FM/FFM 概述 [1]

leolinuxer

July 23, 2020

Contents

1 前置知识

1.1 线性回归模型

线性回归模型如下：

$$y = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n \Rightarrow y = w_0 + \sum_{i=1}^{n} w_ix_i$$

线性回归模型的假设：

- **特征之间相互独立不相关**；
- 特征的作用是可以叠加的；

线性回归模型假设特征之间是相互独立的、不相关的。但在某些场景中，特征之间往往是相关的，而不是相互独立的。比如 < 女性 > 和 < 化妆品 >，< 程序员 > 与 < 计算机类书籍 >，所以需要特征组合。

特征两两组合，得到二阶多项式如下：

$$y = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} x_i x_j$$

其中，n 代表样本的特征数量，x_i 是第 i 个特征的值，w_0, w_i, w_{ij} 是模型参数。则组合特征的参数一共有 $n(n-1)/2$ 个，任意两个参数是独立的。

1.2 非负矩阵分解

非负矩阵：所有元素非负。若 X 是非负矩阵，记作 $X \geq 0$。

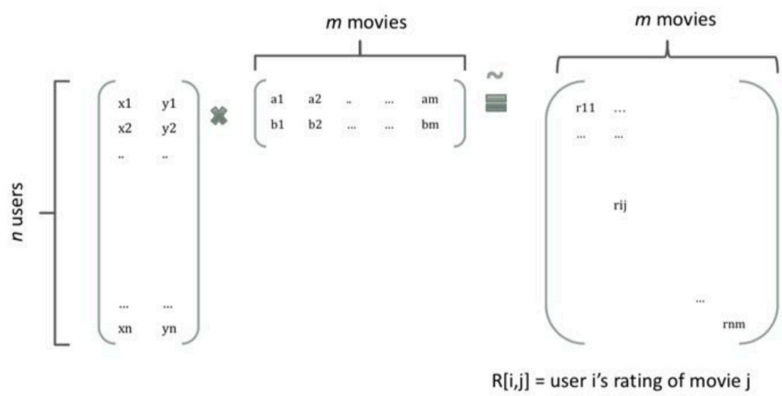
非负矩阵分解：给定一个非负矩阵 $X \geq 0$，找到两个非负矩阵 $W \geq 0$ 和 $H \geq 0$，使得：

$$X \approx WH$$

即将非负矩阵 X 分解为两个非负矩阵 W, H 的乘积形式，称为非负矩阵分解。

非负矩阵分解的例子如下：

我们把每个 user 表示成一个二维向量，同时把每个 item 表示成一个二维向量，两个向量的点积就是矩阵中 user 对 item 的打分，写成非负矩阵分解形式如下：



特征关系的向量化，则得 FM 模型。

2 FM 模型

论文：Factorization Machines

因子分解机模型，即 Factorization Machine Model，简称 FM。

2.1 FM 模型的方程推导

上述二阶多项式的二项式参数 w_{ij} 可以组成一个对称矩阵 W，那么这个矩阵可以分解为：$W = V^T V$，V 的第 j 列便是第 j 维特征的特征向量，即 W_{ij} 的每个参数都可以表示为：

$$W_{ij} = < V_i, V_j >$$

这就是 FM 的核心思想。

即上述二阶多项式可以写成：

$$y = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} < v_i, v_j > x_i x_j$$

这边是 FM 的方程。其中，v_i 是第 i 维特征的隐向量，隐向量的长度为 k，包含 k 个描述特征的因子；$< \cdot, \cdot >$ 代表向量点积，即 $< v_i, v_j > = \sum_{f=1}^{k} v_{i,f} v_{j,f}$。

2.2 FM 模型的假设

- **特征之间两两相关**；
- 特征的作用是可以叠加的；

2.3 FM 模型的计算

利用矩阵直观化推导 FM 模型的计算，具体推导如下。

根据 FM 的二阶多项式，扩充后得到如下矩阵：

$< v_1, v_1 > x_1 x_1$	$< v_1, v_2 > x_1 x_2$	$\cdots$	$< v_1, v_{n-1} > x_1 x_{n-1}$	$< v_1, v_n > x_1 x_n$
$< v_2, v_1 > x_2 x_1$	$< v_2, v_2 > x_2 x_2$	$\cdots$	$< v_2, v_{n-1} > x_2 x_{n-1}$	$< v_2, v_n > x_2 x_n$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$< v_{n-1}, v_1 > x_{n-1} x_1$	$< v_{n-1}, v_2 > x_{n-1} x_2$	$\cdots$	$< v_{n-1}, v_{n-1} > x_{n-1} x_{n-1}$	$< v_{n-1}, v_n > x_{n-1} x_n$
$< v_n, v_1 > x_n x_1$	$< v_n, v_2 > x_n x_2$	$\cdots$	$< v_n, v_{n-1} > x_n x_{n-1}$	$< v_n, v_n > x_n x_n$

矩阵上三角(绿色部分) 的元素之和即为 FM 的二阶部门：

$$\sum_{i=1}^{n} \sum_{j=i+1}^{n} < v_i, v_j > x_i x_j$$

等于**矩阵的全部元素之和减去对角线元素之和，再除以二**：

$$\sum_{i=1}^{n}\sum_{j=i+1}^{n} < v_i, v_j > x_i x_j = \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} < v_i, v_j > x_i x_j - \sum_{i=1}^{n} < v_i, v_i > x_i x_i\right)$$

FM 模型的二次项化简推导过程如下：

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=i+1}^{n} < v_i, v_j > x_i x_j &= \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} < v_i, v_j > x_i x_j - \sum_{i=1}^{n} < v_i, v_i > x_i x_i\right) \\
&= \frac{1}{2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{f=1}^{k} v_{i,f} v_{j,f} x_i x_j - \sum_{i=1}^{n}\sum_{f=1}^{k} v_{i,f} v_{i,f} x_i x_i\right) \\
&= \frac{1}{2}\sum_{f=1}^{k}\left[\left(\sum_{i=1}^{n} v_{i,f} x_i\right)\left(\sum_{j=1}^{n} v_{j,f} x_j\right) - \sum_{i=1}^{n} v_{i,f}^2 x_i^2\right] \\
&= \frac{1}{2}\sum_{f=1}^{k}\left[\left(\sum_{i=1}^{n} v_{i,f} x_i\right)^2 - \sum_{i=1}^{n} v_{i,f}^2 x_i^2\right] \qquad \text{(i 和 j 是等价的)}
\end{aligned}$$

即 FM 模型方程 (时间复杂度为 $O(kn^2)$)：

$$y = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n} < v_i, v_j > x_i x_j$$

可以转化为 (时间复杂度为 $O(kn)$)：

$$y = w_0 + \sum_{i=1}^{n} w_i x_i + \frac{1}{2}\sum_{f=1}^{k}\left[\left(\sum_{i=1}^{n} v_{i,f} x_i\right)^2 - \sum_{i=1}^{n} v_{i,f}^2 x_i^2\right]$$

同时，在稀疏数据场景下，很多特征为 0，所以只需要计算非零特征就行，这样可以极大提升计算效率。

2.4 FM 模型训练

FM 可以采用随机梯度下降的方法进行训练。

FM 模型各个参数的梯度如下：

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

2.5 FM 模型总结

2.5.1 FM 模型优点

- 适用于高度稀疏化的数据场景;
- 具有线性复杂度;

2.5.2 线性回归 vs FM

FM 模型由线性回归模型演化出来。二者的最大区别是：线性回归模型的特征独立，而 FM 模型的特征两两相关。

References

[1] “Fm 因子分解机模型的原理、推导、代码和应用.” [Online]. Available: https://zhuanlan.zhihu.com/p/145436595