图是  $X_{k-1}$ ，现在我们要根据  $X_0$  和  $X_{k-1}$  得到  $X_k$ 。方法就是，对于  $X_0$  的**每一行**，都去和  $X_{k-1}$  的**每一行**进行哈达玛乘积，经过一个矩阵变换之后相加，得到  $X_k$  中的一行。这样， $X_k$  中每行就都包含了k+1阶的信息。

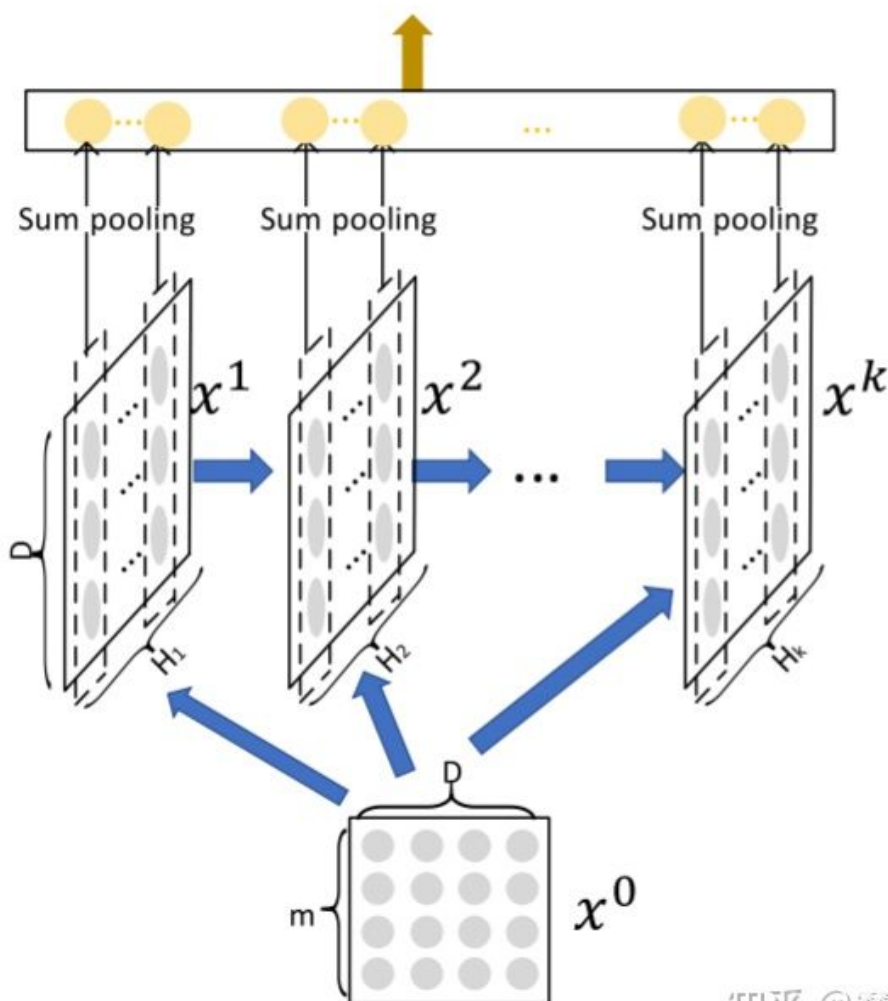
名字中的compressed（压缩）主要就体现在这里。在  $k-1$  这一层，与原始  $X_0$  的embedding作用后，通过加权求和，最终剩下有限个embedding拼接的矩阵。交叉的结果并不会无限膨胀。**因此求和这里其实就是做了压缩**。比如我们可以令每一层的  $H$  都相同，而且是一个比较小的数字。

## 2.2 再次压缩

像这样操作之后，每一步都会得到一个矩阵

$X_k$

，先对特征那个维度(D对应的维度)求和，得到一个向量（这一步是不是和内积很像？）。然后把每一层的这个向量都拼在一起，最后再用一层线性层+Sigmoid输出即可。



知乎 @魔法学院的Chilia

这一步看起来有点RNN的意思。

## 0x02. 时间复杂度

$$X^k[h, :] = \sum_{i=1}^{H_{k-1}} \sum_{j=1}^m W_{i,j}^{k,h} (X^{k-1}[i, :] \odot X^0[j, :])$$

假设第0个feature map有m个特征，之后所有的feature map都是H个特征。

那么，计算第k个feature map的一行就需要 $O(HmD)$ 的时间，那么假设所有feature map都有H行，就需要 $O(H^2mD)$ 时间；假设总共有L层，就需要 $O(H^2mDL)$ 时间。复杂度还是比较高的。

## 0x03. 关于高阶特征交叉的caveat

---

数学告诉我们，所有阶交叉都加上肯定会好。但是实践中我们要知道，在引入高阶交叉的同时，也需要付出很高复杂度的代价。高阶交叉虽然能够涨点，但是存在投入产出比不高的问题。在实践中，需要明确的是做哪些特征交叉比较好，交叉到几阶性价比最高。xDeepFM的复杂度就有些太高了，相比之下，DCN-v2的时间复杂度尚可接受。