

# DNN可以进行高阶特征交互，为什么Wide&Deep和DeepFM等模型仍然需要显式构造Wide部分？

阿泽的学习笔记 1周前

作者：王鸿伟

链接：<https://www.zhihu.com/question/364517083>

理论上来说DNN可以拟合任意函数，因此可以拟合低阶特征+高阶特征组合；但是在实际中并不会将原始特征输入全连接的DNN同时捕捉低阶+高阶特征组合，而是使用FM显式构造二阶项或者Wide&Deep的方式分别建模低阶和高阶特征。

使用原始特征作为DNN输入有什么缺点呢

同理可问：DNN可以进行高阶特征交互，为什么CNN模型仍然需要卷积操作来处理图片？DNN学不会卷积操作吗？

DNN可以进行高阶特征交互，为什么RNN模型仍然需要各种门控单元来处理序列？DNN学不会遗忘信息或者保留信息吗？

DNN可以进行高阶特征交互，为什么GNN模型仍然需要一层一层地从一个节点的邻居中融合信息呢？DNN学不会哪个邻居更重要吗？

题主也提到了“理论上来说DNN可以拟合任意函数”，关键就在于这个“理论上”。Universal Approximation Theorem只是存在性定理，并没有提供任何关于如何找到这样一个拟合完美的函数的保证。为了更好地学习出拟合函数，人们从三个方面在努力。一是根据具体的问题场景提出更好的模型，这些特定的模型会比较适合各自的场景，从而降低拟合难度，比如各类CNN，RNN，GNN等。二是提出更好的优化方法，使得学习拟合的过程更快速高效。

三是提取更好的特征，从而在源头上降低拟合的难度，提升性能的上限。

对于推荐系统这种非常偏工程的场景而言，特征提取做的好不好，对模型的效果影响巨大。夸张一点地说，特征工程决定了推荐系统最终性能的上限，各种推荐系统的模型只是在逼近这个上限而已。通俗一点地说，特征提得好，LR (Logistic Regression) 都能跑。

题主不要迷信DNN的拟合能力。在很多情况下，DNN都不能太好地拟合很多简单的函数。Wide&Deep和FM里手动构造二阶项，也只是想为DNN提供更多的“输入素材”，让DNN可以更好地发挥而已。这些二阶项能不能真的提高模型性能，并没有理论上的保证，因为这取决于具体的推荐系统场景中是否真的有很多这种二阶相关性（例如说“英语的用户喜欢看科幻电影”）。真实的推荐系统场景中的逻辑可能远比这种二阶相关性复杂，比如一个真实（但是大家并不知道）的逻辑可能是“说英语的男用户在周末喜欢看科幻电影”。那这就是一个四阶相关性，在Wide&Deep或者FM的框架下，你需要构造四阶项之后才能发现这种潜在的逻辑。但是构造四阶项的代价实在太大了，所以在Wide&Deep或者FM中只有二阶项，我们期望可以借此覆盖大多数的简单逻辑，而又不至于让模型过于复杂。

喜欢此内容的人还喜欢

数据科学竞赛：你从未见过的究极进化秘笈

kaggle竞赛宝典

---

ICDM2020 Top3方案分享

kaggle竞赛宝典

---

算法工程师研发技能表

机器学习实验室