

# 推荐系统课程实验报告

## 针对 CTR 任务的深度学习模型

霍超凡

### 1 实验概述

CTR 任务是在给定用户信息（历史点击或用户标签）的前提下，预测他是否会点击推荐给他的广告，CTR 任务极具有商业价值，广告收费标准就是根据广告点击率计算的，一个好的 CTR 模型能够为广告商带来丰厚的利润，本次实验讨论了几种解决 CTR 任务的深度学习模型，由于对模型的理解程度不够，只能通过实验来对这几种模型有一个直观上的认识。

### 2 深度学习模型

#### 2.1 深度前向网络

如图 1 所示，DNN（Deep Neural Network）是由几层全连接层堆叠形成，输入特征经过 embedding 层被表示成向量，稀疏向量被转换成密集向量，特征向量经过层层全连接层进行特征提取与组合，最后一层的全连接层的输出节点只有两个，分别表示广告是否被用户点击。这个深度网络需要根据问题的规模指定网络内部隐藏层的层数和每一层网络的“宽度”，由于对网络各个层的含义不理解，调节这些参数是困难的。实验计划探讨以下两个问题：①一个深层的网络可以获得较好的效果，相比浅层网络，性能的提升究竟是参数的增加还是网络结构的优势导致的？②深度网络也可以综合各级特征，那么 DeepFM 中特征两两相乘综合特征的方法相对于它有什么优势？在参数基本相同的情况下，DeepFM 网络真的比 DNN 网络好吗？

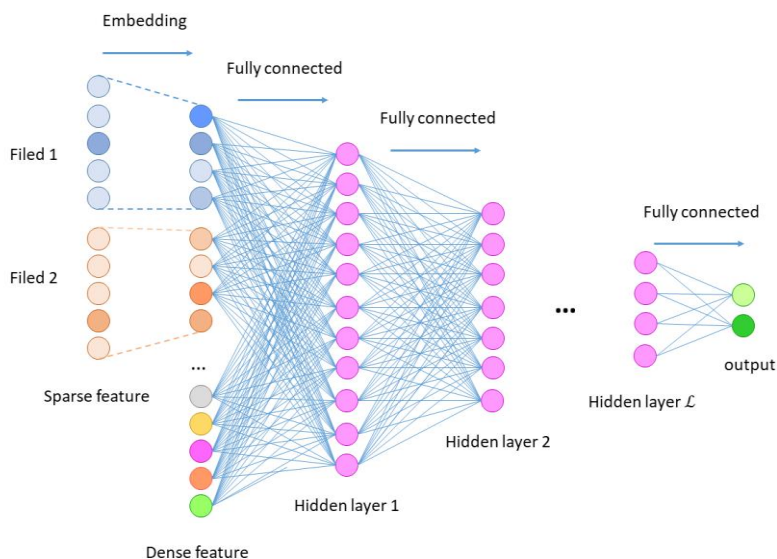


图 1：DNN 网络结构

## 2.2 具有跳跃连接的 DNN

在图像识别领域，经过了 VGG、ResNet、DenseNet、FPN 等网络结构的发展，这些网络逐步提升关键在于引入了不同程度的跳跃连接，这些连接使底层特征能够和高层特征相融合，特征复用增加了模型的表达力，使模型即使在较少的参数的前提下可以拟合复杂的函数。那么类似的思想是否可以应用到解决 CTR 任务的模型中呢？如图 2 所示，将跳跃连接引入到 DNN 中得到如下几种网络结构，原本计划验证带有跳跃连接之后的特征提取网络的确可以利用很少的参数达到很好的性能，而后发现数据集的问题并且时间也不够了，所以带有跳跃连接的模型没有实现。

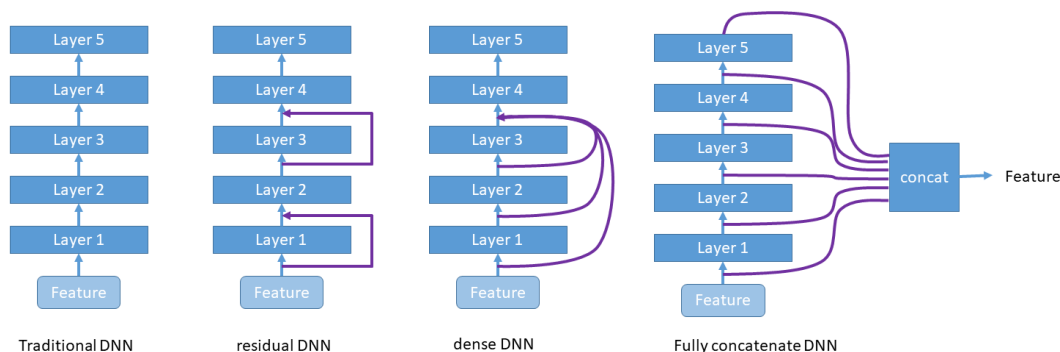


图 2：带有跳跃连接 DNN 网络结构

## 2.3 其它综合特征的模型

像 Deep & Cross Net、DIN、DeepFM 和 xDeepFM 这些网络都是试图通过特征之间的交互进而提升模型的性能，将特征两两相乘，再加权求和，我现在还想不明白这种特征直接相乘的交互方式有什么优势。DeepFM 在 FM 的基础上增加了一个 DNN，而 xDeepFM 增添了 external product FM 来实现无限阶特征交互，实验只再现了 DeepFM，或许是数据集的问题也或者模型的参数的问题，DeepFM 去掉 FM 之后的 DNN 和原 DeepFM 具有不相上下的性能。

## 3 实验细节

实验中，使用 Criteo 数据集验证各个模型的性能。衡量指标像原始论文一样取对数损失和 AUC，使用 Paddle 框架将实验环境搭建在 Baidu 所提供的服务器上。实验首先探讨了 DNN 深度和宽度对结果的影响，参考[1]，为了控制宽度和深度模型参数一致，使用类似于 RNN 的神经网络结构，这种结构如图 3 所示，数据流重复通过一层网络多次来隐式增加模型的深度，把这种结构的网络暂且称作 DRNN (Deep Recurrent Neural Network)，由于层与层之间共享权重参数，所以这种网络在参数量较少的情况下也能达到很好的性能。为了方便之后的说明，这里为说明以下网络结构的命名，例如 DRNN-16-512-3，表示 DRNN 的深度为 3，宽度为 512，这里的 16 表示的是 embedding 密集向量的维度，结合在 2.1 所说的网络结构，DRNN-16-512-3 网络结构第一层是 embedding 层将稀疏向量映射成 16 维密集向量，中间层改为 DRNN 结构，最后一层仍然是全连接层。之后的实验通过调整 DRNN 网络结构的宽度和深度来探究 2.1 节提出的问题。

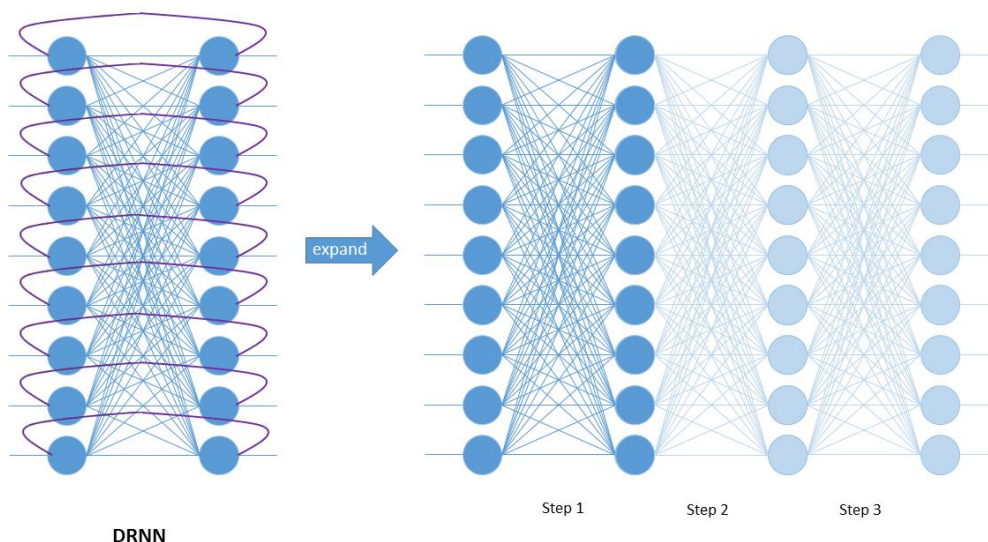


图 3: DRNN 网络结构及其时序展开图

## 4 实验结果

目前还在调试中，遇到的困难是：Criteo 数据量比较大，跑一遍整个训练集需要将近 50 分钟，这意味着在短时间内无法对这些模型进行大量的实验，在训练模型的过程中也遇到一些疑惑，模型在第一轮训练中，损失就可以减小到 0.66 左右，AUC 达到 0.76 左右，之后再训练后 AUC 仅仅是 0.01 数量级的提升，一些模型达到 0.78 就基本收敛，我看到论文中这些模型的 AUC 也在 0.8 上下，不同模型之间的 AUC 差异很小，我实现 deepFM 论文中的 DNN 模型时，性能没有达到论文所说的 0.8+，从第 4 轮训练之后就基本稳定在 0.78，具体问题还在检查。

我重新参考了另外一个示例程序，把数据集预处理、数据集加载修改了之后，似乎是恢复正常了，训练出 DeepFM 模型在验证集上的 AUC 为 0.80+ 和原论文的差不多，但是把 FM 去掉的 DNN 的 AUC 也为 0.80+，原始论文为 0.79+，或许我的参数调节的不同，我设置的 embedding 的维度为 16 维，三层 DNN 的宽度依次为 400, 400, 400。我调节参数时还发现，单层 DRNN-16-1024-1 的 AUC 都可以达到 0.804+，把 embedding 的维度减为 1 的 DRNN-1-512-1 的 AUC 都可以达到 0.800+，我开始怀疑 Criteo 数据集本身就不存在问题，目前我看到的论文中 xDeepFM 在该数据集上的 AUC 为 0.805+ 最高。

经过一系列实验之后，得到表 1 所示的结果，一开始没有经验把 embedding 的维度设成 16，模型明显会过拟合，模型训练一轮后在验证集上的 AUC 就达到最高，之后逐渐降低。对比第一组数据去掉 FM 的 DNN 要比 DeepFM 略高，这不排除因为过拟合而导致 DeepFM 在验证集上的性能下降，DNN++ 是 DeepFM 和 DNN 的一个折中，把 DeepFM 的 FM 层换成全连接层，FM 的一阶交互不变仍然是单个特征加权求和，二阶交互改为 DNN 层中隐藏层的第一层输出特征的加权求和。第二组实验对比了不同宽度的 DRNN 模型，embedding 层所占的参数比重较大，所以宽度从 1024 减小到 2 后，模型的大小基本维持在 100MB 左右。第三组实验对比了 embedding 层向量的维度对结果的影响，从 DRNN-16 到 DRNN-1，模型大小减少了大约 15 倍，但是模型的 AUC 才减少了 0.003。第四组实验对比深度对结果的影响，基本没影响，从第 5 组的实验数据发现，深度加深可以略微提高模型的性能。

越做越发现，在数据集上为了争夺那零点零几的 AUC 似乎意义不大。

表 1: 各种模型在 Criteo 数据集上的性能

	Model	Size(MB)	LogLoss	AUC
#1	DeepFM-16-400-400-400	109.10	0.4480	0.8023
	DNN-16-400-400-400	109.07	<b>0.4468</b>	<b>0.8037</b>
	DNN+-16-400-400-400	108.94	0.4468	0.8036
#2	DRNN-16-1024-1	109.25	<b>0.4464</b>	<b>0.8040</b>
	DRNN-16-512-1	103.19	0.4468	0.8035
	DRNN-16-256-1	100.71	0.4475	0.8028
	DRNN-16-128-1	100.23	0.4483	0.8019
	DRNN-16-2-1	98.769	0.4607	0.7880
#3	DRNN-16-512-1	103.19	<b>0.4468</b>	<b>0.8035</b>
	DRNN-8-512-1	52.434	0.4474	0.8031
	DRNN-4-512-1	26.883	0.4476	0.8030
	DRNN-1-512-1	7.7479	0.4501	0.8004
#4	DRNN-4-512-3	26.693	0.4475	0.8030
	DRNN-4-512-2	26.806	<b>0.4474</b>	<b>0.8031</b>
	DRNN-4-512-1	26.883	0.4476	0.8030
#5	DRNN-1-512-4	7.8114	<b>0.4497</b>	<b>0.8009</b>
	DRNN-1-512-1	7.7479	0.4501	0.8004

之后我由增添了几组实验，之前由于 embedding size 设得太大导致模型过拟合，表 1 中 DNN 要比 DeepFM 强，似乎是由于 DeepFM 过拟合程度高于 DNN 导致的，我将 embedding size 下调至 1，重复训练模型，得到图 1 所示的结果，DeepFM 在训练时拟合训练数据的能力强，但是 DeepFM 在测试集上的泛化能力不如 DNN++。

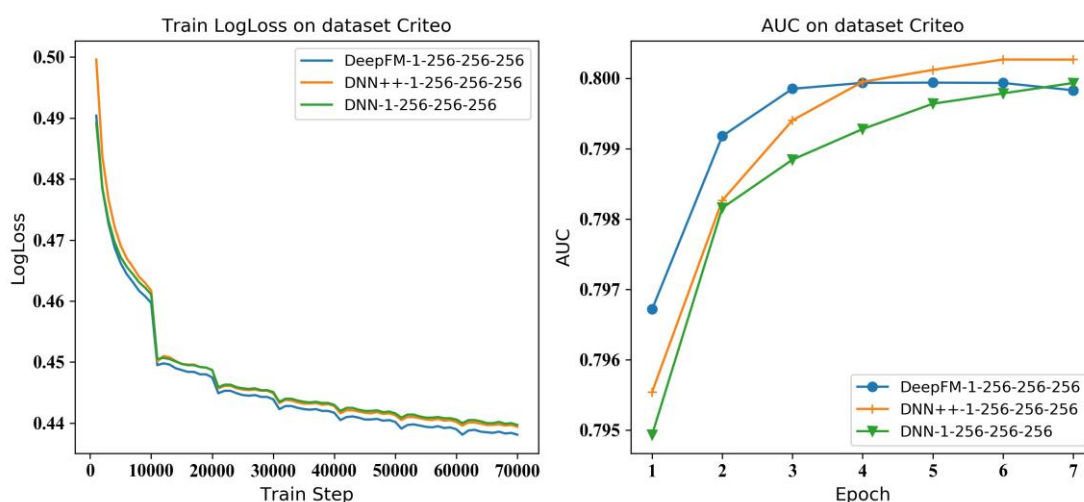


图 1: DeepFM、DNN++、DNN 这三种网络结构在数据集 Criteo 上的训练损失和测试 AUC

调节 DRNN 的深度和宽度，得到如图 2 所示的结果，这张图验证了深度增加或宽度增加提高模型的表现力和泛化能力。

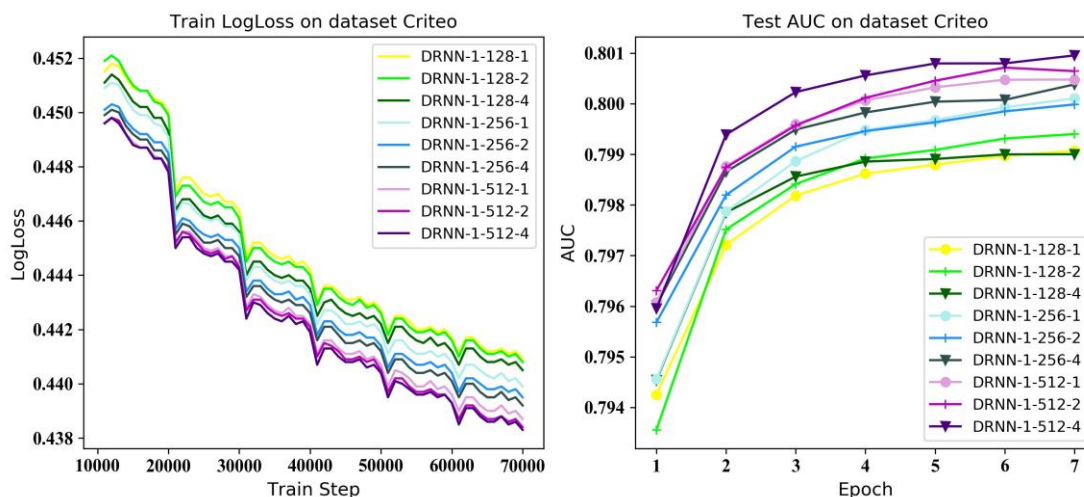


图 2：DRNN 这三种网络结构在数据集 Criteo 上的训练损失和测试 AUC

## 参考文献

- [1] Eigen, David & Rolfe, Jason & Fergus, Rob & Lecun, Yann. (2013). Understanding Deep Architectures using a Recursive Convolutional Network.
- [2] Guo, Huifeng & Tang, Ruiming & Ye, Yunming & Li, Zhenguo & He, Xiuqiang. (2017). DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. 1725-1731. 10.24963/ijcai.2017/239.
- [3] Rendle, Steffen. (2010). Factorization Machines. 995-1000. 10.1109/ICDM.2010.127.
- [4] Lian, Jianxun & Zhou, Xiaohuan & Zhang, Fuzheng & Chen, Zhongxia & Xie, Xing & Sun, Guangzhong. (2018). xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems.
- [5] Wang, Ruoxi & Fu, Bin & Fu, Gang & Wang, Mingliang. (2017). Deep & Cross Network for Ad Click Predictions. 1-7.
- [6] Paddle CTR models, site:  
<https://github.com/PaddlePaddle/models/tree/release/1.8/PaddleRec/ctr>