FM算法简单梳理

杨航锋

1 FM 算法的建模过程

在传统的线性模型中，各个特征之间都是独立考虑的，并没有涉及到特征与特征之间的交互关系，但实际上大量的特征之间是相互关联的。如何寻找相互关联的特征，基于上述思想 FM 算法应运而生。传统的线性模型为

$$y = w_0 + \sum_{i=1}^{n} w_i x_i$$

在传统的线性模型的基础上中引入特征交叉项可得

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

在数据非常稀疏的情况下很难满足 x_i、x_j 都不为 0，这样将会导致 w_{ij} 不能够通过训练得到，因此无法进行相应的参数估计。可以发现参数矩阵 w 是一个实对称矩阵， w_{ij} 可以使用矩阵分解的方法求解，通过引入辅助向量 V

$$V = \begin{bmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1k} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & v_{n3} & \cdots & v_{nk} \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{bmatrix}$$

然后用 $w_{ij} = \mathbf{v}_i \mathbf{v}_j^T$ 对 w 进行分解

$$w = VV^T = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T & \mathbf{v}_2^T & \cdots & \mathbf{v}_n^T \end{bmatrix}$$

综上可以发现原始模型的二项式参数为 $\frac{n(n-1)}{2}$ 个，现在减少为 $kn(k \ll n)$ 个。引入辅助向量 V 最为重要的一点是使得 $x_t x_i$ 和 $x_i x_j$ 的参数不再相互独立，这样就能够在样本数据稀疏的情况下合理的估计模型交叉项的参数

$$\langle \mathbf{v}_t, \mathbf{v}_i \rangle = \sum_{f=1}^{k} \mathbf{v}_{tf} \cdot \mathbf{v}_{if}$$

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{f=1}^{k} \mathbf{v}_{if} \cdot \mathbf{v}_{jf}$$

$x_t x_i$ 和 $x_i x_j$ 的参数分别为 $\langle \mathbf{v}_t, \mathbf{v}_i \rangle$ 和 $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ ，它们之间拥有共同项 $\mathbf{v}_i$ ，即所有包含 $\mathbf{v}_i$ 的非零组合特征的样本都可以用来学习隐向量 $\mathbf{v}_i$ ，而原始模型中 w_{ti} 和 w_{ij} 却是相互独立的，这在很大程度上避免了数据稀疏造成的参数估计不准确的影响。因此原始模型可以改写为最终的 FM 算法

$$y = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j$$

由于求解上述式子的时间复杂度为 $\mathcal{O}(n^2)$ ，可以看出主要是最后一项计算比较复杂，因此从数学上对该式最后一项进行一些改写可以把时间复杂度降为 $\mathcal{O}(kn)$

$$\begin{aligned}
& \sum_{i=1}^{n-1} \sum_{j=1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \\
=& \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j - \frac{1}{2} \sum_{i=1}^{n} \langle \mathbf{v}_i, \mathbf{v}_i \rangle x_i x_i \\
=& \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{f=1}^{k} \mathbf{v}_{if} \mathbf{v}_{jf} x_i x_j - \sum_{i=1}^{n} \sum_{f=1}^{k} \mathbf{v}_{if} \mathbf{v}_{if} x_i x_i \right) \\
=& \frac{1}{2} \sum_{f=1}^{k} \left(\left(\sum_{i=1}^{n} \mathbf{v}_{if} x_i \right) \left(\sum_{j=1}^{n} \mathbf{v}_{jf} x_j \right) - \sum_{i=1}^{n} \mathbf{v}_{if}^2 x_i^2 \right) \\
=& \frac{1}{2} \sum_{f=1}^{k} \left(\left(\sum_{i=1}^{n} \mathbf{v}_{if} x_i \right)^2 - \sum_{i=1}^{n} \mathbf{v}_{if}^2 x_i^2 \right)
\end{aligned}$$

2 FM 算法小结

- FM 算法降低了因数据稀疏，导致特征交叉项参数学习不充分的影响；
- FM 算法提升了参数学习效率和模型预估的能力。