FFM模型在新浪网点击率预估业务的实践总结

2016年3月

徐远东

CTR预估模型

- LR：google、腾讯、百度
- Mixed-LR：阿里巴巴
- GBDT-LR：facebook
- FFM：criteo比赛第一
- GBDT-FFM：avazu比赛第一
- FM/FFM尝试中：百度、新美大、新浪

LR模型

特征交叉项的稀疏性，参数个数：1+n+n(n-1)/2

Clicked?	Country	Day	Ad_type
1	USA	26/11/15	Movie
0	China	1/7/14	Game
1	China	19/2/15	Game

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

Clicked?	Country=USA	Country=China	Day=26/11/15	Day=1/7/14	Day=19/2/15	Ad_type=Movie	Ad_type=Game
1	1	0	1	0	0	1	0
0	0	1	0	1	0	0	1
1	0	1	0	0	1	0	1

问题1：扶翼投放日志中，有女性看尿布的广告，男性看啤酒的广告，是否可以预估男性看尿布广告点击率？

FM模型

缓解稀疏性：交叉特征参数 -> 单特征参数
参数个数：1 + n + nk

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

问题1：扶翼投放日志中，有女性看尿布的广告，男性看啤酒的广告，是否可以预估男性看尿布广告点击率？

问题2：FM的k取多少合适？

FFM模型

$$y(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_{i,f_j}, \mathbf{v}_{j,f_i} \rangle x_i x_j$$

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

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

FFM模型

不增加稀疏性，降低复杂性：单特征参数 -> 单特征分域参数

参数个数：1 + n + nfk

$$y(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_{i,f_j}, \mathbf{v}_{j,f_i} \rangle x_i x_j$$

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

问题3：FFM的参数个数增加了吗？

FFM模型在新浪网点击率预估业务的实践总结

FFM实现

https://github.com/guestwalk/libffm

1.随机梯度下降
2.自适应学习率AdaGrad
3.样本特征梯度分步计算和更新
4.OpenMP多线程扫描样本更新
5.SSE指令加速向量计算
6.内存紧凑减少换页
7.样本归一化

$$
\begin{aligned}
&model = \textbf{init}(tr.n, tr.m, pa)\\
&R_{tr} = \mathbf{1}, R_{va} = \mathbf{1}\\
&\textbf{if } pa.norm \textbf{ then}\\
&\quad R_{tr} = \textbf{norm}(tr), R_{va} = \textbf{norm}(va)\\
&\textbf{end if}\\
&\textbf{for } it = 1, \cdots, pa.itr \textbf{ do}\\
&\quad \textbf{if } pa.rand \textbf{ then}\\
&\quad\quad tr.X = \textbf{shuffle}(tr.X)\\
&\quad \textbf{end if}\\
&\quad \textbf{for } i = 1, \cdots, tr.l \textbf{ do}\\
&\quad\quad \phi = \textbf{calc}\mathbf{\Phi}(tr.X[i], R_{tr}[i], model)\\
&\quad\quad e\phi = \exp\{-tr.Y[i] * \phi\}\\
&\quad\quad L_{tr} = L_{tr} + \log\{1 + e\phi\}\\
&\quad\quad g_{\Phi} = -tr.Y[i] * e\phi/(1 + e\phi)\\
&\quad\quad model = \textbf{update}(tr.X[i], R_{tr}[i], model, g_{\Phi})\\
&\quad \textbf{end for}\\
&\quad \textbf{for } i = 1, \cdots, va.l \textbf{ do}\\
&\quad\quad \phi = \textbf{calc}\mathbf{\Phi}(va.X[i], R_{va}[i], model)\\
&\quad\quad L_{va} = L_{va} + \log\{1 + \exp\{-va.Y[i] * \phi\}\}\\
&\quad \textbf{end for}\\
&\textbf{end for}
\end{aligned}
$$

FFM衍生品

$$y(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_{i,f_j}, \mathbf{v}_{j,f_i} \rangle x_i x_j$$

1.FFM-ftrl
2.FFM-no−self−field-cross
3.FFM-no-1-order-bias
4.FFM-no-bias
5.FFM-only-cross-with-ad
6.FFM-hierarchy-ad

7.FFM-l1, l2, l1-v, l2-v
8.FFM k=1, 4, 8, 16
9.FFM norm-vs-nonorm
+.FFM init params

FFM：ada-grad vs ftrl

	FFM-adag	FFM-ftrl	FM	LR
L2-v	敏感	不敏感		
训练时间	140	160	100	20
Logloss/AUC	好	较好	中	非常一般
收敛速度（迭代次数）	慢	中	类FFM	快

FFM：ada-grad vs ftrl

FM/FFM	Logloss/AUC	真实环境准确率
All bias	好	中
No 1-order bias	中	好
No bias	差	差

FFM：structure variety

	standard	No-self-field-cross	Only-cross-with-ad	Cross−with-hierarchy-ad
Logloss/AUC	好	好	中	比中略好一点
训练时间	160+	160	100	110
收敛速度	慢	慢	快	最快
模型大小	1G	400M-800M	100M	100M

FFM模型在新浪网点击率预估业务的实践总结

FFM：K，norm，init，field

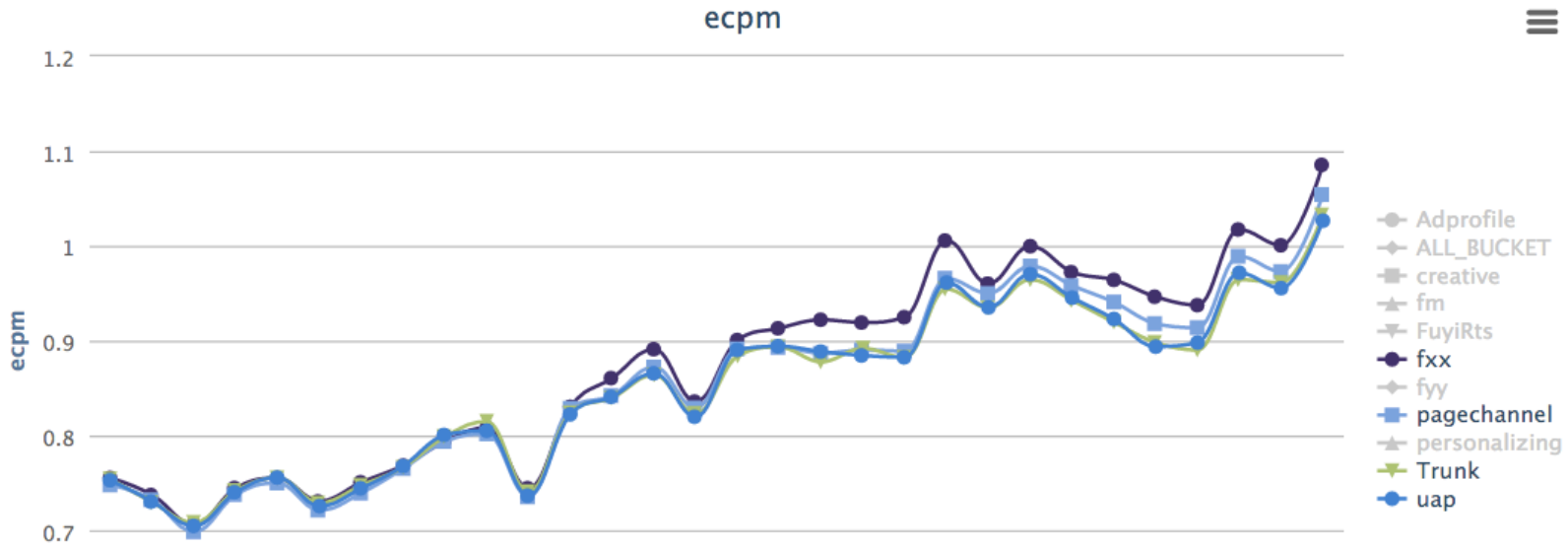
	K 1->4->8->16	Norm 0 -> 1	Init 常数->随机	Field 独立->共享
Logloss/AUC	差->好->差	差 -> 好	中 -> 随机	好->差
训练时间	递增			
过拟合	越来越易过拟合	易 -> 略不易	略不易 -> 易	略不易->易

Init：不同的随机初始点对最终结果非常大

FXX、FXXH：ABTesting

K=4, norm=1, L1=L2=L1v=0, L2v=0.0002, eta=0.2, init=常数/sqrt(K)
FXX: only-cross-with-ad
FXXH: only-cross-with-ad, and with ad hierarchy

http://10.210.228.84/dashboard/bucket_info.php



FXX训练流程

Gitlab：http://10.210.228.76/hero/sinaad_algo_ea_new_ctr_pipeline

1. 生成每日样本
2. 生成分广告位训练数据、测试数据、编码词典
3. MR streaming训练分广告位模型：
    1. 编码生成规范格式<label field:index:value …>
    2. 训练模型
    3. 测试AUC
    4. FFM模型文件到线上格式转换

注意：需要测试数据用于做验证集，和训练数据无交集，为什么？

FFM模型在新浪网点击率预估业务的实践总结

F4M工具包

Gitlab: http://10.210.228.76/hero/f4m

./fm –flagfile=params.conf –train=train.txt –test=test.txt
--model=m.txt –method=fxx –iterations=10 --negsampling=0.2

Params.conf：
--alpha=0.1
–beta=1
--alpha_fm=0.2
--l2_fm=0.0002
--init_dev=0.1
–nbit=16
–kdim=4
–mfield=18
--ad_fields=14,15,16,17
–stop_threshold=0.005

```
Use FFMXX (ada-grad) learner.
--------------------------------------------------------------------------------
normalization =1
dropout rate =0
neg sampling logr =-1.60944
--------------------------------------------------------------------------------
factorization K=4 M=18
2-order l2=2e-05 eta=0.1
w uniform init_coef=0.05
Use ad field = 14, 15, 16, 17,
--------------------------------------------------------------------------------
train from 57683.tr.
test from 57683.te.
feature hash space: 65536
Use neg sampling, rate=0.20
--------------------------------------------------------------------------------
pre-read elapsed time = 0.0s
iter tr_loss  va_loss  time
 1   0.47188  0.19559  80.4s
 2   0.46846  0.19552  80.2s
training samples = 8484455
testing samples = 3681310
elapsed time = 160.5s
```

F4M工具包

Gitlab: http://10.210.228.76/hero/f4m

./fm –flagfile=params.conf –test=test.txt –predict=pred.txt --model=m.txt –method=fxx –mode=1

cat pred.txt | ./fm –method=auc

```
Use FFM (ada-grad) learner.
====================================>
test from 56239.test.
set output predict file: c
feature hash space: 65536
====================================>
test samples = 2128558
logloss = 2.7133
elapsed time = 12.9s
```

```
Calculate AUC.
pv count: 2128558
click count: 25448
auc: 0.5798
logloss: 0.06718
```

F4M工具包

flags	含义	举例
mode	训练 or 测试	0(0)
train	训练数据路径	Train.txt，可以是/dev/stdin
cache	训练数据缓存路径	--train=/dev/stdin 时可用
test	测试集 or 验证集	Test.txt
predict	测试结果输出路径	Pred.txt，mode=0时可用
model	训练模型 保存 or 加载 路径	Model.txt
iterations	迭代次数	10(1)
method	模型选择	lr, fma, fmf, ffma, ffmf, fxx, fxxh, auc
ad_fields	标出哪些域是广告特征域	“14,15,16,17”，fxx或fxxh时可用，层次从高到低

F4M工具包

flags	含义	举例
l1	一次项一阶正则系数	0.1(0)
l2	一次项二阶正则系数	1.0(0，无)
l1_fm	二次项一阶正则系数	0(0)
l2_fm	二次项二阶正则系数	0.00002(0)
alpha	ftrl bias项学习率	0.1(0.1)
beta	用于ftrl bias项学习	1.0(1.0)
alpha_fm	ftrl，ada-grad 二次项学习率	0.2(0.1)
beta_fm	用于ftrl 二次项学习	1.0(1.0)
Init_dev	初始化随机标准差	0.1(0.1)

flags	含义	举例
nbit	特征空间2^nbit	16(0)
kdim	分解隐向量长度	4(4)
mfield	Ffm域个数	18(0)
normalization	是否样本归一化	1(1)
auto_stop	提前停止	1(1)
stop_threshold	提前停止阈值	0.005(0.01)
negsampling	负样本采样	0.5(1.0)
drop_rate	Dropout特征比例	0.0(0.0)
low_freq_filter	低频特征过滤	5(0)

提问环节

预祝各位带动新浪点击率再创新高！