

# 推荐系统与深度学习（六）——NFM模型原理

数据森麟 8月26日

以下文章来源于数据python与算法，作者livan



**数据python与算法**

用技术力量探索行为密码，让大数据助跑每个人；技术贴请关注：<https://my.csdn.net/>



作者：livan

来源：数据python与算法

## NFM原理介绍

FM模式的提出主要是为了解决特征关系的充分应用和稀疏的问题，但是FM模型不能完整的表现深层次的数据特性，所以，某大佬们将FM结合在了DNN结构上，两个结构明确分工：FM处理浅层次特征，DNN处理深层次特征。

这一结合一经推出，风靡全网，引来了很多大佬对其进行改造，主要改造方向分两个：

### 1）并行结构：

FM部分和DNN部分分开计算，只在输出层进行一次融合得到结果，例如：DeepFM、DCN和Wide&Deep结构。

### 2）串行结构：

将FM的一次项和二次项结果(或其中之一)作为DNN部分的输入，经DNN得到最终结果，例如：PNN、AFM和NFM结构。

本文了解一下NFM的模式：

在探索NFM之前我们先看一下FM结构：

$$\hat{y}_{FM(x)} = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n v_i^T v_j \cdot x_i x_j$$

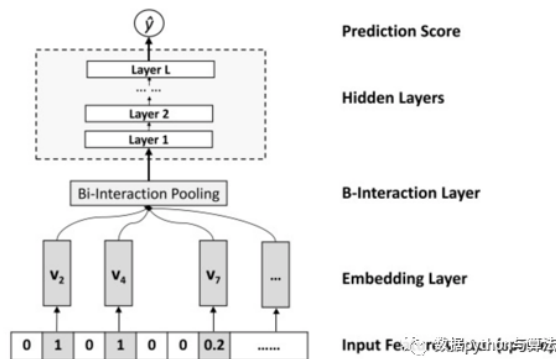
看过前面文章的小伙伴一定很熟悉这一公式吧，所有的故事都是从这个公式讲起的。

跳过推导阶段，我们直接来看一下NFM的公式：

$$\hat{y}_{NFM(x)} = w_0 + \sum_{i=1}^n w_i x_i + f(x)$$

比较一下上面的样式，是不是感觉一个样？

一起围观一下NFM模型的图形：



图形中我们看到了四块：

1) Embedding Layer：这一层就不用多讲了，特征降维和解决稀疏问题的，上一文章中讲过：

2) Bi-Interaction Layer：这一层是NFM的核心部分，本质是一个 pooling 操作，将 embedding vector 集合归并为一个向量，其公式为：

$$f_{BI}(V_x) = \sum_{i=1}^n \sum_{j=i+1}^n x_i v_i \odot x_j v_j$$

有没有感觉很熟悉，是不是与FM模型中的二次项部分很相似，我们要做的就是解这个方程：

⊙表示两个向量对应元素相乘，其结果为一个向量。所以，Bi-Interaction Layer将embedding vectors 进行两两交叉⊙运算，然后将所有向量进行对应元素求和，最终 $f_{BI}(V_x)$ 为pooling之后的一个向量。

经过运算，上面的公式改写为：

$$f_{BI}(V_x) = \frac{1}{2} \left[ \left( \sum_{i=1}^n x_i v_i \right)^2 - \sum_{i=1}^n (x_i v_i)^2 \right]$$

3) Hidden Layer：这一层是对上一层的进一步处理，每一层都可以计算出一个结果并相互叠加，如图：

$$\begin{aligned} z_1 &= \sigma_1(W_1 f_{BI}(V_x) + b_1) \\ z_2 &= \sigma_2(W_2 z_1 + b_2) \\ &\dots\dots \\ z_L &= \sigma_L(W_L z_{L-1} + b_L) \end{aligned}$$

有L层layer，则会形成L个上面的公式，即含有L层嵌套。

4) Prediction Score：这一层是在深度模型运算的基础上添加了FM的前两层运算，得到一个结果值：

首先对上一层的结果 $z_L$ 运算为：

$$f(X) = h^T z_L$$

进一步运算得到一个完整的模型：

$$\hat{y}_{NFM}(X) = w_0 + \sum_{i=1}^n w_i x_i + f(X) = w_0 + \sum_{i=1}^n w_i x_i + h^T \sigma_L(W_L(\dots \sigma_1(W_1 f_{BI}(V_x) + b_1) + b_L))$$

一个完整的NFM的模型就出现了～

有没有很吃惊，这个深奥的理论原来如此简单～

### 注意事项

1) 如果将中间的隐藏层去掉，仅保留最终的prediction layer，那么NFM即复现了FM，可以认为NFM就是FM的推广。

2) 将模型复杂度提高，不可避免的会面临训练过拟合的问题，NFM使用了dropout与batch normalization技术来缓解过拟合问题，但需要注意使用的顺序。

3) NFM的Bi-Interaction Layer能充分的捕获二阶交叉特征信息，所以只需要接入较浅的隐层结构就能捕获更高阶的信息，从而取得不错的效果。NFM对于参数初始化不敏感，更容易进行训练优化。

4) NFM与Wide&Deep与DeepCrossing效果差异不大，不过NFM训练较为简便。



当当网开学季活动来袭，满100减50的基础上，实付满200的基础上使用优惠码

**DYU6NM可以再减40，相当于400-240，**

使用方法如下：

步骤一，挑选心仪的图书至购物车点击结算

步骤二，点击优惠券/码处

步骤三，输入优惠码**DYU6NM**（注意全部要大写）

**需要注意的是：优惠码全场自营图书可用（教材、考试类除外）**