

推荐系统与深度学习（四）——AFM模型原理

原创 livan 数据python与算法 7月15日

缘起

AFM是基于Attention机制的FM模型，主要是在FM模型求和的过程中添加了一个权重因子，而这个权重因子是按照Attention机制计算出来的。

我们可以先看一下AFM的公式：

$$\hat{y}_{AFM}(X) = w_0 + \sum_{i=1}^n w_i x_i + p^T \sum_{i=1}^n \sum_{j=i+1}^n a_{i,j} (v_i \odot v_j) x_i x_j$$

比较FM的公式：

$$Y = w_0 + \sum_{i=1}^n w_i * x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle * x_i * x_j$$

发现两个公式的主要差异在于第三项。

AFM中第三项的计算方法为：

$$\begin{aligned} v_i &= (v_1, v_2, \dots, v_k) \\ v'_j &= (v'_1, v'_2, \dots, v'_k) \\ v_i \odot v'_j &= (v_1, v_2, \dots, v_k) \odot (v'_1, v'_2, \dots, v'_k) \end{aligned}$$

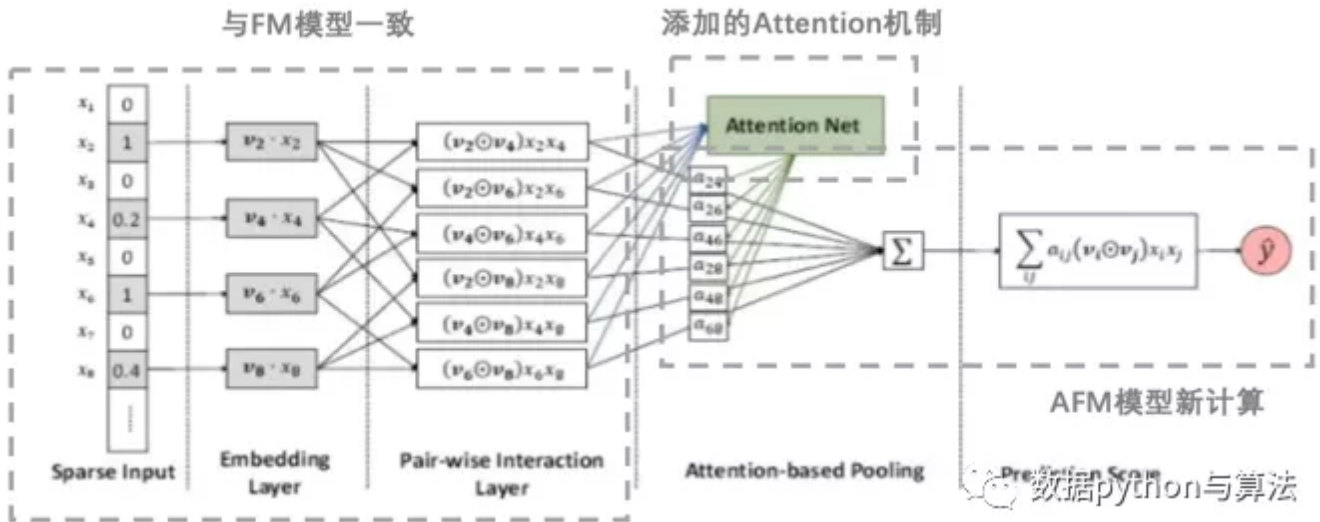
而对应的权重参数 $a_{i,j}$ 即为通过Attention机制计算出的权重因子。

Attention机制背后的数据公式为：

$$\begin{aligned} a'_{ij} &= h^T \text{Relu}(W(v_i \odot v_j) x_i x_j + b) \\ a_{ij} &= \frac{\exp(a'_{ij})}{\sum_{i,j} \exp(a'_{ij})} \end{aligned}$$

从公式中可以看出模型使用了一个双层神经网络，通过 $x_i x_j$ 特征关联计算出了各个特征关联对应的权重参数 a ，并融入到公式中。

如下图为AFM模型的作者给出的完整图：



AFM的不足之处

不难看出AFM只是在FM的基础上添加了Attention的机制，但是实际上，由于最后的加权累加，二次项并没有进行更深的网络去学习非线性交叉特征，所以AFM并没有发挥出DNN的优势，也许结合DNN可以达到更好的结果。

性能方面，AFM表现出较好的优势，即使在AFM不使用dropout时，其效果仍大幅优于FM，说明Attention Network的有效性。

前情回顾：

推荐系统与深度学习（三）——Attention机制原理

欢迎大家关注公众号：