FFM

1、背景

FFM 全称 Field-aware **Factorization Machine** for CTR Prediction，看到全名之后，必须有所感触，“**Factorization Machine**”这是什么鬼玩意呀！！哈哈哈，这时候就必须回到之前提到的FM，**Factorization Machine** 硬翻称 **“因子分解的机器”**

首先先来回顾一下FM的正确姿势，上公式:

$$y = b + \sum_{i=1}^{n} w_i^* x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} < V_i, V_j > x_i \times x_j \tag{1}$$

在FM(公式1)中，主要分为三个部分之和，从左->右依次是，偏置、一阶线性特征、二阶线性特征；偏置就是一个常量、一阶线性特征容易理解，就是特征x_i前面乘以一个权重、然后再次再简单回顾下二阶特征的方法论：

所谓二阶特征，其实就是不同特征之间的组合，比如人的眼睛与鼻子、眼睛与嘴巴、眼睛与耳朵等等，这种组合的表现形式以乘积的形式表示，即$x_i \times x_j$,其中x_i可以粗暴认为是眼睛的大小**数值**、x_j再来粗暴的认为是鼻子的高低程度**数值**，则它们的权重则使用对应特征的向量乘积表示，即$< V_i, V_j >$详情如下：

$$< V_i, V_j > := \sum_{f=1}^{k} v_{i,f} \cdot v_{j,f} \tag{2}$$

ok,下面再具体看一下FM中对公式(2)重写的过程，

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

所以FM的最终的公式可以表示为：

$$y = b + \sum_{i=1}^{n} w_i^* x_i + \frac{1}{2} \sum_{k}^{f=1} \left(\left(\sum_{i=1}^{n} v_{i,f} \times x_i \right)^2 - \sum_{i=1}^{n} n_{i,f}^2 \times x_i^2 \right) \tag{3}$$

哈哈哈，这里直接摆出公式了，不在详细解释缘由，**详见FM**

```
叨叨了半天下面终于到了正文环节FFM了
```

2、FFM的方法论

一个新的算法的提出，必定是有它的创新之处！！！

FFM的方法论是站在FM的基础上进行改进与创新、那么我们还是得从另外角度看FM:

首先FM的核心是一个$n \times k$ 的参数矩阵 **V**，n是单条记录的特征维度、k是超参表示 V_i 的维度，矩阵详情如下:

$$\begin{bmatrix} v_{1,1} & v_{1,1} & \cdots & & v_{1,k} \\ \vdots & & & \cdots & \\ & & & & \vdots \\ \color{red}{v_{i,1}} & \color{red}{v_{i,1}} & \color{red}{\cdots} & & \color{red}{v_{i,k}} \\ \vdots & & & & \vdots \\ v_{n,1} & v_{n,1} & \cdots & & v_{n,k} \end{bmatrix}_{n \times k} \tag{4}$$

任意两个特征的组合的权重都可以使用$< V_i, V_j >$表示，详情见公式(2)

注意，注意啦！！！ 此处需要细品，需要提提FM的不足之处

这里引用原论文中的一句话：**In FM, every feature has only latent vector to learn the latent effect with any other features**

这里结合公式（4）来看，FM中每个变量只有一个**隐向量**，这个隐向量对应公式(4) 中的**一行**，因为 $参数矩阵V的行数 == 特征的数量$，所以 V_i就是特征 i 的隐向量，当 想要 表示 特征 i 与 其他 $n-1$ 特征组合的权重时， 特征 i 有且只能使用 V_i 一个隐向量

为了更好的理解这句话，我假设**老师上课**场景：

```
一个老师需要带40+个学生，同时在正式上课往往为了照顾到班级中大部分的学生，因此会选择一个合适的教学进度，但是这个会造成不可避免的问题：
（1） 优秀的学生吃不饱，觉得学的东西easy
（2） 基础薄弱的学生，觉得太难，必须加班加点才能跟得上老师的进度
这是因为老师必须照顾大部分学生的学习进度，而不得不选择的统一的教学进度
```

在老师上课的场景中，老师选择统一的**教学进度**就可以理解成FM中的 唯一的**隐向量**

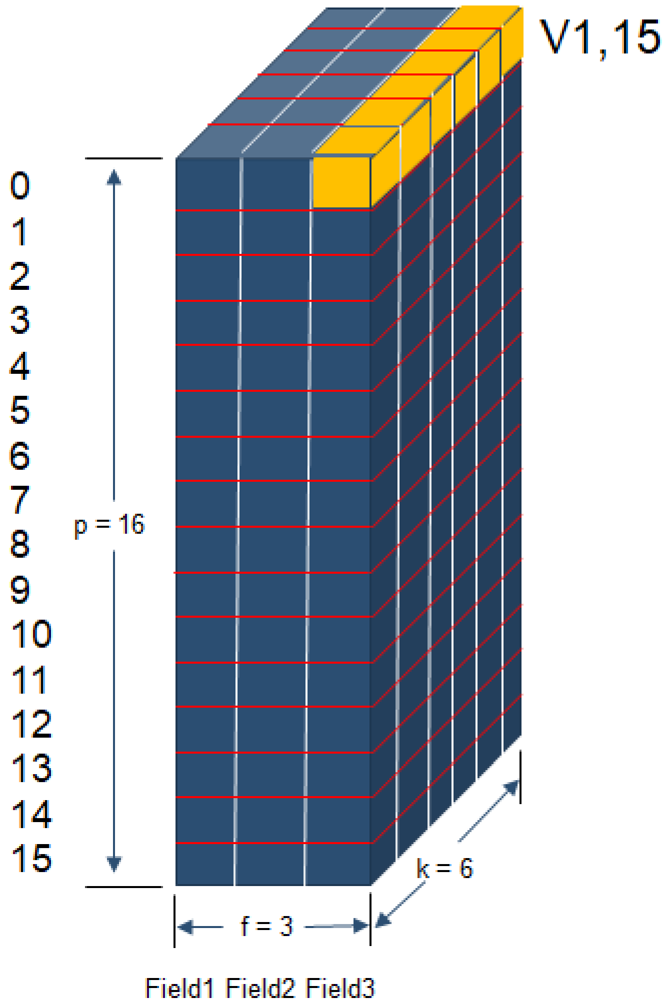
如果是**私教上课**:

```
私教上课就是一对一，根据每一位学生的具体状态进行调整合适的**教学进度(教学方案)**， 即有40个学生，就会有 F 个不同的教方案，F <= 40
```

哈哈哈，上述的**私教上课**就可以理解成 **FFM**，其中 F 就是 **Field-aware**，以下都会使用 F 表示field的数量

3、FFM原理

首先上FFM的参数矩阵图:



此图来自网络， 其中 p就是 我们这里说的 特征数量 $n(n=16)$ ，f 特征类别数量（私教老师的教学方案数），k 与FM中保持一致，表示 隐向量的 维度

与FM相比，FFM的参数变成了 三维， $n \times f \times k$

所以FFM的核心就可以利用私教上课来侧面理解了：特征 x_i 针对不同类别（Field)的特征，将会使用 指定类别（Filed）的 隐向量 $V_{i,f}$。

如果再次与FM相比，那么FFM就是有了 **千人多面**的性质，而不是FM的**千人一面**， 哈哈哈，这里的人在我们这里是指 n 个 特征。

到此，我们将会摆出FFM的最终的公式:

$$y = b + \sum_{i=1}^{n} w_i^* x_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n} < V_{i,fi}, V_{j,fj} > x_i \times x_j \tag{5}$$

其中，f_i 表示特征 x_i 对应的Field，同理，f_j 表示特征 x_j 对应的Field

总的来说就是知道 索引号 i, f_i 然后再去 FFM的参数矩阵中找到 指定的 k 维的 隐向量

4、FFM 特征组合demo

哈哈，一下的demo依然是来自网络 ,我就不重复造轮子了

原始数据

User	Movie	Genre	Price
YuChin	3Idiots	Comedy, Drama	$9.99

特征编号

Field name	Field index	Feature name	Feature index
User	1	User=YuChin	1
Movie	2	Movie=3Idiots	2
Genre	3	Genre=Comedy	3
Price	4	Genre=Drama	4
		Price	5

特征组合

$$\begin{aligned}\langle \mathbf{v}_{1,2}, \mathbf{v}_{2,1} \rangle \cdot 1 \cdot 1 + \langle \mathbf{v}_{1,3}, \mathbf{v}_{3,1} \rangle \cdot 1 \cdot 1 + \langle \mathbf{v}_{1,3}, \mathbf{v}_{4,1} \rangle \cdot 1 \cdot 1 + \langle \mathbf{v}_{1,4}, \mathbf{v}_{5,1} \rangle \cdot 1 \cdot 9.99 \\ + \langle \mathbf{v}_{2,3}, \mathbf{v}_{3,2} \rangle \cdot 1 \cdot 1 + \langle \mathbf{v}_{2,3}, \mathbf{v}_{4,2} \rangle \cdot 1 \cdot 1 + \langle \mathbf{v}_{2,4}, \mathbf{v}_{5,2} \rangle \cdot 1 \cdot 9.99 \\ + \langle \mathbf{v}_{3,3}, \mathbf{v}_{4,3} \rangle \cdot 1 \cdot 1 + \langle \mathbf{v}_{3,4}, \mathbf{v}_{5,3} \rangle \cdot 1 \cdot 9.99 \\ + \langle \mathbf{v}_{4,4}, \mathbf{v}_{5,3} \rangle \cdot 1 \cdot 9.99\end{aligned}$$

具体就不解释了，详情可以连接跳转一下，或者自己依据上面的 **栗子** 就行对对应

5、上代码

睡不着，起的太早，只能后续补上了，去睡个回笼觉了