



北京师范大学 珠海校区

BEIJING NORMAL UNIVERSITY AT ZHUHAI

# 另一个角度看FM

马静





- MF: 利用共现矩阵, 迭代出用户矩阵和物品矩阵

核心思想是通过两个低维小矩阵 (一个代表用户embedding矩阵, 一个代表物品embedding矩阵) 的乘积计算, 来模拟真实用户点击或评分产生的大的协同信息稀疏矩阵

- 用户A的隐向量表达/embedding=[1.2,0.8]

物品

	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

用户

共现矩阵

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

用户矩阵

×

	W	X	Y	Z
1	1.5	1.2	1.0	0.8
2	1.7	0.6	1.1	0.4

物品矩阵

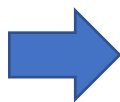
- MF可以被认为只有User ID 和Item ID这两个特征的FM模型



既然ID类特征可以用隐向量表达，其他特征要怎么用隐向量表达？

MF

$$\text{Score} = V_{\text{itemID}} * V_{\text{userID}}$$



FM

$$\text{Score} = F(V_{\text{itemID}}, V_{\text{userID}}, \dots, V_{\text{else}})$$

其他维度特征

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

单特征

二阶交叉特征



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

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \langle v_i, v_j \rangle x_i x_j - \frac{1}{2} \sum_{i=1}^n \langle v_i, v_i \rangle x_i x_i \\ &= \frac{1}{2} \left( \sum_{i=1}^n \sum_{j=1}^n \sum_{f=1}^k v_{i,f} v_{j,f} x_i x_j - \sum_{i=1}^n \sum_{f=1}^k v_{i,f} v_{i,f} x_i x_i \right) \\ &= \frac{1}{2} \sum_{f=1}^k \left( \left( \sum_{i=1}^n v_{i,f} x_i \right) \left( \sum_{j=1}^n v_{j,f} x_j \right) - \sum_{i=1}^n v_{i,f}^2 x_i^2 \right) \\ &= \frac{1}{2} \sum_{f=1}^k \left( \left( \sum_{i=1}^n v_{i,f} x_i \right)^2 - \sum_{i=1}^n v_{i,f}^2 x_i^2 \right) \end{aligned}$$

$$\frac{\partial}{\partial \theta} \hat{y}(x) = \begin{cases} 1, & \text{if } \theta \text{ is } w_0 \\ x_i, & \text{if } \theta \text{ is } w_i \\ x_i \sum_{j=1}^n v_{j,f} x_j - v_{i,f} x_i^2, & \text{if } \theta \text{ is } v_{i,f} \end{cases}$$

与MF一样，利用梯度下降法迭代出隐向量



最终预测效果:

## FM模型

$$\text{FM: } \hat{y}(x) := w_0 + \underbrace{\sum_{i=1}^n w_i x_i}_{\text{LR模型}} + \underbrace{\sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j}_{\text{Dense化两两特征}}$$

LR模型

Dense化两两特征

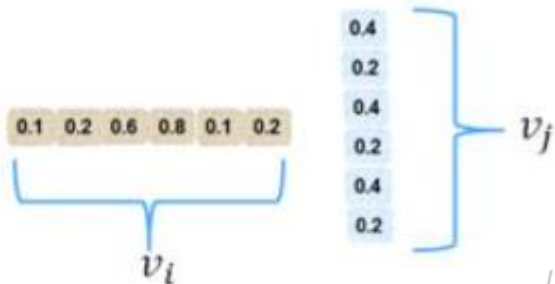
$$\langle v_i, v_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f} =$$

$v_i$   $v_j$



优势：训练集中没有出现的样本，也可以被拟合出来。

## 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



对比暴力组合方式：

## 线性模型改进：加入特征组合

改进版本  $\hat{y}(x) := w_0 + \sum_{i=1}^n w_i x_i + \underbrace{\sum_{i=1}^n \sum_{j=i+1}^n w_{i,j} x_i x_j}_{\text{两两特征组合}}$

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

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





北京師範大學 珠海校区

BEIJING NORMAL UNIVERSITY AT ZHUHAI

[github.com/MadelineMa/Recommender-System](https://github.com/MadelineMa/Recommender-System)

THANKS

[github.com/MadelineMa/Recommender-System](https://github.com/MadelineMa/Recommender-System)

DESIGNED BY 2xh