CTR - Factorization Machines

FM和多项式核的支持向量机类似，可以学习交叉特征，但解决了SVM在数据稀疏的情况下必须要求交叉项全部非零才可以进行训练，难以准确找到超平面的问题，同时，FM仅依赖于线性多个参数而不依赖任何支持向量之类的实际样本。
FM的具体做法是将原本稀疏的输入进行因式分解，每个维度都乘以一个embedding

优势

i. 可以在数据极其稀疏的条件下对交叉特征进行训练；
ii. FM的计算可以控制在线性复杂度内，可以采用SGD进行训练；
iii. FM是一种普适性较强的预测模型，可以适用于任何实值特征输入；

FM结构

二阶FM模型的公式如下：

$$\hat{y}(x) := w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} < V_i, V_j > x_i x_j$$

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

其中，k是一个超参数，代表因子个数，由于数据比较稀疏，所以k值不应过大，否则将难以训练。

由上式容易发现，上式的复杂度是 $O(kn^2)$, 通过一系列的数学推导，可以将上式交叉项转换为：

$$\sum_{j_1=1}^{d} \sum_{j_2=j_1+1}^{d} < V_i, V_j > x_{j_1} x_{j_2} = \frac{1}{2} \sum_{f=1}^{k} ((\sum_{i=1}^{n} v_{i,f} x_i)^2 - \sum_{i=1}^{n} v_{i,f}^2 x_i^2)$$

采用上式进行替换可以讲FM的复杂度从 $O(kn^2)$ 降到 $O(kn)$

FM的参数包括 w_0, W, V，可以通过随机梯度下降法进行求解。

CTR - Field-aware Factorization Machines

FFM是FM的一个变种模型，FFM模型将所有的离散变量转化为二进制列，每个离散属性转化为的二进制列称为一个filed，假设要考虑 j_1, j_2 两个属性之间的交叉，其中$j_1 \in f_1, j_2 \in f_2$ ，FM当中是计算

$$< V_i, V_j > x_{j_1} x_{j_2}$$

而FFM是计算 j_1 和 f_2 之间的相互作用与 j_2 和 f_1 之间相互作用：

$$(w_{j_1,f_2} \cdot w_{j_2,f_1}) x_{j_1} x_{j_2}$$

FFM模型结构

$$\min_w \frac{\lambda}{2} ||w||_2^2 + \sum_{i=1}^{m} log(1 + exp(-y_i \phi_{FFM}(w, x_i)))$$

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

使用FFM的小技巧

- 使用AdaGrad进行优化
- 连续值标准化
- 连续属性离散化
- 将每个离散属性看做一个field
- FFM对epoch敏感，需要确定一个比较好的epoch
- 可以采用hash trick进行one hot
- 超参数选择时，选择较小的 λ, 较大的 η
- 容易过拟合，早停策略能够更好的抑制过拟合
- 适用于包含较多离散属性的数据，适用于转换后稀疏的数据

待改善的地方

- 抑制过拟合的措施
- 梯度下降的优化方法

启发

模型结构确定后，在多个数据集上进行实验，确定该模型适用的数据特征。