

技术解读 | Deep和Cross不得不说的秘密

原创 c0de 云脑科技CloudBrain 2018-07-04

Deep & Cross Network(DCN)[1]是来自于 2017 年 google 和 Stanford 共同完成的一篇工作，对比同样来自 google 的工作 Wide & Deep[2]，DCN **不需要特征工程来获得高阶的交叉特征**，对比 FM 系列[3][4]的模型，DCN **拥有更高的计算效率并且能够提取到更高阶的交叉特征**。

模型简介

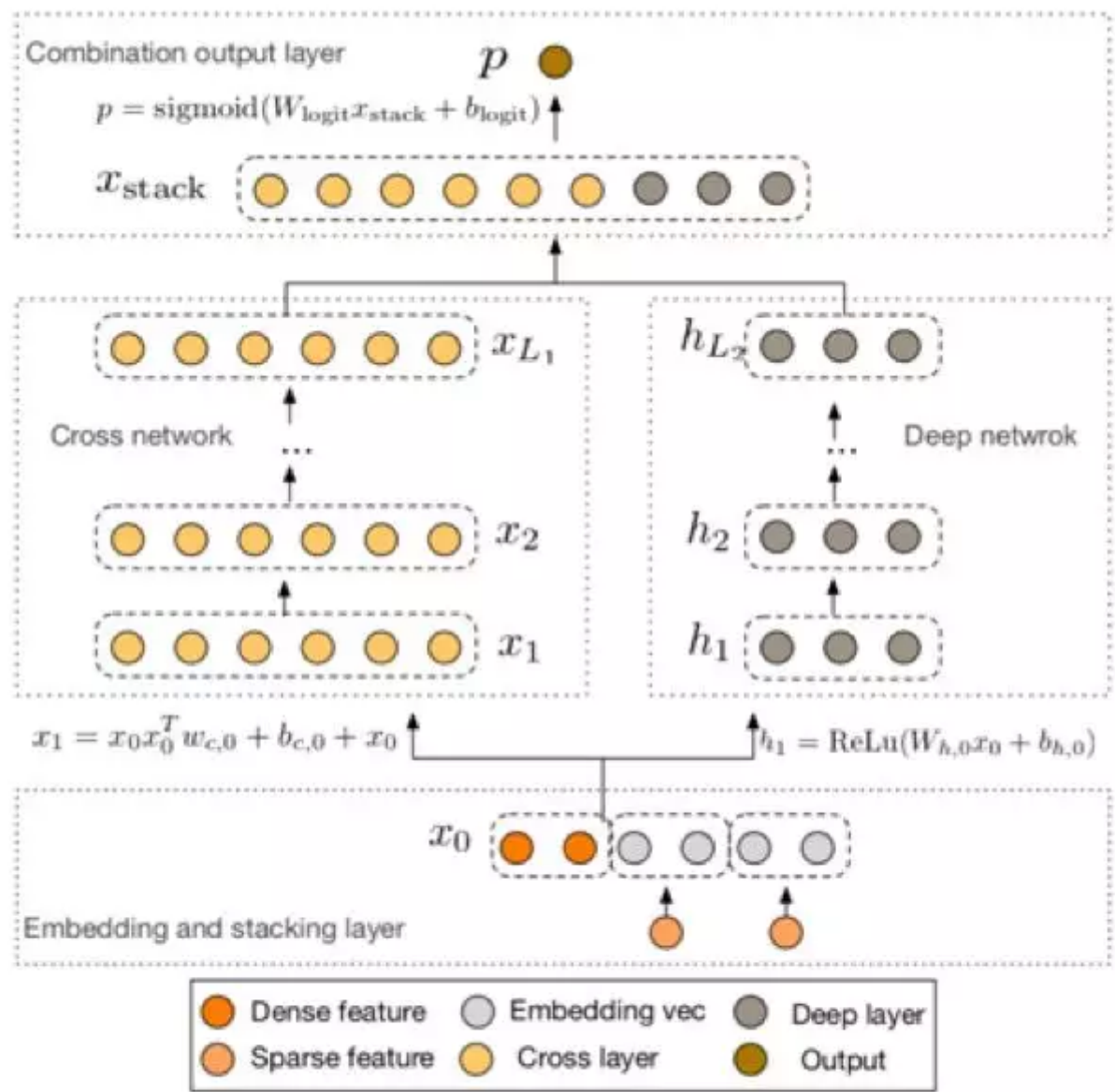


Figure 1: The Deep & Cross Network

从网络结构上面来看，该模型是非常简单明了的，特征分为类别型与数值型，类别型特征经过 embedding 之后与数值型特征直接拼接作为模型的输入。所有的特征分别经过 cross 和 deep 网络，如果把这两个网络看作特征提取的话，经过提取后的特征向量拼接之后是常规的二分类，如果训练数据是曝光和点击，最后输出的就可以看作点击率了。

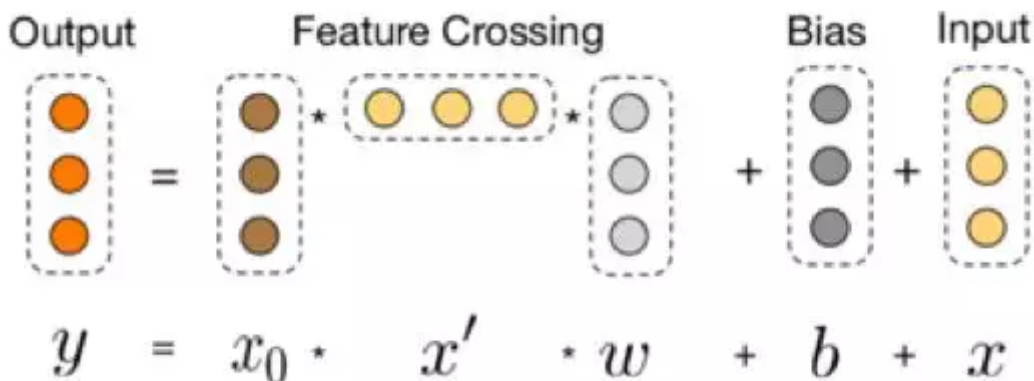
离散特征嵌入这个想法最初来自于 Mikolov 的 word2vec[5] 系列文章。最初解决的问题是词的独热表示过于稀疏，并且不同词之间的向量形式表示完全没有联系。具体思路在此不赘述，最终的实现是将一个上万维的词独热表示嵌入到了只有几百维的稠密向量中。而嵌入的本质其实是构建一张随机初始化的向量查找表，通过我们的训练目标做有监督学习来得到不同词在特定目标下，处于向量空间中的位置。

将词嵌入的思路推广到其它的离散特征处理中，我们可以用同样的方法将各种类别特征如“用户性别”、“城市”、“日期”嵌入到稠密的向量空间中。经过这样处理之后，自然就解决了原本 FM 遇到的特征稀疏问题。

高阶交叉特征

在广告场景下，特征交叉的组合与点击率是有显著相关的，例如，“USA”与“Thanksgiving”、“China”与“Chinese New Year”这样的关联特征，对用户的点击有着正向的影响。换句话说，来自“China”的用户很可能在“Chinese New Year”有大量的浏览、购买行为，而在“Thanksgiving”却不会有特别的消费行为。这种关联特征与label的正向相关性在实际问题中是普遍存在的，如“化妆品”类商品与“女性”，“球类运动配件”的商品与“男性”，“电影票”的商品与“电影”品类偏好等。因此，引入特征的组合是非常有意义的。看到这种需要，我们很容易就能想到 SVM 里面的多项式核函数，FM 就是在多项式核的基础上，优化了稀疏问题以及计算复杂度的产物。

而本文开发了一个新的算子，用另一种形式来得到交叉特征：

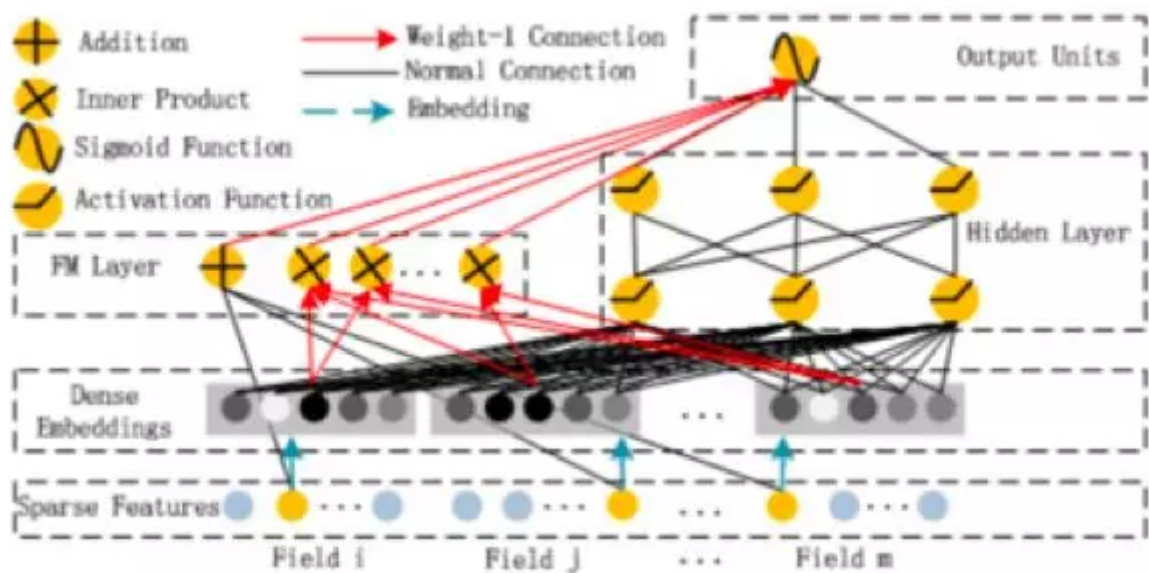


考虑 x_0 为输入的特征及第一层的输入， x 为第 L 层的输入，我们可以看到它的基本思路还是用矩阵乘法来实现特征的组合。

这是个递推形式算子，所以使用它很容易能得到高于二阶的交叉特征；并且该模型还用了残差的思想，解决网络性能退化的问题；此公式还有一个小的优化技巧，三矩阵相乘那个算子，用乘法结合律先计算后面两个矩阵的积，这样可以减少三分之一的计算复杂度。

DCN和同场景模型对比

在此我们略过 FM，因为它已经包含在 deepFM 内了。



可以看到 deepFM 进行了离散特征嵌入的操作，并且还将嵌入前的离散特征加入到了 FM 层；所以该网络可以看作是传统的 FM、离散特征嵌入之后的 FM 和基本 DNN 三个模型融合的结果。

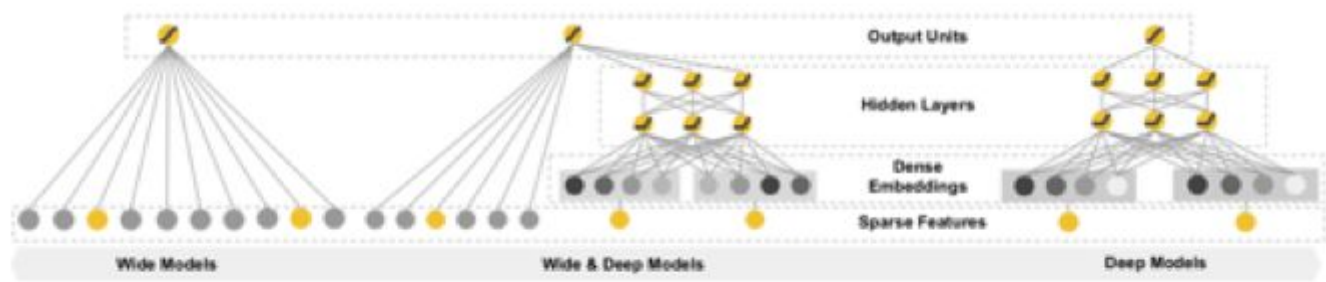


Figure 1: The spectrum of Wide & Deep models.

可以看到 wide & deep 的思路中，deep 部分的做法和 deepFM 是大相径庭的，关键的 wide 部分其实是离线的特征工程，根据业务场景提前完成了特征交叉等处理，该模型可以看作是 DNN 与离线特征模型的融合结果。

而从 DCN 的网络中我们可以发现，deep 部分网络除了使用离散嵌入特征外，还拼接了数值型特征；cross 部分网络直接完成了特征组合，对比 FM 层它可以学到更高阶的组合特征，对比 wide 网络它不需要做线下的特征工程。

小结

DCN 已经在云脑科技内部多个项目中得到了实践运用，包括信息流和应用商店等多个场景下的点击率预测任务。在深度定制的预测服务支持下，得益于该模型不需要复杂的特征工程，该模型能以很低的时延和较高的并发量运行。

实践经验表明，手工做的特征工程有时并不能很好的反映我们的预测目标，在实际完成的生产项目中我们更是发现，减少人工特征工程的使用反而会提高模型的效果，也是更加耐人寻味。

参考资料

- 1.1708.05123, Deep & Cross Network for Ad Click Predictions
- 2.1606.07792, Wide & Deep Learning for Recommender Systems
- 3.Factorization Machines
- 4.1703.04247, DeepFM: A Factorization-Machine based Neural Network for CTR Prediction
- 5.1310.4546, Distributed Representations of Words and Phrases and their Compositionality

作者介绍