

推荐系统入门系列(二)-FFM算法理论与实践

何无涯 何无涯的技术小屋 5月8日

点击蓝字，带你发现更大的世界

海纳百川，有容乃大；壁立千仞，无欲则刚！

—— 林则徐



一、FFM算法背景

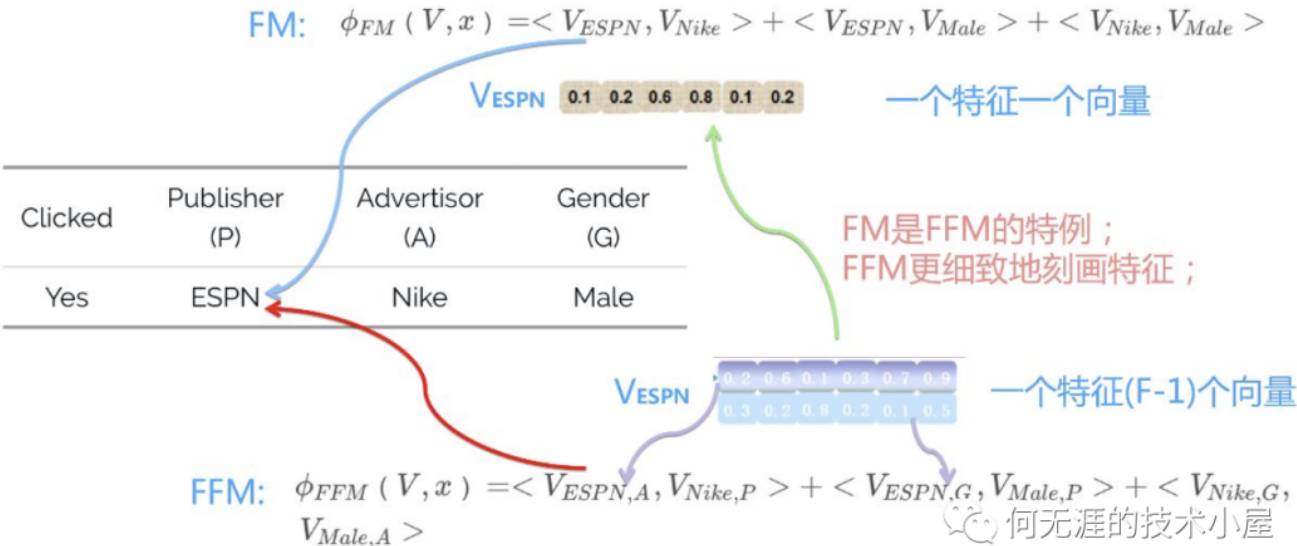
FFM算法（Field-aware Factorization Machine）是在FM算法的基础上所改进的一个算法，具体详情请见上一篇文章《推荐系统入门系列(一)-FM算法理论与实践》。FFM算法的效果比FM算法要好，然而参数量太大，接下来会详细介绍。

二、FFM算法思想

在讲FFM算法之前，先来回顾一下FM算法。FM算法是**为每一个特征学习到一个特征向量，组合特征的特征向量的内积为组合特征的一个权重**。如下面这个例子：

clicked	Publisher (P)	Advertiser (A)	Gender (G)
Yes	ESPN	Nike	Male

该例子中是要往ESPN网站投放广告，广告主Advertiser是Nike，读者Gender是Male男性，那么他会不会点击？该例子中是点击的。对于这个例子来说，FM算法干了什么？



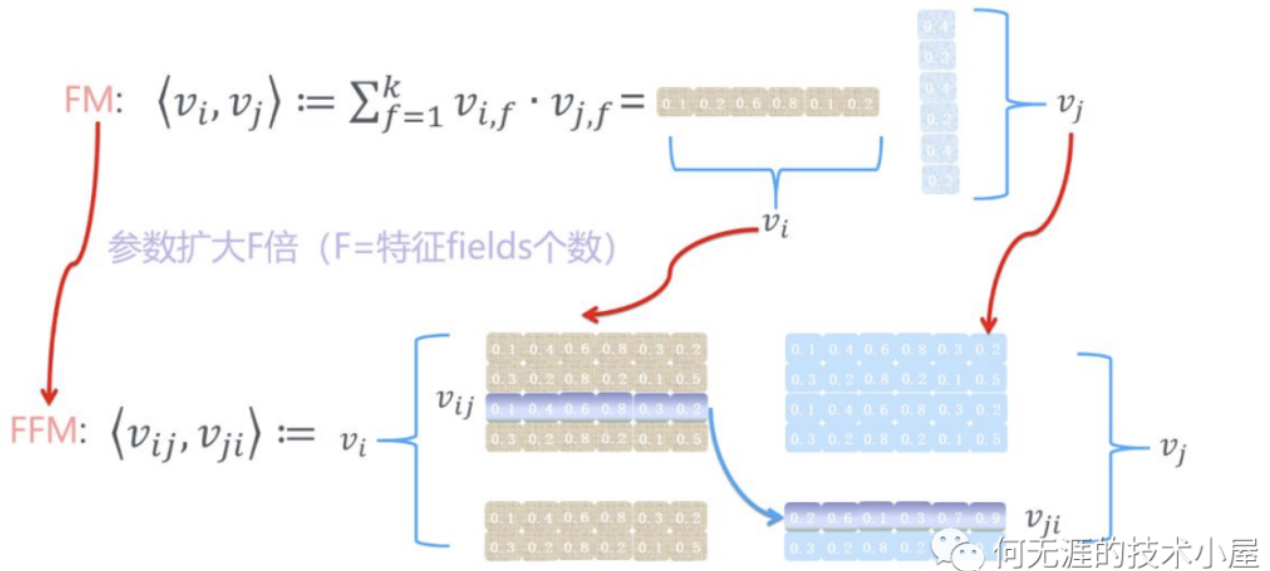
FM算法为每一个特征学习1个embedding，并且将组合的特征的embedding内积作为组合特征的权重。

那么FFM算法是怎么干的呢？在不同的特征组合中，FFM算法会为每一个特征会学习不同的embedding，也就是说为每一个特征学习到F个embedding（F为特征fields的

个数，如例子中 $F=3$ ），以与不同的特征进行组合。如上面例子，特征P与A、G组合时会有不同的embedding。

所以，对于一个特征来说，FM学习一个vector，而FFM学习到 $F$ 个vector，只要有跟其它特征的任意组合，就有一个vector来代表，这么做更细致的刻画了组合特征的信息，然而也导致参数量直接扩增 $F$ 倍。

## FFM模型



### 三、FFM算法实战

接下来，我们来使用PyTorch深度学习框架实现FFM算法，如下：

```
1 class FieldAwareFactorizationMachine(torch.nn.Module):
2     """Field Aware Factorization Machine layer"""
3     def __init__(self, field_dims, embed_dim):
4         super().__init__()
5         self.num_fields = len(field_dims)
6         # one field has F vectors
7         self.embeddings = nn.ModuleList([
8             torch.nn.Embedding(sum(field_dims), embed_dim) for _ in range(self
9         ])
10        self.offsets = np.array((0, *np.cumsum(field_dims)[: -1]), dtype=np.long
11        for embedding in self.embeddings:
12            torch.nn.init.xavier_uniform_(embedding.weight.data)
13
14        def forward(self, x):
15            """
16            :param x: Long tensor of size ``(batch_size, num_fields)``
17            """
```

```
18     x = x + x.new_tensor(self.offsets).unsqueeze(0)
19     xs = [self.embeddings[i](x) for i in range(self.num_fields)]
20     ix = list()
21     for i in range(self.num_fields - 1):
22         for j in range(i + 1, self.num_fields):
23             ix.append(xs[j][:, i] * xs[i][:, j])
24     ix = torch.stack(ix, dim=1)
25     return ix
```

完整的代码可以参考我的github: <https://github.com/yyHaker/RecommendationSystem>。

小结：FFM算法是FM算法的升级版，为每一个特征学习到F个向量，以代表不同的特征组合，效果好于FM，然而参数量巨大，直接扩大为FM算法的F倍。



何无涯的技术小屋

微信号码: leyanyuanyu

机器学习 | 深度学习 | 推荐算法 | NLP | 投资