

编 程 学 习 笔 记

推荐算法

Code

作 者 姓 名： 颜佳

目 录

[1 第一章 1](#_Toc76209995)

[1.1 第一节 1](#_Toc76209996)

[1.1.1 第一段 1](#_Toc76209997)

[1.2 1](#_Toc76209998)

[2 1](#_Toc76209999)

AutoInt（Automatic Feature Interaction Learning via Self-Attentive Neural Networks）

# BST（Behavior Sequence Transformer）

# DADCN

# DALE

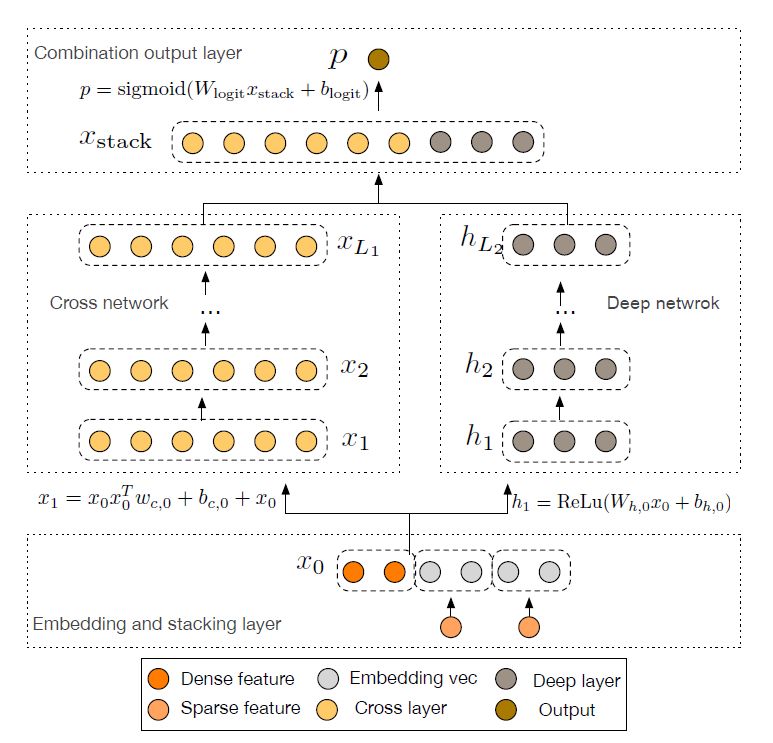
# DAPLE

# DCN（Deep & Cross Network）

## 简介

DCN全称Deep & Cross Network，是谷歌和斯坦福大学在2017年提出的用于Ad Click Prediction的模型（《Deep & Cross network for Ad Click Predictions》）。DCN在学习特定阶数组合特征的时候效率非常高，而且同样不需要特征工程，引入的额外的复杂度也是微乎其微的。DCN是一个可以同时高效学习低维特征交叉和高维非线性特征的深度模型。

## 架构



DCN模型分成4部分：最底下是Embedding and Stacking Layer，中间部分是并行的Cross Network和Deep Network，最上面是Combination Layer，把Cross Network和Deep Network的结果组合得到Output。

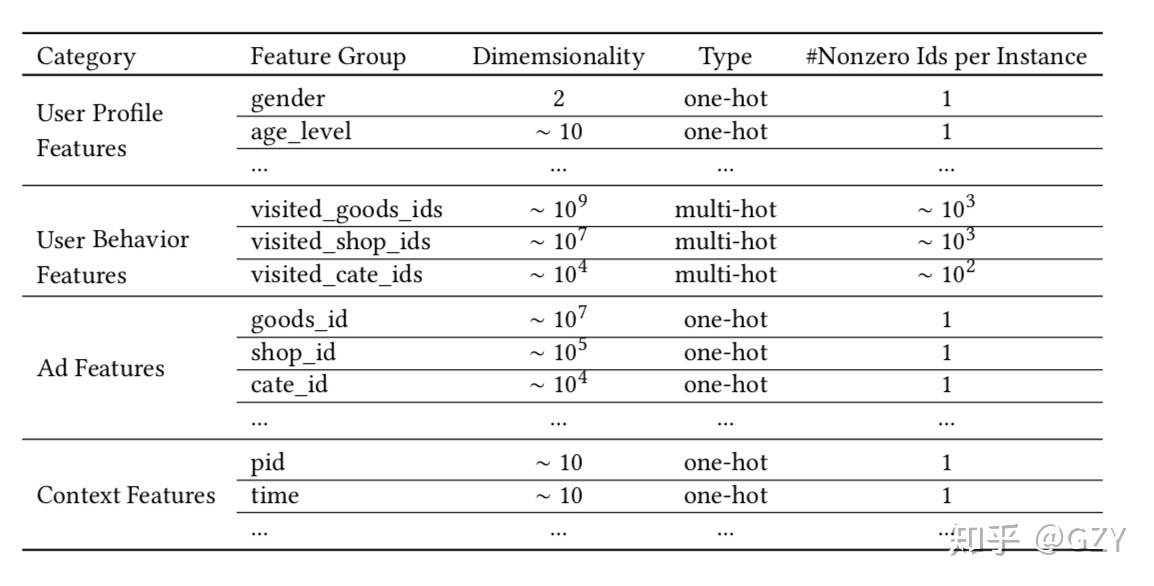
### Embedding and Stacking Layer（特征预处理输入）

这一层是模型的输入层。输入的特征分为稠密和稀疏特征。对于稀疏特征，我们一般采用one hot encoding的方式进行编码，然后再经过一层embedding将one hot embedding转换成稠密向量。最后，将稠密特征和经过转换的稀疏特征对应的稠密向量concat起来组成模型的最终输入，也就是图中的x0。

#### Embedding

在web-scale的推荐系统比如CTR预估中，输入的大部分特征都是类别型特征，通常的处理办法就是one-hot，但是one-hot之后输入特征维度非常高。所以需要使用Embedding来降低输入的维度，就是把这些binary features转换成dense vectors with real values。

Embedding操作其实就是用一个矩阵和one-hot之后的输入相乘，也可以看成是一次查询（lookup）。这个Embedding矩阵跟网络中的其他参数是一样的，是需要随着网络一起学习的。



假设第个特征组的维度为，为第个embedding字典，是维度的embedding向量。

如果是一个one-hot向量，且，则的embedding表示是一个单一向量。

如果是一个multi-hot向量，且，则的embedding表示是一个向量。

不同的用户有不同数量的历史行为，即multi-hot行为特征的向量会导致所产生的embedding向量列表的长度不同，而全连接需要固定长度的输入。一个通用的方法去转化embedding向量列表是通过Pooling层去获得固定长度：

#### Stacking

处理完类别型特征，把连续型特征规范化之后，和嵌入向量stacking到一起，就得到了原始的输入：



### Cross Network（自动化特征显式交叉）

该层用来高效地学习组合特征。形式如下：



其中：

——第层交叉网络的输出；

、——第层交叉网络的权重参数和偏置项；

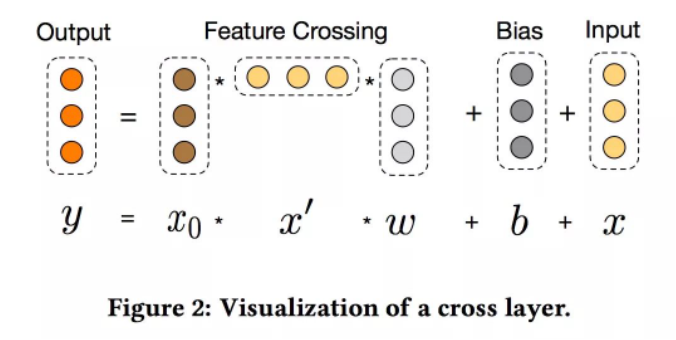
每一层的输出都是上一层的输出加上特征交叉函数。即，就是在拟合该层输出和上一层输出的残差。

为了方便说明交叉特征的形式，是2维向量，令



从上式可以看出Cross Network叠加到层时，可以得到阶的特征交叉，并包含所有的交叉组合。因此Cross Network可以通过对叠加层数的控制，可以高效地学习出低维的特征交叉组合，避免了人工特征工程，同时克服了传统DNN提取高阶特征效率不高的缺点。

Cross Network通过调整结构层数能够构造出有限阶（bounded-degree）交叉特征，对特征进行显式交叉编码，在精简模型参数的同时有效的提高了模型的表征能力。



复杂度分析

Cross Network特殊的网络结构使得交叉特征的阶数随着网络层数的增加而增加。假设一共有Lc层cross layer，起始输入x0的维度为d，因为每一层的W和b都是d维度的，那么整个cross network的参数个数为：



从上式可以发现，复杂度是输入维度d的线性函数。所以相比于deep network，cross network引入的复杂度微不足道。

因为，先计算得到一个标量，然后再计算与的乘积，这样就不用计算并存储整个的矩阵就可以得到所有的cross terms。

### Deep Network（特征隐式交叉）

因为cross network的参数比较少导致它的表达能力受限，为了能够学习高度非线性的组合特征，DCN并行的引入了Deep Network。Deep Network是一个前向传播的全连接神经网络。



其中：

——第层网络的输出；

、——第层网络的权重参数和偏置项；

——relu激活函数；

假设输入x0维度为d，一共有Ld层神经网络，每一层的神经元个数都是m个。那么总的参数或者复杂度为：



### Combination Layer

Combination Layer把Cross Network和Deep Network的输出拼接起来，然后经过一个加权求和后得到logits，然后经过sigmoid函数得到最终的预测概率。



p是最终的预测概率；XL1是d维的，表示Cross Network的最终输出；hL2是m维的，表示Deep Network的最终输出；Wlogits是Combination Layer的权重；最后经过sigmoid函数，得到最终预测概率。

模型损失函数为：



# ESMM（Entire Space Multi-Task Model）

# FEFM（Field-Embedded Factorization Machines）

# FIBINET

# FINAL

# MLDAG

# MMOE

word如何复制粘贴文本时不带自动编号和项目符号

# MMOE2

# MVKE

# PLE

# PLE2

# PNN

# PUAL