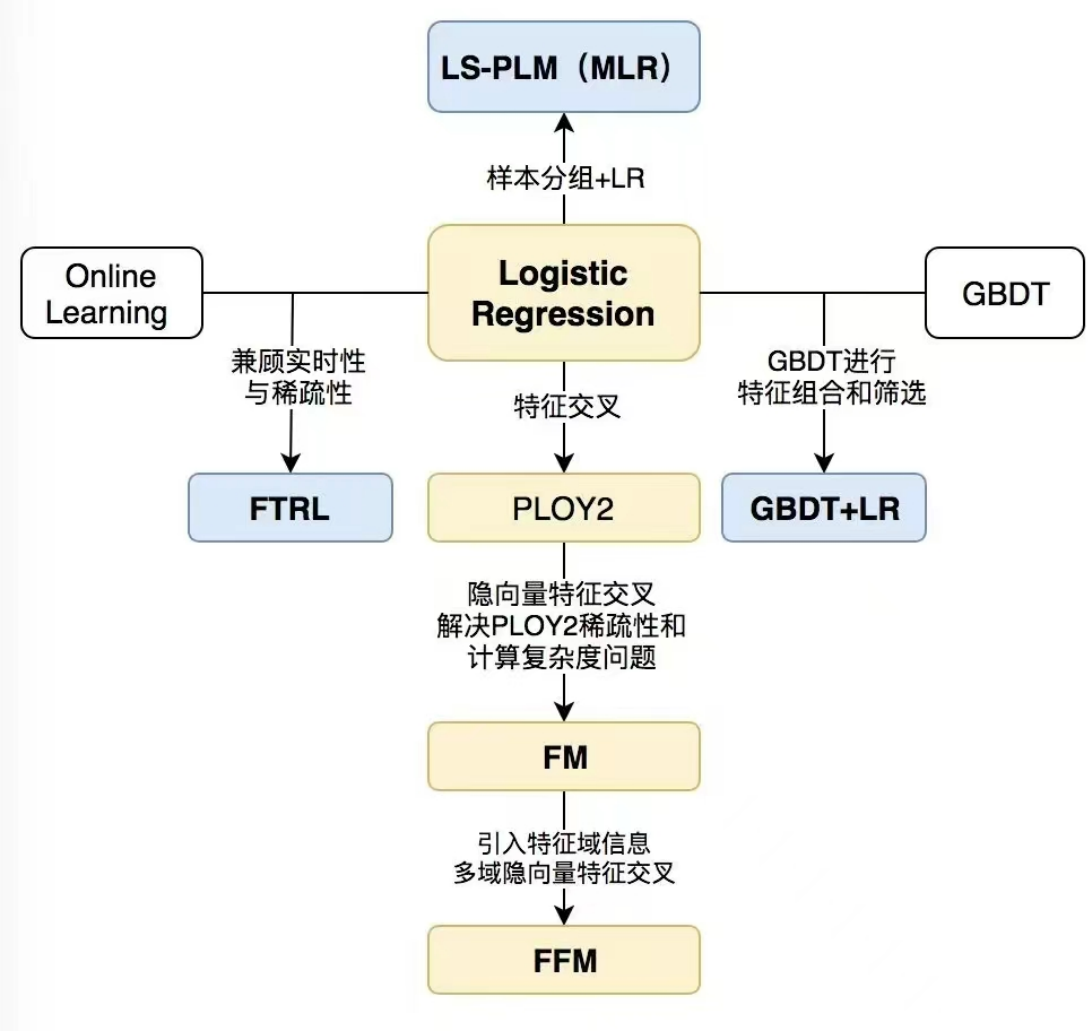
2023/9/19 Week1

张雨桐

Done：

围绕前深度学习时代传统推荐模型的演化，阅读经典论文。



协同过滤

• [Earliest CF] Using Collaborative Filtering to Weave an Information Tapestry (PARC 1992)

• [CF] Amazon Recommendations Item-to-Item Collaborative Filtering (Amazon 2003)

• [ItemCF] Item-Based Collaborative Filtering Recommendation Algorithms (UMN 2001)

协同大家的反馈、评价和意见一起过滤海量信息，从中筛选出目标用户可能感兴趣的信息。

步骤：生成共现矩阵，计算相似度，分为UserCF和ItemCF进行Top N推荐。

UserCF:用户相似度矩阵存储开销大&用户历史数据向量稀疏，适用发现热点, 新闻推荐场景.

ItemCF: 适用于兴趣变化较为稳定的应用，e.g.Amazon电商场景，Netflix视频推荐场景.

CF缺点：处理稀疏矩阵能力不足，头部效应明显，泛化能力较弱，仅利用用户和物品的交互信息。

矩阵分解

• [MF] Matrix Factorization Techniques for Recommender Systems (Yahoo 2009)

分解协同过滤生成的共现矩阵得到用户和物品的隐向量，对共现矩阵进行全局拟合。利用用户的隐向量与所有物品的隐向量进行逐一的内积运算。

求解方法：特征值分解、奇异值分解、梯度下降（主要方法）

优点：泛化能力强，空间复杂度低，更好的扩展性和灵活性。

缺点：不方便加入用户、物品和上下文相关的特征。

逻辑回归

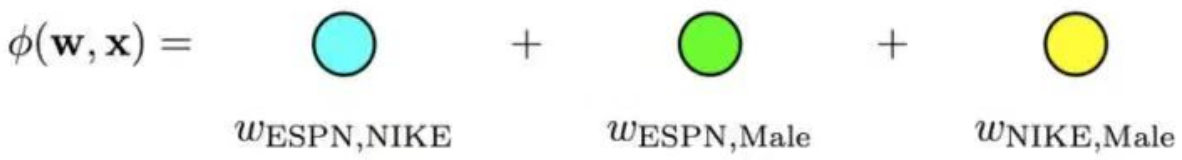
综合利用多种不同特征，将推荐看作分类问题，预测正样本的概率。CTR预估

利用样本的特征向量进行模型训练和在线推断，各特征的加权和，sigmoid映射到0-1区间。

缺点：表达能力不强造成有效信息的损失，无法进行特征交叉、特征筛选。

POLY2

$$\phi_{\mathrm{Poly2}}(\boldsymbol{w}, \boldsymbol{x}) = \sum_{j_1=1}^{n} \sum_{j_2=j_1+1}^{n} w_{h(j_1,j_2)} x_{j_1} x_{j_2}$$



特征交叉的开始，通过暴力组合特征的方式一定程度上解决了特征组合的问题。

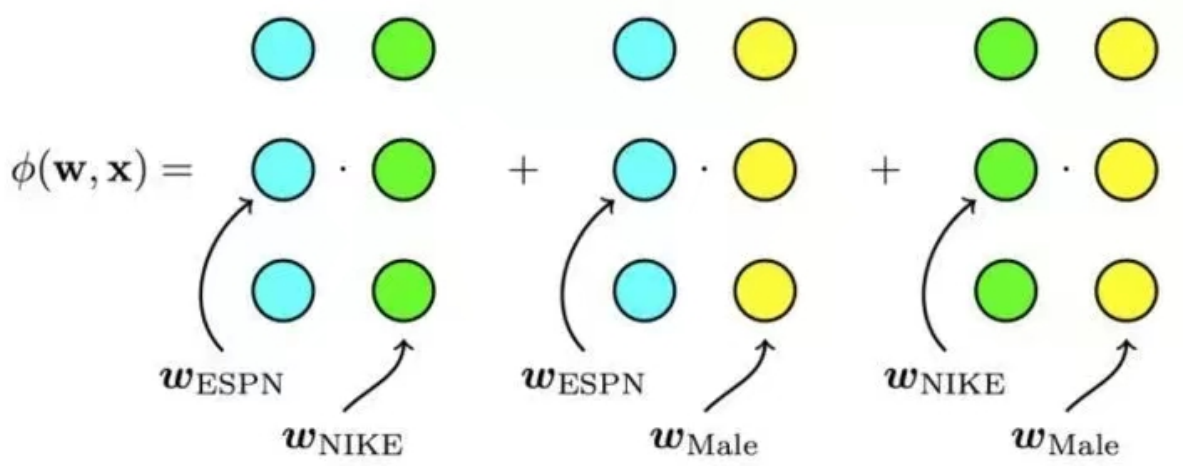
若特征数量为n，则权重数量为n^2级。

FM

https://www.csie.ntu.edu.tw/~b97053/paper/Rendle2010FM.pdf

[FM]Factorization Machines, 2010.

$$\phi_{\text{FM}}(\boldsymbol{w}, \boldsymbol{x}) = \sum_{j_1=1}^{n} \sum_{j_2=j_1+1}^{n} (\boldsymbol{w}_{j_1} \cdot \boldsymbol{w}_{j_2}) x_{j_1} x_{j_2}$$



使用两个特征隐向量的内积作为交叉特征的权重，具备了计算特征组合权重的能力。

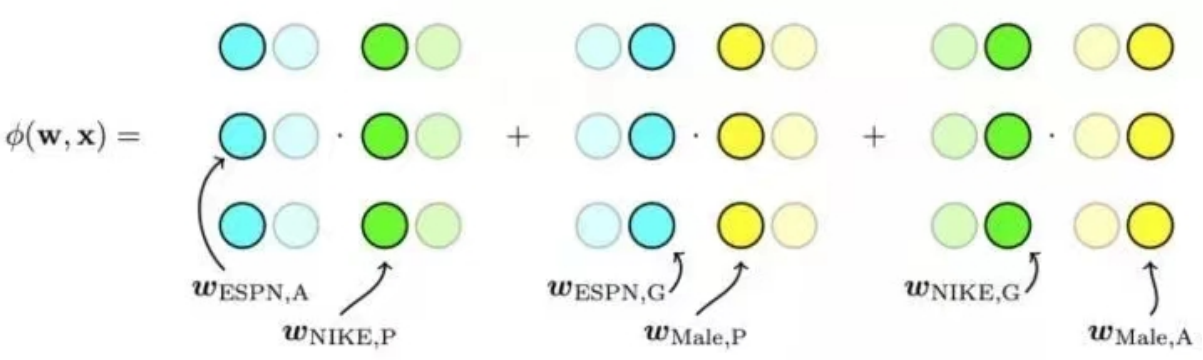
将POLY2 n^2级别的权重参数数量减少到了nk级别，极大的降低了训练开销。12-14业界

FFM

https://www.csie.ntu.edu.tw/~cjlin/papers/ffm.pdf

[FFM]Field-aware factorization machines for CTR prediction, 2016.

$$\phi_{\mathrm{FFM}}(\boldsymbol{w}, \boldsymbol{x}) = \sum_{j_1=1}^{n} \sum_{j_2=j_1+1}^{n} (\boldsymbol{w}_{j_1, f_2} \cdot \boldsymbol{w}_{j_2, f_1}) x_{j_1} x_{j_2}$$



引入特征域感知，使模型的表达能力更强。每个特征对应一组隐向量，每个特征与对方域对应的隐向量做内积运算，得到交叉特征的权重。

学习n个特征在f个域上的k维隐向量，参数数量共nkf个。

三阶特征交叉：权重数量&训练复杂度过高，难以工程实现。

GBDT+LR

https://quinonero.net/Publications/predicting-clicks-facebook.pdf

[GBDT+LR]Practical lessons from predicting clicks on ads at facebook, 2014.

特征工程由一个独立的模型来完成。把GBDT所有子树的特征向量连接起来，形成了LR模型输入的离散特征向量，并输入LR模型，进行CTR预估。开启特征工程新趋势。

LS+PLM

1704.05194

[LS+PLM]Learning piece-wise linear models from large scale data for ad click prediction, 2017.

加入了注意力机制的三层神经网络模型，具备较强的表达能力。

对样本进行分片，在每个分片内部构建逻辑回归模型，将每个样本的各分片概率与逻辑回归的得分进行加权平均，得到最终的预估值。

端到端的非线性学习能力，稀疏性强。

To Do:

深度学习推荐模型的演化+推荐系统的评估